UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Genetic variation and clinical variables contributing to Schizophrenia in a Founder Population from South Africa

Johannes Lodewikus Roos

MBChB, MMed(Psych), FCPsych(SA), MD

Published work submitted to the University of Pretoria for the Degree of Doctor of Science in Medicine

PROMOTER

Professor Doctor J Ker

Emeritus Professor, Department Internal Medicine

Faculty of Health Sciences

University of Pretoria

CO-PROMOTER

Professor Doctor WFP van Heerden

Head, Department of Oral Pathology and Oral Biology

Faculty of Health Sciences

University of Pretoria

January 2014

© University of Pretoria

CONTENT

Declaration	3
Acknowledgement	4
Dedication	5
Introduction	6 - 8
Summary	9 - 10
List of Published Submissions	11
- Group A Submissions	11 - 13
- Group B Submissions	14 - 15
Declaration of Publications	16
- Group A Publications	16
- Group B Publications	17 - 18
Abstract	19
- A: Genetic/Laboratory Findings	19 - 30
- B: Clinical Related Findings	30 - 35
Published Submissions	36 - 267

© University of Pretoria

DECLARATION

I, the undersigned, declare that the work contained in this presentation of publications is my original work, as set forth in the statements which precede the published articles.

.....

JL ROOS

I certify that on this ..... day of .....2014, Johannes Lodewikus Roos signed this declaration in my presence

.....

COMMISSIONER OF OATHS

© University of Pretoria

ACKNOWLEDGEMENT

I am deeply appreciative of the ongoing support and friendship of Prof M Karayiorgou and Prof J Gogos from the University of Columbia, New York. Their ongoing enthusiasm, hard work and research integrity has been a constant motivation to continue with this work.

The research participants who endured long interviews and not getting any immediate gratification for their inputs. My hope is that in the future we will have a better understanding of your illness and be able to treat you more effectively. Thank you for giving me a better understanding of a bad illness.

Thank you to the National Research Foundation (NRF) NRF for a rating and the continuous incentive grants over the past 5 years.

The support and encouragement of Professors Gerhard Lindeque, Head of School of Medicine, Prof James Ker, Emeritus Professor, Department of Internal Medicine, and Prof Willie van Heerden, Head of Department of Oral Pathology and Oral Biology, Faculty of Health Sciences, University of Pretoria

Me Erna Fourie for constant administrative support at all times.

Mr Stephen James for assisting with the printing of the thesis.

Nursing sister Ria van Wyk and Hester van den Berg for gaenealogical tracing and obtaining blood samples.

My wife, Ina Roos, who is a constant motivation in my life and career. Thank you for the ongoing emotional and editorial support.

© University of Pretoria

DEDICATION

To my wife Ina, who has been there for me and the patients who participated in these studies and who gave me some insight into schizophrenia.

© University of Pretoria

Introduction

The author graduated from the Medical School at UP (University of Pretoria, South Africa) in 1975 and specialized at that institution in Psychiatry. He obtained the MMed(Psych) degree in 1982. Later he obtained a fellowship by peer review from the College of Medicine of SA in the College of Psychiatrists in 2002.

Since his early years of study he always had an interest in schizophrenia. He published review papers in Geneeskunde 'Medicine' on schizophrenia whilst specialising. He stayed on at the University of Pretoria and Department of Psychiatry. As a specialist he initially worked in Community Psychiatry and closely involved with a support group for schizophrenia patients in Pretoria (YANA - You Are Not Alone).

He completed his MD thesis with Prof W Bodemer as supervisor in 1990. The title of this thesis was: "An investigation on causes of suicide in patients suffering from schizophrenia". His interest in this topic was triggered whilst working in the community and at Weskoppies Psychiatric Hospital (at that stage a 1400 bed hospital) in Pretoria, South Africa. A high number of patients he was managing suffered from schizophrenia. A subsequent paper was published in 1992 in the South African Medical Journal. This study was included in an international systematic review of risk factors for suicide in patients suffering from schizophrenia in 2005, by Hawton K *et al.* The control group of the Roos *et al.* 1992 study was followed up after 20 years and was reported on in two subsequent publications in 2009 in the SA Journal of Psychiatry (SAJP). SAJP is the SA Journal of Psychiatry affiliated to the SA Medical Journal and mouthpiece of the SA Society of Psychiatrists. During his early years at the department he was also involved in more than 15 randomized controlled trials on the treatment of schizophrenia. He was co-investigator in these trials and had formal training from pharmaceutical companies to do these trials according to Good Clinical Practice.

In 1997 whilst the author was working at the Department of Psychiatry, University of Pretoria as associate professor and principal specialist, Prof Maria Karayiorgou from the Neurogenetics Laboratory at the Rockefeller University in New York visited the department to start a collaborative study on the Genetics of Schizophrenia in an Afrikaner Founder Population. The hospital

© University of Pretoria

admitted Afrikaner schizophrenia patients and the author's first language was Afrikaans. The author was appointed by the Head of Department to be the principal clinical investigator of this collaborative study. He has been doing this research since 1997 and at present it is still ongoing. Initially the University of Cape Town; Department of Psychiatry was also involved with recruitment of patients for a few years. When the author was appointed head of the department in 2001 he then emphasized the fact that research will be a prominent mission in the department.

The 30 papers presented for this thesis is forthcoming from this ongoing collaborative study for the past 15 years which started in 1997. Prof Maria Karayiorgou who heads the Human Neurogenetics laboratory was initially at Rockefeller University and then moved to Columbia University.

The author's role and involvement in this collaborative genetic research of schizophrenia in the Afrikaner founder population is as follows: The goal of the research is to identify genes associated with this risk for schizophrenia in the Afrikaner population. Tackling the genetics of schizophrenia is a problem of paramount importance and unparalleled difficulty. It requires multi-disciplinary teams of dedicated researchers with complementary expertise. The author is the principal clinical investigator in this research. More than 850 patients have been recruited. Of the 850 Diagnostic Interview for Genetic Studies (DIGS) interviews and summary reports completed, 764 were done by the author and 86 by Prof HW Pretorius. These more than 850 probands that have been collected include triads, multiple affected family members and individual cases. The author took great care to personally recruit and evaluate them carefully and in great detail. The results of this work are obvious in several of our studies. For example, the care we took to obtain detailed family history allowed for the accurate distinction between sporadic and familial cases in the CNV work, which turned out to be an important variable. In the early stage of this study there was specific emphasis on the genealogical tracings of recruited patients. To aid the genealogist, information had to be obtained from the National Archives in Pretoria. A group of students were supervised by the author to gain this information. Diagnostic accuracy is of prime importance in schizophrenia genetic research. The same is true with the distinction between the specific mood subtypes in all our schizoaffective cases, since the bipolar-type cases turn out to be important for the linkage on chromosome 13. The author's involvement in the project as summarized in this thesis is deep, intense, and meaningful. He has taken the lead in the clinical studies, from conception to study design, execution, and data analysis and manuscript preparation. In the

© University of Pretoria

Department of Psychiatry at UP this collaborative study platform formed an important source for research training of registrars completing their MMed(Psych) studies. Several of these registrars won research awards from the Faculty of Health Sciences at UP. The theme of the author's clinical studies has been on characterizing variables that influence genetic vulnerability or endophenotypes that could influence genetic liability. He has also been deeply involved with all genetic studies and a key contributor to them.

In September 2012 the author received a special acknowledgement from the SA Society of Psychiatrists for his contribution to the discipline of psychiatry and society and for a significant contribution to SA Psychiatry with reference to a national expert in schizophrenia research. Awards were also received for articles published in the SA Psychiatry Review and for the best clinical scientific article in Afrikaans by the ATKV in 2012.

Prof Maria Karayiorgou on the website of the Neurogenetics Laboratory at Columbia University (schizophreniaresearch.columbia.edu/content/research-program-o) refers to her partnerships during her research career and says: "Another valuable early research partnership is the one with Louw Roos, who has been the cornerstone of the clinical recruitment component of our research program".

Thirty publications are submitted for the purpose of this thesis. Sixteen of the papers deal with the issue of genetic/laboratory findings of the genetics of schizophrenia in a founder population from SA (Group A papers). The other 14 articles pertain to clinical variables regarding the genetics of schizophrenia in this population (Group B papers).

© University of Pretoria

SUMMARY

Thirty publications are submitted. They deal with findings of the genetic architecture of schizophrenia in an Afrikaner founder population and clinical related variables pertaining to this population.

The initial research findings supported the appropriateness of the Afrikaner population for mapping complex traits using both linkage and linkage disequilibrium (LD) approaches. Basic sample descriptors and cardinal symptoms of schizophrenia in the US and South African populations were equivalent. It was concluded that the results from our genetic study of schizophrenia in the Afrikaner sample will be applicable to other populations.

It was found that early non-psychotic childhood deviance (in the first ten years of life) distinguished a distinct subtype of schizophrenia patient, and that the form of early deviance manifested, was meaningful linked to later disease outcome; and that it may be a possible endophenotypic marker in schizophrenia but not in bipolar disorder.

Schizophrenia genetic research used linkage analysis, association studies and exome sequencing studies as it became available in the last few years. We addressed the role of the individual genes from the 22q11 locus (prototype CNV described in schizophrenia). Systematic screening of the 26 genes residing in this locus identified PRODH2, ZDHHC8, NOGO Receptor 1 (RTN4R) gene as contributing to schizophrenia risk associated with this region. Linkage genome-wide scans, using both less dense (10cM) and more dense scans (2cM), identified a locus on chromosome 1 and 13. Recent fine mapping on chromosome 13q32-34 and brain expression analysis implicates MYO16 in schizophrenia (not included in the 30 publications). For the first time a probound with a uniparental disomy (UPD) of the entire chromosome 1, was identified, which further support the involvement of chromosome 1 in schizophrenia.

We confirmed the previous reported rate of 2% frequency of 22q11 deletions in adult schizophrenia Afrikaner patients and provided a two-stage screening protocol to identify these patients in clinical practice.

© University of Pretoria

As more patients were recruited for this study certain comorbid conditions became obvious including marijuana use/abuse and obsessive compulsive disorder (OCD) and obsessive compulsive symptoms (OCS).

Approximately half of the male patients and a quarter of the female patients used or abused marijuana. Male users of marijuana with prominent early non-psychotic deviant behaviour in the first 10 years of life had the lowest mean age of criteria onset (18.4 years) with a poor prognosis. The prevalence of OCD/OCS in this population was 13.2% and differs from other ethnic groups in South Africa, and was associated with significant psychopathology and poor prognosis.

As the research progressed the emphasis has changed from familial cases with the disease to sporadic cases (non-familial). We offered the first clear view of the genetic landscape of schizophrenia. We found that rare de novo structural mutations at many different loci are significantly enriched and contribute to schizophrenia vulnerability in sporadic cases with the disease. We also demonstrated that genes contribute to familial schizophrenia, while new mutations are less prominent.

The study of schizophrenia in the Afrikaner founder population has helped to clear the view of the genetic landscape of schizophrenia.

© University of Pretoria

List of published submissions

GROUP A - GENETIC / LABORATORY FINDINGS:

A1. Liu H., Abecasis GR., Heath SC., Knowles A., Demars S., Chen YJ., Roos JL., Rapaport JL., Gogos JA., Karayiorgou M. Genetic variation in the 22q11 locus and susceptibility to schizophrenia. Proceedings of the National Academy of Sciences of the United States of America. 99(26): 16859-64, 2002

A2. Liu H., Heath SC., Sobin C., Roos JL., Galhe BL., Blundell ML., Lenane M., Robertson B., Wysman EM., Papaport JL., Gogos JA., Karayiorgou M. Genetic variation at the 22q11 PRODH2/DGCR6 locus presents an unusual pattern and increases susceptibility to schizophrenia. Proceedings of the National Academy of Sciences of the United States of America. 2002:99(6): 3717-3722

A3. Hall D., Wijsman EM., Roos JL., Gogos JA., Karayiorgou M. Extended intermarker linkage disequilibrium in the Afrikaners. Genome Research. 12(6): 956-61, 2002

A4. Karayiorgou M., Rottington M., Abecasis GR., Pretorius HW., Robertson E., Kaliski S., Lay S., Sobin C., Maller N., Lundy SL., Blundell ML., Gogos JA., Roos JL. Phenotypic characterization and genealogical tracing in an Afrikaner schizophrenia database. American Journal of Medical Genetics. 2004; 124B(1): 20-28

A5. Abecasis GR., Burt R., Hall D., Bochum S., Doheny KF., Lundy LS., Torrington M., Roos JL., Gogos JA., Karayiorgou M. Genome wide scan in schizophrenia families from the founder population of Afrikaners reveals evidence for linkage and uniparental disomy on chromosome 1. American Journal of Human Genetics. 2004; 74:403-417

A6. Wang H., Bedoya G., Ospina J., Ruiz-Linares A., Macedo A., Palha JA., Heutink P., Aulchenko Y., Oostra B., Van Duijn C., Jarvelin MR., Lin CH., Varilo T., Peddle L., Rahman P., Piras G., Monne M., Peltonan., Servier S., chen Y., Freimer N., Sabatte C., Karayiorgou M., Roos JL., Pretorius HW. Linkage disequilibrium and haplotype homozygosity in population samples genotyped at a high marker density. Human Heredity 2006; 62(4): 175-189

© University of Pretoria

A7. Servier S., De Young J., Karayiourgou M., Roos JL., Pretorius HW., Bedoyta G., Ospina J., Ruiz-Linares A., Macedo A., Palha JA., Heutink P., Aulchenko Y., Oostra B., Van Duijn C., Jarvelin M-R., Varilo T., Sabatri C., Collins A., Freimer N. Magnitude and distribution of linkage disequilibrium in population isolates and implication for genome wide association studies. Nature Genetics. 2006; 38(5): 556-60

A8. Servier S., Macedo a., Palha JA., Heutink P., Aulchenko Y., Oostra B., Van Duijn C., Jarvelin MR., Varilo T., Peddle L., Rahman P., Sabatti C., Piras G., Monne M., Peltonen L., Freimer N., Karayiorgou M., Roos JL., Pretorius HW., Bodoya G., Ospina J., Riuz-Linares A. Tag SNPs chosen from hapMap perform well in several population isolates. Genetic Epidemiology. 2007; 31(3)189-194

A9. Hsu R., Woodroffe A., Lai W., Cook MN., Mukai J., Dunning JP., Swanson JF., Roos JL., Abecasis GR., Karayiorgou M., Gogos JA. Nogo Receptor 1 (RTN4R) as a candidate gene for schizophrenia: Analysis using human and mouse genetic approaches. PLoS ONE. 2007; 140-144

A10. Francks C., Maegawa S., Lauren J., Abrahams BS., Velayos-Baeza A., Medland SE., Colella S., Groszer M., McAuley EZ., Caffrey TM., Timmusk T., Pruunsild P., Koppel I., Lind PA., Matsemoto-Itaba N., Nocod J., Ziong L., Joober R., Ernard W., Krinsky B., Roos JL., Nanba E., Muglia AP., Fisher SE., Wade-Martins R., Rouleau GA., Stein JF., Karayiorgou M., Geschwind DH., Ragoussis J., Kendler KS., Airaksinen MS., Oshimura M., DeLisi LE., Monaco AP. LRRTM1 on chromosome 2p12 is a maternally suppressed gene that is associated paternally with handedness and schizophrenia. Molecular Psychiatry. 2007; 12(12):1129-39

A11. Francks C., Caffrey TM., Timmusk T., Pruunslid P., Lind PA., Matsomoto-Itaba N., Nocod J., Xiong L., joober R., Enard W., Maegawa S., Krinsiky B., Nanba E., Richardson AJ., Riley BP., Martin NG., Strittmatter SM., Moller HJ., Rugesco D., St Clair D., Muglai P., Lauren J., Roos JL., Fisher SE., Wade-Martens R., Rouleau GA., Stein JF., Karayiorgou M., Geschwind DH., Ragoussis J., Kendler KS., Airaksinen MS., Abrahams BS., Oshimura M., Delisi LE., Monaco AP., Velayos-Baeza A., Medland SE., Colleba S., Groszer M., McAuley EZ. LRRTM1 protein is located in the endoplasmic reticulum (ER) in mammalian cells. Molecular Psychiatry. 2007; 12(12):1057

© University of Pretoria

A12 Xu B., Roos JL., Levy S., Janse van Rensburg E., Gogos JA., Karayiorgou M. Strong association of de novo copy number mutations with sporadic schizophrenia. Nature Genetics. 2008;7(40):880-885

A13. MYM Ng., DF Levinson., SV Faraone., BK Suarez., LE De Lisi., T ARinami., B Riley., T Paunio., AE Pulver., Irmansyah., PA Holmans., M Escamilla., BD Wildenauer., NM Williams., C Laurent., BJ Mowry., LM Brzustowicz., M Maziade., P Skiar., DL Garver., GR Abecasis.,, B Lerer., MD Fallin., HMD Gurling., PV Gejman., E Lindholm., HW Moises. W Beyerlye, EM Wijsman., P Forabossco., MT Tsuang. MGH Hwu., Y Okazaki., KS Kendler, B Wormley., A Fanous., D Walsh., FA O'Neill., L Peltonen., G NEstadt., VK Lasseta., KY Liang., GM Papadimitriou., DG Dikeos. SG Schwab., MJ Owen., MC O'Donovan., N Norton., E Hare, H Raventos., H Nicolini., M Albus., W Maier., VL Nimgaonkar. L Terenius., J Mallet., M Jay., S Godard., D Nertney., M Alexander., RR Crowe., JM Silverman., AS Bassett.,, MA Roy., C Merette., CN Pato., MT Pato., J Louw Roos., Y Kohn, D Amann-Zakenstein., G Kalski., A McQuillin., D Curtis., J Brynjolfson., T Sigmundsson., H Petursson., AR Sanders., J Duan., E Jazin., M Myles-Worsley., M Karayiorgou., CM Lewis. Meta-analysis of genome-wide linkage studies of schizophrenia. Molecular Psychiatry. 2009; 14(8):774-785

A14. Bin Xu., Abigail Woodroffe., Laura Rodriquez-Murillo., J Louw Roos., Elizabeth van Rensburg., Goncalo Abecasis., Joseph Gogos., Maria Karayiourgou. Elucidating the genetic architecture of familial schizophrenia using rare copy number variant and linkage scans. Proceedings of the National Academy of Sciences of the United States of America. 2009; 106(39)16746-16751

A15. Xu B., Roos JL., Dexheimer P., Boone B., Plummer B., Levy S., Gogos JA., Karayiorgou M. Exome sequencing supports a de novo mutational paradigm for schizophrenia. Nature Genetics. 2011; 43(9):864-868

A16. Xu B., Ionita-Laza I., Roos JL., Boone B., Woodrick S., Sun Y., Levy S., Gogos JA., Karayiorgou M. De novo gene mutations highlight pattern of genetic and neural complexity in schizophrenia. Nature Genetics 2012;44(12):1356-1370

© University of Pretoria

GROUP B - CLINICAL RELATED FINDINGS:

B1. Sobin C., Roos JL., Pretorius HW., Lundy LS., Karayiorgou M. A comparison study of early non-psychotic deviant behaviour in Afrikaner and US patients with schizophrenia and schizoaffective disorder. Psychiatry Research. 117(2): 113-25, 2003

B2. Roos JL., Honey EM., Pretorius HW., Sobin C., Karayiorgou M. Velocardiofacial syndrome case report: Is this a homogeneous genetic subtype of schizophrenia? SA Family Practice. 45(2): 20-23, 2003

B3. Wiehahn GJ., Bosch GP., Du Preez RR., Pretorius HW., Karayiorgou M., Roos JL. Assessment of the frequency of the 22q11 deletion in Afrikaner schizophrenic patients. The American Journal of Medical Genetics. Part B (Neuropsychiatric Genetics). 2004; 129B: 20-22

B4. Roos JL., Pretorius HW., Karayiorgou M. Multiple affected Afrikaner families in a schizophrenia genetic study: environmental risk factors in interaction with genotypes. SA Psychiatry Review. 2004; 7(4): 10-14

B5. Scholtz MC., Janse van Rensburg MS., Roos JL., Pretorius HW., Karayiorgou M., Levin JB. Early non-psychotic deviant behaviour as an endophenotypic marker in bipolar disorder, schizoaffective disorder and schizophrenia. SA Psychiatry Review. 2005; 8:153-159

B6. Roos JL., Pretorius HW., Karayiorgou M., Boraine H. Cannabis and other variables affecting age at onset in a schizophrenia founder population. SA Psychiatry Review. 2006; 9:99-103

B7. Seedat F., Roos JL., Pretorius HW. Prevalence and clinical characteristics of obsessive-compulsive disorder and obsessive-compulsive symptoms in Afrikaner schizophrenia and schizoaffective disorder patients. African Journal of Psychiatry. 2007; 10:219-24

B8. Roos JL. Die Afrikaner en Skisofrenie: 'n Genetiese studie. Die SA Tydskrif vir Natuurwetenskap en Tegnologie. 2008;27(2):73-87

B9. Roos JL., Pretorius HW., Karayiorgou M. Clinical characteristics of an Afrikaner Founder Population recruited for a schizophrenia genetic study. The Year in Human and Medical Genetics. 2009: Annals of the New York Academy of Sciences. 1151:85-101

© University of Pretoria

B10. Maydell RJ., Van der Walt C., Roos JL., Scribante L., Ladikos A. Clinical characteristics and premorbid variables in childhood onset schizophrenia: a descriptive study of 12 cases from a schizophrenia founder population. African Journal of Psychiatry. 2009;12(2):144-148

B11. Lombard M., Roos JL. Velocardiofacial syndrome - A syndrome with many faces. SA Journal of Psychiatry 2011; 17(3):82-88

B12. Roos JL., Lombard M., Ladikos A. Can we close the barn door before the horses get out? A case study of high genetic loading and subsequent development of psychosis. African Journal of Psychiatry 2011; 14(3):242-244

B13. Roos JL. Genetics of schizophrenia: Communicating scientific findings in the medical setting. African Journal of Psychiatry 2011; 14(2):105-111

B14. Roos JL. Nuwe bevindings oor die genetiese argitektuur van skisofrenie: Die bydrae van genetiese studies onder Afrikaners (New findings on the Genetic architecture of schizophrenia: The contribution of genetic studies in Afrikaners). Die Suid-Afrikaanse Tydskrif vir Natuurwetenskappe en Tegnologie. 2014; 33(1):1-9.

© University of Pretoria

DECLARATION

Group A Publications: Genetic / Laboratory Findings

A1. I was the only clinical psychiatrist from SA as co-author and recruited and clinically evaluated most patients from the Pretoria site.

A2. I was one of two clinical psychiatrists from SA who is a co-author. I recruited and clinically evaluated most patients from the Pretoria site.

A3. I was the only clinical psychiatrist from SA included as co-author. I recruited and clinically evaluated most patients from the Pretoria site.

A4. Four clinical psychiatrists from SA is co-authors. I recruited and clinically evaluated most patients from the Pretoria site. I also supervised data collection for the genealogical tracing from the National Archives in Pretoria.

A5. As co-author of this paper, I was the only clinical psychiatrist from SA, who recruited and clinically evaluated most patients from the Pretoria site. I recruited the patients with the uniparental disomy on chromosome 1.

A6, A7 & A8. I was one of the two co-authors. As clinical psychiatrist I recruited and clinically evaluated most patients from Pretoria

A9, A10 & A11. I was the only clinical psychiatrist from SA as co-author. I recruited and clinically evaluated most of the patients from Pretoria that were included in this study.

A13. This paper is a meta-analysis of genome-wide linkage studies of schizophrenia. I was co-author, clinical psychiatrist from SA who recruited and clinically evaluated most patients from Pretoria.

A12, A14, A15 & A16. In these four recent papers we offered the first clear view of the genetic landscape of schizophrenia. In all four papers I was the only clinical psychiatrist from SA as co-author. I was involved with the recruitment and clinical evaluation of most patients from Pretoria, SA.

© University of Pretoria

Group B Publications: Clinical Related Findings:

B1 I was the co-author with Prof HW Pretorius as clinical psychiatrist. I supervised and collected data on the SA side of the comparison study.

B2. I was the first author - I recruited the patients and wrote the case report with inputs from the co-authors.

B3. The first author was a registrar in the Department of Psychiatry at UP. I was the research supervisor and involved in the design of the study, wrote up the paper with co-authors and was the corresponding author of this publication.

B4. I was the principal clinical investigator of this project that led to the publication. I wrote the publication with input from the other two co-authors. I was the corresponding author.

B5. The first author was a registrar in the Department of Psychiatry at UP. I was the research supervisor and involved in the design of the study and wrote up the paper with the co-authors.

B6. I designed the study, collected most of the data and wrote the paper with contributions from the co-author. I was the corresponding author.

B7. The first author was a registrar in the Department of Psychiatry at UP. I was the research supervisor and involved in the design of the study and wrote up the paper with co-authors.

B8. I was the only author of the paper in Afrikaans. It gives an updated overview of the genetic findings in schizophrenia in the Afrikaner founder population

B9. The focus of this paper was initiated by myself. I wrote the paper with contributions from the co-authors. I was the first and corresponding author of this paper.

B10. The first author was a final year medical student at the Department of Psychiatry at UP. I was the research supervisor, helped with design of the study and the collection of the data. I was involved with writing up the paper with the co-authors.

B11. The first author was a registrar in the Department of Psychiatry at UP. I was the research supervisor and recruited the VCFS patients included in the study. I wrote up the paper with the first author.

© University of Pretoria

B12 & B13. I was the first author. I designed the concept of the papers and wrote the papers to communicate some of the scientific findings in the medical setting.

B14. I was the first and corresponding author of this paper in Afrikaans. This paper is an overview of paper A12, A14, A15 & A16 regarding the new findings on the genetic architecture of schizophrenia, emphasizing the contribution of the genetic studies in Afrikaners.

© University of Pretoria

Abstract

A collaborative genetic research of schizophrenia in the Afrikaner founder population has been ongoing since 1997 between the Human Genetics Laboratory first at Rockefeller University and later at Columbia University in New York and the Department of Psychiatry, University of Pretoria in South Africa. The author is the head of this department and the principal clinical investigator of this study. Prof Maria Karayiorgou heads the Human Neurogenetics laboratory at Columbia University in New York.

The 30 papers presented for this thesis is forthcoming from this ongoing collaborative study for the past 15 years which started in 1997. More than 850 patients have been recruited for this study so far. Of the 850 Diagnostic Interview for Genetic Studies (DIGS) interviews and summary reports completed, 764 were done by the author.

The accomplished results published thus far cover two broad fields: Genetic/Laboratory findings and clinical related findings.

A. Genetic/Laboratory findings:

The 22q11 locus is actually the prototype CNV described in schizophrenia, as early as 1995. This is a locus previously implicated in schizophrenia susceptibility through deletion in 2% of schizophrenia patients. A set of important studies has studied the Afrikaner sample to address the role of individual genes from the 22q11 locus.

The first publication from our systematic screening of genes residing in the critical 22q11 locus appeared in 2002 (paper A1). Individuals with a deleted part of chromosome 22, called 22q11, are at high risk for developing schizophrenia. Their risk is the highest known risk associated with a genetic defect, and is comparable to the risk of an identical twin of a schizophrenic patient or of a child born to two schizophrenic parents. Previous studies revealed that 25 to 30 percent of people with microdeletions in 22q11 develop schizophrenia. The case for the role of 22q11 microdeletions in

© University of Pretoria

schizophrenia was strengthened by additional studies showing that these microdeletions occur in up to 2% of adult schizophrenic patients and up to 6% of patients with childhood-onset schizophrenia. Childhood-onset schizophrenia is a particular aggressive form of the disease that causes symptoms in children by age 13. We hypothesized that while deletions of genes from chromosome 22q11 might account for only a small portion of schizophrenia cases in the general population, other types of more subtle genetic flaws such as common variants of individual genes from the 22q11 region may make a larger contribution in non-deleted schizophrenic patients in the wider population. In Paper A2 we examined 22q11 in 242 non-deleted people affected by schizophrenia and their parents, as well as people not diagnosed with schizophrenia, focusing on variations in 13 genes in that region. We found variants in two genes, PRODH2 and DGCR6, enriched in patients with schizophrenia as compared to unaffected individuals who served as normal controls for the study. Two additional lines of evidence showed that of the two genes, PRODH2 might play the more important role. By analysing the publicly available sequence of chromosome 22, we discovered a linked "pseudogene" of PRODH2. The pseudogene, which is 97 percent similar to the true gene in its DNA sequence, did not produced a functional protein. We discovered that the PRODH2 pseudogene served to introduce in the gene missense mutations that confer susceptibility through a process called gene conversion. Pseudogene-like variations of PRODH2 might prevent synthesis of a fully functional enzyme and seemed important for increasing susceptibility to schizophrenia because they were present either exclusively or in higher frequencies in people with schizophrenia as compared to controls. The sequencing of the human genome has revealed thousands of pseudogenes as a possible constant and renewable source of variation that can be effectively transferred to linked susceptibility genes via gene conversion. PRODH2 encodes for a brain enzyme called proline dehydrogenase, which is involved in neuron signal transmission as well as cell death.

The follow up of paper A1 we published paper A2 in the same year that reported on the genetic variation in the 22q11 locus and susceptibility to schizophrenia.

An increased prevalence of microdeletions at the 22q11 locus has been reported in samples of patients with schizophrenia. 22q11 microdeletion represents the highest known genetic risk factor for the development of schizophrenia, second only to that of the monozygotic co-twin of an affected individual or the offspring of two schizophrenic parents. It is therefore clear

© University of Pretoria

that a schizophrenia susceptibility locus maps to chromosome 22q11. In light of evidence for suggestive linkage for schizophrenia in this region, we hypothesized that, whereas deletions on chromosome 22q11 may account for only a small proportion of schizophrenia cases in the general population (up to ~ 2%), nondeletion variants of individual genes within the 22q11 region may make a larger contribution to susceptibility to schizophrenia in the wider population. By studying a dense collection of markers (average one single nucleotide polymorphism / 20 kb over 1.5 Mb) in the vicinity of the 22q11 locus, in both family- and population - based samples, we presented here results consistent with this assumption. Moreover, our results were consistent with contribution from more than one gene to the strikingly increased disease risk associated with the locus. Finer-scale haplotype mapping had identified two sub regions within the 1.5 Mb locus that are likely to harbour candidate schizophrenia genes.

The combination of the study of Liu *et al.* (paper A1) and the present study (paper A2) represented a comprehensive application of LD mapping across the 22q11 susceptibility locus in relation to schizophrenia. An implication of these studies was that more than one gene from this region might contribute to schizophrenia susceptibility. Functional variants of these genes might independently or synergistically exert modest increases in the disease risk in nondeleted patients. In 22q11 microdeletion carriers, drastic reduction of dosage to half in more than one gene might result in the dramatic 25- to 30-fold increase in disease risk observed. Alternatively, genes in the identified proximal and distal subregions might contribute only partially to the 25- to 30-fold increase in the risk associated with this locus, the largest part of it being due to an unidentified gene that lacks common functional variants and was variably mutated in a small portion of nondeleted schizophrenic patients. Such rare and possibly highly penetrant mutations would be invisible to our present experimental strategy, which focussed on relatively common variants. Nevertheless, our current analysis revealed a limited number of candidate genes (including two membrane-associated proteins, K1AAA1292 and NOGO-R, highly expressed in brain regions implicated in schizophrenia) which needed to be tested in additional samples and whose biological functions needed to be explored in more detail by a generation of animal models.

In 2002 we reported on extended intermarker linkage disequilibrium in the Afrikaners. (Paper A3). In this study we conducted an investigation of the background level of linkage disequilibrium (LD) in the Afrikaner population to evaluate the appropriateness of this genetic isolate for mapping complex traits. We analysed intermarker LD in 62 nuclear families using microsatellite markers

© University of Pretoria

covering extended chromosomal regions. The markers were selected to allow the first direct comparison of long-range LD in the Afrikaners to LD in other demographic groups. Using several statistical measures, we found significant evidence for LD in the Afrikaners extending remarkably over a 6-cM range. In contrast, LD decayed significantly beyond 3-cM distances in the other founder and outbred populations examined. This study strongly supported the appropriateness of the Afrikaner population for genome-wide scans that exploit LD to map common, multigenic disorders.

Paper A4 we reported on the phenotypic characterization and genealogical tracing in an Afrikaner schizophrenia database. Here, we examined the potential to conduct genealogical research in the Afrikaner founder population and presented the initial results indicating that accurate genealogical tracing for up to 17 generations was feasible. We also examined clinical similarities of schizophrenia patients diagnosed in South Africa and those diagnosed in other heterogeneous populations, specifically the US. We found that, with regard to basic sample descriptors and cardinal symptoms of disease, the two populations were equivalent. It was therefore likely that results from our genetic study of schizophrenia would be applicable to other populations. Based on the results presented here, the history and current size of the population, as well as our previous analysis addressing the extent of background linkage disequilibrium (LD) in the Afrikaners; we concluded that the Afrikaner population is likely to be an appropriate founder population to map genes for schizophrenia using both linkage and LD approaches.

Also in 2004 a report on genomewide scan in families with schizophrenia in the Afrikaner founder population was published. (Paper A5). Findings revealed evidence for linkage and uniparental disomy on chromosome 1. Here we reported on our initial genetic linkage studies of schizophrenia in the genetically isolated population of the Afrikaners from South Africa. A 10-cM genomewide scan was performed in 143 small families, 34 of which were informative for linkage. Using both nonparametric and parametric linkage analysis, we obtained evidence for a small number of disease loci on chromosomes 1, 9 and 13. These results suggested that few genes of substantial effect exist for schizophrenia in the Afrikaner population, consistent with our previous genealogical tracing studies. The locus on chromosome 1 reached genomewide significance levels (nonparametric LOD score of 3.30 at marker D1S1612) corresponding to an empirical p-value of 0.012) and represented a novel susceptibility locus for schizophrenia. In addition to providing evidence for linkage for chromosome 1, we also identified a proband with a uniparental disomy (UPD) of the entire chromosome 1. This is the

© University of Pretoria

first time a UPD had been described in a patient with schizophrenia, lending further support to involvement of chromosome 1 in schizophrenia susceptibility in the Afrikaners.

We also participated in a multicentre collaboration where the LD in large samples (200 individuals) from each of 11 well-described population isolates and an outbred European-derived sample. SNP markers spaced across chromosome 22 were assayed. The Afrikaners was one of the eleven population isolates. It was concluded in a letter to Nature Genetics that these populations offer substantial advantages for genome-wide association mapping (paper A6, A7 & A8). At that stage I was a member of the international collaborative group on isolated populations.

As a follow-up to the first two publications in 2002 there was a report on NOGO Receptor 1 (RTN4R) gene that maps to the 22q11.2 schizophrenia locus and is thus a strong functional and positional candidate gene. NOGO Receptor 1 (RTN4R) regulates axonal growth, as well as axon regeneration after injury. The analysis made use of human and mouse genetic approaches (paper A9).

We also participated in a study that was reported on in Molecular Psychiatry (paper A10 & A11) and showed that LRRTM1 on chromosome 2p12 is a gene that is associated paternally with handedness and schizophrenia. SNPs were genotyped in four family samples of White European descent that included individuals with schizophrenia or schizoaffective disorder. These comprised the following samples:

- Subset of the HY/Oxford sample (226 families)
- An Irish 'High Density' sample (236 families)
- A sample collected in Montreal (124 families)
- An Afrikaner sample (416 families)

In 2009 the Afrikaner population isolate was included in a meta-analysis of genome wide linkage studies of schizophrenia (paper A13). A genome scan

© University of Pretoria

meta-analysis (GSMA) was carried out on 32 independent genome-wide linkage scan analyses that included 3255 pedigrees with 7413 genotyped cases affected with schizophrenia or related disorders. The primary GSMA divided the autosomes into 120 bins, rank-ordered the bins within each study according to the most positive linkage result in each bin, summed these ranks (weighted for study size) for each bin across studies and determined the empirical probability of a given summed rank (P_{SR}) by simulation. Suggestive evidence for linkage was observed in two single bins, on chromosomes 5q (142-168 Mb) and 2q (103-134 Mb). Genome-wide evidence for linkage was detected on chromosome 2q (119-152 Mb) when bin boundaries were shifted to the middle of the previous bins. The primary analysis met empirical criteria for 'aggregate' genome-wide significance, indicating that some or all of 10 bins are likely to contain loci linked to schizophrenia, including regions of chromosomes 1, 2q, 3q, 4q, 5q, 8p and 10q. In a secondary analysis of 22 studies of European-ancestry samples, suggestive evidence for linkage was observed on chromosome 8p (16-33 Mb). Although the newer genome-wide association methodology has greater power to detect weak associations to single common DNA sequence variants, linkage analysis could detect diverse genetic effects that segregate in families, including multiple rare variants within one locus or several weakly associated loci in the same region. Therefore, the regions supported by this meta-analysis deserve close attention in future studies.

In four recent papers (paper A12, A14, A15 & A16) we offered the first clear view of the genetic landscape of schizophrenia. In paper A12 we found a strong association of de novo copy number mutations with sporadic schizophrenia. Here we examined the possibility that rare de novo copy number (CN) mutations with relatively high penetrance contribute to the genetic component of schizophrenia. We carried out a whole genome scan and implemented a number of steps for finding and confirming CN mutations. Confirmed de novo mutations were significantly associated with schizophrenia ($p=0.00078$) and collectively ~ 8 times more frequent in sporadic (but not familial) cases with schizophrenia than in unaffected controls. In comparison, rare inherited CN mutations were only modestly enriched in sporadic cases. Our results suggested that rare de novo germ line mutations contribute to schizophrenia vulnerability in sporadic cases and that rare genetic lesion at many different loci could account, at least in part, for the genetic heterogeneity of this disease.

© University of Pretoria

Irrespective of the precise pathogenic models, our findings had two noteworthy implications for our understanding of the genetic basis of schizophrenia.

1. They strongly suggested that rare de novo germ line mutations contributed to schizophrenia vulnerability in sporadic cases with the disease, thereby extending and complementing our initial observation at the 22q11 locus. Because of the limited resolution of our genome microarray scan and its inability to identify balanced abnormalities, the rate of new CN mutations that we detected was likely an underestimate. Such new mutations could explain, at least in part, how schizophrenia persists in the population despite the low fecundity of affected individuals.

2. Our findings suggest that de novo genetic lesions at many different loci could contribute to schizophrenia, and this heterogeneity might account for the difficulties in finding genetic variants with a major effect on disease risk. Heterogeneity might be more pronounced among sporadic cases, although a portion of de novo events (depending on disease severity or age at onset) was likely to be inherited, thus contributing to the genetic heterogeneity among familial cases as well.

In a further paper (paper A14) the genetic architecture of familial schizophrenia was elucidated by using rare copy number variant and linkage scans. To elucidate the genetic architecture of familial schizophrenia we combined linkage analysis with studies of fine-level chromosomal variation in families recruited from the Afrikaner population in SA. We demonstrated that individually rare inherited copy number variants (CNVs) were more frequent in cases with familial schizophrenia as compared to unaffected control and affect almost exclusively genic regions. Interestingly, we found that while the prevalence of rare structural variants was similar in familial and sporadic cases, the type of variants was markedly different. In addition using a high-density linkage scan with a panel of nearly 2000 markers we identified a region on chromosome 13q34 that showed genome-wide significant linkage to schizophrenia and showed that in the families not linked to this locus, there was evidence for linkage to chromosome 1p36. No causative CNVs were identified in either locus. Overall, our results from approaches designed to detect risk variants with relatively low frequency and high penetrance in a well-defined and relatively homogeneous population, provided strong empirical evidence supporting the notion that multiple genetic variants, including individually rare ones, that affected many different genes contribute to the genetic risk of familial schizophrenia. They also highlighted differences in the genetic architecture of the familial and sporadic forms of the disease.

© University of Pretoria

As a background to these findings one must realize that up to now the genetic architecture of the disease remained largely unknown. The strongest predictor of schizophrenia was having an affected first-degree relative. In addition to the familial forms, non-familial (sporadic) forms of the disease also existed. The exact proportion of each form was largely unknown, but it was thought that at least 60% of cases are sporadic. Genome-wide linkage scans had been conducted to identify loci harbouring rare mutations / variants that increased susceptibility to familial schizophrenia. Loci had been identified on almost every chromosome, but only a few regions had been replicated across studies. One such region was near the telomere of chromosome 13q. Across studies the region of peak linkage was broad, from 13q12 to 13q34. This region had also been linked to bipolar disorder. In addition to linkage studies a number of earlier as well as more recent studies had provided strong evidence supporting the importance of rare structural variants / mutation in schizophrenia vulnerability. Rare inherited structural lesions, in particular, were expected to be prominent in familial schizophrenia, but their full contribution to transmitted liability in familial SCZ cases had not been examined so far in a systematic manner.

To understand the genetic architecture of familial schizophrenia and the pattern of transmission of rare risk alleles in affected families we combined high-resolution linkage analysis with studies of fine-level chromosomal variation in families recruited from the Afrikaner population in SA. In addition to the genetic homogeneity, the Afrikaners were valuable for genetic studies because they presented a close-knit family structure and after the potential to perform detailed genealogical analysis, which afforded reliable discrimination of familial and non-familial forms of the disease and facilitates family-based genetic studies. Here, using approaches designed to detect risk variants with relatively low-frequency and high penetrance, we provided strong empiral evidence supporting the notion that multiple genetic variants, including individually rare ones that affected many different genes, contributing to the genetic risk of familial schizophrenia. This heterogeneity (present to some degree even in founder populations) was consistent with the hypothesis that there were many genes that contributed to schizophrenia and might account for past and present difficulties in finding bone fide genetic variants. Because there were significant clinical similarities of schizophrenia cases diagnosed in Afrikaners and those diagnosed in more heterogeneous populations (such as the US) our results were likely to have general implications regarding the genetic architecture of schizophrenia.

© University of Pretoria

In paper A15 it was reported that exome sequencing supported a de novo mutational paradigm for schizophrenia. Here we examined the possibility that rare de novo protein-altering mutations contributed to the genetic component of schizophrenia by sequencing the exomes of 53 sporadic cases, 22 unaffected controls and their parents. We identified 40 de novo mutations in 27 cases affecting 40 genes, including a potentially disruptive mutation in DGCR2, a gene located in the schizophrenia-predisposing 22q11 micro deletion region. A comparison to rare inherited variants indicated that the identified do novo mutations showed a large excess of non-synonymous changes in schizophrenia cases, as well as a greater potential to affect protein structure and function. Our analyses suggested a major role for the de novo mutations in schizophrenia as well as a large mutational target, which together provided a plausible explanation for the high global incidence and persistence of the disease. Although largely ignored in earlier efforts to model disease risk, de novo germ line mutations might account for a substantial fraction of sporadic schizophrenia cases. In agreement with this hypothesis, rare de novo copy number variants (CNVs) were emerging as an important genomic cause of schizophrenia and other psychiatric diseases, and the variant with the strongest statistical support for association with schizophrenia namely the 22q11.2 micro deletion is a de novo and recurrent mutation. In this study the identification of a disruptive de novo SNV in DGCR2 in a case with structurally intact 22q11.2 chromosomes suggested that disruption of this gene might contribute to the elevated schizophrenia risk associated with the 22q11.2 locus whether heterozygous deletions or point mutations in the DGCR2 were sufficient to render the susceptibility to schizophrenia observed in 22q11.2 microdeletion carriers or whether additional genetic interactions required could not be resolved until more DGCR2 mutations were identified and their penetrance determined. Our work showed that de novo protein-altering mutations contribute substantially to the genetic component of schizophrenia and taken together with previous estimates of the de novo CNV rate in the same population indicated that de novo mutations account for more than half of the sporadic cases of schizophrenia. Our findings were also in line with results for genome-wide scans for de novo CNV or CNVs in general, supporting the notion that multiple de novo genetic variants that affect many different genes contribute to the genetic risk of schizophrenia. The complexity of the neural substrates affected in schizophrenia and other psychiatric disorders offered a large mutational target comprised of many genes. We propose that this large number of targets that, when mutated, could give rise schizophrenia, along with the relatively high rate of protein-altering mutations empirically showed in this study, provided a plausible explanation for both the high global incidence and the persistence of schizophrenia despite extremely variable environmental factors, severely reduced fecundity and increased mortality. Our findings were an important step towards understanding the

© University of Pretoria

pathogenesis of the disease and emphasized the challenge in determining the neural substrates that these diverse genetic risk factors converged upon to generate a common pattern of clinical dysfunction and symptoms.

In 2012 it was reported in paper A16 that de novo gene mutations highlight patterns of genetic and neural complexity in schizophrenia. We sequenced at high coverage the exomes of families recruited from two populations with distinct demographic structures and history, to evaluate evidence for de novo etiologies in schizophrenia. We sequenced a total of 795 exomes from 231 parent-proband trios enriched for sporadic schizophrenia cases, as well as 34 unaffected trios. We observed in cases an excess of de novo nonsynonymous single-nucleotide variants as well as a higher prevalence of gene-disruptive de novo mutations relative to controls. We found four genes (LAMA2, DPYD, TRRAP and VPS39) affected by recurrent de novo events within or across the two populations, which is unlikely to have occurred by chance. We showed that de novo mutations affect genes with diverse functions and developmental profiles, but we also find a substantial contribution of mutations in genes with higher expression in early foetal life. Our results helped define the genomic and neural architecture of schizophrenia. The contribution of de novo nucleotide-level variants to schizophrenia risk had not been probed extensively. We completed exome sequencing of 146 Afrikaner proband-parent family trios of subjects with a diagnosis of SCZ or SAD and 34 unaffected control trios (53 and 22 trios respectively, have been described previously). We also sequenced the exomes of 85 US family trios of subjects with schizophrenia or schizoaffective disorder. None of the subjects carried rare de novo CNVs (of $\geq$ 30 kb in size) as determined in previous CNV scans of these cohorts. Considering phenotypic correlates, we observed a correlation between paternal age at the proband's birth and the number of de novo events per offspring, but did not find any other significant differences between proband carriers and non-carriers of de novo functional mutations. Notably, there was a functional correlation between the prenatal, expression bias of the mutated genes and the neurodevelopmental impact of the corresponding mutations. Specifically, among cases carrying de novo mutations, those with mutations in genes with a bias towards prenatal expression were more likely to have had multiple ($\geq$3) behavioural abnormalities in childhood (before the age of 10) compared to the cases carrying de novo mutations in genes with no bias towards fetal brain expression as well as worse functional outcome following disease onset. In addition, comparison of all genes with functional de novo events identified in our families with schizophrenia (n=145) with those identified in families with autism spectrum disorder (ASD) (n=675) identified 15 shared genes an overlap within the range expected by chance (p=0.29). However, 11 of the 15

© University of Pretoria

shared genes (73%) were included in our list of genes that showed a bias towards prenatal expression. The probability that this overlap arose by chance is very low (p=0.004). We also compared the identified functional de novo mutations to the de novo CNVs identified previously in our two cohorts (22 CNVs affecting 156 genes). Five genes (DGCR2, TOP3B, CIT, STAG1 and SMAP2) were altered by both de novo SNVs and CNVs, two of them in affected individuals across the two different populations tested. Two of these genes were within the 22q11.2 schizophrenia susceptibility locus. Our findings which characterized a diverse set of de novo mutations represented an original contribution to unravelling the genomic architecture of schizophrenia in the context of a mutation-selection balance model and highlighted the importance of using family samples where disease history had been thoroughly ascertained. Focusing on our comprehensively ascertained Afrikaner cohort we estimate that at least 17.6% of sporadic cases carried a de novo pathogenic exonic mutation, and at least 9.9% carried a de novo. Thus, such mutations might contribute to the risk in approximately one-fourth to one-third of all sporadic cases. Given that results of non-exonic regions were still forthcoming; this was likely to be an underestimate. Our findings also contributed to the understanding of the neural architecture of schizophrenia risk. Given that we estimated the number of schizophrenia risk loci at more than 850, our findings implied an exquisite sensitivity of the neural circuits underlying susceptibility to schizophrenia to precise levels or activity of many diverse proteins and signalling modules and suggested that focusing on circuits might be more commensurate with the heterogeneity of schizophrenia than other proposed mechanisms that concentrate on specific neurotransmitters or cell types. In addition, we showed that, in determining disease risk, not only the functioning of the mutated gene but also the timing of the genetic insult might be of critical importance. Specifically, although de novo mutations affected genes with diverse functions and developmental profiles, we described a substantial contribution of mutations in developmentally regulated genes with higher expression during early and mid-stage foetal life and showed that such mutations were enriched among adult cases with prominent early pre-psychotic, deviant behaviours. Notably, a bias towards prenatal expression was also demonstrated for genes affected by multiple types of schizophrenia-associated genetic variation. Our findings provided a mechanistic context to interpret epidemiological correlations among various prenatal environmental insults during the first and second trimesters of pregnancy and risk for schizophrenia. Moreover, the fact that expression might explain emerging links between miRNA dysregulation and psychiatric disorders. The challenge remained to identify the affected biological processes and neural circuits and to determine how they are affected. Unbiased network-based approaches as well as animal and cellular models of recurrent mutations will be invaluable in reaching this goal.

© University of Pretoria

B. Clinical Related Findings

Different sources in the literature have suggested that schizophrenia is a neurodevelopmental disorder with roots in early childhood. Early deviant behaviour patterns in children who go on to develop schizophrenia have been consistently reported in high-risk, prospective and retrospective behavioural studies of schizophrenia. Early behavioural deviance appears to occur in only a subgroup of patients. Early behaviour deviance and the various neurodevelopmental anomalies may in fact be both a subtype and disease-onset marker. These findings have important implications. It was important for the researchers to gain information regarding the early non-psychotic behaviour in the Afrikaner patients. In paper B1 a comparison study of early non-psychotic deviant behaviour in Afrikaner and US patients with schizophrenia and schizoaffective disorder was done. The analysis of early non-psychotic deviant behaviours before the age of ten years, broadly replicated the previous findings in the US study. The overall frequency of behaviours and their distributions were strikingly consistent in the two samples, with 68.8% of Afrikaners compared to 67.8% of US participants reporting early deviance. The populations differed with regard to the specific associations found. Initial molecular findings supported a genetic basis for this subtype. It was concluded that early non-psychotic childhood deviance in the Afrikaner population distinguished a distinct subtype of patients and that the form of early deviance manifested was meaningfully linked to later disease outcome.

Individuals with a deleted part of chromosome 22, called 22q11 deletion syndrome are at high risk for developing schizophrenia. Their risk is the highest known risk associated with a genetic defect, and is comparable to the risk of identical twin of a schizophrenic patient or of a child born to two schizophrenic patients. Previous studies revealed that 25 to 30 percent of people with microdeletions (genetic flaws that remove chunks of human chromosomes) in 22q11 develops schizophrenia. The case for the role of 22q11 microdeletions in schizophrenia was strengthened by previous studies showing that these microdeletions occur in up to 2% of adult schizophrenic patients. When the first patient in this study was identified with a 22q11 deletion syndrome (Velocardiofacial Syndrome) I published a case report in a clinical journal (Paper B2). Although these facts regarding 22q11 deletion syndrome was well known in the genetic literature, very little information was provided in the clinical literature. In this case report of a 21 year old Afrikaner male, a diagnosis of velocardiofacial syndrome was made on clinical grounds. The question was asked if this is a homogeneous subtype of schizophrenia? Later a FISH analysis test for 22q11 microdeletion confirmed

© University of Pretoria

this clinical diagnosis. The complete history, physical and psychiatric examination of the 21 year old male schizophrenic patient with velocardiofacial syndrome was described.

As more patients with the 22q11 deletion syndrome were identified in the study the question arose if the frequency of the 22q11 deletion is different in a founder population compared to the known frequency in non-founder populations. In paper B3 the frequency of the 22q11 deletion in Afrikaner schizophrenic patients were assessed. Our study confirmed the previously reported rate of 2% frequency of the 22q11 deletion in adult schizophrenic patients and provided a two-stage screening protocol to identify these patients. This screening protocol would help to cut costs to identify these patients with FISH analysis.

As the recruitment in this study progressed, more multiple affected Afrikaner families were identified. In paper B4 I reported on six multiple affected Afrikaner families suffering from schizophrenia or schizoaffective disorder. Three or more first degree relatives were affected in these families. In each family the following was reported on: a family tree, sociodemographic data, diagnostic features, substance abuse, early insults, early deviant behaviour and longitudinal course of illness. Environmental risk factors in interaction with genotype were discussed and it was emphasised that diseases will tend to cluster in families not because of a direct genetic effect, but because relatives are more vulnerable to the risk-increasing effect of a prevalent environmental risk factor. It was concluded that progress in the study of environmental factors that interact with genes needs to go hand-in-hand with development in molecular genetics.

In our clinical practice at Weskoppies Hospital (training facility for Department of Psychiatry at the University of Pretoria) information regarding early non-psychotic deviant behaviour in schizophrenia and schizoaffective disorder patients were becoming routine when a psychiatric history was obtained. Then the question arose whether the high incidence of early non-psychotic deviant behaviour, as found in previous studies in patients with schizophrenia and schizoaffective disorder, is unique to these illnesses, or whether it also occurs in bipolar disorder patients. Bipolar patients with psychotic features who are difficult to manage are frequently admitted to our tertiary psychiatric hospital. Paper B5 is a further follow-up study on early non-psychotic deviant behaviour in other diagnostic categories including bipolar disorder in Afrikaner

© University of Pretoria

patients. We determined and compared the incidence of early non-psychotic deviant behaviour in Afrikaner patients with bipolar disorder, schizoaffective disorder and schizophrenia. The findings suggested that early deviant behaviour might be a possible endophenotypic marker in schizophrenia and schizoaffective disorder but not in bipolar disorder. The high prevalence of unmarried subjects in schizophrenia emphasized the devastating effect of schizophrenia on social functioning. The adverse effect of schizophrenia relative to schizoaffective disorder and bipolar disorder on school functioning was also reflected in our results, with the average schizophrenia patient having never completed school. Bipolar patients were significantly likelier to attempt suicide in their lifetime.

Co-morbid psychiatric conditions became prominent as recruitment progressed. In paper B6 and paper B7 two comorbid conditions namely marijuana use/abuse and obsessive compulsive disorder (OCD)/obsessive compulsive symptoms (OCS) were further investigated. Marijuana use/abuse is often a comorbid factor in the management of schizophrenia. In paper B6 we examined how marijuana use/abuse might affect factors such as age at onset of illness and longitudinal course of illness. We found that early non-psychotic deviant behaviour was the most important factor determining age of onset and in male patient with schizophrenia it might be a possible endophenotypic marker. The interactive effect of gender and marijuana use was also significant in explaining the age of disease onset in males. Further suggestions were made to identify more sophisticated endophenotypes to aid the identification of susceptibility and modifying genes.

In paper B7 we investigated the prevalence and clinical characteristics of OCD and OCS in Afrikaner schizophrenia and schizoaffective disorder patients. We found that the prevalence rate of 13.2% of comorbid OCD/OCS in these patients differed from findings in other ethnic groups in SA, suggesting the possible role of genetic and cultural factors in the prevalence of comorbid OCD/OCS. Furthermore, the clinical characteristics of Afrikaner schizophrenia and schizoaffective disorder patients with and without comorbid OCD/OCS were the same; both groups were associated with significant psychopathology and poor prognosis.

In paper B9 I was a contributor with fellow authors to The Year in Human and Medical Genetics. Highlights of 2007 - 2008. Here we had the opportunity to describe the epidemiology and characteristics of schizophrenia in

© University of Pretoria

the Afrikaner genetic cohort in SA, and we compared this to characteristics of schizophrenia within the United States. This was done against the background where population isolates were considered particularly advantageous for the study of genetically complex and polygenic disorders. The theoretical basis for advantage was the proposition that specific allelic variants predisposing to complex disorders might be more readily identified because the overall range of allelic variation was less than in an outbred population.

The study of phenotypic extremes with an early-onset has been an important approach to the understanding of disease in all areas of medicine. In different medical illnesses, the combination of greater severity, relative treatment resistance, greater heritability and earlier age of onset has been observed. In Alzheimer's disease and breast cancer, important molecular genetic discoveries have been observed in subpopulations with an earlier age of illness onset. Childhood onset schizophrenia (12 years and younger) is a scarce condition. In 2007 the Afrikaner genetic cohort included more than 500 probands of Afrikaner descent who met the diagnostic criteria for schizophrenia or schizoaffective disorder according to the DSM IVTR. In paper B10 we described the clinical characteristics and premorbid variables in childhood-onset schizophrenia. From our database all participants that had an onset of schizophrenia or schizoaffective disorder at age of 12 years or younger were selected and their DIGS summaries were obtained. Twelve subjects with childhood onset schizophrenia were selected. Different selected parameters were then tabulated for comparison. Certain specific findings were made in this study, whilst other findings compare accurately with what is known about childhood-onset schizophrenia as was found in US samples.

In paper B11 a follow-up paper regarding velocardiofacial syndrome (VCFS) was done. The emphasis here was that this syndrome could present with many different features. Genetic counselling in these patients and what the future holds in terms of treatment were mentioned. Patients with VCFS can have covert symptoms that can be easily overlooked. Guidelines to the detection of the abnormalities in these patients were given.

In paper B12 a case study of higher genetic loading and subsequent development of psychosis was discussed. Patients in this genetic study quite often present with specific clinical issues that can be of great value in everyday practice. In this case study a 10 year old girl was referred to the genetic study because of a high familial genetic loading. At the age of 16

© University of Pretoria

years she subsequently presented with a full blown psychotic disorder. The initial presentation at 10 years of age as well as the course of illness until the diagnosis of psychosis was made was reported on. Possible predictors of the ultimate development of psychosis were highlighted and the follow-up of high-risk cases was emphasized.

The expected identification of susceptibility genes for psychiatric disorders may bring new opportunities and expectations from patients and families for the clinical translation of research findings in psychiatric genetics. In paper B13 the emphasis was on communicating scientific findings of the genetics of schizophrenia in the clinical setting. Information was provided about familial risk of schizophrenia with the theory behind individualizing risk of recurrence highlighted. Recent new findings regarding the new genetic frontier, copy number variations (CNV) were summarized and the genetic architecture of familial and sporadic schizophrenia applicable to the clinical situation was reviewed. A scenario in which genetic testing could be applied in velocardiofacial syndrome (VCFS) schizophrenia was debated. Referring to genetic discrimination in mental disorders, reference is made to the implementation of the Federal Genetic Information Non-Discrimination Act (GINA) of 2008 in the USA and the Mental Health Care Act of 2002 in SA.

I would like to comment on the reasoning behind the two papers published in the Afrikaans language. This was paper B8 and paper B14.

Paper B8: Die Afrikaner en Skisofrenie: 'n Genetiese studie

(The Afrikaner and Schizophrenia: A Genetic study)

Paper B14: Nuwe bevindings oor die genetiese argitektuur van skisofrenie:

Die bydrae van genetiese studies onder Afrikaners

(Newer findings on the genetic architecture of schizophrenia:

The contribution of genetic studies in Afrikaners)

After this genetic collaboration study was ongoing for 10 years I decided to publish an Afrikaans article in an Afrikaans Journal, Suid-Afrikaanse Tydskrif vir Natuurwetenskap en Tegnologie (SA Journal for Nature Science and Technology) on the findings of this study. My reasoning was to plough back knowledge in the Afrikaans science literature as a payback to the Afrikaner

© University of Pretoria

population I was working with. I am Afrikaans speaking, but this endeavour became quite a nightmare as the newer genetic nomenclature is not common knowledge in the Afrikaans language.

In paper B14 after 15 years of research in the Afrikaner population on the genetics of schizophrenia I offered an overview of the genetic landscape of schizophrenia in Afrikaans. This is a review of paper A12, A14, A15 & A16.

These Afrikaans papers could also be used as a source to provide information regarding research findings to patients who participated in the study and interested family members.

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Genetic variation in the 22q11 locus and susceptibility to schizophrenia

Hui Liu*, Gonçalo R. Abecasis†, Simon C. Heath‡, Alyson Knowles*, Sandra Demars*, Ying-Jiun Chen§, J. Louw Roos¶, Judith L. Rapoport/, Joseph A. Gogos§, and Maria Karayiorgou*,**

*The Rockefeller University, Human Neurogenetics Laboratory, New York, NY 10021; †University of Michigan, Department of Biostatistics, Ann Arbor, MI 48109; ‡Memorial Sloan–Kettering Cancer Center, New York, NY 10021; §Columbia University, College of Physicians and Surgeons, Department of Physiology and Cellular Biophysics, Center for Neurobiology and Behavior, New York, NY 10032; ¶University of Pretoria Department of Psychiatry and Weskoppies Hospital, Pretoria 0083, Republic of South Africa; and /National Institute of Mental Health, Child Psychiatry Branch, Bethesda, MD 20892

Edited by David E. Housman, Massachusetts Institute of Technology, Cambridge, MA, and approved September 19, 2002 (received for review March 29, 2002)

An increased prevalence of microdeletions at the 22q11 locus has been reported in samples of patients with schizophrenia. 22q11 microdeletions represent the highest known genetic risk factor for the development of schizophrenia, second only to that of the monozygotic cotwin of an affected individual or the offspring of two schizophrenic parents. It is therefore clear that a schizophrenia susceptibility locus maps to chromosome 22q11. In light of evidence for suggestive linkage for schizophrenia in this region, we hypothesized that, whereas deletions of chromosome 22q11 may account for only a small proportion of schizophrenia cases in the general population (up to =2%), nondeletion variants of individual genes within the 22q11 region may make a larger contribution to susceptibility to schizophrenia in the wider population. By studying a dense collection of markers (average one single nucleotide polymorphism/20 kb over 1.5 Mb) in the vicinity of the 22q11 locus, in both family- and population-based samples, we present here results consistent with this assumption. Moreover, our results are consistent with contribution from more than one gene to the strikingly increased disease risk associated with this locus. Finer-scale haplotype mapping has identified two subregions within the 1.5-Mb locus that are likely to harbor candidate schizophrenia susceptibility genes.

Microdeletions of chromosome 22q11 are associated with variable phenotypic expression (1) that includes relatively high frequency of severe mental illness (2, 3). Approximately one-third to one-fourth of patients with the 22q11 microdeletion develop schizophrenia or schizoaffective disorder, as defined in the *Diagnostic and Statistical Manual of Mental Disorders, 4th Edition* (DSM-I V) (4). Although the microdeletion occurs in the population at a rate of 0.025% (5), it has been found in up to 2% of adult schizophrenic patients (6) and up to 6% of cases with childhood-onset schizophrenia (7). Taken together, these studies suggest that the risk of schizophrenia for a patient with a 22q11 microdeletion may be =25–30 times the general population risk of 1% and that the rate of 22q11 microdeletions in schizophrenia, although relatively low, may be up to 12–80 times the estimated general population rate. In addition, at least two independent studies reported suggestive linkage results in the 22q11 region (8, 9). Finally, a series of recent studies suggest that schizophrenic patients carrying the 22q11 deletion bear the hallmark neuropsychological and neuroanatomical features of classical schizophrenia (10–12). It therefore seems likely that the 22q11 region harbors genes that, alone or in combinations, are causally implicated in schizophrenia in a subset of patients.

The overwhelming majority of the 22q11 deletions (=87%) are 3 Mb in size, whereas a smaller percentage of them (=8%) involve the same proximal breakpoint but a different distal breakpoint, resulting in a smaller 1.5 Mb deletion. All deletions are mediated by low copy repeat (LCR) sequences (13, 14). At least one schizophrenic patient has been described to carry the smaller 22q11 microdeletion (6), and therefore the "schizophrenia critical region" has been defined to 1.5 Mb (LCR-A to -B). The majority of the genes in the region are known (http://genome.ucsc.edu), making this locus amenable to a molecular genetic analysis.

The magnitude of the risk attributed to this deletion is unprecedented for a single genetic lesion and is comparable only to the risk among children of two schizophrenic parents or monozygotic cotwins of an affected individual. In both of these cases, the increased risk is due to the contribution of more than one susceptibility gene. It is therefore possible that the increased risk associated with this microdeletion is due to the contribution of more than one physically linked gene at the 22q11 locus. To address the role of individual genes from this chromosomal region, we have undertaken linkage disequilibrium (LD) studies in family samples (trios) that test for preferential transmission of common (>10% frequency) variants and multivariant haplotypes from parents to affected individuals. These studies are based on (and therefore test) the assumptions that, whereas deletions of chromosome 22q11 may account for only a small proportion of schizophrenia cases in the general population, nondeletion common variants of individual genes within the 22q11 region may make a larger contribution to susceptibility to schizophrenia in the wider population. Using this approach, we analyzed initially 13 of the 23 known genes residing in the deleted locus, delineated a region of significant LD within a segment in which *PRODH2* and *DGCR6* are the only known genes, and identified a relatively common haplotype where the schizophrenia susceptibility variant(s) likely reside (15). One additional result of this analysis was that the observed haplotypic association was particularly pronounced among children with schizophrenia and adults with early onset features of the disease. Further evidence implicating the *PRODH2* gene itself was provided by the identification of rare structural variants of this gene, especially in cases with early onset of the disease (15), as well as by behavioral and neurochemical studies of a mouse model (16). However, comparison of the increase in the morbid risk of schizophrenia associated with the identified *PRODH2* variation to the risk associated with the 22q11 microdeletion could not exclude contribution from other genes in the region. We therefore extended our analysis to include all of the known genes in the 1.5-Mb 22q11 locus.

Methods

Patient Samples. The Adult Schizophrenia (AS) Sample, the Childhood Onset Schizophrenia (COS) Sample, and the South African Afrikaner (SAA) case/control sample have been previously described (15). The SAA family sample included 93 patients of Afrikaner origin (age 19 or older) who met lifetime criteria for DSM-IV (4) schizophrenia or schizoaffective disorder and both their biological parents (trios). Participants were interviewed by

This paper was submitted directly (Track II) to the PNAS office.

Abbreviations: LD, linkage disequilibrium; AS, adult schizophrenia; SNP, single nucleotide polymorphism; COS, childhood onset schizophrenia; SAA, South African Afrikaner; HHRR, haplotype-based haplotype relative risk.

**To whom correspondence should be addressed. E-mail: karayim@mail.rockefeller.edu.

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

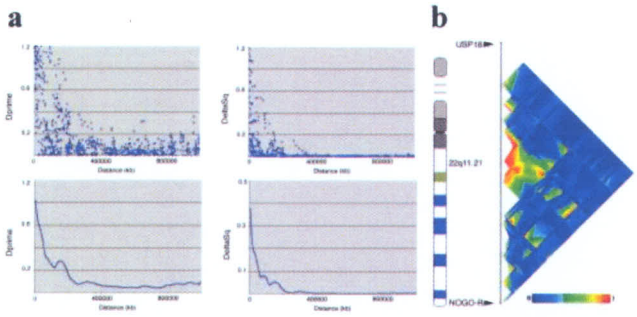


Fig. 1. (*a*) Decay of LD at the 22q11 locus. (*Upper*) Scatter plot of D' and L^2 values vs. physical distance. Values were calculated on the basis of the haplotype frequencies estimated by using partial phase information in the 106 unrelated AS trios. (*Lower*) Sliding window averages of disequilibrium coefficients. (*b*) Distribution of pair-wise disequilibrium (D') across the 22q11 locus. Multiple pair-wise comparisons are shown with D' statistics color coded (bright red and bright blue are opposite ends of the scale) and plotted at the marker locations.

specially trained clinicians using the Diagnostic Interview for Genetic Studies (17). Close genetic relatedness among the patients included in this analysis was excluded up to the fifth degree by family history information.

All participants provided written informed consent before participation in the study. The protocol and the consent forms were approved by the Rockefeller University Institutional Review Board and the Institutional Review Boards of all participating sites.

Single Nucleotide Polymorphism (SNP) Identification and Genotyping/PCR Primers and Conditions. See *Appendix* and Table 6, which are published as supporting information on the PNAS web site, www.pnas.org.

Statistical Analysis

Pairwise Background LD. See *Appendix*, which is published as supporting information on the PNAS web site (18–22).

Association Analysis. Transmissions of single SNPs, as well as multilocus SNP haplotypes, were analyzed by using the haplotype-based haplotype relative risk (HHRR) test (23), as described (15). Tightly linked markers that are in LD provide redundant information, making Bonferroni significance levels conservative. To estimate a global significance level, we randomly permuted transmitted and nontransmitted chromosomes for each trio (24). We generated 1,000,000 such permutations of the data by using the QTDT computer program (25, 26) and calculated the maximum observed allelic transmission score for each permutation.

Results

Pairwise Background LD. The measurement, extent, and variability of marker-to-marker disequilibrium are important factors in deciding the number of markers required for association scans and for interpreting marker-disease association results. We have therefore assessed the extent and characteristics of LD within the 1.5-Mb 22q11 locus. We calculated pairwise disequilibrium coefficients for all marker pairs within the region. The two distributions are summarized in Fig. 1*a*. Strong disequilibrium between a marker and a nearby disease allele can allow marker alleles to serve as surrogates for nearby markers. Multiple definitions of strong or "useful" disequilibrium have been proposed. For example, $L^2 > 0.10$ corresponds to a 10-fold increase in the sample size required to detect an association compared with testing disease alleles directly. L^2 takes maximum values only for markers that have identical allele frequencies; at any particular distance, markers with similar allele frequencies can exhibit higher values of L^2 than those with divergent

Table 1. Extent of significant LD in the 22q11 locus

	L^2			D'		
Distance	Avrg.*	>0.80†	>0.10†	Avrg.*	>0.90†	>0.33†
<10 kb	0.45	0.19	0.90	0.88	0.61	0.97
10–20 kb	0.25	0.06	0.68	0.70	0.24	0.91
20–50 kb	0.15	0.00	0.41	0.55	0.16	0.65
50–100 kb	0.09	0.00	0.28	0.36	0.10	0.35

Average (Avrg.) L^2 and D'* and the percentage of SNP pairs showing LD coefficients† above specific thresholds at different inter-SNP intervals are shown.

allele frequencies. Alternatively, $D' > 0.33$ also corresponds to a 10-fold increase in sample size, provided marker and trait allele frequencies are well matched. At 10, 20, and 50 kb, =97, 91, and 65% of markers in our sample exhibit $D' > 0.33$, respectively. In addition, 90, 68, and 41% of these also show $L^2 > 0.10$, indicating they are well matched in allele frequency (Table 1). We estimate the half-length of D' in the region to 48 kb, which compares well with the pattern observed in a recent study of LD along chromosome 22, which overlaps with our region (19). We identified 17 sets of three or more markers that define haplotypes with limited diversity. Some of these sets are overlapping, reflecting the difficulties in placing ancestral recombination events (data not shown; see also below). Finally, we examined the regional distribution of LD (Fig. 1*b*). The distribution of LD was found to be highly irregular, with significant LD concentrated in the middle of the locus over a distance of =250 kb (Fig. 1*b*).

Association Studies in Patient Samples. We used a three-step strategy to search for a signal of LD from the 22q11 locus. Our goals were: (*i*) to detect a significant and reproducible LD signal; (*ii*) to bind the critical region within the limits imposed by the density of our genetic map; and (*iii*) to identify the causal genetic variant. We first studied 106 trios from the AS sample and 26 trios from the COS sample. A primary scan included SNPs distributed throughout the 1.5-Mb region. Only relatively common (minor allele >10%), informative SNPs were further pursued at this stage. The average marker density is =1 SNP/20 kb, although the distribution of intermarker distances follows the distribution of genes and shows considerable variation.

Phase I of our association studies focused primarily in the region between the *DGCR6* and *CLDN5* genes, whereas Phase II focused in the region between the *CLDN5* and *NOGO-R* genes. We examined each allele of each marker or a combination of two adjacent markers for evidence of transmission disequilibrium using the HHRR test (23). The results of the primary association screen identified two clusters of neighboring SNPs having alleles with nominally significant HHRR results (Fig. 2*a*): Cluster A at the proximal end of the locus included three SNPs (15), and Cluster B at the distal end of the locus included four SNPs (V24, -26, -34, -35) (Table 2). Notably, the two clusters are separated by =1 Mb, and they are not in LD. We have reported separately our results from the detailed analysis of the proximal (centromeric) Cluster A (15).

Delineation of the Associated Region. To investigate further the primary genetic evidence at Cluster B, we studied a denser collection of markers in its vicinity (Table 2, yellow shading). Unlike Cluster A (15), there was no evidence for preferential overtransmission of the associated alleles in the patient group with early onset features (Table 2/COS sample column and data not shown). We therefore restricted our subsequent analysis in the 106 adult trios. Additional SNPs were typed, and significant HHRR results occurred at SNPs V23 ($P \leq 0.017$), V25 ($P \leq 0.017$), and V27 ($P \leq 0.002$) surrounding the previous significant SNPs (Fig. 2*b*; Table 2). In all, seven neighboring SNPs having alleles with nominally

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Table 2. HHRR *P* values

SNPs	Gene	Position		*IS* 1-SNP	*IS* 2-SNP	COS 1-SNP	COS 2-SNP	SAA 1-SNP	SAA 2-SNP	*IS*+ SAA 1-SNP	*IS*+ SAA 2-SNP	SAA<I< 1-SNP
V1	*CLDN5*	3' UTR		038		1		ND		ND		ND
V2		Intronic	NIA	0.123	0.19	0.49	0.4	ND	ND	ND	ND	ND
V3		Intergenic	NIA	0.24	0.08	0.13	O.S1	ND	ND	ND	ND	ND
V4	*TBX1*	Coding	NIA	0.13	*0.S9*	034	0.84	ND	ND	ND	ND	ND
V5	*TBX1*	Coding	NIA		*09*	0.84	*0.S*	ND	ND	ND	ND	ND
V6	*TBX1*	Coding	NIA	0.058	0.18	0.48	0.22	ND	ND	ND	ND	ND
V7	*FKSG1*	3•UTR	NIA	*0.65*	0.18	0.81	0.69	ND	ND	ND	ND	ND
V8	*FKSG1*	Coding	NIA	*0.25*	*015*	*0.18*	0.49	ND	ND	ND	ND	ND
V9	*TRXR2*	Coding	NIA	*0.18*	*018*	0.18	*039*	ND	ND	ND	ND	ND
V10	*TRXR2*	Coding	NIA	0.41	*098*	*036*	*036*	ND	ND	ND	ND	ND
V11	*COMT*	Coding	NIA	*091*	O.S3	*011*	*036*	ND	ND	ND	ND	ND
V12	*COMT*	Coding	NIA	*0.85*	*09*	O.S3	*0.11*	ND	ND	ND	ND	ND
V13	*COMT*	3' UTR	NIA	0.42	094	0.83	011	ND	ND	ND	ND	ND
V14	*ARVCF*	3' UTR	NIA	0.13	0.004	O.S8	033	ND	ND	ND	ND	ND
V15	*ARVCF*	Coding	NIA	*0.011*	*0.29*	0.43	0.43	ND	ND	ND	ND	ND
V16	*ARVCF*	Coding	NIA	*0.065*	0.18	0.19		0.89	ND	ND	ND	ND
V17	*T10*	intronic	*881205*	0.61	0.08	ND	ND	0.1 4	0.68	ND	ND	0.2
V18	*T10*	Intronic	NIA	0.42	*0.02*	0.13	ND	0.83	0.24	ND	ND	ND
V19	*IL1:1:JJ:J7*	Promoter	*1013118*	0.13	*0.29*	ND	ND	0.4	0.2	OD41	ND	*0.16*
V20	*FLJ22137*	3' UTR			0.1		ND		0.13		0.013	
V21			*1640299*	*0.26*	0.43	ND	ND	0.12	*0.26*	0.1	*0.28*	*0.01*
V22	*HTF9C*	Intergenic	NIA	*032*	*0.01*	0.43	ND	0.13	O.S1		0.49	0.09
V23	*HTF9C*	intronic	1633445	0.18	*0.005*	ND	ND	0.89	0.06	0.49	0.00044	0. 14
V24	*RANBP1*	Promoter	395446	*0.011*	0.011	ND	ND	*0.059*	OD4	0.001	0.1)0048	*0.02*
V25	*KIAA1292*	Intronic	115163	0.035	0.016	*032*	ND	*0.041*	0.028	0.003	0.001	*0.05*
V26	*KIAA1292*	Promoter	115164	0.01	*0.006*	ND	ND	0.033	0.034	0.001	0.0001	*0.05*
V27	*KIAA1292*	Intronic	*115114*	0.0008	*0.0015*	034	*0.0016*	ND	0.013	0.000027	0.000036	*0.02*
		3' UTR	NIA	*0.002*	0.036	ND	0.019	ND	0.016	0.000083	0.00054	*0.05*
\128		Intergenic	*649131*	*09*	*0 11*	ND	0.66	ND	*098*	*092*	O.S3	*039*
\129		Intergenic	*105213*	0.19	0.21	ND		ND	*091*	0.23	ND	0.48
V30		Intergenic	1O 10063	*0.62*	0.28	ND	0.89	ND	*098*	ND	ND	0.46
V31		intergenic	*115194*	0.058	*0.11*	ND	*0.82*	ND	0.83	ND	ND	0.44
V32		intergenic	*665180*	*0.19*	*096*	ND	0.69	ND	OD4	ND	ND	0.29
V33		intergenic	NIA		0.1	ND	*0.001*	ND	0.028	ND	ND	*035*
V35		intergenic	*101428*	0.034	*OD26*	0.19		ND	*0.S6*	ND	ND	O.11
V36	*NOGOR*	Intergenic	NIA	*0.02*	0.035	0.49	0.48	ND	ND	ND	ND	*035*
V37	*NOGOR*	Intronic	*101421*		0D14	ND	ND	ND	ND	ND	ND	ND
V38	*NOGOR*	intronic	*696880*	0.109	0.014	*0.06*	ND	038	ND	ND	ND	*0.12*
V39	*NOGOR*	Intronic	NIA	*0.25*	*032*	ND	ND	ND	ND	ND	ND	ND
		Intronic	*855051*	0.19		0.45		ND		ND		ND

Significance levels (*P* values) <0.05 are highlighted. Location within genes or intergenic regions is indicated. SNPs from public databases are indicated by reference SNP identifications. For simplicity,abbreviated Variant numbers (VII) are assigned and used in the text. Inter-SNP distances and PCR primers and conditions are shown in the *Appendix*.

significant 1ffiR R results were identified in both primary and secondary screen, distributed over a region of ""250 kb from the *HTF9C* gene ID the *NOGO-R* gene supporting the existence of ▮▮▮ with the disease over the investigated region.

Using the relatively high-density SNP map, it was possible to determine the fine structure of the haplotypes in this region. Specifically, the analysis shows that the region can be parsed into two major haplotype BLOCKs (I and II), with each block having

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

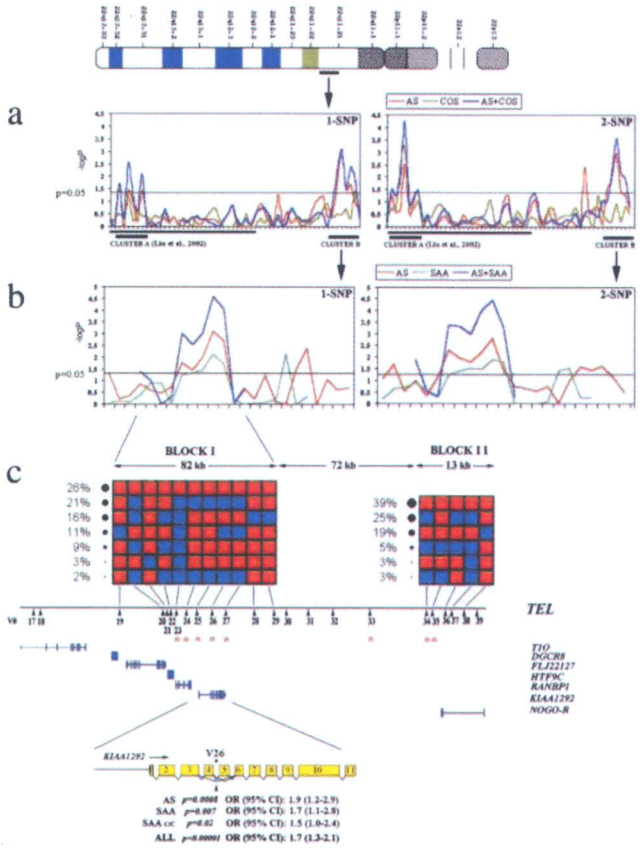


Fig. 2. (*a*) Single SNP and 2-SNP HHRR *P* values (−log10) across the 22q11 locus. The solid line underneath the *x* axis indicates the SNPs presented in Liu *et al.* (15). The remaining SNPs are detailed in Table 2 (blue shading). The two clusters of significant SNP association across the entire 22q11 locus (Clusters A and B) are also underlined. The AS and COS samples were analyzed both separately and as a combined sample, because they originated from the same geographical area (U.S.). Distances are not to scale. (*b*) Detailed view of HHRR significance levels at Cluster B when additional SNPs were genotyped in the vicinity of the initial association shown in *a*. Detailed description can be found in Table 2. The family samples used for this analysis were the AS and SAA, as well as a combination of the two (see text). Distances are not to scale. (*c*) Haplotypic structure of Cluster B. Two haplotypic BLOCKs (I and II) are shown. Red squares indicate major and blue squares minor alleles. Observed haplotype frequencies are shown on the side of each block. The variant number (V#) is indicated at the bottom, and each variant's position relative to the genes residing in this region is also approximately indicated. Asterisks indicate SNPs showing significant association with schizophrenia in the AS and/or SAA samples. The position of the most significantly associated SNP (V26) within intron 4 of the *KIAA1292* candidate gene is indicated along with the associated *P* values and odds ratios (approximate relative risks). Blue bars represent the approximate distribution of exons.

limited diversity (<5 haplotypes account for >90% of all chromosomes) with only a small number of SNPs between blocks. SNPs V23–27 reside in BLOCK I, whereas SNPs V34 and V35, at the 3′ end of the *NOGO-R* gene, reside in BLOCK II. There is no strong evidence for LD between the significant SNPs in BLOCK I and the significant SNPs in BLOCK II (Table 3). On the other hand, within-block SNPs demonstrated highly significant LD with each other (data not shown), indicating that the observed associations in Cluster B could represent two independent disease susceptibility-conferring polymorphisms. Consistent with this prediction, when families were divided into two groups according to the presence or absence in the probands of the risk allele of SNP V26 (the most significant SNP in BLOCK I), significance of the schizophrenia associations with BLOCK I SNPs V23–27 increased dramatically in families with the V26 risk allele present. By contrast, associations with BLOCK II SNPs V34 and V35 remained unchanged or even decreased modestly in the same families (data not shown), indicating a minimal contribution of the V26 risk allele to these latter associations.

Table 3. LD (D′) between SNPs in BLOCKs I and II

	BLOCK II	
BLOCK I	V34	V35
V23	0.22	0.04
V24	0.07	0.09
V25	0.05	0.07
V26	0.01	0.04
V27	0.02	0.07

In conclusion, the strongest observed association at this stage ($P < 0.0008$) occurs at SNP V26, with four nearby SNPs (V23, -24, -25, -27) also exhibiting nominal evidence for association. All five SNPs fall in one haplotypic block (Fig. 2*c*), and thus the evidence for association that they provide is not independent. In these circumstances, a Bonferroni correction is conservative. In fact, the haplotype and disequilibrium structure of this region implies that the 72 markers investigated here correspond to much less than 72 independent tests. We therefore generated 100,000 permutations of the data (24) and recorded the most significant association for each one. Overall, the probability of observing an association as strong as the one presented here is $P < 0.043$, after adjusting for multiple testing at 63 markers (nine markers with low minor allele frequencies and a correspondingly small number of informative trios were not included in this analysis). Similarly with the Cluster A variation (15), no association was detected with the representative SNP V26 in a sample of 80 North American families not afflicted with schizophrenia (not shown), arguing against a generalized transmission distortion at the 22q11 locus.

Replication of the Association in Two Independent Samples. We sought additional confirmation of our findings by examining both the primary and secondary set of SNPs in another family sample derived from an independently collected population (Fig. 2*b*; Table 2) consisting of 93 trios from the Tshwane (formerly known as Pretoria) area in South Africa (SAA sample). Five of the eight SNPs demonstrating significant association in the AS sample also showed significant association in the SAA sample (Fig. 2*b*; Table 2). All significant SNPs are located in BLOCK I. Moreover, single and multilocus analyses within BLOCK I showed that all of the same alleles and the same multiallelic haplotypes were overtransmitted in the two independent family data sets (Fig. 2*b*; Tables 2 and 4). Finally, we sought confirmation of the associations observed in both family samples in an independent case-control sample consisting of 93 unrelated schizophrenics of Afrikaner origin and 75 ethnically matched controls. We observed a similar pattern of association, consistent with involvement of an identical common disease variant in all three samples (Table 2). Combined analysis of all three samples at SNPs residing in BLOCK I resulted in a $P < 10^{-5}$ and revealed an odds ratio (approximate relative risk associated with the causative variant) of =1.7 at SNP V26. We did not replicate the association with SNPs V34 and V35, from the 3′ end of the *NOGO-R* gene located in BLOCK II, in the SAA family set, although significant one- and two-SNP associations were observed =15 kb proximal, in the vicinity of SNP V33 (Fig. 2*b*; Table 2). Because different causal variants may be involved in this case in different populations, confirmation or exclusion of the *NOGO-R* gene (27) (an excellent candidate gene from this region) from association with schizophrenia will require the identification of all polymorphisms in the region and the assaying of most of these polymorphisms in both populations.

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Table 4. Five-SNP risk haplotypes in BLOCK I

V23	V24	V25	V26	V27	T/nT ratio AS	T/nT ratio SAA
2	2	2	2	1	1.5	1.4
1	2	2	2	1	1.3	1.3
1	1	1	1	2	0.6	0.6
1	2	2	1	2	0.6	0.8
				Other	0.5	0.5

T, transmitted; nT, non-transmitted.

Table 5. Coding region variants in BLOCK I genes

Gene	Nucleotide position	Nucleotide change	Amino acid change	No. carriers*	Associated† allele
DGCR8	610	A ⟶ G	I ⟶ V	1 (10)	I
HTF9C	497	G ⟶ T	A ⟶ S	2 (10)	S
	683	C ⟶ A	L ⟶ I	1 (10)	I
KIAA1292	1707	G ⟶ A	A ⟶ T	5 (25)‡	A

DGCR8 : GI:12225241; *HTF9C* : GI:15426493; *KIAA1292* : GI:17484904.
*The number in parentheses indicates the number of patients sequenced. Unless otherwise stated, all the patients are heterozygous for the change.
†Allele associated to V26 allele 2.
‡Information from more than 10 patients was available for this change.

Genetic Interactions and Parent-of-Origin Analysis. Having established a reproducible pattern of association represented by SNP V26, which serves as a tag of the risk haplotypes, we then checked whether the proximal and distal disease risk-associated variation contributes to distinct or overlapping sets of patients. We stratified the AS families according to the presence or absence of the Cluster A (*PRODH2/DGCR6*) risk haplotype (15) in the proband (carriers vs. noncarriers). This stratified analysis provided significant evidence for association with SNP V26 only in the subset of noncarrier AS families (HHRR x^2 = 15.5, $P < 0.001$; transmitted (T)/nontransmitted (nT): allele 1, 36/68; allele 2, 106/74). In this subset, transmission of the V26 risk allele was associated with an odds ratio of 1.9 [95% confidence interval: 1.3–2.6], statistically indistinguishable when compared with the entire AS sample. By contrast, there was no evidence of association with SNP V26 in the carrier subset (HHRR x^2 = 0, P = 1; T/nT: allele 1, 27/27; allele 2, 36/36). As expected, further analysis of the distribution of the V26 alleles in the two proband groups provided similar results (alleles 1/2: noncarriers, 36/106; carriers, 27/27; x^2 = 10.9, $P < 0.001$). This observation is consistent with a disease-risk contribution from the proximal or distal subregions to distinct subsets of patients, possibly with distinct syndromic features (such as early onset features). However, the significance of this finding is, at present, unclear, because it could not be replicated in the South African samples (data not shown). A more conclusive analysis of interaction will require considerable increase in the sample size. We also conducted a parent-of-origin analysis, which did not provide any significant evidence for a parent-of origin effect. For example, when marker V26 is considered, parental origin can be discerned in 74 of 96 transmissions/nontransmissions. In this case, maternal and paternal transmissions appear to contribute equally to the V26 association (8 vs. 24 paternal and 13 vs. 29 maternal).

Candidate Genes from the Associated Region. The replication of the association between BLOCK I SNPs and schizophrenia in an independent family sample and an independent case-control sample, in an identical pattern, adds considerable support to the genetic evidence associating genes at this region with schizophrenia. We analyzed the preferential transmission of specific haplotypes using genotypes from the five BLOCK I SNPs that gave the lowest P values in the association study (Table 4). We identified two overtransmitted risk haplotypes that are distinguished by the presence of allele 2 of SNP V26 and allele 1 of SNP V27 (Table 4). All of the other common haplotypes in this region are undertransmitted. Although there may be separate mutations in the two risk haplotypes, the simplest explanation is that a putative causative allele is strongly linked to these two SNPs rather than to a specific haplotype. On the basis of this latter assumption, we attempted to identify a linked causal missense allele, examining the coding region of all known genes within BLOCK I (as well as that of the *T10* gene located immediately adjacent to it) (Fig. 2*c*) for missense or nonsense allelic variants, by sequencing and comparing a panel of five homozygous carriers and five homozygous noncarriers of the V26 risk allele 2. We identified four missense variants in three of the seven tested genes (Table 5). Because of their relatively low frequency, it is unclear whether they can explain the full effect associated with the very common SNP V26 risk allele. The causal SNP or SNPs may, therefore, affect regulation of one or more of the resident genes, many of which are plausible candidates. Four known genes found in the RefSeq database are located in BLOCK I (Fig. 2*c*): the *DGCR8* gene (which encodes for a protein of undetermined function), the *FLJ22127* gene (which encodes for a protein with a double-strand RNA-binding domain), the *HTF9C* gene (28) (which likely encodes for a methyltransferase), and the *RANBP1* gene (29) (which encodes for a protein that interacts with the cell-cycle regulatory protein Ran/TC4). Two more genes predicted by analysis of the syntenic region on mouse chromosome 16 and supported by available EST and mRNA databases are located either within BLOCK I (*KIAA1292*) (30) or immediately outside it (*T10*) (31) (Fig. 2*c*).

We note that three of the five SNPs that show evidence of association in SAA and AS samples (V25–27) are located within the transcription unit and the putative 5′ control region of *KIAA1292* (Fig. 2*c*). In addition, of all SNPs that show evidence of association, only V26 and V27 are exclusively present in the risk haplotypes (Table 4). SNP V26, located in intron 4 of *KIAA1292*, shows the strongest and most consistent evidence of association of all SNPs in either primary or secondary panels in all tested samples. For that reason and because all three tested SNPs within the *KIAA1292* gene unit demonstrate significant association with schizophrenia in both the AS and the SAA samples, we consider the *KIAA1292* gene a prime candidate worthy of further investigation. KIAA1292 is predicted to have four transmembrane domains and a new form of cysteine-rich domain called DHHC-CRD (32) (Fig. 3 *a* and *b*). The only member of the family that has been characterized to some extent is yeast *Erf2*, a component of a previously uncharacterized Ras subcellular localization pathway (33). The *KIAA1292* gene is

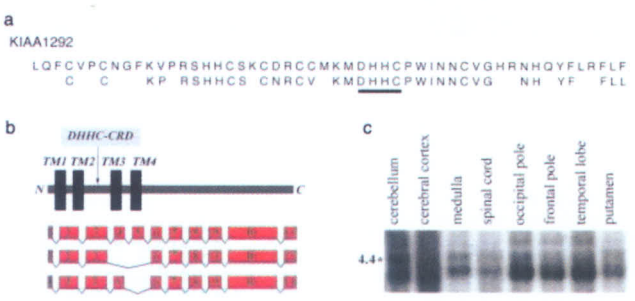


Fig. 3. (*a*) Examination of the deduced KIAA1292 polypeptide sequence shows a new form of cysteine-rich domain called DHHC-CRD that includes a DHHC motif consisting of the amino acids Asp-His-His-Cys (DHHC) and a cysteine-rich domain (CRD) that includes a Cys4 zinc-finger-like metal binding site. The sequence for the KIAA1292 DHHC-CRD as well as a consensus sequence for this domain (32) are indicated. (*b*) Alternative splice forms of the *KIAA1292* gene. Note that the DHHC-CRD is largely part of the loop between adjacent TM2 and TM3 (TM, transmembrane). (*c*) Brain expression of the *KIAA1292* gene.

GENETICS

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

highly expressed in the human brain (Fig. 3*c*), and *in situ* hybridization analysis of expression in the mouse brain demonstrated high expression levels especially in the cortex and hippocampus, areas expected to play an important role in the pathophysiology of the disease (Y.-J.C., M.K., and J.A.G., unpublished work). It should be noted that analysis of EST databases as well as RT-PCR experiments on RNA from human and mouse brains identified two major alternatively spliced variants of *KIAA1292* involving exons 4 and 5 (Fig. 3*b*), which encode for protein forms with truncated DHHC-CDR domains and, most likely, reduced or altered function. Of interest, V26 is located in the intron between the alternatively spliced exons 4 and 5, and a potential effect of V26 genotype on the function of putative splicing enhancers and the relative levels of the two splicing variants warrants further investigation.

Discussion

Using a relatively dense genetic map of SNPs across the entire 1.5-Mb 22q11 region, we found strong evidence of LD with schizophrenia. In our study, D' decayed halfway between maximum and minimum values at 48 kb. The extent of disequilibrium observed in this region of chromosome 22q is similar to that previously described for this (19) and other regions (22, 34). Among the markers investigated, 56 of 72 were in strong LD with surrounding markers and included in a common haplotype block. Nevertheless, a small proportion of the markers we investigated exhibited only low levels of LD with surrounding markers, and in these regions, it would be desirable to investigate additional markers. Given this limitation, our analysis identified two associated subregions (Fig. 2*a*), suggesting that the 22q11 microdeletion-associated schizophrenia may have the characteristics of a contiguous gene syndrome.

In the proximal part of the locus, we delineated a subregion of significant association within a segment of ≈30 kb in which *PRODH2* and *DGCR6* are the only known genes, and we identified one relatively common haplotype where the schizophrenia susceptibility variant(s) likely reside (15). In the distal part of the locus, we bound the associated subregion to a haplotypic block that spans at least 80 kb and shows consistent association with the disease in three independent samples from two different populations.

One difference between the two subregions of association is that in the proximal subregion (Cluster A), the association was more pronounced among children with schizophrenia and adults with early onset features of the disease. Similar association was not observed in the distal subregion. These observations seem to suggest that genetic variation at the distal subregion contributes to schizophrenia independently of the onset of the disease, whereas the proximal subregion contains genes (most likely *PRODH2*) contributing to early onset forms of the disease. It might be relevant that previous studies indicated a relatively wide range of age at onset among 22q11-deleted patients with schizophrenia ranging from childhood onset to up to 26 years of age (2).

The combination of the study of Liu *et al.* (15) and the present study represents a comprehensive application of LD mapping across the 22q11 susceptibility locus in relation to schizophrenia. An implication of these studies is that more than one gene from this region may contribute to schizophrenia susceptibility. Functional variants of these genes may independently or synergistically exert modest increases in the disease risk in nondeleted patients. In 22q11 microdeletion carriers, drastic reduction of dosage to half in more than one gene may result in the dramatic 25- to 30-fold increase in disease risk observed. Alternatively, genes in the identified proximal and distal subregions may contribute only partially to the 25- to 30-fold increase in the risk associated with this locus, the largest part of it being due to an unidentified gene that lacks common functional variants and is variably mutated in a small portion of nondeleted schizophrenic patients. Such rare and possibly highly penetrant mutations would be invisible to our present experimental strategy, which focuses on relatively common variants. Nevertheless, our current analysis revealed a limited number of candidate genes (including two membrane-associated proteins, *KIAA1292* and *NOGO-R*, highly expressed in brain regions implicated in schizophrenia), which need to be tested in additional samples and whose biological function needs to be explored in more detail by generation of animal models.

We thank all of the families who participated in this research. Support for this work was provided in part by National Institute of Mental Health R01 MH61399-01 (to M.K.). Clinical work at the Rockefeller University Hospital General Clinical Research Center was supported in part by National Center for Research Resources at the National Institutes of Health, Grant M01 RR00102.

1. Scambler, P. J. (2000) *Hum. Mol. Genet.* **9,** 2421–2426.
2. Pulver, A. E., Nestadt, G., Goldberg, R., Shprintzen, R. J., Lamacz, M., Wolyniec, P. S., Morrow, B., Karayiorgou, M., Antonarakis, S. E., Housman, D., *et al.* (1994) *J. Nerv. Ment. Dis.* **182,** 476–478.
3. Murphy, K. C., Jones, L. A. & Owen, M. J. (1999) *Arch. Gen. Psychiatry* **56,** 940–945.
4. American Psychiatric Association (1994) *Diagnostic and Statistical Manual of Mental Disorders* (Am. Psychiatric Press, Washington, DC), 4th Ed.
5. du Montcel, S. T., Mendizabal, H., Ayme, S., Levy, A. & Philip N. (1996) *J. Med. Genet.* **33,** 719.
6. Karayiorgou, M., Morris, M. A., Morrow, B., Shprintzen, R. J., Goldberg, R., Borrow, J., Gos, A., Nestadt, G., Wolyniec, P. S., Lasseter, V. K., *et al.* (1995) *Proc. Natl. Acad. Sci. USA* **92,** 7612–7616.
7. Usiskin, S. I., Nicolson, R., Krasnewich, D. M., Yan, W., Lenane, M., Wudarsky, M., Hamburger, S. D. & Rapoport, J. L. (1999) *J. Am. Acad. Child Adolesc. Psychiatry* **38,** 1536–1543.
8. Blouin, J. L., Dombroski, B. A., Nath, S. K., Lasseter, V. K., Wolyniec, P. S., Nestadt, G., Thornquist, M., Ullrich, G., McGrath. J., Kasch, L., *et al.* (1998) *Nat. Genet.* **20,** 70–73.
9. Shaw, S. H., Kelly, M., Smith, A. B., Shields, G., Hopkins, P. J., Loftus, J., Laval, S. H., Vita, A., De Hert, M., Cardon, L. R., *et al.* (1998) *Am. J. Med. Genet.* **81,** 364–376.
10. Bearden, C. E., Woodin, M. F., Wang, P. P., Moss, E., McDonald-McGinn, D., Zackai, E., Emannuel, B. & Cannon, T. D. (2001) *J. Clin. Exp. Neuropsychol.* **23,** 447–464.
11. Chow, E. W., Mikulis, D. J., Zipursky, R. B., Scutt, L. E., Weksberg, R. & Bassett, A. S. (1999) *Biol. Psychiatry* **46,** 1436–1442.
12. Chow, E. W., Zipursky, R. B., Mikulis, D. J. & Bassett, A. S. (2002) *Biol. Psychiatry* **51,** 208–215.
13. Edelmann, L., Pandita, R. K., Spiteri, E., Funke, B., Goldberg, R., Palanisamy, N., Chaganti, R. S., Magenis, E., Shprintzen, R. J. & Morrow B. E. (1999) *Hum. Mol. Genet.* **8,** 1157–1167.
14. Shaikh, T. H., Kurahashi, H., Saitta, S. C., O'Hare, A. M., Hu, P., Roe, B. A., Driscoll, D. A., McDonald-McGinn, D. M., Zackai, E. H., Budarf, M. L., *et al.* (2000) *Hum. Mol. Genet.* **9,** 489–501.
15. Liu, H., Heath, S. C., Sobin, C., Roos, J. L., Galke, B. L., Blundell, M. L., Lenane, M., Robertson, B., Wijsman, E. M., Rapoport, J. L., Gogos, J. A. & Karayiorgou, M. (2002) *Proc. Natl. Acad. Sci. USA* **99,** 3717–3722.
16. Gogos, J. A., Santha, M., Takacs, Z., Beck, K. D., Luine, V., Lucas, L. R., Nadler, J. V. & Karayiorgou, M. (1999) *Nat. Genet.* **21,** 434–439.
17. Nurnberger, J. I., Jr., Blehar, M. C., Kaufmann, C. A., York-Cooler, C., Simpson, S. G., Harkavy-Friedman, J., Severe, J. B., Malaspina, D. & Reich, T. (1994) *Arch. Gen. Psychiatry* **51,** 849–859.
18. Abecasis, G. R., Cherny, S. S., Cookson, W. O. & Cardon, L. R. (2002) *Nat. Genet.* **30,** 97–101.
19. Dawson, E., Abecasis, G. R., Bumpstead, S., Chen, Y., Hunt, S., Beare, D. M., Pabial, J., *et al.* (2002) *Nature* **418,** 544–548.
20. Hedrick, P. W. (1987) *Genetics* **117,** 331–341.
21. Devlin, B. & Risch, N. (1995) *Genomics* **29,** 311–322.
22. Abecasis, G. R., Noguchi, E., Heinzmann, A., Traherne, J. A., Bhattacharyya, S., Leaves, N. I., Anderson, G. G., Zhang, Y., Lench, N. J., Carey, A., *et al.* (2001) *Am. J. Hum. Genet.* **68,** 191–197.
23. Terwilliger, J. & Ott, J. (1992) *Hum. Hered.* **42,** 337–346.
24. McIntyre, L. M., Martin, E. R., Simonsen, K. L. & Kaplan, N. L. (2000) *Genet. Epidemiol.* **19,** 18–29.
25. Abecasis, G. R., Cardon, L. R. & Cookson, W. O. C. (2000) *Am. J. Hum. Genet.* **66,** 279–292.
26. Abecasis, G. R., Cookson, W. O. & Cardon, L. R. (2000) *Eur. J. Hum. Genet.* **8,** 545–551.
27. Fournier, A. E., GrandPre, T. & Strittmatter, S. M. (2001) *Nature* **409,** 341–346.
28. Guarguaglini, G., Battistoni, A., Pittoggi, C., Di Matteo, G., Di Fiore, B. & Lavia, P. (1997) *Biochem. J.* **325,** 277–286.
29. Coutavas, E., Ren, M., Oppenheim, J. D., D'Eustachio, P. & Rush, M. G. (1993) *Nature* **366,** 585–587.
30. Nagase, T., Ishikawa, K., Kikuno, R., Hirosawa, M., Nomura, N. & Ohara, O. (1999) *DNA Res.* **6,** 337–345.
31. Halford, S., Wilson, D. I., Daw, S. C., Roberts, C., Wadey, R., Kamath, S., Wickremasinghe, A., Burn, J., Goodship, J., Mattei, M. G., *et al.* (1993) *Hum. Mol. Genet.* **2,** 1577–1582.
32. Putilina, T., Wong, P. & Gentleman, S. (1999) *Mol. Cell. Biochem.* **195,** 219–226.
33. Bartels, D. J., Mitchell, D. A., Dong, X. & Deschenes, R. J. (1999) *Mol. Cell. Biol.* **19,** 6775–6787.
34. Reich, D. E., Cargill, M., Bolk, S., Ireland, J., Sabeti, P. C., Richter, D. J., Lavery, T., Kouyoumjian, R., Farhadian, S. F., Ward, R. & Lander, E. S. (2001) *Nature* **411,** 199–204.

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Genetic variation at the 22q11 PRODH2/DGCR6 locus presents an unusual pattern and increases susceptibility to schizophrenia

Hui Liu[†], Simon C. Heath[‡], Christina Sobin[†], J. Louw Roos[§], Brandi L. Galke[†], Maude L. Blundell[†], Marge Lenane[¶], Brian Robertson[||], Ellen M. Wijsman[**], Judith L. Rapoport[¶], Joseph A. Gogos[††], and Maria Karayiorgou[†‡‡]

[†]Human Neurogenetics Laboratory, The Rockefeller University, New York, NY 10021; [‡]Memorial Sloan Kettering Cancer Center, New York, NY 10021; [§]Department of Psychiatry and Weskoppies Hospital, University of Pretoria, Pretoria 0083, Republic of South Africa; [¶]Child Psychiatry Branch, National Institute of Mental Health, Bethesda, MD 20892; [||]Department of Psychiatry and Valkenberg Hospital, University of Capetown, Capetown 7925, Republic of South Africa; [**]Division of Medical Genetics and Department of Biostatistics, University of Washington, Seattle, WA 98195; and [††]College of Physicians and Surgeons, Department of Physiology and Cellular Biophysics, Center for Neurobiology and Behavior, Columbia University, New York, NY 10032

Communicated by David E. Housman, Massachusetts Institute of Technology, Cambridge, MA, December 26, 2001 (received for review September 25, 2001)

The location of a schizophrenia susceptibility locus at chromosome 22q11 has been suggested by genome-wide linkage studies. Additional support was provided by the observation of a higher-than-expected frequency of 22q11 microdeletions in patients with schizophrenia and the demonstration that ≈20–30% of individuals with 22q11 microdeletions develop schizophrenia or schizoaffective disorder in adolescence and adulthood. Analysis of the extent of these microdeletions by using polymorphic markers afforded further refinement of this locus to a region of ≈1.5 Mb. Recently, a high rate of 22q11 microdeletions was also reported for a cohort of 47 patients with Childhood Onset Schizophrenia, a rare and severe form of schizophrenia with onset by age 13. It is therefore likely that this 1.5-Mb region contains one or more genes that predispose to schizophrenia. In three independent samples, we provide evidence for a contribution of the *PRODH2/DGCR6* locus in 22q11-associated schizophrenia. We also uncover an unusual pattern of *PRODH2* gene variation that mimics the sequence of a linked pseudogene. Several of the pseudogene-like variants we identified result in missense changes at conserved residues and may prevent synthesis of a fully functional enzyme. Our results have implications for understanding the genetic basis of the 22q11-associated psychiatric phenotypes and provide further insights into the genomic instability of this region.

Schizophrenia is a common severe mental illness that affects 1% of the population and is characterized by disordered thinking as well as deficits in emotional and social behavior. Microdeletions of chromosome 22q11 are associated with variable phenotypic expression that often includes learning disabilities, palatal abnormalities, congenital heart defects, and mildly dysmorphic facial features (1). One of the first clues that psychiatric symptomatology may also be associated with this microdeletion was a report of expressionless face, monotonous speech, and flattened affect among children with the microdeletion (2). In light of the evidence for suggestive linkage for schizophrenia on chromosome 22 (reviewed in ref. 3), patients with the 22q11 microdeletion were evaluated for psychiatric symptoms or disorders, and a relatively high frequency of severe mental illness has since been reported. Specifically, two independent studies have reported that 25–31% of patients with the 22q11 microdeletion met diagnostic criteria for schizophrenia or schizoaffective disorder (4, 5). Although the microdeletion occurs in the population at a rate of 0.025%, it has been found in 2% of adult schizophrenic patients (6) and in 6% of cases with childhood onset schizophrenia (COS) (7). In addition, several studies have described 22q11 microdeletions among schizophrenic patients of various ethnic origins (8). These studies collectively suggest that the morbid risk of schizophrenia for a patient with a 22q11 microdeletion may be approximately 20–30 times the general population risk of 1%, and that the rate of 22q11 microdeletions in schizophrenia, although relatively low, may be approximately 80 times the estimated general population rate. It therefore seems likely that the 22q11 region harbors genes that alone, or in combination, are causally implicated in schizophrenia in a certain proportion of patients. The overwhelming majority of the 22q11 deletions are 3 Mb in size, whereas a smaller percentage (≈8%) involve the same proximal breakpoint but a different distal breakpoint resulting in a smaller 1.5-Mb deletion (9, 10). At least one schizophrenic patient has been described as carrying the smaller 22q11 microdeletion, and therefore the "schizophrenia critical region" has been defined to 1.5 Mb (ref. 6). The majority of the genes in the region are known (11). To address the role of individual genes from this chromosomal region, we are undertaking systematic mutational screening of these genes in combination with linkage disequilibrium (LD) studies in family samples (triads) that test for preferential transmission of single-nucleotide polymorphisms (SNPs) and multi-SNP haplotypes from parents to affected individuals.

Materials and Methods

Patient Samples. Detailed information about the adult schizophrenic (AS) sample is provided in ref. 12. The South African sample is part of our ongoing collection of schizophrenia patients of Afrikaner origin and will be described in detail elsewhere (M.K., M. Torrington, C.S., B.R., S.C.H., M.L.B., H. Pretorius, S. Lay, J.A.G., and J.L.R., unpublished work). Probands in both samples met lifetime criteria for Diagnostic and Statistical Manual of Mental Disorders, 4th Ed. (DSM-IV) (13) schizophrenia or schizoaffective disorder. Participants were interviewed by specially trained clinicians by using the Diagnostic Interview for Genetic Studies (DIGS) (14). Detailed information about the COS sample is provided in refs. 7 and 15. All COS probands met unmodified criteria for schizophrenia with onset of psychotic symptoms before their 13th birthday and mean age of onset of psychosis at 10.1 (±1.8 yr). The protocol and the consent forms were approved by the Institutional Review Boards (IRBs) at all participating sites. The National Institute of Mental Health (NIMH) samples have been obtained from the NIMH Human Genetics Inititative dataset (http://zork.wustl.edu/nimh). [Acknowledgment for this sample can be found in the *Appendix*, which is published as supporting information on the PNAS web site (www.pnas.org)].

Abbreviations: LD, linkage disequilibrium; COS, childhood onset schizophrenia; SNP, single-nucleotide polymorphism; AS, adult schizophrenic; DSM-IV, Diagnostic and Statistical Manual of Mental Disorders, fourth edition; NIMH, National Institute of Mental Health; LCR, low copy repeat sequence; RFLP, restriction fragment length polymorphism; HHRR, haplotype-based haplotype relative risk; TDT, Transmission Disequilibrium Test; T/nT, Transmitted/non-Transmitted; CI, confidence interval.

[‡‡]To whom reprint requests should be addressed. E-mail: karayim@mail.rockefeller.edu.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

SNP Identification and Genotyping. Methods (16, 17), primers, and PCR conditions can be found in the *Appendix*, which is published as supporting information on the PNAS web site, www.pnas.org.

Somatic Cell Hybrids. Somatic cell hybrid methods are as described in ref. 6.

Statistical Analysis. Haplotype probabilities for each individual were calculated assuming no recombination between SNPs, summing over ambiguous haplotypes caused by missing phase information or genotype data. Haplotype frequences were estimated by using the Expectation Maximization algorithm (18) simultaneously with the calculation of haplotype probabilities. For the analyses on multilocus SNP haplotypes, trios with ambiguous haplotype reconstructions were used only if there was one reconstruction with a probability of >0.95. Transmissions of single SNPs were analyzed by using both the haplotype-based Haplotype Relative Risk (HHRR) statistic (19) and the Transmission Disequilibrium Test (TDT) (20) (where the TDT considered only transmissions from heterozygous parents). The significance of the TDT analyses was assessed by using a standard 1 degree of freedom x^2 test, and the significance of the HHRR tests was assessed by using Fisher's exact test. Multilocus SNP haplotype transmissions were analyzed by using the HHRR, with the significance being estimated by using Fisher's exact test. Because of the relatively small size of the COS sample and the stratified adult samples, exact *P* values are given only for $P > 0.001$.

Results

Association Between PRODH2/HsPOX2 Locus Variation and Schizophrenia. A first-stage scan of the 22q11 region involved a set of 18 SNPs from nine gene loci distributed across the entire 1.5 Mb (Fig. 1*b*, green shading). The inheritance of these variants by probands affected with schizophrenia from parents was examined in 107 independent triads from the U.S. (AS sample). Analysis of the transmission by using the HHRR statistic (19) revealed a modest nominally significant association between schizophrenia and *PRODH2**1945 allele 2 ($P = 0.04$, uncorrected for multiple testing). The TDT test (20), a test for association in the presence of linkage, provided similar results [($P = 0.035$; Transmitted/non-Transmitted (T/nT), 41/24]. No significant evidence of transmission disequilibrium was observed for any of the other 17 SNPs (Fig. 1*b*).

The *PRODH2* gene encodes for proline dehydrogenase, a mitochondrial enzyme that converts proline to L^1-pyrroline-5-carboxylate and is involved in transfer of redox potential across the mitochondrial membrane (21). The gene is widely expressed in the brain as well as in other tissues (21), consists of 14 coding exons, and is located in the most centromeric part of the 22q11 critical region, where, along with the *DGCR6* gene, it resides within a 75-kb module (previously referred to as sc11.1), which is part of a ≈350-kb low copy repeat (LCR) sequence (LCR-A) (10). This module is duplicated ≈1.5 Mb distally as part of another ≈135-kb LCR (LCR-B) (10). The duplicated copy of the gene, *PRODH2-P* (AC007663.29), has accumulated point mutations and small deletions and insertions, as well as a large deletion that removes exons 2, 3, 4, 5, and part of exon 6 (ref. 11). Reverse transcription–PCR analysis from total cellular RNA extracted from lymphocytes indicates that the duplicated copy is not transcribed and therefore has become a nonexpressed pseudogene (unpublished observations). *PRODH2* and *PRODH2-P* share over 95% sequence similarity.

We sought to follow up the initial finding of association between schizophrenia and the *PRODH2**1945 SNP by, (*i*) identifying additional SNPs from the *PRODH2* locus, (*ii*) better defining the nature of the potential association, and (*iii*) attempting to replicate it in two additional samples. Exons 2–14 and surrounding intronic sequences were examined at this second stage in 48 AS patients, resulting in the identification of five additional SNPs (Fig. 1 *a* and *b*, blue shading). Significant evidence for pairwise LD ($P < 0.01$)

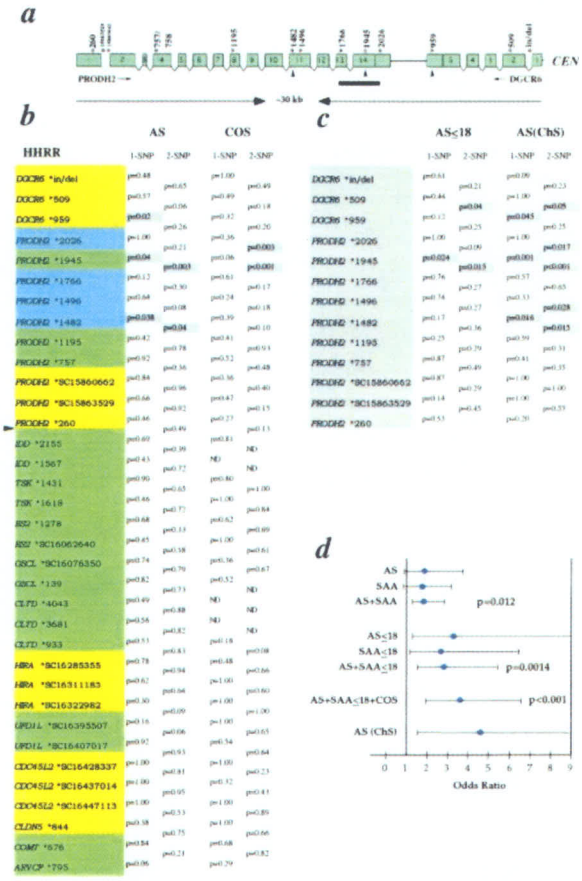


Fig. 1. (*a*) Genomic structure of the *PRODH2/DGCR6* locus. The approximate distribution of the identified polymorphisms in both genes is indicated (*Top*). Polymorphisms demonstrating significant overtransmission in the AS sample are indicated by arrowheads. Arrows indicate direction of transcription. The black horizontal bar signifies the *PRODH2**1766/1945/2026 SNP combination that defines the risk haplotype presented in *d*. The approximate size of the locus is also indicated. CEN, centromere. (*b* and *c*) Linkage disequilibrium results in the AS and COS (*b*), as well as stratified AS families (*c*). In a total of 133 families, transmission to the affected probands of alleles and haplotypes defined by the *PRODH2**1482, *1766, *1945, *2026, and *DGCR6**959 SNPs deviates significantly from expectation under the null hypothesis and provides substantial support for a contribution of the *PRODH2/DGCR6* locus variation to schizophrenia susceptibility. By contrast, no such association was detected in a sample of 80 North American families not afflicted with schizophrenia (not shown), arguing against a generalized transmission distortion at this locus. Color coding reflects the three stages of our analysis (see text). Markers are listed in order from centromere (*Top*) to telomere (*Bottom*). SNPs located in expressed sequences are referenced according to their position in corresponding cDNAs (accession nos. are available in the *Appendix*). Intronic SNPs are referenced according to their position in the University of California, Santa Cruz public database (October 7, 2000, freeze; http://genome.ucsc.edu). Polymorphism *PRODH2**758 is not included in the analysis, because it was in complete LD with *PRODH2**757 (see Fig. 3, which is published as supporting information on the PNAS web site). ND, the polymorphism was not typed in this group. Significant associations are shaded. Arrowhead indicates the relative position of microsatellite marker D22S1638. (*d*) Estimated risk (with 95% CI) for *PRODH2**1766/1945/2026 2–2–1 haplotype relative to the other combinations. For each study, the circle represents the estimated risk, and the line indicates the extent of the 95% CI around this estimate. Data shown for odds ratios are based on haplotype counts (for the case-control study) and the transmission (HHRR) ratios in trios. AS(ChS), subset of AS patients with early childhood deviant behaviors (12); AS < 18, subset of AS patients with DSM-IV disease onset before or at age 18; SAA, South African Afrikaner case/control sample.

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

between the majority of all eight *PRODH2* SNP loci was found over a distance of ≈12 kb by analysis of the 214 unrelated parents of the AS sample. A total of 17 haplotypes were detected, but only 4 haplotypes accounted for ≈75% of all those seen. The most common haplotype (30%) was comprised of the most common alleles of all eight markers (see Fig. 3). Two of the additional variants demonstrated evidence for transmission disequilibrium: *PRODH2**1482 (Asp-4263 Asp) in exon 11 showed a significant 2-fold overtransmission of the less common *PRODH2**1482 allele 2 (HHRR, $P = 0.038$; TDT, $P = 0.025$; T/nT, 30/15), whereas *PRODH2**1766 (Gln-5213 Arg) in exon 13 showed a trend for overtransmission of the most common *PRODH2**1766 allele 2 (HHRR, $P = 0.12$; TDT, $P = 0.07$; T/nT, 6/14). To incorporate information from adjacent SNPs, we analyzed transmission of haplotypes by using a two-locus moving-window strategy (Fig. 1*b*). Of the seven haplotype combinations distributed over the entire gene from exon 2 to the 3′-UTR, only one combination (*PRODH2**1766/1945) showed evidence of transmission disequilibrium more significant than each SNP by itself ($P = 0.003$), primarily because of preferential transmission of the *PRODH2**1766/1945 2–2 haplotype (T/nT, 36/17).

PRODH2 Locus Variation and COS. Given the apparent enrichment of 22q11 microdeletions in patients with COS (7), we sought to replicate the identified association in an independent sample afflicted with COS. Twenty-nine triads (probands with both biological parents) were used in this analysis. In three of these families, COS probands were previously shown to carry a 22q11 microdeletion by fluorescence *in situ* hybridization (7). In addition, polymorphic marker analysis showed that, in all three probands, the *PRODH2* gene is hemizygously deleted (data not shown). Analysis of the inheritance of the *PRODH2* alleles and two-locus haplotypes in the remaining 26 families revealed a pattern of transmission distortion remarkably similar to that observed in the AS sample (Fig. 1*b*). Specifically, and despite the small size of the COS sample, a preferential transmission of *PRODH2**1945 allele 2 to COS probands was observed and provided substantiating evidence for an allelic association (HHRR, $P = 0.06$; TDT, $P = 0.03$; T/nT, 11/3). The *PRODH2**1766/1945 combination showed again significant evidence for transmission disequilibrium ($P < 0.001$) because of preferential transmission of the 2–2 haplotype (T/nT, 10/0). Transmission of the adjacent combination also deviated significantly from neutral ($P = 0.003$) because of overtransmission of the *PRODH2**1945/2026 2–1 haplotype (T/nT, 11/1). No other two-SNP haplotypes provided significant association results.

PRODH2 Locus Variation and Schizophrenia with Early Onset Features. Early behavioral and cognitive abnormalities are a fundamental component of the 22q11 microdeletion syndrome and are present more consistently than the physical features of the syndrome (such as palatal abnormalities and heart defects) among schizophrenic patients carrying the 22q11 microdeletions (22). On the basis of these observations and the apparent enrichment of 22q11 microdeletions in patients with COS (7), we reasoned that the contribution of the *PRODH2* locus to schizophrenia susceptibility may be more significant among the subgroup of our AS patients with early disease onset or early deviant behaviors. To address this possibility, we stratified the AS families according to two criteria: First, by age at which full diagnostic criteria for schizophrenia according to the DSM-IV (13) are met. We set the cutoff point before or equal to age 18. Because age at clinical syndrome onset might not be an accurate indicator of the biological onset of the disease, we also stratified the families by history of early deviant behaviors occurring several years before the onset of the syndrome (before age 10) (ref. 12). The overall significant association in the entire AS sample justified further HHRR analysis including only AS probands stratified according to early onset profiles. Despite the decrease in sample size, this stratified analysis provided significant evidence for association (Fig. 1*c*). In the subset of AS probands with early childhood deviant behavioral profiles, transmission of the three-marker haplotype *PRODH2**1766/1945/2026 2–2-1 was associated with an odds ratio (approximate relative risk) of 4.6 [95% confidence interval (CI), 1.4–14.2], a 2.4-fold increase compared with the entire AS sample. When families were stratified according to age of DSM-IV schizophrenia onset (AS < 18), transmission of the 2–2-1 risk haplotype was also associated with an increase in the relative risk to 3.3 (95% CI, 1.1–9.8) (Fig. 1*d*).

Delineation of the Associated Region and Analysis of the Contribution of DGCR6. Evidence for association is weak or absent at the 5′ half of the *PRODH2* gene. By screening a sample of 18 patient carriers of the risk haplotype (Fig. 3), we identified three additional SNPs from the 5′ end of the gene [*PRODH2**260 (Pro-193 Gln) from coding exon 1, *PRODH2**SC15863529 and *SC15860662 from the adjacent intron; Fig. 1 *a* and *b*, yellow shading] and typed them in the AS and COS families. Although one of these variants (*SC15860662) showed strong disequilibrium with *PRODH2**1945 ($P < 10^{-4}$), none yielded any evidence for association. We also typed both family samples for the D22S1638 microsatellite marker, which maps ≈50 kb distal to *PRODH2*. This marker is not in LD with *PRODH2**1945 ($P = 0.59$), and it did not yield any evidence for association (data not shown). DNA sequence analysis (not shown), as well as a search of public databases (http://genome.ucsc.edu), indicates that the ≈100-kb region immediately distal to the *PRODH2* gene is devoid of functional genes (a HERV-K101 provirus and several expressed sequence tags (ESTs) representing noncoding mRNAs, likely originating from expressed pseudogenes, have been mapped in the region between *PRODH2* and *IDD*; ref. 11 and data not shown). By contrast, the adjacent region, from *IDD* to *CLDN5* (spanning 488 kb) is highly gene-rich (11). In addition to the 13 SNPs from this interval tested in the first stage of our screening, we identified and typed seven additional SNPs (Fig. 1*b*, yellow shading). Analysis of transmission of all 20 individual SNPs and two-SNP combinations did not provide any significant results in both AS and COS samples. Therefore, this analysis does not support the region distal to the *PRODH2* as the source of the observed association.

Evidence for transmission disequilibrium appears to be stronger at the 3′ end of the *PRODH2* gene. This region is particularly interesting, because it contains the *DGCR6* gene, which encodes for a product that shares homology with the *Drosophila melanogaster* gonadal protein and has weak characteristics of a cell adhesion molecule (23). The gene consists of at least five exons distributed over a region of ≈5 kb in LCR-A (10), well within the LD range for the identified *PRODH2* risk haplotype (the two genes share an ≈1.5 kb 3′ intergenic region; Fig. 1*a*). A duplicated copy of the *DGCR6* gene [designated *DGCR6*-like (L) (AC008103.22)] exists in LCR-B adjacent to *PRODH2-P*. *DGCR6-L* shares over 95% sequence similarity with *DGCR6* and has accumulated several point mutations. At least five of them result in missense changes (not shown), but none introduces a premature stop codon. Analysis of EST databases indicates that *DGCR6-L* is transcribed in a variety of tissues, and it is unclear at present whether it represents a nonfunctional pseudogene. We scanned the ≈1.5-kb intergenic region, as well as all exons and flanking intronic sequences of the *DGCR6* gene in 18 patient carriers of the risk haplotype, for common sequence variations. This analysis identified two SNPs (*DGCR6**509/G3 A, *DGCR6**959/C3 T) and one intronic 9-bp insertion/deletion (in/del) polymorphism (*DGCR6**in/del) 9 bp from the splice donor site of exon 1 (Fig. 1*b*, yellow shading). Allelic frequencies ranged from 3.9 to 12.3%. Not surprisingly, both common variants (*DGCR6**959, *DGCR6**in/del) were in LD with each other (not shown) and also in strong LD with *PRODH2**1945 ($P < 10^{-6}$). Only the rare allele (T, 12.3%) and associated haplotypes of one of these variants (*DGCR6**959, located at the 3′ UTR of the *DGCR6* gene, 1,267 bp proximal to *PRODH2**1945; Fig. 1*a*)

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

demonstrated evidence for association (Fig. 1*b*). Overall, this stage of our analysis extended the associated region to include the 3 end of the *DGCR6* gene and delineated the range of significant disequilibrium within a segment in which *PRODH2* and *DGCR6* are the only known genes.

Genetic Variation at the PRODH2 Locus Presents an Unusual Pattern. In the course of this work, we noticed that the sequence of the infrequent allele of the exon 11 SNP *PRODH2**1496 (a1 = A, 11%, Fig. 3) is identical to the published pseudogene sequence. To test the possibility that the *PRODH2-P* pseudogene can serve as a reservoir of mutations (24), we aligned the gene/pseudogene exonic sequences (exons 7–14) and initially assayed for a subset of differences, namely those detected by simple PCR-RFLP assays predicted to result in missense changes (see Table 3, which is published as supporting information on the PNAS web site).

We analyzed genomic DNA from AS (n = 84–106) and COS (n = 41) samples, and several affected individuals were discovered to be heterozygous carriers of pseudogene-like single base-pair substitutions or clusters of them (Fig. 2 *a* and *b* and Table 1). Interestingly, the majority of the identified pseudogene-like alleles were located in exon 11 of the *PRODH2* gene where the originally identified *PRODH2**1496 polymorphism resides. In all cases where parents were available for genotyping, we confirmed inheritance of pseudogene-like alleles from parental chromosomes, and no *de novo* changes were identified. In addition, several control experi- ments confirmed that the observed changes are not the result of artifactual hybrid amplicons. First, control PCR amplification and sequencing of the equivalent of exon 11 from all affected carriers by using *PRODH2-P* (pseudogene)-specific primers produced a sequence reflecting the entire set of multiple substitutions expected by the published pseudogene sequence. Second, we used reverse transcription–PCR on total RNA isolated from available trans- formed lymphocytes to amplify exon 11 sequences by using primers located in flanking exons (data not shown). In all cases tested, the presence of single or multiple adjacent pseudogene-like substitu- tions was confirmed in the cDNA where *PRODH2-P* is not repre- sented. Third, in one case, we confirmed the presence of a pseu- dogene-like allele in the paternal chromosome of a heterozygous carrier (haplotype D; see below) by qualitative Southern blot hybridization on DNA isolated from hamster– human somatic cell hybrids retaining only one of the two chromosome 22 copies (Fig. 2*c*). In a similar manner, we also aligned the coding sequences from both copies of the *DGCR6* gene and checked *DGCR6*, by PCR-RFLP assays, for four of the five differences present in the *DGCR6-L* predicted to result in missense changes (see Table 3). We analyzed the same AS and COS samples, but no affected individuals were discovered to carry the predicted mutations.

We determined *PRODH2* exon 11 pseudogene-like haplotypes by sequencing the entire exon from affected carriers and their parents (when available). Seven distinct haplotypes were identified carrying single or clusters of multiple adjacent substitutions (Fig. 2*b*). Conversion of proline to L^1-pyrroline-5-carboxylate is a rate-limiting step, and therefore the *PRODH2* gene can be haploinsufficient, in the sense that heterozygous carriers of debilitating mutations may demonstrate plasma proline accumulation (25). For practical reasons, it was not feasible to test the plasma proline levels of all affected carriers. Nevertheless, in the small number of heterozygous carriers tested, plasma proline levels were at the higher end of the normal distribution. The plasma proline level of three adult female carriers of exon 11 haplotypes A, B, and E, respectively, was ≈1.5- to 2-fold higher when compared with the mean value of sex- and age-matched noncarrier controls (Fig. 2*d*). However, it should be emphasized that it is unclear at present whether the plasma proline levels accurately reflect the effect of the identified mutations on the function of the PRODH2 enzyme and the local concentration of proline in brain subregions, given the high-level expression of a homologous proline dehydrogenase

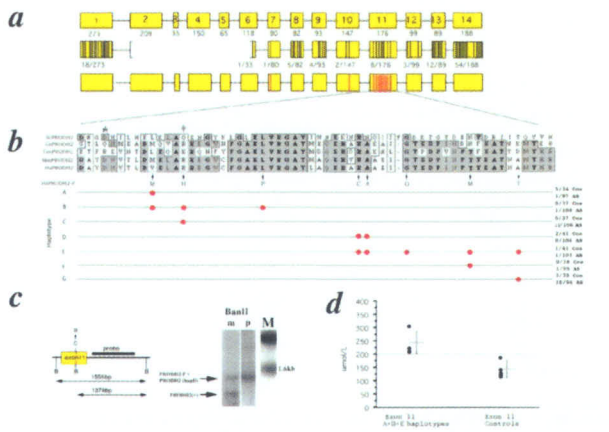


Fig. 2. (*a*) Genomic structure of the *PRODH2* (*Top* and *Bottom*) and *PRODH2-P* locus (*Middle*); colored boxes indicate predicted coding exons; brackets indicate a large deletion that removes predicted exons 2, 3, 4, 5, and part of 6 from *PRODH2-P*. Introns are not to scale. The size of the *PRODH2* exons (in nucleotides), as well as the number of exonic differences between *PRODH2-P* and *PRODH2*, is indicated. Black vertical lines indicate the approximate distribution of these differences. *PRODH2* exons 7, 9, 10, 11, 13, and 14 were assayed for presence of a subset of these differences (n = 12; see Table 3) in the AS and COS patient samples. Red vertical lines (*Bottom*) indicate pseudogene-like substitutions detected in both samples. (*b Top*) Cross-species comparison of the *Homo sapiens PRODH2* exon 11 predicted protein sequence by using the program CLUSTAL-W (as implemented in MacVector, Accelrys, San Diego). Boxed and shaded areas highlight identities or similarities. Bottom row indicates the predicted effect of pseudogene-like base changes on the exon 11 encoded protein sequence. * and † indicate the codons where the synonymous SNP *PRODH2**1482 and the nonsynonymous SNP *PRODH2**1496 reside, respectively. (*Bottom*) Exon 11 haplotypes determined by sequencing the exon from affected carriers and their parents. Seven distinct haplotypes were identified carrying single or clusters of multiple adjacent pseudogene-like substitutions. Filled circles indicate codons carrying pseudogene-like, synonymous (n = 2), and nonsynonymous (n = 6) base-pair substitutions. Solid lines indicate regions of identity to the *PRODH2* reference sequence, shared by all haplotypes. COS patients from all ethnic backgrounds (21 European Americans, 12 African Americans, 8 of other ethnic background) were included in this analysis, whereas only European and African Americans were included in the analysis described in Table 1. (*c*) Southern blot hybridization on DNA isolated from hamster–human somatic cell hybrids retaining only one of the two chromosome 22 copies from a heterozygous carrier of haplotype D (m, maternal copy; p, paternal copy). The Arg-453Cys substitution, present in the *PRODH2-P* and the paternal copy of *PRODH2* in this patient, is detected by a *Ban*II RFLP (*Ban*II site absent). Genomic DNA was digested with *Ban*II restriction endonuclease, and a 1-kb intronic fragment, indicated by a solid line, was used as a probe. (M, 1-kb ladder DNA marker; B, *Ban*II). (*d*) Distribution of plasma proline level of three adult female carriers of exon 11 haplotypes A, B, and E and five sex- and age-matched noncarriers controls. Blood was drawn from the subjects in the morning, after overnight fasting. Proline levels were determined by the same qualified diagnostic laboratory (Mayo Clinic, Rochester, MN). Horizontal lines denote mean values and vertical lines, standard errors.

(*PRODH1* or *HsPOX1*) (U80018.1) in peripheral tissues but not in the brain.

Transmission and Distribution of Pseudogene-Like Alleles in Patients. In addition to *PRODH2**1496, only one (Ala-4723 Thr) of the eight missense pseudogene-like variants identified was sufficiently prevalent in the combined AS and COS sample to be used in family-based association studies. Analysis of the transmission of the Ala-4723 Thr variant by using the TDT statistic revealed a modest but significant association between schizophrenia and the rare pseudogene-like allele (P = 0.03; T/nT, 18/7:13/6 in AS, 5/1 in COS). In addition to the preferential inheritance, an enrichment of the Thr-472 substitution is observed in the European American AS (5.8%) and COS (7.1%) probands compared with a European

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Table 1. Distribution and Inheritance of *PRODH2* pseudogene-like substitutions

		European American			African American			AS + COS triads
Exon no.	Amino acid change	AS Sample	COS Sample	Controls	NIMH/AS Sample	COS Sample	Controls	T/nT
7	Leu-289 → Met	1/103	0/21	3/93	ND	0/12	ND	ND
10	Pro-406 → Leu	1/84	0/21	0/108	ND	1/12	0/37	ND
11	Val-427 → Met	1/97	1/21	4/108 (1.8%)	7/36 (11%)*	4/12 (16.6%)†	1/37 (1.3%)	ND
11	Arg-431 → His	ND (see Fig. 1)						
11	Leu-441 → Pro	0/105	0/21	ND	2/36	0/12	0/37	ND
11	Arg-453 → Cys	1/102 (0.5%)	3/21 (7.1%)‡	1/115 (0.4%)	ND	0/12	ND	ND
11	Thr-466 → Met	2/95	1/21	0/78	ND	0/12	ND	ND
11	Ala-472 → Thr	11/95 (5.8%)§	3/21 (7.1%)¶	4/93 (2.1%)	ND	0/12	ND	18/7

Number of carriers over the total number of patients tested for each substitution is shown. The NIMH/AS sample includes 35 schizophrenia probands of African American descent from the NIMH Human Genetics Initiative dataset and the one African American proband from the AS sample. Allelic frequencies are given in parentheses. ND, not done. *, $P = 0.1$; †, $P = 0.01$; ‡, $P = 0.015$; §, $P = 0.048$; ¶, $P = 0.1$; , $P = 0.03$.

American control sample (2.1%; $P = 0.048$ and 0.1, respectively) (Table 1). The Ala-4723 Thr variant is not in disequilibrium with *PRODH2**1945 ($P = 0.124$), indicating an independent association with schizophrenia susceptibility. The distribution of the other rare variants was compared between cases and two racially matched U.S. control samples. The Arg-4533 Cys change that affects a highly conserved amino acid seems to be enriched in the European American COS sample (7.1%) compared with both the AS sample (0.5%; $P = 0.02$) and the European American control sample (0.4%; $P = 0.015$). The Thr-4663 Met change is present in two AS patients and one COS patient from the European American sample and absent from all of the control chromosomes tested.

The allelic frequency of the Val-4273 Met change is significantly higher in the African American subset of the COS sample (16.6%) compared with both an African American control sample (1.3%; $P = 0.01$) and the European American control sample (1.8%; $P = 0.008$). There is no significant difference in the distribution of the Val-4273 Met change between the two control samples ($P > 0.7$). Interestingly, the Val-4273 Met is present in the only African American patient of the AS sample. Further sequencing revealed that this patient also carries a Leu-4413 Pro substitution (Fig. 2*b*, haplotype B) not present in any chromosome in the control African American sample. Leu-441 is a highly conserved amino acid, and the Leu-4413 Pro mutation is predicted to result in a protein with drastically reduced enzymatic activity. The apparent enrichment of these two variants among African American patients was investigated further by genotyping 35 independent African American probands extracted from the NIMH Human Genetics Initiative collection of families afflicted with schizophrenia. We identified six additional patient carriers of the Val-4273 Met variant (a total of 7/36 as opposed to 1/37 controls; $P = 0.1$) and one additional patient carrier of the Leu-4413 Pro variant.

Finally, one additional rare pseudogene-like Pro-4063 Leu substitution, which affects a relatively conserved Pro residue at exon 10, was identified in one European American adult and one African American COS patient and was not detected in any control subject.

Replication of the Association Between PRODH2 Locus Variation and Schizophrenia in a Third Sample. We sought further confirmation of all these results in a third independent sample of 109 unrelated individuals affected with schizophrenia and 75 healthy controls (with no family history of a major psychiatric disorder) of South African Afrikaner origin. We first genotyped *PRODH2**1766, *1945, *2026 SNPs, estimated haplotype frequencies for the three-SNP combinations for cases and controls separately, and compared the distribution of the 2–2-1 risk haplotype in chromosomes of affecteds and controls (Table 2). This follow-up sample demonstrated a strong nonsignificant trend in the same direction and similar odds ratios with the U.S. AS sample (Table 2 and Fig. 1*d*). The frequency of the 2–2-1 haplotype was 9% among control chromosomes, very similar to the frequency observed in the untransmitted chromosomes of the U.S. family sample (8.8%). This frequency increased to 15% among the chromosomes of affecteds, corresponding to a relative risk of 1.7 [95% CI, 0.9–3.3] ($P = 0.055$, one-tailed). Most importantly, and consistent with the pattern observed in the U.S. sample, when only the subset of affecteds with age at onset <18 years was considered ($n = 21$), the frequency of the associated haplotype increased to 20%, corresponding to a risk of 2.6 (95% CI, 1.05–6.6) ($P = 0.01$, one-tailed). Only three haplotypes were observed (2–1-1, 2–2-1, and 1–2-2) in this sample, and no other haplotype showed evidence of association. Finally, we compared the distribution in case and control chromosomes of the three exon 11 pseudogene-like variants that were apparently enriched among the European American subset of our U.S. samples. One of these variants (Thr-4663 Met) was not present in either case or control chromosomes. However, the other two variants (Arg-4533 Cys and Ala-4723 Thr) showed a striking enrichment among the disease chromosomes (Table 2). Interestingly, eight of

Table 2. Replication of association in the SAA sample

	*PRODH2**1766		*PRODH2**1945		*PRODH2**2026		*PRODH2**1766/1945/2026		
Allele	1	2	1	2	1	2	221	Other	
SAA	9	209	177	41	209	9	32 (15%)	184	$P = 0.055$*
SAA < 18	2	40	31	11	40	2	9 (20%)	33	$P = 0.01$†
SA controls	5	145	131	19	145	5	14 (9%)	136	
	Arg-453 → Cys				Thr-466 → Met		Ala-472 → Thr		
SAA	5/75 (3.3%)		$P = 0.04$		0/75 (0%)		2/75 (8%)		$P = 0.008$
SA controls	0/50 (0%)				0/50 (0%)		1/50 (1%)		

Number of carriers over the total number of patients tested is shown for pseudogene-like substitutions; one patient carried both variants. Eight of sixteen carriers had age at disease onset <18 yr. Allelic frequencies are given in parentheses. *P* values are one-tailed. SAA, South African Afrikaner.
*OR = 1.7 (95% CI, 0.9–3.3); †OR = 2.6 (95% CI, 1.05–6.6).

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

the sixteen carriers of either variant (50%) had an early age at disease onset (<18), a 2.5-fold increase over the observed rate of early onset cases in this sample ($P = 0.05$). Moreover, none of the two variants was found in the context of the 2–2-1 risk haplotype, indicating an independent association with schizophrenia susceptibility in this population as well.

Discussion

Using LD analysis in three independent nondeleted patient samples, we provide here strong evidence for an association between variation at the 22q11 locus and schizophrenia with early onset features.

Through fine-scale mapping in heterogeneous U.S. family samples, we delineated the region of significant LD within a segment in which *PRODH2* and *DGCR6* are the only known genes, and we identified a relatively common haplotype where the schizophrenia susceptibility variant(s) likely reside. Until the functional consequences of the associated variation are established in brain samples, our LD studies alone cannot unequivocally distinguish between a contribution from the *PRODH2* gene itself and/or from the neighboring *DGCR6* gene.

Further evidence implicating the *PRODH2* gene itself is provided by the identification of rare pseudogene-like structural variants of the *PRODH2* gene, especially in cases with early onset of the disease. No such variants were identified in the *DGCR6* gene. Several of the variants identified result in missense changes at residues conserved to a different extent among species. Observation of clustering of point mutations suggests that at least some of the pseudogene-like mutations arose as the result of gene conversion events between the *PRODH2* gene and the homologous nonfunctional linked copy (24). Exon 11 of the *PRODH2* gene seems to be a "hot spot" for microconversion events, presumably because of the proximity of a strand exchange- promoting sequence (24). Analysis of pseudogene-like haplotypes at this exon indicates that sequence exchanges between the two copies often involve conversion tracts with a minimum size of 18–57 bp. Identification of pseudogene-like alleles may have additional implications. First, it is of interest that recombination between LCR-A (the *PRODH2* locus) and LCR-B (the *PRODH2-P* locus) results in a 1.5-Mb deletion, which accounts for ≈8% of the total 22q11 deletions (9, 10). This rearrangement may involve intrachromosomal pairing between the duplicated modules and formation of a "stem–loop" intermediate (9, 10). Therefore, a possible relation between the apparent gene conversion observed in this study and the LCR-A/B-mediated recombination warrants further study. Second, in relation to the origin of mutations in schizophrenia (and other common complex psychiatric disorders), our results implicate the dynamic instability of the genome and the resulting processes of gene duplication and generation of nonfunctional pseudogenes. Pseudogenes may, in turn, serve as a constant and renewable reservoir of variation that can be effectively transferred to linked susceptibility genes via gene conversion, thus partly counteracting the negative effect on reproduction (26).

How does genetic variation in the *PRODH2/DGCR6* locus affect susceptibility to schizophrenia? Some clues are provided by our recent studies of a PRODH-deficient animal model. This analysis was initiated soon after our first positive genetic findings with *PRODH2* to determine the relevance of this gene to the psychiatric phenotype under study. Specifically, comparison of 42 mice homozygous for a truncated form of PRODH with ≈7-fold decreased activity and 26 wild-type littermates revealed a deficit in prepulse inhibition, a measure of sensorimotor gating, a central inhibitory/filtering mechanism affected in patients with psychiatric disorders including schizophrenia (21). Proline itself may serve as a direct modulator of glutamatergic transmission in the brain, a role suggested primarily by the selective expression of a brain-specific high-affinity proline transporter (SLC6A7) in a subset of glutamatergic synapses (27) and supported by neurochemical and electrophysiological studies (21). An alternative model is suggested by recent observations that implicate *PRODH2* in apoptosis (28).

In the present study, we typed one to three SNPs from 13 of the 23 known genes mapping in the entire 22q11 critical region (Fig. 1*b*). The genes examined represent 85% of the known genes located at the proximal half of the critical region, within 780 kb of the *PRODH2/DGCR6* locus. Analysis of the remaining genes is ongoing, with the ultimate goal of examining individual contributions, as well as possible interactions, of all known genes that map in this region.

We are deeply grateful to all of the families who have participated in this research. We thank Alyson Knowles, Lubomira Malinova, and Shoshanna Friedman for technical assistance, Diana Hall for help with the LD calculations, and Markus Stoffel for DNA samples from an African American control group. Support for this work was provided in part by grants from the New York Academy of Medicine, the EJLB Foundation, and the Essel Foundation (M.K.), as well as the Norman and Rosita Winston Foundation (postdoctoral fellowship to H.L.). Clinical work at the Rockefeller University Hospital General Clinical Research Center was supported in part by Grant M01 RR00102.

1. Scambler, P. J. (2000) *Hum. Mol. Genet.* **9,** 2421–2426.
2. Golding-Kushner, K. J., Weller, G. & Shprintzen, R. J. (1985) *J. Cranio-facial Genet. Dev. Biol.* **5,** 259–266.
3. Baron, M. (2001) *Am. J. Hum. Genet.* **68,** 299–312.
4. Pulver, A. E., Nestadt, G., Goldberg, R., Shprintzen, R. J., Lamacz, M., Wolyniec, P. S., Morrow, B., Karayiorgou, M., Antonarakis, S. E., Housman, D. & Kucherlapati, R. (1994) *J. Nerv. Mental Dis.* **182,** 476–478.
5. Murphy, K. C., Jones, L. A. & Owen, M. J. (1999) *Arch. Gen. Psychiatry* **56,** 940–945.
6. Karayiorgou, M., Morris, M. A., Morrow, B., Shprintzen, R. J., Goldberg, R., Borrow, J., Gos, A., Nestadt, G., Wolyniec, P. S., Lasseter, V. K., *et al.* (1995) *Proc. Natl. Acad. Sci. USA* **92,** 7612–7616.
7. Usiskin, S. I., Nicolson, R., Krasnewich, D. M., Yan, W., Lenane, M., Wudarsky, M., Hamburger, S. D. & Rapoport, J. L. (1999) *J. Am. Acad. Child Adolesc. Psychiatry* **38,** 1536–1543.
8. Sugama, S., Namihira, T., Matsuoka, R., Taira, N., Eto, Y. & Maekawa, K. (1999) *J. Neurol. Neurosurg. Psychiatry* **67,** 803–806.
9. Edelmann, L., Pandita, R. K., Spiteri, E., Funke, B., Goldberg, R., Palanisamy, N., Chaganti, R. S., Magenis, E., Shprintzen, R. J. & Morrow, B. E. (1999) *Hum. Mol. Genet.* **8,** 1157–1167.
10. Shaikh, T. H., Kurahashi, H., Saitta, S. C., O'Hare, A. M., Hu, P., Roe, B. A., Driscoll, D. A., McDonald-McGinn, D. M., Zackai, E. H., Budarf, M. L., *et al.* (2000) *Hum. Mol. Genet.* **9,** 489–501.
11. Dunham, I., Shimizu, N., Roe, B. A., Chissoe, S., Hunt, A. R., Collins, J. E., Bruskiewich, R., Beare, D. M., Clamp, M., Smink, L. J., *et al.* (1999) *Nature (London)* **402,** 489–495.
12. Sobin, C., Blundell, M. L., Conry, A., Weiller, F., Gavigan, C., Haiman, C. & Karayiorgou, M. (2001) *Psychiatry Res.* **101,** 101–113.
13. American Psychiatric Association (1994) *Diagnostic and Statistical Manual* (Am. Psychiatric Assoc., Washington, DC), 4th Ed.
14. Nurnberger J. I., Jr., Blehar, M. C., Kaufmann, C. A., York-Cooler, C., Simpson, S. G., Harkavy-Friedman, J., Severe, J. B., Malaspina, D. & Reich, T. (1994) *Arch. Gen. Psychiatry* **51,** 849–859.
15. Nicolson, R., Lenane, M., Singaracharlu, S., Malaspina, D., Giedd, J. N., Hamburger, S. D., Gochman, P., Bedwell, J., Thaker, G. K., Fernandez, T., *et al.* (2000) *Am. J. Psychiatry* **157,** 794–800.
16. Underhill, P. A., Jin, L., Lin, A. A., Mehdi, S. Q., Jenkins, T., Vollrath, D., Davis, R. W., Cavalli-Sforza, L. L. & Oefner, P. J. (1997) *Genome Res.* **7,** 996–1005.
17. Chen, X., Levine, L. & Kwok, P. Y. (1999) *Genome Res.* **9,** 492–498.
18. Dempster, A. P., Laird, N. M. & Rubin, D. M. (1997) *J. R. Stat. Soc.* **39B,** 1–38.
19. Terwilliger, J. & Ott, J. (1992) *Hum. Hered.* **42,** 337–346.
20. Spielman, R. S., McGinnis, R. E. & Ewens, W. J. (1993) *Am. J. Hum. Genet.* **52,** 506–516.
21. Gogos, J. A., Santha, M., Takacs, Z., Beck, K. D., Luine, V., Lucas, L. R., Nadler, J. V. & Karayiorgou, M. (1999) *Nat. Genet.* **21,** 434–439.
22. Bassett, A. S., Hodgkinson, K., Chow, E. W., Correia, S., Scutt, L. E. & Weksberg, R. (1998) *Am. J. Med. Genet.* **81,** 328–337.
23. Demczuk, S., Thomas, G. & Aurias, A. (1996) *Hum. Mol. Genet.* **5,** 633–638.
24. Collier, S., Tassabehji, M., Sinnott, P. & Strachan, T. (1993) *Nat. Genet.* **3,** 260–265.
25. Phang, J. M., Yeh, G. C. & Scriver, C. R. (1995) in *The Metabolic and Molecular Bases of Inherited Disease*, eds. Scriver, C. R., Beaudet, A. L., Sly, W. S. & Valle, D. (McGraw–Hill, New York), Vol. 1, pp. 1125–1146.
26. Erlenmeyer-Kimling, L. (1978) in *Annual Review of the Schizophrenic Syndrome*, ed. Cancro, R. (Brunner/Mazel, New York), Vol. 5, pp. 298–333.
27. Renick, S. E., Kleven, D. T., Chan, J., Stenius, K., Milner, T. A., Pickel, V. M. & Fremeau, R. T., Jr. (1999) *J. Neurosci.* **19,** 21–33.
28. Maxwell, S. A. & Davis, G. E. (2000) *Proc. Natl. Acad. Sci. USA* **97,** 13009–13014.

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Letter

Extended Intermarker Linkage Disequilibrium in the Afrikaners

Diana Hall,[1] Ellen M. Wijsman,[2] J. Louw Roos,[3] Joseph A. Gogos,[4] and Maria Karayiorgou[1,5]

[1]*The Rockefeller University, Human Neurogenetics Laboratory, New York, New York 10021, USA;* [2]*University of Washington, Division of Medical Genetics, Department of Medicine; and Department of Biostatistics, Seattle, Washington 98195, USA;* [3]*University of Pretoria Department of Psychiatry & Weskoppies Hospital, Pretoria, 0083, Republic of South Africa;* [4]*Columbia University, College of Physicians & Surgeons, Department of Physiology and Cellular Biophysics and Center for Neurobiology & Behavior, New York, New York 10032, USA.*

In this study we conducted an investigation of the background level of linkage disequilibrium (LD) in the Afrikaner population to evaluate the appropriateness of this genetic isolate for mapping complex traits. We analyzed intermarker LD in 62 nuclear families using microsatellite markers covering extended chromosomal regions. The markers were selected to allow the first direct comparison of long-range LD in the Afrikaners to LD in other demographic groups. Using several statistical measures, we find significant evidence for LD in the Afrikaners extending remarkably over a 6-cM range. In contrast, LD decays significantly beyond 3-cM distances in the other founder and outbred populations examined. This study strongly supports the appropriateness of the Afrikaner population for genome-wide scans that exploit LD to map common, multigenic disorders.

The positional cloning of genes underlying common diseases is one of the greatest challenges in biomedical research. However, traditional analysis of the cosegregation of disease and genetic markers within pedigrees (linkage analysis) shows limited power when complex inheritance patterns are present. A number of approaches that exploit linkage disequilibrium (LD) have been proposed to circumvent this problem, including population-based allelic association methods or joint analysis of linkage and allelic association (Jorde 1995; Risch and Merikangas 1996; Freimer et al. 1997). Use of families from founder populations is expected to facilitate both designs. The measured LD level between a marker and a disease allele is determined by several elements, including the distance between the marker and the disease gene, the age of the disease allele, the frequency of the linked marker allele, and other less predictable forces such as genetic drift, rate of population expansion, genetic bottlenecks, and admixture. As expected theoretically and observed empirically, LD decays as distance increases. However, several recent studies indicate that within a short distance range, the contribution of recombination to the level of LD is negligible compared with the other forces (Shifman and Darvasi 2001). As distance increases, however, the increased recombination frequency is expected to erode any association and therefore recombination and distance become the primary elements that determine the level of LD.

Patients from founder populations likely inherited, from a common ancestor, segments of chromosomes containing a disease gene and surrounding marker loci for which a shared haplotype may be identified. The relatively recent origins of a certain founder population make it likely that chromosomal regions identical by descent surrounding a disease allele will be larger than in outbred populations. Therefore, optimally designed genome-wide scans that exploit LD and involve founder populations of recent ancestry should require significantly fewer markers or significantly smaller sample size than similar scans in outbred populations (Chapman and Wijsman 1998). Many questions still remain open regarding the optimal design of such genome-wide mapping studies, including the appropriate demographic history of the founder population, the appropriate spacing of genetic markers, and the appropriate thresholds for significance testing. Addressing these issues depends partly on understanding the forces governing the distribution of background LD (bLD) in the genome. In the present study, we examined the distribution of bLD in a set of Afrikaner families from Tshwane (formerly known as Pretoria), Limpopo (formerly Northern Transvaal), and Mpumalanga (formerly the Eastern Transvaal) that were collected for a genetic study of schizophrenia.

The Afrikaners, who total approximately 3 million, descended from a small number of original settlers (who soon became known as "Boere", or farmers). In 1652, the first immigrants (just >1000), primarily of Dutch origin, settled in the Cape. They later spread inland, among other places, to Tshwane and Limpopo. The communities that were eventually established were geographically isolated. Cultural considerations, including language differences (Afrikaans language derived from Dutch) and religious practices (most Afrikaners were members of the Dutch Reformed Church), further maintained the isolation. Consanguinity was common, especially in early generations. Although in more recent years some admixture has occurred, the population growth over 13–15 generations was almost entirely through reproduction, as immigration subsequent to the founding was minimal. The demographic history of this population is reflected in the unusually high frequency of certain rare Mendelian disorders, the unusually low diversity of the associated allelic variants, and the unusually large extent (8–11 cM) of conserved haplotypes

[5]**Corresponding author.**
E-MAIL karayim@mail.rockefeller.edu; FAX (212) 327-7329.
Article and publication are at http://www.genome.org/cgi/doi/10.1101/gr.136202.

12:956–961 ©2002 by Cold Spring Harbor Laboratory Press ISSN 1088-9051/01 $5.00; www.genome.org

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Table 1. Features and Relative Locations of Microsatellite Markers

Marker	Distance to the next marker (Kosambi cM)[a]	Distance to the next marker (Mb)[b]	Heterozygosity observed
D18S851	0.4	0.1	0.70
D18S484	0.6	n.r.	0.73
AFM318xd5	0.2	n.r.	0.40
D18S1156	1.8	1.1	0.59
D18S487	1.8	1.8	0.75
D18S35	0.4	0.4	0.68
D18S69	0.6	0.7	0.76
D18S39	0.7	n.r.	0.78
D18S41			0.58
D4S425	1.1	0.7	0.72
D4S404	5.6	2.6	0.77
D4S1091	0.9	1.2	0.41
D4S391	1.1	n.r.	0.82
D4S230	0.0	n.r.	0.87
D4S418	1.3	0.2	0.77
D4S1618	55	66	0.54
D4S423	0.0	0.7	0.86
D4S410	0.0	1.1	0.50
D4S414			0.87

[a]Data from Eaves et al. 2000 for chromosome 18 map; Marshfield database for chromosome 4.
[b]Human Genome Working Draft-Santa Cruz; April 1, 2001, freeze.
n.r., not reported.

around disease genes (Hayden et al. 1980; Brink et al. 1987; Rosendorff et al. 1987; Leitersdorf et al. 1989; Torrington and Viljoen 1991; Brink et al. 1995; Pronk et al. 1995; Goldman et al. 1996; Warnich et al. 1996; Groenewald et al. 1998; Roby et al. 1999).

We wished to compare patterns of LD in the Afrikaners with other founder populations, as well as outbred populations. To assess the level of bLD in the Afrikaner population, we used a total of 62 families collected from Tshwane, Limpopo, and Mpumalanga as part of an ongoing genetic study of schizophrenia.

Table 2. Associations Observed Between Pairs of Loci

Locus 1	Locus 2	FET p value < .01	cM
AFM318xd5	D18S1156	.0001*	0.2
D18S851	D18S484	<.00001*	0.4
D18S69	D18S39	<.00001*	0.6
D18S39	D18S41	<.00001*	0.7
D18S851	D18S1156	<.00001*	1.2
D18S69	D18S41	<.00001*	1.3
D18S35	D18S41	.0022	1.7
D18S1156	D18S487	.0088	1.8
AFM318xd5	D18S487	.0090	2
D18S487	D18S69	.0094	2.2
D18S487	D18S39	<.00001*	2.8
D18S487	D18S41	<.00001*	3.5
D18S1156	D18S35	<.00001*	3.6
D18S1156	D18S69	<.00001*	4
D18S1156	D18S39	<.00001*	4.6
D18S851	D18S69	.0096	5.2
D18S1156	D18S41	<.00001*	5.3
AFM318xd5	D18S41	<.00001*	5.5

Shown are locus associations detected using the Fisher exact test (FET).
The reported distances were computed using genetic maps indicated in Table 1.
*Locus pairs significant at p value < .01 after Bonferroni correction.

RESULTS

We examined nine microsatellite markers from a region in chromosome 18, investigated by Eaves et al. (2000) in samples from two genetically isolated populations (Finland and Sardinia) and two outbred populations (Britain and USA). The markers we used and their genetic map distances are reported in Table 1. The choice of the marker set was dictated by the possibility to compare directly the pattern of bLD on chromosome 18 between the Afrikaners and the four other populations reported in Eaves et al. 2000 (See Web Site References for location of raw data). We first evaluated all pairs of loci on chromosome 18 in the Afrikaner sample for the presence of allelic associations using Fisher exact test (FET). To describe the extent of nonrandom allelic association between pairs of loci we computed the tail probability (FET p value) using the `Arlequin` software (Schneider et al. 2000). A total of 36 locus pairs were considered for chromosome 18 with an average distance of 3.1 cM. A large number of pairs (50%) with a p value <.01 were identified (Table 2): 4/7 pairs within 1 cM, 4/7 between 1 and 2 cM, 3/4 between 2 and 3 cM, 2/4 between 3 and 4 cM, 2/7 between 4 and 5 cM, and 3/7 between 5 and 6.5 cM. Multiple testing is an issue with regard to the two locus tests. However, when the conservative Bonferroni correction is applied for the 36 tests performed, 14/18 (78%) tests reported in Table 2 still remain significant at $p < .01$. Under this stringent criterion, 85% (6/7) of the tests for marker pairs separated by more than 3 cM remain significant, including two marker pairs at a distance >5 cM.

We also quantified the strength of bLD rather than the significance of the association between all possible pairs of loci. We used two different methods to quantify LD. First, we used the widely employed multiallelic extension of

Table 3. Comparison of LD between Locus Pairs at Increasing Ranges of Genetic Distance across Populations

Population	0–1 cM	1–2 cM	2–3 cM	3–4 cM	4–5 cM	5–6.5 cM
	Proportion of locus pairs with $D'_m > 0.30$					
Afrikaners	3/7	3/7	0/4	1/4	0/7	2/7
Finns	1/7	0/7	0/4	0/4	0/7	0/7
Sardinians	0/7	0/7	0/4	0/4	0/7	0/7
British	0/7	0/7	0/4	0/4	0/7	0/7
USA	0/7	0/7	0/4	0/4	0/7	0/7
	Proportion of locus pairs with $G^2 > 0.05$					
	0–1 cM	1–2 cM	2–3 cM	3–4 cM	4–5 cM	5–6.5 cM
Afrikaners	7/7	7/7	4/4	4/4	7/7	7/7
Finns	6/7	4/7	2/4	0/4	0/7	0/7
Sardinians	3/7	1/7	0/4	0/4	0/7	0/7
British	4/7	0/7	0/4	0/4	0/7	0/7
USA	3/7	0/7	0/4	0/4	0/7	0/7

LD, linkage disequilibrium.

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

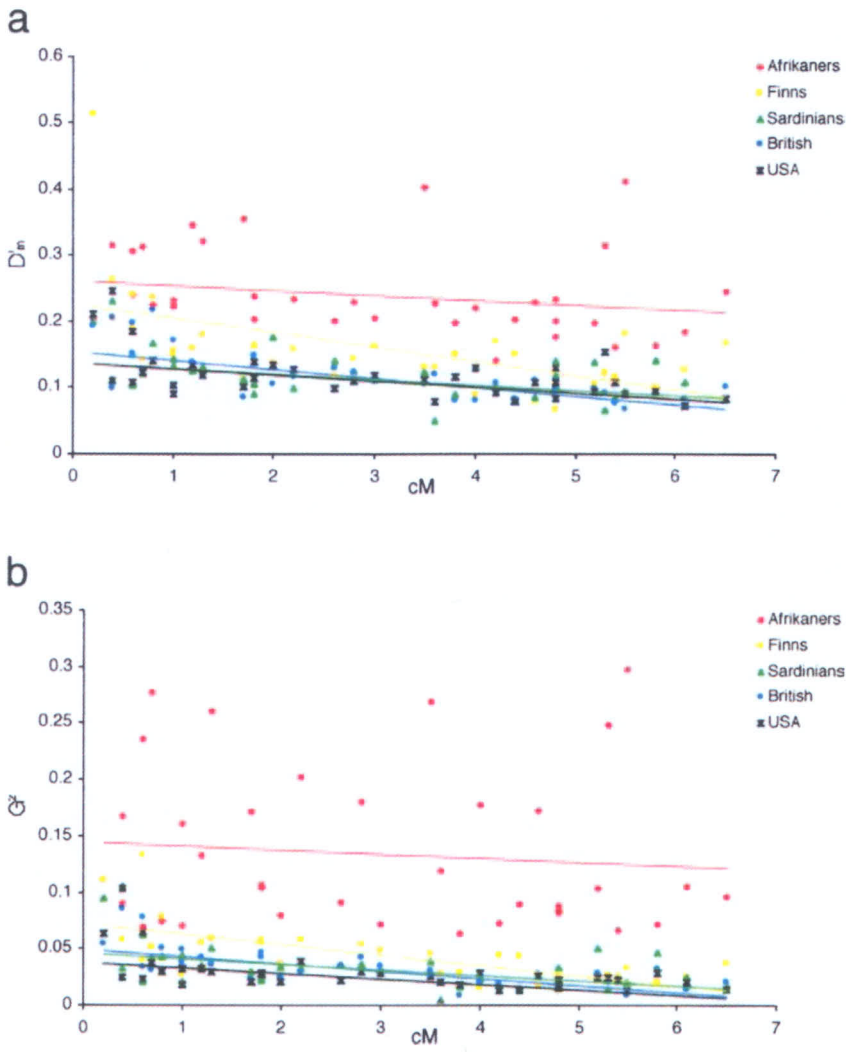


Figure 1 Relationship between multiallelic D'_m (*a*) and G^2 (*b*) values and genetic distance (cM) across populations. Solid lines are linear regression best fit lines.

Lewontin's standardized measure of disequilibrium (D' measure, Lewontin 1964; Hedrick 1987). This statistic may be affected by sample size (with smaller size samples demonstrating larger D'_m values) and allele frequencies (Mohlke et al. 2001) and therefore may not be appropriate for comparison of LD between studies that use different sample sizes and markers with differing numbers of alleles and differing allele frequencies. To facilitate comparison of bLD across populations, we also used the pairwise G^2 measure, originally proposed by Balakrishnan and Sangvi (1968) and then explored by Chapman and Wijsman (1998), as a measure of LD for loci with multiple alleles. G^2 values are not affected by changes in sample sizes.

Figure 1 reports the distribution of pairwise D'_m and G^2 values for all marker pairs tested for all five populations compared. We find that the extent of bLD in the Afrikaners, when either G^2 or D'_m values are considered across chromosome 18, is consistently higher than in the other populations tested. Over short intervals (0–3 cM), analysis of G^2 values distribution indicates high bLD in the Afrikaners and to a lesser extent in the Finns, with fewer differences among all the other populations. At this range, the Afrikaners show $G^2 > 0.05$ for 18/18 marker pairs compared with 12/18 in the Finnish sample, 4/18 in the Sardinian and British sample, and 3/18 in the USA sample (Fig. 1b, Table 3). When using the D'_m statistic, the Afrikaners show $D'_m > 0.30$ for 6/18 marker pairs compared with 1/18 in the Finns and 0/18 in all other populations tested (Fig. 1a, Table 3). Although the D'_m statistic may be affected by sample size (the Afrikaner sample is almost one third of all other demographic groups), the similar distribution pattern of the G^2 values strongly indicates that higher levels of bLD exist in the Afrikaner dataset in the entire 0–6.5-cM range at chromosome 18.

Because different pairs of loci in the genome will show variable amount of LD, we compared the distribution of bLD across two genomic regions within the Afrikaner sample by typing 10 additional markers on chromosome 4 (Table 1).

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

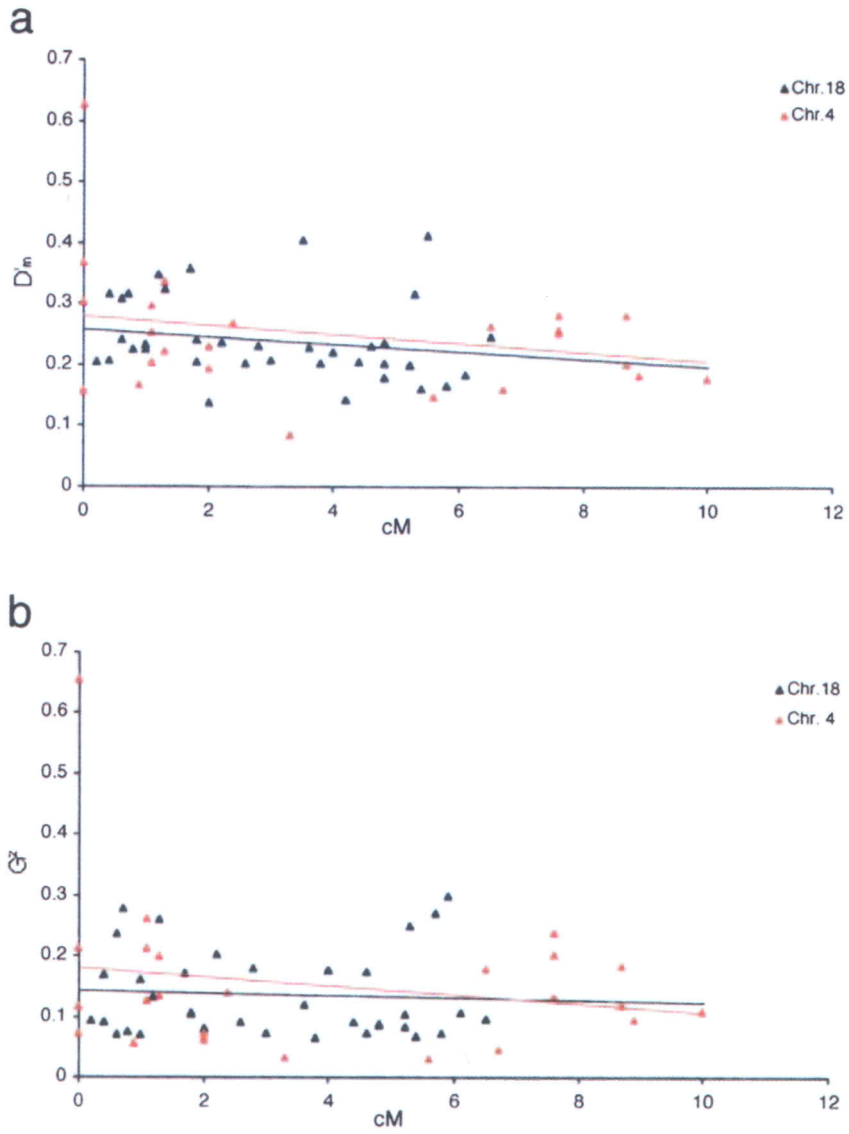


Figure 2 Comparison of the strength of linkage disequilibrium (LD) on chromosomes 18 and 4 in the Afrikaners. Multiallelic D'_m (*a*) and G^2 (*b*) are plotted against genetic distance (cM). Each point represents one pair of markers. Solid lines are linear regression best fit lines.

Analysis of the distribution of D'_m and G^2 values (Fig. 2) indicates that LD does not seem to vary significantly between the two genomic regions analyzed in this study. Marker pair comparisons within the common genetic range tested for both chromosomes (0–6.5 cM) indicate a similar proportion of pairs with $D'_m > 0.30$ or $G^2 > 0.05$, as well as similar D'_m and G^2 mean values (Table 4).

DISCUSSION

In the present study, we conducted an investigation of the level of bLD in the Afrikaner population to test the appropriateness of this genetic isolate for genome-wide linkage and linkage disequilibrium studies designed to identify genes for complex traits. We took care to avoid strategies previously proven confusing in LD data interpretation. We used microsatellite markers as opposed to commonly used di-allelic single nucleotide polymorphisms (SNPs) for two reasons: (1) because of the large number of dense microsatellite marker panels already in use in genetic linkage studies and (2) because of the notion that di-allelic markers may have less power to detect LD than multiallelic markers (Chapman and Wijsman 1998). In addition, we did not collapse alleles to simulate di-allelic markers, as is often done to facilitate the statistical analysis but can give highly misleading results (Goddard et al. 1999). Haplotypes were determined with precision using primarily family data rather than data inferred through statistical methods. Finally, alongside the most commonly used measures, such as D'_m (Lewontin 1964) and *p* values from pairwise significance tests, we also applied the G^2 statistic that fa-

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Table 4. Comparison of LD between Two Genomic Regions in the Afrikaners

Chromosome	$D'_m > 0.30$ ratio (%)	D'_m average	$G^2 > 0.05$ ratio (%)	G^2 average
18	9/36 (25)	0.24	36/36 (100)	0.14
4	4/16 (25)	0.26	14/16 (87.5)	0.16

cilitated the comparison of data from different populations with different sample sizes (Chapman and Wijsman 1998).

One important finding of our study is the remarkable number of significant marker-to-marker associations at 3–6 cM of genetic distance observed in the Afrikaners (Table 2). This finding is in agreement with a previous, more limited study by Gordon et al. (2000) that found significant evidence for LD in the Afrikaners extending over a 5-cM region in other chromosomal regions. The validity of this finding is further supported in the current study by a direct bLD comparison of chromosome 18 between the Afrikaners and four other populations, including two genetic isolates and two mixed populations. This comparison indicates consistently higher mean LD levels for the Afrikaners at genetic distances over 3 cM when either the mean G^2 values or D'_m values are considered. The lower LD observed at this range in the Finns could be attributable to a larger number of generations of exponential growth compared with the Afrikaners who are characterized by a recent bottleneck event (13–15 generations) and have maintained the initial level of LD over larger genetic distances. Retention of bLD over large distances may also be related to a limited number of original founders for the families examined or to previously undetected admixture. Analysis of the population structure (Pritchard and Rosenberg 1999) and detailed genealogical research (M. Karayiorgou et al. in prep.) will help us understand the forces controlling bLD in this population.

Taken together, these results indicate that the Afrikaner population conforms to the standards desired in a population appropriate for a genome-wide scan requiring extensive LD. In comparison with other older founder populations with similar size (such as the Finns), significantly fewer markers may be sufficient for genome-wide LD analysis of complex disorders in the Afrikaners. This LD, coupled with the possibility of recruitment of large samples, could result in considerable power to detect LD and thus easier detection of genes with modest contribution to disease. In that respect, the Afrikaners, who total approximately 3 million, do not suffer from a common drawback encountered in founder populations with extended LD but small population size, namely, the availability of a limited number of cases even for common disorders (Zavattari et al. 2000). It is the combination of extended LD and availability of a large number of cases for ascertainment that makes the Afrikaners an ideal genetic isolate for mapping complex traits.

A common concern in using founder populations for mapping complex, multigenic traits is the relevance of the study's findings to other cosmopolitan populations. Although currently there is no empirical data arguing for or against this issue, the Northern European/Dutch origin of the Afrikaners is likely to render findings from studies in this population relevant to many other Western populations. It is of course likely that only a subset of susceptibility genes for any given complex disorder will segregate in a founder population. Nevertheless, identification of these genes is likely to provide important insight into the affected molecular and cellular pathways and inspire novel drug development approaches. Finally, an additional concern, applicable particularly in genetic mapping of phenotypes diagnosed as clinical syndromes rather than as quantitative traits, is the clinical similarity of a phenotype across populations. In that respect, we can mention that in our initial analyses comparing the clinical similarity between schizophrenia in the Afrikaner sample (N = 159) and our US sample (N = 223), we find that the basic sample descriptors and cardinal symptoms of the disease are equivalent (M. Karayiorgou et al. in prep.).

Our analysis used bLD estimates as an indication of disease-associated LD. It should be emphasized, however, that extended bLD may, in principle, confuse the interpretation of LD mapping unless it is accounted for (McPeek and Strahs 1999) and distinguished from disease-associated LD. It should be noted, however, that some of the limitations imposed by extended bLD can be overcome by following meiotic recombination events within linked families with multiple affected individuals, thus reducing the size of disease-associated haplotypes and facilitating fine mapping of the disease-causing gene. Alternatively, once LD is obtained and replicated, the fine mapping of the region could be pursued in the general Northern European/Dutch population of origin.

Finally, it is important to point out that the usefulness of founder populations in the dissection of complex traits also includes aspects other than increased bLD levels. These include a more uniform environment, good genealogical records, more intact families, and a phenotype definition that is easier to standardize.

METHODS

Sample

Of the 62 families used in the current study, 55 are three-member families composed of one child and both biological parents, and seven are four-member families composed of two children and both biological parents. All families studied are of Afrikaner heritage, as determined by our extensive research of genealogical records dating back to the early sixteenth century. DNA was extracted from blood samples from all study participants.

Our sampling approach is unlikely to make these families nonrepresentative of the general Afrikaner population for two reasons. First, separate analysis of transmitted and nontransmitted parental chromosomes provided identical results. Second, given the existing literature on genome-wide searches for schizophrenia susceptibility loci, any ascertainment bias introduced would have been minimal because the risk conferred by a putative schizophrenia susceptibility locus in the areas examined would be small.

The study was approved by Institutional Review Boards (IRBs) at both the Rockefeller University and University of Pretoria. Appropriate informed consent was obtained from all study participants.

Markers

All microsatellite markers we used are dinucleotide repeats except D18S851 that contains tetranucleotide repeats. Primer sequences are available from the Genome Database (GDB) with the exception of D18S1156, in which forward primer sequence was changed to CTTGCACCCTGCAAGTT, and D4S418, in which primer sequence was changed to GGATCA CAGAGTGAAG because of the presence of an undetermined nucleotide (N) in the primer sequence given at GDB. Observed heterozygosity values in the Afrikaner population range from 0.40 to 0.78 with an average value of 0.66.

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Tests of Linkage Disequilibrium between Pairs of Markers

For each pair of loci, a contingency table of gamete frequencies was formed and 100,000 tables with the same marginal totals were generated on the basis of a Monte Carlo Markov-chain algorithm (Guo and Thompson 1992). The *p* value is the fraction of tables that were at least as extreme as the observed table.

D' is the standardized disequilibrium value that takes the usual disequilibrium coefficient P (A_iB_j)-P (A_i) P (B_j) and divides it by its maximal possible value. Values range from 0 to 1, with 0 reflecting perfect independence between alleles at the two loci compared and 1 reflecting complete LD. We calculated a multiallelic extension of the normalized association measure D' (Lewontin 1964) as $D'_m = \sum_i\sum_j p_i q_j |D'_{ij}|$ where p_i and q_j are the observed frequencies for alleles i and j, respectively, at the two loci. We computed the pairwise G^2 as $\sum_i \sum_i (P_i - Q_i)^2/(P_i + Q_i)$ where P_i is the estimated two-locus frequency for haplotype i and Q_i is the expected haplotype frequency based on independence of individual allele frequencies at the same two loci. Both statistics were calculated using haplotypes phased from family data. Haplotypes were manually constructed for the four parental chromosomes by inferring phase from their genotypes and those of the child. Before genotyping the 7 four-member families, one child was selected at random. Final haplotype frequencies were estimated using an EM algorithm.

ACKNOWLEDGMENTS

We thank all the study participants and Sister Ria van Wyk for assistance with the recruitment of the families. We also thank Gonçalo Abecasis for useful comments. Support for this work was provided by grants from the EJLB Foundation, the Essel Foundation, and NIH R01 MH61399–01 (to MK).

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

REFERENCES

Balakrishnan, V. and Sanghvi, L.D. 1968. Distance between populations on the basis of attribute data. *Biometrics* **24:** 859–865.

Brink, P.A., Steyn, L.T., Coetzee, G.A., and van der Westhuyzen, D.R. 1987. Familial hypercholesterolemia in South African Afrikaners. PvuII and StuI DNA polymorphisms in the LDL-receptor gene consistent with a predominating founder gene effect. *Hum. Genet.* **77:** 32–35.

Brink, P.A., Ferreira, A., Moolman, J.C., Weymar, H.W., van der Merwe, P.L., and Corfield, V.A. 1995. Gene for progressive familial heart block type I maps to chromosome 19q13. *Circulation* **91:** 1633–1640.

Chapman, N.H. and Wijsman, E.M. 1998. Genome screens using linkage disequilibrium tests: Optimal marker characteristic and feasibility. *Am. J. Hum. Genet.* **63:** 1871–1885.

Eaves, I.A., Merriman, T.R., Barber, R.A., Nutland, S., Tuomilehto-Wolf, E., Tuomilehto, J., Cucca, F., and Todd, J.A. 2000. The genetically isolated populations of Finland and Sardinia may not be a panacea for linkage disequilibrium mapping of common disease genes. *Nat. Genet.* **25:** 320–323.

Freimer, N.B., Service, S.K., and Slatkin, M. 1997. Expanding on population studies. *Nat. Genet.* **17:** 371–373.

Goddard, K.A.B. 1999. 'Design issues in the analysis of complex genetic traits.' PhD diss., University of Washington, Seattle.

Goldman, A., Krause, A., Ramsay, M., and Jenkins, T. 1996. Founder effect and the prevalence of myotonic dystrophy in South Africans: Molecular studies. *Am. J. Hum. Genet.* **59:** 445–452.

Gordon, D., Simonic, I., and Ott, J. 2000. Significant evidence for linkage disequilibrium over a 5-cM region among Afrikaners. *Genomics* **66:** 87–92.

Groenewald, J.Z., Liebenberg, J., Groenewald, I.M., and Warnich, L. 1998. Linkage disequilibrium analysis in a recently founded population: Evaluation of the variegate porphyria founder in South African Afrikaners. *Am. J. Hum. Genet.* **62:** 1254–1258.

Guo, S.W. and Thompson, E.A. 1992. Performing the exact test of Hardy-Weinberg proportion for multiple alleles. *Biometrics* **48:** 361–372.

Hayden, M.R., MacGregor, J.M., and Beighton, P.H. 1980. The origin of Huntington's chorea in Afrikaner population. *S. Afr. Med. J.* **58:** 197–200.

Hedrick, P.W. 1987. Gametic disequilibrium measures: Proceed with caution. *Genetics* **117:** 331–341.

Jorde, L.B. 1995. Linkage disequilibrium as a gene-mapping tool. *Am. J. Hum. Genet.* **56:** 11–14.

Leitersdorf, E., van der Westhuyzen, D.R., Coetzee, G.A., and Hobbs, H.H. 1989. Two common low density lipoprotein receptor gene mutations cause familial hypercholesterolemia in Afrikaners. *J. Clin. Invest.* **84:** 954–961.

Lewontin, R.C. 1964. The interaction of selection and linkage. I. General considerations; heterotic models. *Genetics* **49:** 49–67.

McPeek, M.S. and Strahs, A. 1999. Assessment of linkage disequilibrium by the decay of haplotype sharing, with application to fine-scale genetic mapping. *Am. J. Hum. Genet.* **65:** 858–875.

Mohlke, K.L., Lange, E.M., Valle, T.T., Ghosh, S., Magnuson, V.L., Silander, K., Watanabe, R.M., Chines, P.S., Bergman, R.N., Tuomilehto, J., et al. 2001. Linkage disequilibrium between microsatellite markers extends beyond 1 cM on chromosome 20 in Finns. *Genome Res.* **11:** 1221–1226.

Pritchard, J.K. and Rosenberg, N.A. 1999. Use of unlinked genetic markers to detect population stratification in association studies. *Am. J. Hum. Genet.* **65:** 220–228.

Pronk, J.C., Gibson, R.A., Savoia, A., Wijker, M., Morgan, N.V., Melchionda, S., Ford, D., Temtamy, S., Ortega, J.J., Jansen, S., et al. 1995. Localisation of the Fanconi anaemia complementation group A gene to chromosome 16q24.3. *Nat. Genet.* **11:** 338–340.

Risch, N. and Merikangas, K. 1996. The future of genetic studies of complex human disease. *Science* **273:** 1516–1517.

Roby, P., Eyre, S., Worthington, J., Ramesar, R., Cilliers, H., Beighton, P., Grant, M. and Wallis, G. 1999. Autosomal dominant (Beukes) premature degenerative osteoarthropathy of the hip joint maps to an 11-cM region on chromosome 4q35. *Am. J. Hum. Genet.* **64:** 904–908.

Rosendorff, J., Bernestein, R., Macdougall, L. and Jenkins, T. 1987. Fanconi anaemia: Another disease of high prevalence in the Afrikaans population of South Africa. *Am. J. Med. Genet.* **27:** 793–797.

Schneider, S., Roessli, D., and Excoffier, L. 2000. Arlequin: A software for population genetic data analysis. User manual ver 2.000. Ver. 2.000. Geneva: Genetics and Biometry Lab, Dept. of Anthropology, University of Geneva, Geneva.

Shifman, S. and Darvasi, A. 2001. The value of isolated populations. *Nat. Genet.* **28:** 309–310.

Torrington, M. and Viljoen, D.L. 1991. Founder effect in Afrikaners kindreds with pseudoxanthoma elasticum. *S. Afr. Med. J.* **79:** 7–11.

Warnich, L., Kotze, M.J., Groenewald, I.M., Groenewald, J.Z., van Brakel, M.G., van Heerden, C.J., de Villiers, J.N., van de Ven, W.J., Schoenmakers, E.F., Taketani, S. et al. 1996. Identification of three mutations and associated haplotypes in the protoporphyrinogen oxidase gene in South Africa. *Hum. Mol. Genet.* **5:** 981–984.

Zavattari, P., Deidda, E., Whalen, M., Lampis, R., Mulagria, A., Loddo, M., Eaves, I., Mastio, G., Todd, J.A., and Cucca, F. 2000. Major factors influencing linkage disequilibrium by analysis of different chromosome regions in distinct populations: Demography, chromosome recombination frequency and selection. *Hum. Mol. Genet.* **9:** 2947–2957.

WEB SITE REFERENCES

http://anthropologie.unige.ch/arlequin; Arlequin program.

http://research.marshfieldclinic.org/genetics; Haplotypes from Eaves et al. 2000.

http://genome.UCSC.edu; Human Genome Project Working Draft, Santa Cruz.

http://www.gdb.org; Genome Database (GDB).

http://www-gene.cimr.cam.ac.uk/todd/public_data/chr18/LD webpage.html; Marshfield genetic maps.

Received January 29, 2002 ; accepted in revised form March 29, 2002.

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

American Journal of Medical Genetics Part B (Neuropsychiatric Genetics) 124B:20–28 (2004)

Phenotypic Characterization and Genealogical Tracing in an Afrikaner Schizophrenia Database

Maria Karayiorgou,[1]* Marie Torrington,[2] Gonçalo R. Abecasis,[3] Herman Pretorius,[2] Brian Robertson,[4] Sean Kaliski,[4] Stephen Lay,[4] Christina Sobin,[1] Natalie Möller,[2] S. Laura Lundy,[1] Maude L. Blundell,[1] Joseph A. Gogos,[5] and J. Louw Roos[2]

[1]Human Neurogenetics Laboratory, The Rockefeller University, New York, New York
[2]Department of Psychiatry and Weskoppies Hospital, University of Pretoria, Pretoria, Republic of South Africa
[3]Department of Biostatistics, University of Michigan, Ann Arbor, Michigan
[4]Department of Psychiatry and Valkenberg Hospital, University of Capetown, Capetown, Republic of South Africa
[5]Department of Physiology and Cellular Biophysics and Center for Neurobiology and Behavior, College of Physicians and Surgeons, Columbia University, New York, New York

Founder populations hold tremendous promise for mapping genes for complex traits, as they offer less genetic and environmental heterogeneity and greater potential for genealogical research. Not all founder populations are equally valuable, however. The Afrikaner population meets several criteria that make it an ideal population for mapping complex traits, including founding by a small number of initial founders that likely allowed for a relatively restricted set of mutations and a large current population size that allows identification of a sufficient number of cases. Here, we examine the potential to conduct genealogical research in this population and present initial results indicating that accurate genealogical tracing for up to 17 generations is feasible. We also examine the clinical similarities of schizophrenia cases diagnosed in South Africa and those diagnosed in other, heterogeneous populations, specifically the US. We find that, with regard to basic sample descriptors and cardinal symptoms of disease, the two populations are equivalent. It is, therefore, likely that results from our genetic study of schizophrenia will be applicable to other populations. Based on the results presented here, the history and current size of the population, as well as our previous analysis addressing the extent of background linkage disequilibrium (LD) in the Afrikaners, we conclude that the Afrikaner population is likely an appropriate founder population to map genes for schizophrenia using both linkage and LD approaches.
© 2003 Wiley-Liss, Inc.

KEY WORDS: founder population; schizophrenia genetics; South Africa; genealogical research; gene mapping

INTRODUCTION

The identification of susceptibility genes for common, complex disorders has been a task of unprecedented difficulty. A very high degree of genetic heterogeneity, as well as clinical heterogeneity in certain diseases, such as the psychiatric syndromes, are suspected to be at the core of the difficulty. There are certain expectations that founder populations may be useful to circumvent some of these difficulties by offering reduced genetic heterogeneity, more uniform environment, and the potential for genealogical research. Such populations have proven useful for mapping genes that underlie Mendelian disorders and multigenic disorders that segregate as Mendelian conditions in specific populations [Hastbacka et al., 1994; Kalaydjieva et al., 1996; Nystuen et al., 1996; Visapaa et al., 2002] and particular features of these populations should also facilitate the search for genes that influence susceptibility to complex traits.

Founder populations have originated from a relatively small number of individuals and have expanded over several generations in relative isolation. As a result, it may be reasonable to expect that a decreased number of

Grant sponsor: Essel Foundation; Grant sponsor: EJLB Foundation; Grant sponsor: National Institute of Mental Health (to MK); Grant number: R01 MH 61399-01.

*Correspondence to: Maria Karayiorgou, M.D., Associate Professor, Head, Human Neurogenetics Laboratory, The Rockefeller University, 1230 York Avenue, Box #313, New York, NY 10021. E-mail: karayim@mail.rockefeller.edu

Received 4 February 2003; Accepted 19 May 2003
DOI 10.1002/ajmg.b.20090

© 2003 Wiley-Liss, Inc.

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

independent susceptibility loci are segregating in such a population. At any susceptibility locus, all affected individuals in the population may carry the same or, more likely, a limited set of alleles derived identical-by-descent (IBD) from a few common ancestors. The reduced locus and allelic heterogeneity in founder populations means that genetic mapping strategies may require smaller sample sizes for success. Furthermore, the relatively recent origin of the population ensures that the chromosomal region surrounding the disease allele is larger than in outbred populations, allowing relatively sparse genetic maps in these populations to be as informative as denser maps in outbred populations. As a result, founder populations are more amenable to linkage disequilibrium (LD) approaches, which may be more powerful than traditional linkage analysis for complex traits. It is important to note that the degree of resolution of mapping strategies based on LD between loci will vary between founder populations and depend on the number of generations since the founding. While initial gene localization may be more feasible in young founder populations because of the extensive LD, fine scale localization of a gene may be problematic for the same reason.

It is not known at this point which population history and structure is more advantageous for mapping complex traits. Current genetic linkage studies of schizophrenia include patient samples from founder populations of different ages (such as from Finland, Costa Rica, French Quebec, Palau, Kosrae) [Hovatta et al., 1999; Paunio et al., 2001; DeLisi et al., 2002; Devlin et al., 2002; Wijsman et al., 2003]. Here, we present the Afrikaner population of South Africa as an ideal genetic isolate for our gene mapping studies in schizophrenia. The Afrikaner population meets several criteria that make it an ideal population for mapping complex traits, including (a) founding by a small number of initial founders that likely allowed for a relatively restricted set of mutations to have been introduced in the population, (b) large current population size to allow identification of sufficient number of cases, and (c) strong infrastructure to allow genealogical research as well as collection of good quality clinical data.

The Afrikaners were founded by approximately 2,000 initial settlers who came from The Netherlands and other parts of Northern Europe in consecutive waves, starting at 1,652. The first immigrants settled in the Cape. They later spread inland, founding geographically isolated communities. Besides geography, other factors, such as culture, language differences, and religious practices, contributed to the isolation of the Afrikaner. Most Afrikaners were members of the Dutch Reformed Church. They lived in large families scattered over the country, so that consanguinity was common, especially in early generations. Population growth was almost entirely through reproduction and immigration subsequent to the founding was minimal. In more recent years, some admixture has occurred. The current size of the Afrikaner population is approximately 3 million and their age of ancestry is on average 12.5 generations old.

The demographic history of this population is reflected in the unusually high frequency of certain rare Mendelian disorders (as much as five to ten times higher than in most other population groups), such as variegate porphyria (VP), Huntington's chorea, familial hypercholesterolemia (FH), Fanconi anemia, myotonic dystrophy [Hayden et al., 1980a,b; Brink et al., 1987; Rosendorff et al., 1987; Goldman et al., 1996; Warnich et al., 1996; Groenewald et al., 1998]. Moreover, unusually low allelic diversity at the associated disease loci has been observed in the Afrikaner. This is an expected result of genetic drift in a population of this size. Indeed, studies on FH identified only three founder mutations responsible for more than 95% of FH cases in the Afrikaner population [Leitersdorf et al., 1989]. These Afrikaner mutations were also identified in FH patients in The Netherlands [Defesche et al., 1993] and one of them was shown to have originated there and been introduced to South Africa by a single individual of Dutch origin [Defesche et al., 1996]. Finally, unusually large extent (8–11 cM) of conserved haplotypes around rare disease genes has also been described in the Afrikaners [Pronk et al., 1995; Roby et al., 1999].

In a previous study, supporting the appropriateness of the Afrikaner population for mapping complex traits, we have shown that the level of background LD in the Afrikaner sample is detectable at 6 cM [Hall et al., 2002], whereas it decays rapidly beyond 3 cM distances in the other founder and outbred populations examined, including the Finns, previously proven a valuable population for monogenic disease mapping. Here, we address two additional issues regarding this population, namely, the feasibility of detailed genealogical research and also the clinical similarity of schizophrenia diagnosed in this population and in more cosmopolitan populations.

We show that it is feasible in this population to trace each proband to the 16th century ancestors, an approach that promises to be very powerful when assessing alleles or chromosomal segments shared by descent. Initial results from our genealogical research are presented. Finally, we address the clinical comparability between Afrikaner schizophrenic patients and US schizophrenic patients and conclude that the cardinal symptoms of the disorder are stable across the two populations. This is important because while there is general agreement that using founder populations to map genes for complex traits may be meritorious, concern exists about the generalization of findings to the population at large, given the uncertainty about the accuracy of the diagnosis in disorders diagnosed exclusively as clinical syndromes, as all psychiatric disorders are at present.

METHODOLOGY

Sample Collection

Since 1997, in collaboration with clinicians from two of the major psychiatric hospitals in South Africa (Weskoppies Hospital in Pretoria and Valkenberg Hospital in Capetown), we ascertained and evaluated probands with a history of psychotic illness. Approval for the protocol was obtained from the Rockefeller IRB, as well as from the IRB committees at each site. All participants gave written informed consent.

At each site, an Afrikaans-speaking mental health worker (recruiter) identified individuals potentially ap-

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

propriate for enrollment in the study through patient referral from hospital staff, chart reviews in local hospitals, presentations to local support groups, and advertisements in local newspapers and magazines. After identifying a potential proband, the recruiters complete a brief psychiatric screen, explain the study, ascertain the patient's interest and motivation, determine the Afrikaner lineage and availability of both biological parents, obtain informed consent from the patient and parents, and record detailed pedigree information. A blood sample is then obtained (from proband and parents) and a clinical interview with the proband is scheduled. For probands who are severely affected or actively psychotic, an informant interview with a knowledgeable first-degree relative (usually a parent) is conducted, and a brief, unstructured interview is administered to the proband.

Our current sample includes 319 probands of Afrikaner descent (251 from Pretoria and 68 from Capetown), who meet diagnostic criteria for schizophrenia or schizoaffective disorder according to the American Psychiatric Association [1994]. Of the 319 probands, 71 are members of families with two or more affected individuals, while the rest are members of triads (affected individual and both biological parents).

Clinical Evaluation

The Diagnostic Instrument for Genetic Studies (DIGS) [Nurnberger et al., 1994], after it was translated and back translated into Afrikaans, was used for all diagnostic evaluations. Two senior psychiatrists (HP and JLR) from the Weskoppies Hospital in Pretoria and a senior Psychologist (SL) from the Valkenberg Hospital in Capetown, with a minimum of 10 years clinical experience, conducted all clinical interviews after they were trained in the use of the DIGS and the research application of DSM-IV. Reliability interviews with each clinical interviewer were conducted at the beginning of the study and the clinical validity of hospital chart diagnoses was also assessed. In order to ensure continued interviewer reliability and to reduce interviewer 'drift,' additional reliability interviews were conducted at regular intervals (once a year). In addition, each site videotaped two interviews per interviewer each year for the purposes of continuing quality control. Reliability of final diagnoses was found to be excellent at both sites. All clinical interviewers were fluent in both English and Afrikaans. The diagnostic interviews were conducted in Afrikaans. Following each interview, diagnosticians completed DSM-IV checklists for each category in which positive symptoms were identified in the DIGS; a narrative chronologic summary of prodromal traits, symptom onset, and functional impairment; and an English-language data summary form of selected DIGS items.

Genealogical Research

The psychiatric nurse-study recruiter at each site collected initial pedigree information from all probands who met study criteria and were enrolled in the study. The genealogical query includes questions about place and date of birth, place, and date of death for each great grandparent, as well as name and order of birth for all children in each generation. The proband and/or parents supplied the information. Typically, informants were able to provide pedigree information for three to four generations [up to great grandparents]. A systematic search was then conducted to identify the generations prior to the great grandparents. This search involved reviewing records in the State Archives, Church Archives, the Master's office, the Deeds office, the office of the Registrar of Births, Marriages, and Deaths, museums. The painstaking archival research continued until the pedigree was dated to the early 1800s. From that point, the Genealogies of Old South African Families [De Villiers, 1981], which record in detail the history of all families up until the initial 2,000 founders, are used to bring the pedigree back to its original founders. The genealogical information is then set out in the form of a family tree, or pedigree chart, and entered into a secure searchable File Maker Pro-Computerized database, where each person is assigned a unique numerical ID. This genealogical ID is different than the clinical ID and the DNA ID. Before tracing of an individual begins, the database is searched (by last name, first name, or both, and by year of birth) to confirm that the individual has not already been traced.

Predicting the Extent of Shared Haplotypes

We used simulation to predict the extent of the original founder haplotype that might surround a disease mutation carried by affected individuals in our sample. Carrying out parametric gene dropping simulations [Ploughman and Boehnke, 1989] that include a particular disease model, is impractical in a large pedigree with many loops such as the Afrikaner pedigree described here. Instead, we used a simple non-parametric procedure to study the fraction of a risk haplotype introduced by a single founder chromosome that might be transmitted to each affected individual.

In each simulation, we first labeled a single founder chromosome as the disease chromosome. We then iterated through the pedigree, examining each individual before any of his descendants. Each individual was then assigned either two normal chromosomes (if neither parent carried the disease chromosome) or one normal and one disease chromosome (if at least one parent carried a disease chromosome). Whenever a disease chromosome was transmitted, we sampled the location of flanking recombination events from an exponential distribution and tracked the erosion of the ancestral haplotype. Finally, we recorded the extent of the original ancestral haplotype carried by each affected individual in the pedigree.

This simulation represents a simplification of the processes by which disease alleles segregate through a pedigree and shared haplotypes are generated. Two simplifications are particularly important: (1) it assumes that all affected individuals share a particular founder allele and (2) it does not account for shared haplotype stretches, other than the original founder

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

chromosome. In the presence of genetic heterogeneity and phenocopies it is unlikely that all affected individuals will share a particular mutation, and subsets of affected individuals likely to share a particular founder mutation (because they are linked to the same locus or exhibit similar phenotypes) must be identified. Nevertheless, our simulation provides an accurate estimate of the length of the original founder chromosome that will flank the disease allele in each affected carrier. Since we do not account for sharing of haplotypes other than the original founder chromosome, this estimate serves as a lower bound for the extent of haplotype sharing among carriers of a shared disease allele.

RESULTS AND DISCUSSION Clinical

Comparison of the Afrikaner Schizophrenia Sample to a US Schizophrenia Sample

In the analyses reported here, we characterized the demographic and clinical features of a subset of our Afrikaner patients (N ¼ 260) and assessed their clinical comparability to the manifestations of schizophrenia in North America by comparing them to our current US schizophrenia sample (N ¼ 238). In both samples, only one affected individual per family (proband) was included. Affected family members were excluded from these analyses, as were probands who were unavailable for interview and thus diagnosed solely on the basis of medical records and/or family report. All US patients have completed in full the same DIGS interview administered to South African patients [Sobin et al., 2001]. Our results are reported in Table I.

Eleven demographic and clinical diagnostic characteristics were indistinguishable across both populations, including mean age at time of interview, gender distribution, percent of left-handed patients, percent of patients with a family history of schizophrenia, or other psychiatric diagnosis, percent of patients with co-morbid substance abuse or dependence, and the reported rates of delusions, auditory hallucinations, visual hallucinations, sensory hallucinations, disordered behavior, and alogia/flat affect. Thus with regard to basic sample descriptors and cardinal symptoms of disease, the two populations are equivalent.

Three demographic and four clinical variables were different between the Afrikaner and the US samples: (a) Afrikaner patients completed fewer years of education (12 vs. 14 years, P ::; 0.01); (b) fewer Afrikaner patients were single (i.e., never married) (74 vs. 83%, P ::; 0.01); (c) more Afrikaner patients were living with parents or care-taking family members (living dependently) (80 vs. 59%, P ::; 0.01); (d) mean onset age of DSM-IV schizophrenia or schizoaffective disorder was reported to be approximately 3 years later in the Afrikaner sample (23 vs. 20 years, P ::; 0.01); (e) rates of reported schizoaffective disorder were lower in the Afrikaner sample (26 vs. 46%, P ::; 0.01); (f) suicide attempts were reported to occur less frequently among Afrikaner patients with schizophrenia or schizoaffective disorder (28 vs. 52%, P ::; 0.01); and finally (g) rates of reported thought disorder were higher in the Afrikaner sample (68 vs. 50%, P ::; 0.01). The differences in level of education, marital status, and dependent living arrangements likely reflect cultural variation. For example, the rate of high school completion among the Afrikaner population overall is

TABLE I. Comparison of Demographic and Clinical Characteristics of Afrikaner and US Probands

	Afrikaners (N ¼ 260)		US (N ¼ 238)	
	Mean	SD	Mean	SD
Demographics				
Interview age	32.63	10.52	33.63	8.93
Years of education	11.87	2.06	13.60	2.18*
Female	31.54%	(82)	38.66%	(92)
Left-handed	9.77%	(21)	9.40%	(22)
Single (never married)	73.54%	(189)	83.41%	(191)*
Living dependently	80.16%	(206)	58.95%	(135)*
Family history/schizophrenia	33.33%	(86)	31.03%	(54)
Family history/other psychiatric diagnosis	68.99%	(178)	62.64%	(109)
Syndrome features				
Age DSM-IV schizophrenia onset	23.19	6.78	20.01	5.31*
Schizoaffective	25.58%	(66)	45.99%	(109)*
Comorbid subst. abuse/dependence	33.33%	(85)	36.13%	(86)
Attempted suicide	27.76%	(68)	52.34%	(123)*
Symptom history				
Delusions	97.26%	(248)	97.40%	(225)
Auditory hallucinations	79.17%	(190)	73.93%	(173)
Visual hallucinations	38.89%	(84)	46.36%	(102)
Sensory hallucinations	27.32%	(59)	35.71%	(80)
Disordered behavior	69.03%	(156)	63.32%	(145)
Thought disorder	67.87%	(150)	50.00%	(115)*
Alogia/flat affect	63.85%	(136)	68.12%	(156)

All parenthetical values indicate the value of 'N.'
*Fisher's exact test P ::; 0.01 or t-test P ::; 0.01.

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

lower than that of the US. Also, the increased percentages of patients living at home as adults may grossly reflect the greater proportion of nuclear intact families in South Africa.

On the other hand, differences in clinical variables may reflect a higher allelic homogeneity in this population, that predisposes to a subtype of the disease with fewer mood symptoms, higher rates of thought disorder and somewhat later onset. For example, recent analysis of families from the Irish Study of High-Density Schizophrenia Families (a more heterogeneous dataset) indicated that affected individuals from families with evidence for linkage to chromosome 8p had significantly more affective deterioration, poorer outcome, more thought disorder, and fewer depressive symptoms than affected individuals from the other families in the study [Kendler et al., 2000]. However, we cannot exclude currently the possibility that the differences found in clinical variables may be attributable to sample differences. Since diagnosticians for both the US and Afrikaner samples were trained and monitored identically by the same supervising clinician (CS), and used identical clinical instruments, it is unlikely that methodologic ascertainment bias accounts for the differences. It is possible that the populations themselves may differ. For example, the later onset age may be a secondary effect of the closer-knit family structure in the Afrikaner population. Unlike the US sample, Afrikaner probands tend to stay in the family home longer, eliminating the stress associated with leaving one's family of origin, and perhaps temporarily delaying the onset of symptoms. Similarly, the lower rates of schizoaffective disorder in the Afrikaner sample may reflect greater current severity of the patient population. The US sample was comprised almost entirely of patients who, at the time of their diagnostic interviews, were stabilized on medication, while the Afrikaner population included many more severely affected patients who were either chronically hospitalized or whose medication was only partially effective. A less well-stabilized sample of patients may be less able to report the details of past episodes that would provide the clinical substantiation for a diagnosis of schizoaffective disorder. Finally, the decreased rates of suicide may be the result of the closer supervised care provided to Afrikaner patients living with family members or in long-term institutionalized/hospital settings, an option that still exists in South Africa.

Genealogical Tracing in the Afrikaner Dataset

Genealogical approaches are best practiced in the context of a population that has been isolated over many centuries with little migration and where extensive genealogy is available and accessible to allow for tracking of the flow of genetic information. Information about genealogical relationships may decrease the locus heterogeneity for the disease and will significantly improve our ability to assess the evidence for increased allele sharing among affected individuals. Through detailed archival research (see Materials and Methods), we are able to trace genealogical relations back to the original founders. At present, we have traced 98 of our probands who meet diagnostic criteria for schizophrenia or schizoaffective disorder (rv31% of the current Afrikaner sample). Remarkably, of the 98 traced affecteds, 87 descend from a single couple (Fig. 1).

In Table II, we list the maximum number of affected descendants for any one individual in each generation. It is of interest to note that we only observe individuals with a large number of affected descendants ten generations ago from present, indicating relatively distant connections between the individual pedigrees. Importantly, this does not pose a problem as the early portion of the genealogy is extremely well-documented in multiple Archival books. Currently, the largest connected portion of the pedigree includes 8,935 individuals in 17 generations. Of the 2,000 original settlers there are 644 represented in the master pedigree (making the fact that the majority of our probands link back to one settler and his wife even more impressive). There are 1,414 inbred individuals. The average coefficient of inbreeding for the pedigree is 0.00389. Among inbred individuals this rises to 0.0246. The maximum inbreeding coefficient for an individual is 0.15625. There are >2,500 loops within the pedigree, the majority of which are marriage loops. The majority of the consanguineous relationships occur in the early generations. To evaluate the completeness of the genealogy, we calculated the number of individuals and the proportion of founders in each generation. As seen in Table III, the genealogy seems complete, since we have identified the parents for the vast majority of individuals in generations 1–11 from present. One notable exception is the 3rd generation from present, where the fraction of individuals with no known ancestors is 44.4%, a dramatic increase compared to earlier and later generations. This reflects the history of the population, which in the period between 1900 and 1950 (coinciding with 3rd generation from present) experienced a dramatic movement following the discovery of minerals. The ensuing economic growth brought about a breakdown of tradition resulting to changes of spelling of surnames from Afrikaans to more anglicized versions, as well as to changes in the naming system of children, which up until then followed a very strict and predictable pattern, allowing inferences about lineage. These changes pose serious difficulties in the tracing of people in the 3rd generation from present.

We expect that a substantial fraction of the affected individuals in our pedigree may share risk alleles introduced by one or a few founders. We used simulation to estimate the length of the haplotype surrounding each of these alleles. Two individuals in our current extended pedigree have a total of 87 affected descendants, using strict diagnostic criteria (schizophrenia or schizoaffective disorder, depressed type), or 121 descendants, using the broadest classification criteria (any psychiatric disorder). We generated 1,000 pedigrees where a risk allele introduced by one of these individuals was transmitted to all affected descendants. In each case, we recorded the proportion of the original founder haplotype carried by each affected descendant. On average, there are 12.5 (range 10–17) intervening generations between this founder couple and their affected descendants so that

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

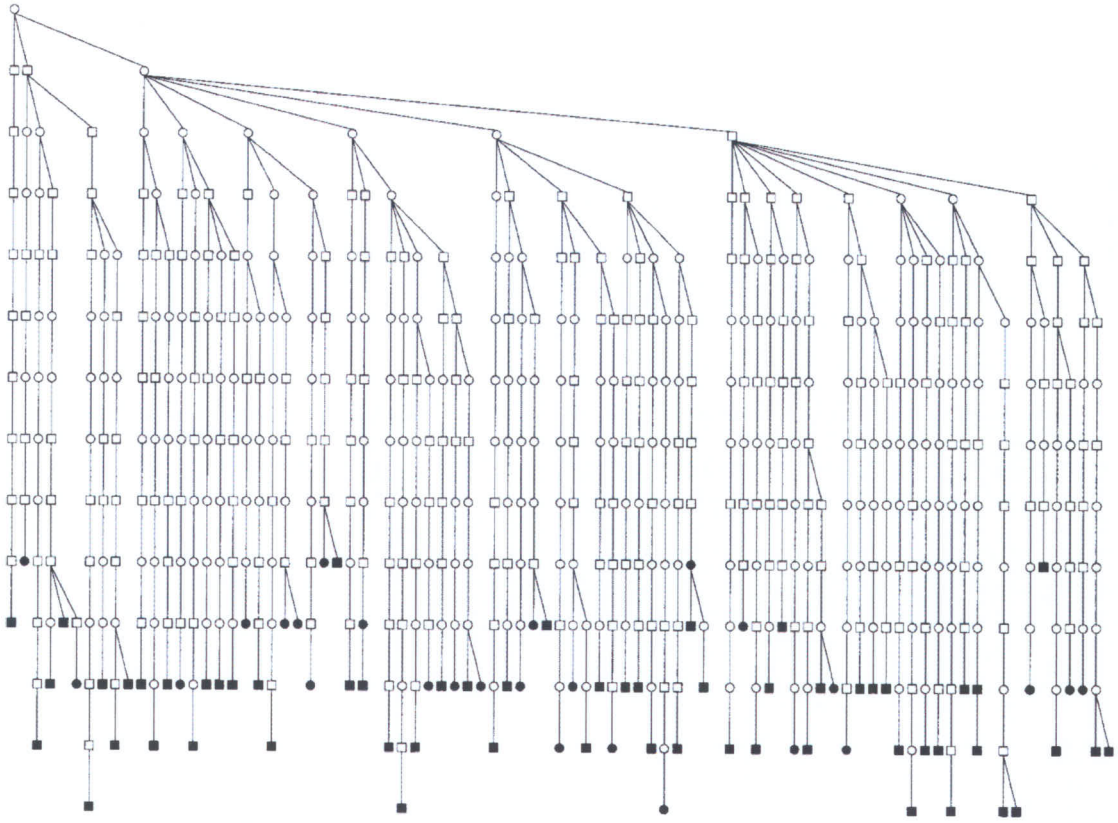


Fig. 1. Eighty seven of the affected individuals in our sample descend from a single couple, who lived 16 generations from present. Typically, each affected individual is connected to this couple through multiple lines of descent. The figure summarizes the shortest path (smallest number of meioses) between this couple and each affected descendant in our sample. On average, there are four alternative lines of descent connecting this founder couple and each affected individual. When multiple alternative lines of the same length were present, a single line was selected at random. Filled squares (males) and circles (females) represent affected individuals in our sample.

TABLE II. Maximum Number of Affected Descendants for a Single Individual in Each Generation

Generations from present[a]	Maximum affected descendants (using strict diagnostic criteria)
1	2
2	3
3	3
4	3
5	3
6	4
7	8
8	8
9	16
10	26
11	39
12	52
13	63
14	78
15	83
16	87

[a]Our genealogical database includes overlapping generations. For the purpose of this table individuals in generation 0 are those with no descendants in the pedigree. For other individuals, "generations from present" is defined as the maximum number of meioses required for connecting each individual to a descendant in generation zero.

TABLE III. Completeness of the Sample

Generations from present	People[a]	Founders[b]	Fraction (%)[c]
0	268	0	0.0
1	287	19	6.6
2	471	109	23.1
3	662	294	44.4
4	680	154	22.6
5	904	138	15.3
6	1,131	112	9.9
7	1,251	139	11.1
8	1,098	137	12.5
9	834	149	17.9
10	583	168	28.8
11	374	173	46.3
12	202	134	66.3
13	92	53	57.6
14	60	42	70.0
15	32	29	90.6
16	6	6	100.0

[a]Total number of individuals in each generation from present (people).
[b]The number of individuals who have no other known ancestors in the pedigree (founders).
[c]The proportion of all individuals who are founders (fraction).

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

disease alleles in each affected descendant should be flanked by an rv8 cM (100 cM/12.5) stretch of founder chromosome [Boehnke, 1994]. The results of Boehnke [1994] assumed independent lines of descent between the original founder and each affected descendant. In contrast, the affected individuals in our sample share a significant proportion of their genealogy and a single recombination event can affect the extent of the original haplotype present in a large number of individuals, generating great variability between different replicates of our pedigree (see Fig. 2a,b). Despite this, our simulations are in good agreement with the results of Boehnke [1994] and show that, on average, disease alleles are flanked by 8.8 cM (±8.3 cM, standard deviation) of founder chromosome, although this value is inflated by a small number of descendants carrying very long stretches of the original chromosome (>40 cM).

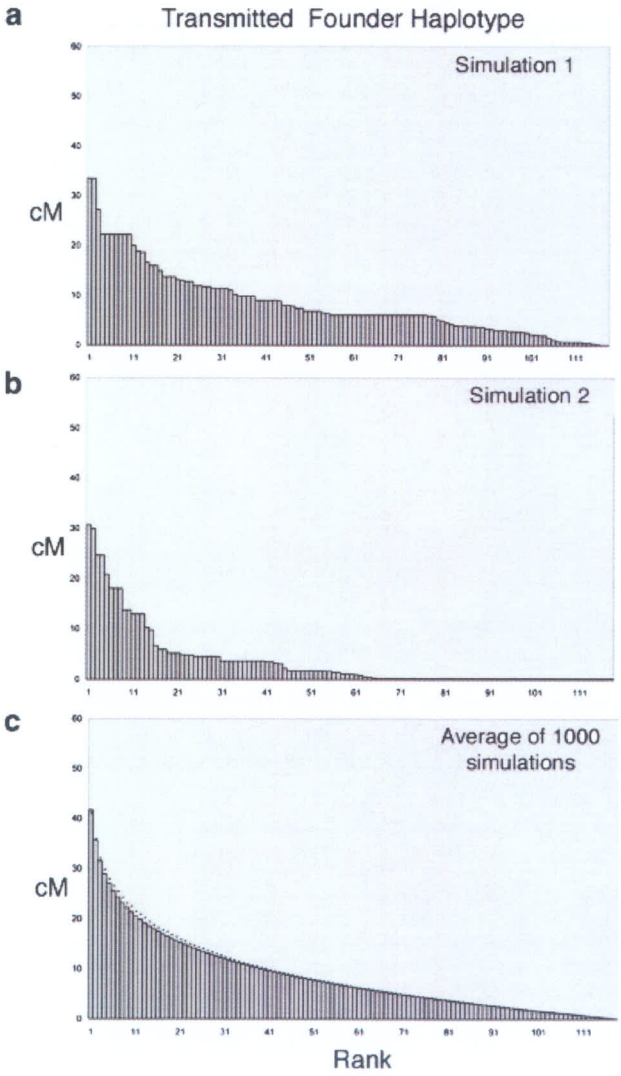


Fig. 2. The histogram illustrates the extent of the original founder chromosome flanking the shared disease allele in each affected individual. Panels (a) and (b) are the results of two representative replicates. Note how there are a number of discrete steps in the extent of the shared haplotype in each simulation. Each of these steps results from truncation of the ancestral haplotype by a single recombination event. Panel (c) summarizes the results of our 1,000 simulations. Filled bars represent the average extent of the original founder haplotype. Squares and lines represent the median and inter-quartile range.

On average (Fig. 2c), we found that the disease allele was flanked by rv6.4 cM or more of the original founder chromosome in half (60 of 121) of all affected individuals and by rv2.2 cM or more of the same chromosome in 80% of all affected individuals (96 of 121). We note that the shared genealogy between these individuals means that a small number of recombination events in key meioses can dramatically reduce the proportion of the founder haplotype transmitted to affected individuals (e.g., Fig. 2b). Nevertheless, in 95% of replicates the disease allele was flanked by >2.9 cM or more of the original founder chromosome in half of all affected carriers. In summary, our results suggest that shared disease alleles are likely to be flanked by >2 cM of the original founder chromosome in this pedigree. We found similar results for narrower phenotype definitions and for a small number of other founders with large numbers of affected descendants. These results suggest that approaches designed to search for haplotypes shared among affected individuals are feasible in this population using maps of microsatellite markers at 1–2 cM spacing.

CONCLUSIONS

The first set of results we present here is a clinical comparison between schizophrenia cases identified and diagnosed in the Afrikaner population and a sample of schizophrenia cases identified and diagnosed in the US. This comparison suggests that our genetic findings in the Afrikaner study will be applicable to other populations and schizophrenia in general. The diagnostic evaluations at the two sites were carried out by independent groups of clinicians specially trained in administering the DIGS. Eleven demographic and clinical diagnostic characteristics were stable across both populations and, with regard to basic sample descriptors and cardinal symptoms of disease, the two populations are equivalent. Three demographic and four clinical variables were different. It is not known whether these differences reflect genetic or cultural features of the Afrikaner sample.

The second set of results summarizes the genealogical research we have been able to perform for a large (and increasing) number of the schizophrenia families recruited to date. The consistency of our results suggests that it is feasible to trace genealogical relations with accuracy in the Afrikaner population. In addition, our genealogical data supports the hypothesis that a substantial portion of the schizophrenia cases we have identified will share the same or a small number of susceptibility alleles, since 87 affecteds traced so far link back to one founder couple. Establishment of detailed genealogical relationships between affecteds will allow us to cluster together patients who are more likely to share the genetic basis of the disease and search for segments of DNA shared IBD on a genome-wide level.

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

In principle, we expect to undertake this search in two stages. In a first stage, a map of widely spaced microsatellite markers (10–20 cM) should identify regions of the genome shared between close relatives, such as siblings. Statistical methods for evaluating linkage findings in small pedigrees are now mature and well-developed [e.g., Abecasis et al., 2002]. In a second stage, a search for shared founder haplotypes in genomic regions previously linked to schizophrenia, should allow us to hone in on disease alleles. Since disease alleles are likely to be flanked by >2 cM of the original founder haplotype, a moderate panel of microsatellite markers (at 1–2 cM spacing) should allow us to identify any founder haplotypes that are shared between more distant relatives. Statistical methods for comprehensively evaluating evidence for haplotype sharing are less well-developed and evaluating any evidence for a shared ancestral haplotype among affected individuals is likely to require an empirical standard, such as the extent of shared haplotypes in parental chromosomes not transmitted to affected individuals.

The number of meioses, patients, and segregating disease alleles in a pedigree are related to mapping power and resolution. In appropriate populations, the genealogical approach may represent an effective way of calibrating these quantities and a very powerful method for dissection of complex traits provided that appropriate statistical tests designed to harness this power are developed.

ACKNOWLEDGMENTS

The authors thank all individuals who have participated in this research. They also gratefully acknowledge nursing sisters Ria van Wyk, Cheryl Botha, and Glenda Jansen for assistance with family identification and recruitment, as well as Christina Seubert and Samantha Brenner for assistance with genealogical research and data entering.

REFERENCES

Abecasis GR, Cherny SS, Cookson WO, Cardon LR. 2002. Merlin-rapid analysis of dense genetic maps using sparse gene flow trees. Nat Genet 30:97–101.

American Psychiatric Association. 1994. Diagnostic and statistical manual. 4th edition. Washington, DC: American Psychiatric Association.

Boehnke M. 1994. Limits of resolution of genetic linkage studies, implications for the positional cloning of human disease genes. Am J Hum Genet 55:379–390.

Brink PA, Steyn LT, Coetzee GA, Van der Westhuyzen DR. 1987. Familial hypercholesterolemia in South African Afrikaners. PvuII and StuI DNA polymorphisms in the LDL-receptor gene consistent with a predominating founder gene effect. Hum Genet 77:32–35.

De Villiers CC. 1981. Genealogies of old South African families. Completely revised edition augmented and rewritten by Pama C. Rotterdam: A.A. Balkema.

Defesche JC, van Diermen DE, Lansberg PJ, Lamping RJ, Reymer PW, Hayden MR, Kastelein JJ. 1993. South African founder mutations in the low-density lipoprotein receptor gene causing familiar hypercholeserolemia in the Dutch population. Hum Genet 92:567–570.

Defesche JC, Van Diermen DE, Hayden MR, Kastelein JP. 1996. Origin and migration of an Afrikaner founder mutation FHAfrikaner-2. V408M. causing familial hypercholesterolemia. Gene Geogr 10:1–10.

DeLisi LE, Mesen A, Rodriguez C, Bertheau A, LaPrade B, Llach M, Riondet S, Razi K, Relja M, Byerley W, Sherrington R. 2002. Genome-wide scan for linkage to schizophrenia in a Spanish-origin cohort from Costa Rica. Am J Med Genet 114:497–508.

Devlin B, Bacanu SA, Roeder K, Reimherr F, Wender P, Galke B, Novasad D, Chu A, Cveno KT, Tiobek S, Otto C, Byerley W. 2002. Genome-wide multipoint linkage analyses of multiplex schizophrenia pedigrees from the oceanic nation of Palau. Mol Psychiatry 7:689–694.

Goldman A, Krause A, Ramsay M, Jenkins T. 1996. Founder effect and the prevalence of myotonic dystrophy in South Africans, Molecular studies. Am J Hum Genet 59:445–452.

Groenewald JZ, Liebenberg J, Groenewald IM, Warnich L. 1998. Linkage disequilibrium analysis in a recently founded population, evaluation of the variegate porphyria founder in South African Afrikaners. Am J Hum Genet 62:1254–1258.

Hall D, Wijsman EM, Roos JL, Gogos JA, Karayiorgou M. 2002. Extended intermarker linkage disequilibrium in the Afrikaners. Genome Res 12:956–961.

Hastbacka J, de la Chapelle A, Mahtani MM, Clines G, Reeve-Daly MP, Daly M, Hamilton BA, Kusumi K, Trivedi B, Weaver A, Coloma A, Lovett M, Buckler A, Kaitila I, Lander ES. 1994. The diastrophic dysplasia gene encodes a novel sulfate transporter, positional cloning by fine structure linkage disequilibrium mapping. Cell 78:1073–1087.

Hayden MR, MacGregor JM, Beighton PH. 1980a. The prevalence of Huntington's chorea in South Africa. S Afr Med J 58:193–196.

Hayden MR, Hopkins HC, Macrea M, Beighton PH. 1980b. The origin of Huntington's chorea in the Afrikaner population of South Africa. S Afr Med J 58:197–200.

Hovatta I, Varilo T, Suvisaari J, Terwilliger JD, Ollikainen V, Arajarvi R, Juvonen H, Kokko-Sahin ML, Vaisanen L, Mannila H, Lonnqvist J, Peltonen L. 1999. A genomewide screen for schizophrenia genes in an isolated Finnish subpopulation, suggesting multiple susceptibility loci. Am J Med Genet 65:1114–1124.

Kalaydjieva L, Hallmayer J, Chandler D, Savov A, Nikolova A, Angelicheva D, King RHH, Ishpekova B, Honeyman K, Calafell F, Shmarov A, Petrova J, Turnev I, Hristova A, Moskov M, Stancheva S, Petkova I, Bittles AH, Georgieva V, Middleton L, Thomas PK. 1996. Gene mapping in Gypsies identifies a novel demyelinating neuropathy on chromosome 8q24. Nat Genet 14:214–217.

Kendler KS, Myers JM, O'Neill FA, Martin R, Murphy B, MacLean CJ, Walsh D, Straub RE. 2000. Clinical features of schizophrenia and linkage to chromosomes 5q, 6p, 8p, and 10p in the Irish Study of High-Density Schizophrenia Families. Am J Psychiatry 157: 402–408.

Leitersdorf E, Van der Westhuyzen DR, Coetzee GA, Hobbs HH. 1989. Two common low density lipoprotein receptor gene mutations cause familial hypercholesterolemia in Afrikaners. J Clin Invest 84:954–961.

Nurnberger JI, Blehar MC, Kaufmann CA, York-Cooler C, Simpson SG, Harkavy-Friedman J, Severe JB, Malaspina D, Reich T, NIMH Genetics Initiative. 1994. Diagnostic interview for genetic studies, rationale, unique features, and training. Arch Gen Psychiatry 51:949.

Nystuen A, Benke PJ, Merren J, Stone EM, Sheffield VC. 1996. A cerebellar ataxia locus identified by DNA pooling to search for linkage disequilibrium in an isolated population from the Cayman Islands. Hum Mol Genet 5:525–531.

Paunio T, Ekelund J, Varilo T, Parker A, Hovatta I, Turunen JA, Rinard K, Foti A, Terwilliger JD, Juvonen H, Suvisaari J, Arajarvi R, Suokas J, Partonen T, Lonnqvist J, Meyer J, Peltonen L. 2001. Genome-wide scan in a nationwide study sample of schizophrenia families in Finland reveals susceptibility loci on chromosomes 2q and 5q. Hum Mol Genet 10:3037–3048.

Ploughman LM, Boehnke M. 1989. Estimating the power of a proposed linkage study for a complex genetic trait. Am J Hum Genet 44: 543–551.

Pronk JC, Gibson RA, Savoia A, Wijker M, Morgan NV, Melchionda S, Ford D, Temtamy S, Ortega JJ, Jansen S, Havenga C, Cohn RJ, de Ravel TJ, Roberts I, Westerveld A, Easton DF, Joenje H, Mathew CG, Arwert F. 1995. Localisation of the Fanconi anaemia complementation group A gene to chromosome 16q24.3. Nat Genet 11:338–340.

Roby P, Eyre S, Worthington J, Ramesar R, Cilliers H, Beighton P, Grant M, Wallis G. 1999. Autosomal dominant 'Beukes' premature degenerative osteoarthropathy of the hip joint maps to an 11-cM region on chromosome 4q35. Am J Hum Genet 64:904–908.

Rosendorff J, Bernstein R, Macdougall L, Jenkins T. 1987. Fanconi anaemia, another disease of high prevalence in the Afrikaans population of South Africa. Am J Med Genet 27:793–797.

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Sobin C, Blundell ML, Conry A, Weiller F, Gavigan C, Haiman C, Karayiorgou M. 2001. Early, non-psychotic deviant behavior in schizophrenia: A possible endophenotypic marker for genetic studies. Psychiatry Res 101:101 –113.

Visapaa I, Fellman V, Vesa J, Dasvarma A, Hutton JL, Kumar V, Payne GS, Makarow M, Van Coster R, Taylor RW, Turnbull DM, Suomalainen A, Peltonen L. 2002. Gracile syndrome, a lethal metabolic disorder with iron overload, is caused by a point mutation in BCS1L. Am J Hum Genet 71:863 –876.

Warnich L, Kotze MJ, Groenewald IM, Groenewald JZ, van Brakel MG, van Heerden CJ, de Villiers JN, van de Ven WJ, Schoenmakers EF, Taketani S, Retief AE. 1996. Identification of three mutations and associated haplotypes in the protoporphyrinogen oxidase gene in South Africa. Hum Mol Genet 5:981 –984.

Wijsman EM, Rosenthal EA, Hall D, Blundell ML, Sobin C, Heath SC, Williams R, Brownstein MJ, Gogos JA, Karayiorgou M. 2003. Genome-wide scan in a large complex pedigree with predominantly male schizophrenics from the island of Kosrae: Evidence for linkage to chromosome 2q. Mol Psychiatry 8:695 –705.

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

A5:63

Am. J. Hum. Genet. 74:403–417, 2004

Genomewide Scan in Families with Schizophrenia from the Founder Population of Afrikaners Reveals Evidence for Linkage and Uniparental Disomy on Chromosome 1

Gonçalo R. Abecasis,[1] Rachel A. Burt,[2] Diana Hall,[2] Sylvia Bochum,[1] Kimberly F. Doheny,[4] S. Laura Lundy,[2] Marie Torrington,[5] J. Louw Roos,[5] Joseph A. Gogos,[3] and Maria Karayiorgou[2]

[1]Department of Biostatistics, University of Michigan, Ann Arbor; [2]Human Neurogenetics Laboratory, The Rockefeller University, and [3]Department of Physiology and Cellular Biophysics and Center for Neurobiology and Behavior, College of Physicians and Surgeons, Columbia University, New York; [4]Center for Inherited Disease Research, Institute of Genetic Medicine, Johns Hopkins School of Medicine, Baltimore; and [5]University of Pretoria Department of Psychiatry and Weskoppies Hospital, Pretoria

We report on our initial genetic linkage studies of schizophrenia in the genetically isolated population of the Afrikaners from South Africa. A 10-cM genomewide scan was performed on 143 small families, 34 of which were informative for linkage. Using both nonparametric and parametric linkage analyses, we obtained evidence for a small number of disease loci on chromosomes 1, 9, and 13. These results suggest that few genes of substantial effect exist for schizophrenia in the Afrikaner population, consistent with our previous genealogical tracing studies. The locus on chromosome 1 reached genomewide significance levels (nonparametric LOD score of 3.30 at marker D1S1612, corresponding to an empirical *P* value of .012) and represents a novel susceptibility locus for schizophrenia. In addition to providing evidence for linkage for chromosome 1, we also identified a proband with a uniparental disomy (UPD) of the entire chromosome 1. This is the first time a UPD has been described in a patient with schizophrenia, lending further support to involvement of chromosome 1 in schizophrenia susceptibility in the Afrikaners.

Introduction

Schizophrenia (MIM 181500) is a severe mental disorder affecting 1% of the world population and is characterized by psychotic symptoms and by cognitive, affective, and psychosocial impairment. There is substantial evidence that schizophrenia has a genetic basis (Gottesman and Shields 1982; Tsuang and Faraone 1994). For example, concordant diagnosis of schizophrenia is more than two times higher for MZ twins (41%–65% concordance) than for DZ twins (0%–28% concordance) and other first-degree relatives of schizophrenic individuals (for a review, see Cardno and Gottesman [2000]). It seems clear that multiple gene loci and environmental factors influence susceptibility to schizophrenia. Understanding of the role of individual genes in the disorder, however, remains limited.

Over the past few years, several searches for susceptibility loci have been undertaken. Genomewide linkage scans have suggested that susceptibility loci may be present on several chromosomes (Coon et al. 1994; Moises et al. 1995; Blouin et al. 1998; Faraone et al. 1998; Levinson et al. 1998; Hovatta et al. 1999; Rees et al. 1999; Williams et al. 1999; Bailer et al. 2000; Brzustowicz et al. 2000; Ekelund et al. 2000; Camp et al. 2001; Gurling et al. 2001; Paunio et al. 2001; DeLisi et al. 2002*a*, 2002*b*; Devlin et al. 2002; Straub et al. 2002). For most implicated regions, both positive and negative replication studies have been reported. In addition to locus heterogeneity, differences in diagnostic and ascertainment procedures, analytical methods, and sample configuration and size may help explain differences in the results. Ethnic heterogeneity might also be a contributing factor, since families studied in the initial genome scans were drawn mostly from the genetically diverse populations of Europe and the United States.

More recently, genetic studies of schizophrenia have employed genetically isolated populations from Finland (Hovatta et al. 1999; Ekelund et al. 2000; Paunio et al. 2001), Sweden (Lindholm et al. 2001), Iceland (Stefansson et al. 2002), Quebec (Maziade et al. 2002), Costa Rica (DeLisi et al. 2002*a*), and two Micronesian islands: Palau (Camp et al. 2001; Devlin et al. 2002) and Kosrae (Wijsman et al. 2003). It is possible that, in these populations, there will be fewer susceptibility loci and alleles (de la Chapelle and Wright 1998; Ober and Cox 1998). In addition, founder populations often

Received August 11, 2003; accepted for publication November 20, 2003; electronically published January 28, 2004.

Address for correspondence and reprints: Dr. Maria Karayiorgou, Human Neurogenetics Laboratory, The Rockefeller University, 1230 York Avenue, Box 313, New York, NY 10021. E-mail: karayim@rockefeller.edu

© 2004 by The American Society of Human Genetics. All rights reserved.
0002-9297/2004/7403-0010$15.00

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

exhibit less environmental heterogeneity than do other populations and allow detailed genealogical research and reconstruction of extended multigenerational pedigrees (Peltonen et al. 2000). Availability of such large pedigrees should facilitate the detection of linkage (Wijsman and Amos 1997; Dyer et al. 2001). Despite these potential advantages, it must be noted that the outcome of studies of complex traits in an isolated population will depend greatly on the particular demographic characteristics and historical circumstances surrounding the founding of the population.

Here, we report the results of our initial genetic linkage studies of schizophrenia in the genetically isolated population of the Afrikaners from South Africa. The Afrikaners descended from a small number (1,000–2,000) of immigrants, primarily of Dutch origin, who settled in the Cape in 1652. These immigrants spread inland, with splinter groups forming small, geographically isolated communities. Because of language differences (the Afrikaans language is derived from Dutch) and religious practices (most Afrikaners were members of the Dutch Reformed Church), the Afrikaners remained isolated. The population expanded over 13–15 generations almost entirely through reproduction, since immigration subsequent to the founding was minimal. In the early generations, consanguinity was common. The Afrikaner population, which is currently 3 million, has expanded in relative genetic isolation, as is evident from the unusually high frequency of certain rare Mendelian disorders (e.g., variegate porphyria), the low diversity of the associated allelic variants, and the large extent (8–11 cM) of conserved haplotypes around disease genes (Hayden et al. 1980; Brink et al. 1987; Rosendorff et al. 1987; Leitersdorf et al. 1989; Torrington and Viljoen 1991; Brink et al. 1995; Pronk et al. 1995; Goldman et al. 1996; Warnich et al. 1996; Groenewald et al. 1998; Roby et al. 1999).

In a previous study, we described our ongoing efforts to document the genealogical history of a sample of Afrikaner individuals with schizophrenia (Karayiorgou et al. 2004). We traced the ancestry of 98 probands with schizophrenia. It was remarkable that we found that 87 of them descended from a single couple who immigrated to South Africa ~12.5 generations ago. This particular line of descent has not, to our knowledge, been described previously for any other disorder studied in the Afrikaner population (M.T., unpublished data). These results suggest that the affected individuals in this population may share a small number of disease alleles. In this setting, linkage and association scans hold exceptional promise, and the present article reports results from an initial scan for linkage in this population.

Material and Methods

Sample Recruitment

A collaborative project between the Laboratory of Human Neurogenetics at the Rockefeller University and two major psychiatric hospitals in South Africa was undertaken, with the goal of ascertaining and evaluating Afrikaner probands with a history of psychotic illness and their families for a genetic study. The participating hospitals were the Weskoppies Hospital in Pretoria, an academic hospital affiliated with the University of Pretoria, and the Valkenberg Hospital, which is affiliated with the University of Capetown. The Afrikaner families in our study were recruited from the regions of Tshwane (formerly known as Pretoria), Limpopo (formerly Northern Transvaal), and Mpumalanga (formerly the Eastern Transvaal), as well as from the greater Cape area. The study was approved by institutional review boards at all sites (Rockefeller University, University of Pretoria, and University of Capetown). Appropriate written informed consent was obtained from all study participants. A detailed description of our sample is available elsewhere (Karayiorgou et al. 2004).

Clinical Evaluation

The Diagnostic Interview for Genetic Studies (DIGS) (Nurnberger et al. 1994), after being translated and back-translated into Afrikaans, was the diagnostic instrument of choice. The DIGS was developed at the National Institute of Mental Health Diagnostic Centers for Psychiatric Linkage Studies and was specifically designed for the diagnosis of schizophrenia, schizoaffective disorder, major affective disorders, and anxiety disorders in genetic research. All diagnostic interviews were conducted in person by specially trained clinicians. Specifically, two psychiatrists and one clinical psychologist with a minimum of 10 years of clinical experience each were specially trained in the use of the DIGS and in the research application of the Diagnostic and Statistical Manual–4th Edition (DSM-IV) (American Psychiatric Association 1994). On the basis of information gathered in the DIGS, the clinical interviewers assigned appropriate diagnoses according to the DSM-IV. After each interview, diagnosticians completed (*a*) DSM-IV checklists for each category in which positive symptoms were identified in the DIGS; (*b*) a narrative chronologic summary of prodromal traits, symptom onset, and functional impairment; and (*c*) an English-language data summary form of selected DIGS items.

Ongoing reliability studies between the Afrikaner clinicians, the U.S. clinicians who conducted the diagnostic evaluations for our U.S.-based genetic schizophrenia sample, and the U.S. investigator who trained all the clinicians

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

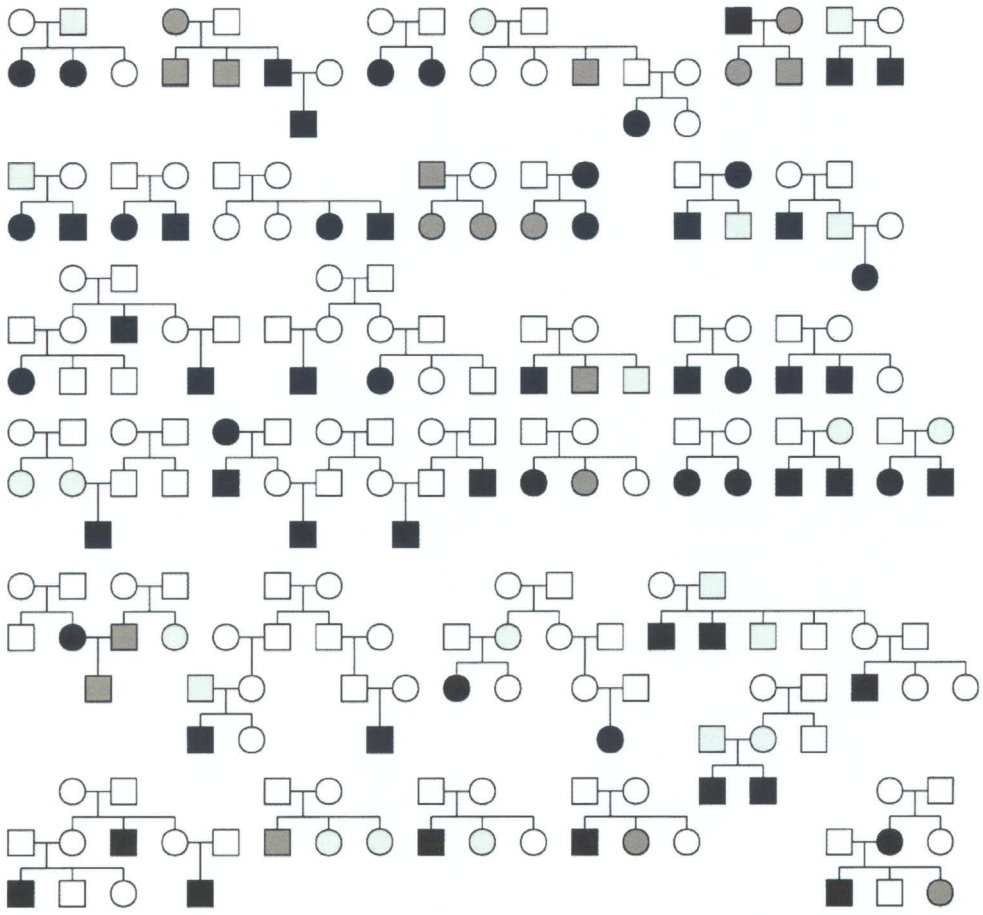


Figure 1 Informative pedigrees included in linkage analyses. Black coloring indicates affected individuals in LC I, and dark and light gray coloring indicate individuals in LCs II and III, respectively.

show >90% agreement for DSM-IV diagnoses. Reliability exercises consisted of yearly reliability interviews between all the clinicians, as well as review of videotapes of an additional two interviews per interviewer.

Affected Status Definitions

For the genetic linkage analysis, we used three phenotypic liability classes (LCs) (Straub et al. 1995; Kendler et al. 1996). LC I comprises the "core schizophrenic phenotypes" of schizophrenia and depressed-type-only schizoaffective disorder (family data suggest that the two diagnoses are alternative expressions of the same genotypes [Cloninger 1989]). LC II combines the LC I cases with all psychotic disorders, including primarily cases with schizoaffective disorder of mainly affective course, as well as some cases with bipolar disorder with psychotic symptoms, psychotic depression, and psychosis not otherwise specified (NOS). LC III combines the LC II cases with all psychiatric disorders, including depression, anxiety disorders, and ethanol dependence.

Family Sample

The entire genotyped sample includes 173, 205, and 253 individuals in LCs I, II, and III, respectively. These individuals are organized into 143 small families and >137 relative pairs (including 79 parent-child pairs, 35 sib pairs, 15 avuncular pairs, 4 grandparent-grandchild pairs, and 4 cousin pairs). Thirty-four of these families are informative for linkage (fig. 1). The remaining 110 are useful for allele frequency estimation and transmission disequilibrium testing. This analysis will be described elsewhere. The 34 families that are informative for linkage include 62, 81, and 100 individuals in LCs I, II, and III, respectively. Of the 19 cases included in LC II but not LC I, 10 meet criteria for schizoaffective

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

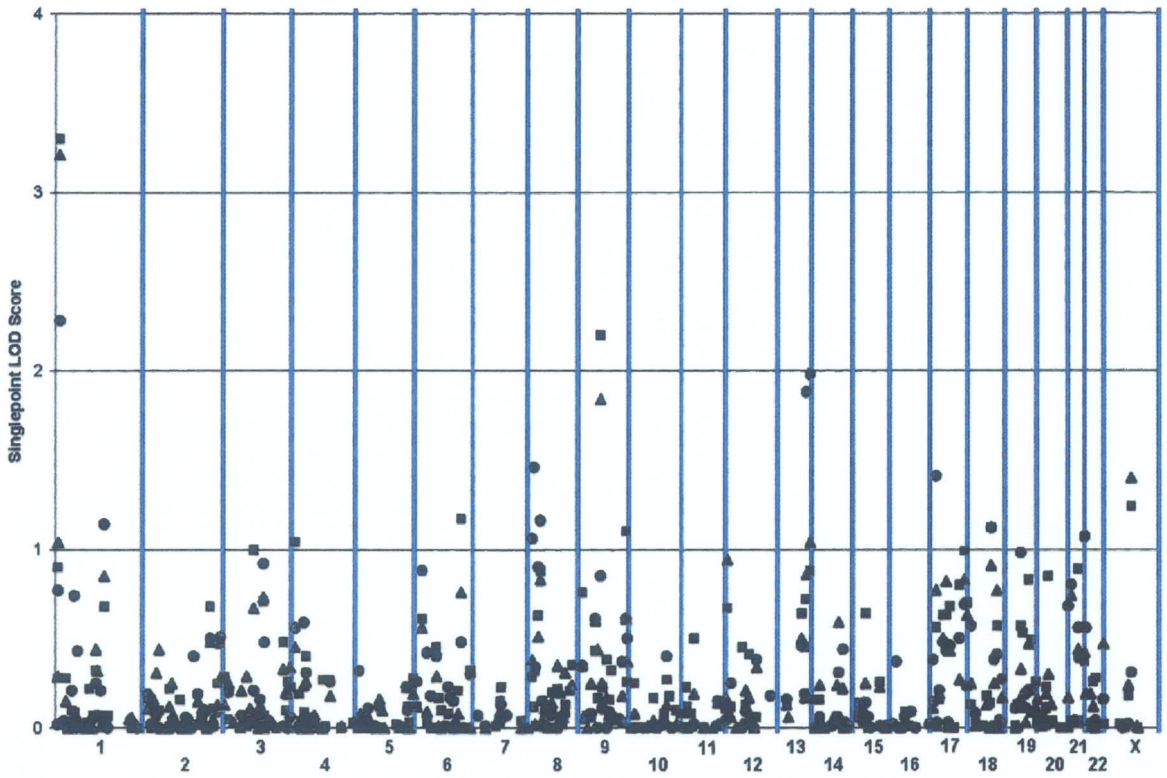


Figure 2 Single-point analysis results. Single-point Kong and Cox (1997) nonparametric LOD scores calculated using the S_{ALL} scoring statistic are shown. Results are summarized throughout the genome for three phenotypic LCs: LC I (*circles*), LC II (*triangles*), and LC III (*squares*).

disorder of mainly affective course, and 9 meet the criteria for other psychotic disorders.

Genomewide Screening

DNA from all study participants was extracted from 24 ml of EDTA-treated blood, according to standard procedures (Ciulla et al. 1988). In addition to affected individuals, we collected DNA samples for genotyping from first-degree relatives (parents and siblings) whenever possible. A genomewide screen was performed at the Center for Inherited Disease Research (CIDR). A total of 602 samples was screened using 388 fluorescent microsatellite markers. The marker set was composed primarily of trinucleotide and tetranucleotide repeat markers and was based on the Marshfield Genetics version 8 screening set, available for purchase from Research Genetics (~10% of the marker loci are different between the CIDR marker set and Marshfield version 8) (Center for Medical Genetics Web site; CIDR Human Marker Set Web site). The average spacing of the markers used was 9 cM, with one gap >20 cM. The average heterozygosity of the markers used was 0.76. CEPH samples 1331-1 and 1331-2 were included as size standards in all PCRs and gel runs. Internal standards were also included in each gel lane. Gel electrophoresis was performed on an ABI Prism 377 DNA sequencer instrument (Perkin Elmer), and genotypes were assigned using Genotyper 2.0 software (Perkin Elmer). Two individuals scored the alleles independently. Discrepancies were flagged and resolved.

Ten samples did not amplify for >25% of the markers and were dropped from the analysis. Of these 10, only one sample was from an affected proband; the other 9 were samples from unaffected relatives or parents. The overall error rate was 0.09%, and the inconsistency rate was 0.30%. The missing data rate was 6.6%.

Error Checking

We verified familial relationships through use of the GRR (Abecasis et al. 2001*b*) and RELPAIR (Epstein et al. 2000) programs. Four families showed segregation problems that were consistent with a sample switch in two of the families and with nonpaternity in the other two. After resolving problems in family structure, we

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Table 1

Peak Single-Point LOD Scores Observed in the Nonparametric Linkage Analysis

Chromosome and Marker[a]	Location[b] (cM)	LOD Score for LC I	LC II	LC III
1:				
D1S2660	11	...	1.04	...
D1S1612	16	2.28	3.21	3.30
D1S1653	164	1.14	...	...
3:				
D3S2406	103	...	...	1.00
4:				
D4S2366	13	...	...	1.04
6:				
D6S2436	155	1.08	...	...
8:				
D8S1469	16	1.06	...	...
D8S1130	22	1.46	...	...
D8S136	44	1.16	...	...
9:				
D9S1122	76	...	1.84	2.20
D9S1826	160	...	...	1.10
13:				
D13S1265	99	1.88	1.04	...
D13S285	111	1.98	...	...
17:				
D17S974	22	1.41	...	...
18:				
D18S858	80	1.12	...	1.12
21:				
D21S1446	58	1.07	...	...
X:				
DXS6800	93	...	1.40	1.24

[a] All markers with a LOD score (Kong and Cox 1997) >1.0 in the single-point analyses are shown.

[b] Marker locations are given according to the Marshfield genetic map.

used MERLIN (Abecasis et al. 2002) to check for Mendelian inconsistencies and to identify genotypes associated with an excessive number of recombination events.

Nonparametric Linkage Analysis

We performed genomewide, nonparametric linkage analyses using the S_{ALL} statistic (Whittemore and Halpern 1994), as recommended by Sengul et al. (2001). This statistic combines information from all affected individuals in a pedigree and can detect regions of excess identity-by-descent sharing. We conducted these analyses with the MERLIN computer program (Abecasis et al. 2002) and calculated LOD scores as described elsewhere (Kong and Cox 1997). We performed both multipoint and single-point analyses throughout the genome. Although multipoint linkage analysis is theoretically more powerful, it is also more sensitive to errors in genotype data or genetic map specification (Abecasis et al. 2001*a*; Sullivan et al. 2003). In contrast, single-point analysis is less efficient and more prone to random stochastic fluctuations, but it is more robust to genotyping error and does not require specification of genetic maps (Abecasis et al. 2001*a*; Sullivan et al. 2003).

Parametric Linkage Analysis

We repeated the linkage analyses using a set of four parametric models for schizophrenia, as described by Brzustowicz et al. (2000), corresponding to dominant and recessive models for LCs I and II. For LC I, when a dominant model was assumed, we set the allele frequency at $p = 0.0045$ with penetrance vector $\mathbf{f} = [0.75, 0.50, 0.001]$. If we assume a recessive model, $p = 0.065$ and $\mathbf{f} = [0.50, 0.0015, 0.0015]$. Elements of the penetrance vector $\mathbf{f}$ are ordered to denote the probability of disease for genotypes with frequency p^2, $2(1-p)p$, and $(1-p)^2$, respectively. For LC II, when a dominant model was assumed, $p = 0.007$ and $\mathbf{f} = [0.90, 0.80, 009]$. For the recessive model, $p = 0.10$ and $\mathbf{f} = [0.60, 0.01, 0.01]$. These analyses were conducted with the Allegro computer program (Gudbjartsson et al. 2000).

Empirical Significance Levels

We calculated genomewide significance levels through use of 5,000 gene-dropping simulations. In each simulation, we retained the original phenotypes and generated a new data set with the same allele frequencies, marker spacing, phenotypes, and missing data pattern. Any evidence for linkage in these simulated data is due to chance. We analyzed each replicate and recorded the highest peak for each chromosome, to evaluate the false-positive rate. These simulations allowed us to account for uneven marker spacing and informativeness and for the diversity of family structures in our study, as well as to calculate the probability of observing multiple linkage peaks of a certain height. For a discussion of the utility of empirical significance levels in linkage analysis, see Kruglyak and Daly (1998).

Fine Mapping

A denser marker map was generated for the three chromosomal regions where markers showed a LOD score >1.5 in the original scan. The additional markers analyzed were selected from the Marshfield genetic maps, and their order was cross checked with the current genome assembly prior to analysis. CEPH samples 1331-1 and 1331-2 were included as size standards in all PCRs and gel runs. Internal standards were also included in each gel lane. Gel electrophoresis was performed on an ABI 3700 DNA sequencer instrument at the Starr Center Genotyping Facility at the Rockefeller University. Genotypes were assigned using Genotyper 2.5 software (Perkin Elmer). Two individuals scored the alleles independently. Discrepancies were flagged and resolved.

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Table 2

Empirical Genomewide *P* Values for Markers with Observed Nonparametric Peak LOD Scores >1.0

LC and Rank (r)[a]	Marker	Location[b]	Observed LOD Score (Z)	Expected[c] E(No. of LOD Scores ≥Z)	Expected[c] SD	P (r LODs ≥Z)[d]
LC I:						
1	D1S1612	16	2.28	.07	.27	.072
2	D13S285	111	1.98	.40	.62	.060
3	D8S1130	22	1.46	1.67	1.23	.230
4	D17S974	22	1.41	1.94	1.32	.122
5	D18S858	80	1.12	4.16	1.81	.406
6	D6S2436	155	1.08	4.56	1.88	.294
7	D21S1446	58	1.07	4.66	1.90	.163
LC II:						
1	D1S1612	16	3.21	.01	.08	.006
2	D9S1122	76	1.84	.75	.84	.171
3	D13S1265	99	1.04	4.74	1.89	.887
LC III:						
1	D1S1612	16	3.30	.01	.08	.007
2	D9S1122	76	2.20	.20	.54	.035
3	D18S858	80	1.12	3.85	1.74	.771
4	D4S2366	13	1.04	4.54	1.83	.698
5	D3S2406	103	1.00	4.93	1.89	.571

[a] For each phenotypic liability class, the table ranks markers exhibiting the strongest evidence for linkage.

[b] Chromosomal locations are given in centimorgans only; chromosome assignments are implicit in the marker name.

[c] The expected number of chromosomes exhibiting equal or greater LOD scores (and its SD).

[d] The probability that ≥r LOD scores of equal or greater magnitude would be observed by chance. The probabilities and expected values were estimated from 5,000 sets of genotypes generated under the null (see text). This simulation considered only autosomes 1–22.

Overall, 19 additional markers (D1S2893, D1S2795, D1S2694, D1S1160, D1S503, D1S2736, D1S434, D1S507, D9S175, D9S153, D9S152, D13S174, D13S248, D13S778, D13S1295, D13S1315, D13S261, D13S293, and D13S1825) were investigated in regions showing evidence of linkage on chromosomes 1, 9, and 13. Marker D1S2694 showed clear evidence for departure from Hardy-Weinberg equilibrium ($P < .0007$) and was excluded from further analyses.

Results

Nonparametric Linkage Analysis

We initially performed single-point, nonparametric linkage analysis (Whittemore and Halpern 1994) using the NPL-ALL statistic (as recommended by Sengul et al. 2001). NPL Z scores were converted into LOD scores, as described elsewhere (Kong and Cox 1997). A summary of all estimated LOD scores is given in figure 2. In addition, observed LOD scores of ≥1.0 are summarized in table 1. Overall, we observed the strongest evidence for linkage on chromosome 1 at marker D1S1612 (LOD score 2.28, asymptotic $P = .0006$ for LC I; LOD score 3.21, $P = .00006$ for LC II; and LOD score 3.30, $P = .00005$ for LC III). Markers D9S1122 on chromosome 9 (LOD score 1.84, $P = .002$ for LC II; and LOD score 2.20, $P = .0007$ for LC III) and D13S1265 on chromosome 13 (LOD score 1.88, $P = .002$ for LC I) also exhibit strong evidence for linkage.

Empirical Significance Levels

We used gene-dropping simulations to estimate genomewide P values for our observed linkage signals. The results of these simulations are summarized in table 2. In our sample, the probability of observing one or more chromosomes with a LOD score ≥3.30 for LC III (the LOD score we observed at D1S1612) by chance is $P = .007$ (see table 2). Less than 1 such signal is expected to occur by chance in every 100 genome scans. Even after accounting for our use of three different phenotypic liability classes, the probability of observing a LOD score of ≥3.30 is only .012.

For LCs I and II, the highest observed LOD scores are 2.28 and 3.21, respectively, and the probability of observing one or more chromosomes with equal or greater LOD scores by chance is $P = .072$ and $P = .006$, respectively. It is noteworthy that, for LC I, we observed evidence for linkage in more chromosomes than ex-

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Table 3

Peak Multipoint LOD Scores from the Nonparametric Linkage Analysis

Chromosome and Nearest Marker	LC I			LC II			LC III		
	LOD	Location[a] (cM)	Interval[b] (cM)	LOD	Location[a] (cM)	Interval[b] (cM)	LOD	Location[a] (cM)	Interval[b] (cM)
1:									
D1S1612	1.48	16	13–21	2.99	16	9–25	2.44	16	9–25
9:									
D9S1118	...	...	...	1.19	61	59–64	...	...	...
D9S1122	...	...	...	1.32	76	69–78	1.57	76	71–92
13:									
D13S285	2.23	109	93–111	1.53	110	101–111	1.31	109	102–111
17:									
D17S2196	...	...	...	...	...	...	1.13	45	42–46
D17S928	...	...	...	...	...	...	1.04	123	120–126
21:									
D21S1446	1.40	56	46–58	...	...	...	...	...	...

Note.—The table lists all linkage peaks with a LOD score (Kong and Cox 1997) >1.0 in the multipoint analyses.
[a] LOD score location (according to the Marshfield genetic map) for each peak.
[b] Interval in which the LOD score is >1.0.

pected by chance. For example, at LOD score thresholds of 2.28, 1.98, 1.41, and 1.08, we observed evidence for linkage in one, two, four, and six chromosomes, respectively. In contrast, we would expect to observe evidence for linkage in only 0.07, 0.40, 1.97, and 4.56 chromosomes at each of these thresholds. Although this excess of linkage peaks does not reach genomewide significance levels, it is consistent with an oligogenic basis for schizophrenia, which would produce modestly increased allele sharing at a number of loci rather than large increases in sharing at a few loci (Cookson et al. [2001] present a similar analysis for eczema).

Table 2 also illustrates the utility of calculating analysis- and data-specific thresholds when evaluating genomewide significance levels. For example, at threshold 1.12, 4.16 linkage peaks are expected to occur by chance for LC I, but only 3.85 such peaks are expected for LC III. This is the result of differences in informativeness between the two phenotypic liability classes in our sample (for example, there are more affected individuals per family for LC III, and the proportion of different types of relative pairs is different for each LC).

Multipoint Nonparametric Linkage Analysis

Multipoint results are largely consistent with the single-point analyses (table 3; fig. 3). Strong evidence for linkage is observed on chromosome 1 (near marker D1S1612 at 16 cM), with LOD scores of 1.48 (P = .005), 2.99 (P = .0001), and 2.44 (P = .0004), for LCs I, II, and III, respectively. As in the single-point analyses, two other modest linkage peaks were observed on chromosomes 9 and 13, with peak LOD scores of 1.57 (P = .004) for LC III and 2.23 (P = .0007) for LC I, respectively. Concordance between the multipoint and single-point analyses increases our confidence that these linkage signals are real and not artifacts generated by genotyping or map error.

In inbred samples, LOD scores may be inflated (Genin and Clerget-Darpoux 1996). This is a phenomenon that could result in misleading inferences about linkage. To ensure that this was not the case in the sample of nuclear families we selected for genotyping, we calculated average nonparametric LOD scores (Kong and Cox 1997) at 1-cM intervals throughout the genome, through use of the LC III phenotype definition (data not shown). At each location, we set the sign of the LOD score to positive if there was evidence for excess sharing among affected individuals and to negative otherwise. We observed an average LOD score of 0.03, suggesting that inbreeding is not a serious cause for concern in interpreting our results.

Parametric Linkage Analysis

We first calculated parametric LOD scores for markers on chromosomes 1, 9, and 13, which exhibited the strongest evidence for linkage in multipoint and single-point nonparametric analyses. We considered four previously described disease models (Brzustowicz et al. 2000) and performed analyses both with and without a heterogeneity parameter. We speculated that these analyses might help evaluate different modes of action for susceptibility genes that may be segregating in our population.

Our peak LOD scores in single-marker analyses were 3.26 for marker D1S1612 (chromosome 1, 16 cM, LC II), 2.17 for marker D9S1122 (chromosome 9, 76 cM, LC II), and 1.79 for marker D13S285 (chromosome 13, 111 cM, LC I). Multipoint results also confirmed the

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

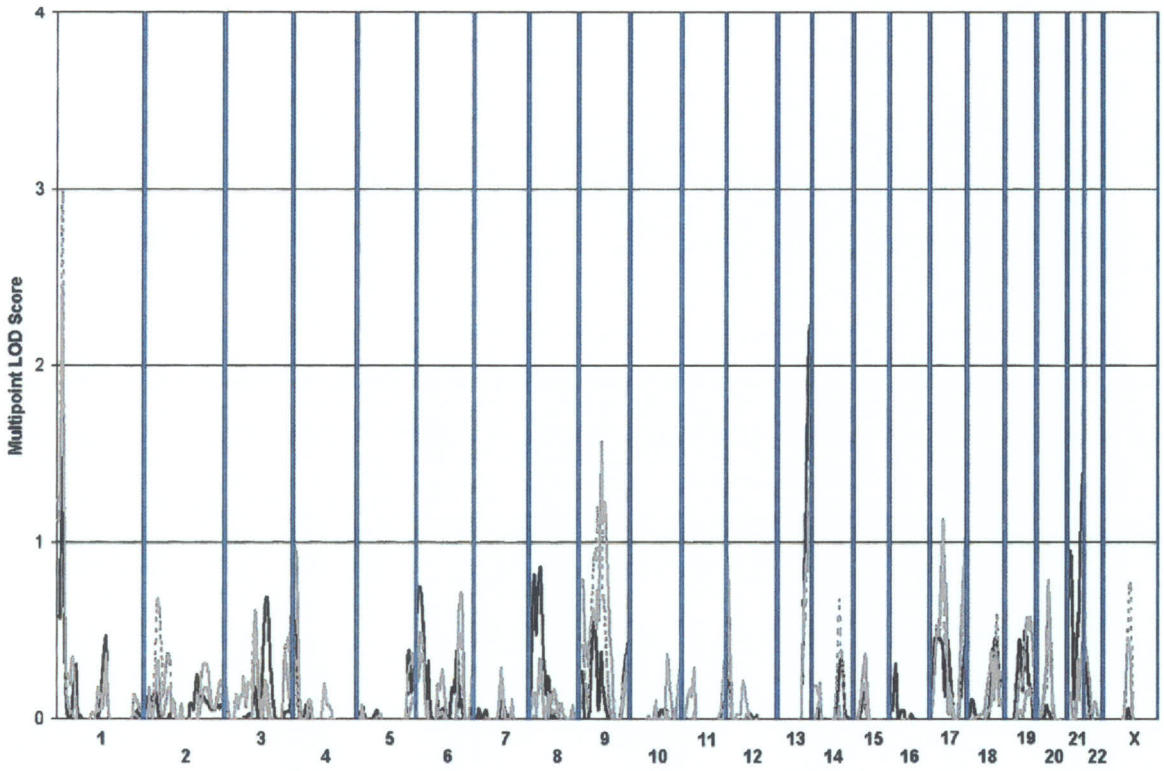


Figure 3 Multipoint analysis results. Multipoint Kong and Cox (1997) nonparametric LOD scores calculated using the S_{ALL} scoring statistic are shown. Results are summarized throughout the genome for three phenotypic LCs: LC I (*solid black line*), LC II (*dashed line*), and LC III (*solid gray line*).

results of our nonparametric analyses, although observed LOD scores were somewhat higher on chromosome 13, where the peak LOD score increased to 3.11 (chromosome 13, 106 cM, LC I), and somewhat lower on chromosomes 1 and 9, with peak LOD scores of 2.95 (chromosome 1, 16 cM) and 1.59 (chromosome 9, 75 cM). Analyses allowing for heterogeneity raised observed LOD scores only slightly.

A summary of our parametric analyses, including details of linkage peaks for each chromosome and analysis model, is provided in table 4. Although we observed the strongest evidence for linkage when we assumed a recessive disease model with high penetrance, it is important to note that small pedigrees have very limited utility for distinguishing between different disease models (for a detailed discussion of the issue, see Göring and Terwilliger [2000]), and one must be cautious in assuming that this model is closest to the truth.

We also performed parametric linkage analyses for the rest of the other autosomes. In single-marker analyses, the three markers exhibiting the strongest evidence for linkage were D3S2460 (heterogeneity LOD score [HLOD] 2.34 with $\alpha = 0.29$, using LC I and the recessive model), D18S858 (HLOD 2.28 with $\alpha = 0.10$, using LC II and the recessive model), and D8S1130 (HLOD 1.70 with $\alpha = 0.33$, using LC I and the recessive model). Parametric multipoint analyses did not identify any regions outside chromosomes 1, 9, and 13 where LOD scores were >1.5 or HLODs were >1.5. Detailed tables, including all our parametric linkage results, are available online (authors' Web site).

Fine Mapping

We genotyped additional markers around our chromosome 1, 9, and 13 linkage peaks. We reasoned that these markers might help extract additional linkage information in these regions. In total, we genotyped eight new markers around each of our two highest peaks, on chromosomes 1 and 13, and three additional markers around the chromosome 9 linkage peak. These chromosomes were selected because they provided suggestive evidence for linkage (LOD score >1.5) in parametric and nonparametric analyses, whether single-point or multipoint.

Overall, evidence for linkage to chromosome 1 weak-

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Table 4

Multipoint and Single-Point Parametric Linkage Analysis Results for Chromosomes 1, 9, and 13

	Homogeneity				Heterogeneity					
	Single-Point		Multipoint		Single-Point			Multipoint		
Chromosome, LC, and Model	Marker (Position in cM)	LOD[a]	Position (cM)	LOD[a]	Marker (Position in cM)	α[b]	HLOD[c]	Position (cM)	α[b]	HLOD[c]
1:										
LC I:										
Dominant	D1S1653 (164)	1.94	14	.26	D1S1653 (164)	.00	1.94	14	.37	.86
Recessive	D1S1612 (16)	2.97	16	1.34	D1S1612 (16)	.11	3.04	16	.33	2.17
LC II:										
Dominant	D1S1612 (16)	1.67	14	1.49	D1S1612 (16)	.20	1.98	14	.21	1.74
Recessive	***D1S1612 (16)***	***3.26***	***16***	***2.95***	***D1S1612 (16)***	***.00***	***3.26***	***16***	***.13***	***3.04***
9:										
LC I:										
Dominant	D9S1122 (76)	−.24	50	−2.15	D9S1122 (76)	.34	1.06	76	.73	.28
Recessive	D9S1826 (160)	.69	51	−.35	D9S1826 (160)	.27	1.27	164	.51	1.42
LC II:										
Dominant	D9S1122 (76)	.84	50	−1.99	D9S1122 (76)	.28	1.50	76	.62	.55
Recessive	***D9S1122 (76)***	***2.17***	***75***	***1.59***	***D9S1122 (76)***	***.08***	***2.20***	***75***	***.24***	***1.88***
13:										
LC I:										
Dominant	D13S1265 (99)	.64	105	1.84	D13S1265 (99)	.32	1.48	104	.12	1.86
Recessive	***D13S285 (111)***	***1.79***	***106***	***3.11***	***D13S1265 (99)***	***.34***	***2.26***	***106***	***.09***	***3.15***
LC II:										
Dominant	D13S894 (33)	−1.78	106	−.59	D13S1265 (99)	.60	.66	104	.54	.78
Recessive	D13S1265 (99)	1.78	107	2.48	D13S1265 (99)	.28	2.15	107	.13	2.54

Note.—Boldface italic type indicates the combinations of LC and disease model that resulted in the strongest evidence for linkage.

[a] Parametric LOD score calculated under the assumption of no heterogeneity.

[b] Estimated value at each location.

[c] Parametric LOD score maximized over a heterogeneity parameter α.

ened somewhat, with the peak nonparametric multipoint LOD score decreasing from 2.99 (for LC II at 16 cM) to 2.36 (also for LC II, but at 15 cM). In contrast, evidence for linkage on chromosomes 9 and 13 increased. On chromosome 9, the peak nonparametric LOD score increased from 1.57 (for LC III at 76 cM) to 1.83 (also for LC III, but at 87 cM). On chromosome 13, the peak nonparametric LOD score increased from 2.23 (for LC I at 109 cM) to 3.01 (also for LC I, at 111 cM). A comparison of our linkage peaks in the original and fine-mapping panels is provided in figure 4. We cautiously interpret these results as corroboration of our linkage peaks.

Chromosome 1 Isodisomy

Close inspection of the Mendelian inconsistencies in our sample revealed that 13 of them occurred in a single family, all for markers on chromosome 1. Mendelian consistency for all markers on the X chromosome and for autosomes 2–22 suggests that nonpaternity (or nonmaternity) is not an issue. In this family, the proband is homozygous for all chromosome 1 markers and does not carry a maternally derived allele at 13 of these markers (fig. 5*A*). The genotypes for 30/32 chromosome 1 markers are consistent with a paternal isodisomy for the entire chromosome; genotypes are not available for the other two markers. This observation is intriguing, in light of the fact that our strongest linkage signal is also on chromosome 1 and our parametric analyses suggest a recessive mode of inheritance (all our peak LOD scores were observed with the recessive model). It is interesting that genealogical tracing revealed that the isodisomic individual belongs to a previously described cluster (Karayiorgou et al. 2004) of 87 individuals meeting our strict definition for schizophrenia (LC I) and descending from a single founder couple. The same cluster currently includes 12 of the 34 families we used in our linkage analyses (fig. 5*B*).

Discussion

Current attempts to identify genetic loci for schizophrenia include the use of founder populations as a way to limit the genetic and environmental heterogeneity. Isolated populations established by a limited number of founders have proven useful for mapping genes of rare monogenic disorders (Hastbacka et al. 1994; Kalaydjieva et al. 1996; Nystuen et al. 1996; Visapaa et al. 2002). Their usefulness in mapping complex traits is still largely untested. Although it is perhaps unreasonable to

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

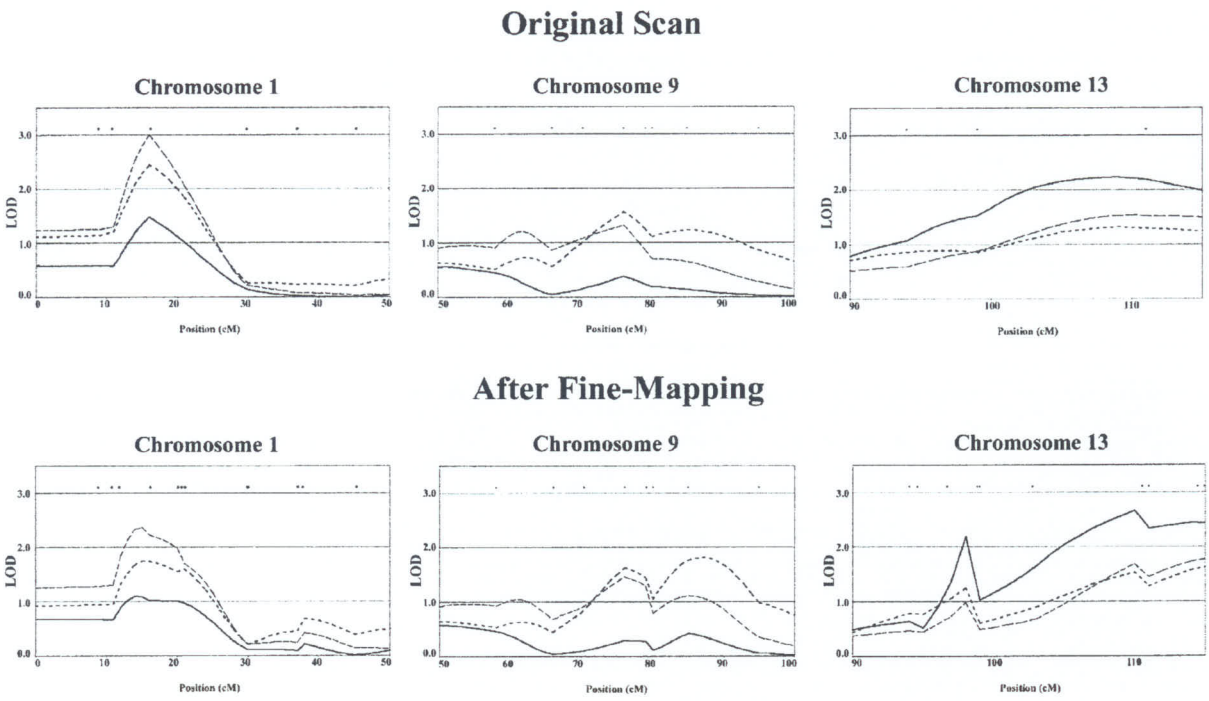


Figure 4 Comparison of original and fine-mapping results. Multipoint Kong and Cox (1997) nonparametric LOD scores calculated using the S_{ALL} scoring statistic are shown. Results are summarized for three regions exhibiting strongest evidence for linkage, on chromosomes 1, 9, and 13. The three phenotypic LCs are LC I (*solid black lines*), LC II (*lines with long dashes*), and LC III (*lines with short dashes*).

expect that there be one major (monogenic disorder–like) susceptibility locus in genetic studies of complex traits such as schizophrenia, one can reasonably expect a smaller number of contributing loci and a smaller number of disease alleles at each locus for at least some founder populations.

Here, we report initial results from a 10-cM genomewide scan of families with schizophrenia from the Afrikaner population. Thirty-four two- to four-generation pedigrees with at least one individual meeting strict DSM-IV criteria for schizophrenia or schizoaffective disorder were informative for linkage. Using both nonparametric and parametric linkage analyses, we obtained evidence for a small number of disease loci on chromosomes 1, 9, and 13. These results are consistent with the notion that genes of substantial effect exist for schizophrenia in the Afrikaner population.

Our chromosome 9 and 13 findings do not reach genomewide significance levels but do show suggestive overlap with other studies. The locus on chromosome 9 overlaps with the locus described by Hovatta et al. (1999) in families from Finland. In the Finnish study, a maximum LOD score of 1.95 was obtained with marker D9S922 (3.4 cM telomeric to our most significant marker, D9S1122). It is interesting that positive findings in this region have been reported in the genome scans of Moises et al. (1995) and Levinson et al. (1998), 6 cM centromeric and 16 cM telomeric to our marker D9S1122, respectively.

The locus on chromosome 13 is slightly distal to the locus described by Blouin et al. (1998). Our most positive marker, D13S285, maps 14 cM telomeric to D13S174, the most positive marker in the Blouin et al. (1998) study. Recent results from a linkage disequilibrium–based positional cloning effort centered around marker D13S174 suggest an association between schizophrenia susceptibility and the *G72* gene, which encodes a protein that may interact and modulate the activity of D-amino-acid-oxidase (Chumakov et al. 2002). Recent association studies suggest a potential contribution of *G72* to bipolar disorder as well (Hattori et al. 2003). *G72* is located 3.2 Mb and 7 Mb proximally from markers D13S1265 and D13S285, respectively, which gave the best linkage results in our studies. Association studies in an extended sample of Afrikaner families revealed a weak contribution of the *G72* gene in our sample (D.H., J.A.G., and M.K., unpublished data), but it is not yet clear whether this can account for the observed linkage signal.

We also obtained strong evidence for an apparently novel locus on chromosome 1. Although other studies

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

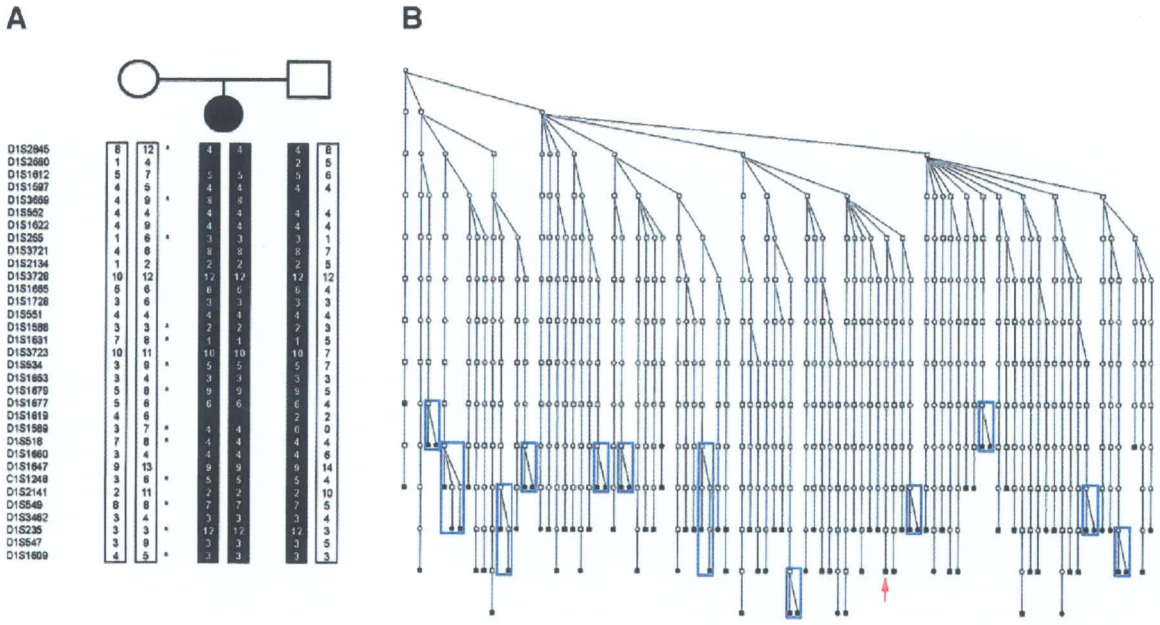


Figure 5 Isodisomic individual. *A*, Chromosome 1 genotypes for isodisomic individual and two parents. The isodisomic chromosome is shaded in black, and asterisks (*) indicate markers where mother and offspring genotypes are incompatible. *B*, Shortest connection between individuals in LC I and a single ancestor who emigrated to South Africa ~12 generations before present (Karayiorgou et al. 2004). Most affected individuals in our sample are connected to this founder through multiple descent routes. Twelve small pedigrees suitable for linkage analysis and included in our sample are boxed in the diagram. The cluster includes additional informative families when LCs II and III are used. The isodisomic individual is indicated by an arrow (↑).

have implicated chromosome 1 in schizophrenia, the region identified in the present study is quite distant from the previously implicated regions, ~148–154 cM from the locus described by Brzustowicz et al. (2000) and ~208 cM from the locus described by Hovatta et al. (1999), both of which are in the q arm of chromosome 1. Our study demarcates a region on the p arm of chromosome 1 at ~16 cM, which has not been implicated previously, with the possible exception of a recent genome scan in Icelandic families (Stefansson et al. 2002), in which a region on chromosome 1 that appears to overlap with ours provided a weak linkage signal.

It is most interesting that we identified one proband with uniparental disomy (UPD) for paternal chromosome 1. UPD, which was first described by Engel (1980), is a rare condition in which a diploid individual inherits two copies of a specific chromosome from one parent and no copy from the other parent. UPD may lead to rare recessive disorders or to developmental disturbances due to aberrant imprinting effects (Ledbetter and Engel 1995), or it may have no apparent phenotypic effects. There are very few reported cases of UPD of chromosome 1 (UPD 1), involving full or partial isodisomy. These include a small number of cases of maternal UPD 1 (Pulkkinen et al. 1997; Field et al. 1998; Dufourcq-Lagelouse et al. 1999), as well as a few cases of paternal UPD 1 (Gelb et al. 1998; Chen et al. 1999; Miura et al. 2000; Takizawa et al. 2000; Thompson et al. 2002). Among the latter, UPD 1 has been associated with (*a*) retinal dystrophy due to mutations in the *RPE65* gene (Thompson et al. 2002), (*b*) congenital insensitivity to pain with anhidrosis due to mutations of the *TRKA* gene (Miura et al. 2000), (*c*) Herlitz junctional epidermolysis bullosa due to mutation in *LAMC2* (Takizawa et al. 2000), or (*d*) pycnodysostosis due to mutations in the cathepsin K gene (Gelb et al. 1998).

All described individuals with UPD 1 developed normally, suggesting that chromosome 1 does not carry developmentally important maternally imprinted genes, although the existence of maternally imprinted genes with subtler effects on brain function and behavior cannot be excluded. The proband described here also had normal development (J.L.R. and M.K., unpublished data). To our knowledge, this is the first time a UPD of any chromosome has been described in a patient with schizophrenia and the second time a UPD has been discovered during a genome screen linkage study (the other one was described by Field et al. [1998]). On the basis

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

of the striking coincidence of linkage and UPD on chromosome 1 and the genealogical evidence, it is tempting to speculate that this individual might be homozygous for a chromosome 1 schizophrenia-susceptibility locus segregating in the Afrikaner population. If this were the case, we might be able to exploit this fortuitous chromosomal abnormality in our attempts to identify and refine an Afrikaner risk haplotype on chromosome 1. Specifically, additional genotyping might allow us to identify regions of this individual's haplotype that are shared by other individuals with schizophrenia. This focused search for sharing of a particular haplotype is simpler and more powerful than a search for any shared haplotype because it reduces multiple testing. This work is currently in progress.

It has been suggested that genes influencing susceptibility to schizophrenia and other psychiatric disorders may have pleiotropic effects (Collier and Sham 1997). In addition to the standard definitions of schizophrenia used in genetic studies (LC I and LC II), we also considered other psychiatric disorders found in relatives of schizophrenic individuals in the sample used in our study (LC III). The latter classification results in small increases in evidence for linkage for the chromosome 1 and 9 loci. This observation is compatible with the hypothesis that schizophrenia-susceptibility genes might lead to other psychiatric disorders in the relatives of patients with schizophrenia and that the clinical manifestation of these genes depends on interactions with the environment, other genes, and/or stochastic factors.

In summary, in the current study, we obtained evidence for schizophrenia loci on three chromosomes: 1, 9, and 13. Genealogical evidence shows that the affected individuals in our sample shared a relatively recent ancestor and are likely to share schizophrenia-susceptibility genes. If any schizophrenia-susceptibility alleles were inherited through the shared line of descent we have identified in our sample (Karayiorgou et al. 2004), they are likely to be surrounded by a relatively large haplotype, which will facilitate fine-mapping efforts. Nevertheless, it is possible that different lines of descent are important and that identifying the susceptibility alleles we are seeking will require extensive fine mapping with hundreds of markers. It is fortunate that these approaches are rapidly becoming feasible (Cardon and Abecasis 2003). Although we are cautious about overinterpreting the significance of these loci, at a minimum, our findings warrant further studies using our data set and independent data sets from the Afrikaner population. It is most important that, given the advantage of accurate genealogical tracing in this population, future studies will attempt to identify any extensive shared haplotypes by using denser 1–2-cM microsatellite maps, both at the identified loci and at the genomewide level.

Acknowledgments

The authors thank all the families that participated to the study; the clinicians Christina Sobin, Herman Pretorius, Brian Robertson, Sean Kaliski, and Stephen Lay; and the nurses Ria van Wyk, Cheryl Botha, and Glenda Jansen. We would also like to thank Yu Zhao for help with figure preparation. Genotyping services were provided by CIDR. CIDR is fully funded through a federal contract from the National Institutes of Health to The Johns Hopkins University, contract number N01-HG-65403. This research was supported by National Institutes of Health grant MH61399 and by the EJLB Foundation (support to M.K.).

Electronic-Database Information

The URLs for data presented herein are as follows:

Authors' Web site, http://www.sph.umich.edu/csg/abecasis/public/ (for results of parametric linkage analyses)
Center for Medical Genetics, Marshfield Clinic, http://research.marshfieldclinic.org/genetics/ (for the Marshfield genetic map)
CIDR Human Marker Set, http://www.cidr.jhmi.edu/markerset.html (for marker screening set)
Online Mendelian Inheritance in Man (OMIM), http://www.ncbi.nlm.nih.gov/Omim/ (for schizophrenia)

References

Abecasis GR, Cherny SS, Cardon LR (2001*a*) The impact of genotyping error on family-based analysis of quantitative traits. Eur J Hum Genet 9:130–134
Abecasis GR, Cherny SS, Cookson WO, Cardon LR (2001*b*) GRR: graphical representation of relationship errors. Bioinformatics 17:742–743
——— (2002) Merlin—rapid analysis of dense genetic maps using sparse gene flow trees. Nat Genet 30:97–101
American Psychiatric Association (1994) DSM-IV: diagnostic and statistical manual of mental disorders, 4th ed. American Psychiatric Association, Washington, DC
Bailer U, Leisch F, Meszaros K, Lenzinger E, Willinger U, Strobl R, Gebhardt C, Gerhard E, Fuchs K, Sieghart W, Kasper S, Hornik K, Aschauer HN (2000) Genome scan for susceptibility loci for schizophrenia. Neuropsychobiology 42:175–182
Blouin JL, Dombroski BA, Nath SK, Lasseter VK, Wolyniec PS, Nestadt G, Thornquist M, et al (1998) Schizophrenia susceptibility loci on chromosomes 13q32 and 8p21. Nat Genet 20:70–73
Brink PA, Ferreira A, Moolman JC, Weymar HW, van der Merwe PL, Corfield VA (1995) Gene for progressive familial heart block type I maps to chromosome 19q13. Circulation 91:1633–1640
Brink PA, Steyn LT, Coetzee GA, Van der Westhuyzen DR (1987) Familial hypercholesterolemia in South African Afrikaners. PvuII and StuI DNA polymorphisms in the LDL-receptor gene consistent with a predominating founder gene effect. Hum Genet 77:32–35
Brzustowicz LM, Hodgkinson KA, Chow EW, Honer WG,

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Bassett AS (2000) Location of a major susceptibility locus for familial schizophrenia on chromosome 1q21-q22. Science 288:678–682

Camp NJ, Neuhausen SL, Tiobech J, Polloi A, Coon H, Myles-Worsley M (2001) Genomewide multipoint linkage analysis of seven extended Palauan pedigrees with schizophrenia, by a Markov-chain Monte Carlo method. Am J Hum Genet 69:1278–1289

Cardno AG, Gottesman II (2000) Twin studies of schizophrenia: from bow-and-arrow concordances to star wars Mx and functional genomics. Am J Med Genet 97:12–17

Cardon LR, Abecasis GR (2003) Using haplotype blocks to map human complex trait loci. Trends Genet 19:135–140

Chen H, Young R, Mu X, Nandi K, Miao S, Prouty L, Ursin S, Gonzalez J, Yanamandra K (1999) Uniparental isodisomy resulting from 46,XX,i(1p),i(1q) in a woman with short stature, ptosis, micro/retrognathia, myopathy, deafness, and sterility. Am J Med Genet 82:215–218

Chumakov I, Blumenfeld M, Guerassimenko O, Cavarec L, Palicio M, Abderrahim H, Bougueleret L, et al (2002) Genetic and physiological data implicating the new human gene G72 and the gene for D-amino acid oxidase in schizophrenia. Proc Nat Acad Sci USA 99:13675–13680

Ciulla TA, Sklar RM, Hauser SL (1988) A simple method for DNA purification from peripheral blood. Analyt Biochem 174:485–488

Cloninger CR (1989) Schizophrenia: genetic etiological factors. In: Kaplan HI, Saddock BJ (eds) Comprehensive textbook of psychiatry, 5th ed. William & Wilkins, Baltimore, pp 732–744

Collier DA, Sham PC (1997) Catch me if you can: are catechol- and indoleamine genes pleiotropic QTLs for common mental disorders? Mol Psychiatry 2:181–183

Cookson WO, Ubhi B, Lawrence R, Abecasis GR, Walley AJ, Cox HE, Coleman R, Leaves NI, Trembath RC, Moffatt MF, Harper JI (2001) Genetic linkage of childhood atopic dermatitis to psoriasis susceptibility loci. Nat Genet 27:372–373

Coon, H, Holik J, Hoff M, Reimherr F, Wender P, Myles-Worsley M, Waldo M, Freedman R, Byerley W (1994) Analysis of chromosome 22 markers in nine schizophrenia pedigrees. Am J Med Genet 54:72–79

de la Chapelle A, Wright FA (1998) Linkage disequilibrium mapping in isolated populations: the example of Finland revisited. Proc Natl Acad Sci USA 95:12416–12423

DeLisi LE, Mesen A, Rodriguez C, Bertheau A, LaPrade B, Llach M, Riondet S, Razi K, Relja M, Byerley W, Sherrington R (2002*a*) Genome-wide scan for linkage to schizophrenia in a Spanish-origin cohort from Costa Rica. Am J Med Genet 114:497–508

DeLisi LE, Shaw SH, Crow TJ, Shields G, Smith AB, Larach VW, Wellman N, Loftus J, Nanthakumar B, Razi K, Stewart J, Comazzi M, Vita A, Heffner T, Sherrington R (2002*b*) A genome-wide scan for linkage to chromosomal regions in 382 sibling pairs with schizophrenia or schizoaffective disorder. Am J Psychiatry 159:803–812

Devlin B, Bacanu SA, Roeder K, Reimherr F, Wender P, Galke B, Novasad D, Chu A, TCuenco K, Tiobek S, Otto C, Byerley W (2002) Genome-wide multipoint linkage analyses of multiplex schizophrenia pedigrees from the oceanic nation of Palau. Mol Psychiatry 7:689–694

Dufourcq-Lagelouse R, Lambert N, Duval M, Viot G, Vilmer E, Fischer A, Prieur M, de Saint Basile G (1999) Chediak-Higashi syndrome associated with maternal uniparental isodisomy of chromosome 1. Eur J Hum Genet 7:633–637

Dyer TD, Blangero J, Williams JT, Goring HH, Mahaney MC (2001) The effect of pedigree complexity on quantitative trait linkage analysis. Genet Epidemiol 21 Suppl 1:S236–S243

Ekelund J, Lichtermann D, Hovatta I, Ellonen P, Suvisaari J, Terwilliger JD, Juvonen H, Varilo T, Arajarvi R, Kokko-Sahin ML, Lonnqvist J, Peltonen L (2000) Genome-wide scan for schizophrenia in the Finnish population: evidence for a locus on chromosome 7q22. Hum Mol Genet 9:1049–1057

Engel E (1980) A new genetic concept: uniparental disomy and its potential effect, isodisomy. Am J Med Genet 6:137–143

Epstein MP, Duren WL, Boehnke M (2000) Improved inference of relationship for pairs of individuals. Am J Hum Genet 67:1219–1231

Faraone SV, Matise T, Svrakic D, Pepple J, Malaspina D, Suarez B, Hampe C, Zambuto CT, Schmitt K, Meyer J, Markel P, Lee H, Harkavy Friedman J, Kaufmann C, Cloninger CR, Tsuang MT (1998) Genome scan of European-American schizophrenia pedigrees: results of the NIMH Genetics Initiative and Millennium Consortium. Am J Med Genet 81:290–295

Field LL, Tobias R, Robinson WP, Paisey R, Bain S (1998) Maternal uniparental disomy of chromosome 1 with no apparent phenotypic effects. Am J Hum Genet 63:1216–1220

Gelb BD, Willner JP, Dunn TM, Kardon NB, Verloes A, Poncin J, Desnick RJ (1998) Paternal uniparental disomy for chromosome 1 revealed by molecular analysis of a patient with pycnodysostosis. Am J Hum Genet 62:848–854

Genin E, Clerget-Darpoux F (1996) Consanguinity and the sib-pair method, an approach using identity by descent between and within individuals. Am J Hum Genet 59:1149–1162

Goldman A, Krause A, Ramsay M, Jenkins T (1996) Founder effect and prevalence of myotonic dystrophy in South Africans: molecular studies. Am J Hum Genet 59:445–452

Göring HH, Terwilliger JD (2000) Linkage analysis in the presence of errors IV: joint pseudomarker analysis of linkage and/or linkage disequilibrium on a mixture of pedigrees and singletons when the mode of inheritance cannot be accurately specified. Am J Hum Genet 66:1310–1327

Gottesman II, Shields J (1982) Schizophrenia: the epigenetic puzzle. Cambridge University Press, Cambridge

Groenewald JZ, Liebenberg J, Groenewald IM, Warnich L (1998) Linkage disequilibrium analysis in a recently founded population: evaluation of the variegate porphyria founder in South African Afrikaners. Am J Hum Genet 62:1254–1258

Gudbjartsson DF, Jonasson K, Frigge ML, Kong A (2000) Allegro, a new computer program for multipoint linkage analysis. Nat Genet 25:12–13

Gurling HM, Kalsi G, Brynjolfson J, Sigmundsson T, Sherrington R, Mankoo BS, Read T, Murphy P, Blaveri E, McQuillin A, Petursson H, Curtis D (2001) Genomewide

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

genetic linkage analysis confirms the presence of susceptibility loci for schizophrenia, on chromosomes 1q32.2, 5q33.2, and 8p21-22 and provides support for linkage to schizophrenia, on chromosomes 11q23.3-24 and 20q12.1-11.23. Am J Hum Genet 68:661–673

Hastbacka J, de la Chapelle A, Mahtani MM, Clines G, Reeve-Daly MP, Daly M, Hamilton BA, Kusumi K, Trivedi B, Weaver A, Coloma A, Lovett M, Buckler A, Kaitila I, Lander ES (1994) The diastrophic dysplasia gene encodes a novel sulfate transporter: positional cloning by fine-structure linkage disequilibrium mapping. Cell 78:1073–1087

Hattori E, Liu C, Badner JA, Bonner TI, Christian SL, Maheshwari M, Detera-Wadleigh SD, Gibbs RA, Gershon ES (2003) Polymorphisms at the G72/G30 gene locus, on 13q33, are associated with bipolar disorder in two independent pedigree series. Am J Hum Genet 72:1131–1140

Hayden MR, Hopkins HC, Macrea M, Beighton PH (1980) The origin of Huntington's chorea in the Afrikaner population of South Africa. S Afr Med J 58:197–200

Hovatta I, Varilo T, Suvisaari J, Terwilliger JD, Ollikainen V, Arajärvi R, Juvonen H, Kokko-Sahin ML, Väisänen L, Mannila H, Lönnqvist J, Peltonen L (1999) A genomewide screen for schizophrenia genes in an isolated Finnish subpopulation, suggesting multiple susceptibility loci. Am J Hum Genet 65:1114–1124

Kalaydjieva L, Hallmayer J, Chandler D, Savov A, Nikolova A, Angelicheva D, King RH, Ishpekova B, Honeyman K, Calafell F, Shmarov A, Petrova J, Turnev I, Hristova A, Moskov M, Stancheva S, Petkova I, Bittles AH, Georgieva V, Middleton L, Thomas PK (1996) Gene mapping in Gypsies identifies a novel demyelinating neuropathy on chromosome 8q24. Nat Genet 14:214–217

Karayiorgou M, Torrington M, Abecasis GR, Pretorius H, Robertson B, Kaliski S, Lay S, Sobin C, Moller N, Lundy LS, Blundell ML, Gogos JA, Roos JL (2004) Phenotypic characterization and genealogical tracing in an Afrikaner schizophrenia database. Am J Med Genet 124B:20–28

Kendler KS, O'Neill FA, Burke J, Murphy B, Duke F, Straub RE, Shinkwin R, Ni Nuallain M, MacLean CJ, Walsh D (1996) Irish study on high-density schizophrenia families: field methods and power to detect linkage. Am J Med Genet 67:179–190

Kong A, Cox NJ (1997) Allele-sharing models: LOD scores and accurate linkage tests. Am J Hum Genet 61:1179–1188

Kruglyak L, Daly MJ (1998) Linkage thresholds for two-stage genome scans. Am J Hum Genet 62:994–997

Ledbetter DH, Engel E (1995) Uniparental disomy in humans: development of an imprinting map and its implications for prenatal diagnosis. Hum Mol Genet 4:1757–1764

Leitersdorf E, Van der Westhuyzen DR, Coetzee GA, Hobbs HH (1989) Two common low density lipoprotein receptor gene mutations cause familial hypercholesterolemia in Afrikaners. J Clin Invest 84:954–961

Levinson DF, Mahtani MM, Nancarrow DJ, Brown DM, Kruglyak L, Kirby A, Hayward NK, Crowe RR, Andreasen NC, Black DW, Silverman JM, Endicott J, Sharpe L, Mohs RC, Siever LJ, Walters MK, Lennon DP, Jones HL, Nertney DA, Daly MJ, Gladis M, Mowry BJ (1998) Genome scan of schizophrenia. Am J Psychiatry 155:741–750

Lindholm E, Ekholm B, Shaw S, Jalonen P, Johansson G, Pettersson U, Sherrington R, Adolfsson R, Jazin E (2001) A schizophrenia-susceptibility locus at 6q25, in one of the world's largest reported pedigrees. Am J Hum Genet 69:96–105

Maziade M, Fournier A, Phaneuf D, Cliche D, Fournier JP, Roy MA, Merette C (2002) Chromosome 1q12-q22 linkage results in eastern Quebec families affected by schizophrenia. Am J Med Genet 114:51–55

Miura Y, Hiura M, Torigoe K, Numata O, Kuwahara A, Matsunaga M, Hasegawa S, Boku N, Ino H, Mardy S, Endo F, Matsuda I, Indo Y (2000) Complete paternal uniparental isodisomy for chromosome 1 revealed by mutation analyses of the TRKA (NTRK1) gene encoding a receptor tyrosine kinase for nerve growth factor in a patient with congenital insensitivity to pain with anhidrosis. Hum Genet 107:205–209

Moises HW, Yang L, Kristbjarnarson H, Wiese C, Byerley W, Macciardi F, Arolt V, et al (1995) An international two-stage genome-wide search for schizophrenia susceptibility genes. Nat Genet 11:321–324

Nurnberger JI Jr, Blehar MC, Kaufmann CA, York-Cooler C, Simpson SG, Harkavy-Friedman J, Severe JB, Malaspina D, Reich T (1994) Diagnostic interview for genetic studies. Rationale, unique features, and training. NIMH Genetics Initiative. Arch Gen Psychiatry 51:849–859

Nystuen A, Benke PJ, Merren J, Stone EM, Sheffield VC (1996) A cerebellar ataxia locus identified by DNA pooling to search for linkage disequilibrium in an isolated population from the Cayman Islands. Hum Mol Genet 5:525–531

Ober C, Cox NJ (1998) The genetics of asthma: mapping genes for complex traits in founder populations. Clin Exp Allergy 28 Suppl 1:101–105

Paunio T, Ekelund J, Varilo T, Parker A, Hovatta I, Turunen JA, Rinard K, Foti A, Terwilliger JD, Juvonen H, Suvisaari J, Arajarvi R, Suokas J, Partonen T, Lonnqvist J, Meyer J, Peltonen L (2001) Genome-wide scan in a nationwide study sample of schizophrenia families in Finland reveals susceptibility loci on chromosomes 2q and 5q. Hum Mol Genet 10:3037–3048

Peltonen L, Palotie A, Lange K (2000) Use of population isolates for mapping complex traits. Nat Rev Genet 1:182–190

Pronk JC, Gibson RA, Savoia A, Wijker M, Morgan NV, Melchionda S, Ford D, Temtamy S, Ortega JJ, Jansen S, Havenga C, Cohn RJ, de Ravel TJ, Roberts I, Westerveld A, Easton DF, Joenje H, Mathew CG, Arwert F (1995) Localisation of the Fanconi anaemia complementation group A gene to chromosome 16q24.3. Nat Genet 11:338–340

Pulkkinen L, Bullrich F, Czarnecki P, Weiss L, Uitto J (1997) Maternal uniparental disomy of chromosome 1 with reduction to homozygosity of the LAMB3 locus in a patient with Herlitz junctional epidermolysis bullosa. Am J Hum Genet 61:611–619

Rees MI, Fenton I, Williams NM, Holmans P, Norton N, Cardno A, Asherson P, Spurlock G, Roberts E, Parfitt E, Mant R, Vallada H, Dawson E, Li MW, Collier DA, Powell JF, Nanko S, Gill M, McGuffin P, Owen MJ (1999) Autosome search for schizophrenia susceptibility genes in multiply affected families. Mol Psychiatry 4:353–359

Roby P, Eyre S, Worthington J, Ramesar R, Cilliers H, Beighton P, Grant M, Wallis G (1999) Autosomal dominant (Beukes)

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

premature degenerative osteoarthropathy of the hip joint maps to an 11-cM region on chromosome 4q35. Am J Hum Genet 64:904–908

Rosendorff J, Bernstein R, Macdougall L, Jenkins T (1987) Fanconi anemia: another disease of unusually high prevalence in the Afrikaans population of South Africa. Am J Med Genet 27:793–797

Sengul H, Weeks DE, Feingold E (2001) A survey of affected-sibship statistics for nonparametric linkage analysis. Am J Hum Genet 69:179–190

Stefansson H, Sigurdsson E, Steinthorsdottir V, Bjornsdottir S, Sigmundsson T, Ghosh S, Brynjolfsson J, et al (2002) Neuregulin 1 and susceptibility to schizophrenia. Am J Hum Genet 71:877–892

Straub RE, MacLean CJ, Ma Y, Webb BT, Myakishev MV, Harris-Kerr C, Wormley B, Sadek H, Kadambi B, O'Neill FA, Walsh D, Kendler KS (2002) Genome-wide scans of three independent sets of 90 Irish multiplex schizophrenia families and follow-up of selected regions in all families provides evidence for multiple susceptibility genes. Mol Psychiatry 7:542–559

Straub RE, MacLean CJ, O'Neill FA, Burke J, Murphy B, Duke F, Shinkwin R, Webb BT, Zhang J, Walsh D, Kendler KS (1995) A potential vulnerability locus for schizophrenia on chromosome 6p24-22: evidence for genetic heterogeneity. Nat Genet 11:287–293

Sullivan PF, Neale BM, Neale MC, van den Oord E, Kendler KS (2003) Multipoint and single point non-parametric linkage analysis with imperfect data. Am J Med Genet 121B: 89–94

Takizawa Y, Pulkkinen L, Chao SC, Nakajima H, Nakano Y, Shimizu H, Uitto J (2000) Mutation report: complete paternal uniparental isodisomy of chromosome 1: a novel mechanism for Herlitz junctional epidermolysis bullosa. J Invest Dermatol 115:307–311

Thompson DA, McHenry CL, Li Y, Richards JE, Othman MI, Schwinger E, Vollrath D, Jacobson SG, Gal A (2002) Retinal dystrophy due to paternal isodisomy for chromosome 1 or chromosome 2, with homoallelism for mutations in RPE65 or MERTK, respectively. Am J Hum Genet 70:224–229

Torrington M, Viljoen DL (1991) Founder effect in 20 Afrikaner kindreds with pseudoxanthoma elasticum. S Afr Med J 79:7–11

Tsuang MT, Faraone SV (1994) The genetic epidemiology of schizophrenia. Compr Ther 20:130–135

Visapaa I, Fellman V, Vesa J, Dasvarma A, Hutton JL, Kumar V, Payne GS, Makarow M, Van Coster R, Taylor RW, Turnbull DM, Suomalainen A, Peltonen L (2002) GRACILE syndrome, a lethal metabolic disorder with iron overload, is caused by a point mutation in BCS1L. Am J Hum Genet 71:863–876

Warnich L, Kotze MJ, Groenewald IM, Groenewald JZ, van Brakel MG, van Heerden CJ, de Villiers JN, van de Ven WJ, Schoenmakers EF, Taketani S, Retief AE (1996) Identification of three mutations and associated haplotypes in the protoporphyrinogen oxidase gene in South African families with variegate porphyria. Hum Mol Genet 5:981–984

Whittemore AS, Halpern J (1994) A class of tests for linkage using affected pedigree members. Biometrics 50:118–127

Wijsman EM, Amos CI (1997) Genetic analysis of simulated oligogenic traits in nuclear and extended pedigrees: summary of GAW10 contributions. Genet Epidemiol 14:719–735

Wijsman EM, Rosenthal EA, Hall D, Blundell ML, Sobin C, Heath SC, Williams R, Brownstein MJ, Gogos JA, Karayiorgou M (2003) Genome-wide scan in a large complex pedigree with predominantly male schizophrenics from the island of Kosrae: evidence for linkage to chromosome 2q. Mol Psychiatry 8:695–705

Williams NM, Rees MI, Holmans P, Norton N, Cardno AG, Jones LA, Murphy KC, Sanders RD, McCarthy G, Gray MY, Fenton I, McGuffin P, Owen MJ (1999) A two-stage genome scan for schizophrenia susceptibility genes in 196 affected sibling pairs. Hum Mol Genet 8:1729–1739

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Original Paper

Human Heredity

Hum Hered 2006;62:175–189
DOI: 10.1159/000096599

Received: May 3, 2006
Accepted after revision: August 7, 2006
Published online: October 30, 2006

Linkage Disequilibrium and Haplotype Homozygosity in Population Samples Genotyped at a High Marker Density

Hui Wang[a] Chia-Ho Lin[a] Susan Service[b] Yuguo Chen[d] Nelson Freimer[b]
Chiara Sabatti[a, c] The international collaborative group on isolated populations

[a]Department of Statistics, [b]Center for Neurobehavioral Genetics, and [c]Department of Human Genetics, UCLA, Los Angeles, Calif., and [d]Department of Statistics, University of Illinois at Urbana-Champain, Champain, Ill., USA

Key Words
Linkage disequilibrium measures · Inbreeding · Copy number variation · Population genetics · Genomic loss

Abstract
Objective: Analyze the information contained in homozygous haplotypes detected with high density genotyping. ***Methods:*** We analyze the genotypes of ~2,500 markers on chr 22 in 12 population samples, each including 200 individuals. We develop a measure of disequilibrium based on haplotype homozygosity and an algorithm to identify genomic segments characterized by non-random homozygosity (NRH), taking into account allele frequencies, missing data, genotyping error, and linkage disequilibrium. ***Results:*** We show how our measure of linkage disequilibrium based on homozygosity leads to results comparable to those of R^2, as well as the importance of correcting for small sample variation when evaluating D'. We observe that the regions that harbor NRH segments tend to be consistent across populations, are gene rich, and are characterized by lower recombination. ***Conclusions:*** It is crucial to take into account LD patterns when interpreting long stretches of homozygous markers.

Copyright © 2006 S. Karger AG, Basel

The international collaborative group on isolated populations members that are not listed separately as authors of this manuscript are: Maria Karayiorgou[a], J. Louw Roos[b], Herman Pretorious[b], Gabriel Bedoya[c], Jorge Ospina[d], Andres Ruiz-Linares[c, e], António Macedo[f], Joana Almeida Palha[g], Peter Heutink[h, i], Yurii Aulchenko[j], Ben Oostra[j], Cornelia van Duijn[j], Marjo-Riitta Jarvelin[k, l], Teppo Varilo[m, n], Lynette Peddle[o], Proton Rahman[p], Giovanna Piras[q], Maria Monne[q], Leena Peltonen[m, n], and the affiliations:
[a]Rockefeller University, New York, N.Y, USA; [b]University of Pretoria Weskoppies Hospital, Pretoria, Republic of South Africa; [c]Laboratorio de Genetica Molecular, Universidad de Antioquia, and [d]Departamento de Psiquiatria, Universidad de Antioquia, Medellin, Colombia; [e]The Galton Laboratory, Department of Biology (Wolfson House), University College London, London, UK; [f]Instituto de Psicologia Médica, Faculdade de Medicina, Coimbra, and [g]Life and Health Sciences Research Institute (ICVS), School of Health Sciences, University of Minho, Braga, Portugal; [h]Department of Human Genetics, Section of Medical Genomics, VU University Medical Center, and [i]Center for Neurogenomics and Cognitive Research, VU University and VU University Medical Center, Amsterdam and [j]Genetic Epidemiology Unit, Departments of Epidemiology, Biostatistics and Clinical Genetics, Erasmus Medical Center, Rotterdam, The Netherlands; [k]Department of Public Health Science and General Practice, University of Oulu, Oulu, Finland; [l]Department of Epidemiology and Public Health, Imperial College London, London, UK; [m]Department of Medical Genetics, University of Helsinki, and [n]Department of Molecular Medicine, National Public Health Institute, Biomedicum Helsinki, Helsinki, Finland; [o]Newfound Genomics, and [p]Memorial University of Newfoundland, Newfoundland, Canada; [q]Division of Haematology, San Francesco Hospital, Nuoro, Italy.

KARGER
Fax +41 61 306 12 34
E-Mail karger@karger.ch
www.karger.com

© 2006 S. Karger AG, Basel
0001–5652/06/0624–0175$23.50/0

Accessible online at:
www.karger.com/hhe

Chiara Sabatti
Department of Human Genetics, UCLA School of Medicine
695 Charles E. Young Drive South
Los Angeles, CA 90095-7088 (USA)
Tel. +1 310 794 9567, Fax +1 310 794 5446, E-Mail csabatti@mednet.ucla.edu

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Identification of excess homozygosity in genotype data may provide valuable information for a wide range of population genetic investigations and also may indicate genomic variations that are associated with diseases. The level of homozygosity at a single marker reflects the age of the polymorphism, population structure and history, and possibly the effect of selection. Excess joint homozygosity of contiguous markers (haplotype homozygosity) can be a signature of selective pressure, linkage disequilibrium, inbreeding, and variations in copy numbers.

Until recently it was not feasible to assess fine-scale homozygosity in the human genome in a systematic manner or on a large scale. It is now evident that high resolution genotype data provide the information needed for such assessment. For example, several datasets of high density genotypes have shown that long stretches of homozygosity are unexpectedly common, and possibly reflect the effects of either inbreeding or selection. Estimates of homozygosity in genotype datasets can also be used for a variety of statistical analyses, for example, as a measure of linkage disequilibrium. One of the beauties of homozygosity-based measures is that, often, they do not require phasing of the data. This is a particularly appealing feature for analyzing large genotype datasets that are now available (see for example references [1–3] for a review of undergoing studies).

To explore several questions involved in identifying and interpreting homozygosity patterns in large genotype datasets, we analyzed a dataset consisting of about 2,500 SNPs on human chromosome 22 in 12 different population samples, each with about 200 individuals [4]. In particular, we were interested in exploring the relationship between homozygosity and linkage disequilibrium (LD) (see, for a review, our previous work in [5]). With this goal in mind, we conducted a two-fold analysis.

On one front, we define a new measure, *Hvol*, which is constructed by normalizing the levels of homozygosity using a volume test approach [6, 7] and is robust to small sample variation. For comparison purposes, we also consider a volume-test version of D'. We find out that measures based on homozygosity are successful in capturing the pattern of LD detected by D' and, especially, by R^2. Additionally, we are able to verify that volume measures avoid the inflation due to small sample sizes that characterizes D'.

On another front, we propose an algorithm to identify genomic regions that, on the basis of their homozygosity levels, are likely candidates for IBD, genomic loss, and/or selection. Our approach is based on a model that leaves a certain degree of ambiguity on the origins of homozygosity (in particular, IBD or genomic loss can be equivalent interpretations). This approach allows us to identify segments of 'non random homozygosity' (NRH) whose genetic origin will be investigated on the basis of data other than genotypes.

The paper is organized as follows. In section 1, we briefly describe the dataset with particular reference to the levels of homozygosity. In section 2, we introduce the measure *Hvol* and describe the results of its evaluation on this dataset. Section 3 introduces a model for the identification of NRH segments, and discusses the possible interpretations for NRH segments together with the results of its application to our dataset.

1 The Genotyped Samples

In this analysis, we study eleven isolated populations and one European-derived (CAU) population. The eleven isolated populations are Antioquia, Colombia (ANT), Ashkenazi (ASH), Azores (AZO), Costa Rican Central Valley (CR), Southwestern Netherlands (ERF), Finland – mainly early settlement – (FIP), Finland mixture of early and late settlement (FIC), late settlement Finland, Kuusamo (FIK), Newfoundland (NFL), Afrikaner (SAF), and Sardinia, province of Nuoro (SAR). The Caucasian outbred sample consists of 60 parents from CEPH trios and 140 Caucasians from the Coriell Institute Human Variation Panel (CIHVP). The sample consists of 200 individuals from each population. Samples were collected through ongoing genetic studies of various common disorders; subjects included in this analysis were either controls or parents of probands. In most populations, genealogy of the subjects was assessed, and it was documented that they were unrelated for at least several generations.

A total of 2,486 SNPs were successfully genotyped, covering 34.2 Mb of chromosome 22, with an average (median) spacing of one marker every 13.8 (8.5) kb. Specifically, 78% of the gaps between markers were less than 20 kb, and only 3.5% of gaps were greater than 50 kb. These markers were in Hardy-Weinberg equilibrium in all populations (after correcting for multiple testing) and were not monomorphic in any population. The twelve populations had similar numbers of markers with minor allele frequency (MAF) $\leq$10%, with percentages ranging from 10% in Antioquia to 14% in the Sardinian isolate. A detailed description of the genotype data, comparisons in

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

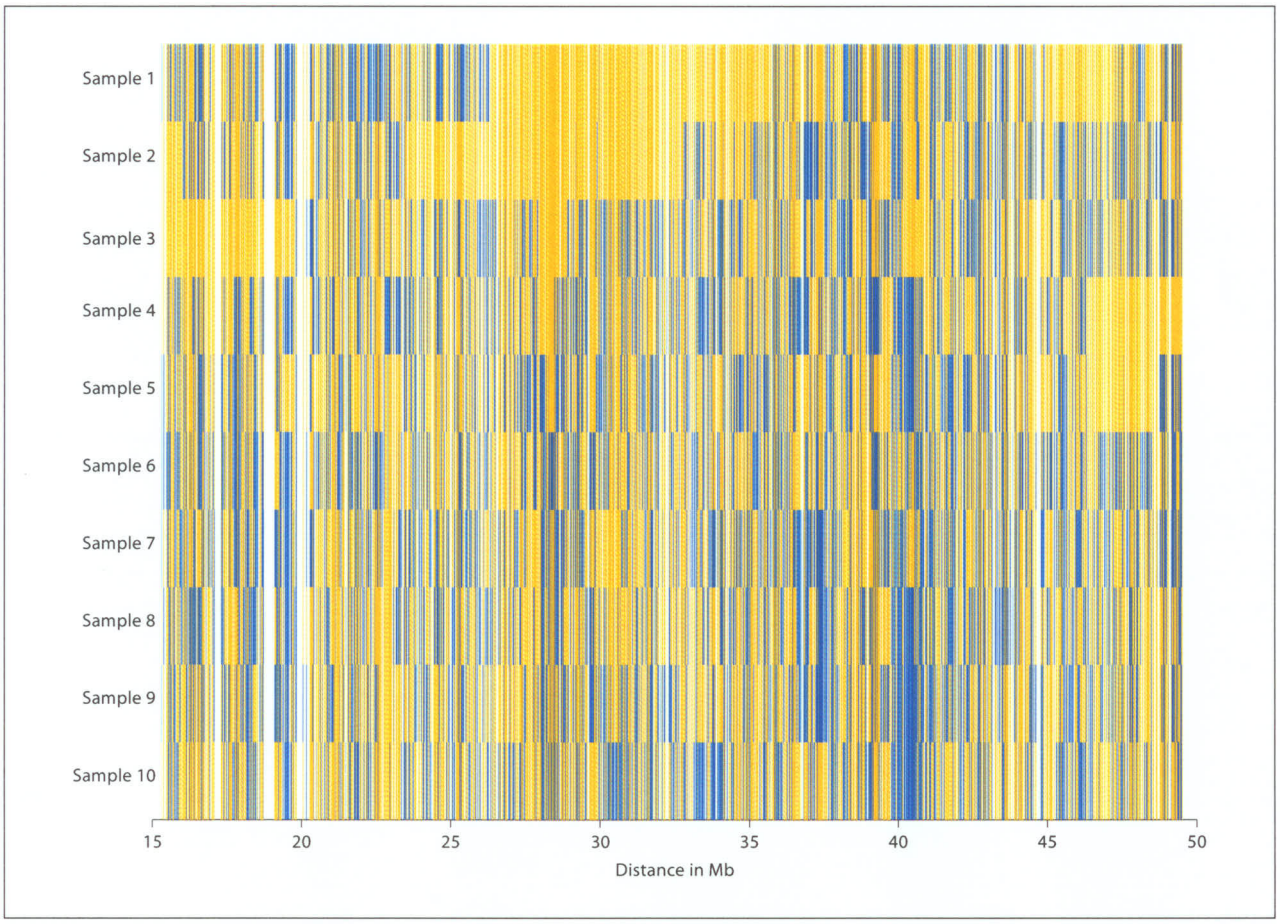


Fig. 1. Representation of genotypes in 10 individuals from Costa Rica. Each horizontal line corresponds to an individual, and each vertical line represents a SNP – whose position in Mb is given on the x axis. The white spaces represent the gaps between SNPs. An orange segment is used to denote a homozygous genotype, i.e. either AA or aa; blue and cyan correspond, respectively, to heterozygous and missing genotypes. Samples 1 to 3 have unusually long homozygous stretches; sample 4 and 5 have suspiciously long homozygous stretches; and samples 6 to 10 are representatives of the overall population. Note that in these samples, the number of ungenotyped SNPs is so small, that practically no cyan can be detected.

allele frequencies between populations, and evaluation of linkage disequilibrium patterns using LD maps can be found in Service et al. [4].

Analysis of overall heterozygosity did not reveal any interesting pattern. The mean heterozygosity of the markers was similar in different populations and ranged from 0.359 (Sardinia) to 0.373 (Antioquia). A moving average of homozygosity values across the chromosomal region also appeared fluctuating randomly around the mean value. Once, however, we started considering joint homozygosity of near-by markers, we were able to observe some interesting features. Firstly, we noticed that when one inspects the homozygosity status of markers along the chromosome in one individual, homozygous markers tend to cluster (see fig. 1). To some degree, this is to be expected because of linkage disequilibrium, as described by Sabatti and Risch [5]. Furthermore, we noticed that in a number of individuals we were able to observe rather long stretches of homozygous markers (see fig. 1). Such long haplotype homozygosity can be due to long range disequilibrium, can be the signature of selection [8], or indicate genomic regions that are identical by descent [9], or again be due to an imprecise genotype call that assigns the status of homozygous to a portion of the genome that

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

are really monozygous due to genomic loss [10]. We were naturally interested in the reason behind the homozygosity observed in our sample and decided to investigate it further. The next two sections will illustrate first the analysis based on pairwise disequilibrium measures and then the study of stretches of homozygous markers, including a discussion of their origins.

2 Measuring Linkage Disequilibrium

The connection between haplotype homozygosity and linkage disequilibrium has long been noted [11], and has been described in detail by Sabatti and Risch [5]. Let us consider the case of two markers. The alleles at each of these markers can be used to define a partition of the haplotypes. An element of the partition induced by one marker is defined by all the haplotypes that have the same allele at that marker. The level of haplotype homozygosity can be used to define a measure of agreement between the partitions defined by markers one and two, with excess homozygosity (heterozygosity) corresponding to the case of more (less) agreement than expected by chance. Different levels of agreement among these partitions translate into different prediction power, given the allele at one marker, of the allele at the other marker in the haplotype. When agreement is high, we have good prediction power, while low agreement makes prediction more challenging. Excess of haplotype homozygosity, then, corresponds to cases where the R^2 between the two SNPs is high. The relation between haplotype homozygosity values and D' is, instead, more complex, as D' does not really track 'predictability' of one SNP given the other.

While Sabatti and Risch [5] introduced measures of LD based on haplotype homozygosity with values in the range [–1,1], their applicability was hindered by the lack of explicit formulas for the required maximum and minimum values of haplotype homozygosity given the allele frequencies. A solution to this impasse is provided by the use of 'volume measures' of linkage disequilibrium. Sabatti [6] discusses these measures and their precedents in statistical literature [12, 13]. Volume measures can be defined directly on the table of observed haplotype counts, which results in their robustness to the effects of small sample sizes [7]. For simplicity, we consider only the case of two biallelic markers (which is the relevant one here). Let A and B be two SNPs with alleles A_1, A_2 and B_1, B_2, and let n be the number of observed haplotypes. We can summarize these in the table $\mathcal{N}$:

$$\mathcal{N} = \begin{array}{ccccc} & & B_1 & B_2 & \\ & A_1 & n_{11} & n_{12} & n_{1\cdot} \\ & A_2 & n_{21} & n_{22} & n_{2\cdot} \\ & & n_{\cdot 1} & n_{\cdot 2} & n \end{array}$$

where $n_{i\cdot}$ ($n_{\cdot j}$) indicated the sum of the elements in row (column) i (j). To define a volume measure of disequilibrium, one firstly needs to select a dissimilarity function, with which compare any table $\mathcal{N}$ against E, the one expected under equilibrium, with entries $e_{ij} = n_{i\cdot}n_{\cdot j}/n$. We consider two such dissimilarities. One is defined only for biallelic markers and is the same evaluated in D'; the other is excess of homozygosity. Volume measures are, then, calculated considering the number of tables T, with the same margins $n_{i\cdot}$, $n_{\cdot j}$ as $\mathcal{N}$, and lower dissimilarity with E.

Specifically, let Ω_1 denote the set of all contingency tables with the same row and column sums as the observed table. Looking at the excess of homozygosity H

$$H(\mathcal{N}) = \sum_{i,j} n_{ij}^2 - \sum_i n_{i\cdot}^2 \sum_j n_{\cdot j}^2 / n^2,$$

we can define the measure *Hvol*:

$$Hvol = \operatorname{sign}\left(H(\mathcal{N})\right) \frac{\sum_{\mathcal{N}' \in \Omega_1} 1_{\{|H(\mathcal{N}')| < |H(\mathcal{N})|\}} 1_{\{H(\mathcal{N})H(\mathcal{N}') \geq 0\}}}{\sum_{\mathcal{N}' \in \Omega_1} 1_{\{H(\mathcal{N})H(\mathcal{N}') \geq 0\}}}. \quad (1)$$

When $c = r = 2$, let Ω_2 denote the set of all contingency tables with the same row and column sums as the observed table and the same sign of $(n_{11} - n_{1\cdot}n_{\cdot 1}/n)$ as in the observed table. Let

$$M(\mathcal{N}) = \sum_{i,j} \frac{\left(n_{ij} - n_{i\cdot}n_{\cdot j}/n\right)^2}{n_{i\cdot}n_{\cdot j}/n}.$$

We then define *Dvol* as

$$Dvol = \frac{1}{|\Omega_2|} \sum_{\mathcal{N}' \in \Omega_2} 1_{\{M(\mathcal{N}') < M(\mathcal{N})\}}. \quad (2)$$

Note that the above definitions use the strict inequality sign. The choice of $<$ over $\leq$ is irrelevant for large n, but it makes a difference in the case of small n, where strict inequality allows us to better discriminate against apparent associations due to small samples.

Since LD measures based on homozygosity are probably not very familiar to the reader, we illustrate the relation between known measures such as D', $|R|$, and *Hvol*, by calculating them on a subset of our data. We considered the haplotypes defined by consecutive SNPs in one of our population of interest and evaluated the three mea-

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

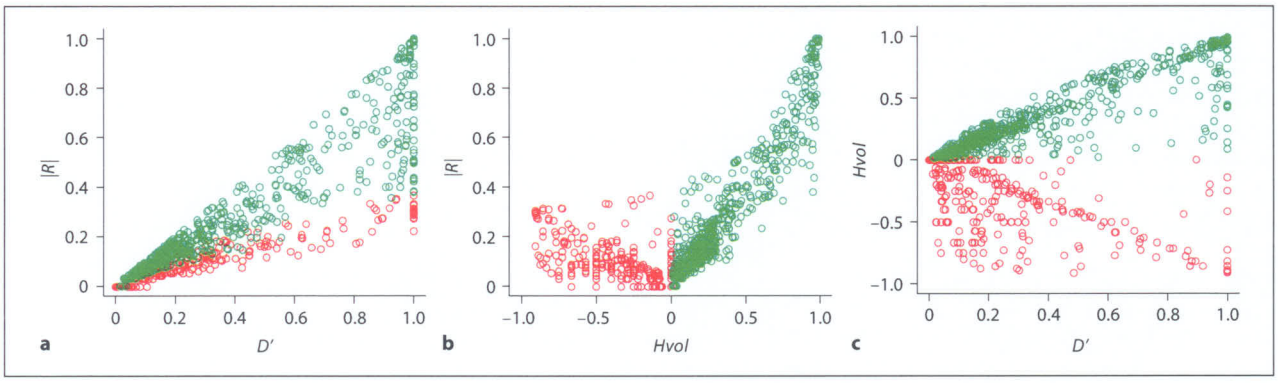


Fig. 2. Illustration of relationships among D', $|R|$, and *Hvol*. Panels **a–c** respectively show the relationships between D' and $|R|$, *Hvol* and $|R|$, D' and *Hvol*. The points are 1,000 random samples from marker pairs with a distance less than 500 kb. The red circles represent tables with excess heterozygosity, i.e. $Hvol \leq 0$, and green circles represent tables with excess homozygosity, i.e. $Hvol > 0$.

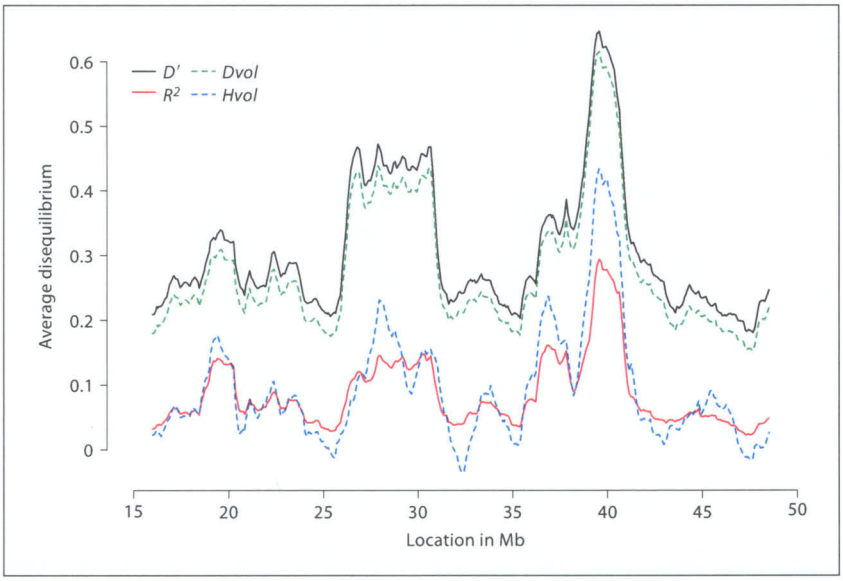


Fig. 3. Linkage disequilibrium on chromosome 22 in the Costa Rican population according to four different measures. D', R^2, *Dvol* and *Hvol* are represented, respectively, with a solid black, a solid red, a broken red, and a broken blue line. The average value of the measures, between markers that are within a 1.7 Mb window, is plotted against the middle point of the window, with the x axis representing the length of chromosome 22.

sures on the sample haplotype frequencies. Figure 2 presents the results. In this display we use $|R|$ rather than R^2, as its values are more directly comparable with D'. In the data analysis that follows, instead, we will use R^2, as it has a more direct interpretation in terms of prediction power. In panel (a) the relation between $|R|$ and D' is illustrated: there are a number of tables with high D' that have low $|R|$. In this display red circles correspond to tables with excess heterozygosity and green circles to tables with excess homozygosity: clearly excess heterozygosity correlates with low values of $|R|$. The positive correlation between $|R|$ and *Hvol* is explicitly illustrated in panel (b): high values of $|R|$ translate in high values of *Hvol*; negative values of *Hvol* correspond to low values of $|R|$; and there are a number of cases where *Hvol* is equal or close to 1, while $|R|$ does not have a very high value. These empirical relations between the values of *Hvol* and $|R|$ mirror the similarity and differences between the notions of prediction and agreement. Panel (c) indicates that the information in *Hvol* is largely orthogonal to the one contained in

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

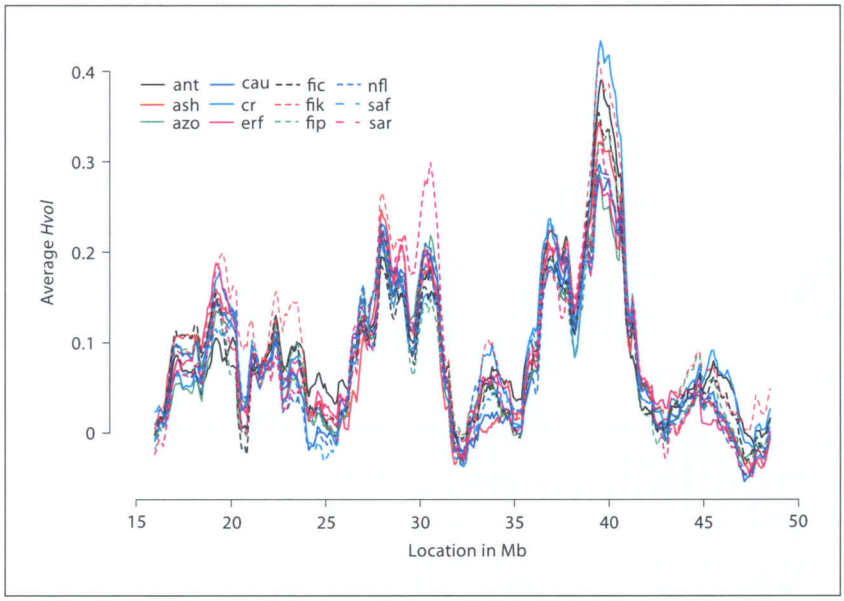


Fig. 4. Linkage disequilibrium on chromosome 22 in the 12 populations as measured with *Hvol*. The figure displays *Hvol* for all 12 populations. The average value of the measure, between markers that are within a 1.7 Mb window, is plotted against the middle point of the window, with the x axis representing the length of chromosome 22.

D'. Overall, the fact that *Hvol* is quite close to $|R|$ suggests that one could use it as a substitute for $|R|$ with multiallelic markers.

To conduct a complete analysis of the linkage disequilibrium patterns in the 12 population samples, we restricted our attention to the SNPs with sample minor allele frequencies larger than 0.1. We did so for uniformity with previous studies, such as the one conducted by Hinds [2], and to make sure that our results were not strongly influenced by the rare markers with exceptionally high homozygosity. This leads us to work with 1920 SNPs. Four measures, D', *Dvol*, R^2 and *Hvol* are calculated for each of the 1,842,240 pairs of SNPs. The results were summarized by averaging the measured disequilibrium within windows of 1.7 Mb sliding along chromosome 22.

Figure 3 reports the values of the four measures in the Costa Rican population. The observed relation between the measures is consistent across populations. In particular it can be noted that the average values of *Dvol* are lower than the ones of D', while clearly exhibiting very similar patterns. This testifies that even if the sample size is moderately large (200 individuals) and only markers with MAF >0.1 are considered, D' is inflated due to unobserved rare haplotypes. Turning now our attention to *Hvol*, one can note that its values are closer to the ones of R^2 than to any other measure – not surprisingly, given that both measures are related to predictability. *Hvol* appears to have a higher dynamic range, with larger fluctuations than R^2: this can be explained considering that *Hvol* takes negative values when the agreement is less than expected under independence, while R^2 has its minimum value at zero, in correspondence with independence.

Figure 4 presents the pattern of *Hvol* across all the 12 populations. The most striking aspect is the consistency, across populations, of the LD patterns, suggesting a fundamental role of recombination frequencies. Regions where disequilibrium is higher, allow to better identify differences across populations. In terms of mean values, FIK exhibits the highest level of disequilibrium and AZO the lowest. All of the observations above are consistent with the patterns detected for D' and R^2, described in Service et al. [4].

3 Long Stretches of Homozygous Genotypes

As mentioned in the introduction, stretches of adjacent homozygous markers can be signatures of a number of interesting genetic phenomena. In the previous section we have already underscored how linkage disequilibrium between markers in a region can significantly increase their joint homozygosity. An analysis of extended haplotype homozygosities in different populations can be found in Sabatti and Risch [5]. In addition to this, it is

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

important to keep in mind that selective sweeps substantially reduce diversity in the neighborhood of the selected gene, and thus increase the frequency and length of homozygous segments [8]. Furthermore, the genome of inbred individuals contains regions that are identical by descent and that result in stretches of homozygous genotypes [9, 14]. It is important to recover the boundaries of IBD segments, for example, for gene mapping purposes. Yet another phenomenon that gives rises to stretches of homozygous genotypes is genomic loss, and, more generally, copy number variation. Genotyping technology is unable to accurately determine the number of copies of a given genomic region, so that a haploid segment will be classified as a homozygous diploid [15], and segments that have high copy numbers are also likely to be scored as homozygous as the signal coming from the multiple copies on one chromosome overpowers the one from the normal chromosome. Recently a number of studies [16, 17] have underscored the higher than expected prevalence of variation in copy number in the genome. Their identification is important to correctly interpret results highlighting linkage and/or association between a phenotype of interest and genomic regions.

As all of these different genetic processes result in homozygous segments, it is difficult to distinguish them through the analysis of genotype data alone. Fortunately other data are usually available. Information on raw values recorded during the genotyping reaction can help identify copy number variations. A comparison of the location of homozygous segments across individuals can help separate subject specific effects (likely due to inbreeding) from population ones (that may reflect variation in recombination rate and/or selection). In order to engage in this more complex analysis, however, one needs to initially identify homozygous stretches that are warranted the researcher's attention, as adjacent markers may happen to be jointly homozygous by no more exciting reasons than random chance. Indeed, a quick look at figure 1 makes quite evident that, within each individual, homozygous and heterozygous markers tend to cluster. In order to identify homozygous segments that are reasonable candidates for regions that are identical by descent or that vary in copy number, for example, one needs a model that guides us in the definition of what is to be considered usual or not, as well as an algorithm that allows efficient scanning of the entire collection of genotypes. The rest of this section is devoted to the description of such model and the summary of the results of its application to our dataset.

Interestingly – and reflecting the ambivalence that we have already outlined – there is one model that can be used both to describe IBD status [9] and genomic loss status [15] along the chromosome of one individual. This, combined with a formalization of the genotype generating mechanism, provides a useful tool to analyze our data and identify interesting homozygous segments. Consider data from one individual and let m be the total number of markers on a chromosome (different chromosomes can be considered as independent realizations of the same process). It is useful to introduce a set of hidden variables $\Pi = \{\pi_i\}_{i=1}^{m}$ that indicate the IBD (or loss) status of the individual at the positions corresponding to each of the m markers on a chromosome: $\pi_i = 0$ indicates no IBD (no genomic alterations) at the location of the i-th marker and $\pi_i = 1$ indicates IBD (or an alteration). Leutenegger et al. [9] described the joint distribution of $\{\pi_i\}_{i=1}^{m}$ for an inbred individual using a Markov model of order one with transition probabilities that depend on the distance d_i in centimorgans (cM) between marker i and $i + 1$. In particular, we have the following transition matrix:

$$\begin{pmatrix} t(\pi_{i+1}=1|\pi_i=1) & t(\pi_{i+1}=0|\pi_i=1) \\ t(\pi_{i+1}=1|\pi_i=0) & t(\pi_{i+1}=0|\pi_i=0) \end{pmatrix} = \begin{pmatrix} 1-(1-\delta)(1-e^{-\eta d_i}) & (1-\delta)(1-e^{-\eta d_i}) \\ \delta(1-e^{-\eta d_i}) & 1-\delta(1-e^{-\eta d_i}) \end{pmatrix}, \quad (3)$$

where the parameter δ corresponds to inbreeding coefficient and η is related to the number of meiotic steps separating the two chromosomes from their most recent common ancestor. Newton and colleagues [10, 18] used the same model to represent the genomic instability of cancer cells, with δ indicating the sporadic loss rate, and the parameter η modelling the dependency among the π_is.

In order to link the unobserved variables Π with the observed genotypes $X = \{x_i\}_{i=1}^{m}$ both Leutenegger et al. [9] and Wang et al. [15] used a hidden Markov model framework. The genotype x_i has four possible values: *AA*, *Aa*, *aa*, –, with – indicating missing data. In the HMM, conditional on π_i, the probability of the four possible values of x_i is independent of the genotype values at the other markers and of the values of π_j with $i \neq j$. If $\pi_i = 1$, we expect a homozygous genotype (modulo genotyping error), while if $\pi_i = 0$, all the genotypes are observed according to their population frequencies. According to the HMM terminology, then, we have the following emission probabilities:

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

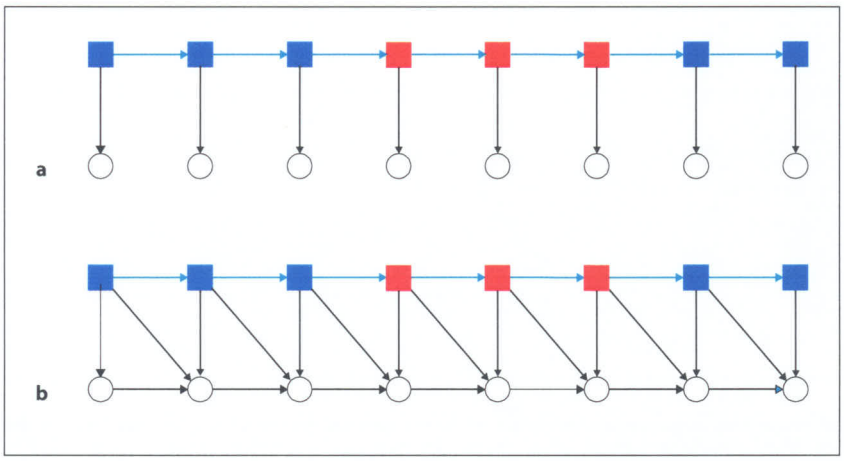


Fig. 5. Hidden Markov Models. Each filled square represents the hidden state π_i at SNP i, red squares correspond to $\pi_i = 1$, blue squares to $\pi_i = 0$, and each circle represents the genotype x_i. The arrows denote the dependence relationships of the HMM. Panel (**a**) gives an illustration of the standard HMM. The next hidden state only depends on the previous hidden state, and the genotype at each SNP only depends on the hidden state of that SNP. Panel (**b**) is an illustration of our generalized HMM. The hidden state still preserves the property of a Markov chain, while the genotype of a SNP depends not only on its hidden state, but also on the hidden state and the genotype of its previous SNP.

$$\begin{pmatrix} e(x_i = AA \mid \pi_i = 1) & e(x_i = Aa \mid \pi_i = 1) & e(x_i = aa \mid \pi_i = 1) & e(x_i = - \mid \pi_i = 1) \\ e(x_i = AA \mid \pi_i = 0) & e(x_i = Aa \mid \pi_i = 0) & e(x_i = aa \mid \pi_i = 0) & e(x_i = - \mid \pi_i = 0) \end{pmatrix}$$
$$= \begin{pmatrix} p_{A^i}(1-\kappa) & 0 & (1-p_{A^i})(1-\kappa) & \kappa \\ p_{A^i}^2(1-\kappa) & 2p_{A^i}(1-p_{A^i})(1-\kappa) & (1-p_{A^i})^2(1-\kappa) & \kappa \end{pmatrix} \quad (4)$$

where p_{A^i} is the frequency of allele A for the i-th marker and κ is the missing rate.

While the genotype model in (4) was adequate for the marker density considered in the studies of Leutenegger et al. [9] and Wang et al. [15], it is not for the high density data we collected on chromosome 22. The average inter-marker distance in our data-set is such that we expect nearby markers to be in linkage disequilibrium (and indeed we observed this to be the case in the analysis of the previous section). Model (4) assumes, instead, markers to be in linkage equilibrium, and applying it to our data would result in an excess identification of IBD/loss regions. In order to develop an effective screening tool for homozygous segments, we then modified model (4) to take into account linkage disequilibrium. Figure 5 gives a graphical illustration of the original HMM model (a) and of our generalization (b). For conceptual and computational simplicity, we limited ourselves to a description of linkage disequilibrium that relies on a first order Markov model (see later discussion). The genotype x_i, conditionally on π_i, depends also on the genotype values of the immediate neighboring markers x_{i-1}, x_{i+1}, and actually, the specific form of this link is determined by the values of π_{i-1} and π_{i+1}. Conditionally on (π_i, π_{i-1}, π_{i+1}, x_{i-1}, x_{i+1}), the genotype x_i is independent of every other value of X and Π. We can describe the joint probability of the entire sequence X, by specifying $Pr(x_i \mid x_{i-1}, \pi_{i-1}, \pi_i)$ (which we will denote $e(x_i \mid x_{i-1}, \pi_{i-1}, x_i)$) – as illustrated by the arrows in figure 5 (b). These probabilities are defined combining two elements: the expectation of a homozygous genotype when $\pi_i = 1$, and the dependence between alleles at neighboring markers on the same chromosome according to a Markov model of order 1. To indicate the precise form of these modified emission probabilities, we need to introduce some notation. Let us indicate with $\underline{AA}$ the haplotype consisting of alleles A at marker $i-1$ (succinctly A^{i-1}) and allele A at marker i (succinctly A^i). The population haplotype frequency at the two markers can then be written out as:

$$\begin{array}{ccccc} & & A^i & a^i & \\ A^{i-1} & & p_{\underline{AA}} & p_{\underline{Aa}} & p_{A^{i-1}} \\ a^{i-1} & & p_{\underline{aA}} & p_{\underline{aa}} & p_{a^{i-1}} \\ & & p_{A^i} & p_{a^i} & 1 \end{array} \quad (5)$$

To fix ideas, let's exclude for the moment the possibility of genotyping error and missing data, and consider the case when $\pi_{i-1} = \pi_i = 1$: the genotypes x_{i-1}, x_i are homozygous and represent the alleles at two neighboring markers on the same chromosome, so that given x_{i-1}, the value of x_i is determined by the conditional population frequency of alleles at marker i given the observed allele at marker $i-1$. We then have:

$$e(x_i = AA \mid x_{i-1} = AA, \pi_{i-1} = \pi_i = 1) = p_{\underline{AA}}/p_{A^{i-1}}$$
$$e(x_i = aa \mid x_{i-1} = AA, \pi_{i-1} = \pi_i = 1) = p_{\underline{Aa}}/p_{A^{i-1}}$$
$$e(x_i = AA \mid x_{i-1} = aa, \pi_{i-1} = \pi_i = 1) = p_{\underline{aA}}/p_{a^{i-1}}$$
$$e(x_i = aa \mid x_{i-1} = aa, \pi_{i-1} = \pi_i = 1) = p_{\underline{aa}}/p_{a^{i-1}}.$$

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

We can then introduce a probability ε of genotyping error and κ of missing data. If we assume that, in presence of genotyping error, alleles are independent and observed according to their population frequencies, one obtains the following expressions for the conditional probabilities of the genotypes x_i given $\pi_{i-1} = \pi_i = 1$ (that can be used to define the appropriate $e(x_i \mid x_{i-1}, \pi_i = \pi_{i-1} = 1)$):

$$p(x_i = AA \mid x_{i-1} = AA) =$$
$$\frac{(1-\varepsilon)^2 p_{\underline{AA}} + \varepsilon(1-\varepsilon) p_{A^{i-1}} p_{A^i} \left(p_{A^{i-1}} + p_{A^i}\right) + \varepsilon^2 p^2_{A^{i-1}} p^2_{A^i}}{(1-\varepsilon) p_{A^{i-1}} + \varepsilon p^2_{A^{i-1}}} (1-\kappa)$$
$$p(x_i = Aa \mid x_{i-1} = AA) = 2\varepsilon p_{A^i} p_{a^i} (1-\kappa)$$
$$p(x_i = aa \mid x_{i-1} = AA) =$$
$$\frac{(1-\varepsilon)^2 p_{\underline{Aa}} + \varepsilon(1-\varepsilon) p_{A^{i-1}} p_{a^i} \left(p_{A^{i-1}} + p_{a^i}\right) + \varepsilon^2 p^2_{A^{i-1}} p^2_{a^i}}{(1-\varepsilon) p_{A^{i-1}} + \varepsilon p^2_{A^{i-1}}} (1-\kappa)$$
$$p(x_i = AA \mid x_{i-1} = Aa) = \left((1-\varepsilon) p_{A^i} + \varepsilon^2 p^2_{A^i}\right)(1-\kappa)$$
$$p(x_i = Aa \mid x_{i-1} = Aa) = 2\varepsilon p_{A^i} p_{a^i} (1-\kappa)$$
$$p(x_i = aa \mid x_{i-1} = Aa) = \left((1-\varepsilon) p_{a^i} + \varepsilon^2 p^2_{a^i}\right)(1-\kappa)$$
$$p(x_i = - \mid x_{i-1}) = \kappa$$

Replacing A with a in $p(x_i \mid x_{i-1} = AA)$ will generate probablities $p(x_i \mid x_{i-1} = aa)$. Note that $p(x_i \mid x_{i-1} = Aa)$ is actually the marginal distribution of $p(x_i)$, which is a function of allele frequencies of the i-th marker. This will be the case no matter what the values of (π_{i-1}, π_i) are. When $x_{i-1} = -$, the genotype probability distribution of x_i is the marginal.

Let us now consider how the expression above would need to be modified for other values of π_{i-1}, π_i: in the case where $\pi_{i-1} = \pi_i = 0$, there are two haplotypes contributing to the genotype; when $\pi_{i-1} = 0$, $\pi_i = 1$ or $\pi_{i-1} = 1$, $\pi_i = 0$, only one chromosome contributes alleles at both markers. The probabilities $e(x_i \mid x_{i-1}, \pi_{i-1}, \pi_i)$ that can be derived taking into account these observations and the haplotype distribution (5) are given in the appendix.

Given our modified emission probabilities, our model no longer has precisely the structure of a HMM; however, it is still possible to evaluate the probability of the entire genotype sequence X using a recursive algorithm. For analogy with the HMM literature, we refer to forward and backward algorithm to indicate the two recursions that can be defined on our model.

Forward algorithm:

$$\alpha(\pi_i) = P(x_1, \ldots, x_i, \pi_i) = \sum_{\pi_{i-1}} \alpha(\pi_{i-1}) t(\pi_i \mid \pi_{i-1}) e(x_i \mid x_{i-1}, \pi_{i-1}, \pi_i) \quad (6)$$

with initial conditions:

$$\alpha(\pi_1 = 1) = P(x_1 \mid \pi_1 = 1) P(\pi_1 = 1) = e(x_1 \mid \pi_1 = 1)\delta$$
$$\alpha(\pi_1 = 0) = P(x_1 \mid \pi_1 = 0) P(\pi_1 = 1) = e(x_1 \mid \pi_1 = 0)(1-\delta)$$

Backward algorithm:

$$\beta(\pi_i) = P(x_{i+1}, \ldots, x_m \mid x_i, \pi_i) =$$
$$\sum_{\pi_{i+1}} \beta(\pi_{i+1}) t(\pi_{i+1} \mid \pi_i) e(x_{i+1} \mid x_i, \pi_i, \pi_{i+1}) \quad (7)$$

with initial conditions:

$$\beta(\pi_m = 1) = \beta(\pi_m = 0) = 1$$

With $\alpha()$ and $\beta()$ defined as above, we also obtain other expressions that are typical of HMM:

$$P(X, \pi_i, \pi_{i+1}) = \alpha(\pi_i) \beta(\pi_{i+1}) t(\pi_{i+1} \mid \pi_i) e(x_{i+1} \mid x_i, \pi_i, \pi_{i+1})$$

and

$$P(X, \pi_i) = \alpha(\pi_i) \beta(\pi_i).$$

These similarities with the HMM formulation allow us to calculate efficiently the probability of a genotype sequence X and to reconstruct with a Viterbi style algorithm the unobserved values Π. The portion in the genomes where $\pi_i = 1$ correspond to 'interesting' homozygous segments: that is, when the data is evaluated taking into account allele frequency and linkage disequilibrium (even if in a quite minimal form) these regions appear as IBD or subject to genomic losses. As we have already pointed out, it is impossible to distinguish between these two phenomena on the basis of the genotypes of one individual alone; however, this is not our goal here. We are simply aiming at defining an algorithm that processes genotypes of one individual and highlights interesting homozygous segments. These can be due to IBD status, genomic loss, selection, or even linkage disequilibrium patterns that are not well described by a Markov model of order one, but rather imply much longer range dependence. To see how the described model can be used with this purpose, we need to clarify how we specify the parameter values.

We assume that the values of κ and ε are obtained by prior genotyping studies conducted using the same technology. The collection of two-marker haplotype population frequencies (5) is obtained from the genotype data using a standard EM algorithm. Once these parameters have been fixed, we estimate – for each individual – δ and η using a maximum likelihood approach. The numerical optimization is performed using a gradient algorithm described in the appendix.

Applying the algorithm to the dataset, we are able to identify the regions of non-random homozygosity (NRH)

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Table 1. The 5-number summary of length of NRH segments (cM) and overall percentage of the chromosomes covered by NRH segments in the 12 populations

	NRH segment length (cM)				NRH chromosome proportion			
	1st Qu.	median	3rd Qu.	max.	1st Qu.	median	3rd Qu.	max.
azo	0.0000	0.1473	0.4665	23.63	0.0043	0.0129	0.0276	0.2565
fip	0.0000	0.2272	0.5752	39.36	0.0059	0.0148	0.0299	0.5522
ant	0.0563	0.2885	0.6546	24.24	0.0071	0.0162	0.0304	0.4694
ash	0.1057	0.2892	0.6324	7.937	0.0068	0.0142	0.0258	0.1086
erf	0.0327	0.3164	0.6565	20.90	0.0063	0.0153	0.0299	0.2268
nfl	0.1290	0.3164	0.6325	20.97	0.0061	0.0108	0.0222	0.2471
cau	0.1966	0.3621	0.6888	9.186	0.0064	0.0160	0.0270	0.0997
sar	0.1422	0.3661	0.8127	25.10	0.0077	0.0158	0.0319	0.3042
saf	0.1667	0.3687	0.7266	17.00	0.0062	0.0131	0.0231	0.3167
cr	0.1962	0.4219	0.7171	19.54	0.0057	0.0169	0.0326	0.2685
fic	0.2800	0.4845	0.8968	12.14	0.0095	0.0184	0.0334	0.1473
fik	0.2651	0.5168	1.0370	22.90	0.0083	0.0197	0.0431	0.2645

We do not display the minimum values of the NRH segments and overall percentages in the 12 populations in the table because they are all zeros.

in each of the individuals in the 12 populations. Table 1 illustrates some summary statistics across populations. We report the five number summaries of the NRH segments lengths, and the proportion of chromosome 22 on which they span in each individual. (A more accurate description of the distributions of these two quantities is available in the supplementary material). In terms of population comparisons, the Finnish isolate (FIK) and the Finnish cohort sample (which include individuals from isolate) (FIC) show the longest and most frequent NRH segments. This is consistent with the higher values of LD in FIK, as measured both here and in Service et al. [4]. The nature of this analysis is substantially exploratory and hence we do not claim that differences between the distributions illustrated in table 1 are statistically significant. This question can be addressed with appropriate confirmatory techniques; a first step in this direction is provided in the supplementary material.

Once the regions of NRH are reconstructed, it is of interest to try to interpret what genetic process contributed to their formation. The first hypothesis we considered is inbreeding. The Costa Rica population, for which we had careful genealogical records, offered an ideal test case. A total of 6 individuals were known to have positive – albeit small – inbreeding coefficient. Our algorithm estimated a positive δ for each of them, and a likelihood ratio test, whose significance was evaluated with simulations, rejected the hypothesis $\delta = 0$ for all cases. These results testify that our methodology can effectively detect inbreeding. However, the distribution of the estimated values of δ in CR, as well in the other populations, makes it unrealistic to attribute to inbreeding the bulk of the detected NRH segments: there are too many $\delta > 0$ and too large values of δ for these to be due only to previously undetected inbreeding. For example, figure 1 presents some individuals that are homozygous for as much as 30% of their genotypes. This distortion could be due to the fact that we are observing a small fraction of the entire genome of these individuals, but it is also important to explore other possible causes of NRH.

Chromosome 22 is known to harbor one of the most well-studied microdeletions: a deletion on 22q11.2 is known to generate susceptibility to DiGeorge [19] and velocardiofacial [20] syndromes and has been associated to schizophrenia [21]. It was then natural to investigate the possibility that some of the genotyped individuals harbored a deletion in this region. We did not find, however, any evidence of this. A few individuals from the Costa Rica showed a NRH region overlapping 22q11.2, but when we inspected the raw intensity values of the genotyping reaction, we saw no evidence of deletion. Moreover, it is worth emphasizing that the NRH segments we identify are generally quite long: shorter segments as the ones resulting from a microdeletion may not be judged as interesting by our methodology, as their length is entire-

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

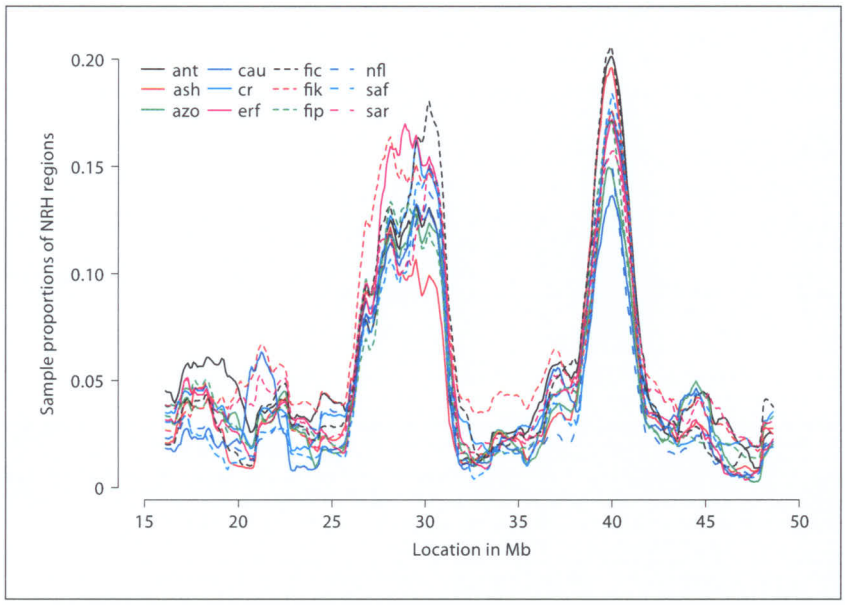


Fig. 6. Localization of NRH regions across populations. For each population, we evaluated the proportion of individuals that exhibit NRH at each SNP. The average of these proportions across a 1.7 Mb sliding window is plotted against the position of the window center, with the x axis representing the length of chromosome 22.

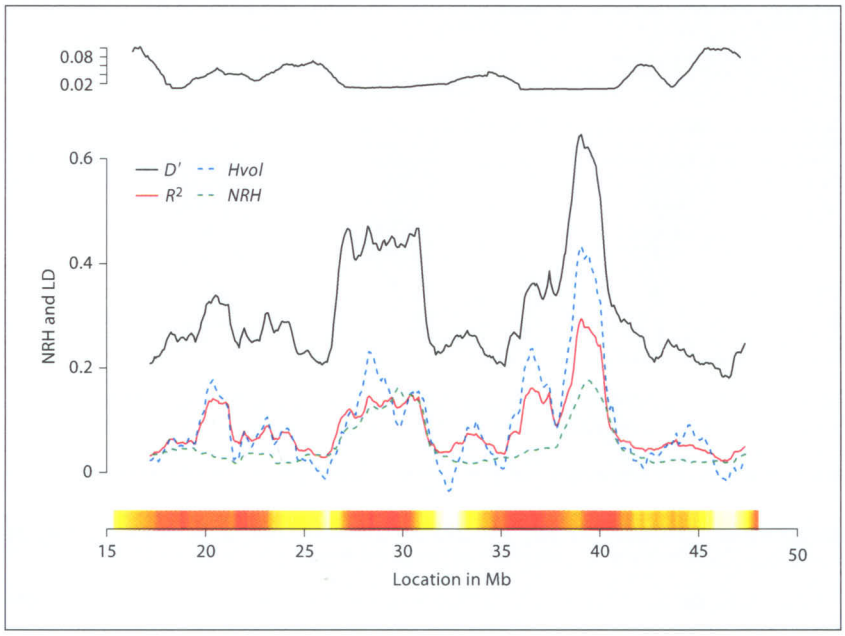


Fig. 7. Relation between NRH regions and patterns of linkage disequilibrium in Costa Rica. D', R^2, *Hvol* and NRH proportions are represented, respectively, with a solid black, a solid red, a broken blue, and a broken green line (curves are obtained as in previous pictures). The heat-colored bar at the bottom displays the known gene density on chromosome 22, as obtained averaging information from the UCSC genome browser within 1.7 Mb sliding windows. On the top of the picture, a smoothed derivative of the genetic map of Chr. 22.

ly comparable to that of random homozygosity segments that are found all over the genome.

Segment of NRH could reflect the action of selection. In such case, we would expect NRH segments to occupy roughly the same genomic location in different individuals. To explore this possibility, we tallied, within each population, the proportion of individuals whose genotypes were interpreted as NRH at each genomic position. The results, presented in figure 6 revealed a consistent pattern across populations: while there are some specific population differences, these may very well be attributed to sample variation and are much less striking than the overall consistency. There are clearly two regions on chr 22 where NRH is prevalent: between 26 and 31 Mb and 38 and 41 Mb. These are comparatively gene rich regions, with, respectively 38 and 36 genes per Mb, versus the 20

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

genes per Mb of chr 22 as a whole. This is consistent with the hypothesis of selection playing a role in determining these NRH segments. However, there is one more possible reason for the aggregation of these NRH segments that we need to explore.

Figure 7 illustrates the relation between patterns of NRH and linkage disequilibrium. For clarity we have focused on the sample from one population only – Costa Rica – but the results are consistent across populations. As it can be seen, the regions of high density of NRH correspond very well with the regions of high LD. Recall that our methods for the identification of NRH takes into account LD in as much as it can be modeled with a Markov process of order one. Effects of higher order LD are not incorporated in the model and hence could lead to the identification of NRH segments: these would be signatures of long range linkage disequilibrium. Figure 7 clearly seems to suggest this to be the case. Linkage disequilibrium can indeed be caused by the effect of selection, however, it is interesting to note how this pattern correlates well also with variation in recombination fraction across chromosome 22, as captured by the genetic map and illustrated in figure 7. This suggests that the high LD due to low recombination may suffice to explain a large portion of the NRH regions. A similar finding was recently reported by Gibson et al. [22] in an analysis of hap-map data.

4 Conclusion

Homozygosity is a very natural genetic concept. We illustrate here how it can be constructively used to measure linkage disequilibrium, as well as a first pass detection tool for other genetic phenomena.

A number of studies have recently noted the presence of long stretches of homozygous markers in highly densely genotyped individuals. We provide a model and an algorithm to identify such segments characterized by non-random homozygosity. Our analysis of 12 population samples comprising 200 individuals each genotyped at 2,500 markers on chromosome 22 suggests that the vast majority of these NRH segments is due to linkage disequilibrium, that could result from selection effects, but that is actually well correlated with variation in recombination frequency. Caution has to be used, then, when homozygous haplotypes are used to estimate inbreeding or localize disease genes, that an accurate consideration is given to the specific linkage disequilibrium levels.

Acknowledgments

Chiara Sabatti and HuiWang were partially supported by NSF grant DMS0239427 and NIH/NIDOCD grant DC04224. Chiara Sabatti also acknowledges support from ASA/Ames grant NCC2-1364, USPHS grant GM53275, NIH grants R01NS037484 and R01MH049499. Nelson Freimer is partially supported by NIH grants R01NS037484 and R01MH049499. Yuguo Chen is partially supported by NSF grant DMS0503981. We thank two anonymous reviewers for their significant input.

Appendix: The IBD Reconstruction Model and Algorithm

Model with Differential Probabilities of Missing Genotype

In the main text, in the interest of readability, we have considered only one missing rate. However, it is quite reasonable to assume that the probability of a missing genotype depends on the hidden status Π, at least when this indicates genomic alterations. To accommodate this case we have considered a model with differential missing rate κ for genotypes corresponding to $\pi_i = 0$, and τ for genotypes corresponding to $\pi_i = 1$. In the detailed description of our model and algorithm that follows we will use these two differential missing rates. As in the main text, κ is assumed as given (estimated on the base of prior data), while τ is estimated with maximum likelihood.

Modified Emission Probabilities

Below we list under different hidden states, the emission probabilities of the genotypes.

When $\pi_{i-1} = 1$, $\pi_i = 1$, we have

$$p\left(x_i = AA \,|\, x_{i-1} = AA\right) = \frac{(1-\varepsilon)^2 p_{\underline{AA}} + \varepsilon(1-\varepsilon) p_{A^{i-1}}\left(p_{A^{i-1}} + p_{A^i}\right) + \varepsilon^2 p^2_{A^{i-1}} p^2_{A^i}}{(1-\varepsilon) p_{A^{i-1}} + \varepsilon p^2_{A^{i-1}}}(1-\tau)$$

$$p\left(x_i = Aa \,|\, x_{i-1} = AA\right) = 2\varepsilon p_{A^i} p_{a^i}(1-\tau)$$

$$p\left(x_i = aa \,|\, x_{i-1} = AA\right) = \frac{(1-\varepsilon)^2 p_{\underline{Aa}} + \varepsilon(1-\varepsilon) p_{A^{i-1}} p_{a^i}\left(p_{A^{i-1}} + p_{a^i}\right) + \varepsilon^2 p^2_{A^{i-1}} p^2_{a^i}}{(1-\varepsilon) p_{A^{i-1}} + \varepsilon p^2_{A^{i-1}}}(1-\tau)$$

$$p\left(x_i = AA \,|\, x_{i-1} = Aa\right) = \left((1-\varepsilon) p_{A^i} + \varepsilon^2 p^2_{A^i}\right)(1-\tau)$$

$$p\left(x_i = Aa \,|\, x_{i-1} = Aa\right) = 2\varepsilon p_{A^i} p_{a^i}(1-\tau)$$

$$p\left(x_i = aa \,|\, x_{i-1} = Aa\right) = \left((1-\varepsilon) p_{a^i} + \varepsilon^2 p^2_{a^i}\right)(1-\tau)$$

Similarly, when $\pi_{i-1} = 1$, $\pi_i = 0$,

$$p\left(x_i = AA \,|\, x_{i-1} = AA\right) = \frac{(1-\varepsilon)^2 p_{\underline{AA}} p_{A^i} + \varepsilon(1-\varepsilon) p_{A^{i-1}} p^2_{A^i} + \varepsilon p^2_{A^{i-1}} p^2_{A^i}}{(1-\varepsilon) p_{A^{i-1}} + \varepsilon p^2_{A^{i-1}}}(1-\kappa)$$

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

$$p\left(x_i = Aa \,|\, x_{i-1} = AA\right) =$$
$$\frac{(1-\varepsilon)^2\left(p_{\underline{Aa}}p_{A^i} + p_{\underline{AA}}p_{a^i}\right) + 2\varepsilon(1-\varepsilon)p_{A^{i-1}}p_{A^i}p_{a^i} + 2\varepsilon p^2_{A^{i-1}}p_{A^i}p_{a^i}}{(1-\varepsilon)p_{A^{i-1}} + \varepsilon p^2_{A^{i-1}}}(1-\kappa)$$
$$p\left(x_i = aa \,|\, x_{i-1} = AA\right) =$$
$$\frac{(1-\varepsilon)^2 p_{\underline{Aa}}p_{a^i} + \varepsilon(1-\varepsilon)p_{A^{i-1}}p^2_{a^i} + \varepsilon p^2_{A^{i-1}}p^2_{a^i}}{(1-\varepsilon)p_{A^{i-1}} + \varepsilon p^2_{A^{i-1}}}(1-\kappa)$$
$$p\left(x_i = AA \,|\, x_{i-1} = Aa\right) = p^2_{A^i}(1-\kappa)$$
$$p\left(x_i = Aa \,|\, x_{i-1} = Aa\right) = 2p_{A^i}p_{a^i}(1-\kappa)$$
$$p\left(x_i = aa \,|\, x_{i-1} = Aa\right) = p^2_{a^i}(1-\kappa)$$
$$p\left(x_i = -|x_{i-1}\right) = \kappa$$

When $\pi_{i-1} = 0$, $\pi_i = 1$,

$$p\left(x_i = AA|x_{i-1} = AA\right) = \left((1-\varepsilon)^2\frac{p_{\underline{AA}}}{p_{A^{i-1}}} + \varepsilon(1-\varepsilon)p_{A^i} + \varepsilon p^2_{A^i}\right)(1-\tau)$$
$$p\left(x_i = Aa \,|\, x_{i-1} = AA\right) = 2\varepsilon p_{A^i}p_{a^i}(1-\tau)$$
$$p\left(x_i = aa|x_{i-1} = AA\right) = \left((1-\varepsilon)^2\frac{p_{\underline{Aa}}}{p_{A^{i-1}}} + \varepsilon(1-\varepsilon)p_{a^i} + \varepsilon p^2_{a^i}\right)(1-\tau)$$
$$p\left(x_i = AA|x_{i-1} = Aa\right) =$$
$$\left((1-\varepsilon)^2\frac{\left(p_{\underline{AA}}p_{a^{i-1}} + p_{\underline{aA}}p_{A^{i-1}}\right)}{2p_{A^{i-1}}p_{a^{i-1}}} + 2\varepsilon(1-\varepsilon)p_{A^i} + 2\varepsilon p^2_{A^i}\right)(1-\tau)$$
$$p\left(x_i = Aa \,|\, x_{i-1} = Aa\right) = 2\varepsilon p_{A^i}p_{a^i}(1-\tau)$$
$$p\left(x_i = aa \,|\, x_{i-1} = Aa\right) =$$
$$\left((1-\varepsilon)^2\frac{\left(p_{\underline{Aa}}p_{a^{i-1}} + p_{\underline{aa}}p_{A^{i-1}}\right)}{2p_{A^{i-1}}p_{a^{i-1}}} + 2\varepsilon(1-\varepsilon)p_{a^i} + 2\varepsilon p^2_{a^i}\right)(1-\tau)$$
$$p\left(x_i = -|x_{i-1}\right) = \tau$$

When $\pi_{i-1} = 0$, $\pi_i = 0$,

$$p\left(x_i = AA|x_{i-1} = AA\right) = \left((1-\varepsilon)^2\frac{p^2_{\underline{AA}}}{p^2_{A^{i-1}}} + \varepsilon(2-\varepsilon)p^2_{A^i}\right)(1-\kappa)$$
$$p\left(x_i = Aa|x_{i-1} = AA\right) = \left((1-\varepsilon)^2\frac{2p_{\underline{AA}}p_{\underline{Aa}}}{p^2_{A^{i-1}}} + 2\varepsilon(2-\varepsilon)p_{A^i}p_{a^i}\right)(1-\kappa)$$
$$p\left(x_i = aa \,|\, x_{i-1} = AA\right) = \left((1-\varepsilon)^2\frac{p^2_{\underline{Aa}}}{p^2_{A^{i-1}}} + \varepsilon(2-\varepsilon)p^2_{a^i}\right)(1-\kappa)$$
$$p\left(x_i = AA|x_{i-1} = Aa\right) = \left((1-\varepsilon)^2\frac{p_{\underline{AA}}p_{\underline{aA}}}{p_{A^{i-1}}p_{a^{i-1}}} + \varepsilon(2-\varepsilon)p^2_{A^i}\right)(1-\kappa)$$
$$p\left(x_i = Aa|x_{i-1} = Aa\right) =$$
$$\left((1-\varepsilon)^2\frac{p_{\underline{AA}}p_{\underline{aa}} + p_{\underline{Aa}}p_{\underline{aA}}}{p_{A^{i-1}}p_{a^{i-1}}} + 2\varepsilon(2-\varepsilon)p_{A^i}p_{a^i}\right)(1-\kappa)$$
$$p\left(x_i = aa|x_{i-1} = Aa\right) = \left((1-\varepsilon)^2\frac{p_{\underline{Aa}}p_{\underline{aa}}}{p_{A^{i-1}}p_{a^{i-1}}} + \varepsilon(2-\varepsilon)p^2_{a^i}\right)(1-\kappa)$$
$$p\left(x_i = -|x_{i-1}\right) = \kappa$$

MLE and Derivation of the Gradient Algorithm

The dataset we analyze consists in the genotypes of multiple individuals. Depending on the interpretation that one is willing to give to δ and η it may be appropriate to estimate a separate value of these parameters for each individual or a single value for the entire population. In particular, if inbreeding is perceived as the most likely cause of extended segments of homozygosity, it is important to estimate a subject specific value of the two parameters describing the distribution of the unknown states, as every individual will have his/her own inbreeding coefficient. If, instead, the genotypes under analysis come from cancer cell lines, so that the most likely cause of extended homozygous segments is genomic loss, it may be more appropriate to estimate only one value of the parameters describing the loss process across all individuals. The gradient algorithm we describe can be easily adapted to both cases. As specified in the main text, to analyze our data we estimated individual specific values of δ and η. Generally speaking, when one decides to use the model with differential missing rate, the value of τ should be assumed constant across individuals. We now give the details of the gradient algorithm.

When the hidden state $\Pi = (\pi_i)$ is known, the likelihood is

$$P(X) = \sum_{\Pi} P\left(X, \Pi\right) =$$
$$\sum_{\Pi} t\left(\pi_1\right)\prod_{i=1}^{m-1} t\left(\pi_{i+1}\,|\,\pi_i\right)e\left(x_1\,|\,\pi_1\right)\prod_{i=1}^{m-1} e\left(x_{i+1}\,|\,x_i, \pi_i, \pi_{i+1}\right)$$

Note that

$$\frac{\partial P\left(X, \Pi\right)}{\partial t\left(\pi_{i+1}\,|\,\pi_i\right)} = \frac{P\left(X, \Pi\right)}{t\left(\pi_{i+1}\,|\,\pi_i\right)}$$
$$\frac{\partial P\left(X, \Pi\right)}{\partial e\left(x_{i+1}\,|\,x_i, \pi_i, \pi_{i+1}\right)} = \frac{P\left(X, \Pi\right)}{e\left(x_{i+1}\,|\,x_i, \pi_i, \pi_{i+1}\right)}$$

The derivation of the first derivatives of the parameters are similar to the non-LD case:

$$\frac{\partial \log P(X)}{\partial \eta} = \frac{1}{P(X)}\sum_{\Pi}\frac{\partial P\left(X, \Pi\right)}{\partial \eta}$$
$$= \frac{1}{P(X)}\sum_{\Pi}\sum_{i=1}^{m-1}\frac{\partial P\left(X, \Pi\right)}{\partial t\left(\pi_{i+1}\,|\,\pi_i\right)}\frac{\partial t\left(\pi_{i+1}\,|\,\pi_i\right)}{\partial \eta}$$
$$= \frac{1}{P(X)}\sum_{i=1}^{m-1}\sum_{\Pi}\frac{P\left(X, \Pi\right)}{t\left(\pi_{i+1}\,|\,\pi_i\right)}\frac{\partial t\left(\pi_{i+1}\,|\,\pi_i\right)}{\partial \eta}$$

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

$$= \frac{1}{P(X)} \sum_{i=1}^{m-1} \sum_{\pi_i} \sum_{\pi_{i+1}} \frac{P(X, \pi_i, \pi_{i+1})}{t(\pi_{i+1}|\pi_i)} \frac{\partial t(\pi_{i+1}|\pi_i)}{\partial \eta}$$

$$= \frac{1}{P(X)} \sum_{i=1}^{m-1} \sum_{\pi_i} \sum_{\pi_{i+1}} \alpha(\pi_i)\beta(\pi_{i+1}) e(x_{i+1}|x_i, \pi_i, \pi_{i+1}) \frac{\partial t(\pi_{i+1}|\pi_i)}{\partial \eta}$$

$$\frac{\partial \log P(X)}{\partial \delta} = \frac{1}{P(X)} \sum_{\Pi} \frac{\partial P(X, \Pi)}{\partial \delta}$$

$$= \frac{1}{P(X)} \left(\sum_{\Pi} \frac{\partial P(X, \Pi)}{\partial t(\pi_1)} \frac{\partial t(\pi_1)}{\partial \delta} + \sum_{\Pi} \sum_{i=1}^{m-1} \frac{\partial P(X, \Pi)}{\partial t(\pi_{i+1}|\pi_i)} \frac{\partial t(\pi_{i+1}|\pi_i)}{\partial \delta} \right)$$

$$= \frac{1}{P(X)} \left(\sum_{\Pi} \frac{P(X, \Pi)}{t(\pi_1)} \frac{\partial t(\pi_1)}{\partial \delta} + \sum_{i=1}^{m-1} \sum_{\Pi} \frac{P(X, \Pi)}{t(\pi_{i+1}|\pi_i)} \frac{\partial t(\pi_{i+1}|\pi_i)}{\partial \delta} \right)$$

$$= \frac{1}{P(X)} \left(\sum_{\pi_1} \frac{P(X, \pi_1)}{t(\pi_1)} \frac{\partial t(\pi_1)}{\partial \delta} + \sum_{i=1}^{m-1} \sum_{\pi_i} \sum_{\pi_{i+1}} \frac{P(X, \pi_i, \pi_{i+1})}{t(\pi_{i+1}|\pi_i)} \frac{\partial t(\pi_{i+1}|\pi_i)}{\partial \delta} \right)$$

$$= \frac{1}{P(X)} \left(\sum_{\pi_1} \frac{\alpha(\pi_1)\beta(\pi_1)}{t(\pi_1)} \frac{\partial t(\pi_1)}{\partial \delta} + \sum_{i=1}^{m-1} \sum_{\pi_i} \sum_{\pi_{i+1}} \alpha(\pi_i)\beta(\pi_{i+1}) e(x_{i+1}|x_i, \pi_i, \pi_{i+1}) \frac{\partial t(\pi_{i+1}|\pi_i)}{\partial \delta} \right)$$

The gradient algorithm to maximize δ, η, τ is as following, t is the time, and λ is the step size.

$$\delta^{(t+1)} = \delta^{(t)} + \lambda_\delta \frac{\partial \log P(X)}{\partial \delta}_{|\delta=\delta^{(t)}}$$

$$\eta^{(t+1)} = \eta^{(t)} + \lambda_\eta \frac{\partial \log P(X)}{\partial \eta}_{|\eta=\eta^{(t)}}$$

$$\tau^{(t+1)} = \tau^{(t)} + \lambda_\tau \frac{\partial \log P(X)}{\partial \tau}_{|\tau=\tau^{(t)}}$$

These partial derivatives are additive, and the gradient algorithm can easily be extended to cases with multiple sequences where the δ and η are the same for the entire population. Suppose we have $j = 1, \ldots, n$ independent sequences, each with length m_j, and we denote each of them by X_j. Then the gradient algorithm becomes:

$$\frac{\partial \log P(X)}{\partial \eta} = \sum_j \frac{1}{P(X_j)} \left(\sum_{i=1}^{m_j-1} \sum_{\pi_{ji}} \sum_{\pi_{j,i+1}} \alpha(\pi_{ji})\beta(\pi_{j,i+1}) e(x_{j,i+1}|\pi_{j,i+1}, \pi_{ji}, x_{ji}) \frac{\partial t(\pi_{j,i+1}|\pi_{ji})}{\partial \eta} \right)$$

$$\frac{\partial \log P(X)}{\partial \delta} = \sum_j \frac{1}{P(X_j)} \left(\sum_{\pi_{j1}} \alpha(\pi_{j1})\beta(\pi_{j1}) \frac{\partial t(\pi_{j1})}{\partial \delta} + \sum_{i=1}^{m_j-1} \sum_{\pi_{ji}} \sum_{\pi_{j,i+1}} \alpha(\pi_{ji})\beta(\pi_{j,i+1}) e(x_{j,i+1}|x_{ji}, \pi_{ji}, \pi_{j,i+1}) \frac{\partial t(\pi_{j,i+1}|\pi_i)}{\partial \delta} \right)$$

$$\frac{\partial \log P(X)}{\partial \tau} = \sum_j \frac{1}{P(X_j)} \left(\sum_{\pi_{j1}} \alpha(\pi_{j1})\beta(\pi_{j1}) \frac{\partial e(x_{j1}|\pi_{j1})}{\partial \tau} + \sum_{i=1}^{m_j-1} \sum_{\pi_{ji}} \sum_{\pi_{j,i+1}} \alpha(\pi_{ji})\beta(\pi_{j,i+1}) t(\pi_{j,i+1}|\pi_{ji}) \frac{\partial e(x_{j,i+1}|x_{ji}, \pi_{ji}, \pi_{j,i+1})}{\partial \tau} \right)$$

References

1 The International HapMap Consortium: A haplotype map of the human genome. Nature 2005;437:1299–1320.

2 Hinds DA, Stuve LL, Nilsen GB, Halperin E, Eskin E, Ballinger DG, Frazer KA, Cox DR: Whole-genome patterns of common DNA variation in three human populations. Science 2005;307:1072–1079.

3 Thomas DC, Haile RW, Duggan D: Recent developments in genomewide association scans: a workshop summary and review. Am J Hum Genet 2005;77:337–345.

4 Service S, DeYoung J, Karayiorgou M, Louw Roos J, Pretorious H, Bedoya G, Ospina J, Ruiz-Linares A, Macedo A, Palha J, Heutink P, Aulchenko Y, Oostra B, van Duijn C, Jarvelin M, Varilo T, Peddle L, Rahman P, Piras G, Monne M, Murray S, Galver L, Peltonen L, Sabatti C, Collins A, Freimer N: Magnitude and distribution of linkage disequilibrium in population isolates and implications for genome-wide association studies. Nature Genetics 2006;38:556–560.

5 Sabatti C, Risch N: Homozygosity and linkage disequilibrium. Genetics 2002;160:1707–1719.

6 Sabatti C: Measuring dependence with volume tests. The American Statistician 2002;50:191–195.

7 Chen Y, Lin C, Sabatti C: Volume measures for linkage disequilibrium. Submitted, 2006.

8 Sabeti PC, Reich DE, Higgins JM, Levine HZ, Richter DJ, Schaffner SF, Gabriel SB, Platko JV, Patterson NJ, McDonald GJ, Ackerman HC, Campbell SJ, Altshuler D, Cooper R, Kwiatkowski D, Ward R, Lander ES: Detecting recent positive selection in the human genome from haplotype structure. Nature 2002;419:832–837.

9 Leutenegger A, Prum B, Genin E, Verny C, Lemainque A, Clerget-Darpoux F, Thompson E: Estimation of the inbreeding coefficient through use of genomic data. Am J Hum Genet 2003;73:516–523.

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

10 Newton M, Gould M, Reznikoff C, Haag J: On the statistical analysis of allelic-loss data. Stat Med 2000;17:1425–1445.
11 Ohta T: Linkage disequilibrium between amino acid sites in immunoglobulin genes and other multigene families. Genet Res 1980;36:181–197.
12 Hotelling H: Tubes and spheres in n-spaces, and a class of statistical problems. American Journal of Mathematics 1939;61:440–460.
13 Diaconis P, Efron B: Testing for independence in a two-way table: new interpretations of the Chi-square statistics. The Annals of Statistics 1985;13:845–874.
14 Woods CG, Cox J, Springell K, Hampshire DJ, Mohamed MD, McKibbin M, Stern R, Raymond FL, Sandford R, Sharif SM, Karbani G, Ahmed M, Bond J, Clayton D, Inglehearn CF: Quantification of homozygosity in consanguineous individuals with autosomal recessive disease. Am J Hum Genet 2006;78:889–896.
15 Wang H, Lee Y, Nelson S, Sabatti C: Inferring genomic loss and location of tumor suppressor genes from high density genotypes. Journal of the French Statistical Society 2005; 146:153–171.
16 Iafrate A, Feuk L, Rivera M, Listewnik M, Donahoe P, Qi Y, Scherer SW, Lee C: Detection of large-scale variation in the human genome. Nat Genet 2004;36:949–951.
17 Sebat J, Lakshmi B, Troge J, Alexander J, Young J, Lundin P, Maner S, Massa H, Walker M, Chi M, Navin N, Lucito R, Healy J, Hicks J, Ye K, Reiner A, Gilliam T, Trask B, Patterson N, Zetterberg A, Wigler M: Large-scale copy number polymorphism in the human genome. Science 2004;305:525–528.
18 Newton M, Lee Y: Inferring the location and effect of tumor suppressor genes by instabilityselection modeling of allelic-loss data. Biometrics 2000;56:1088–1097.
19 de la Chapelle A, Herva R, Koivisto M, Aula P: A deletion in chromosome 22 can cause DiGeorge syndrome. Hum Genet 1981;57: 253–256.
20 Scambler PJ, Kelly D, Lindsay E, Williamson R, Goldberg R, Shprintzen R, Wilson DI, Goodship JA, Cross IE, Burn J: Velo-cardio-facial syndrome associated with chromosome 22 deletions encompassing the DiGeorge locus. Lancet 1992;339:1138–1139.
21 Karayiorgou M, Morris MA, Morrow B, Shprintzen RJ, Goldberg R, Borrow J, Gos A, Nestadt G, Wolyniec PS, Lasseter VK, Eisen H, Childs B, Kazazian HH, Kucherlapati R, Antonarakis SE, Pulver AE, Housman DE: Schizophrenia susceptibility associated with interstitial deletions of chromosome 22q11. Proc Nat Acad Sci 1995;92:7612–7616.
22 Gibson J, Morton N, Collins A: Extended tracts of homozygosity in outbred human populations. Human Molecular Genetics 2006;15:789–795.

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

nature genetics

© 2006 Nature Publishing Group http://www.nature.com/naturegenetics

Magnitude and distribution of linkage disequilibrium in population isolates and implications for genome-wide association studies

Susan Service[1], Joseph DeYoung[2], Maria Karayiorgou[3], J Louw Roos[4], Herman Pretorious[4], Gabriel Bedoya[5], Jorge Ospina[6], Andres Ruiz-Linares[5,7], António Macedo[8], Joana Almeida Palha[9], Peter Heutink[10,11], Yurii Aulchenko[12], Ben Oostra[12], Cornelia van Duijn[12], Marjo-Riitta Jarvelin[13,14], Teppo Varilo[15,16], Lynette Peddle[17], Proton Rahman[18], Giovanna Piras[19], Maria Monne[19], Sarah Murray[20], Luana Galver[20], Leena Peltonen[15,16], Chiara Sabatti[21,22], Andrew Collins[23] & Nelson Freimer[1,24,25]

The genome-wide distribution of linkage disequilibrium (LD) determines the strategy for selecting markers for association studies, but it varies between populations. We assayed LD in large samples (200 individuals) from each of 11 well-described population isolates and an outbred European-derived sample, using SNP markers spaced across chromosome 22. Most isolates show substantially higher levels of LD than the outbred sample and many fewer regions of very low LD (termed 'holes'). Young isolates known to have had relatively few founders show particularly extensive LD with very few holes; these populations offer substantial advantages for genome-wide association mapping.

There remains a great deal of uncertainty about how to design and carry out genome-wide association studies of common diseases[1]. It is not yet feasible to directly assay the entire genome for association between diseases and candidate variants. Current genome-wide studies must therefore be indirect, identifying associations by assaying SNPs in LD with untyped variants that are causal for a given trait. The extent and distribution of LD determines the number of SNPs required for such studies. Variations in demographic histories between populations generate substantial differences in the extent and distribution of LD[2] and therefore cause variability between populations in the number of SNPs required for association studies. It has been suggested that population isolates, particularly those founded recently, have longer stretches of LD than outbred populations[3–5]. Hence, in comparison with outbred populations, such isolates may require fewer markers for genome-wide association studies or may achieve better genome-wide coverage with equivalent numbers of markers. However, these possible advantages of isolates have not yet been quantified.

We evaluated LD in 11 populations that have been proposed as isolates, together with an outbred European-derived sample. The demographic histories of these isolates are well documented (**Supplementary Note** and **Supplementary Table 1** online). Additionally, the histories of these isolates differ according to factors hypothesized to influence LD; for example, some populations experienced admixture at the time of founding, whereas others differed markedly in the size of the founding population or in the time since their founding. Notably, we used a large sample size in this study (200 persons in each population); small samples generally overestimate the strength of LD, because they may not include rare, recombinant haplotypes[6,7].

We analyzed LD in these populations across the entire length of chromosome 22. We chose a marker density that previous smaller-scale surveys of LD suggested would adequately assay LD in the

[1]Center for Neurobehavioral Genetics and [2]Southern California Genotyping Consortium, University of California, Los Angeles, California 90095, USA. [3]Rockefeller University, New York, New York 10021, USA. [4]University of Pretoria, Department of Psychiatry & Weskoppies Hospital, Pretoria, Republic of South Africa 0001. [5]Laboratory of Molecular Genetics and [6]Department of Psychiatry, University of Antioquia, Medellin, Colombia. [7]The Galton Laboratory, Department of Biology (Wolfson House), University College London, London NW1 2HE, UK. [8]Institute of Medical Psychology, Faculty of Medicine, University of Coimbra, 3000-033 Coimbra, Portugal. [9]Life and Health Sciences Research Institute (ICVS), School of Health Sciences, University of Minho, 4704-553 Braga, Portugal. [10]Department of Human Genetics, Section of Medical Genomics and [11]Center for Neurogenomics and Cognitive Research, Vrije University and Vrije University Medical Center, 1081 BT, Amsterdam, The Netherlands. [12]Genetic Epidemiology Unit, Departments of Epidemiology & Biostatistics and Clinical Genetics, Erasmus Medical Center, 3000 DR Rotterdam, The Netherlands. [13]Department of Public Health Science and General Practice, University of Oulu, 90220 Oulu, Finland. [14]Department of Epidemiology and Public Health, Imperial College London, London W2 1PG, UK. [15]Department of Medical Genetics, University of Helsinki, 00290 Helsinki, Finland. [16]Department of Molecular Medicine, National Public Health Institute, Biomedicum Helsinki, 00290 Helsinki, Finland. [17]Newfound Genomics, 187 Lemarchant Road, St. John's, Newfoundland, Canada A1C 2H5. [18]Memorial University of Newfoundland, St. John's, Newfoundland, Canada A1C 5S7. [19]Division of Haematology, 'San Francesco' Hospital, ASL3 Nuoro, Italy. [20]Illumina, Inc., 9885 Towne Centre Dr., San Diego, California 92121, USA. [21]Department of Human Genetics and [22]Department of Statistics, University of California, Los Angeles, California 90095 USA. [23]Human Genetics Research Division, University of Southampton, Southampton General Hospital, Tremona Road, Southampton SO16 6YA, UK. [24]The Jane and Terry Semel Institute for Neuroscience and Human Behavior and [25]Department of Psychiatry, University of California, Los Angeles, California 90095 USA. Correspondence should be addressed to N.F. (nfreimer@mednet.ucla.edu).

Received 8 November 2005; accepted 28 February 2006; published online 2 April 2006; doi:10.1038/ng1770

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

© 2006 Nature Publishing Group http://www.nature.com/naturegenetics

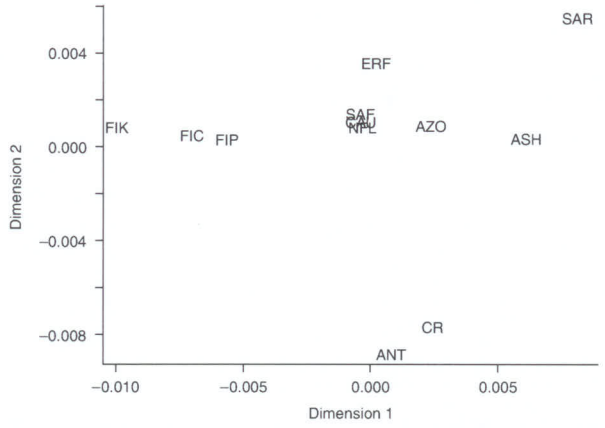


Figure 1 Result of non-metric multidimensional scaling applied to Nei's genetic distance between all pairs of populations. Population abbreviations are as in **Table 1**.

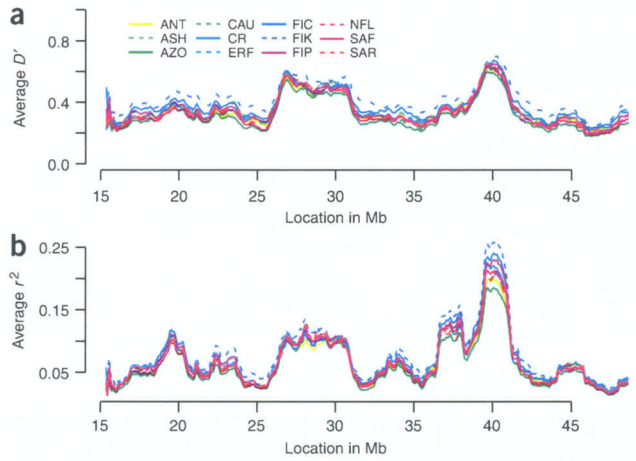


Figure 2 Distribution of linkage disequilibrium on chromosome 22. Average D' (**a**) and r^2 (**b**) coefficients plotted in sliding windows containing all markers separated by less than 500 kb in successive 1.7-Mb segments (1.6-Mb overlap). Population abbreviations are as in **Table 1**.

isolates[8–13] and that was similar to the density in currently available assays for genome-wide association studies[14]. The 2,486 SNPs that met the criteria for successful genotyping (see Methods) spanned 34.2 Mb of chromosome 22, at an average spacing of one marker every 13.8 kb (median, one marker per 8.5 kb). Seventy-eight percent of gaps between markers were under 20 kb, and only 3.5% of gaps were greater than 50 kb. These markers were in Hardy-Weinberg equilibrium in all populations (after correcting for multiple testing) and were not monomorphic in any population. Most markers (98%) had over 180 persons genotyped in each population.

The mean heterozygosity of the markers was similar in the different populations, ranging from 0.359 (Sardinia) to 0.373 (Antioquia). The 12 populations also had similar numbers of markers with minor allele frequency (MAF) $<10\%$, ranging from 10% of markers in Antioquia to 14% in Sardinia. We estimated the genetic distance between all pairs of populations with all marker data using Nei's D[15], and we applied nonmetric multidimensional scaling to this distance matrix to obtain a two-dimensional representation of interpopulation distances (**Fig. 1**). The outbred sample was most similar to the Newfoundland and Afrikaner populations and least similar to the populations of Antioquia, Kuusamo and Sardinia. The Costa Rican and Antioquian populations were as similar to each other as were the three Finnish samples.

We calculated D' and r^2 between all pairs of markers. The profile of LD for both of these measures was similar to that previously observed on chromosome 22 in European-derived samples[16]. Interpopulation differences were apparent from the moving average of D' and r^2 in sliding windows along the chromosome (**Fig. 2**), with the highest LD in the three Finnish samples and the lowest LD in the Azores sample.

We constructed LD maps[17] for all populations. In this method, distances are estimated in linkage disequilibrium units (LDUs) between adjacent pairs of SNPs. The LDU represents the product of the physical distance between markers and a parameter that reflects the exponential decline in the probability of association between markers according to physical distance. The LDU scale has additive distances, and these maps are analogous to recombination maps. The LD map is more robust than haplotype block methods to variation in marker density[18], and previous studies suggest that spacing markers evenly on the LDU map is a simple but effective strategy for determining appropriate marker density for association studies.

The profiles of the LD maps were very similar in all 12 populations examined, as has been observed previously[19,20] (**Fig. 3**). This pattern corresponds to regions of elevated and diminished recombination that are common to all populations, as evidenced by comparison with the linkage map[21]. Although the LD maps for the 12 populations had similar shapes, they differed substantially in their overall length (**Fig. 3** and **Table 1**). The shortest map (and hence, the population with the most extensive LD) was in the Finnish isolate from Kuusamo, and the longest maps were in the Azorean, Newfoundland, Afrikaner and outbred samples.

It has been previously observed that LD between markers is indeterminate when the LDU distance between them is ≥ 3 (ref. 19). Therefore, in common with previous studies using this methodology, we defined an LD 'hole' to be a chromosomal segment with a gap of ≥ 2.5 LDU between adjacent SNPs[22]. In our data set, the

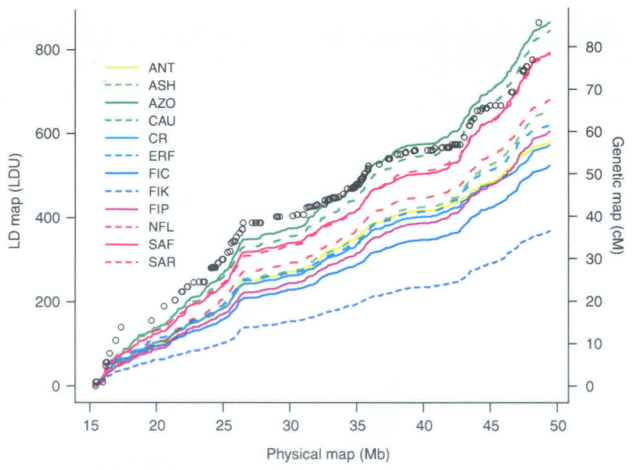


Figure 3 Quantitative comparison of the LD maps of the 12 populations and physical and genetic maps. Open circles represent the genetic map plotted against the physical map. Note the second y-axis showing the scale of the genetic map. Population abbreviations are as in **Table 1**.

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

© 2006 Nature Publishing Group http://www.nature.com/naturegenetics

Table 1 LD metric map length

Population	Length of LD map in LDU	s.e.m.	LDU/Mb ratio	Number of LD holes	Total size LD holes (kb)	Swept radius
ANT	581.9	34.2	17.01	31	1,092	164.62
ASH	656.5	26.1	19.19	26	975	155.84
AZO	864.5	26.2	25.27	84	2,709	130.51
CAU	845.1	29.6	24.70	84	2,574	129.55
CR	572.1	31.1	16.72	23	821	161.95
ERF	620.8	30.4	18.15	29	1,166	156.11
FIC	523.9	15.2	15.31	21	821	161.52
FIK	368.3	19.4	10.77	5	146	202.75
FIP	606.5	20.9	17.73	26	731	149.88
NFL	790.4	23.5	23.10	69	2,014	138.93
SAF	794.1	23.7	23.21	76	2,633	141.29
SAR	681.4	26.0	19.92	38	1,241	145.51

Third column shows the delete-d jackknife estimate of the s.e.m. of the length of the LD map. A hole is defined as a gap in the LD map of >2.5 LDU. Swept radius indicates the mean useful extent of LD for association mapping on chromosome 22 (in kb). ANT: Antioquia, Colombia; ASH: Ashkenazi; AZO: Azores; CAU: outbred European-derived sample; CR: Central Valley of Costa Rica; ERF: a village in southwestern Netherlands; FIC: early-settlement Finland; FIK: Finnish subisolate of Kuusamo; FIP: Finland nationwide; NFL: Newfoundland; SAF: Afrikaner; SAR: province of Nuoro in Sardinia.

median physical distance between markers defining a hole is larger than the median distance between non-hole markers (28.1 kb versus 8.4 kb), even though most of the larger gaps between markers do not contain LD holes (only 8.3% of intermarker gaps >50 kb contain LD holes). The correlation between the total number of LD holes in a 3-Mb window and the recombination rate for that window (downloaded from HapMap) is 0.83 (n = 21 comparison points). The number of holes is not uniformly distributed across the 12 populations (χ^2 = 197, 11 degrees of freedom (d.f.), P = 0) and is positively related to the length of the LD maps (**Table 1**, **Fig. 4**).

Figure 4 Distribution of LD holes (defined as a gap in the LD map of ≥2.5 LDU) by population. The null hypothesis of uniformity of the distribution of the total number of holes cannot be rejected for populations with the same color coding. Within each group, one cannot reject the hypothesis of a uniform distribution of LD holes (blue group: χ^2 = 6.9, 6 d.f., P = 0.33; pink group χ^2 = 2, 3 d.f., P = 0.57). Population abbreviations are as in **Table 1**.

As r^2 is the standard measure of whether the LD between two markers is sufficient for detecting associations, for each population, we calculated r^2 between pairs of markers that defined a hole. We compared this value to the r^2 value for a randomly selected pair of markers (matched to be approximately the same physical distance apart). The r^2 values are much higher between marker pairs that do not flank holes than between marker pairs that flank a hole. The results show that the holes are not simply segments with a large physical distance between markers (**Fig. 5a**). In fact, marker pairs within 100 kb of either side of an LD hole also have lower estimated r^2 than do non-hole areas of a similar size, indicating reduced power to detect association in a larger region around the hole-defining markers (**Fig. 5b**).

To assess the impact of increasing marker density of the number of LD holes, we constructed and evaluated chromosome 22 LD maps for the Centre d'Etude du Polymorphisme Humain (CEPH) samples from successive releases of HapMap data. The LD map constructed from release #19 of the HapMap data used nearly three times as many markers as the LD map constructed from release #13 of the HapMap data (27,060 SNPs versus 9,658 SNPs), yet the number of LD holes decreased only by one-half (30 LD holes versus 64 LD holes).

There are substantial differences between the populations in the overall magnitude of LD; the length of the LD map in the majority of the isolates was ~20–45% lower than in the outbred sample. The length of the LD map provides a guide to the minimum number of markers needed for association analyses, which require equal marker spacing on the LDU scale[19,20]. Considering the length of the LD maps only, our data suggest that association analyses in samples from Finland, the Dutch isolate, Costa Rica, Antioquia, Sardinia and the Ashkenazim would require at least 30% fewer markers than studies in outbred populations. The advantages of working in such isolates, which display several-fold fewer LD holes than the outbred sample, are also apparent in considering the additional markers needed to fill such holes. Marker sets currently being proposed for genome-wide LD mapping studies are spaced at similar densities to that used in the current analysis, suggesting that, in these isolates, such mapping sets may provide much better genome coverage for detecting LD than in outbred populations.

Two of the most important determinants of the extent of LD in a population are the effective population size (influenced strongly by the number of founders and expansion rate of the population) and the

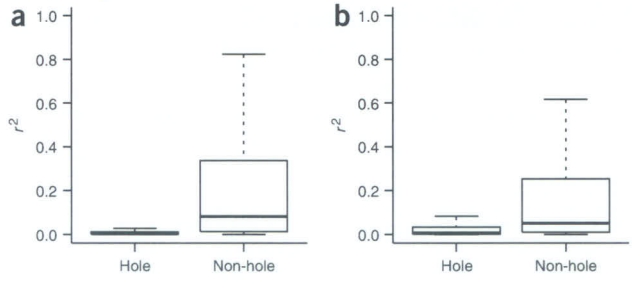


Figure 5 Box plots of r^2 values between markers surrounding LD holes and in non-hole regions. The dark line in the center of the box is the median, the edges of the box are located at the quartiles and the range of non-outlier points is indicated by the dashed lines. (**a**) r^2 between pairs of markers that define a hole and between randomly selected marker pairs a similar physical distance apart. (**b**) r^2 between pairs of markers within 100 kb of either marker that defined a hole and between pairs of markers in randomly selected regions.

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

© 2006 Nature Publishing Group http://www.nature.com/naturegenetics

time since founding. These factors affect, respectively, the number of different haplotypes present in the population and the number of opportunities for recombination between markers. In the Kuusamo isolate, the precise historical evidence documenting recent founding by a few individuals, subsequent population bottleneck and then rapid expansion is consistent with the observation that LD in this sample is much more extensive than in any of the other samples. Conversely, the relatively large number of founders of the Newfoundland and Azorean populations and low rates of expansion provide reasonable explanations for why less LD is seen in these samples. Admixture and inbreeding are two other possible explanations for the variation in length of LD map and number of LD holes among the isolates. Admixture at the time of founding is a possible source of enhanced LD in a population. Two isolates with well-documented initial admixture (Costa Rica and Antioquia[23]) display a greater extent of LD and fewer LD holes than the outbred sample. It is difficult to untangle, retrospectively, the relative contributions of admixture from other aspects of population history. Investigation of the extent and distribution of LD in Native American groups from the vicinities of Costa Rica and Antioquia may provide useful background information for the design of association studies in these two isolates. Inbreeding is likely to be higher in isolates than in outbred populations; however, in our study, marker heterozygosities were similar in the different populations, indicating that this was probably not a large factor in creating variability among populations in the extent of LD (**Supplementary Note**).

In summary, our comparison of 12 populations demonstrates that population isolates that have experienced marked recent population growth, starting from a small founding population, have higher overall levels of LD than do outbred populations, and they have far fewer regions of very low LD. Therefore, in such isolates it will be possible to conduct genome-wide association studies using fewer SNPs than in outbred samples. Similarly, these isolates will have fewer regions in which associations are undetectable by off-the-shelf marker sets now available for genome-wide studies.

METHODS

Samples. We genotyped 200 independent persons from each population. The samples from the population isolates were collected as part of genetic studies of common diseases. Subjects were chosen mostly from control groups, except in four populations, where they were drawn from the unrelated parents of affected individuals (**Supplementary Note**). The outbred sample consisted of 60 parents from CEPH trios (the same samples used in ref. 24) and 140 apparently healthy, unrelated self-declared European-Americans from the Coriell Institute Human Variation Panel. Informed consent was obtained from all study subjects, and the University of California, Los Angeles Institutional Review Board approved this study.

Before we began genotyping, we determined concentrations of genomic DNA samples by a Quant-iT PicoGreen (Invitrogen) dsDNA assay. Samples were then normalized to a concentration of 40–70 ng/μl. Samples containing <40 ng/μl genomic DNA were concentrated by partial dehydration to 90 °C using a 96-well heat block. Samples containing more than 70 ng/μl genomic DNA were diluted using 1× TE.

SNP markers, genotyping and scoring. Genotyping reactions were performed using Illumina BeadLab kit reagents and protocols[25]. We performed genotyping, using 250 ng of normalized genomic DNA from each sample as template for the Illumina GoldenGate genotype assay, on the Illumina BeadLab 1000 platform. Samples were biotinylated and then captured by avidin-coated paramagnetic particles. Pooled oligonucleotides were then hybridized to the bound genomic DNA. After removal of unbound or mismatched oligonucleotides by post-hybridization washes, a polymerase-mediated extension of the allele-specific oligonucleotides (ASOs) was performed to fill in the gap between the ASOs and locus-specific oligonucleotides (LSOs). A separate ligation step sealed the nick between the end of the extension and the LSO(s). These single-stranded templates were then used in a multiplex PCR reaction. Simultaneous PCR amplification of 1,536 loci per sample per oligonucleotide pool was performed using Titanium Taq (BD Biosciences) and 0.5 units UDG (Invitrogen) per 60 μl reaction.

After post-PCR washes, the amplification products were hybridized to derivatized beads arranged in a Sentrix Array Matrix (SAM). Each array contains 96 bundles of 50,000 fiber-optic filaments. Each filament is chemically etched to create a 3-μm well. Each well houses a single bead, and each bead contains the complementary address sequence of one of the oligonucleotide pool loci. With 50,000 beads and 1,536 bead types, each bead type is represented in the array approximately 30 times. Each genotype is, therefore, an average of 30 independent measurements of fluorescence intensities. Collection of fluorescence intensities was performed on a BeadArray Reader (Illumina) confocal fluorescence scanner. Data collection was accomplished using the BeadScan (Illumina) software package.

GenCall version 6.0.7 (Illumina) was used to cluster fluorescence intensities. Automated clustering was reviewed by hand and corrected where appropriate. The clustering for each population was reviewed and adjusted separately. After clustering, genotypes were called and quality scores assigned by GenCall. GTS reports (Illumina) were used to remove failed markers, failed samples and individual genotypes with quality scores below the 0.25 threshold from the final data set.

Error checking and selection of SNPs for LD analysis. Only the 2,589 markers scored in all 12 populations were considered for analysis. We then evaluated these markers by testing for Hardy-Weinberg equilibrium, estimating homozygosity and checking for mendelian transmission (**Supplementary Methods** online); the 2,486 markers that met all criteria were used in subsequent analyses.

Comparison of allele frequencies and calculation of pairwise LD measurements. Allele frequency distributions in each population were compared with the outbred sample using a χ^2 test. A marker was considered to have a significantly different allele frequency distribution from the outbred sample if $\chi^2 > 18$, to correct for multiple comparisons. The genetic distance between all pairs of populations was estimated using Nei's D[15], using all marker data and the GENDIST program of the PHYLIP package. Nei's D is estimated by

$$D = -\ln\left(\frac{\sum_m \sum_i p_{1mi} p_{2mi}}{(\sum_m \sum_i p_{1mi}{}^2)^{0.5} (\sum_m \sum_i p_{2mi}{}^2)^{0.5}}\right) \quad (1)$$

where m indexes the number of markers, i indexes the alleles at each marker and p_{1mi} is the frequency of allele i of marker m in population 1 (similarly for population 2). The matrix of genetic distances was used as a difference matrix in Kruskal's non-metric multidimensional scaling to obtain a two-dimensional representation of the proximity of populations to one another. Estimates of D' and r^2 between all marker pairs in each population were calculated using GOLD[26].

Construction of metric LD maps. For LD map construction[17] we computed the pairwise association, ρ, which equates to the absolute value of D', for pairs of SNPs and the corresponding information K_ρ under the null hypothesis that $\rho = 0$, such that $\rho^2 K_\rho = \chi^2_1$ (ref. 27). The predicted association between two SNPs separated by distance $\sum d_i$ (in kb) is modeled as

$$\rho = (1 - \mathrm{L})\mathrm{M}e^{-\sum \varepsilon_i d_i} + \mathrm{L}.$$

Here, d_i is the length in kb of the ith map interval between adjacent SNPs, ε_i is the exponential decline of association with distance and the intercept M is <1 if founding haplotypes have polyphyletic origin. The asymptote L reflects association at large distance that is not due to linkage and is predicted as the weighted mean deviation for a normal distribution[28]. The estimation of ε_i is achieved by composite likelihood using the informative pairwise data that include the ith interval. SNP pairs separated by more than 500 kb and pairs separated by more than 100 map intervals were excluded as uninformative. The

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

M parameter was re-estimated at regular intervals during map construction, which proceeds by iteratively computing the additive LD unit distances $\sum \varepsilon_i d_i$. The model is fitted from the composite likelihood

$$\exp[-\sum K_\rho(\hat{\rho} - \rho)^2/2],$$

where $\hat{\rho}$ is an estimate with prediction ρ and the summation is over all SNP pairs used for analysis. LDU maps were assembled in segments of ~500 markers, with a 25-marker overlap between adjacent segments and the overlap LDU distance averaged. These defaults give rapid construction of LD maps with the flanking intervals contributing efficiently to each interval estimate.

A delete-d jackknife estimate[29] of the s.e.m. around the length of each population's LD map was found by taking 100 samples of 180 persons and recomputing the LD map for each jackknife sample. The s.e.m. for *m* draws of size $n - d$ from a sample of *n* subjects is found by

$$\left\{\frac{n-d}{dm}\sum_{i=1}^{m}\left(L_{n-d,i} - \frac{1}{m}\sum_{j=1}^{m} L_{n-d,j}\right)^2\right\}^{0.5} \quad (2)$$

In our application, $m = 100$, $d = 20$, $n = 200$ and $L_{n-d,i}$ represents the length of the LD map from the *i*th sample of $n - d$ persons. We chose to use a jackknife estimate of the s.e.m. rather than a bootstrap estimate, because resampling persons with replacement (as in bootstrap sampling) may result in more LD than actually observed and may not accurately reflect the distribution of the length of the LD map.

LD hole distribution. To assess the significance of different total number of LD holes (defined as gaps between adjacent SNPs in the LD map of ≥2.5 LDU) across populations, we considered a model where each hole is an observation from a random element that can take on 12 possible values corresponding to the 12 populations under study. Fixing the number of holes to the observed 512, we considered the expected counts for each of the populations under the hypothesis of equal probability of all the populations and contrasted them with the observed totals for each population using a χ^2 test. We then analyzed population subgroups easily identifiable from the data to determine if the hypothesis of uniform probability would hold within these groups.

Recombination and LD holes. We downloaded the recombination data (calculated using the method of ref. 30) generated for the CEPH samples from release #16 of the HapMap data and compared the recombination rate from these data (averaged over windows of 3 Mb, with overlap of 1.5 Mb) to the total number of LD holes (in the same 3-Mb windows, with 1.5-Mb overlap).

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS

We thank all of the study participants; the Center of Medical Systems Biology (CMSB); D. Ruano (Life and Health Sciences Research Institute (ICVS), University of Minho, Braga, Portugal); M.J. Soares, J. Valente and M.H. Azevedo (Instituto de Psicologia Médica, Faculdade de Medicina, Coimbra, Portugal); C. Pato and M.T. Pato (Center for Psychiatric and Molecular Genetics, and Department of Psychiatry, State University of New York, Syracuse, New York, and the Veterans Administration Medical Center, Washington, D.C.); A. Gabbas (Haematology Division and Bone Marrow Transplantation Unit, San Francesco Hospital, Nuoro, Italy); R. van Wyk, C. Botha and G. Valencia (Universidad de Antioquia) and P. Snijders for recruitment of subjects; M. Almonte and E. Slaten for genotyping assistance; R. Ophoff for comments on the manuscript and M. Levinson for assistance with graphics. We acknowledge funding from the Biotechnology and Biological Sciences Research Council (A.C.), the US National Institutes of Health (N.F., M.K., A.R.-L.), Newfound Genomics (P.R.), Colciencias (A.R.-L.), the Universidad de Antioquia (A.R.-L.), the National Alliance for Research in Schizophrenia and Depression (A.R.-L.), the Wellcome Trust (A.R.-L.), the Netherlands Organisation for Scientific Research (C. van D.), the Netherlands Diabetes Fund (C. van D.), the Netherlands Kidney Fund (C. van D.), the Netherlands Heart Foundation (C. van D.), the International Alzheimer Organisation (C. van D.), the Netherlands Brain Fund (C. van D.), the Center of Excellence of the Academy of Finland (L.P.), Nordic Center of Excellence in Disease Genetics (L.P.) and Biocentrum Helsinki, Finland (L.P.).

COMPETING INTERESTS STATEMENT

The authors declare competing financial interests (see the *Nature Genetics* website for details).

Published online at http://www.nature.com/naturegenetics
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/

1. Thomas, D.C., Haile, R.W. & Duggan, D. Recent developments in genomewide association scans: A workshop summary and review. *Am. J. Hum. Genet.* **77**, 337–345 (2005).
2. Stumpf, M.P.H. & Goldstein, D.B. Demography, recombination hotspot intensity, and the block structure of linkage disequilibrium. *Curr. Biol.* **13**, 1–8 (2003).
3. Peltonen, L., Palotie, A. & Lange, K. Use of population isolates for mapping complex traits. *Nat. Rev. Genet.* **1**, 182–190 (2000).
4. Wright, A.F., Carothers, A.D. & Piratsu, M. Population choice in mapping genes for complex disease. *Nat. Genet.* **23**, 397–404 (1999).
5. Pritchard, J.K. & Przeworski, M. Linkage disequilibrium in humans: models and data. *Am. J. Hum. Genet.* **69**, 1–14 (2001).
6. Teare, M.D., Dunning, A.M., Durocher, F., Rennart, G. & Easton, D.F. Sampling distribution of summary linkage disequilibrium measures. *Ann. Hum. Genet.* **66**, 223–233 (2002).
7. Tenesa, A. *et al.* Extent of linkage disequilbirum in a Sardinian sub-isolate: sampling and methodological considerations. *Hum. Mol. Genet.* **13**, 25–33 (2004).
8. Groenewald, J.Z., Liebenberg, J., Groenewald, I.M. & Warnich, L. Linkage disequilibrium analysis in a recently founded population: evaluation of the variegate porphyria founder in South African Afrikaners. *Am. J. Hum. Genet.* **62**, 1254–1258 (1998).
9. Gordon, D., Simonic, I. & Ott, J. Significant evidence for linkage disequilibrium over a 5-cM region among Afrikaners. *Genomics* **66**, 87–92 (2000).
10. Service, S.K., Ophoff, R. & Freimer, N.B. The genomewide distribution of background linkage disequilbrium in a population isolate. *Hum. Mol. Genet.* **10**, 545–551 (2001).
11. Hall, D., Wijsman, E.M., Roos, J.L., Gogos, J.A. & Karayiorgou, M. Extended inter-marker linkage disequilibrium in the Afrikaners. *Genome Res.* **12**, 956–961 (2002).
12. Varilo, T. *et al.* The interval of linkage disequilibrium (LD) detected with microsatellite and SNP markers in chromosomes of Finnish populations with different histories. *Hum. Mol. Genet.* **12**, 51–59 (2003).
13. Aulchenko, Y.S. *et al.* Linkage disequilibrium in young genetically isolated Dutch population. *Eur. J. Hum. Genet.* **12**, 527–534 (2004).
14. Maraganore, D.M. *et al.* High-resolution whole-genome association study of Parkinson disease. *Am. J. Hum. Genet.* **77**, 685–693 (2005).
15. Nei, M. Genetic distance between populations. *Am. Nat.* **106**, 283–292 (1972).
16. Dawson, E. *et al.* A first-generation linkage disequilibrium map of human chromosome 22. *Nature* **418**, 544–548 (2002).
17. Maniatis, N. *et al.* The first linkage disequilibrium (LD) maps: Delineation of hot and cold blocks by diplotype analysis. *Proc. Natl. Acad. Sci. USA* **99**, 2228–2233 (2002).
18. Ke, X. *et al.* The impact of SNP density on fine-scale patterns of linkage disequilibrium. *Hum. Mol. Genet.* **13**, 577–588 (2004).
19. Tapper, W.J., Maniatis, N., Morton, N.E. & Collins, A. A metric linkage disequilibrium map of a human chromosome. *Ann. Hum. Genet.* **67**, 487–494 (2003).
20. De La Vega, F.M. *et al.* The linkage disequilibrium maps of three human chromosomes across four populations reflect their demographic history and a common underlying recombination pattern. *Genome Res.* **15**, 454–462 (2005).
21. Kong, X. *et al.* A combined linkage-physical map of the human genome. *Am. J. Hum. Genet.* **75**, 1143–1148 (2004).
22. Maniatis, N. *et al.* Positional cloning by linkage disequilibrium. *Am. J. Hum. Genet.* **74**, 846–855 (2004).
23. Carvajal-Carmona, L.G. *et al.* Genetic demography of Antioquia (Colombia) and the central valley of Costa Rica. *Hum. Genet.* **112**, 534–541 (2003).
24. The International HapMap Consortium. The International HapMap Project. *Nature* **426**, 789–796 (2003).
25. Fan, J.-B. *et al.* Highly parallel SNP genotyping. *Cold Spring Harb. Symp. Quant. Biol.* **68**, 69–78 (2004).
26. Abecasis, G.R. & Cookson, W.O. GOLD–graphical overview of linkage disequilibrium. *Bioinformatics* **16**, 182–183 (2000).
27. Collins, A. & Morton, N.E. Mapping a disease locus by allelic association. *Proc. Natl. Acad. Sci. USA* **95**, 1741–1745 (1998).
28. Morton, N.E. *et al.* The optimal measure of allelic association. *Proc. Natl. Acad. Sci. USA* **98**, 5217–5221 (2001).
29. Efron, B. & Tibshirani, R.J. *An Introduction to the Bootstrap* (Chapman and Hall, New York, 1993).
30. McVean, G.A. *et al.* The fine-scale structure of recombination rate variation in the human genome. *Science* **304**, 581–584 (2004).

© 2006 Nature Publishing Group http://www.nature.com/naturegenetics

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Genetic Epidemiology 31: 189–194 (2007)

Tag SNPs Chosen From HapMap Perform Well in Several Population Isolates

Susan Service,[1] The International Collaborative Group on Isolated Populations,[2] Chiara Sabatti,[3,4] and Nelson Freimer[1,5,6]*

[1]Center for Neurobehavioral Genetics, University of California, Los Angeles, California
[2]The International Collaborative Group on Isolated Populations members that are not listed separately as authors of this manuscript are: Maria Karayiorgou, J. Louw Roos, Herman Pretorious, Gabriel Bedoya, Jorge Ospina, Andres Ruiz-Linares, António Macedo, Joana Almeida Palha, Peter Heutink, Yurii Aulchenko, Ben Oostra, Cornelia van Duijn, Marjo-Riitta Jarvelin, Teppo Varilo, Lynette Peddle, Proton Rahman, Giovanna Piras, Maria Monne and Leena Peltonen
[3]Department of Human Genetics, University of California, Los Angeles, California
[4]Department of Statistics, University of California, Los Angeles, California
[5]The Jane and Terry Semel Institute for Neuroscience and Human Behavior, University of California, Los Angeles, California
[6]Department of Psychiatry, University of California, Los Angeles, California

Population isolates may be particularly useful for association studies of complex traits. This utility, however, largely depends on the transferability of tag SNPs chosen from reference samples, such as HapMap, to samples from such populations. Factors that characterize population isolates, such as widespread genetic drift, could impede such transferability. In this report, we show that tag SNPs chosen from HapMap perform well in several population isolates; this is true even for populations that differ substantially from the HapMap sample either in levels of linkage disequilibrium or in SNP allele frequency distributions. Genet. Epidemiol. 31:189–194, 2007. © 2007 Wiley-Liss, Inc.

Key words: tag SNP transferability; linkage disequilibrium; genome-wide association

Contract grant sponsor: NIH; Contract grant numbers: MH001375, MH049499, NS037484 and NS040024.
*Correspondence to: Nelson B Freimer, UCLA Center for Neurobehavioral Genetics, Gonda Center, Room 3506, 695 Charles E Young Drive South, Box 951761, Los Angeles CA 90095-1761. E-mail: nfreimer@mednet.ucla.edu
Received 21 July 2006; Accepted 8 November 2006
Published online 23 February 2007 in Wiley InterScience (www.interscience.wiley.com).
DOI: 10.1002/gepi.20201

INTRODUCTION

How to choose markers is a primary question in the design of genetic association studies. This is becoming less of a practical issue for whole-genome association studies where the cost advantages of off-the-shelf arrays typically over-ride other considerations. Most investigations of well-delineated regions of interest, however, such as fine mapping studies or evaluations of specific candidate genes, use custom designed sets of SNPs. Due to the efforts of large-scale SNP discovery programs such as the International HapMap Project [The International HapMap Consortium, 2005], for most such projects there is an overabundance of SNPs available. Therefore, in choosing SNPs for such studies investigators typically take advantage of the fact that, because of linkage disequilibrium (LD) among the SNPs, many SNPs present redundant information. Tag SNPs are those subsets of SNPs that together capture the majority of the information contained in the entire set of genotyped SNPs. Most algorithms use r^2 as a metric by which to choose tag SNPs [Carlson et al., 2004]. An $r^2 \geq 0.8$ is a common threshold for selecting tags, so that all variation is either typed directly as a tag, or is in LD with a tag SNP at a level of $r^2 \geq 0.8$ or greater.

As the values for r^2 for large numbers of pairs of SNPs are currently known only for reference samples such as those of HapMap, these datasets provide the basis for choosing tag SNPs for genotyping in association studies. Implicit in such use of HapMap data is the assumption that the tags chosen from HapMap populations — European Americans, West Africans, Japanese, and Han Chinese — will perform similarly in other samples from populations with similar ancestry.

© 2007 Wiley-Liss, Inc.

WILEY InterScience DISCOVER SOMETHING GREAT

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Genetic Epidemiology 31: 189–194 (2007)

Several factors have been proposed to influence the transferability of tag SNPs between populations, including genetic differences between populations, and differences in the degree of LD between populations.

It has been shown previously that tag SNPs selected from the HapMap European-derived reference sample (CEU — independent members of the CEPH families from Utah) perform relatively well in samples from a wide variety of populations, in that "'75–80% of non-tag SNPs are in LD with CEU selected tag SNPs at a level of $r^2 \geq 0.8$. This observation holds true for samples drawn from or descended from a wide variety of European populations [Mueller et al., 2005; Montpetit et al., 2006; Ribas et al., 2006; Stankovich et al., 2006; Willer et al., 2006]. de Bakker et al. [2006] have shown that CEU tag SNPs perform at a similar level in samples of Chinese and Japanese derivation and in a Latino population from Southern California (with presumed mixed European and Amerindian descent). They found, however, that the CEU tag SNPs performed substantially more poorly (50.4% showed LD at a level of $r^2 \geq 0.8$) only in an African-American sample, consistent with the substantial differences between African and non-African populations in SNP allele frequency distributions and with the markedly lower levels of LD observed in African populations compared to other populations. It is important to note that most of the studies cited above have been conducted in specific, limited genome regions, primarily in gene regions of interest.

González-Neira et al. [2006] studied the portability of tags chosen from 144 SNPs in small samples (14–70 individuals) from 38 different populations. Three of the populations are considered to be reference populations: Yoruban as a reference for Africa, French as a reference for European, and Han Chinese as a reference for Asian. On average, tags selected from the reference samples performed well, in terms of r^2, in test populations that were most like the reference population. This was particularly true for tags selected from the French reference sample and tested in the other European populations. Gonzalez-Neira et al. [2006] found many of the African populations to be particularly difficult to tag, even when choosing tags from the Yoruban reference sample. The authors conceded that their conclusions are limited by small sample sizes and limited genome region (144 SNPs in 1 Mb of chromosome 22), and suggested that further studies in other genome regions, in additional populations is needed.

While González-Neira et al. [2006] examined a wide diversity of populations, none of them were selected specifically to be population isolates. Isolated populations are of tremendous interest for gene mapping due to the fact that they demonstrate increased LD compared to outbred populations, and due to the expectation that founder effects in isolates may reduce the genetic complexity of disease risk in such populations. These population-specific factors might also affect the transferability of tag SNPs derived from reference populations to the isolate populations; for example, genetic drift may have generated allele frequencies in isolates that differ substantially from those in large, heterogeneous populations. We have assayed the transferability of HapMap (CEU) tag SNPs over an extended chromosomal region in relatively large samples (200 individuals) from each of 11 populations that have been considered isolates (all of which have substantial European derivation). We have used these samples previously to evaluate the extent and distribution of LD in comparison with each other and with 200 individuals from a European-derived reference sample [Service et al., 2006]. We previously showed that many of these populations demonstrated dramatically increased LD compared to the reference population as well as significantly different allele frequencies. We now analyze the genotypes from our previous study (2,486 SNPs spaced at a average density of 13.8 kb across chromosome 22) to evaluate the performance of tag SNPs chosen from the HapMap CEU sample in each of the 12 populations.

MATERIALS AND METHODS

We first identified which of the 2,486 SNPs studied by Service et al. [2006] were genotyped in CEU as part of HapMap ($n = 2{,}215$). Data for these SNPs in the CEU sample were downloaded from HapMap and we then chose tags only from this subset of SNPs that were genotyped in both the CEU and the isolate populations. The Haploview program [version 3.2, Barrett et al., 2005] was used to select tag SNPs in CEU from the set of 2,215, using pairwise tagging and an r^2 threshold of 0.8, as implemented in Haploview using the Tagger program [de Bakker et al., 2005]. In each of our 12 populations, we then used Haploview, including as tags the same SNPs chosen as tags in the

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

HapMap CEU sample, and excluding as tags the same SNPs that were not chosen as tags in the CEU sample. The distribution of the maximum r^2 between each of the non-tag SNPs and the tag SNPs was evaluated.

In order to assess the impact of sampling variability on the performance of tag SNPs, we randomly created two samples each of size 60 persons from our 200-person CAU sample. One sample of 60 was considered to be the reference set, and tag SNPs were chosen from this sample using Haploview, and the other sample of 60 was considered to be the test set, and performance of tags selected from the reference set assessed in the test set as described above. We repeated this procedure 100 times.

We also assessed the ability of the HapMap CEU tags to predict variation not seen in the HapMap sample. There were 271 SNPs genotyped in Service et al. [2006] that were not found in HapMap. The distribution of maximum r^2 values between the CEU tags and these 271 SNPs was also evaluated.

Informed consent was obtained from all study subjects, and the UCLA IRB approved this study.

RESULTS

A total of 1,485 tags were selected from the entire set of 2,215 in the CEU sample. The average r^2 value between tags and the remaining 730 SNPs was 0.94. In each of our 12 populations, we then used Haploview, including as tags the 1,485 SNPs chosen in the HapMap CEU sample, and excluding as tags the same 730 SNPs that were not chosen as tags in the CEU sample. The maximum r^2 between each of the 730 non-tag SNPs and the 1,485 CEU tag SNPs was noted, and summary statistics on the distribution of maximum r^2 values for each population are presented in Table I.

The CAU sample includes the 60 CEU individuals from HapMap as well as 140 additional individuals from European-derived outbred populations. As expected, the tag SNPs performed better in this sample than in any of the other populations that we investigated. Nevertheless, in the CAU sample 12% of the r^2 values with non-tag SNPs are less than the threshold of 0.8. We evaluated whether the decreased performance of the tag SNPs in the CAU sample compared to the CEU sample could be explained by the larger size of the former sample (200 compared to 60), as it has been shown that r^2 is biased upwards in small samples [Teare et al., 2002; Terwilliger and Hiekkalinna, 2006]. We took a random sample of 60 individuals from CAU and found that the tag SNPs performed very similarly in the subsample as in the larger CAU sample.

We then considered the possibility that, because of chance differences in the composition of the two samples, tag SNPs perform more poorly in a test sample than in a reference sample even when the two samples are drawn from the same population. Barrett and Cardon [2006] suggested that tags identified in a reference set are "overfitted" to that reference set, and as such overestimate the coverage of the tag SNP set. We examined the impact of this possibility on our results by randomly creating two samples of 60 persons each (a reference sample and a test sample) from the 200-person CAU sample, as described in the Materials and Methods. On average, the r^2 between tags and non-tags in the randomly created reference sets was 0.94 (SD = 0.003), and by definition 100% of r^2 values were ≥ 0.80. The average (SD) r^2 between tags and non-tags in the randomly created test sets was 0.92 (0.006), and the average percent of marker pairs to have $r^2 \geq 0.8$ was 87% (range: 82.5–91.8%) (Table III). In all 100 replicates, the average r^2 between tags and non-tags was greater in the reference set than in the test set. These results are very similar to the difference we observed between the CEU reference set and our CAU sample, suggesting that the tags are overfitted to the CEU reference set.

In most of the other populations the tag SNPs performed almost as well as they did in the CAU sample. In all of the 11 populations the average r^2 between tags and non-tag markers is greater than 0.8, and in 8 of the 11 populations more than 80% of the marker pairs displayed r^2 above this value. In three populations the tag SNPs performed more poorly than in the CAU sample; in CR about 75% of marker pairs showed LD at $r^2 \geq 0.8$, and in AZO and ANT less than 70% of marker pairs showed LD at this threshold. Tag SNP performance in the 11 isolates, as measured by the percent of non-tags to have r^2 with a tag at a level of 0.80 or greater, was not well correlated with either the length of the LD map [a measure of the extent of LD, presented in Service et al., 2006] or with the percent of SNPs to have significantly different allele frequency distributions from the CAU reference sample ($r = 0.28$ and 0.25, respectively).

The ability of the CEU tags to predict variation not seen in HapMap was uniformly low for all

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

TABLE I. Summary information on the distribution of the maximum r^2 between the 1,485 HapMap CEU tags and the 730 non-tags in the CEU sample in 12 populations examined in Service et al. [2006]

Population[a]	Average r^2	% r^2≥0.8[b] (%)	Min	First Q	Median	Third Q	Max
CEU	0.94	100.00	0.80	0.88	0.96	1.00	1.00
CAU	0.92	88.08	0.37	0.87	0.95	0.99	1.00
CAU-60	0.92	87.81	0.25	0.87	0.96	1.00	1.00
FIP	0.91	86.58	0.24	0.86	0.96	0.99	1.00
FIC	0.92	86.30	0.37	0.89	0.97	1.00	1.00
ERF	0.91	85.07	0.14	0.87	0.96	0.99	1.00
NFL	0.91	84.93	0.36	0.87	0.95	0.98	1.00
FIK	0.92	83.97	0.12	0.88	0.98	1.00	1.00
ASH	0.89	83.84	0.27	0.85	0.94	0.98	1.00
SAR	0.89	80.68	0.20	0.85	0.94	0.98	1.00
SAF	0.89	80.27	0.37	0.84	0.93	0.97	1.00
CR	0.86	75.62	0.13	0.81	0.90	0.97	1.00
AZO	0.81	65.48	0.14	0.76	0.85	0.91	0.99
ANT	0.81	63.15	0.10	0.75	0.85	0.92	1.00

[a]Population abbreviations: CEU; CEPH parents from Utah families; CAU: Caucasian reference sample from the Human Variation Panel of the Coriell Institute together with the CEU sample; CAU-60: 60 randomly selected persons from the full CAU sample; FIP: Finland nationwide, mostly early settlement Finland; FIC: Northern Finland Birth Cohort, a mix of early and late settlement Finland; ERF: a village in Southwestern Netherlands; NFL: Newfoundland; FIK: Finnish subisolate of Kuusamo, from late settlement Finland; ASH: Ashkenazi; SAR: Province of Nuoro in Sardinia; SAF: Afrikaner; CR: Central Valley of Costa Rica; AZO: Azores; ANT: Antioquia, Colombia.
[b]% r^2≥0.8 is the percentage of non-tag SNPs to have r^2 values with a CEU tag SNP above the indicated thresholds.
The N for the comparisons is 730.

TABLE II. Summary information on the distribution of the maximum r^2 between the 1,485 HapMap CEU tags and the 271 SNPs not found in the HapMap sample but genotyped in the 12 populations examined in Service et al. [2006]

Population[a]	Average r^2	% r^2≥0.8[b] (%)	Min	First Q	Median	Third Q	Max
CAU	0.53	35.79	0.02	0.22	0.52	0.90	1.00
FIP	0.54	35.79	0.03	0.18	0.53	0.89	1.00
FIC	0.54	36.90	0.02	0.19	0.51	0.94	1.00
ERF	0.54	35.42	0.03	0.20	0.57	0.91	1.00
NFL	0.53	32.84	0.02	0.18	0.56	0.89	1.00
FIK	0.56	35.79	0.03	0.21	0.55	0.96	1.00
ASH	0.53	34.32	0.02	0.19	0.49	0.88	1.00
SAR	0.53	32.10	0.01	0.19	0.52	0.89	1.00
SAF	0.52	33.21	0.02	0.20	0.52	0.87	1.00
CR	0.52	31.36	0.03	0.19	0.52	0.86	1.00
AZO	0.47	21.40	0.02	0.18	0.46	0.77	0.96
ANT	0.48	25.09	0.03	0.16	0.46	0.80	1.00

[a]Population abbreviations: see Table I.
[b]% r^2≥0.8 is the percentage of non-HapMap SNPs to have r^2 values with a CEU tag SNP above the indicated thresholds.
The N for the comparisons is 271.

TABLE III. Summary information on the distribution, over 100 replicates, of the average r^2 between the tags chosen from a randomly created reference set, and tested in a randomly chosen test set (see Methods for details)

Sample	Average r^2	% r^2≥0.8[a] (%)	Min	First Q	Median	Third Q	Max
Reference Set	0.936	100	0.929	0.934	0.936	0.938	0.942
Test Set	0.918	87.32	0.899	0.914	0.919	0.922	0.933

[a]% r^2≥0.8 is the average percentage of non-tag SNPs in the test set to have r^2 values with a tag SNP chosen from the reference set above the indicated thresholds.

isolates (Table II). The average (SD) of the mean minor allele frequency, across the 12 populations for these 271 SNPs was 0.275 (0.005), which is not significantly different from the average mean minor allele frequency for the 2,215 SNPs that were found in HapMap (average = 0.279, SD = 0.003). The percent of SNPs to have minor allele frequency less than 5% was slightly higher,

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

however, in the set of 271 SNPs (range over the 12 populations: 3.3–8.1%) than in the set of 2,215 SNPs (range: 2.5–5.2%), which may account in part for the poorer performance of the CEU tags on these 271 SNPs (see also Table III).

DISCUSSION

Overall there is no substantial difference in the performance of the tag SNPs between populations characterized [in Service et al., 2006] by particularly high LD (e.g. the three Finnish samples) and populations characterized by much less extensive LD, similar to that seen in CAU (e.g. NFL and SAF, population abbreviations are found in Table I). Similarly, the tag SNPs performed equally well in populations shown [in Service et al., 2006] to have allele frequency distributions substantially different from that of CAU (e.g. the three Finnish samples and ASH) and populations whose allele frequency distributions were not different from that of CAU (e.g. NFL and ERF).

In samples from the populations of CR, ANT and AZO, the tag SNPs performed noticeably more poorly than in the CAU sample, in terms of the percent of non-tag SNPs to have maximum r^2 greater than 0.8 with a tag SNP. As in the other populations, there is no obvious relation between tag SNP performance and overall levels of LD [CR and ANT showed high LD and AZO showed relatively low levels of LD in Service et al., 2006] or deviations in allele frequency compared to CAU [ANT and CR showed substantially different allele frequencies from CAU, and AZO showed a similar allele frequency to CAU in Service et al., 2006]. One possible explanation for the poorer performance of the tag SNPs in these populations is that they may be characterized by more extensive derivation from non-European populations compared to the other populations analyzed; evaluation of historical information and of mtDNA, and Y-chromosome polymorphisms indicate a component of sub-Saharan African ancestry in all three populations ("'5–7% in ANT and CR, and from 2 to 10% in AZO), and extensive Amerindian ancestry in ANT and CR [Carvajal-Carmona et al., 2003; Fernando et al., 2005; Pereira et al., 2005; Branco et al., 2006]. In admixed populations, one might consider a strategy of supplementing the tags selected in the CEU sample with those chosen from the one of the other HapMap reference populations (Yoruban, Chinese or Japanese).

Our results suggest that tag SNPs chosen from European-derived reference samples may perform well for association studies in a wide range of population isolates; this is true even for populations that display substantial deviation in allele frequencies and in the extent and distribution of LD compared to reference populations. For practical purposes, the performance of tag SNPs in these populations is comparable to that observed in the reference populations themselves; as we have shown, even in reference populations, there is a modest diminution of performance of tag SNPs when tags are applied in a different sample from the same population. In fact, the percent of non-tag SNPs in 100 randomly created test sets to have maximum r^2 of at least 0.80 with a tag chosen from a reference set of the same population ranged from 82.5% to 91.8%, overlapping the results for most of the populations in Table I. Furthermore, a substantial majority of such tag SNPs perform well even in population isolates known to have experienced admixture between European and non-European populations.

One difference between our study and most previous comparisons of tag SNP transferability between populations is that we have evaluated an extended genome region rather than gene-based segments. The extended region we have examined is all of chromosome 22. As this chromosome has regions of both low and high LD (one parameter critical to the choice and number of tag SNPs), it is likely that results may be generalizable to other portions of the genome. An important limitation of our results is the SNP density employed in our study (on average, 1 SNP every 13.8 kb). This lower density, in comparison to other studies, results in the majority of SNPs being chosen as tags (1,485 of 2,215 SNPs are tags), and very little redundancy in tag selection: the average r^2 between tag SNPs varies from 0.052 (AZO) to 0.07 (FIK). If we limit our observations only to tags with 100 kb of each other, the average r^2 values double, but are still very low (AZO: 0.11, FIK: 0.14). Our conclusions about HapMap tag SNP performance are based on the ability of these tags to characterize the 730 non-tags, and may limit the ability of this study to accurately determine the performance of the tags with respect to the complete spectrum of unknown variation in these populations. The comparability of our results to those of other studies, however, suggests that these differences may be of little practical effect. As with the other studies conducted so far that

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

have evaluated in other populations the performance of tag SNPs chosen from HapMap samples, our evaluation was limited to SNPs chosen for other purposes (in this case investigation of the magnitude and distribution of LD in population isolates). The field is still awaiting the results of studies testing the performance in various populations of tag SNPs chosen from all of the SNPs available from HapMap. Given the numerous large-scale association studies now being undertaken, such results should be available in the not too distant future.

WEB RESOURCES

The International HapMap Project: www.hapmap.org

Haploview Program: www.broad.mit.edu/mpg/haploview/index.php

REFERENCES

Barrett JC, Cardon LR. 2006. Evaluating coverage of genome-wide association studies. Nat Genet 38:659–662.

Barrett JC, Fry B, Maller J, Daly MJ. 2005. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 21:263–265.

Branco CC, Palla R, Lino S, Pachecho PR, Cabral R, De Fez L, Peixoto BR, Mota-Vieira L. 2006. Assessment of Azorean ancestry by Alu insertion polymorphisms. Am J Hum Biol 18:223–226.

Carlson C, Eberle MA, Rieder MJ, Yi Q, Kruglyak L, Nickerson DA. 2004. Selecting a maximally informative set of single-nucleotide polymorphisms for association analysis using linkage disequilibrium. Am J Hum Genet 74:106–120.

Carvajal-Carmona LG, Ophof f R, Service S, Hartiala J, Molina J, Leon P, Ospina J, Bedoya G, Freimer N, Ruiz-Linares A. 2003. Genetic demography of Antioquia (Colombia) and the central valley of Costa Rica. Hum Genet 112:534–511.

de Bakker PIW, Yelensky R, Pe'er I, Gabriel SB, Daly MJ, Altshuler D. 2005. Efficiency and power in genetic association studies. Nat Genet 37:1217–1223.

de Bakker PIW, Graham RR, Altshuler D, Henderson BE, Haiman CA. 2006. Transferability of tag SNPs to capture common genetic variation in DNA repair genes across multiple populations. Pacific Symp Biocomput 11:478–486.

Fernando O, Mota P, Lima M, Silva C, Montiel R, Amorim A, Prata MJ. 2005. Peopling of the Azores Islands (Portugal): data from the Y chromosome. Hum Biol 77:189–199.

Gonzá lez-Neira A, Ke X, Lao O, Calafell F, Navarro A, Comas D, Cann H, Bumpstead S, Ghori J, Hunt H, et al. 2006. The portability of tagSNPs across populations: a worldwide survey. Genome Res 16:323–330.

Montpetit A, Nelis M, Laflamme P, Magi R, Ke X, Remm M, Cardon L, Hudson TJ, Metspalu A. 2006. An evaluation of the performance of tag SNPs derived from HapMap in a Caucasian population. PLOS Genet 2:282–290.

Mueller JC, Lohmussaar E, Magi R, Remm M, Bettecken T, Lichtner P, Biskup S, Illig T, Pfeufer A, Luedemann J, et al. 2005. Linkage disequilibrium patterns and tagSNP transferability among European populations. Am J Hum Genet 76: 387–398.

Pereira L, Cunha C, Alves C, Amorim A. 2005. African female heritage in Iberia: a reassessment of mtDNA lineage distribution in present times. Hum Biol 77:213–229.

Ribas G, Gonzalez-Neira A, Salas A, Milne RL, Vega A, Carracedo B, Gonzalez E, Barrosa E, Fernandez LP, Yankilevich P, et al. 2006. Evaluating HapMap SNP data transferability in a large-scale genotyping project involving 175 cancer-associated genes. Hum Genet 118:669–679.

Service S, DeYoung J, Karayiorgou M, Roos JL, Pretorious H, Bedoya G, Ospina J, Ruiz-Linares A, Macedo A, Almeida Palha J, et al. 2006. Extent, distribution and magnitude of linkage disequilibrium in eleven population isolates. Nat Genet 38: 556–560.

Stankovich J, Cox CJ, Tan RB, Montgomery DS, Huxtable SJ, Rubio JP, Ehm MG, Johnson L, Butzkueven H, Kilpatrick TJ, et al. 2006. On the utility of data from the International HapMap project for Australian association studies. Hum Genet 119: 220–222.

Teare MD, Dunning AM, Duroche F, Rennart G, Easton DF. 2002 Sampling distribution of summary linkage disequilibrium measures. Ann Hum Genet 66:223–233.

Terwilliger JD, Hiekkalinna T. 2006. An utter refutation of the 'Fundamental Theorem of the HapMap'. European J of Hum Genet 14:426–437.

The International HapMap Consortium. 2005. A haplotype map of the human genome. Nature 437:1299–1320.

Willer CJ, Scott LJ, Bonnycastle LL, Jackson AU, Chines P, Prium R, Bark CW, Tsai YY, Pugh EW, Doheny KF, et al. 2006. Tag SNP selection for Finnish individuals based on the CEPH Utah HapMap database. Genet Epidemiol 30:180–190.

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

OPEN ACCESS Freely available online

PLoS one

Nogo Receptor 1 (RTN4R) as a Candidate Gene for Schizophrenia: Analysis Using Human and Mouse Genetic Approaches

Ruby Hsu[1], Abigail Woodroffe[2], Wen-Sung Lai[3], Melloni N. Cook[4], Jun Mukai[3], Jonathan P. Dunning[4], Douglas J. Swanson[5], J. Louw Roos[6], Gonçalo R. Abecasis[7], Maria Karayiorgou[8]*, Joseph A. Gogos[1,3]*

1 Department of Neuroscience, Columbia University, New York, New York, United States of America, 2 Department of Epidemiology, Center for Statistical Genetics, University of Michigan, Ann Arbor, Michigan, United States of America, 3 Department of Physiology and Cellular Biophysics, College of Physicians and Surgeons, Columbia University, New York, New York, United States of America, 4 Department of Psychology, The University of Memphis, Memphis, Tennessee, United States of America, 5 Department of Anatomy and Neurobiology, University of Tennessee Health Science Center, Memphis, Tennessee, United States of America, 6 Department of Psychiatry, University of Pretoria and Weskoppies Hospital, Pretoria, Republic of South Africa, 7 Department of Biostatistics, Center for Statistical Genetics, University of Michigan, Ann Arbor, Michigan, United States of America, 8 Department of Psychiatry, College of Physicians and Surgeons, Columbia University, New York, New York, United States of America

Background. NOGO Receptor 1 (RTN4R) regulates axonal growth, as well as axon regeneration after injury. The gene maps to the 22q11.2 schizophrenia susceptibility locus and is thus a strong functional and positional candidate gene. Methodology/Principal Findings. We evaluate evidence for genetic association between common RTN4R polymorphisms and schizophrenia in a large family sample of Afrikaner origin and screen the exonic sequence of RTN4R for rare variants in an independent sample from the U.S. We also employ animal model studies to assay a panel of schizophrenia-related behavioral tasks in an Rtn4r-deficient mouse model. We found weak sex-specific evidence for association between common RTN4R polymorphisms and schizophrenia in the Afrikaner patients. In the U.S. sample, we identified two novel non-conservative RTN4R coding variants in two patients with schizophrenia that were absent in 600 control chromosomes. In our complementary mouse model studies, we identified a haploinsufficient effect of Rtn4r on locomotor activity, but normal performance in schizophrenia-related behavioral tasks. We also provide evidence that Rtn4r deficiency can modulate the long-term behavioral effects of transient postnatal N-methyl-D-aspartate (NMDA) receptor hypofunction. Conclusions. Our results do not support a major role of RTN4R in susceptibility to schizophrenia or the cognitive and behavioral deficits observed in individuals with 22q11 microdeletions. However, they suggest that RTN4R may modulate the genetic risk or clinical expression of schizophrenia in a subset of patients and identify additional studies that will be necessary to clarify the role of RTN4R in psychiatric phenotypes. In addition, our results raise interesting issues about evaluating the significance of rare genetic variants in disease and their role in causation.

Citation: Hsu R, Woodroffe A, Lai W-S, Cook MN, Mukai J, et al (2007) Nogo Receptor 1 (RTN4R) as a Candidate Gene for Schizophrenia: Analysis Using Human and Mouse Genetic Approaches. PLoS ONE 2(11): e1234. doi:10.1371/journal.pone.0001234

INTRODUCTION

Axonal growth and regeneration is restricted during the maturation of the central nervous system (CNS) as well as after injury, and several lines of evidence indicate that myelin-associated proteins play a critical role in these processes. Early seminal studies by David and Aguayo, 1981 [Ref. 1] demonstrated that the CNS environment contains inhibitory factors which limit growth, and a number of myelin-associated growth inhibitory factors have since been identified, including oligodendrocyte-myelin glycoprotein (OMgp), myelin-associated glycoprotein (MAG), Nogo-A (RTN4), and most recently, Ephrin-B3. Notably, despite their distinct molecular structures, OMgp, MAG, and RTN4 all share a common receptor, Nogo Receptor 1 (RTN4R) [2–4]. RTN4R, a glycosylphosphatidylinositol-linked (GPI-linked) cell surface molecule, forms a heteromeric receptor complex with either LINGO-1 and p75, or LINGO-1 and TROY (a tumor necrosis factor (TNF) receptor family member). Activation of RTN4R initiates a cascade that leads to the activation of RhoA and, ultimately, the inhibition of axonal growth [5–7]. Given that it acts as a convergence point for three separate factors, which inhibit neurite outgrowth and regeneration, RTN4R has generated much interest as a target for therapeutic intervention following CNS injury [2–4,8].

The human RTN4R gene is located within the 22q11.2 locus where relatively common hemizygous microdeletions occur at a frequency of 1 in 5000 live births [9]. The majority of these deletions are de novo events and occur on different haplotype backgrounds [10]. The physical phenotype of the 22q11.2 deletion is broad and variable and most frequently includes congenital heart defects, velopharyngeal defects, and thymic hypoplasia. In addition, most patients display a pattern of cognitive impairments and behavioral deficits, including deficits in working memory, conflict monitoring, visuospatial short-term memory, and executive visual attention [11–14].

Academic Editor: Bernhard Baune, James Cook University, Australia

Received May 20, 2007; Accepted November 1, 2007; Published November 28, 2007

Copyright: © 2007 Hsu et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This research was supported in part by the National Institutes of Health (grant MH67068 to MK and JAG, MH61399 to MK, and U01-MH61971 to Tennessee Mouse Genome Consortium) and by the New York Academy of Sciences (JAG). RH is supported by an HHMI pre-doctoral fellowship. JM is supported by a NARSAD Young Investigator Award. JAG is also the recipient of a McKnight Brain Disorders Award, and a NARSAD Award.

Competing Interests: The authors have declared that no competing interests exist.

* To whom correspondence should be addressed. E-mail: mk2758@columbia.edu (MK); jag90@columbia.edu (JG)

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

In addition, a range of neuroanatomical abnormalities such as reduced total brain volume, enlarged ventricles, and white matter abnormalities [15,16] has been described in some 22q11.2 microdeletion carriers. Both syndromic and non-syndromic patients with the deletion also show extremely high frequencies of psychiatric illness, especially schizophrenia: children with the deletion are 25–30 times more likely to develop schizophrenia or schizoaffective disorder by early adulthood [17,18], and 22q11.2 microdeletions account for ~2% of schizophrenia cases in Caucasian populations [19]. Children with this microdeletion are also reported to have impaired sensorimotor gating [20], which is considered an endophenotype of several psychiatric disorders including schizophrenia, as well as a high incidence of emotional problems including anxiety, depression, social withdrawal, and obsessive-compulsive behaviors [13].

The possibility that RTN4R deficiency contributes to the psychiatric symptoms associated with the 22q11.2 microdeletion, in particular, is intriguing and supported by some preliminary human genetic and gene expression studies. Although, collectively, the human genetic and animal model studies designed to identify schizophrenia susceptibility genes from the 22q11.2 region have implicated primarily three genes: proline dehydrogenase (PRODH), ZDHHC8 and catechol-O-methyltransferase (COMT) [21–25] and their interactions [26,27], a study by Liu et al. [22] presented suggestive evidence that common variants located at the 39 end of the RTN4R gene are associated with schizophrenia in patient samples from the U.S. and South Africa. There has been one attempt to replicate these initial findings in the Han Chinese population, with negative results [28]. However, two preliminary reports described sex-specific associations between schizophrenia and common variants in the RTN4R gene [29], as well as its ligand RTN4 [30]. In addition, Sinibaldi et al. [31] reported two rare non-conservative sequence variants in the RTN4R gene in an Italian sample of 120 schizophrenia patients that were absent in a sample of 200 controls. Moreover, alterations in the levels of RTN4R or two RTN4R ligands have been described in postmortem analyses of brains from individuals with schizophrenia. Specifically, microarray expression studies and single-gene quantitative RT-PCR studies indicate a down-regulation of MAG in at least some schizophrenia cohorts [32,33] and one preliminary study suggested that levels of RTN4 mRNA are also increased in the cortex of some individuals with chronic schizophrenia [34]. Finally, a recent meta-analysis of several expression profiling studies revealed a ~10% decrease ($P = 0.019$) of RTN4R transcript levels in brains of individuals with schizophrenia (www.stanleygenomics.org). It is unclear, however, whether the observed changes constitute part of the genetic diathesis in schizophrenia or represent a reactive response. Here, we undertake a comprehensive multi-pronged approach to explore the possibility that RTN4R is a schizophrenia susceptibility gene from the 22q11.2 locus.

RESULTS

Search for rare RTN4R coding variants

The RTN4R protein is encoded by two exons: exon I, which codes for the first 7 amino acids, and exon II, which encodes amino acids 8–473 (Fig. 1A). We sequenced exon II of RTN4R in 208 individuals with schizophrenia from the U.S. (European ancestry) and identified five variants. Of these five variants, three were synonymous changes (L18L, S192S, P394P), and two resulted in non-conservative changes (T134M, L347R) in evolutionary conserved amino acids (Fig. 1B, C). Each of these five variants was identified in single individuals except for S192S, which was found in two individuals. We also sequenced exon II of RTN4R in a sample of 300 control individuals from the U.S. and found only two synonymous changes (L18L, S192S). The S192S variant was identified in a single individual, whereas the L18L variant was found in two individuals. Previously, Sinibaldi et al. [31] reported two non-conservative sequence variants in the RTN4R gene (R119W and R196H) in an Italian sample of 120 schizophrenia patients that were absent in 200 controls. In our sample, we did not find these two variants in our patient sample, but we did find the synonymous L18L variant reported in the same study.

The RTN4R protein is comprised of two domains: an extracellular N-terminal ectodomain containing eight leucine-rich repeats (LRRs) and a membrane-anchored C-terminal domain containing a glycosylphosphatidylinositol (GPI) anchor sequence [35]. The ectodomain is both necessary and sufficient for binding the RTN4 ligand [35] and the C-terminal domain containing the GPI sequence and the connecting ''stalk'' is required for interaction with at least one of the co-factors, p75 [36]. The T134M variant we identified in one female patient is located in the concave face of the ligand-binding ectodomain of RTN4R, within the fourth LRR. Based on the crystal structure of the soluble ectodomain [37,38] the T134 residue is found on the concave face within a negatively charged patch of amino acids including D111, D114, C138, D139, which forms a slight local depression which could be important for ligand or partner interactions [38] (Fig. 1D). The other non-synonymous variant we identified in one male patient, L347R, is positioned within the ''stalk'' of RTN4R that is required for p75 interaction.

Association analysis

Given the difficulties in assessing the significance of association to rare variants, primarily due to a lack of power [39], we sought to obtain additional supportive data by testing for independent association with common variants of RTN4R (Fig. 1A) in families with schizophrenia. Previously, we provided weak evidence for an association between RTN4R variation and susceptibility to schizophrenia by identifying three distinct variants located at the 39 end of the RTN4R gene that were preferentially transmitted in individuals with schizophrenia in two independent, family-based samples from the U.S. ($N = 106$ triads) and South Africa ($N = 93$ triads) [22]. Here, we expand the South African sample to 312 families by including an additional 219 families with at least one schizophrenic subject (see Table S1 for study population characteristics). Since it has been suggested that variants in RTN4R may affect schizophrenia susceptibility differently in males and females we conducted both combined and sex-stratified analyses.

To facilitate comparison, three of the five single nucleotide polymorphisms (SNPs) tested in the present study (rs665780, rs701427, and rs696880) were also included in the earlier study [22]. Two of these SNPs previously showed weak evidence for association in that study: rs665780 in the Afrikaner families only ($P = 0.04$) and rs701427 in the U.S. families only ($P = 0.02$). Four of the five SNPs in this study are within the RTN4R intron and one is located 39 of the gene. All SNPs are common with a minor allele frequency of at least 20% (see Table 1 for SNP positions and minor allele frequencies). The average (6SD) R^2 and D' between the ten pairs of SNPs are 0.33 (60.42) and 0.62 (60.48), respectively. In the HapMap subjects of European descent (CEU), these five SNPs tag the common (minor allele frequency ~10%) SNPs in RTN4R with an average R2 (6SD) of 0.75 (60.39). Three of the five SNPs (rs701427, rs854971, and rs696880) are separated by less than 2-kb and are in high linkage disequilibrium (LD) with each other ($R^2$$0.90), (see Table 2 for pair-wise values). This group of SNPs, however, shows relatively low LD with the two flanking SNPs (rs665780 and rs1567871).

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

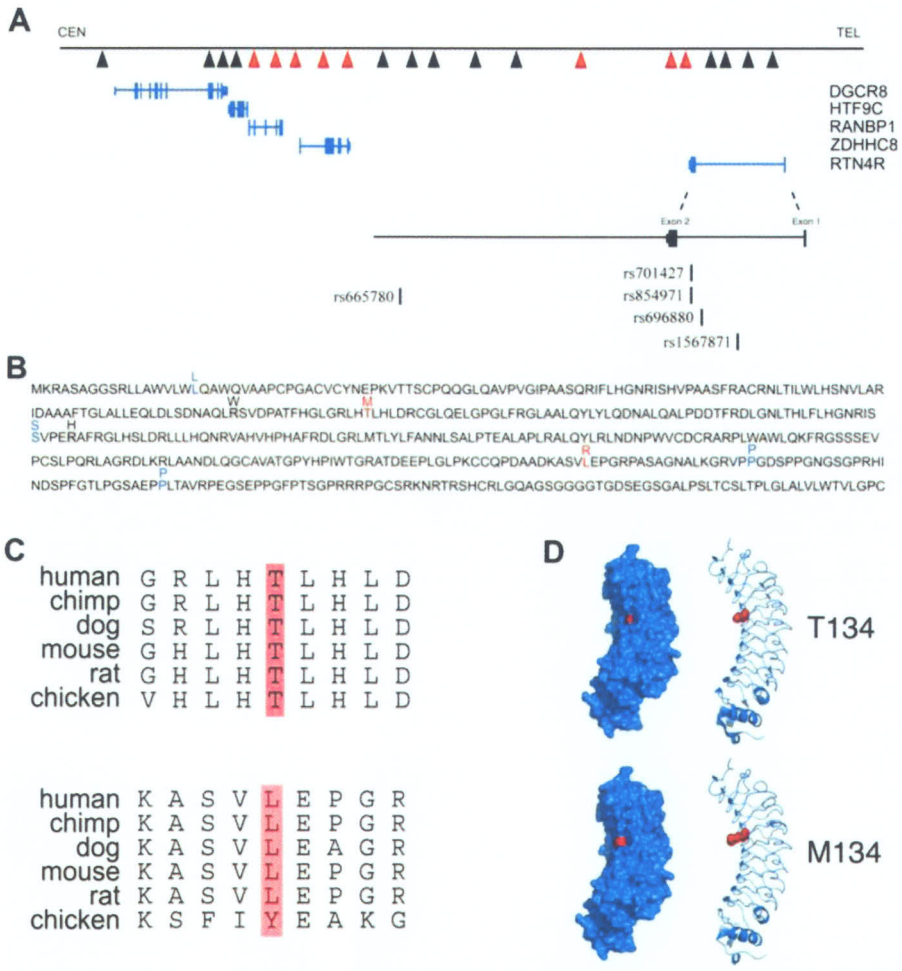


Figure 1. Genetic variation at the RTN4R locus. (A) Location of human RTN4R within chromosomal region 22q11.2. Exons are represented by blue bars. The rs numbers of the SNPs typed in the present study are shown. SNPs genotyped by Liu et al. [22] are indicated by arrowheads (those SNPs that showed significant association with schizophrenia are shown in red arrowheads). The association observed with the three SNPs at the 3′ end of RTN4R was independent from the association observed more proximally in the vicinity of the ZDHHC8 gene [22]. (B) Complete amino acid sequence of RTN4R. Rare coding variants identified in the U.S. schizophrenia sample are shown in blue (synonymous) or red (missense). Two missense variants reported by Sinibaldi et al. [31] are shaded in grey. (C) Evolutionary conservation across species of the missense variants identified in this study at amino acid positions 134 (top) and 347 (bottom). (D) Surface (left) and cartoon (right) representation of the crystal structure of the ligand-binding ectodomain of RTN4R. The variant at residue 134, drawn in red, is shown as the wild-type threonine (top) and the mutated methionine (bottom). Prepared using Pymol (DeLano Scientific, LLC, San Francisco, CA).
doi:10.1371/journal.pone.0001234.g001

Transmission disequilibrium test (TDT) analysis did not find evidence for unequal allelic transmission ratios at any of the tested variants in the combined sample (data not shown). Application of the more powerful linkage and association modeling in pedigrees (LAMP) analysis (Table 3) on the combined sample also provided no evidence for unequal allelic transmission. In a sex-stratified analysis, SNP rs696880 showed the most significant association in females (Sch1: $RR_A = 0.74$, $P = .064$; Sch2: $RR_A = 0.73$, $P = .046$)

Table 1. Position and minor allele frequency of SNPs

SNP	Position (in Mb)	Position (in Gene)	Major/Minor Alleles	Minor Allele Frequency
rs665780	18.560	3′	T/C	.21
rs701427	18.608	intronic	C/A	.34
rs854971	18.608	intronic	G/A	.34
rs696880	18.610	intronic	A/G	.36
rs1567871	18.616	intronic	C/T	.24

doi:10.1371/journal.pone.0001234.t001

Table 2. Pair-wise LD values

	rs665780	rs701427	rs854971	rs696880	rs1567871
rs665780	-	>.01	>.01	>.01	>.01
rs701427	.05	-	.99	.90	.16
rs854971	.05	1.00	-	.91	.16
rs696880	.06	.99	1.00	-	.17
rs1567871	.08	.99	.99	.98	-

doi:10.1371/journal.pone.0001234.t002

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Table 3. LAMP results for individual SNPs

	SNP	Allele	Sch1		Sch2	
			RR	P-Val	RR	P-Val
All	rs665780	T	1.12	-	1.10	-
	rs701427	C	0.97	-	0.98	-
	rs854971	G	0.97	-	0.98	-
	rs696880	A	0.96	-	0.96	-
	rs1567871	C	0.98	-	1.01	-
Females	rs665780	T	1.23	-	1.23	-
	rs701427	C	0.73	.062	0.74	.054
	rs854971	G	0.74	.070	0.74	.064
	rs696880	A	0.74	.064	0.73	.046
	rs1567871	C	1.12	-	1.18	-
Males	rs665780	T	1.02	-	1.00	-
	rs701427	C	1.19	-	1.21	.019
	rs854971	G	1.18	-	1.20	.021
	rs696880	A	1.17	-	1.18	.029
	rs1567871	C	0.91	-	0.91	-

*RR = risk ratio; P-Val = P-value of T_{LD} statistic.
*Bonferroni corrected P-value cutoff = .0050.
doi:10.1371/journal.pone.0001234.t003

Table 4. LAMP results for haplotype analysis (frequency ≥ .05)

	Haplotype*	Freq	Sch1		Sch2	
			RR	P-Val	RR	P-Val
All	TCGAC	.34	0.94	-	0.94	-
	TAAGC	.27	1.05	-	1.01	-
	TCGAT	.17	1.10	-	1.10	-
	CAAGC	.07	0.97	-	1.06	-
	CCGAC	.07	0.95	-	0.98	-
	CCGAT	.06	0.89	-	0.82	-
Females	TCGAC	.34	0.93	-	0.92	-
	TAAGC	.27	1.28	.053	1.29	.055
	TCGAT	.17	0.89	-	0.87	-
	CAAGC	.07	1.30	-	1.27	-
	CCGAC	.07	0.46	.054	0.59	-
	CCGAT	.06	0.86	-	0.77	-
Males	TCGAC	.34	0.96	-	0.98	-
	TAAGC	.27	0.90	-	0.84	-
	TCGAT	.17	1.20	.058	1.23	.061
	CAAGC	.07	0.78	-	0.90	-
	CCGAC	.07	1.32	.083	1.28	.081
	CCGAT	.06	0.94	-	0.91	-

*Alleles for rs665780, rs701427, rs854971, rs696880, and rs1567871, respectively.
*Rare haplotypes (not included in analysis: TCGGC (.018); CCGGC (.007); TCGGT (.001); TAGAC (.001); TAAGT (≥.001).
*RR = risk ratio; P-Val = P-value of T_{LD} statistic.
*At a-level of .05, Bonferroni corrected P-value is .0042.
doi:10.1371/journal.pone.0001234.t004

and non-significant trends were also obtained with the other two linked SNPs, rs701427 and rs854971. In affected males, suggestive evidence for association was observed at rs701427 (Sch1: $RR_C = 1.19$, $P = .10$; Sch2: $RR_C = 1.21$, $P = .019$), as well as at the two SNPs (rs696880 and rs854971) in high LD with it. Interestingly, the alleles of rs701427, rs854971, and rs696880 that appear to increase susceptibility to schizophrenia in males show a trend towards decreasing the risk in females, consistent with the observed lack of association in the combined sample. A Bonferroni correction for multiple SNPs and affection statuses results in a P-value threshold of 0.0042. Therefore, after correcting for multiple testing, none of these results are significant.

When we estimate the haplotype frequencies in our population using an Expectation Maximization (E-M) algorithm, we observe six common haplotypes (frequency ≥5%) and one additional haplotype with a frequency ≥1%. The TDT analysis reveals a proportionate transmission of all haplotypes (data not shown). Similarly, the LAMP analysis does not detect any association between a haplotype and schizophrenia either in the combined or in the sex-stratified sample (see Table 4 for the results from the LAMP analyses).

Generation of Rtn4r-deficient mice

In order to complement our human genetic studies (see Discussion), we generated a mouse model that is deficient in the Rtn4r gene and probed for deficits in phenotypic components, which include endophenotypes associated with schizophrenia that can be modeled reliably in mice [40].

The mouse ortholog of RTN4R is located in the syntenic region of the human 22q11.2 locus that lies on mouse chromosome 16. All human genes except for one are represented in this region, although the order of the genes is different [41]. As a result of this rearrangement, the Rtn4r mouse gene is located between the Prodh gene (≥40-kb proximal) and the Zdhhc8 gene (≥70-kb distal) (Fig. 2A), both of which have been implicated by human genetic studies as strong candidate susceptibility genes for schizophrenia [21–23]. To minimize the possibility that genetic modification of the Rtn4r locus could affect expression of the two neighboring genes, we designed and implemented a gene targeting strategy (Fig. 2B–F) to replace exon II of Rtn4r with the self-excisable selection cassette that includes the neo gene selectable marker (pACN) [40]. Excision of the neo gene, following germline transmission, ensures that any observed phenotype is due to the deletion rather than any long-range transcriptional effects of the selection cassette [42]. Homozygous Rtn4r mutant mice were viable and fertile and have normal brain morphology by gross morphological inspection, as well as finer morphometric calculations of cell densities, laminal thickness and the thickness of the corpus callosum (see Text S1 and Fig. S1).

Behavioral characterization of Rtn4r-deficient mice at baseline

We characterized wild-type, heterozygous, and homozygous Rtn4r mutant mice at baseline using tasks designed to assess domains known to be perturbed in schizophrenia, including anxiety, sensorimotor gating, learning and memory, and behavioral despair. Rtn4r-deficient mice demonstrated normal neurological signs and normal pain threshold levels using the hot plate assay (data not shown). In addition to locomotor activity and anxiety-related tasks, we assessed a) sensorimotor gating using the prepulse inhibition (PPI) assay; b) cognitive performance using contextual and cued fear conditioning, as well as a working memory task; and c) a depression-relevant behavior using the tail-suspension assay. For a description of the equipment and procedures used, see Text S2.

Open field and anxiety-relevant tasks In the open field assay there was a significant effect of genotype on the total distance traveled in the open field [$F(2, 60) = 4.192$, $P \geq 0.02$] and number

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

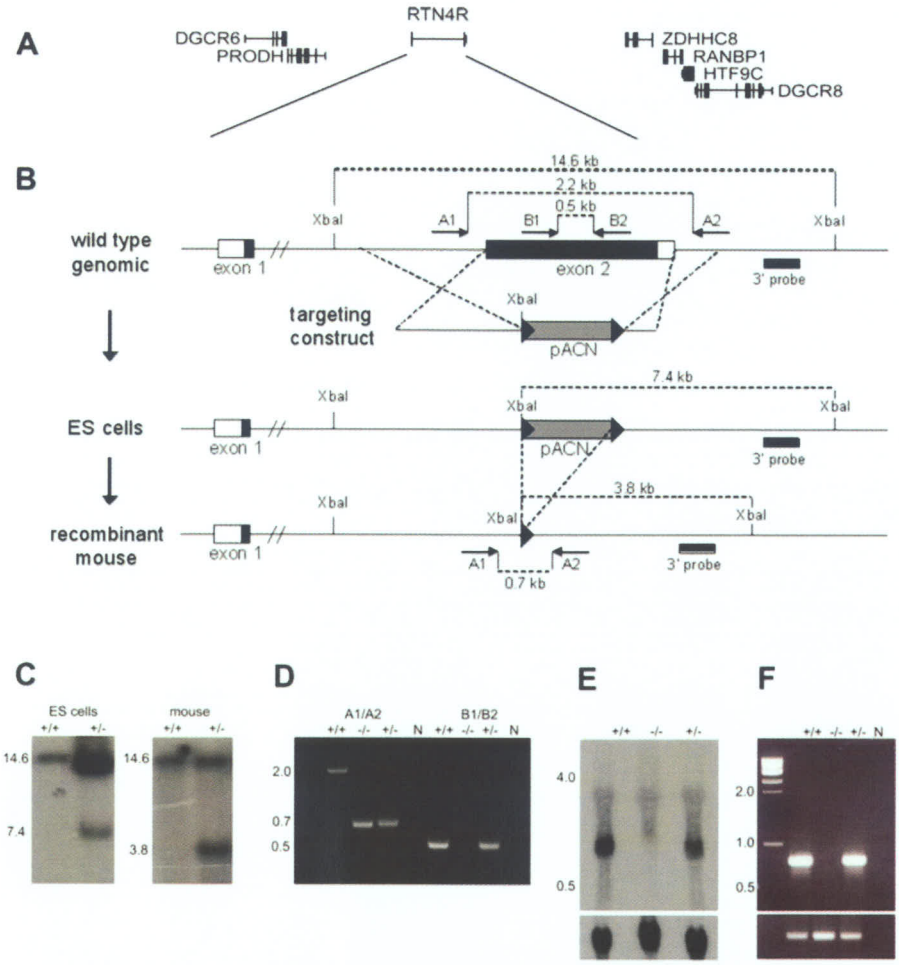


Figure 2. Targeting the Rtn4r locus for homologous recombination. (A) Gene map of the Rtn4r locus at mouse chromosome 16. (B) Exon II of Rtn4r was replaced by the self-excisable pACN cassette including the neo gene for selection. Genomic DNA was digested with XbaI and the 3′ probe was used for southern blot analysis. A PCR strategy was designed to genotype offspring using primer sets A1/A2 and B1/B2. (C) Genomic southern blot of ES cell DNA and tail biopsies. Wild-type fragments were 14.6-kb, recombinant fragments from ES cells were 7.4-kb, and recombinant fragments from tail biopsies after pACN cassette excision were 3.8-kb. (D) PCR genotyping of tail biopsies using primer sets A1/A2 and B1/B2. Wild-type mice yielded 2.2-kb and 0.5-kb fragments from sets A1/A2 and B1/B2 respectively; homozygotes yielded 0.7-kb and no fragment, respectively; and heterozygotes yielded 0.7-kb and 0.5-kb fragments, respectively. (E) Northern blot of total brain RNA using a Rtn4r 3′UTR probe. (F) RT-PCR of total brain RNA using isoform-specific primers for Rtn4r exon II.
doi:10.1371/journal.pone.0001234.g002

of rears (vertical movements) [$F(2, 60) = 9.14$, $P > 0.001$] (Fig. 3A, B). Interestingly, reductions in distance traveled and rears were observed in both the homozygous mutant and heterozygous mice in comparison to wild-type littermates suggesting that the gene could be haploinsufficient with respect to these behaviors. There were no differences in habituation of locomotor activity in the open field (not shown). Differences in the percentage of distance traveled in the central area of the open field, an index of anxiety-relevant behavior, fell short of significance [$F(2, 60) = 2.11$, $P = 0.071$] (Fig. 3C). In the light/dark test, which also assays for anxiety-relevant behaviors, we found no genotype or genotype X sex effects on the number of entries into each compartment (Fig. 3D) or on the percentage of time animals spent in the light compartment of the light/dark apparatus (data not shown).

Sensorimotor Gating Sensorimotor gating is typically evaluated using the PPI paradigm, a common measure of pre-attentive processing [40]. Abnormal PPI has been described in individuals with schizophrenia, as well as in children with the 22q11.2 microdeletion syndrome [20]. In this model, the response to a startle-eliciting stimulus, such as a loud tone, is attenuated if the stimulus is preceded within a few hundred milliseconds by a lower-intensity stimulus, or pre-pulse [43,44]. We evaluated startle response to auditory stimuli as well as inhibition of the startle response elicited by pre-pulses. There were no genotype or genotype X sex differences in inhibition of the startle response at any of the prepulse levels of 70, 80 and 85 dB (Fig. 3E). In addition, in control experiments, there were no genotype or genotype X sex effects on acoustic startle response (120 dB) (Fig. 3F). Notably, although there was no significant effect of genotype, homozygous mutants tended to show slightly more inhibition of startle response than wild-type littermate control mice, as well as slightly less startle response.

Associative learning and memory We used the Pavlovian conditioned fear paradigm to assess associative learning and

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

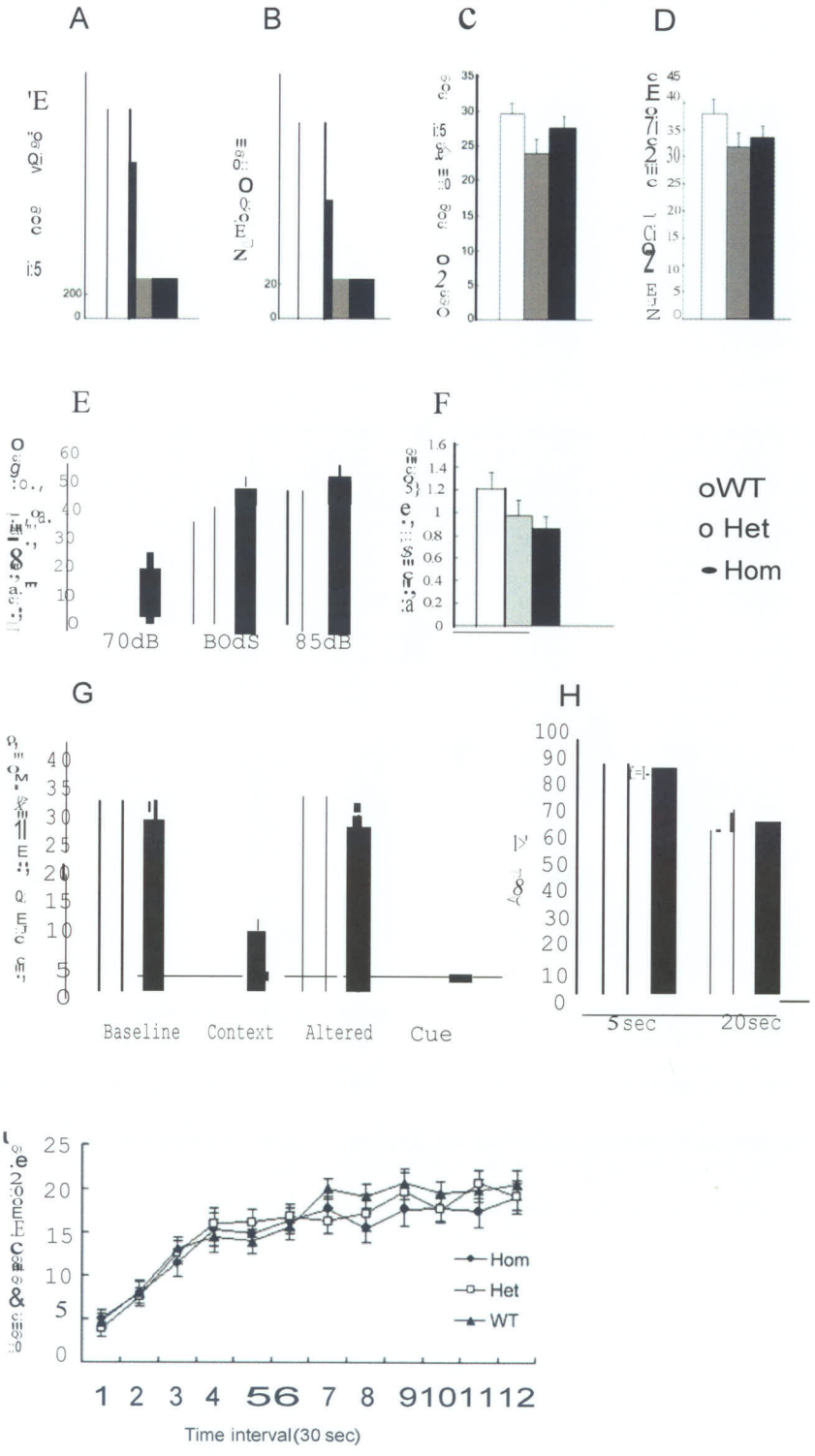


Figure 3. Behavioral characterization of Rtn4r-deficient mice. (A) Total distance traveled, (B) mean number of rears, and (C) ratio of center traveled distance/margin traveled distance in a 2D-min open field test. (D) Number of dark-to-light transitions in a 1o-min dark/light transition test. (E) PPI using combinations of one startle level (120 dB) and three prepulse levels (70, 80 and 85 dB). (F) Startle response at 120 dB. (G) Contextual and cued fear-conditioning paradigm. Activity (beam breaks) was measured at baseline, after reintroduction into the same context where training occurred (context), at baseline in an altered context (altered), and after presentation of the cued conditioned stimulus in the altered environment (cue). (H) Accuracy in a working memory test after a 5-sec or 2D-sec inter-trial delay using a delayed alternation T-maze assay. (I) Mean time spent immobile in a tail suspension test. All results shown are from combined sexes. All data are represented as mean!:S.E.M.
doi:10.1371/journal.pone.0001 234.g003

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

memory in Rtn4r-deficient mice (Fig. 3G). We evaluated both contextual conditioning, as measured by suppressed activity in the context where training occurred, and cued conditioning as measured by suppression of activity in a new environment in which the conditioned stimulus (CS) is presented.

Baseline activity

There were no genotype differences or genotype X sex effects in baseline activity during the training phase of fear conditioning. There were also no differences in activity levels after the third presentation of the unconditioned stimulus (US) and tone pairing.

Contextual Conditioning

There were no genotype or genotype X sex effects in activity during contextual conditioning.

Altered Context

There were no genotypic differences in activity levels in the altered context test, which is both amygdala- and hippocampus-dependent [$F(2, 60) = 2.11$, $P = 0.13$]; however, there was a significant genotype X sex interaction in activity levels in the altered environment [$F(2, 60) = 4.2$, $P > 0.02$].

Cued Conditioning

In the cued version of the test, which requires the amygdala but not the hippocampus [45], there were no genotype or genotype X sex effects in the suppression of activity in response to the conditioned stimulus.

Working memory We used the T-maze delayed alternation task [46,47] to examine whether deficiency in Rtn4r-mediated signaling activity is associated with changes in spatial working memory. Working memory is defined as the ability to maintain and manipulate information transiently in the service of other cognitive processes to guide behavior [48]. Working memory is frequently impaired in patients with schizophrenia (often prior to or at the onset of the illness), as well as in a portion of their non-schizophrenic, first-degree relatives [49,50]. Impaired working memory has also been noted in individuals with the 22q11.2 microdeletion syndrome [12–14]. Rtn4r-deficient mice learned the 5-sec delay T-maze task during 6 consecutive training days and performed as well as wild-type littermate control mice. In addition, there were no genotypic differences or genotype X sex interactions in their T-maze based working memory performance at both 5-sec and 20-sec delays (Fig. 3H).

Behavioral despair High rates of depressive symptoms have been observed in patients with schizophrenia [51] and children with the 22q11.2 microdeletion [12]. We assayed for behavioral despair, an index of depression-like behavior in rodents, using a tail suspension test [52]. The test is based on the fact that animals subjected to the short-term, inescapable stress of being suspended by their tail will develop an immobile posture. Various antidepressant medications reverse the immobility and promote the occurrence of escape-related behavior [53]. No genotypic differences or genotype X sex interactions in total time spent immobile on the tail suspension task were observed (Fig. 3I). A repeated measures ANOVA also showed that there were no genotypic differences or genotype X sex interactions in time spent immobile over intervals of the test session.

Behavioral characterization of Rtn4r-deficient mice following transient postnatal NMDAR blockade

Recent work using an independent animal model suggests that RTN4R may function in the mammalian CNS to stabilize neural circuits in early postnatal development [54]. It is therefore conceivable that RTN4R deficiency facilitates the effect of other schizophrenia susceptibility genes or disease-related pathological processes by destabilizing relevant neural circuits during development. Such a contribution might not be detectable with the study design outlined above but may require additional genetic or pharmacological challenges. Aberrant NMDA receptor-mediated glutamatergic transmission has been implicated in psychiatric disorders such as schizophrenia, and several pharmacological modeling approaches have been established based on this hypothesis [40]. In one of them, transient inhibition of NMDAR signaling during early postnatal development produces long-lasting behavioral deficits including deficits in domains such as cognitive flexibility, working memory and sensorimotor gating [55,56]. We employed this model to ask whether Rtn4r deficiency modulates the long-term behavioral effects of postnatal NMDA receptor hypofunction. In our experimental paradigm, mouse pups were given injections of the NMDA channel blocker MK-801 during two distinct periods: postnatal days (P) 7–10, which corresponds to a critical period of sensitivity for the NMDAR system, and P11–14, which follows the critical period [57,58]. Sensorimotor gating, as well as spontaneous alternation (an index of normal working memory) and motoric activity in the Y-maze were then tested in adulthood.

Spontaneous alternation Adult homozygous Rtn4r-deficient mice and their wild-type littermates showed, as expected, a tendency to spontaneously alternate but there were no significant differences in spontaneous alternation in the Y-maze due to genotype or drug treatment (Fig. 4A). However, we observed a significant genotype X treatment interaction ($P > 0.05$) in the percentage of time moving, a measure of locomotor activity. Specifically, a significant genotype effect was observed in animals that were injected with MK-801 but not saline at P7–10 ($P > 0.05$). Interestingly, this effect was limited to animals treated at P7–10. Animals that underwent the same treatment at P11–14 showed no significant genotype X treatment interaction (Fig. 4B). Further analysis showed that this difference emerged because while neonatal MK-801 treatment at P7–P10 induced an increase in locomotor activity in the wild-type mice in adulthood, it resulted in a decrease in locomotor activity of adult Rtn4r-deficient mice (Fig. 4C).

Acoustic startle and prepulse inhibition There were no overall significant differences in prepulse inhibition or acoustic startle response due to genotype or drug treatment (Fig. 4D, E). Previously, Harris et al. [56] found that neonatal MK-801 administration resulted in a greater decrease in PPI levels in adult females only. However, we found no significant sex effect in both acoustic startle response and prepulse inhibition.

DISCUSSION

Several lines of evidence suggest that oligodendroglial function and myelin maintenance are disturbed in schizophrenia [reviewed in Davis et al. [59]]. These include imaging and neurocytochemical evidence, changes in white matter, myelin-related gene abnormalities, and morphologic abnormalities in the oligodendroglia observed in brains of individuals with schizophrenia. The relation of these findings to the pathogenesis of the disease is still uncertain. Most intriguing is the possibility of a relation between deficits in myelination and function of the prefrontal cortex, a region that modulates many functions impaired in schizophrenia and whose myelination is completed during late adolescence and early adulthood, a period when overt symptomatology of schizophrenia most commonly emerges. In this context, the location of RTN4R within the replicated and highly penetrant 22q11.2 schizophrenia susceptibility locus makes the gene a prime candidate for influencing disease susceptibility.

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

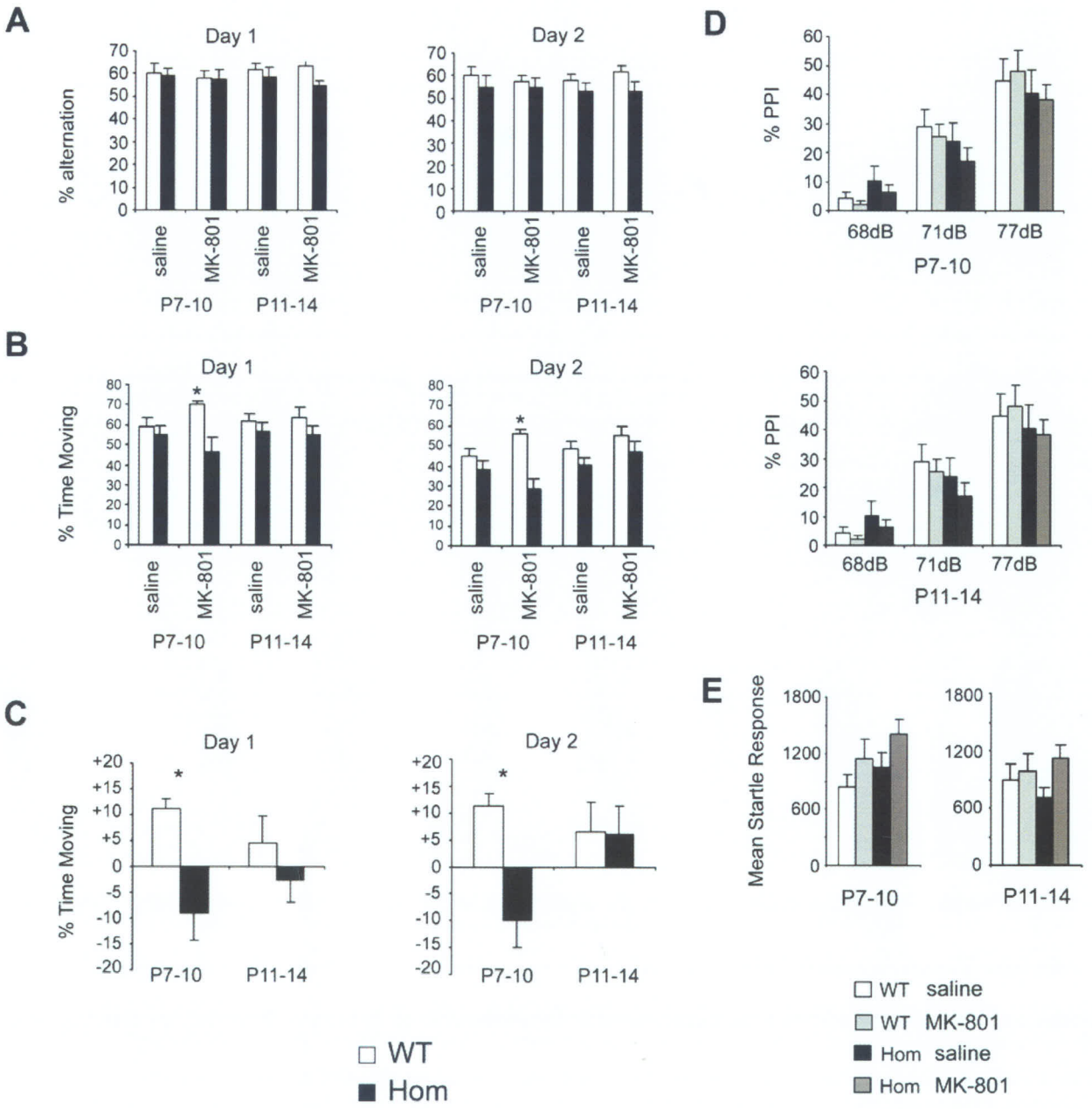


Figure 4. Behavioral characterization of adult Rtn4r-deficient mice following transient postnatal NMDAR blockade. (A, B) Percentage of alternation (A) and % time moving (B) observed in an 8-min session of a Y-maze spontaneous alternation assay, repeated over 2 days. Animals were injected with MK-801 or saline for 4 consecutive days during either P7–10 or P11–14; behavioral testing was conducted at P62 and P63. (C) Relative effect of MK-801 injection normalized to saline injection on time spent moving. (D) Average percentage of PPI (measured at P91) in animals injected during either P7–10 or P11–14. (E) Mean startle response to a 40 ms, 120 dB acoustic stimulus in animals injected during either P7–10 or P11–14. Startle response was measured at P91.
doi:10.1371/journal.pone.0001234.g004

We tested common SNPs and haplotypes in and around RTN4R in a large sample of Afrikaner families. Given that Afrikaners are a relatively genetically and environmentally homogenous population, we expected higher power in our analysis, making disease susceptibility alleles easier to identify. Our analyses did not identify SNPs or haplotypes that were segregating with the disease in the overall sample. Stratification by sex provided weak evidence for an association between RTN4R and schizophrenia, which did not survive a conservative Bonferroni correction.

We also identified rare non-synonymous changes within the RTN4R gene of schizophrenia patients but not in unaffected controls. This observation appears to lend some support for a role of RTN4R in disease susceptibility. However, the major limitation in conducting association analyses of rare variants is that it is

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

difficult to obtain enough statistical power to unambiguously interpret their results. In our study, the sample of several hundred chromosomes does not have sufficient power for our results to be considered unequivocal in proving the association between rare variants of RTN4R and schizophrenia [39,60,61]. Therefore, we cannot exclude the possibility that all four RTN4R non-conservative variants that have been described in patients with schizophrenia (present study and in Sinibaldi et al. [31] study) might actually be rare variants that are neutral in relation to the disease status. In general, interpreting association tests between rare variants and complex traits is an emerging problem in the field [62,63]. The significance of rare variants such as the ones identified in this study can be addressed by screening thousands of both patient and control chromosomes for the presence of non-synonymous variants [61], by searching for rare non-synonymous variants in extended pedigrees (i.e. ones that show linkage to the 22q11.2 locus) where co-inheritance of rare non-conservative variants with disease status can be strictly evaluated, as well as by using model systems to strengthen a functional link between variation in the gene and the disease.

In the context of suggestive human genetic data, the generation of genetic mouse models that mimic the effect of rare hypomorphic variants can be particularly informative [40]. Our animal model studies, designed to facilitate interpretation of our human genetic results, indicated an effect of Rtn4r on locomotor activity in an open field assay. Furthermore, in heterozygous Rtn4r mice, we found a haploinsufficient effect in the open field assay. However, in schizophrenia-related behavioral tasks, such as PPI and the working memory-dependent T-maze delayed alternation task, our studies revealed an unremarkable behavioral profile with normal performance in these tasks, thus arguing against a major role of RTN4R in susceptibility to schizophrenia. It should be noted that using an independently generated Rtn4r-deficient mouse model, Kim et al. [64] also observed hypoactivity in Rtn4r null mice along with decreased rotarod performance, but normal Basso-Beattie-Bresnahan (BBB) scores, a measure of locomotor function. Schizophrenia-related behavioral tasks were not reported in the Kim et al. [64] study.

Obviously, positive results from accurate genetic mouse models can be instrumental in establishing the ''biological plausibility'' of disease-associated genetic variants. However, any interpretation based on negative results from behavioral studies in animal models suffers from several caveats that reflect the limitations of behavioral approaches and the complex genetic structure of the disease. For example, the behavioral effect of any given mutation, especially in a gene that could be involved in a complex disorder, could depend on the age and genetic background of the organism. In addition, we cannot ignore the possibility that RTN4R deficiency contributes to other phenotypic components not assayed here. In that respect, future studies designed to examine the effect of Rtn4r deficit in younger animals or in inbred genetic backgrounds including additional hypothesis-driven behavioral assays will be informative.

Behavioral assessment at baseline may also overlook modulatory effects of a gene in a given behavior that may be revealed by additional genetic or pharmacological challenges. NMDA hypofunction is considered one of the leading hypotheses for schizophrenia pathogenesis and has inspired the generation of several pharmacological models of the disease. Indeed, by adapting such a pharmacological model we found that neonatal NMDAR disruption during a presumably critical period for NMDA activity (P7–10), differentially affects wild-type and Rtn4r-deficient mice. During critical periods of development, widespread pharmacological or genetic inhibition of NMDAR signaling has been shown to facilitate terminal axon sprouting and to modulate the effects of activity-based competition on axon growth [58,65,66]. Such ''wiring'' alteration may contribute not only to the behavioral deficits associated with this pharmacological model, but also to the disease pathogenesis. Moreover, Rtn4r may modulate this aspect of NMDA hypofunction on early brain wiring.

Interestingly, and in agreement with our behavioral analysis at baseline, the observed effects were specific to the motor domain. Although it is very difficult to extrapolate from motor deficits in mice to deficits in motor skills in humans, it is worth noting that patients with schizophrenia display motor development delays, as well as motor disturbances [67,68]. It is also worth noting that motor deficits have been consistently observed in children with the 22q11.2 deletion syndrome including an impairment in fine and gross motor skills that emerges during adolescence [11,69,70]. Importantly, a recent study has shown that motoric deficits are more prevalent among 22q11.2-deleted individuals with schizophrenia than without schizophrenia and along with deficits in verbal learning and social cognition can be used to distinguish these two groups of patients [71].

Our results do not support a major role of RTN4R in susceptibility to schizophrenia or in the cognitive and behavioral deficits observed in individuals with 22q11 microdeletions. However, a hypothesis emerging from the findings described here, as well as from previous independent work (see Introduction) is that RTN4R may modulate the genetic risk for psychiatric phenotypes in a subset of patients, at least partly by mediating the developmental effects of NMDAR receptor hypofunction on early brain wiring. Further genetic studies in afflicted families, as well as analysis of the knockout mice in additional behavioral, synaptic and molecular phenotypes are warranted and will be necessary to test this hypothesis and further clarify any role of RTN4R in psychiatric phenotypes.

MATERIALS AND METHODS

Sample description

All procedures of subject recruitment and evaluation were approved by the Institutional Review Boards at Rockefeller University, New York State Psychiatric Institute and University of Pretoria. Written informed consent was obtained from all participants. The sample used for direct sequencing consisted of 208 individuals with schizophrenia, of European ancestry, recruited from the U.S. and diagnosed by a clinical team specially trained in the use of the Diagnostic Instrument for Genetic Studies (DIGS) [72] and the research application of the Diagnostic and Statistical Manual–4th Edition (DSM-IV) [73]. On the basis of the information gathered in the DIGS, the clinical interviewers assigned appropriate diagnoses according to the DSM-IV. The control sample used for direct sequencing consisted of 60 parents from CEPH trios (residents of Utah with ancestry from Northern and Western Europe) and an additional set of 240 healthy Caucasians collected from the U.S. by us (MK). Sequencing analysis was performed as described in Text S2.

To evaluate association between the RTN4R gene and susceptibility to schizophrenia, we recruited participants from 312 Afrikaner families with at least one schizophrenic subject. Afrikaners are descendants of mostly Dutch immigrants who settled in South Africa beginning in 1652 [74]. Participants were diagnosed in person by specially trained clinicians again using the DIGS, which had been translated and back-translated into Afrikaans. Patients were classified into two diagnostic categories. The stricter category, Sch1, includes 348 individuals who meet DSM-IV criteria for schizophrenia or depressed-type schizoaffec-

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

tive disorder. Family data suggest that the two diagnoses are alternative expressions of the same genotypes [75] and in the past we have considered them together under one phenotypic liability class, LC I [76]. The broader category, Sch2, includes an additional 52 individuals diagnosed with schizoaffective disorder of mainly affective course. Association analysis was performed as described in Text S2.

Generation of Rtn4r knockout mice

For the construction of the targeting construct, we replaced an Rtn4r genomic fragment encompassing exon II with the self-excisable pACN cassette including the neo gene selectable marker [77]. Cell culture, embryonic stem (ES) cell electroporation and generation of the chimeric mice were performed essentially as previously described [78]. Approximately 5% of the tested ES cell clones were positive for homologous recombination, and one clone was selected for injection into C57BL/6J blastocysts. Chimeric males were mated with C57BL/6J females, and DNA from tail biopsy samples from F1 agouti-coat pups was genotyped by Southern blotting at the Rtn4r genomic locus (wild-type fragment: 14.6-kb; recombinant fragment after pACN cassette excision: 3.8-kb). We mated F1 heterozygous mice and obtained F2 mice of all three genotypes. The mutation was on a hybrid C57BL/6×129Sv background. All animal procedures were performed according to protocols approved by the appropriate Institutional Animal Care and Use Committee under the federal and state regulations.

Histological analysis

See Text S2.

Behavioral Testing: Equipment and Procedures

See Text S2.

Analysis following transient postnatal NMDAR blockade

See Text S2.

SUPPORTING INFORMATION

Text S1
Found at: doi:10.1371/journal.pone.0001234.s001 (0.02 MB DOC)

Text S2
Found at: doi:10.1371/journal.pone.0001234.s002 (0.06 MB DOC)

Table S1
Found at: doi:10.1371/journal.pone.0001234.s003 (0.03 MB DOC)

Figure S1 Brain histology in Rtn4r-deficient mice: Representative microphotographs from Nissl staining of coronal sections through the cerebral cortex (A, B), anterior-dorsal hippocampus (C, D) and corpus callosum (E, F) from 8-wk-old Rtn4r-deficient mice (−/−) and their wild-type littermates (+/+). (G, H) Coronal sections through the cerebral cortex of Rtn4r recombinant mice crossed with Thy1-YFPH expressing transgenics. Scale bars represent 0.4 mm. (I) Average cell density of the retrosplenial agranular cortex at Bregma −2.18 mm, −1.70 mm, and −1.22 mm. (J) Average thickness of the retrosplenial agranular cortex. (K) Average thickness of the corpus callosum. All data are represented as mean±S.E.M.
Found at: doi:10.1371/journal.pone.0001234.s004 (9.10 MB TIF)

ACKNOWLEDGMENTS

The authors wish to acknowledge Dionne Swor, Megan Sribour and Jessica Pellegrino for technical support and assistance with the mouse colony, Ying-Jiun Chen for help with the generation of the targeting construct, and members of the Gogos and Karayiorgou labs for useful suggestions.

Author Contributions

Conceived and designed the experiments: JG MK. Performed the experiments: RH WL JM MC JD DS. Analyzed the data: GA JG MK RH WL MC JD DS AW. Contributed reagents/materials/analysis tools: JR. Wrote the paper: GA JG MK RH WL MC AW.

REFERENCES

1. David S, Aguayo AJ (1981) Axonal elongation into peripheral nervous system "bridges" after central nervous system injury in adult rats. Science 214: 931–933.
2. Domeniconi M, Cao Z, Spencer T, Sivasankaran R, Wang K, et al. (2002) Myelin-associated glycoprotein interacts with the Nogo66 receptor to inhibit neurite outgrowth. Neuron 35: 283–290.
3. Liu BP, Fournier A, GrandPre T, Strittmatter SM (2002) Myelin-associated glycoprotein as a functional ligand for the Nogo-66 receptor. Science 297: 1190–1193.
4. Wang KC, Koprivica V, Kim JA, Sivasankaran R, Guo Y, et al. (2002) Oligodendrocyte-myelin glycoprotein is a Nogo receptor ligand that inhibits neurite outgrowth. Nature 417: 941–944.
5. Mi S, Lee X, Shao Z, Thill G, Ji B, et al. (2004) LINGO-1 is a component of the Nogo-66 receptor/p75 signaling complex. Nat Neurosci 7: 221–228.
6. Park JB, Yiu G, Kaneko S, Wang J, Chang J, et al. (2005) A TNF receptor family member, TROY, is a coreceptor with Nogo receptor in mediating the inhibitory activity of myelin inhibitors. Neuron 45: 345–351.
7. Shao Z, Browning JL, Lee X, Scott ML, Shulga-Morskaya S, et al. (2005) TAJ/TROY, an orphan TNF receptor family member, binds Nogo-66 receptor 1 and regulates axonal regeneration. Neuron 45: 353–359.
8. Fournier AE, GrandPre T, Strittmatter SM (2001) Identification of a receptor mediating Nogo-66 inhibition of axonal regeneration. Nature 409: 341–346.
9. Botto LD, May K, Fernhoff PM, Correa A, Coleman K, et al. (2003) A population-based study of the 22q11.2 deletion: phenotype, incidence, and contributuion to major birth defects in the population. Pediatrics 112: 101–107.
10. Morrow B, Goldberg R, Carlson C, Gupta RD, Sirotkin H, et al. (1995) Molecular definition of the 22q11.2 deletions in velo-cardio-facial syndromes. Amer J Hum Genet 56: 1391–1403.
11. Antshel KM, Kates WR, Roizen N, Fremont W, Shprintzen RJ (2005) 22q11.2 deletion syndrome: genetics, neuroanatomy and cognitive/behavioral features. Child Neuropsych 11: 5–19.
12. Sobin C, Kiley-Brabeck K, Daniels S, Khuri J, Taylor L, et al. (2005) Neuropsychological characteristics of children with the 22q11.2 deletion syndrome: a descriptive analysis. Child Neuropsychol 11: 39–53.
13. Woodin M, Wang PP, Aleman D, McDonald-McGinn D, Zackai E, et al. (2001) Neuropsychological profile of chilren and adolescents with the 22q11.2 microdeletion. Genet Med 3: 34–39.
14. Bearden CE, Woodin MF, Wang PP, Moss E, McDonald-McGinn D, et al. (2001) The neurocognitive phenotype of the 22q11.2 deletion syndrome: selective deficit in visual-spatial memory. J Clin Exp Neuropsychol 23: 447–464.
15. Simon TJ, Ding L, Bish JP, McDonald-McGinn DM, Zackai EH, et al. (2005) Volumetric, connective, and morphologic changes in the brains of children with chromosome 22q11.2 deletion syndrome: an integrative study. Neuroimage 25: 169–180.
16. van Amelsvoort T, Daly E, Robertson D, Suckling J, Ng V, et al. (2001) Structural brain abnormalities associated with deletion at chromosome 22q11.2. Brit J of Psych 178: 412–419.
17. Pulver AE, Nestadt G, Goldberg R, Shprintzen RJ, Lamacz M, et al. (1994) Psychotic illness in patients diagnosed with velo-cardio-facial syndrome and their relatives. J Nerv Ment Dis 182: 476–478.
18. Murphy KC, Jones LA, Owen MJ (1999) High rates of schizophrenia in adults with velo-cardio-facial syndrome. Arch Gen Psychiatry 56: 940–945.
19. Karayiorgou M, Morris MA, Morrow B, Shprintzen RJ, Goldberg R, et al. (1995) Schizophrenia susceptibility associated with interstitial deletions of chromosome 22q11.2. Proc Natl Acad Sci USA 92: 7612–7616.
20. Sobin C, Kiley-Brabeck K, Karayiorgou M (2005) Lower prepulse inhibition in children with the 22q11.2 deletion syndrome. Am J Psychiatry 162: 1090–1099.

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

21. Liu H, Heath SC, Sobin C, Roos JL, Galke BL, et al. (2002) Genetic variation at the 22q11.2 PRODH2/DGCR6 locus presents an unusual pattern and increases susceptibility to schizophrenia. Proc Natl Acad Sci USA 99: 3717–3722.
22. Liu H, Abecasis GR, Heath SC, Knowles A, Demars S, et al. (2002) Genetic variation in the 22q11.2 locus and susceptibility to schizophrenia. Proc Natl Acad Sci USA 99: 16859–16864.
23. Mukai J, Liu H, Burt RA, Swor DE, Lai WS, et al. (2004) Evidence that the gene encoding ZDHHC8 contributes to the risk of schizophrenia. Nat Genet 36: 725–731.
24. Shifman S, Bronstein M, Sternfeld M, Pisante-Shalom A, Lev-Lehman E, et al. (2002) A highly significant association between a COMT haplotype and schizophrenia. Am J Hum Genet 71: 1296–1302.
25. Egan MF, Goldberg TE, Kolachana BS, Callicott JH, Mazzanti CM, et al. (2001) Effect of COMT Val108/158 Met genotype on frontal lobe function and risk for schizophrenia. Proc Natl Acad Sci USA 98: 6917–6922.
26. Paterlini M, Zakharenko SS, Lai WS, Qin J, Zhang H, et al. (2005) Transcriptional and behavioral interaction between 22q11.2 orthologs modulates schizophrenia-related phenotypes in mice. Nat Neurosci 8: 1586–1594.
27. Raux G, Bumsel E, Hecketsweiler B, van Amelsvoort T, Zinkstok J, et al. (2006) Involvement of hyperprolinemia in cognitive and psychiatric features of the 22q11 deletion syndrome. Hum Mol Genet 16: 83–91.
28. Meng J, Shi Y, Zhao X, Guo S, Wang H, et al. (2007) No association between the genetic polymorphisms in the RTN4R gene and schizophrenia in the Chinese population. J Neural Transm 114: 249–254.
29. Fallin MD, Belmonte P, Lasseter VK, Cheng N, Avramopoulos D, et al. (2006) Examination of sex and age of onset as important sources of variation for genetic association findings in schizophrenia. Abstract #2076 ASHG Meeting.
30. Pierce T, Bray N, Williams N, Ivanov D, Owen M, et al. (2006) Support for reticulon 4 (Rtn4) as a potential susceptibility gene for schizophrenia. XIV World Congress on Psychiatric Genetics: 179.
31. Sinibaldi L, De Luca A, Bellacchio E, Conti E, Pasini A, et al. (2004) Mutations of the Nogo-66 receptor (RTN4R) gene in schizophrenia. Hum Mutat 24: 534–535.
32. Hakak Y, Walker JR, Li C, Wong WH, Davis KL, et al. (2001) Genome-wide expression analysis reveals dysregulation of myelination-related genes in chronic schizophrenia. Proc Natl Acad Sci USA 98: 4746–4751.
33. Tkachev D, Mimmack ML, Ryan MM, Wayland M, Freeman T, et al. (2003) Oligodendrocyte dysfunction in schizophrenia and bipolar disorder. Lancet 362: 798–805.
34. Novak G, Kim D, Seeman P, Tallerico T (2002) Schizophrenia and Nogo: elevated mRNA in cortex, and high prevalence of a homozygous CAA insert. Brain Res Mol Brain Res 107: 183–189.
35. Fournier AE, Gould GC, Liu BP, Strittmatter SM (2002) Truncated soluble Nogo receptor binds Nogo-66 and blocks inhibition of axon growth by myelin. J Neurosci 22: 8876–8883.
36. Wang KC, Kim JA, Sivasankaran R, Segal R, He Z (2002) P75 interacts with the Nogo receptor as a co-receptor for Nogo, MAG, and OMgp. Nature 420: 74–78.
37. Barton WA, Liu BP, Tzvetkova D, Jeffrey PD, Fournier AE, et al. (2003) Structure and axon outgrowth inhibitor binding of the Nogo-66 receptor and related proteins. EMBO J 22: 3291–3302.
38. He XL, Bazan JF, McDermott G, Park JB, Wang K, et al. (2003) Structure of the Nogo receptor ectodomain: a recognition module implicated in myelin inhibition. Neuron 38: 177–185.
39. Keen-Kim D, Mathews CA, Reus VI, Lowe TL, Herrera LD, et al. (2006) Overrepresentation of rare variants in a specific ethnic group may confuse interpretation of association analyses. Hum Mol Genet 15: 3324–3328.
40. Arguello PA, Gogos JA (2006) Modeling Madness in mice: one piece at a time. Neuron 52: 179–196.
41. Puech A, Saint-Jore B, Funke B, Gilbert DJ, Sirotkin H, et al. (1997) Comparative mapping of the human 22q11.2 chromosomal region and the orthologous region in mice reveals complex changes in gene organization. Proc Natl Acad Sci USA 94: 14608–14613.
42. Olson EN, Arnold HH, Rigby PW, Wold BJ (1996) Know your neighbors: three phenotypes in null mutants of the myogenic bHLH gene MRF4. Cell 85: 1–4.
43. Gogos JA, Santha M, Takacs Z, Beck KD, Luine V, et al. (1999) The gene encoding proline dehydrogenase modulates sensorimotor gating in mice. Nat Genet 21: 434–439.
44. Paylor R, Crawley JN (1997) Inbred strain differences in prepulse inhibition of the mouse startle response. Psychopharmacology (Berl) 132: 169–180.
45. Phillips RG, LeDoux JE (1992) Differential contribution of amygdala and hippocampus to cued and contextual fear conditioning. Behav Neurosci 106: 274–285.
46. Seeger T, Fedorova I, Zheng F, Miyakawa T, Koustova E, et al. (2004) M_2 muscarinic acetylcholine receptor knock-out mice show deficits in behavioral flexibility, working memory, and hippocampal plasticity. J Neurosci 24: 10117–10127.
47. Dudchenko PA (2004) An overview of the tasks used to test working memory in rodents. Neurosci Biobehav Rev 28: 699–709.
48. Baddeley A (1992) Working memory. Science 255: 556–559.
49. Cannon TD, Huttunen MO, Lonnqvist J, Tuulio-Henriksson A, Pirkola T, et al. (2000) The inheritance of neuropsychological dysfunction in twins discordant for schizophrenia. Amer J Hum Genet 67: 369–382.
50. Elvevag B, Goldberg TE (2000) Cognitive impairment in schizophrenia is the core of the disorder. Crit Rev Neurobiol 14: 1–21.
51. Ellenbroek BA, Cools AR (2000) Animal models for the negative symptoms of schizophrenia. Behav Pharmacol 11: 223–233.
52. Porsolt RD (2000) Animal models of depression: utility for transgenic research. Rev Neurosci 11: 53–58.
53. Cryan JF, Mombereau C, Vassout A (2005) The tail suspension test as a model for assessing antidepressant activity: review of pharmacological and genetic studies in mice. Neurosci Biobehav Rev 29: 571–625.
54. McGee AW, Yang Y, Fischer QS, Daw NW, Strittmatter SM (2005) Experience-driven plasticity of visual cortex limited by myelin and Nogo receptor. Science 309: 2222–2226.
55. Stefani MR, Moghaddam B (2005) Transient N-methyl-D-aspartate receptor blockade in early development causes lasting cognitive deficits relevant to schizophrenia. Biol Psychiatry 67: 433–436.
56. Harris LW, Sharp T, Gartlon J, Jones DN, Harrison PJ (2003) Long-term behavioural, molecular and morphological effects of neonatal NMDA receptor antagonism. Eur J Neurosci 18: 1706–1710.
57. Ikonomidou C, Bosch F, Miksa M, Bittigau P, Vockler J, et al. (1999) Blockade of NMDA Receptors and Apoptotic Neurodegeneration in the Developing Brain. Science 283: 70–74.
58. Colonnese MT, Constantine-Paton M (2006) Developmental period for N-methyl-D-aspartate (NMDA) receptor-dependent synapse elimination correlated with visuotopic map refinement. J Comp Neurol 494: 738–751.
59. Davis KL, Stewart DG, Friedman JI, Buchsbaum M, Harvey PD, et al. (2003) White matter changes in schizophrenia: evidence for myelin-related dysfunction. Arch Gen Psychiatry 60: 443–456.
60. Mitchell AA, Chakravarti A, Cutler DJ (2005) On the probability that a novel variant is a disease-causing mutation. Genome Res 15: 960–966.
61. Cohen JC, Kiss RS, Pertsemlidis A, Marcel YL, McPherson R, et al. (2004) Multiple rare alleles contribute to low plasma levels of HDL cholesterol. Science 305: 869–872.
62. Altshuler D, Hirschhorn JN (2005) MEF2A sequence variants and coronary artery disease: a change of heart? J Clin Invest 115: 831–833.
63. Abelson JF, Kwan KY, O'Roak BJ, Baek DY, Stillman AA, et al. (2005) Sequence variants in SLITRK1 are associated with Tourette's syndrome. Science 310: 317–320.
64. Kim JE, Liu BP, Park JH, Strittmatter SM (2004) Nogo-66 receptor prevents raphespinal and rubrospinal axon regeneration and limits functional recovery from spinal cord injury. Neuron 44: 439–451.
65. Lee LJ, Iwasato T, Itohara S, Erzurumlu RS (2005) Exuberant thalamocortical axon arborization in cortex-specific NMDAR1 knockout mice. J Comp Neurol 485: 280–292.
66. Franks KM, Isaacson JS (2005) Synapse-specific downregulation of NMDA receptors by early experience: a critical period for plasticity of sensory input to olfactory cortex. Neuron 47: 101–114.
67. Manschreck TC, Ames D (1984) Neurologic features and psychopathology in schizophrenic disorders. Biol Psychiatry 19: 703–719.
68. Putzhammer A, Klein HE (2006) Quantitative analysis of motor disturbances in schizophrenic patients. Dialogues Clin Neurosci 8: 123–130.
69. Shprintzen RJ (2000) Velo-cardio-facial syndrome: a distinctive behavioral phenotype. Ment Retard Dev Disabil Res Rev 6: 1142–147.
70. Sobin C, Monk SH, Kiley-Brabeck K, Khuri J, Karayiorgou M (2006) Neuromotor deficits in children with the 22q11 deletion syndrome. Mov Disord 21: 2082–2089.
71. Chow EW, Watson M, Young DA, Bassett AS (2006) Neurocognitive profile in 22q11 deletion syndrome and schizophrenia. Schizophr Res 87: 270–278.
72. Nurnberger JI Jr, Blehar MC, Kaufmann CA, York-Cooler C, Simpson SG, Harkavy-, et al. (1994) Diagnostic interview for genetic studies. Rationale, unique features, and training. NIMH Genetics Initiative. Arch Gen Psychiatry 51: 849–859.
73. American Psychiatric Association (1994) DSM-IV: diagnostic and statistical manual of mental disorders, 4th ed. Washington, DC: American Psychiatric Press.
74. Karayiorgou M, Torrington M, Abecasis GR, Pretorius H, Robertson B, et al. (2004) Phenotypic characterization and genealogical tracing in an Afrikaner schizophrenia database. Am J Med Genet B 124: 20–28.
75. Cloninger CR (1989) Schizophrenia: genetic etiological factors. In: Kaplan HI, Saddock BJ, eds. Comprehensive textbook of psychiatry, 5th ed. Baltimore: William & Wilkins.
76. Abecasis GR, Burt RA, Hall D, Bochum S, Doheny KF, et al. (2004) Genomewide scan in families with schizophrenia from the founder population of Afrikaners reveals evidence for linkage and uniparental disomy on chromosome 1. Am J Hum Genet 74: 403–417.
77. Bunting M, Bernstein KE, Greer JM, Capecchi MR, Thomas KR (1999) Targeting genes for self-excision in the germ line. Genes Dev 13: 1524–1528.
78. Gogos JA, Osborne J, Nemes A, Mendelsohn M, Axel R (2000) Genetic ablation and restoration of the olfactory topographic map. Cell 103: 609–620.

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

A10: 115

Molecular Psychiatry (2007) 12, 1129–1139
& 2007 Nature Publishing Group All rights reserved 1359-4184/07 $30.00
www.nature.com/mp

ORIGINAL ARTICLE

LRRTM1 on chromosome 2p12 is a maternally suppressed gene that is associated paternally with handedness and schizophrenia

C Francks[1,2], S Maegawa[3,21], J Laurén[4,21], BS Abrahams[5], A Velayos-Baeza[1], SE Medland[6,7], S Colella[1], M Groszer[1], EZ McAuley[1], TM Caffrey[1], T Timmusk[8], P Pruunsild[8], I Koppel[8], PA Lind[7], N Matsumoto-Itaba[9], J Nicod[1], L Xiong[10], R Joober[11], W Enard[12], B Krinsky[12], E Nanba[3], AJ Richardson[13], BP Riley[5], NG Martin[6], SM Strittmatter[4], H-J Möller[14], D Rujescu[14], D St Clair[15], P Muglia[2], JL Roos[16], SE Fisher[1], R Wade-Martins[1], GA Rouleau[10], JF Stein[13], M Karayiorgou[17], DH Geschwind[5], J Ragoussis[1], KS Kendler[5], MS Airaksinen[18], M Oshimura[9], LE DeLisi[19,20] and AP Monaco[1]

[1]Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK; [2]GlaxoSmithKline SpA, Research Centre, Verona, Italy; [3]Division of Functional Genomics, Research Center for Bioscience and Technology, Tottori University, Yonago, Tottori, Japan; [4]Department of Neurology, Yale University School of Medicine, New Haven, CT, USA; [5]Department of Neurology, David Geffen School of Medicine, University of California at Los Angeles, Los Angeles, CA, USA; [6]Virginia Institute for Psychiatric and Behavioral Genetics, Department of Human Genetics, Virginia Commonwealth University, Richmond, VA, USA; [7]Queensland Institute of Medical Research, Brisbane, QLD, Australia; [8]Department of Gene Technology, Tallinn University of Technology, Tallinn, Estonia; [9]Department of Biomedical Science, Regenerative Medicine and Biofunction, Graduate School of Medical Science, Tottori University, Yonago, Tottori, Japan; [10]Center for the Study of Brain Diseases, CHUM Research Center, Montreal University, Montreal, QC, Canada; [11]Departments of Psychiatry, Neurology and Human Genetics, McGill University, Douglas Hospital Research Centre, Montreal, QC, Canada; [12]Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany; [13]Department of Physiology, Anatomy and Genetics, University of Oxford, Oxford, UK; [14]Department of Psychiatry, Ludwig Maximilians University, Munich, Germany; [15]Institute of Medical Sciences, University of Aberdeen, Aberdeen, Scotland; [16]Weskoppies Hospital, Department of Psychiatry, University of Pretoria, Pretoria, Republic of South Africa; [17]Laboratory of Human Neurogenetics, Rockefeller University, New York, NY, USA; [18]Neuroscience Center, University of Helsinki, Helsinki, Finland; [19]Department of Psychiatry, New York University, New York, NY, USA and [20]The Nathan S Kline Institute for Psychiatric Research, Orangeburg, New York, NY, USA

Left–right asymmetrical brain function underlies much of human cognition, behavior and emotion. Abnormalities of cerebral asymmetry are associated with schizophrenia and other neuropsychiatric disorders. The molecular, developmental and evolutionary origins of human brain asymmetry are unknown. We found significant association of a haplotype upstream of the gene LRRTM1 (Leucine-rich repeat transmembrane neuronal 1) with a quantitative measure of human handedness in a set of dyslexic siblings, when the haplotype was inherited paternally ($P = 0.00002$). While we were unable to find this effect in an epidemiological set of twin-based sibships, we did find that the same haplotype is overtransmitted paternally to individuals with schizophrenia/schizoaffective disorder in a study of 1002 affected families ($P = 0.0014$). We then found direct confirmatory evidence that LRRTM1 is an imprinted gene in humans that shows a variable pattern of maternal downregulation. We also showed that LRRTM1 is expressed during the development of specific forebrain structures, and thus could influence neuronal differentiation and connectivity. This is the first potential genetic influence on human handedness to be identified, and the first putative genetic effect on variability in human brain asymmetry. LRRTM1 is a candidate gene for involvement in several common neurodevelopmental disorders, and may have played a role in human cognitive and behavioral evolution.
Molecular Psychiatry (2007) 12, 1129–1139; doi:10.1038/sj.mp.4002053; published online 31 July 2007

Keywords: handedness; schizophrenia; association; imprinted gene; brain asymmetry

Correspondence: Dr C Francks, Wellcome Trust Centre for Human Genetics, University of Oxford, Roosevelt Drive, Oxford OX3 7BN, UK.
E-mail: clyde.francks@well.ox.ac.uk
[21]These authors contributed equally to this work.
Received 9 August 2006; revised 6 June 2007; accepted 19 June 2007; published online 31 July 2007

Introduction

Left–right asymmetrical function is a conserved feature of vertebrate central nervous systems.[1] Asymmetrical brain function is pronounced in humans, and underlies much of our behavior, cognition and

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

emotion.[2] Approximately 90% of humans are right-handed.[2] This is the strongest population-level bias in handedness for any primate.[1,2] Left-handedness in humans is associated with reductions or reversals of normal brain asymmetries,[3] particularly of cerebral cortical areas related to language perception and production. This association is likely to be partly genetic in etiology.[4] Handedness and complex cognition in humans may, therefore, be related developmentally and evolutionarily, although the nature and extent of these relationships remain uncertain.

We performed previously the first genome-wide linkage scan for a quantitative measure of human handedness[5] (relative hand skill; assessed with timed peg moving[5]), and found linkage on 2p12–q11.[5] This analysis was performed in 222 reading-disabled siblings,[6] although there was no relationship detectable between reading disability (RD) and relative hand skill in this sample.[5,7] We then replicated the 2p12–q11 linkage,[8] with the same measure of relative hand skill, in 105 pairs of left-handed brothers ($P = 0.0009$).[8] We found that the linkage to relative hand skill, in the RD sibs, was derived entirely from paternal inheritance of 2p12–q11 ($P = 0.0000037$),[9] whereas the maternally inherited locus was not linked.[9] This suggested that the underlying gene was imprinted, and inactivated or downregulated on the maternally inherited chromosome.[9] We did not have parental genotypes to assign parent-of-origin for analysis of the left-handed brother sample.[9] However, under standard linkage analysis, the peak location of linkage to relative hand skill in the left-handed brothers was the same as the peak paternal-specific location in the dyslexic siblings,[9] and was within the gene CTNNA2 (a-N-catenin) on 2p12.

Schizophrenia is a heterogeneous neuropsychiatric disorder that affects roughly 1% of the adult human population. Schizophrenia is associated with reductions or reversals of normal cerebral asymmetries, including the medial temporal lobe, superior temporal gyrus, planum temporale and the overall brain anterior–posterior torque.[10,11] Schizophrenia is also associated with a slightly elevated rate of mixed or left-handedness.[12,13] Crow drew attention to the association of abnormal brain asymmetry with schizophrenia-like psychoses,[14] and was drawn to a human-specific rearrangement of the sex chromosomes and a candidate gene there,[15] Protocadherin X-Y. However, this gene has not so far been shown to be linked to asymmetry, language or psychosis.

In a recent meta-analysis of 20 linkage screens for schizophrenia, 2p12–q22 was the only location to reach significance when adjusted for genome-wide testing.[16] This linkage was derived primarily, but not exclusively, from a sample of affected sibling pairs of European descent that were recruited in New York and Oxford.[17] We showed previously that the 2p12–q11 linkage in this sample was derived predominantly from paternal sharing (log of the odds, LOD = 4.72), not maternal sharing (LOD = 0.6).[9] As parent-of-origin effects are unusual in the genome, we proposed in 2003 that a single paternally expressed gene was responsible for the linkages of 2p12–q11 to handedness and schizophrenia.[9]

In the present study we have used genetic association mapping and gene-functional analysis to identify a novel imprinted gene on 2p12, LRRTM1, that we propose is responsible for causing the linkages of this chromosomal region to human handedness and schizophrenia.

Materials and methods

All work involving human samples and materials was approved by the appropriate institutional review boards, and appropriate informed consent was obtained from all human subjects.

Details of the Materials and methods are given in Supplementary information online.

Results

We genotyped 87 single-nucleotide polymorphisms (SNPs) in 222 RD siblings and their parents, within a region of paternal-specific linkage to relative hand skill that we had detected previously in this sample (Figure 1). The SNPs were targeted within four positional and functional candidate genes on 2p12–p11 (LRRTM4, CTNNA2, LRRTM1 and DNAH6) (Figure 1 and Table S1). Following our parent-of-origin linkage data (Figure 1), we tested for quantitative association of paternally inherited SNP alleles with relative hand skill (using QTDT[18]). Four SNPs in distinct locations (rs1517771, rs290015, rs2063436 and rs723524; Table S1), which were not in significant linkage disequilibrium (LD) with one another, showed nominally significant paternal-specific association with relative hand skill ($0.05 > P > 0.01$). This was no more than the expected false-positive rate.

We then aimed to interrogate the underlying haplotype structure by testing haplotypes for association, that were constructed in 'sliding windows' representing all sets of three consecutive SNPs (disregarding any instances where SNPs were not in significant LD, that is, $P > 0.05$ based on w^2-test of pairwise LD). We used the total, paternal and multiallelic options of QTDT together for the association tests. Significance levels for paternal-specific association were all $P > 0.05$, except for haplotypes derived from the SNPs rs1446109-rs1007371-rs723524 ($P = 0.00002$), and the overlapping window rs1007371-rs723524-rs1025947 ($P = 0.0007$). These SNPs are on 2p12 within a region of strong inter-marker LD that spans at least the first exon and 137 kb upstream of the gene LRRTM1, which is located within an intron of CTNNA2 (Figure 1). There was no significant maternal-specific haplotype association at any location (all $P > 0.05$).

The rs1446109-rs1007371-rs723524 haplotype 2-2-2 (where 2 is the minor allele of each SNP) was responsible primarily for the paternal-specific association observed with haplotypes derived from these

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

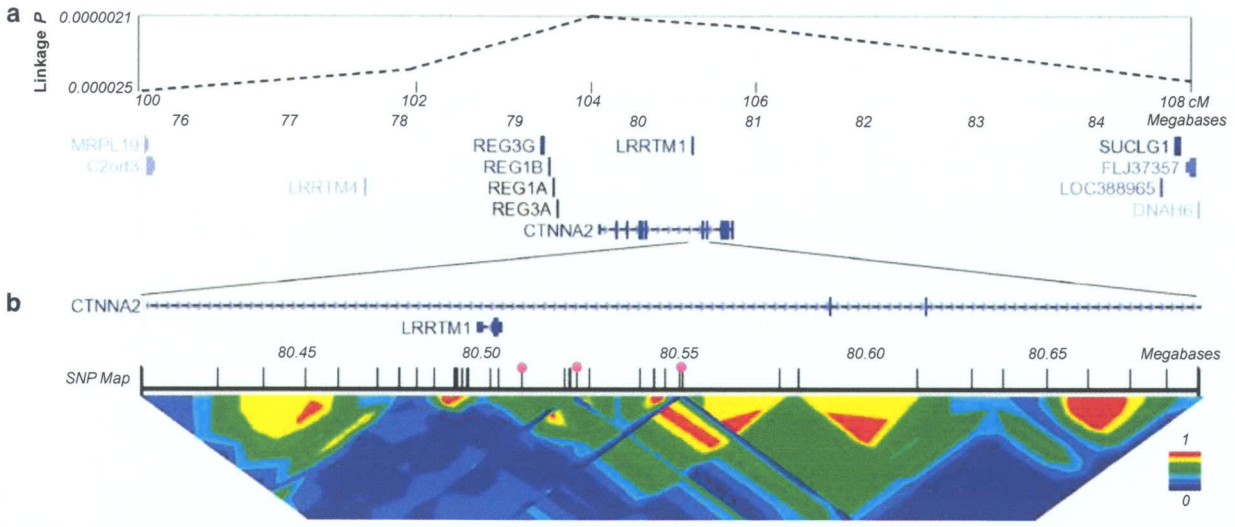


Figure 1 Mapping of an imprinted quantitative-trait-locus (QTL) for human handedness. (a) Top: pointwise significance of linkage derived from paternal sharing across 8 cM of chromosome 2p12-p11.[9] This corresponds roughly to a 1-LOD unit support interval for the QTL. Significance of linkage is shown on a logarithmic scale. Bottom: known genes within the region (adapted from UCSC Genome Browser). Note LRRTM1 (transcribed proximal to distal (right to left)) sited within an intron of CTNNA2 (transcribed distal to proximal). (b) Close up of 286 kb around LRRTM1. Two exons of CTNNA2 are also visible. The final SNP map that we used in this region is shown (see also Table S1), and the three SNPs that showed the initial paternal-specific haplotype associations are highlighted by pink circles (rs1446109-rs1007371-rs723524). Pairwise intermarker LD (Cramer's V) is shown at the bottom.

three SNPs. The 2-2-2 haplotype had 9% frequency, and was associated with a mean shift of 1.1 s.d. toward left-handedness in the relative hand skill distribution[5,7] when inherited paternally, compared to all other haplotypes. Genotyping of 32 more SNPs, including 12 within 137 kb upstream of LRRTM1, confirmed that rs1446109-rs1007371-rs723524 haplotype 2-2-2 represents a distinct haplotype clade for at least 76 kb upstream of, and including, the predicted promoter of LRRTM1 (Figure S1 and Table S1).

We screened the first two exons and predicted promoter of LRRTM1 for polymorphisms in 26 left-handers from the RD sample by denaturing high-performance liquid chromatography and sequencing, but we did not detect any polymorphisms that tagged rs1446109-rs1007371-rs723524 haplotype 2-2-2, or that had overt disruptive effects on the predicted LRRTM1 protein (entirely coded within exon 2). The only novel SNP that we identified was upstream of the first LRRTM1 exon (Table S1).

We analyzed the SNPs rs1007371 and rs723524, which together can be used to construct a close proxy to the 'risk' haplotype described above (see Figure S1), in a sample of normal twin-based sibships from Brisbane, Australia, that was derived from 215 independent families. We found no significant evidence for paternal association of the risk haplotype with our quantitative measure of human handedness ($P > 0.1$).

We genotyped the SNPs rs1446109 and either rs723524 or rs718466 (the latter are equivalent tagging SNPs according to international HapMap data) in four family samples of white European descent that included individuals with schizophrenia or poor-outcome schizoaffective disorder. These comprised a subset of the New York/Oxford sample[17] (226 families), an Irish 'High Density' sample[19] (236 families), a sample collected in Montreal[20] (124 families) and an Afrikaner sample[21] (416 families) (Table 1). The 2-2 haplotype defined by the rare alleles of these SNPs is equivalent to the risk haplotype for left-handedness defined above (Figure S1). The 2-2 haplotype varied in frequency between 7.6 and 12.1% in the four sample sets. We found that haplotype 2-2 was overtransmitted paternally to affected individuals in a combined analysis of the four samples, $P = 0.0014$ (one-tailed test, 38 transmissions to 16 nontransmissions; Table 1). This was a specific hypothesis test that requires no statistical adjustment. There was no significant paternal overtransmission of any other haplotype, nor was there maternal overtransmission of any haplotype. The paternal 2-2 result was derived roughly equally from three of the four samples (Table 1).

We genotyped rs1446109 (almost tagging for the risk haplotype; Figure S1), in two case–control collections of European descent (461 cases and 459 controls from Munich, Germany, and 429 cases and 428 controls from Scotland (all cases had Diagnostic and Statistical Manual of Mental Disorders, fourth edition diagnoses of schizophrenia, and both sample sets were recruited according to the same protocol)).[22] As the parental origin of the alleles could not be established and the paternal and maternal alleles

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Table 1 Transmission disequilibrium statistics for two-marker haplotypes (see main text), for families affected with schizophrenia/schizoaffective disorder

			Paternal			Maternal			Both parents		
	N	%	TR	NT	w^2	TR	NT	w^2	TR	NT	w^2
New York/Oxford											
Haplo.1	504	81.8	18	27	1.80	21	19	0.10	40	47	0.56
Haplo.2	50	8.1	15	6	3.86	8	9	0.06	23	15	1.68
Haplo.3	61	9.9	13	13	0.00	15	14	0.03	29	28	0.02
Haplo.4	1	0.2	0	0	0.00	0	2	2.00	0	2	2.0
Irish high density											
Haplo.1	336	79.6	6	14	3.20	15	17	0.12	23	33	1.79
Haplo.2	51	12.1	10	1	7.36	11	6	1.47	22	8	6.53
Haplo.3	30	7.1	5	5	0.00	5	8	0.69	11	14	0.36
Haplo.4	5	1.2	0	1	1.00	1	1	0.00	1	2	0.33
Montreal											
Haplo.1	327	86.5	4	4	0.00	11	9	0.20	15	13	0.14
Haplo.2	31	8.2	3	3	0.00	5	12	2.88	8	15	2.13
Haplo.3	19	5.0	2	2	0.00	7	2	2.78	9	4	1.92
Haplo.4	1	0.3	0	0	0.00	0	0	0.00	0	0	0.00
Afrikaner											
Haplo.1	1013	81.0	16	19	0.26	21	18	0.23	39	39	0.00
Haplo.2	95	7.6	10	6	1.00	4	14	5.56	15	21	1.00
Haplo.3	118	9.4	11	9	0.20	15	8	2.13	27	18	1.80
Haplo.4	24	1.9	2	5	1.29	3	3	0.00	5	8	0.69
Total sample											
Haplo.1	2180	81.8	44	64	3.70	68	63	0.19	117	132	0.90
Haplo.2	227	8.5	38	16	8.96	28	41	2.45	68	59	0.64
Haplo.3	228	8.6	31	29	0.07	42	32	1.35	76	64	1.03
Haplo.4	31	1.2	2	6	2.00	4	6	0.40	6	12	2.00

w^2, w^2 statistic; haplo.1, haplotype 1-1; haplo.2, haplotype 2-2; haplo.3, haplotype 1-2; haplo.4, haplotype 2-1; N, number of instances of the haplotype in the sample; NT, number of non-transmitted haplotypes from heterozygous parents; TR, number of transmitted haplotypes from heterozygous parents; %, haplotype frequency.
The left-handedness risk haplotype 2-2 is highlighted in bold.

were confounded, we expected this analysis to be low-powered to detect an imprinted effect. Nonetheless, this SNP showed a trend toward association (P = 0.09) when tested under a genotypic 2-df logistic regression model using covariates for gender and the collection site (Munich/Aberdeen). The additive component of this model was significant at P = 0.036, and the direction of allelic effect was the same as in the family samples. When we repeated the analysis using only those cases (151) who reported a positive family history of schizophrenia or bipolar disorder, with the aim of removing sporadic environmental cases, rs1446109 showed significant association with P = 0.013, and the additive component of the model showed P = 0.0038, again in the expected direction. We also tested for association in a sample of 270 Han Chinese families[23] but we found no significant bias in paternal or maternal transmission of any haplotype to schizophrenic people (there were 65 paternal transmissions to 78 paternal nontransmissions of the 2-2 haplotype).

No imprinted genes were known previously on chromosome 2p. We found that human LRRTM1 is imprinted (paternal-only expression) in hybrid A9 cells[24] (mouse cell lines containing single human chromosomes of known parental origin) (Figure 2). Despite certain limitations with hybrid cells in the

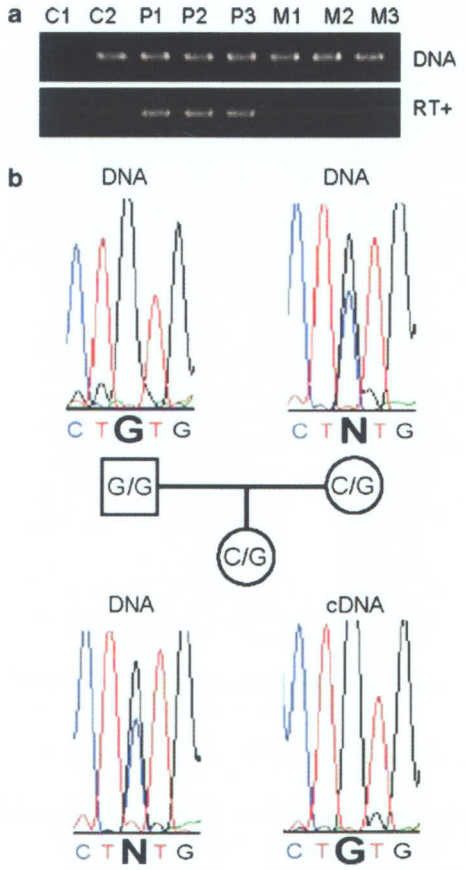


Figure 2 LRRTM1 expression is downregulated maternally in humans. (a) Data are shown from A9 cells that each contains a single human chromosome 2 of known parental origin. Products are shown from PCR using primers specific for human LRRTM1. Upper panel: the human gene is present in genomic DNA (labeled 'DNA') from all cell lines tested, apart from C1 (mouse A9 cell line containing no human chromosome). Lower panel: the human gene is only expressed (detected by RT-PCR) in three cell lines containing paternally derived human chromosome 2s (P1, P2, P3), and not in three cell lines containing maternally derived human chromosome 2s (M1, M2, M3). (C2 is a human fibroblast cell line). (b) Monoallelic paternal expression of LRRTM1 in a human EBV-transformed lymphoblastoid cell line. Sequence traces derived from genomic DNA (labeled 'DNA') surrounding a C/G SNP are shown for a father, mother and child. The child is heterozygous, but the cDNA prepared from mRNA showed expression of only the paternally inherited G allele of LRRTM1.

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

context of imprinting studies,[25] mouse A9 cells have been used for the reliable identification and verification of human imprinted genes.[24] To validate the chromosome 2 hybrid cells, we analyzed the expression of four additional biallelically expressed genes, GGCX (2p11.2), BCL2L11 (2q13), TCF7L1 (2p11.2) and RALB (2q14.2). We detected the expression of all four genes in chromosome 2 hybrid cells, regardless of the parental origin of the human chromosome 2 (Figure S2). No detection of these genes was obtained in RT- controls. We also used transcribed polymorphisms to show mono-allelic LRRTM1 expression in tissue samples from a minority of unrelated post-mortem human brains (3 out of 18) by RT-PCR and sequencing or restriction digestion (not shown). However, biallelic expression was found in 15 out of 18 brains, and LRRTM1 was expressed at similar levels from both alleles in the adult human cerebral cortex of five individuals showing biallelic expression (analyzed by quantitative PCR). All brain tissue was obtained from normal controls that were unaffected by major neurological or psychiatric disease. These results suggest that imprinting of LRRTM1 is either variable between individuals (as for the imprinted genes IMPT1 or IGF2[26]), and/or variable between different brain regions or cell/tissue types (as for UBE3A[27]). We also found mono-allelic paternal expression of LRRTM1 in four out of four unrelated EBV-transformed human lymphoblastoid cell lines that were heterozygous for a transcribed SNP (Figure 2).

By use of in situ hybridization in the mouse, we found that Lrrtm1 is expressed predominantly in the nervous system by postmitotic neurons, but also in some nonneuronal tissues (Figure 3; Figure S3). Expression is upregulated in the mouse brain during embryonic development and early postnatally. In adult brain, Lrrtm1 expression is most prominent in the forebrain, particularly in the thalamus (in most or all nuclei), and in cortical areas including hippocampus, piriform and posterior cingulate (Figure 3; Figure S3).

In northern blot analysis of adult human brain, LRRTM1 also showed predominant expression in forebrain regions including thalamus and cerebral cortex (Figure S4). By in situ hybridization in coronal sections of the post-mortem developing human brain (14–16 weeks gestation; Figure 4), strong expression was observed in anterior sections throughout the cortical plate and in septum, caudate and putamen. The absence of signal in the subventricular zone argues against a direct involvement in neurogenesis (Figure 4). Transcript distribution was similar in more caudal sections with the addition of signal in dorsolateral thalamus (Figure 4). More caudal still, thalamic signal shifted ventrally to a structure consistent with the lateral geniculate body (Figure 4). The striking signal in caudate and putamen in human (Figure 4) was not present in mouse, at least not at E15 and in adult. In addition, expression within the human thalamus was more restricted as compared to the mouse, with staining relatively limited to dorsomedial regions.

No consistent asymmetric expression was observed in any of three human developing brains examined by in situ hybridization (14–16 weeks' gestation), regardless of whether cerebral hemispheres were analyzed in aggregate or cortical sub-regions were examined in isolation (dorsolateral, temporal, ventrolateral or cingulated). Similarly to embryonic brain, LRRTM1 was expressed at similar levels (that is, symmetrically) in all analyzed regions of left and right adult human cerebral cortex (several different cortical regions from five individuals were analyzed by quantitative PCR). We also quantified left- and right-brain Lrrtm1 mRNA expression levels in rats and embryonic mice (see Materials and methods), but did not detect evidence for asymmetrical expression in rodents. In addition, we tested cerebral cortex, cerebellum, brain stem, olfactory bulb, thymus, heart, lung, liver, intestine, pancreas, spleen, kidney, muscle and testis of two reciprocal crossed F1 mice between C57BL/6J and JF1 strains for allele-specific expression. Each of the F1 mice was 30 weeks old. Expression of Lrrtm1 was detected in cerebral cortex, cerebellum and brain stem, but it was biallelic (Figure S5).

In rodent primary sensory (DRG) and cortical neurons (Figure 3), and in cerebellar granular neurons (data not shown), overexpressed LRRTM1 localized to the cell soma, neurites and lamellipodia of growth cones, suggesting a function in axon guidance and/or synaptogenesis. Unexpectedly, in transfected MRC5, Cos-7 and Neuro-2a cells, LRRTM1 colocalized with endoplasmic reticulum (ER) markers (Figure 5 and Figure S6). Live-cell staining for overexpressed LRRTM1 in DRG neurons revealed that the protein is not accessible on the plasma membrane under conditions that allowed surface detection of a related member of the LRR protein superfamily, Lingo1 (Figure 5). These results suggest that endogenous LRRTM1 may have a role in intracellular trafficking within axons. However, it remains possible that LRRTM1 is localized to plasma membrane in cells expressing an unidentified LRRTM1 chaperone or coreceptor protein that promotes its processing and/or transport.

We analyzed methylation within 2 CpG islands that correspond to the predicted promoter and coding exon of LRRTM1, and a third island roughly 18 kb upstream of LRRTM1, in 17 lymphoblastoid cell lines and 17 human post-mortem brain samples, but we did not find evidence that these CpG islands are differentially methylated regions (DMRs) (data not shown).

We tested CTNNA2 for mono-allelic expression in 10 post-mortem brain samples from normal, unrelated individuals, by use of a transcribed SNP in exon 12, but expression was always biallelic, in contrast to LRRTM1 (data not shown). We did not detect CTNNA2 expression in the A9 hybrids, or in human lymphoblastoid cell lines, by RT-PCR. The first CTNNA2 exon is over 750 kb from LRRTM1 (Figure 1), and several LD blocks away, further

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

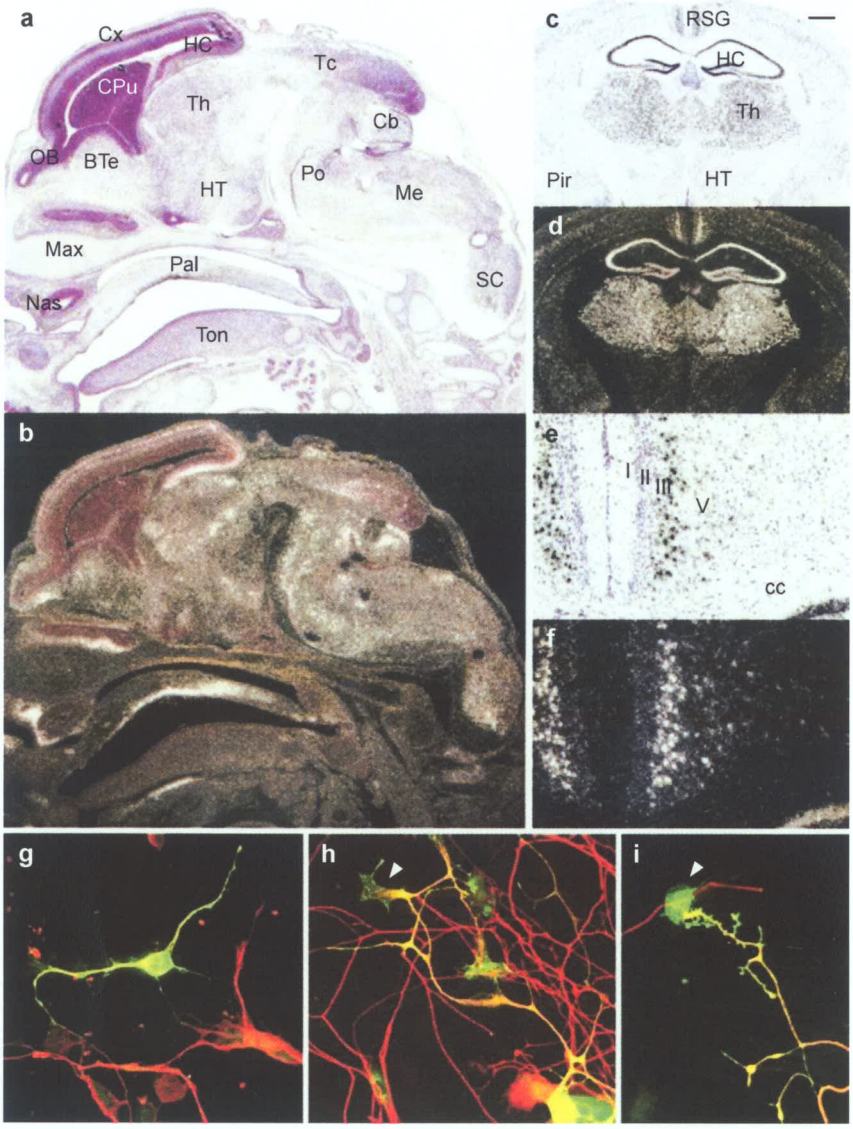


Figure 3 Developmental increase of Lrrtm1 mRNA expression in mouse thalamus, hippocampus and retrosplenial cortex, and localization of LRRTM1 in neurons. In situ hybridization analysis in sagittal sections of E15 mouse embryos (a and b) and coronal sections of adult mouse brain (c–f). Bright-field images counterstained with hematoxylin (a, c and e) and corresponding dark-field autoradiographs are shown (b, d and f). The white signal in the dark-field images indicates Lrrtm1 expression. (e and f) Higher magnification of the retrosplenial granular cortex (boxed area in d) shows a high level of Lrrtm1 expression in layer III–IV neurons. (g–i) Myc-LRRTM1 was transfected into E17 mouse cortical neurons (g) and nucleofected into rat P18 dorsal root ganglion neurons (h and i) and detected 40 h later with anti-myc antibody. Neurons were visualized with anti-b-tubulin antibody. LRRTM1 is localized to the cell soma and in the neurites; in neurites it is also localized to lamellipodia of the growth cones (arrowheads in h and i). BTe, basal telencephalon; Cb, cerebellum; cc, corpus callosum; CPu, caudate-putamen; Cx, cerebral cortex; HC, hippocampus; HT, hypothalamus; Max, maxilla, Me, medulla; Nas, nasal cavity; OB, olfactory bulb; Pal, palatine; Pir, piriform cortex; Po, pons; RSG, retrosplenial granular cortex; Sc, Spinal cord; Tc, tectum; Th, thalamus; Ton, tongue. Scale bar is 500 mm in (a–d), 100 mm in (e–f) and 20 mm in (g–i).

suggesting that the regulation of these genes is distinct. There are no other genes within 1.13 Mb of LRRTM1, and the next closest gene expressed in the forebrain (expression data available via the UCSC genome browser) is over 2.75 Mb distally (LRRTM4; Figure 1).

In order to analyze the evolution of LRRTM1, we compared it between humans and chimpanzees. There is no fixed amino acid difference between human LRRTM1 and chimpanzee LRRTM1, and it is not among the genes that are differently expressed between adult humans and adult chimpanzees in

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

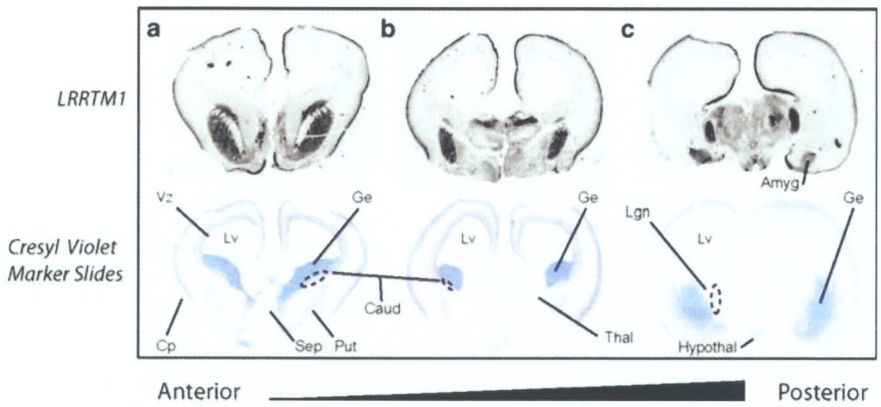


Figure 4 In situ hybridization analysis of LRRTM1 expression in the developing human brain at 15 weeks' gestation. In coronal sections of anterior brain (a), expression is strong throughout the cortical plate and otherwise restricted to septum, caudate, and putamen. Transcript distribution is similar in more caudal sections (b) with the addition of signal in dorsolateral thalamus. More caudal still (c), thalamic signal shifts ventrally to a structure consistent with the lateral geniculate body. Abbreviations are as follows: Amyg (Amygdala), Caud (Caudate), Cp (Cortical), Ge (Ganglionic Eminence), Hypothal (Hypothalamus), Lgn (Lateral Geniculate Nucleus), Lv (Lateral Ventricle), Thal (Thalamus), Vz (Ventricular Zone).

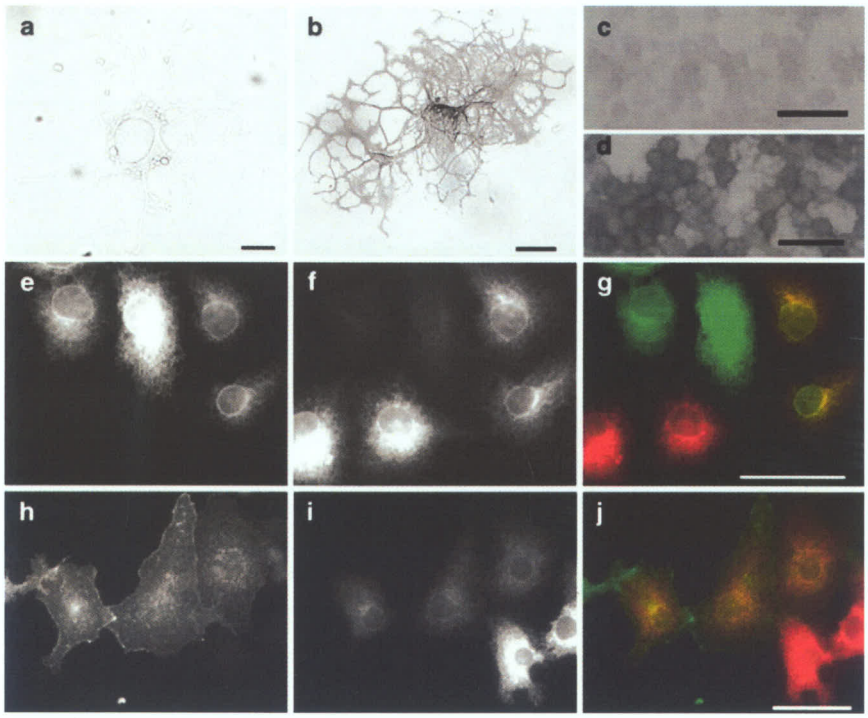


Figure 5 LRRTM1 is not localized on the plasma membrane. No cell-surface expression of LRRTM1 is seen in primary sensory neurons electroporated with myc-LRRTM1 construct and detected by live-cell staining with alkaline phosphatase-conjugated anti-myc antibody (a); strong cell-surface staining of myc-Lingo1 expressing neuron is seen (b). Live-cell staining resulted in 20–30 strongly stained DRG neurons per well for myc-Lingo1 but none for myc-LRRTM1. Similarly, in neuro-2a neuroblastoma cells myc-LRRTM1 (c) is not, but myc-lingo1 is (d) localized to plasma membrane as detected by live-cell staining. In permeabilized Cos-7 cells LRRTM1 (anti-myc antibody; e) is colocalized (merge picture; g) with ER-marked dsRED2-ER (f), whereas lingo1 (anti-myc antibody; h) is only partially colocalized with dsRED2-ER, and it is partially localized on the plasma membrane (i; merge picture: j). Scale bars are 50 mm.

brain, liver, kidney, heart or testis.[28] Furthermore, biallelic expression of LRRTM1 was detected in two out of two chimpanzee brain samples (see Materials and methods).

Discussion

Crow proposed that handedness, brain asymmetry and schizophrenia share an underlying genetic

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

A10: 122

relationship.[14] However, no genes have previously been shown to associate across these traits. We carried out SNP-based association screening for human relative hand skill, under a parent-of-origin model, within four candidate genes close to or within 6 mb of a peak of paternal-specific linkage on 2p12, in a sample of sib pairs recruited for the presence of RD. We found strong paternal-specific association with relative hand skill implicating the gene LRRTM1 (leucine-rich repeat transmembrane neuronal 1) and a region of 80 kb upstream of this gene (which is located within an intron of CTNNA2). We did not find this association in a sample of normal twin-based sibships from Australia. However, consistent with our parent-of-origin association data, we went on to show that LRRTM1 is downregulated on the maternally inherited chromosome, and is therefore a newly identified imprinted locus in the human genome. We also found that the same haplotype that drove the paternal association with handedness was overtransmitted paternally, but not maternally, to schizophrenic/schizoaffective individuals in 1002 affected families of European descent ($P = 0.0014$). This was a specific hypothesis test that requires no statistical adjustment, and indicates that LRRTM1 is a novel susceptibility factor for schizophrenia. We also found consistent evidence for association in a large case–control sample of European descent that was collected in Scotland and Germany.

We did not find that the 'risk' 2-2 haplotype that we identified in European family-based samples was associated with schizophrenia in a Han Chinese family-based sample. The frequency of the 2-2 risk haplotype was somewhat higher in the Han Chinese, at 18%. The failure to replicate may nonetheless be related to modest power in the Chinese sample of 270 families. Alternatively, the nonreplication in the Chinese may indicate that the 2-2 haplotype does not carry risk variation in this population (it may also be relevant that native Chinese speakers show some morphological differences to English speakers in language-related areas of the brain[29]). A further study in a larger Chinese sample will be required to distinguish these possibilities, together with identification of the functional variants in European populations.

Of a total of approximately 35 000 human genes, only roughly 40 were previously known to be imprinted, and the total number is unlikely to exceed this greatly.[30–32] The finding that LRRTM1 is imprinted, and maternally suppressed, was a confirmation of the parent-of-origin genetic data that led to this locus, because the a priori probability that LRRTM1 would be imprinted was very low (roughly 1 in 1000). Our association data suggest that allelic variation upstream of LRRTM1, that may affect the gene's expression, is relevant functionally for brain asymmetry when inherited on the active paternal chromosome, but not on the relatively inactive maternal chromosome, at certain locations and time points during human brain development. However, the 'risk' haplotype that we identified at LRRTM1 cannot explain the majority of the paternal-specific linkage across the genomic region of 2p12 (Figure 1), and therefore additional genetic or epigenetic variability at LRRTM1, or neighboring genes, may also be responsible partly for the linkages of handedness and schizophrenia to 2p12.[9] Although we found evidence that LRRTM1 is variably imprinted in the post-mortem adult human brain (as is the imprinted gene 5HT2A),[33] we cannot conclude directly, on the basis of our data, that variability in imprinting of LRRTM1 is involved in individual differences in human handedness and schizophrenia liability. Our data indicate a role for the 2-2 haplotype in handedness variability and susceptibility to schizophrenia, in European populations, while the data on imprinting are consistent with the paternal-specific nature of this effect.

Our data suggest that a subtype of schizophrenia, linked to misregulation of human LRRTM1, may have its origins in fetal neurodevelopment. Since LRRTM1 appears to underlie the strongest linkage to schizophrenia in the genome, as identified by a meta-analysis of 20 genomewide linkage scans,[16] it is possible that LRRTM1 dysfunction causes a major, common subtype of schizophrenia. Assessing the frequency of this subtype will require studies in further clinical and epidemiological samples, together with a better definition of the functional genetic and epigenetic variation at the LRRTM1 locus. This information may have a substantial impact on pharmacogenetic studies and the development of new treatments for schizophrenia, by allowing patient heterogeneity to be accounted for in clinical trial studies. The receptor-like structure of LRRTM1 also suggests that it may be a drug-tractable target in its own right.

LRRTM1 is one of a four-member family of type I transmembrane proteins containing leucine-rich repeat (LRR) domains, which are commonly involved in protein–protein interactions (LRRs are present in the Slits and Nogo-receptor,[34,35] involved in axonal pathfinding). Each LRRTM member has a specific, dynamically regulated regional brain-expression distribution.[35] Using several gene-functional approaches, we found that LRRTM1 is likely to play a role during the development of specific forebrain structures by influencing neuronal differentiation and connectivity, with a possible role in intracellular trafficking in axons. Thus, LRRTM1 is an ideal candidate gene for having an involvement in subtle developmental abnormalities of the central nervous system.

When comparing expression between human and mouse, it seems that there is a striking absence of signal in caudate and putamen in mouse, at least at e15 and in adult. Also, expression within the fetal human thalamus is more restricted as compared to the mouse, with staining relatively limited to dorsomedial regions that have been implicated in schizophrenia.[36,37] We found no evidence for overtly asymmetrical expression of Lrrtm1 in the developing

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

or adult brain of rodents or humans, using in situ hybridization and/or quantitative PCR. However, we cannot rule out a subtle asymmetry of function or expression at some restricted time point during human brain development, as for the transcription factor LMO4.[38] It remains possible that LRRTM1 is important in the establishment, consolidation or elaboration of the left–right axis during human brain growth and development, particularly before 14 weeks' gestation, which was the earliest stage that we were able to analyze in human. (Population-level morphological asymmetries of the cerebral cortex are already noticeable shortly after this time, and are visible by ultrasound in normal fetuses at 20–22 weeks' gestation.[39]) Therefore, detailed studies of the roles of LRRTM1 and the pathways in which it functions in mammalian brain development are warranted, particularly at developmental time points earlier than 14 weeks in humans, that may reveal critical new insights into the establishment and/or maintenance of normal and abnormal human brain function and asymmetry.

Humans have the strongest population-level bias in handedness of any primate,[1,2] and LRRTM1 is a candidate for having had a role in the evolution of this trait. As the human and mouse predicted LRRTM1 proteins are 96% identical, and the human and chimpanzee proteins are 100% identical, any potential human-specific properties of LRRTM1 may involve the spatiotemporal control of its imprinted expression. (It is interesting that recent analysis of conserved noncoding elements demonstrated that a sequence element 130 kb downstream of LRRTM1 displays accelerated evolution in the human lineage, which may be important for the gene's regulation).[40] We detected only biallelic, nonimprinted expression of Lrrtm1 in cerebral cortex, cerebellum, and brain stem from two 30-week-old reciprocal crossed F1 mice between C57BL/6J and JF1 strains (data not shown), and only biallelic expression in the post-mortem brains of two adult chimpanzees. Larger sample sizes will be needed to test whether imprinting of LRRTM1 is found only in humans. (There is a high level of discordance in imprinting between humans and mice;[32] the gene DLX5 is an example.[41]) It will also be interesting to compare the regulation of LRRTM1 longitudinally during the development of humans and other species, as the imprinted regulation may be restricted to certain developmental periods.[38] In addition, human genetic variation at this locus in nonclinical populations could be analyzed with dense marker genotyping, to test for evidence of selection, particularly with regard to the risk haplotype.

Genomic imprinting can arise when the optimal level of maternal resource investment in offspring differs between the two parental sexes, in polygamous mating systems.[31,42] Paternally inherited alleles of imprinted genes often sequester more maternal resources than do maternally inherited alleles, especially when influencing growth in utero.[31,42] The imprinting of LRRTM1 in humans, therefore, suggests an intriguing evolutionary scenario, in which the parental sexes have conflicting interests in relation to the outcome of lateralized brain development in their offspring, which underlies much of human cognition and behavior. Further study of the neural systems and behaviors that LRRTM1 influences will be required to understand the selective forces that drove LRRTM1 to be imprinted.

We did not detect evidence for differential methylation of the paternal and maternal alleles at the CpG islands associated directly with LRRTM1, and therefore the mechanism underlying imprinted expression of LRRTM1 remains unclear. However, many other imprinted genes are not regulated by differential methylation within their own promoters, but are regulated instead by imprinted control centers which may be located many kilobases distant.[31] Also, allele-specific histone modifications are emerging currently as equally important determinants of imprinted gene expression.[43] The lack of methylation does not, therefore, refute the imprinted expression of LRRTM1, which we observed in three different tissues and experimental settings. We recommend that the mechanism of imprinted regulation at this locus should be investigated as a priority, as epigenetic misregulation of LRRTM1 may have clinical relevance and could underlie much of the paternal linkage of 2p12 to schizophrenia.

We analyzed schizophrenia and schizoaffective disorder jointly in the New York/Oxford sample to maximize the numbers of informative paternal transmissions for the risk haplotype (Table 1). However, this sample was not sufficiently large to test for heterogeneity of the association effect between these two patient groups. Our previous linkage analysis[9] of 2p12 had indicated that the linkage was stronger for the subset of sibling pairs with schizophrenia only (paternal LOD 4.72), but this remains to be investigated more thoroughly by targeted association analysis in additional samples.

LRRTM1 is the first genetic influence on human handedness to be identified. Handedness, brain asymmetry and schizophrenia are likely to be etiologically complex traits with several, or many, genetic and environmental influences. As we detected an effect of LRRTM1 on handedness in dyslexic siblings, but not in normal twin-based sibships, it is possible that LRRTM1 influences behavioral lateralization more strongly in clinically selected populations for psychiatric/neurodevelopmental dysfunctions, than in the general population. This would suggest an interactive, nonadditive effect of LRRTM1 with other genetic and/or environmental effects that predispose to neuropsychiatric dysfunction, such that the effect of LRRTM1 depends on other risk factors in order to manifest fully. However, it is also possible that the failure to replicate the association with handedness was due to low power and a relatively small replication sample in our study. It is a well-described

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

A10: 124

statistical feature of genetic association analysis that follow-up samples should usually be much larger than initial samples to provide adequate power, as the genetic effect in the initial sample is typically inflated.[44] Unfortunately, we are unaware of other family-based collections with accurate quantitative assessments of handedness, that are suitable for genetic studies.

In addition to schizophrenia, other neurodevelopmental disorders including bipolar disorder, autism and language impairment have shown evidence for associations with left-handedness and/or abnormal asymmetrical brain structure/function.[2,10,11,45–49] LRRTM1 is, therefore, also a candidate for involvement in these traits. The LRRTMs are a four-member gene family first described in 2003, and it is interesting to note that LRRTM3 has recently been proposed as a susceptibility factor for late-onset Alzheimer's disease.[50] We recommend, therefore, that the whole LRRTM gene family is investigated in relation to psychiatric and neurological disorders.

Acknowledgments

We thank the families who took part in this study; David Smith, Catherine Joachim and OPTIMA (Oxford Project to Investigate Memory and Ageing) for access to post-mortem brain tissue; Xiayi Ke for testing his screening algorithm for imprinted genes; Phil Burnet for plasmids and samples; Dianne Gerelli and the Human Developmental Brain Resource (London, UK) for pilot in situ hybridization studies; Jordana Tsenova for ideas on genetic analysis; Eris Duro for help with mutation screening; Abigail Woodroffe for handling genotype data; Julie V Perederiy for help with in situ experiments; Drs Dermot Walsh and F Anthony O'Neill made critical contributions to the collection of the Irish Study of High Density Schizophrenia Families (ISHDSF). Drs Enn Jõeste and Liina Kiho from North Estonian Regional Hospital, Tallinn, helped in collecting human brain tissue. David Goldstein, Kevin Shianna and Dongliang Ge provided genotype data. Han Chinese Schizophrenia Sample data and biomaterials were collected as a part of the National Institute of Mental Health (NIMH) Schizophrenia Genetics Initiative. A complete list of the people who contributed on the NIMH project is in the Supplementary material. CF was a NARSAD Young Investigator (National Alliance for Research on Schizophrenia and Depression). APM is a Wellcome Principal Research Fellow. RW-M is a Wellcome Trust Research Career Development Fellow. SEF is a Royal Society Research Fellow. TT is a Wellcome Trust International Senior Research Fellow in Biomedical Science in Central Europe. JN is supported by the Swiss National Science Foundation and by a Marie Curie Intra-European Fellowship. BSA was supported by a Fellowship from the Tourette Syndrome Association. This research was funded additionally by The Wellcome Trust (UK); the Schizophrenia Research Fund (UK); the Academy of Finland and the Sigrid Juselius Foundation (Finland); the Graham Boeckh Chair Program in Schizophrenia to GAR at McGill University; Warner-Lambert, Parke-Davis Pharmaceuticals Company, and National Institute of Mental Health grant R01 MH-44245 (LED); National Institute of Health grants MH41953 (collection and analysis of the ISHDSF) and MH61399 (MK); and by grant 'Research on Psychiatric and Neurological disorders and Mental Health' from the ministry of Health, Labor and Welfare of Japan (H14-Kokoro-002), and Grant-in-Aid for Scientific Research on Priority Areas (C) Medical Genome Science from the Ministry of Education, Culture, Sports, Science and Technology of Japan. Clyde Francks conceived and directed this research.

References

1 Rogers LJ, Andrew R (eds). Comparative Vertebrate Lateralization. Cambridge University Press: Cambridge, UK, 2002.
2 Hugdahl K, Davidson RJ (eds). The Asymmetrical Brain. MIT Press: Cambridge, MA, 2003.
3 Mevorach C, Humphreys GW, Shalev L. Attending to local form while ignoring global aspects depends on handedness: evidence from TMS. Nat Neurosci 2005; 8: 276–277.
4 Geschwind DH, Miller BL, DeCarli C, Carmelli D. Heritability of lobar brain volumes in twins supports genetic models of cerebral laterality and handedness. Proc Natl Acad Sci USA 2002; 99: 3176–3181.
5 Francks C, Fisher SE, MacPhie IL, Richardson AJ, Marlow AJ, Stein JF et al. A genomewide linkage screen for relative hand skill in sibling pairs. Am J Hum Genet 2002; 70: 800–805.
6 Fisher SE, Francks C, Marlow AJ, MacPhie IL, Newbury DF, Cardon LR et al. Independent genome-wide scans identify a chromosome 18 quantitative-trait locus influencing dyslexia. Nat Genet 2002; 30: 86–91.
7 Francks C, Fisher SE, Marlow AJ, MacPhie IL, Taylor KE, Richardson AJ et al. Familial and genetic effects on motor coordination, laterality, and reading-related cognition. Am J Psychiatry 2003; 160: 1970–1977.
8 Francks C, DeLisi LE, Fisher SE, Laval SH, Rue JE, Stein JF et al. Confirmatory evidence for linkage of relative hand skill to 2p12-q11. Am J Hum Genet 2003; 72: 499–502.
9 Francks C, DeLisi LE, Shaw SH, Fisher SE, Richardson AJ, Stein JF et al. Parent-of-origin effects on handedness and schizophrenia susceptibility on chromosome 2p12-q11. Hum Mol Genet 2003; 12: 3225–3230.
10 DeLisi LE, Sakuma M, Kushner M, Finer DL, Hoff AL, Crow TJ. Anomalous cerebral asymmetry and language processing in schizophrenia. Schizophr Bull 1997; 23: 255–271.
11 Shenton ME, Dickey CC, Frumin M, McCarley RW. A review of MRI findings in schizophrenia. Schizophr Res 2001; 49: 1–52.
12 DeLisi LE, Svetina C, Razi K, Shields G, Wellman N, Crow TJ. Hand preference and hand skill in families with schizophrenia. Laterality 2002; 7: 321–332.
13 Orr KG, Cannon M, Gilvarry CM, Jones PB, Murray RM. Schizophrenic patients and their first-degree relatives show an excess of mixed-handedness. Schizophr Res 1999; 39: 167–176.
14 Berlim MT, Mattevi BS, Belmonte-de-Abreu P, Crow TJ. The etiology of schizophrenia and the origin of language: overview of a theory. Compr Psychiatry 2003; 44: 7–14.
15 Williams NA, Close JP, Giouzeli M, Crow TJ. Accelerated evolution of Protocadherin11X/Y: a candidate gene-pair for cerebral asymmetry and language. Am J Med Genet B Neuropsychiatr Genet 2006; 141: 623–633.
16 Lewis CM, Levinson DF, Wise LH, DeLisi LE, Straub RE, Hovatta I et al. Genome scan meta-analysis of schizophrenia and bipolar disorder, part II: schizophrenia. Am J Hum Genet 2003; 73: 34–48.

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

17 DeLisi LE, Shaw SH, Crow TJ, Shields G, Smith AB, Larach VW et al. A genome-wide scan for linkage to chromosomal regions in 382 sibling pairs with schizophrenia or schizoaffective disorder. Am J Psychiatry 2002; 159: 803–812.
18 Abecasis GR, Cookson WO, Cardon LR. Pedigree tests of transmission disequilibrium. Eur J Hum Genet 2000; 8: 545–551.
19 Thiselton DL, Webb BT, Neale BM, Ribble RC, O'Neill FA, Walsh D et al. No evidence for linkage or association of neuregulin-1 (NRG1) with disease in the Irish study of high-density schizophrenia families (ISHDSF). Mol Psychiatry 2004; 9: 777–783.
20 Xiong L, Rouleau GA, DeLisi LE, St Onge J, Najafee R, Riviere JB et al. CAA insertion polymorphism in the 3'UTR of Nogo gene on 2p14 is not associated with schizophrenia. Mol Brain Res 2005; 133: 153–156.
21 Abecasis GR, Burt RA, Hall D, Bochum S, Doheny KF, Lundy SL et al. Genomewide scan in families with schizophrenia from the founder population of Afrikaners reveals evidence for linkage and uniparental disomy on chromosome 1. Am J Hum Genet 2004; 74: 403–417.
22 Van den Oord EJCG, Rujescu D, Robles JR, Giegling I, Birrell C, Bukszar J et al. Factor structure and external validity of the PANSS revisited. Schizophr Res 2006; 82: 213–223.
23 Takahashi S, Cui YH, Kojima T, Han YH, Zhou RL, Kamioka M et al. Family-based association study of markers on chromosome 22 in schizophrenia using African-American, European-American, and Chinese families. Am J Med Genet B Neuropsychiatr Genet 2003; 120: 11–17.
24 Kugoh H, Mitsuya K, Meguro M, Shigenami K, Schulz TC, Oshimura M. Mouse A9 cells containing single human chromosomes for analysis of genomic imprinting. DNA Res 1999; 6: 165–172.
25 Mergenthaler S, Hitchins MP, Blagitko-Dorfs N, Monk D, Wollmann HA, Ranke MB et al. Conflicting reports of imprinting status of human GRB10 in developing brain: how reliable are somatic cell hybrids for predicting allelic origin of expression? Am J Hum Genet 2001; 68: 543–545.
26 Sakatani T, Wei M, Katoh M, Okita C, Wada D, Mitsuya K et al. Epigenetic heterogeneity at imprinted loci in normal populations. Biochem Biophys Res Commun 2001; 283: 1124–1130.
27 Yamasaki K, Joh K, Ohta T, Masuzaki H, Ishimaru T, Mukai T et al. Neurons but not glial cells show reciprocal imprinting of sense and antisense transcripts of Ube3a. Hum Mol Genet 2003; 12: 837–847.
28 Khaitovich P, Hellmann I, Enard W, Nowick K, Leinweber M, Franz H et al. Parallel patterns of evolution in the genomes and transcriptomes of humans and chimpanzees. Science 2005; 309: 1850–1854.
29 Kochunov P, Fox P, Lancaster J, Tan LH, Amunts K, Zilles K et al. Localized morphological brain differences between English-speaking Caucasians and Chinese-speaking Asians: new evidence of anatomical plasticity. Neuroreport 2003; 14: 961–964.
30 Imprinted Gene Catalogue. http://igc.otago.ac.nz/home.html. Web 2006.
31 Reik W, Walter J. Genomic imprinting: parental influence on the genome. Nat Rev Genet 2001; 2: 21–32.
32 Morison IM, Ramsay JP, Spencer HG. A census of mammalian imprinting. Trends Genet 2005; 21: 457–465.
33 Bunzel R, Blumcke I, Cichon S, Normann S, Schramm J, Propping P et al. Polymorphic imprinting of the serotonin-2A (5-HT2A) receptor gene in human adult brain. Mol Brain Res 1998; 59: 90–92.
34 Fournier AE, GrandPre T, Gould G, Wang X, Strittmatter SM. Nogo and the Nogo-66 receptor. Prog Brain Res 2002; 137: 361–369.
35 Lauren J, Airaksinen MS, Saarma M, Timmusk T. A novel gene family encoding leucine-rich repeat transmembrane proteins differentially expressed in the nervous system. Genomics 2003; 81: 411–421.
36 Andrews J, Wang L, Csernansky JG, Gado MH, Barch DM. Abnormalities of thalamic activation and cognition in schizophrenia. Am J Psychiatry 2006; 163: 463–469.
37 Harrison PJ, Weinberger DR. Schizophrenia genes, gene expression, and neuropathology: on the matter of their convergence. Mol Psychiatry 2004; 10: 40–68.
38 Sun T, Patoine C, Abu-Khalil A, Visvader J, Sum E, Cherry TJ et al. Early asymmetry of gene transcription in embryonic human left and right cerebral cortex. Science 2005; 308: 1794–1798.
39 Hering-Hanit R, Achiron R, Lipitz S, Achiron A. Asymmetry of fetal cerebral hemispheres: in utero ultrasound study. Arch Dis Child Fetal Neonat Ed 2001; 85: F194–F196.
40 Prabhakar S, Noonan JP, Paabo S, Rubin EM. Accelerated evolution of conserved noncoding sequences in humans. Science 2006; 314: 786.
41 Kimura MI, Kazuki Y, Kashiwagi A, Kai Y, Abe S, Barbieri O et al. Dlx5, the mouse homologue of the human-imprinted DLX5 gene, is biallelically expressed in the mouse brain. J Hum Genet 2004; 49: 273–277.
42 Wilkins JF, Haig D. What good is genomic imprinting: the function of parent-specific gene expression. Nat Rev Genet 2003; 4: 359–368.
43 Peters AH, Schubeler D. Methylation of histones: playing memory with DNA. Curr Opin Cell Biol 2005; 17: 230–238.
44 Ioannidis JPA, Ntzani EE, Trikalinos TA, Contopoulos-Ioannidis DG. Replication validity of genetic association studies. Nat Genet 2001; 29: 306–309.
45 Csernansky JG, Schindler MK, Splinter NR, Wang L, Gado M, Selemon LD et al. Abnormalities of thalamic volume and shape in schizophrenia. Am J Psychiatry 2004; 161: 896–902.
46 De Fosse L, Hodge SM, Makris N, Kennedy DN, Caviness Jr VS, McGrath L et al. Language-association cortex asymmetry in autism and specific language impairment. Ann Neurol 2004; 56: 757–766.
47 Herbert MR, Harris GJ, Adrien KT, Ziegler DA, Makris N, Kennedy DN et al. Abnormal asymmetry in language association cortex in autism. Ann Neurol 2002; 52: 588–596.
48 Paulesu E, Demonet JF, Fazio F, McCrory E, Chanoine V, Brunswick N et al. Dyslexia: cultural diversity and biological unity. Science 2001; 291: 2165–2167.
49 Sommer IE, Ramsey NF, Mandl RC, Kahn RS. Language lateralization in monozygotic twin pairs concordant and discordant for handedness. Brain 2002; 125: 2710–2718.
50 Majercak J, Ray WJ, Espeseth A, Simon A, Shi XP, Wolffe C et al. LRRTM3 promotes processing of amyloid-precursor protein by BACE1 and is a positional candidate gene for late-onset Alzheimer's disease. Proc Natl Acad Sci USA 2006; 103: 17967–17972.
51 Abecasis GR, Cherny SS, Cookson WO, Cardon LR. Merlin—rapid analysis of dense genetic maps using sparse gene flow trees. Nat Genet 2002; 30: 97–101.
52 Wright M, De Geus E, Ando J, Luciano M, Posthuma D, Ono Y et al. Genetics of cognition: outline of a collaborative twin study. Twin Res 2001; 4: 48–56.
53 Maegawa S, Yoshioka H, Itaba N, Kubota N, Nishihara S, Shirayoshi Y et al. Epigenetic silencing of PEG3 gene expression in human glioma cell lines. Mol Carcinog 2001; 31: 1–9.
54 Stanssens P, Zabeau M, Meersseman G, Remes G, Gansemans Y, Storm N et al. High-throughput MALDI-TOF discovery of genomic sequence polymorphisms. Genome Res 2004; 14: 126–133.
55 Ehrich M, Nelson MR, Stanssens P, Zabeau M, Liloglou T, Xinarianos G et al. Quantitative high-throughput analysis of DNA methylation patterns by base-specific cleavage and mass spectrometry. Proc Natl Acad Sci USA 2005; 102: 15785–15790.
56 Klumperman J, Schweizer A, Clausen H, Tang BL, Hong W, Oorschot V et al. The recycling pathway of protein ERGIC-53 and dynamics of the ER-Golgi intermediate compartment. J Cell Sci 1998; 111: 3411–3425.

Supplementary Information accompanies the paper on the Molecular Psychiatry website (http://www.nature.com/mp)

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

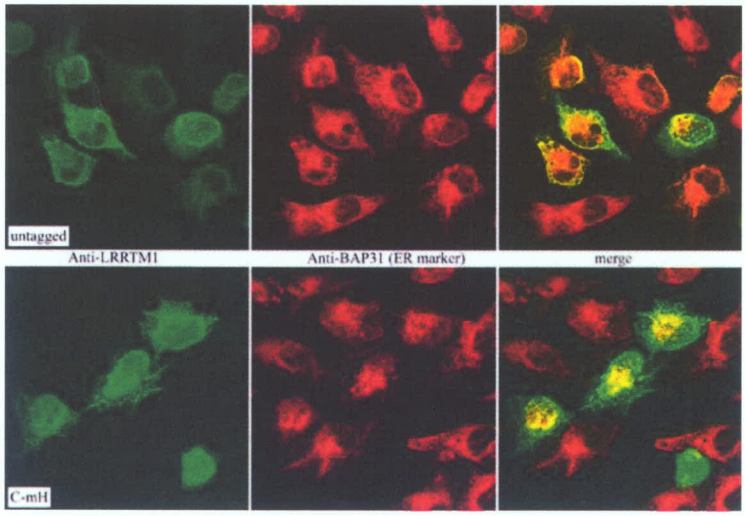
Molecular Psychiatry (2007) 12, 1057
& 2007 Nature Publishing Group All rights reserved 1359-4184/07 $30.00
www.nature.com/mp

IMAGE

LRRTM1 protein is located in the endoplasmic reticulum (ER) in mammalian cells

C Francks[1,2], S Maegawa[3,21], J Laurén[4,21], BS Abrahams[5], A Velayos-Baeza[1], SE Medland[6,7], S Colella[1], M Groszer[1], EZ McAuley[1], TM Caffrey[1], T Timmusk[8], P Pruunsild[8], I Koppel[8], PA Lind[7], N Matsumoto-Itaba[9], J Nicod[1], L Xiong[10], R Joober[11], W Enard[12], B Krinsky[12], E Nanba[3], AJ Richardson[13], BP Riley[5], NG Martin[6], SM Strittmatter[4], H-J Möller[14], D Rujescu[14], D St Clair[15], P Muglia[2], JL Roos[16], SE Fisher[1], R Wade-Martins[1], GA Rouleau[10], JF Stein[13], M Karayiorgou[17], DH Geschwind[5], J Ragoussis[1], KS Kendler[5], MS Airaksinen[18], M Oshimura[9], LE DeLisi[19,20] and AP Monaco[1]

[1]Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK; [2]GlaxoSmithKline SpA, Research Centre, Verona, Italy; [3]Division of Functional Genomics, Research Center for Bioscience and Technology, Tottori University, Yonago, Tottori, Japan; [4]Department of Neurology, Yale University School of Medicine, New Haven, CT, USA; [5]Department of Neurology, David Geffen School of Medicine, University of California at Los Angeles, Los Angeles, CA, USA; [6]Virginia Institute for Psychiatric and Behavioral Genetics, Department of Human Genetics, Virginia Commonwealth University, Richmond, VA, USA; [7]Queensland Institute of Medical Research, Brisbane, QLD, Australia; [8]Department of Gene Technology, Tallinn University of Technology, Tallinn, Estonia; [9]Department of Biomedical Science, Regenerative Medicine and Biofunction, Graduate School of Medical Science, Tottori University, Yonago, Tottori, Japan; [10]Center for the Study of Brain Diseases, CHUM Research Center, Montreal University, Montreal, QC, Canada; [11]Departments of Psychiatry, Neurology and Human Genetics, McGill University, Douglas Hospital Research Centre, Montreal, QC, Canada; [12]Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany; [13]Department of Physiology, Anatomy and Genetics, University of Oxford, Oxford, UK; [14]Department of Psychiatry, Ludwig Maximilians University, Munich, Germany; [15]Institute of Medical Sciences, University of Aberdeen, Aberdeen, Scotland; [16]Weskoppies Hospital, Department of Psychiatry, University of Pretoria, Pretoria, Republic of South Africa; [17]Laboratory of Human Neurogenetics, Rockefeller University, New York, NY, USA; [18]Neuroscience Center, University of Helsinki, Helsinki, Finland; [19]Department of Psychiatry, New York University, New York, NY, USA and [20]The Nathan S Kline Institute for Psychiatric Research, Orangeburg, New York, NY, USA



Molecular Psychiatry (2007) 12, 1057; doi:10.1038/sj.mp.4002116

LRRTM1 is an imprinted gene that shows parent-of-origin association with human handedness and schizophrenia. The left-hand panels (green) show localization of LRRTM1 within MRC-5 SV2 cells, following transfection with untagged (top) or myc þ His-tagged (bottom) LRRTM1, and detection with an LRRTM1 antipeptide. The red panels (middle) show the localization of BAP31, an ER marker. In the right-hand panels, the images for LRRTM1 and BAP31 are overlaid, and the yellow areas indicate the colocalization of the two proteins. LRRTM1 was also detected in the axons and growth cones of neurons but not at the cell surface membrane. Together, these data suggest that LRRTM1 may be involved in intracellular trafficking within axons. For more information on this topic, see the paper by Francks et al. on pages 1129–1139.

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

A12: 127

LETTERS

nature genetics

© 2008 Nature Publishing Group http://www.nature.com/naturegenetics

Strong association of de novo copy number mutations with sporadic schizophrenia

Bin Xu[1,2], J Louw Roos[3], Shawn Levy[4], E J van Rensburg[5], Joseph A Gogos[1,6] & Maria Karayiorgou[2]

Schizophrenia is an etiologically heterogeneous psychiatric disease, which exists in familial and nonfamilial (sporadic) forms[1]. Here, we examine the possibility that rare de novo copy number (CN) mutations with relatively high penetrance contribute to the genetic component of schizophrenia. We carried out a whole-genome scan and implemented a number of steps for finding and confirming CN mutations. Confirmed de novo mutations were significantly associated with schizophrenia (P ¼ 0.00078) and were collectively B 8 times more frequent in sporadic (but not familial) cases with schizophrenia than in unaffected controls. In comparison, rare inherited CN mutations were only modestly enriched in sporadic cases. Our results suggest that rare de novo germline mutations contribute to schizophrenia vulnerability in sporadic cases and that rare genetic lesions at many different loci can account, at least in part, for the genetic heterogeneity of this disease.

Schizophrenia has a strong genetic component[2,3], but the genetic architecture of the disease remains unknown. The predominant common disease–common allele hypothesis[4] has inspired scores of association studies[5], but has failed so far to provide unequivocal statistical support for any gene in schizophrenia and disease-related endophenotypes. Evidence supporting the importance of rare mutations in schizophrenia vulnerability has been provided by the reproducible observation of rare but recurrent and highly penetrant de novo CN mutations at the 22q11.2 locus responsible for introducing 'new' cases of schizophrenia in the population[6,7], and by reports of rare cases of cytogenetic abnormalities that occasionally co-segregate with schizophrenia in families[8]. Recent studies have established that rare de novo CN mutations are collectively significant risk factors for sporadic cases with childhood neurodevelopmental disorders, such as autism spectrum disorders[9,10]. Nevertheless, the full contribution of rare de novo CN mutations to sporadic schizophrenia remains unknown.

To search for de novo genomic 'loss' or 'gain' events, we analyzed 359 subjects and both of their biological parents, representing a total

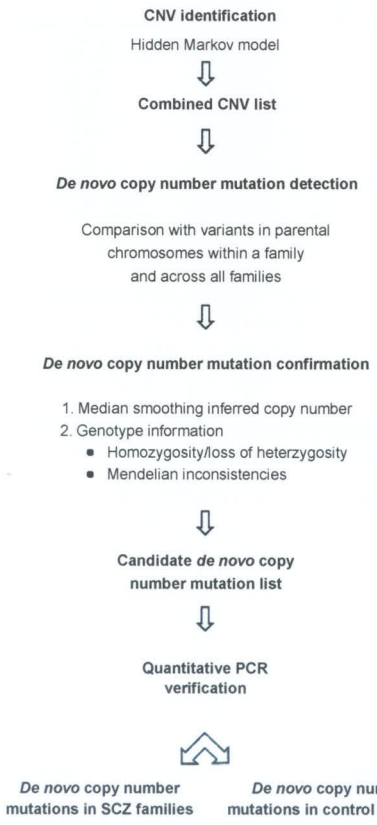


Figure 1 Procedures for CNV identification, de novo CNV detection and verification. The flow diagram describes the step-by-step procedure for identifying genomic regions that had either fewer than or more than two genomic copies (genomic 'loss' or 'gain' events) in the subjects but not their parents.

[1]Department of Physiology and Cellular Biophysics, Columbia University, 630 West 168th Street, New York, New York 10032, USA. [2]Department of Psychiatry, Columbia University, 1051 Riverside Drive, New York, New York 10032, USA. [3]Weskoppies Hospital, Private Bag H113, Pretoria 0001, Republic of South Africa. [4]Vanderbilt University, Microarray Shared Resource, 465 21st Avenue South, Nashville, Tennessee 37232, USA. [5]Department of Genetics, University of Pretoria, P.O. Box 667, Pretoria 0001, Republic of South Africa. [6]Department of Neuroscience, Columbia University, 1051 Riverside Drive, New York, New York 10032, USA. Correspondence should be addressed to M.K. (mk2758@columbia.edu) or J.A.G. (jag90@columbia.edu).

Received 13 March; accepted 29 April; published online 30 May 2008; doi:10.1038/ng.162

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

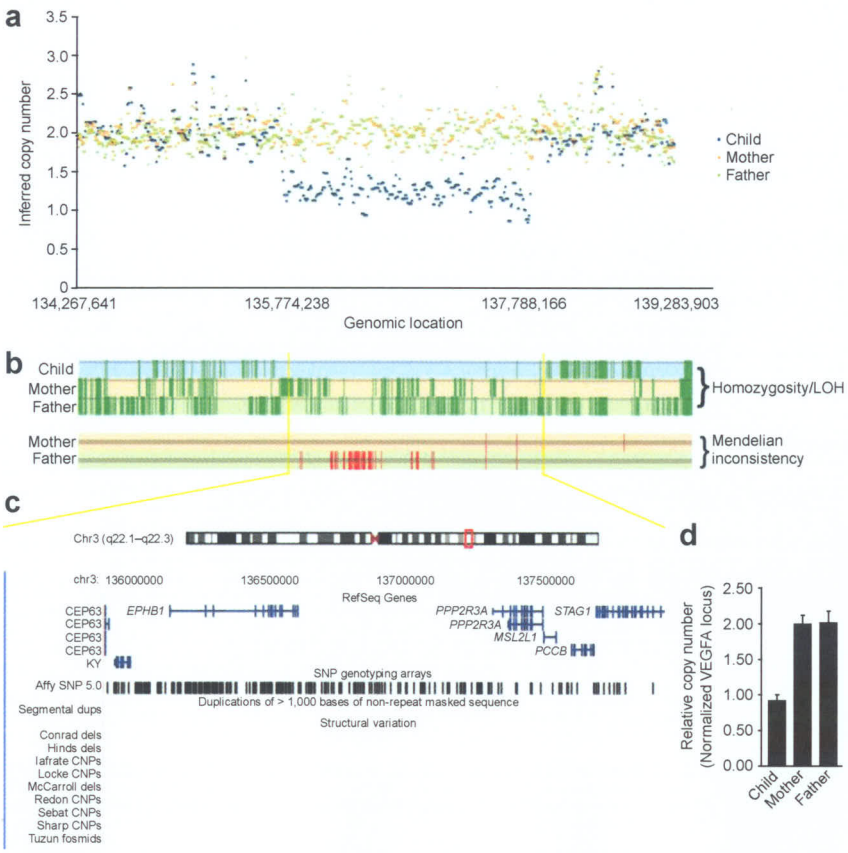


Figure 2 Identification and confirmation of a de novo CN mutation in a subject with schizophrenia. The de novo CN mutation was identified by HMM and confirmed by ascertaining parents for change in copy number. (a) Median smoothing–inferred copy numbers for the proband, mother and father along the genomic locus indicate a deletion in the proband. (b) Genotype information indicates homozygosity or loss-of-heterozygosity in the proband (upper three rows), as well as mendelian inconsistencies between the proband and parents (lower two rows). Green bars indicate heterozygous SNPs. Red bars indicate mendelian inconsistencies in the genotype calls between the proband and the parents. (c) The map of the deletion region indicates annotated known genes, Affymetrix probe set density and structural variation identified by previous studies as shown in the University of California Santa Cruz Genome Browser, Human March 2006 (hg18) assembly. (d) Taqman quantitative PCR (qPCR) confirms the relative copy number change in the proband. qPCR was done in five replicates for each sample. Error bars, s.d.

of 1,077 individuals recruited from the genetically homogeneous Afrikaner population in South Africa[11]. Informed consent was obtained from all participants, and all procedures were approved by the Institutional Review Committees of Columbia University and University of Pretoria. We prepared DNA samples from whole blood and used them in a whole-genome scan employing the Affymetrix 5.0 GeneChip array. We first compared the de novo mutation rate in a sample of 152 individuals diagnosed with schizophrenia or schizoaffective disorder, with no history of the disease in a first- or second-degree relative ('sporadic cases cohort'), to a sample of 159 unaffected individuals recruited from the same population ('control cohort'). Identities of all subjects and their parents were coded so that analysis could be done blind to affected status while maintaining knowledge of the parent–child relations. To maximize copy number variant (CNV) discovery, we merged results from two software packages, DNA-Chip Analyzer (dChip)[10,12,13] and Partek (version 6.3). To improve the signal-to-noise ratio, we only considered CNVs that involved at least ten consecutive probe sets, thus providing an average resolution of B30 kb. We implemented a number of steps to avoid false positives and ensure that events we detected in subjects were in fact de novo. As part of the filtering process, we disregarded CNVs if they showed Z 50% overlap with a CNV detected in any parental chromosome (within a family and across families). This step was done in order to filter out any common CNVs in 'unstable' genomic regions with unusually high new mutation rates, which are present in the unaffected biological parents and in the general population, and also eliminate common inherited CNVs that would incorrectly appear as de novo[9]. A flow chart of our procedures for finding and validating de novo CN mutations is depicted in Figure 1, and an example of an identified CN mutation is shown in Figure 2. In all, 19 CNVs were detected and confirmed to be de novo in 17 individuals, consisting of 15 cases and two controls. Notably, we found that 2 of 15 affected individuals carry two confirmed de novo events each, similar to findings in autism[10]. All validated de novo mutations are listed in Table 1. The frequency of de novo CN mutation carriers was B10% (15 of 152) in our cohort of sporadic cases (Table 2). Fourteen of the 15 affected individuals were diagnosed with strict schizophrenia (99.3%), and only one female subject was diagnosed with schizoaffective disorder, depressed type. The frequency of de novo CN mutation carriers among unaffected individuals was 1.3% (2 of 159), consistent with three previous studies[9,10,14]. Our data demonstrate that de novo CN mutations are collectively B8 times more frequent in sporadic cases with schizophrenia (15 of 152) than in unaffected individuals (2 of 159), with an association that is statistically highly significant (two-tailed Fisher's exact test, P ¼ 0.00078).

To address the specificity and relative strength of this association, we first estimated the de novo mutation rate in 48 individuals diagnosed with schizophrenia or schizoaffective disorder with positive disease history in a first- or second-degree relative ('familial cases cohort'). In contrast to the sporadic cases, we did not find any de novo structural lesions in the familial cases cohort (P ¼ 0.02, compared to sporadic cases; Table 2). Thus, the association between de novo CN mutations and schizophrenia seems to be primarily confined to the sporadic cases. Second, we compared the collective rates of de novo mutation carriers to those carrying rare inherited mutations in the same cohorts of 152 sporadic cases and 159 controls. Similar to our de novo mutation detection procedure, inherited rare CNVs detected in subjects and their parents were considered only if they involved at least ten consecutive probe sets (average resolution of B30 kb) and did not show Z 50% overlap with a CNV detected in any parental chromosome other than those of the biological parents. Rare inherited gains and losses in cases and controls and affected genes are shown in Table 3 and Supplementary Tables 1 and 2 online. The frequency of carriers of rare structural lesions inherited from unaffected parents

© 2008 Nature Publishing Group http://www.nature.com/naturegenetics

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Table 1 Rare de novo copy number mutations identified in this study

Sex	Diagnosis	Cytoband	Start position (bp)	Size (kb)	CN change	Number of genes	Single gene mutations
M	SCZ	14q32.13–14q32.2	93,831,174	2803.6	Gain	24	
		1p34.2	40,063,283	614.1	Gain	10	
M	SCZ	22q11.21	17,404,673	2593.9	Loss	41	
M	SCZ	3q22.2	135,765,543	2022.6	Loss	7	
F	SCZAFF	22q11.21	17,052,138	1533.8	Loss	30	
F	SCZ	22q11.21	17,386,984	1444.5	Loss	27	
F	SCZ	6p22.3	17,575,070	364.3	Gain	4	
M	SCZ	16p13.2	8,116,843	297.1	Gain	3	
M	SCZ	9p23	11,617,789	259.9	Loss	1	
		12q24.23	118,689,478	24.8	Gain	1	CIT
F	SCZ	19q13.12	42,014,819	178.5	Loss	4	
F	SCZ	3p24.1	28,277,792	164.3	Gain	3	
M	SCZ	5q31.1	130,863,839	114.3	Loss	1	RAPGEF6
M	SCZ	3p26.1	8,126,457	67.2	Gain	2	
M	SCZ	4q34.1	175,113,952	43.3	Gain	4	
M	SCZ	16p13.2	6,888,176	30.7	Gain	1	A2BP1
M	SCZ	10q21.1	57,665,722	27.6	Loss	–	
F	Unaffected	2q11.1–2q11.2	94,693,998	4503.7	Gain	51	
F	Unaffected	7q11.23	72,666,695	1105.2	Gain	18	

SCZ, schizophrenia; SCZAFF, schizoaffective disorder, depressed type.

was ~30% (46 of 152) in our cohort of sporadic cases and ~20% (32 of 159) in controls (Table 2). Thus, sporadic cases with schizophrenia are only ~1.5 times more likely than unaffected controls to harbor an inherited rare CN mutation (two-tailed Fisher's exact test, $P = 0.049$). We obtained almost identical results (relative enrichment 1.3, $P = 0.023$, data not shown) when we defined rare inherited variants as those not overlapping with a common CNV collection downloaded from the Database of Genomic Variants (hg18 version 3). The modest collective enrichment of rare inherited CN mutations suggests that only a small portion of them should be expected to contribute to the pathogenesis of the sporadic cases. By contrast, the marked and specific enrichment of rare de novo CN mutations in sporadic schizophrenia strongly suggests that many of the de novo lesions we observed in affected individuals contribute to the disorder.

We observed an approximately equal number of de novo genomic gains and losses in individuals with schizophrenia, whereas both mutations detected in controls were duplications. Of the eight deletions observed, five occurred in paternal and three in maternal chromosomes, whereas of the five duplications for which origin could be estimated (see Supplementary Methods online), three occurred in paternal and two in maternal chromosomes (Supplementary Table 3 online). Segmental duplications were present near 9 of the 17 de novo mutations and especially among ones larger than 300 kb (Supplementary Table 3). Most of the de novo CN mutations we detected in affected individuals affect one or more known genes or transcripts, and three of them affect a single known gene (Table 1).

The male-to-female ratio in affected individuals was 2:1 for CN mutation carriers, similar to the 2.4:1 gender ratio of the entire sporadic cases cohort. It is notable that, with the possible exception of a higher frequency of developmental delays ($P = 0.04$, uncorrected), there were no significant differences ($P > 0.05$) in age at disease onset, history of learning disabilities, presence of mental retardation or parental age at birth of the proband between de novo CN mutation carriers and the rest of the sporadic cases cohort (Supplementary Table 3 and Supplementary Methods). In an independent study, we evaluated the facial features of a subset of our cohort ($n = 85$) and identified six subjects (7%) with mildly dysmorphic facial features[15]. Notably, of these six subjects, four (67%) are among the subjects identified here as carrying de novo structural lesions: two carry 22q11.2 microdeletions, one carries a combination of two duplications at 14q32.13–32.2 and 1p34.2, and another carries a microdeletion at 10q21.1.

Our results confirm the importance of microdeletions in the 22q11.2 locus as the only known recurrent CN mutation associated with schizophrenia[6]. We identified three de novo 22q11.2 microdeletions (1.8% rate) with sizes consistent with the two predominant sizes (3 Mb or 1.5 Mb) reported in the literature (Table 1). All three carriers were diagnosed with typical schizophrenia or schizoaffective disorder. No other recurrent structural mutations were identified. However, several of the de novo losses or gains we observed deserve further discussion. For example, we found a microdeletion involving exons 2–11 of the RAPGEF6 gene (Table 1) within a schizophrenia susceptibility locus at 5q previously highlighted by the identification of a de novo microdeletion in a individual with schizophrenia[16], as well as by multiple linkage studies[17]. Systematic fine mapping of this locus identified a haplotype spanning three neighboring genes (CDC42SE2-RAPGEF6-ACSL6) as associated with schizophrenia in multiple cohorts[18]. Identification of a de novo microdeletion disrupting RAPGEF6 makes this gene a leading candidate from this region. RAPGEF6 is a GTP exchange factor that acts with the small GTPases Rap1 and Rap2 to modulate the active state of G proteins[19]. Two other de novo deletions encompassing multiple genes, one on chromosome 19q13.12 and another on chromosome 3q22.2, remove, respectively, a cluster of C_2H_2 zinc finger transcription factors and the ephrin receptor EphB1 (Fig. 2)[20,21]. An intragenic duplication at 12q24 (a locus frequently reported as linked with bipolar disorder[22]) includes eight exons of the CIT (citron) gene[23,24] and is predicted to affect the structure of the encoded protein product. Finally, we identified in one individual a

Table 2 Increased frequency of rare de novo CN mutations in sporadic schizophrenia

Subject group	Number of families	de novo CN mutation carriers N	de novo CN mutation carriers %	P value[a] vs. controls	P value[a] vs. familial	Inherited CN mutation carriers N	Inherited CN mutation carriers %	P value[a] vs. controls
Sporadic	152	15	9.90	0.00078	0.02	46	30	0.049
Familial	48	0	0	0.97		n.a.	n.a.	
Controls	159	2	1.26			32	20	

n.a., not applicable.
[a]Statistical significance was determined using Fisher's exact probability test.

© 2008 Nature Publishing Group http://www.nature.com/naturegenetics

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

de novo duplication at 14q32.13–32.2 that includes DICER1, a key gene in the miRNA processing pathway[25]. This is notable because the 22q11.2 microdeletion removes one copy of another key gene in this pathway, DGCR8/(PASHA), and the resulting abnormality in miRNA biogenesis contributes to the behavioral and neuronal deficits associated with this CN mutation in a mouse model[26]. The number of genes included within the identified de novo CN mutations (Supplementary Table 4 online) is relatively small, but gene-class testing among these genes revealed a nominally significant enrichment in functionally defined pathways (Gene Ontology (GO) terms) associated with neural development (24 genes), small GTPase activity (4 genes) and RNA binding/processing (6 genes) (Supplementary Table 5 online).

The relatively low structural mutation rate[27] is used as an argument to assign high likelihood of causation to de novo rearrangements found in sporadic cases. In some instances, additional independent risk factors and more inclusive models that account for the variable degree of penetrance associated with de novo events[28,29] should be considered. In such models, de novo mutations represent necessary but not always sufficient genetic 'hits', which lead to a disease state only when a critical threshold of genetic burden, provided in part by additional inherited genetic variation, is surpassed. Irrespective of the precise pathogenic models, our findings have two noteworthy implications for our understanding of the genetic basis of schizophrenia. First, they strongly suggest that rare de novo germline mutations contribute to schizophrenia vulnerability in sporadic cases with the disease, thus extending and complementing our initial observation at the 22q11.2 locus[6]. Because of the limited resolution of our genome microarray scan and its inability to identify balanced abnormalities, the rate of new CN mutations that we detect is likely an underestimate. Such new mutations can explain, at least in part, how schizophrenia persists in the population despite the low fecundity of affected individuals[1]. Second, our findings suggest that de novo genetic lesions at many different loci can contribute to schizophrenia, and this heterogeneity may account for the difficulties in finding genetic variants with a major effect on disease risk. Heterogeneity may be more pronounced among sporadic cases, although a portion of de novo events (depending on disease severity or age at onset) is likely to be inherited, thus contributing to the genetic heterogeneity among familial cases as well. It should be noted that, while this work was under review, Walsh and colleagues[30], using different CN mutation discovery platforms and a case-control design, reported an excess of rare structural variants that influence neurodevelopmental genes in adult cases as well as in atypical childhood-onset schizophrenia cases. The relative contribution of de novo or inherited CN mutations in sporadic versus

Table 3 Rare inherited copy number mutations affecting known genes

Sex	Diagnosis	Cytoband	Start position (bp)	Size (kb)	CN change	Number of genes	Single gene mutations	Inherited from
M	SCZAFF*	9p24.1	4,968,475	1,024.2	Gain	13		Father
F	SCZ	5q15	95,262,168	579.5	Gain	2		Mother
		17p11.2	19,910,327	251.6	Gain	1	SPECC1	Mother
M	SCZ	2q24.1	154,942,524	365.3	Gain	2		Father
F	SCZ	12q12	36,768,808	301.4	Gain	1	ALG10B	Father
M	SCZ	7q11.22	69,953,808	285.5	Gain	1	WBSCR17	Father
M	SCZ	11q14.1	83,623,412	269.2	Loss	1	DLG2	Mother
M	SCZ	4p16.2	3,405,913	249.6	Gain	5		Mother
F	SCZ	11q25	133,779,886	244.1	Gain[a]	1	B3GAT1	Father
F	SCZ	4q28.3	135,150,046	227.7	Loss	1	PABPC4L	Father
		7p22.1	6,005,655	44.2	Loss	3		Mother
M	SCZ	3p21.31	47,755,720	221.7	Gain	3		Father
F	SCZAFF	12q24.31	122,184,497	204.0	Gain[b]	4		Father
M	SCZ	2q35	215,198,940	204.0	Loss	1	BARD1	Father
M	SCZ	7q36.2	153,155,808	203.3	Gain	1	DPP6	Father
F	SCZ	3p14.2	59,845,328	197.1	Loss	1	FHIT	Father
M	SCZ	2q21.2	133,523,822	161.9	Loss	1	NAP5	Father
M	SCZAFF	22q11.23	23,049,604	128.6	Gain[c]	2		Father
M	SCZ	1q21.1	144,316,672	122.4	Loss	5		Father
M	SCZ	2q22.2	142,518,411	97.7	Gain	1	LRP1B	Mother
		4q21.21	79,647,471	87.9	Gain	2		Mother
F	SCZ	13q32.1	94,737,932	65.2	Gain	1	ABCC4	Mother
M	SCZ	17q25.1	72,105,584	59.8	Gain[d]	1	ST6GALNAC1	Mother
M	SCZ	13q33.3	106,785,713	54.6	Loss	1	FAM155A	Father
		2p22.2	38,062,903	54.4	Gain	1	FAM82A	Father
F	SCZ	15q21.3	55,489,977	42.8	Loss[e]	1	CGNL1	Mother
F	SCZAFF*	12p12.1	21,583,158	36.4	Loss	1	GYS2	Mother
M	SCZ	17q24.2	62,132,371	35.8	Loss[f]	1	PRKCA	Father
M	Unaffected	22q11.21	17,242,610	2,516.4	Gain	43		Mother
F	Unaffected	11q13.1	64,728,025	1,730.2	Gain	69		Mother
		6q26	162,636,295	175.6	Loss	1	PARK2	Mother
F	Unaffected	9q34.2–9q34.3	136,493,115	1,197.9	Gain	12		Father
F	Unaffected	16p13.11	14,776,350	1,162.5	Gain	13		Father
M	Unaffected	4q33	170,922,834	884.5	Gain	2		Mother
F	Unaffected	6q23.3–6q24.1	138,841,552	504.2	Gain	3		Father
F	Unaffected	7q11.23	75,983,580	455.9	Loss	2		Mother
F	Unaffected	18q22.1	63,214,284	230.3	Loss	1	DSEL	Father
F	Unaffected	8q11.23	53,669,436	206.0	Gain	1	RB1CC1	Father
		2p16.1	56,275,952	31.0	Loss	1	CCDC85A	Mother
M	Unaffected	9p24.1	6,592,066	195.4	Loss	2		Mother
F	Unaffected	17p13.3	770,573	147.3	Gain	3		Father
F	Unaffected	8q22.2	100,607,246	132.4	Loss	1	VPS13B	Mother
F	Unaffected	11p15.5	932,344	121.4	Gain[g]	2		Father
M	Unaffected	16q24.1	83,398,254	114.8	Loss	1	CRISPLD2	Mother
M	Unaffected	5q11.2	54,524,676	108.9	Gain	2		Father
M	Unaffected	4q35.2	187,326,127	103.3	Gain[h]	4		Mother
F	Unaffected	12q21.32	86,977,302	78.5	Loss	1	CEP290	Mother
M	Unaffected	14q22.2	52,358,200	77.4	Gain	1	FERMT2	Mother
F	Unaffected	6q22.31	124,307,100	75.6	Loss	1	NKAIN2	Father
M	Unaffected	8q23.1	106,880,858	66.9	Loss	1	ZFPM2	Mother
F	Unaffected	18q21.31	53,910,209	53.8	Loss	1	NEDD4L	Father

SCZ, schizophrenia; SCZAFF, schizoaffective disorder, depressed type; SCZAFF*, schizoaffective disorder, mixed type.
[a–h]Denotes the same individual as in Supplementary Table 1.

© 2008 Nature Publishing Group http://www.nature.com/naturegenetics

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

familial cases was not addressed in that study, but, as suggested by our results here, the observed ~3 fold excess reported[30] likely represents the combined but uneven contribution of both inherited and de novo rare CN mutations to familial and sporadic cases of schizophrenia.

METHODS

Subjects. Both affected and control families were recruited from the Afrikaner population in South Africa. Afrikaner heritage was established by surname and by having Afrikaans-speaking grandparents. Diagnostic evaluations were done in person by specially trained clinicians using the Diagnostic Instrument for Genetic Studies (DIGS), which was translated and back-translated into Afrikaans[11]. Upon completion of the DIGS, the interviewers assigned DSM-IV diagnoses[11] and completed a detailed narrative report. All narratives and DIGS were independently evaluated by appropriately trained clinicians in New York City. Family history was obtained by a nursing sister, who recorded detailed pedigree information, and by the clinical interviewer, who inquired in detail about family history during the clinical interview. The sporadic cases cohort consisted of all probands collected from this population who satisfied two criteria: i) both biological parents were available for genotyping, and ii) there was no history of schizophrenia in first- or second-degree relatives. Similarly, our familial cases cohort consisted of all probands collected from this population who satisfied two criteria: i) both biological parents were available for genotyping, and ii) there was a positive history of schizophrenia in first or second-degree relatives. Finally, the control cohort consisted of 159 families (triads) recruited either by E.J.V.R. or J.L.R. from the Afrikaner community. A subset of the control families completed a detailed self-report questionnaire that inquired about several mental conditions, including phobias, anxiety, depression and history of treatment for any of these conditions. Also, family history of any mental illness up to three generations was excluded. Families were excluded only if sample failures occurred at the quality control and CNV identification step as described in Supplementary Methods (even if one of the three DNA samples included in each triad failed, the entire triad was excluded from analysis, as the de novo nature of CNV could not be established). For additional cohort characteristics, see Supplementary Methods.

Genotyping. We genotyped subjects using the Affymetrix Human Genome-Wide SNP Array 5.0 (Affymetrix), which contains 500,568 SNPs, as well as 420,000 additional nonpolymorphic probes that can assess other genetic differences such as CN variation across the human genome. All sample processing was completed at the Vanderbilt Microarray Shared Resource according to manufacturer's protocol. Following hybridization, we used the Affymetrix GeneChip Command Console (AGCC) (Affymetrix) to acquire raw genotyping data. Raw data in the form of CEL files was imported into the Affymetrix Genotyping Console (Affymetrix) for quality control analysis and genotype calling using the BRLMM-P algorithm (Affymetrix). Average call rate on arrays used in this study was 99.43%.

CNV identification, de novo CNV detection and de novo CNV verification. We analyzed CEL files from all chips using two software packages, DNA-Chip Analyzer (dChip)[10,12,13] and Partek Genomics Suite (Partek software, version 6.3 Beta, build #6.07.1127), which enable detection of CNVs by performing scaling, normalization and feature extraction of hybridization signal intensities, as well as hidden Markov model (HMM) analysis for inferring CN. The entire procedure included five steps: (i) quality control; (ii) CNV identification; (iii) de novo CNV detection; (iv) de novo CNV verification by intensity and genotyping filters and (v) de novo CNV confirmation by duplex Taqman quantitative real time PCR (qPCR). The procedure is described in detail in Supplementary Methods.

Rare inherited CNV detection. We considered a CNV as inherited only if there was 50% or more overlap in size with a variant in its biological parental chromosomes. Inherited rare CNVs detected in subjects and their parents were considered only if they involved at least ten consecutive probe sets and did not show Z 50% overlap with a CNV detected in any parental chromosome other than those of the biological parents. An alternative criterion (CNVs not showing Z 50% overlap with the CNV collection downloaded from Database of Genomic Variants (hg18 version 3)) was also applied with almost identical results. For each identified candidate rare inherited CNV, all genotypes and inferred CN of SNPs within the region were retrieved for each triad (the subject and his or her parents). Loss of heterozygosity, mendelian inconsistencies and average median smoothing inferred CN of the region were analyzed in MS-Excel driven by a VBA script and further examined visually. The average size and gene content of deletions was 126 kb (0.3 genes per kb) and 131 kb (0.6 genes per kb) for cases and controls, respectively. The average size and gene content of duplications was 220 kb (1.7 genes per kb) and 492 kb (7.8 genes per kb), respectively, for cases and controls (this last difference was primarily caused by two extended duplications at chromosomes 22q11.21 and 11q13.1 identified in two unaffected controls).

Additional analyses. Determination of the parental origin of de novo CNVs and functional annotation analysis of genes within the CN mutations are described in Supplementary Methods.

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS

We thank all the families who participated in this research. We also thank H. Pretorius and nursing sisters R. van Wyk, C. Botha and H. van den Berg for their assistance with subject recruitment and diagnostic evaluations. We thank A. Abrams-Downey and Y. Sun for expert technical assistance and J. Merriam and R.L. Allikmets for their help. We also thank A. Hagerty, J. Mote and J. Hallock for sample processing support and B. Boone and P. Dexheimer for genotyping data generation and support. All microarray experiments were performed in the Vanderbilt Microarray Shared Resource, which is supported by the Vanderbilt Ingram Cancer Center (P30 CA68485), the Vanderbilt Digestive Disease Center (P30 DK58404) and the Vanderbilt Vision Center (P30 EY08126). This work was supported in part by National Institute of Mental Health (NIMH) grants MH061399 (to M.K.) and MH077235 (to J.A.G.) and the Lieber Center for Schizophrenia Research at Columbia University Medical Center.

AUTHOR CONTRIBUTIONS

B.X. carried out the analysis; J.L.R. collected subjects and participated in the diagnostic evaluations; E.J.V.R. collected a subset of the control subjects; S.L. oversaw the genotyping; J.A.G. and M.K. designed the project and oversaw the analysis; M.K. directed the project; and B.X., J.A.G. and M.K. wrote the paper.

Published online at http://www.nature.com/naturegenetics/
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/

1. Gottesman, I.I. Schizophrenia Genesis. (W.H. Freeman and Company, New York, 1991).
2. Gottesman, I.I. & Shields, J. A polygenic theory of schizophrenia. Proc. Natl. Acad. Sci. USA 58, 199–205 (1967).
3. Sullivan, P.F., Kendler, K.S. & Neale, M.C. Schizophrenia as a complex trait: evidence from a meta-analysis of twin studies. Arch. Gen. Psychiatry 60, 1187–1192 (2003).
4. Pritchard, J.K. & Cox, N.J. The allelic architecture of human disease genes: common disease-common variant...or not? Hum. Mol. Genet. 11, 2417–2423 (2002).
5. Karayiorgou, M. & Gogos, J.A. Schizophrenia genetics: uncovering positional candidate genes. Eur. J. Hum. Genet. 14, 512–519 (2006).
6. Karayiorgou, M. et al. Schizophrenia susceptibility associated with interstitial deletions of chromosome 22q11. Proc. Natl. Acad. Sci. USA 92, 7612–7616 (1995).
7. Liu, H. et al. Genetic variation in the 22q11 locus and susceptibility to schizophrenia. Proc. Natl. Acad. Sci. USA 99, 16859–16864 (2002).
8. MacIntyre, D.J., Blackwood, D.H., Porteous, D.J., Pickard, B.S. & Muir, W.J. Chromosomal abnormalities and mental illness. Mol. Psychiatry 8, 275–287 (2003).
9. Sebat, J. et al. Strong association of de novo copy number mutations with autism. Science 316, 445–449 (2007).
10. Marshall, C.R. et al. Structural variation of chromosomes in autism spectrum disorder. Am. J. Hum. Genet. 82, 477–488 (2008).
11. Abecasis, G.R. et al. Genomewide scan in families with schizophrenia from the founder population of Afrikaners reveals evidence for linkage and uniparental disomy on chromosome 1. Am. J. Hum. Genet. 74, 403–417 (2004).
12. Autism Genome Project Consortium. Mapping autism risk loci using genetic linkage and chromosomal rearrangements. Nat. Genet. 39, 319–328 (2007).
13. Li, C. & Wong, W.H. Model-based analysis of oligonucleotide arrays: expression index computation and outlier detection. Proc. Natl. Acad. Sci. USA 98, 31–36 (2001).
14. Redon, R. et al. Global variation in copy number in the human genome. Nature 444, 444–454 (2006).
15. Wiehahn, G.J. et al. Assessment of the frequency of the 22q11 deletion in Afrikaner schizophrenic patients. Am. J. Med. Genet. B. Neuropsychiatr. Genet. 129, 20–22 (2004).

© 2008 Nature Publishing Group http://www.nature.com/naturegenetics

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

16. Bennett, R.L., Karayiorgou, M., Sobin, C.A., Norwood, T.H. & Kay, M.A. Identification of an interstitial deletion in an adult female with schizophrenia, mental retardation, and dysmorphic features: further support for a putative schizophrenia-susceptibility locus at 5q21–23.1. Am. J. Hum. Genet. 61, 1450–1454 (1997).
17. Lewis, C.M. et al. Genome scan meta-analysis of schizophrenia and bipolar disorder, part II: schizophrenia. Am. J. Hum. Genet. 73, 34–48 (2003).
18. Chen, X. et al. Haplotypes spanning SPEC2, PDZ-GEF2 and ACSL6 genes are associated with schizophrenia. Hum. Mol. Genet. 15, 3329–3342 (2006).
19. Kuiperij, H.B. et al. Characterisation of PDZ-GEFs, a family of guanine nucleotide exchange factors specific for Rap1 and Rap2. Biochim. Biophys. Acta 1593, 141–149 (2003).
20. Kayser, M.S., McClelland, A.C., Hughes, E.G. & Dalva, M.B. Intracellular and trans-synaptic regulation of glutamatergic synaptogenesis by EphB receptors. J. Neurosci. 26, 12152–12164 (2006).
21. Yue, Y. et al. Specification of distinct dopaminergic neural pathways: roles of the Eph family receptor EphB1 and ligand ephrin-B2. J. Neurosci. 19, 2090–2101 (1999).
22. Lyons-Warren, A. et al. Evidence of association between bipolar disorder and Citron on chromosome 12q24. Mol. Psychiatry 10, 807–809 (2005).
23. Di Cunto, F. et al. Defective neurogenesis in citron kinase knockout mice by altered cytokinesis and massive apoptosis. Neuron 28, 115–127 (2000).
24. Furuyashiki, T. et al. Citron, a Rho-target, interacts with PSD-95/SAP-90 at glutamatergic synapses in the thalamus. J. Neurosci. 19, 109–118 (1999).
25. Murchison, E.P. & Hannon, G.J. miRNAs on the move: miRNA biogenesis and the RNAi machinery. Curr. Opin. Cell Biol. 16, 223–229 (2004).
26. Stark, K.L. et al. Altered brain microRNA biogenesis contributes to phenotypic deficits in a 22q11-deletion mouse model. Nat. Genet. 40, 751–760 (2008).
27. Lupski, J.R. Genomic rearrangements and sporadic disease. Nat. Genet. 39, S43–S47 (2007).
28. Murphy, K.C., Jones, L.A. & Owen, M.J. High rates of schizophrenia in adults with velo-cardio-facial syndrome. Arch. Gen. Psychiatry 56, 940–945 (1999).
29. Pulver, A.E. et al. Psychotic illness in patients diagnosed with velo-cardio-facial syndrome and their relatives. J. Nerv. Ment. Dis. 182, 476–478 (1994).
30. Walsh, T. et al. Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia. Science 320, 539–543 (2008).

© 2008 Nature Publishing Group http://www.nature.com/naturegenetics

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

A13: 133

Molecular Psychiatry (2009) 14, 774–785
& 2009 Nature Publishing Group All rights reserved 1359-4184/09 $32.00
www.nature.com/mp

ORIGINAL ARTICLE

Meta-analysis of 32 genome-wide linkage studies of schizophrenia

MYM Ng[1], DF Levinson[2], SV Faraone[3], BK Suarez[4], LE DeLisi[5,6], T Arinami[7], B Riley[8], T Paunio[9,10], AE Pulver[11], Irmansyah[12], PA Holmans[13], M Escamilla[14], DB Wildenauer[15], NM Williams[13], C Laurent[16], BJ Mowry[17], LM Brzustowicz[18], M Maziade[19], P Sklar[20], DL Garver[21], GR Abecasis[22], B Lerer[23], MD Fallin[24], HMD Gurling[25], PV Gejman[26], E Lindholm[27], HW Moises[28], W Byerley[29], EM Wijsman[30], P Forabosco[1], MT Tsuang[31,32], H-G Hwu[33], Y Okazaki[34], KS Kendler[8], B Wormley[8], A Fanous[35,36], D Walsh[37], FA O'Neill[38], L Peltonen[39,40,41,42], G Nestadt[11], VK Lasseter[11], KY Liang[43], GM Papadimitriou[44], DG Dikeos[44], SG Schwab[45,46,47], MJ Owen[13], MC O'Donovan[13], N Norton[13], E Hare[14], H Raventos[48], H Nicolini[49], M Albus[50], W Maier[51], VL Nimgaonkar[52], L Terenius[53], J Mallet[54], M Jay[55], S Godard[56], D Nertney[57], M Alexander[2], RR Crowe[58], JM Silverman[59], AS Bassett[60], M-A Roy[19], C Mérette[19], CN Pato[61], MT Pato[61], J Louw Roos[62], Y Kohn[23], D Amann-Zalcenstein[63], G Kalsi[25], A McQuillin[25], D Curtis[64], J Brynjolfson[65], T Sigmundsson[65], H Petursson[65], AR Sanders[26], J Duan[26], E Jazin[27], M Myles-Worsley[66], M Karayiorgou[67] and CM Lewis[1,68]

[1]King's College London, Department of Medical and Molecular Genetics, London, UK; [2]Department of Psychiatry, Stanford University, Stanford, CA, USA; [3]Departments of Psychiatry and of Neuroscience and Physiology, SUNY Upstate Medical University, Syracuse, NY, USA; [4]Washington University in St Louis, St Louis, MO, USA; [5]Department of Psychiatry, The New York University Langone Medical Center, New York, NY, USA; [6]Nathan S Kline Institute for Psychiatric Research, Orangeburg, NY, USA; [7]Department of Medical Genetics, University of Tsukuba, Tsukuba, Japan; [8]Department of Psychiatry, Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University, Richmond, VA, USA; [9]National Public Health Institute, Helsinki, Finland; [10]Department of Psychiatry, Helsinki University Central Hospital, Helsinki, Finland; [11]Department of Psychiatry & Behavioral Sciences, Johns Hopkins University School of Medicine, Baltimore, MD, USA; [12]Department of Psychiatry, University of Indonesia, Jakarta, Indonesia; [13]Department of Psychological Medicine, School of Medicine, Cardiff University, Cardiff, UK; [14]University of Texas Health Science Center at San Antonio, San Antonio, TX, USA; [15]Center for Clinical Research in Neuropsychiatry, School of Psychiatry and Clinical Neurosciences, University of Western Australia, Perth, WA, Australia; [16]Department of Child Psychiatry, Université Pierre et Marie Curie and Hôpital de la Pitiè-Salpêtrière, Paris, France; [17]Queensland Centre for Mental Health Research and University of Queensland, Brisbane, QLD, Australia; [18]Department of Genetics, Rutgers University, Piscataway, NJ, USA; [19]Department of Psychiatry, Laval University & Centre de recherche Université Laval Robert-Giffard, Québec, QC, Canada; [20]Center for Human Genetic Research, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA; [21]VA Medical Center, Asheville, NC, USA; [22]Department of Biostatistics, University of Michigan School of Public Health, Ann Arbor, MI, USA; [23]Department of Psychiatry, Hadassah-Hebrew University Medical Center, Jerusalem, Israel; [24]Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA; [25]Department of Mental Health Sciences, University College London, London, UK; [26]Center for Psychiatric Genetics, NorthShore University HealthSystem Research Institute and Northwestern University, Evanston, IL, USA; [27]Department of Development & Genetics, Uppsala University, Uppsala, Sweden; [28]University of Kiel, Kiel, Germany; [29]University of California, San Francisco, CA, USA; [30]Departments of Medicine and Biostatistics, University of Washington, Seattle, WA, USA; [31]Center for Behavioral Genomics and Department of Psychiatry, University of California, San Diego, CA, USA; [32]Harvard Institute of Psychiatric Epidemiology & Genetics, Boston, MA, USA; [33]National Taiwan University Hospital and National Taiwan University College of Medicine, Taipei, Taiwan; [34]Tokyo Metropolitan Matsuzawa Hospital, Tokyo, Japan; [35]Washington VA Medical Center, Washington, DC, USA; [36]Department of Psychiatry, Georgetown University Medical Center, Virginia Commonwealth University, Richmond, VA, USA; [37]The Health Research Board, Dublin, Ireland; [38]Department of Psychiatry, Queens University, Belfast, Northern Ireland; [39]Department of Molecular Medicine, National Public Health Institute, Helsinki, Finland; [40]Department of Medical Genetics, University of Helsinki, Helsinki, Finland; [41]The Broad Institute, MIT, Boston, MA, USA; [42]Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Cambridge, UK; [43]Department of Biostatistics, Johns Hopkins University Bloomberg School of Public Health, Baltimore, MD, USA; [44]1st Department of Psychiatry, University of Athens Medical School, and University Mental Health Research Institute, Athens,

Correspondence: Dr CM Lewis, Department of Medical and Molecular Genetics, King's College London, 8th Floor Tower Wing, Guy's Hospital, Great Maze Pond, London SE1 9RT, UK.
E-mail: Cathryn.lewis@kcl.ac.uk
Received 9 October 2008; accepted 11 November 2008; published online 30 December 2008

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Greece; [49]Western Australian Institute for Medical Research, University of Western Australia, Perth, WA, Australia; [46]School of Psychiatry and Clinical Neurosciences, University of Western Australia, Perth, WA, Australia; [47]School of Medicine and Pharmacology, University of Western Australia, Perth, WA, Australia; [48]School of Biology and CIBCM, University of Costa Rica, San Jose, Costa Rica; [49]Carracci Medical Group and Universidad Autonoma de la Ciudad de Mexico, Mexico City, Mexico; [50]State Mental Hospital, Haar, Germany; [51]Department of Psychiatry, University of Bonn, Bonn, Germany; [52]Departments of Psychiatry and Human Genetics, University of Pittsburgh, Pittsburgh, PA, USA; [53]Department of Clinical Neuroscience, Karolinska Hospital, Stockholm, Sweden; [54]Laboratoire de Génétique Moléculaire de la Neurotransmission et des Processus Neurodégénératifs, Centre National de la Recherche Scientifique, Hôpital de la Pitié Salpêtrière, Paris, France; [55]Department of Child Psychiatry, Université Pierre et Marie Curie and Hôpital de la Pitiè-Salpêtrière, Paris, France; [56]INSERM, Institut de Myologie, Hôpital de la Pitiè-Salpêtrière, Paris, France; [57]Queensland Centre for Mental Health Research and University of Queensland, Brisbane, QLD, Australia; [58]Department of Psychiatry, The University of Iowa Carver College of Medicine, Iowa City, IA, USA; [59]Department of Psychiatry, Mount Sinai School of Medicine, New York, NY, USA; [60]Clinical Genetics Research Program, Centre for Addiction and Mental Health and Department of Psychiatry, University of Toronto, Toronto, ON, Canada; [61]Center for Genomic Psychiatry, University of Southern California, Los Angeles, CA, USA; [62]Department of Psychiatry, University of Pretoria, Weskoppies Hospital, Pretoria, Republic of South Africa; [63]Department of Molecular Genetics, Weizmann Institute of Science, Rehovot, Israel; [64]Department of Psychological Medicine, St Bartholomew's and Royal London School of Medicine and Dentistry, London, UK; [65]Department of Psychiatry, General Hospital, Reykjavik, Iceland; [66]Department of Psychiatry, SUNY Upstate Medical University, Syracuse, NY, USA; [67]Departments of Psychiatry and Genetics & Development, Columbia University Medical Center, New York, NY, USA and [68]King's College London, MRC SGDP Centre, Institute of Psychiatry, London, UK

A genome scan meta-a nalysis (GSMA) was carried out on 32 independent genome-wide linkage scan analyses that included 3255 pedigrees with 7413 genotyped cases affected with schizophrenia (SCZ) or related disorders. The primary GSMA divided the autosomes into 120 bins, rank-ordered the bins within each study according to the most positive linkage result in each bin, summed these ranks (weighted for study size) for each bin across studies and determined the empirical probability of a given summed rank (P_{SR}) by simulation. Suggestive evidence for linkage was observed in two single bins, on chromosomes 5q (142–168 Mb) and 2q (103–134 Mb). Genome-wide evidence for linkage was detected on chromosome 2q (119–152 Mb) when bin boundaries were shifted to the middle of the previous bins. The primary analysis met empirical criteria for 'aggregate' genome-wide significance, indicating that some or all of 10 bins are likely to contain loci linked to SCZ, including regions of chromosomes 1, 2q, 3q, 4q, 5q, 8p and 10q. In a secondary analysis of 22 studies of European-ancestry samples, suggestive evidence for linkage was observed on chromosome 8p (16–33 Mb). Although the newer genome-wide association methodology has greater power to detect weak associations to single common DNA sequence variants, linkage analysis can detect diverse genetic effects that segregate in families, including multiple rare variants within one locus or several weakly associated loci in the same region. Therefore, the regions supported by this meta-analysis deserve close attention in future studies.
Molecular Psychiatry (2009) 14, 774–785; doi:10.1038/mp.2008.135; published online 30 December 2008

Keywords: genome; human; humans; schizophrenia/genetics; genetic predisposition to disease; linkage (genetics); meta-analysis

Introduction

We report here on a new genome scan meta-analysis (GSMA)[1–3] of genome-wide linkage scans (GWLS) of schizophrenia (SCZ). We previously published a GSMA of 20 scans that included 1208 pedigrees with 2945 genotyped affected individuals.[4] Here we analyze 32 scans that included 3255 pedigrees with 7413 affected individuals.

Genome-wide association (GWA) study methods have proven more successful than GWLS for common diseases: they detect smaller effects of common single nucleotide polymorphisms (SNPs)[5] as well as some copy number variants (CNVs),[6] and have identified many common disease associations.[7] But an adequately powered GWLS can detect a signal from many kinds of susceptibility variants (common SNPs, multiple rare SNPs, tandem repeats and heritable structural variations) in one or more[8] loci in a region. Meta-analysis can achieve much larger sample sizes than single studies. Consensus linkage regions may deserve further study using methods such as high-throughput resequencing.

Family and twin data[9] suggest that most of the genetic risk to SCZ is conferred by multiple interacting loci, each causing a small increase in risk. As SCZ has similar symptoms and roughly similar prevalence throughout the world,[10] a reasonable hypothesis is that at least some loci have effects in many populations. SCZ linkage findings do not replicate consistently, and although there may be some true positives in small samples from unique populations, in general,

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

power is probably limited by inadequate sample size and differences in ascertainment, marker sets, ancestry and statistical methods. There must also be some measurement error due to limitations in diagnostic methods. Current diagnostic criteria produce the largest known estimates of heritability (twins) and of increased risk to first-degree relatives,[11,12] and there is good diagnostic reliability within and across research groups.[13] But diagnostic judgments are based on data from interview and medical records that vary in completeness and quality. Fortunately diagnostic differences are usually among psychotic disorders that coaggregate in families with similar risks.[14] It is not known whether a further subdivision of cases would increase power.

We present a GSMA of all 32 scan analyses, and a secondary analysis of 22 European-ancestry analyses. The new samples (some of them large) are from European, non-European and isolate populations, but are too diverse to consider a subset other than Europeans. We comment briefly below on an analysis of heterogeneity between Asian and European samples. Statistical significance has been evaluated with permutation tests and simulations. In the primary analysis, 2 bins met empirical genome-wide criteria for suggestive linkage; and a set of 10 bins (in eight chromosomal regions) met an empirical aggregate criterion for genome-wide significant linkage, indicating that loci linked to SCZ are present in some or all of these bins. Five of these bins were consistent with our previous report.[4] When bin boundaries were shifted to the middle of the previous bins, one genome-wide significant result was observed on chromosome 2q, consistent with our previous report.[4]

Materials and methods

Study selection

We identified GWLS of SCZ (in addition to those studied previously[4]) through PubMed literature searches, conference abstracts and personal contacts with investigators, selecting the largest analysis for each study. Study characteristics are summarized in Table 1. Investigators of each of these studies agreed to contribute data to this analysis. All scans ascertained probands with SCZ by one of three closely related sets of diagnostic criteria. Most scans included relatives with SCZ and those with one or all subtypes of schizoaffective disorder (SA). A few scans included only SCZ cases, or included broader diagnoses in alternative models. Scans were included if informative families ascertained through a SCZ case had at least 30 genotyped SCZ and SA cases. We included only primary genome-wide analyses with few gaps in marker coverage, and did not consider subsequent 'fine mapping' with additional markers or pedigrees in selected regions. We excluded a small Utah study[48] whose RFLP markers were difficult to place on the current map; a study of a large Costa Rican pedigree[49] with too few affected cases; a small study from an Italian isolate population[50] that combined SCZ and bipolar cases; and a study of isolated villages in Daghestan[51] that used a very broad phenotype definition. We lacked detailed results for the Icelandic study of Stefansson et al.[52] We included two earlier studies of Icelandic families for which we had complete results,[15,20] and discuss below a secondary analysis substituting results from Stefansson et al. inferred from published graphs.

A companion paper[38] reports on a combined analysis of studies 24–31 (Table 1). Older data for all or parts of six of those data sets were included as separate scans in our previous GSMA, so all eight are treated separately here using the newer data.[38] To eliminate sample overlap, investigators provided us with data for NUH (study 28) that excluded families that were also in the US/International study (study 9); and data for sample 15 (US/Sweden) without families in sample 26 (Cardiff). The investigators determined from genotypes that there were no overlapping families in samples 10 and 22.

There were 14 studies of European-ancestry pedigrees (including partial isolates such as Finland and Ashkenazim); 8 of both European and non-European families; and 10 non-European samples, including 3 Asian (Japanese, Han Chinese and Indonesian), 2 Latino, 1 Arab, 2 Pacific Island isolates (Palau and Kosrae) and 2 entirely (study 20) or predominantly (study 6) African-American samples. Study 6 (Texas) included a few European-ancestry families for which we lacked separate data. The ALL analysis included all 32 samples; the EUR (European-ancestry) analysis included 22 samples or subsamples (Table 1). Data for chromosome X markers were available for 23 studies, so this chromosome was considered in a secondary analysis.

Linkage scores were obtained from investigators or internet postings. Data for each study were rank ordered based on the designated 'primary' analysis, and if this included more than one test or model, we took the most positive linkage score across tests for each chromosomal bin. From our previous GSMA,[4] data were carried over unchanged for studies 1–10 in Table 1; the Utah study[48] was omitted as discussed above; and nine studies were updated with new genotyping results, some with larger samples. Studies 13 and 24–31 used SNP markers from the Illumina version 4 panel; all other studies used microsatellite markers at approximately 8–10 cM density.

Marker mapping

Marker locations were determined from the Rutgers combined linkage map (Build 36),[53] or interpolated based on physical position, other maps or flanking markers. Chromosomes were divided into bins of approximately equal genetic widths—120 30-cM bins (our primary analysis), or (for secondary analyses) 179 20-cM bins and 88 40-cM bins. Six additional 30-cM bins were defined on the X chromosome (female map length). If a study had no marker in a bin (for example,

© University of Pretoria

A13: 136

Table 1 Genome-wide linkage scan characteristics (ALL and European-ancestry analyses)

	Study site or project[a]		ALL analysis		EUR analysis					
	(previous(p)/modified(m)/new(n))	Diagnostic criteria[b]	Ped	Aff	Ped	Aff	Markers	Program	Statistics	Ethnicity[c]
1	Kiel[15] (p)	SCZ, SA (RDC or D3R)	5	37	5	37	413	WRPC	P-values	Icelandic
2	Palau[16] (p)	SCZ, SAS (RDC)	5	33			496	LINKAGE	LOD	Palau
3	Canada[17] (p)	SCZ, SA (D3R)	22	79	22	79	381	FASTLINK	HLOD	Celtic Canada
4	Finland—isolate[18] (p)	SCZ; þ SA (D4)	53	117	53	117	345	MLINK	LOD	Finnish
5	Finland—national[18] (p)	SCZ; þ SA (D4)	185	394	185	394	345	MLINK	LOD	Finnish
6	Texas[19] (p)	SCZ (D4)	27	66			406	Genehunter Plus	NPL	AA, few EA
7	UC London[20] (p)	SCZ,SA,UFP (RDC)	13	56	13	56	365	MFLINK	P-values	Icelandic
8	North Sweden[21] (p)	SCZ; þ SAD; þ PNOS; þ SAB (D4)	1	43	1	43	371	LINKAGE	LOD	Swedish
9	US/International[22] (p)	SCZ, SA (D3R)	294	669	294	669	396	Mapmaker/Sibs	MLS	EA (inc Chile)
10	Costa Rica[23] (p)	SCZ, SA (D4)	60	157			404	Genehunter Plus	KCLOD	Costa Rican
11	Kosrae[24] (n)	SCZ, SA (D4)	9	30			398	FASTLINK	LOD	Kosrae
12	Israeli Arab[25] (n)	SCZ, SAD (D4)	21	59			350	Allegro	NPL	Arab
13	Japan[26] (n)	SCZ, SA (D4)	236	488			5861	Merlin	KCLOD	Japanese
14	Ashkenazi[27] (n)	SCZ, SA (D4)	29	59	29	59	382	Genehunter	LOD	Ashkenazi
15	US/Sweden[28] (m)	SCZ, SA (D4)	135	198	135	198	392	Mapmaker/Sibs	MLS	Swedish, US
16	Portugal[29] (n)	SCZ, SAD (D4)	28	67	28	67	366	Genehunter	NPL	Portuguese
17	South Africa[30] (n)	SCZ, SAD (D4)	34	62	34	62	388	Merlin	KCLOD	Afrikaner
18	Eastern Quebec[31] (n)	SCZ (D3R)	15	71	15	71	350	FASTLINK	LOD	Quebecois
19	VA (EA)[32] (n)	SCZ, SAD (D3R)	61	135	61	135	414	Merlin	KCLOD	EA
20	VA (AA)[32] (n)	SCZ, SAD (D3R)	60	143			414	Merlin	KCLOD	AA
21	MGS[33] (n)	SCZ, SA (D4)	409	924	263	571	400	Genehunter Plus	Zlr	EA, AA
22	US/Mexico/Central America[34] (n)	SCZ, SA (D4)	96	240			404	Merlin	KCLOD	Mexican,Cent Am
23	Taiwan[35] (n)	SCZ (D4)	557	1207			286	Merlin	Zlr	Han Chinese
24	Aust/US[36–38] (m)	SCZ, SA (D3R)	59	143	49	114	5868	Allegro	Zlr	EA, AA, Other
25	UWA/Bonn[38,39] (m)	SCZ, SA (D3R)	96	212	96	212	5868	Allegro	Zlr	German,Israeli
26	Cardiff[28,38] (m)	SCZ, SA (D3R)	113	239	113	239	5868	Allegro	Zlr	UK
27	France/LaRéunion[38,40,41] (n)	SCZ, SA (D3R)	63	186	29	75	5868	Allegro	Zlr	EA, AA, Other
28	NUH[38,42] (n)	SCZ, SA (D3R)	20	42	19	40	5868	Allegro	Zlr	EA, AA
29	Johns Hopkins[38,43] (m)	SCZ, SA (D3R)	133	317	124	295	5868	Allegro	Zlr	EA, AA, Other
30	NIMH[38,44,45] (m)	SCZ, SA (D3R)	106	261	61	142	5868	Allegro	Zlr	EA, AA, Other
31	VCU/Ireland[38,46] (m)	SCZ, SAPO (D3R)	184	417	184	417	5868	Allegro	Zlr	Irish
32	UWA/Indonesia[38,47] (n)	SCZ (D3R)	124	267			400	Genehunter	MLS	Indonesian
		Total	3283	7476	1835	4141				

[a]Sorted by year of publication of the data used in this analysis. The letter in parentheses indicates whether the dataset was included in the previous analysis (p), has been modified with new genotyping and/or additional pedigrees (m), or is new (n).

[b]D3R = DSM-IIIR criteria; D4 = DSM-IV criteria; RDC = Research Diagnostic Criteria; SA = schizoaffective disorder (SAS = RDC mainly schizophrenic type; SAD = depressed type; SAB = bipolar type; SAPO = poor outcome; PNOS = psychosis not otherwise specified; UFP = unspecified functional psychosis (RDC). Diagnoses separated by commas constitute one model; sets separated by semicolons are alternative models over which scores were maximized, with 'þ' indicating that diagnoses were added to the ones in the next narrower model.

[c]EA = mixed European ancestry (usually from the US, but also mixed US, European and Australian samples); AA = African American.

Other abbreviations: Ped = informative pedigrees; Aff = genotyped affected cases. VA = US Veterans' Administration; MGS = Molecular Genetics of Schizophrenia collaboration; Aust/US = Australia/US collaboration; NUH = NorthShore University HealthSystem Research Institute; NIMH = National Institute of Mental Health Schizophrenia Genetics Initiative; VCU = Virginia Commonwealth University; UWA = University of Western Australia. LOD = log of the odds ratio under one or more parametric models; HLOD = heterogeneity LOD scores (generally, maximized over at least two models); NPL = nonparametric linkage Z-score; MLS = maximum LOD score (affected sib-pair analysis, constrained to possible triangle); Zlr = Z-likelihood ratio statistic; KCLOD = Kong-Cox equivalent LOD score computed from Zlr.

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

the narrower 20 cM bins), we used the mean linkage score for the two flanking markers, or the single closest marker for an empty telomeric bin. Bin boundaries in cM and Mb are listed in online Supplementary Table 1.

Statistical analysis

GSMA is a nonparametric method to combine data generated with different maps and statistical tests.[1–3] A GSMA divides the genome into approximately equal-width bins (N_{bins}), labeled ch.K, where ch = chromosome and K = bin number (that is, 1.4 is the fourth bin of chromosome 1). For each study, bins are ranked by their highest LOD, NPL or Z score or minimum P-value. Ranks for each sample was weighted by the square root of the number of genotyped affected cases, and then, for comparison, analyzed without weights (results of unweighted analyses are provided in the online Supplementary Materials). Note that '120' was the best rank, and higher summed ranks indicate stronger evidence for linkage (whereas in our previous report,[4] '1' was the best rank in a study and we reported averaged rather than summed ranks).

We used GSMA software[2] to evaluate nominal (bin-wise) significance by randomly permuting the ranks for bins within each study and then resumming the ranks across studies to determine the empirical summed rank probability (P_{SR})—the nominal probability that a bin would achieve or exceed the observed summed rank under the null hypothesis of no linkage in the bin. We also determined the empirical ordered rank P-value (P_{OR}—the probability that the kth highest summed rank achieves or exceeds an observed summed rank under the null). Thresholds for genome-wide significant and suggestive evidence for linkage[54] were determined empirically (see below). For the primary analysis, these thresholds were $P_{SR} < 0.00037$ and $P_{SR} < 0.0077$ respectively.

Strong linkage signals can cover a wide genetic region in complex disorders, and often produce high summed ranks in adjacent bins,[1] but GSMA can miss a weak linkage signal near the boundary of two bins. Therefore, regions showing suggestive evidence for linkage were reanalyzed using 20 and 40 cM bin widths, and using 30 cM bins starting at the midpoint of the bins used in the primary analysis (Supplementary Figure 1).

We tested for heterogeneity in study ranks between three outbred Asian studies (Taiwan, Japan, Indonesia) and the 22 EUR studies, using the nonparametric Wilcoxon rank sum test. We did not test for heterogeneity arising from an arbitrary set of studies;[55] these tests detect bins where study ranks have a wider spread than expected by chance, but they have low power and do not identify the study subgroups contributing to the heterogeneity.[56] We also carried out sensitivity analyses, dropping one study at a time and analyzing the remaining studies with GSMA (weighted and unweighted analyses, 30 cM bins).

Simulation study of power and significance thresholds

We performed a simulation study to determine: (1) power; (2) permutation-based estimates of type I error and (3) aggregate criteria for genome-wide significance. Details are provided in Supplementary Online Materials and in a previous report.[1] Briefly, genome-wide data were simulated for affected sibling pairs under the assumption of no linkage or assuming a range of genetic models (1–10 linked bins with locus-specific sibling relative risks (λ_S) of 1.15 or 1.3, at the edge or near the middle of chromosomes), and 1000 replicates of the 32-sample data set (or of alternative subsets of the data) were created by replacing each sample with a number of ASPs with roughly comparable linkage information.

For estimates of type 1 error, we determined empirical bin-wise thresholds for genome-wide suggestive linkage (the nominal P-value observed on average once per GSMA replicate) or genome-wide significant linkage (the nominal P-value observed on average in 5% of GSMA replicates).

We also determined criteria for aggregate genome-wide significant linkage, that is, that the total number of bins achieving a specified nominal P-value threshold was greater than that observed in 5% of genome-wide replicates with no linkage present. This criterion was determined for ALL and EUR studies by establishing a P-value threshold, P_0, for which exactly 5% of unlinked replicates had N or more bins with $P_{SR} < P_0$. P_0 was initially set at 0.05, and the number of unlinked replicates with XN bins achieving $P_{SR} < P_0$ was tabulated (for N = 1, 2, 3,y). P_0 was then reduced incrementally until exactly 5% of simulations had at least N bins with $P_{SR} < P_0$, for some value N. For example, for the primary weighted analysis of 32 studies, the 5% genome-wide threshold by this procedure was 10 bins with $P < 0.046$; for EUR studies, it was 10 bins with $P < 0.048$.

For power calculations, we determined how often, on average, the linked bins exceeded the P_{SR} threshold for ALL and for EUR studies, and, to investigate heterogeneity, for ALL studies assuming that linkage arose only in EUR studies.

Results

Evidence for linkage (ALL and EUR studies)

Figure 1 illustrates P-values for all bins in the ALL and EUR analyses, and the figure legend states the empirical P-value thresholds for each study. For ALL studies (3255 pedigrees, 7413 genotyped cases), Table 2 lists the 10 bins (in eight nonadjacent chromosomal regions) that achieved $P_{SR} < 0.05$ in the primary analysis (weighted, 30 cM bins). These 10 bins met the empirical aggregate criterion for genome-wide significant linkage (10 bins with nominal $P_{SR} < 0.046$). Suggestive evidence for linkage was observed for single bins on chromosome 5q (bin 5.6, $P_{SR} = 0.0046$) and chromosome 2q (bin 2.5, $P_{SR} = 0.0075$). Figure 1 illustrates the results for all bins (ALL and EUR analyses), and Figure 2 illustrates

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

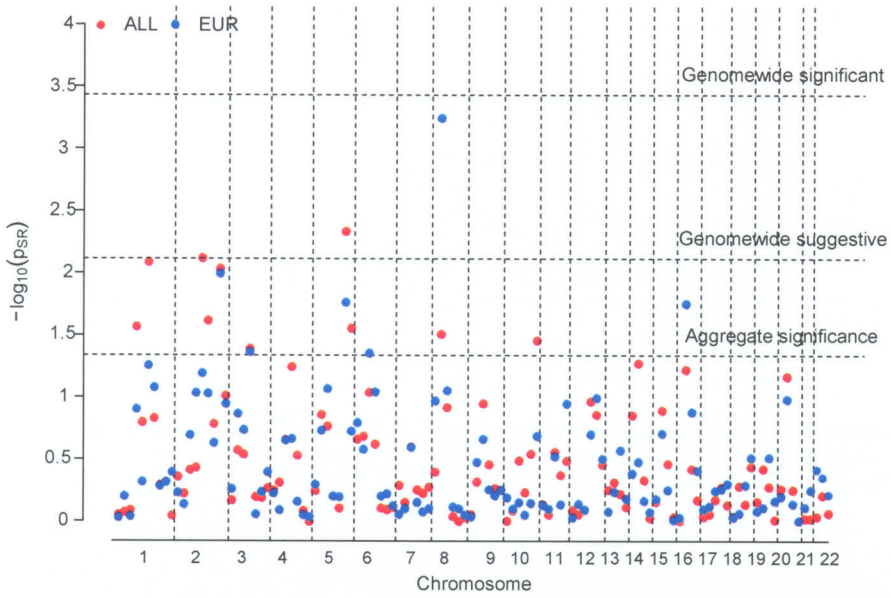


Figure 1 Genomewide results for primary analysis (weighted, 30 cM bins) for ALL studies (red) and EUR studies (blue). The two upper dashed lines represent empirical thresholds for genomewide significant (theoretical $P_{SR} = 0.00037$) and suggestive (theoretical $P_{SR} = 0.0077$) evidence for linkage for a single bin for ALL studies; and the lower dashed line represents the threshold for aggregate significance for ALL studies (theoretical $P_{SR} = 0.046$ in 10 or more bins). The thresholds for EUR studies were slightly less stringent (P_{SR} values of 0.00044, 0.0078 and 0.0475 (in 10 or more bins) respectively), but the Figure shows all EUR bins within the region for their correct threshold of significance.

Table 2 Bins achieving nominally significant evidence for linkage (ALL studies)

Pos	Bin	cM	Mb	P_{SR}	P_{OR}
1	5.6*	148.9–178.7	141.8–167.7	0.00459	0.43156
2	2.5*	117.5–146.9	103.3–134.0	0.00755	0.22798
3	1.6*	143.1–171.7	114.6–162.1	0.00814	0.06639
4	2.8	205.7–235.1	206.3–228.3	0.00916	0.01970
5	2.6*	146.9–176.3	134.0–169.9	0.02395	0.14269
6	1.4	85.8–114.5	57.3–84.6	0.02692	0.08449
7	5.7	178.7–208.5	167.7–180.4	0.02760	0.03430
8	8.2*	28.1–56.2	15.7–32.7	0.03086	0.02070
9	10.6	145.9–175.0	123.1–135.1	0.03487	0.01350
10	3.4	95.9–127.9	71.6–120.2	0.04047	0.01140

Shown are the bins with nominally significant evidence for linkage ($P_{SR} < 0.05$), for ALL studies (30 cM bins, weighted analysis), ordered by their position (1 = best) according to weighted analysis P_{SR} (summed rank P-value). P_{OR} : ordered rank P-value. Bold P_{SR} values achieved an empirical threshold for suggestive evidence for linkage ($P_{SR} < 0.0077$). No single bin achieved an empirical genome-wide significant threshold (0.00037).

Empirically (see text), 10 or more bins with nominal $P_{SR} < 0.046$ was observed in 5% of simulated replicates of this dataset in the absence of linkage. Therefore, it is likely that at least some of the bins listed here contain loci linked to SCZ. Asterisks indicate which regions were consistent with our previous GSMA[4] of 20 schizophrenia scans (i.e., bins that overlap with those in the previous analysis that met aggregate significance criteria—that analysis used the Marshfield rather than the Rutgers map, so bin boundaries were not identical). See Figure 2 for further details.

the relative ranks of bins in each study. (Results for all analyses are available in Online Supplementary Tables; ranks for each study are available online at www.kcl.ac.uk/mmg/ngdata.html.)

Table 3 summarizes results for 22 EUR samples or subsamples (1813 pedigrees, 4094 genotyped cases). Suggestive evidence for linkage was observed on chromosome 8p ($P_{SR} = 0.00057$, close to the genome-wide threshold for 22 EUR studies of 0.00044), with a further five bins in four different regions achieving nominal significance.

On chromosome X, there were no nominally significant bins for ALL studies; there was one such bin (X.5, 130–162 cM, 119–141 Mb, $P_{SR} = 0.0247$) for EUR studies.

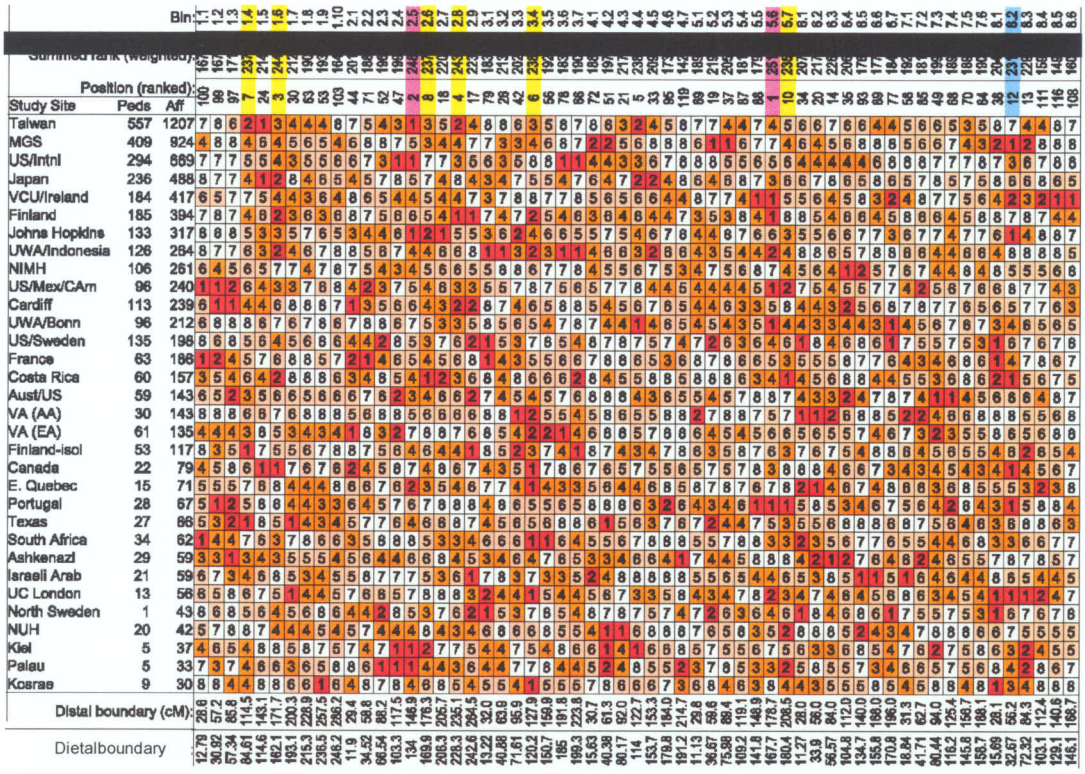
When estimated ranks (interpolated from published graphs) for the Icelandic study of Stefansson et al.[52] were substituted for the results of Gurling et al.[20] and Moises et al.,[15] results were similar to those shown in Table 1 except that bin 3.4 failed to achieve nominal significance. For EUR studies, results again were similar except that bins 8.2 and 2.8 both achieved suggestive significance.

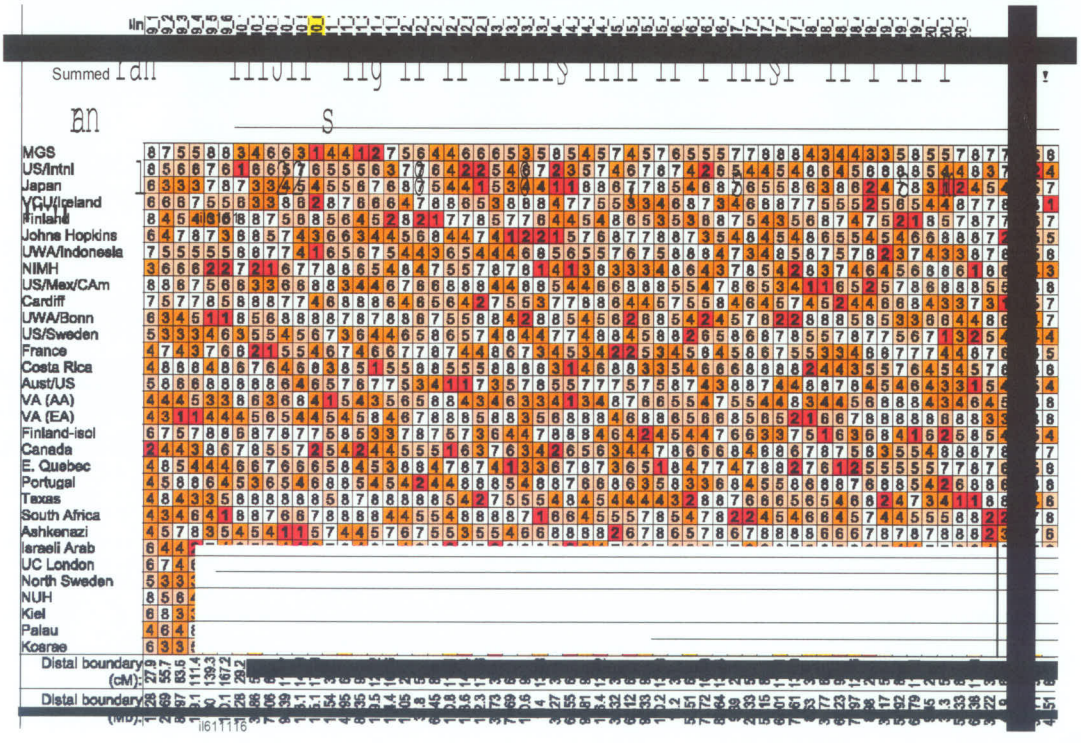
Effect of bin width

Online Supplementary Figures 2–4 illustrate results of analyses using four alternative bin widths (20, 30 and 40 cM, and 30 cM bins shifted by 50% for chromosomes with suggestive evidence for linkage (chromosomes 2 and 5 for ALL, chromosome 8 for EUR). On chromosome 2, genome-wide significance was obtained for the shifted 30 cM bin ($P_{SR} = 0.00035$,

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA





© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

132.2–161.6 cM, 118.7–152 Mb), and suggestive linkage evidence was seen in the adjacent bin (102.8–132.2 cM, 84.9–118.7 Mb). For the region of chromosome 5q that produced suggestive evidence for linkage in the primary analysis, the signal was similar regardless of bin width. On chromosome 8, using 20 cM bins extended the signal by approximately 5 cM in both directions; if linkage is present, this pattern could indicate a broader signal (multiple loci) or the effects of a single strong signal on adjacent bins.

Heterogeneity and sensitivity analyses

In tests of differences between 22 EUR vs 3 Asian studies, none of the 120 bins exceeded nominal $P < 0.01$, and only 4 bins achieved nominal $P < 0.05$, thus, no significant difference was observed. The most significant evidence for heterogeneity between the Asian and European studies was in bin 1.5, but this did not reach the suggestive threshold (online Supplementary Figure 5). Results of sensitivity analyses, omitting each study in turn, are shown in online Supplementary Figures 6 and 7. The largest changes: bin 8.2 (8p) achieves the suggestive linkage threshold without the Taiwan data set; bin 2.8 (2q) becomes substantially more significant (but not achieving genome-wide significance) without the Japan data set; bin 5.6 (5q) becomes substantially more significant (but not genome-wide significant) without the Suarez et al. data set (study 21); and bin 6.3 (6pq) would join the list of nominally significant bins without either the Taiwan or Japan data set.

Table 3 Bins achieving nominally significant evidence for linkage (European studies)

Pos	Bin	cM	Mb	P_{SR}	P_{OR}
1	8.2*	28.1–56.2	15.7–32.7	0.00057	0.06659
2	2.8	205.7–235.1	206.3–228.3	0.01016	0.35077
3	5.6*	148.9–178.7	141.8–167.7	0.01718	0.33717
4	16.2*	33.3–66.7	13.2–51.5	0.01775	0.15808
5	3.4	95.9–127.9	71.6–120.2	0.04359	0.62724
6	6.3*	56.0–84.0	33.9–56.6	0.04433	0.44626

Shown are bins with nominally significant evidence for linkage ($P_{SR} < 0.05$), in 22 samples or subsamples of European-ancestry families (see Table 1); the bolded p-value met empirical criteria for suggestive evidence for linkage ($P_{SR} < 0.0078$). The results here did not meet aggregate criteria for significant genome-wide evidence for linkage in this subset of the data (10 bins with nominal $P_{SR} < 0.0475$). See legend for Table 2 for additional details.

Figure 2 Ranks by study and summed ranks. Shown for each study is the within-study rank of each bin. '1' against a red background indicates bins with ranks of 116–120 (the bin containing the most significant linkage score in a study is ranked 120); 2 indicates ranks 111–115; 3—ranks 101–110; 4—ranks 81–100; 5—ranks 61–80; 6—ranks 41–60; 7—ranks 21–40; and 8—ranks 1–20 (cells with 7 and 8 have a white background). Shown at the top of each part of the figure are the bin number, the weighted summed rank for each bin across studies (i.e., weighted for each sample by the square root of the N of genotyped affected individuals), and the rank-ordered position of that bin in the ALL analysis (i.e., 1 is the best position, see details in Table 2). Purple background indicates bins with empirical suggestive evidence for linkage in the ALL analysis; 10 bins shaded in purple, yellow or blue met the threshold for aggregate genome-wide evidence for linkage (10 bins with nominal $P < 0.046$); and bin 8.2 (shaded blue) also met empirical suggestive evidence for linkage in the EUR analysis. Shown at the bottom is the distal boundary of each bin in Rutgers cM and in Mb (build 36). Note that tied ranks can result in uneven numbers of bins in a grouping, particularly for lower ranks when there are many zero or negative scores. See Table 1 for the references associated with each study.

Type I error

Type I error was calculated for the unlinked GSMA simulations (120 000 30-cM bins) by determining the number of observed bins in which the GSMA program computed P_{SR} values (based on permutation alone) that were less than the theoretical threshold for nominal significance (0.05); for genome-wide suggestive linkage, or a value expected once per GSMA ($1/120 = 0.00833$) or genome-wide significant linkage ($0.05/120 = 0.00042$). The GSMA program's permutation procedure produced type I errors that were slightly liberal for ALL (0.053, 0.0096 and 0.00051) and less liberal for EUR studies (0.051, 0.0091 and 0.00042). The source of the discrepancy is not clear, but we have used the permutation-based suggestive and significant values here for single bins (0.0077 and 0.00037 for ALL, and 0.0078 and 0.00044 for EUR studies) and an empirical threshold for aggregate genome-wide significance as described above.

Power

Table 4 summarizes the results of power calculations from simulation studies. Power is lower for edge bins for multipoint analyses, so we computed a weighted average of power assuming 20% edge and 80% mid-chromosome bins (see Table 4 legend). Power was excellent to detect significant linkage at multiple loci with observed population-wide $l_S = 1.3$ (30% increased risk in sibs) across ALL studies or in EUR studies alone. For $l_S = 1.15$, there was excellent power to detect significant linkage in multiple bins (for ALL studies), or to detect suggestive evidence for linkage in multiple bins (if limited to EUR studies); there was reduced power in the presence of substantial heterogeneity. Genetic effects limited to a defined subset (EUR) were more readily detected in a separate analysis.

Discussion

Meta-analysis of 32 GWLS, encompassing 7413 individuals affected with SCZ or related disorders in 3255 pedigrees, produced evidence for significant linkage in the genome based on empirical aggregate criteria: 10 30-cM bins that achieved bin-wise

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Table 4 Power to detect linkage in simulation studies (weighted GSMA)

	Level of significance								
	Nominal			Suggestive			Genome-wide		
Number of linked bins:	1	5	10	1	5	10	1	5	10
$l_S = 1.3$									
ALL studies	1.000	1.000	1.000	1.000	1.000	0.998	1.000	0.996	0.975
European samples	1.000	0.998	0.993	1.000	0.992	0.953	0.950	0.912	0.745
ALL, Heterogeneity[a]	0.988	0.952	0.918	0.898	0.834	0.705	0.600	0.474	0.325
$l_S = 1.15$									
ALL studies	1.000	1.000	0.989	0.996	0.980	0.932	0.904	0.850	0.705
European samples	0.970	0.948	0.908	0.874	0.822	0.697	0.598	0.468	0.323
ALL, Heterogeneity[a]	0.864	0.774	0.740	0.592	0.546	0.403	0.186	0.180	0.116

Shown is the power of weighted GSMA to detect linkage in the simulation study of samples of affected sibling pairs with information approximately comparable to the 32 studies in the schizophrenia GSMA (120 30-cM bins). l_S = relative risk to siblings of case probands vs population risk. For 1 linked bin per simulated genome, power is a weighted average assuming that 20% of bins lie at a chromosome end ('edge,' see text), and 80% of are mid-chromosomal ('mid'). Simulations with 5 linked loci included 1 edge and 4 mid bins; those with 10 linked included 2 edge and 8 mid bins.
[a]Heterogeneity analyses modelled cases where all samples were included, but European samples were linked as shown while non-European studies were unlinked.

$P_{SR} < 0.046$. Genome-wide evidence for linkage was detected on chromosome 2q (118.7–152 Mb) in a secondary analysis that shifted bin location by 50%, consistent with our previous finding of significant linkage in the same region in our previous GSMA of 20 studies.[4] Suggestive evidence for linkage was detected in ALL studies in two single bins, on chromosomes 5q and 2q. Separate analysis of European-ancestry samples produced linkage evidence falling just short of the genome-wide significance in bin 8.2 (15.7–32.7 Mb).

It is widely assumed that many loci contribute to SCZ susceptibility. Table 4 shows that for the GSMA method, power to detect at least one true linkage goes up with the number of linked loci, but the power to detect any specific locus decreases, because in a rank-ordered test there is a finite number of high ranks in each study. But if there were many loci with locus-specific l_S values of 1.3 or even 1.15 (averaged across all study populations and accounting for the effects of various sources of measurement error), we would expect to detect genome-wide significant linkage in several individual bins, and suggestive evidence for linkage in many more than two bins. It would be reasonable to conclude that there are probably no loci with such strong effects worldwide or across all European populations, although they might exist within particular populations for which samples of this size are not available or within subsets of families, which we cannot yet identify in advance.

The power of linkage analysis for a locus is predicted by its contribution to the relative risk (RR) to siblings of probands (l_S),[57] whereas the power of association analysis is predicted by allelic or genotypic RRs, that is, the increased risk to individual carriers.[5] GWA studies have produced robust association findings for many complex genetic disorders[58–60] typically with RRs of 1.1–1.5. A variant found on 10% of chromosomes conferring a 'large' RR of 1.5 per allele in a multiplicative model (2.25 for homozygous carriers) would produce a population-wide sibling RR of only 1.02 (undetectable by linkage). But GWA chips are unlikely to tag all pathogenic variants. For example, in NOD2 there are three rare variants, each with a frequency of less than 5%, that confer moderate risks of Crohn's disease. Their individual effects generate only weak signals when assayed by the Affymetrix 500K chip because they are poorly tagged; but combined, they confer a sibling RR of 1.16, which is detectable in a very large linkage sample.[58,61] This illustrates a strength of linkage analysis: allelic heterogeneity is difficult to detect in association studies, as the genetic effect is split across variants, but linkage can detect the pooled effect of all variants (not only SNPs) within one or more susceptibility genes or elements in a region.

Readers may be interested to know the location of SCZ candidate genes or regions in relation to bins that produced evidence for linkage here. On chromosome 1q, NOS1AP (160 Mb; previously known as CAPON),[62,63] RGS4 (161.3 Mb)[64,65] and UHMK1 (160.7 Mb)[66] are near the telomeric edge of bin 1.6 (114.6–162.1 Mb), and the rare (usually de novo) SCZ-associated CNV is located in a different part of the same bin (145–146 Mb).[67,68] On chromosome 8p, PPP3CC (22.35 Mb)[69] and NRG1 (31.6–32.7 Mb)[52,70] are both within bin 8.2 (15.7–32.7 Mb). On chromosome 2q, ZNF804A, which contains a SNP that produced genome-wide significant evidence for association in a combined SCZ-bipolar disorder sample,[71] is in bin 2.7 at 185 Mb. This bin was not nominally

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

significant in the primary analysis, but 185 Mb is within nominally significant bins in two secondary analyses (one that shifted bin boundaries by 50% and one that considered 40 cM bin widths). DTNBP1, DAOA, TRAR4/STX7, CHRNA7, COMT/ARVCF, DISC1, AKT1, HTR2A and DRD2 are not in bins listed in Tables 2 or 3. Many of the research groups whose samples are included here are actively investigating association of SCZ to sequence variants in genes identified in linkage regions or by the more recent genome-wide association methodology.

In conclusion, this updated GSMA of 32 SCZ studies showed significant evidence for linkage on chromosome 2q in a secondary analysis, suggestive evidence for linkage on chromosomes 5q and 2q in the primary analysis, and suggestive evidence for linkage on 8p in European samples. Genome-wide significant aggregate evidence for linkage (that is, more modestly significant results than expected by chance) was observed based on results for 10 bins in 8 nonadjacent chromosomal regions. As results from genome-wide association studies emerge, it may be important to keep in mind that there may be SCZ susceptibility loci in some of these consensus linkage regions that cannot be detected by common tag SNPs. It is likely that diverse methods will be required to identify specific DNA sequence variants and to confirm and define their role in the etiology of SCZ.

Acknowledgments

The work reported here was supported by: Medical Research Council (UK) Grants G0400960 (CML) and G9309834 (MO); National Institute of Mental Health Grants 7R01MH062276 (to DFL [Aust/US], CL [France/La Réunion], MO [Cardiff] and DW [Bonn]), 5R01MH068922 (to PG [ENH]), 5R01MH068921 (to AEP [Johns Hopkins]), 5R01MH068881 (to BR [VCU/Ireland]), MH-41953 (to KSK [VCU/Ireland]); MH63356 and MH80299 (to WB [Palau]); MH58586 (to JMS [Aust/US]); MH 56242 (to VLM [US/Sweden]), MH61399 (EMW, MK [Kosrae, South Africa]), NIMH Grant MH062440, Canadian Institutes of Health Research Grants MOP-53216 and MOP-12155, a National Alliance for Research on Schizophrenia and Depression Distinguished Investigator Award and Canada Research Chair in Schizophrenia Genetics (LMB, ASB [Canada]); Australian National Health and Medical Research Council Grants 910234, 941087, and 971095 (to BJM [Aust/US]), MRC project Grant G880473N, The European Science Foundation, SANE, the Iceland Department of Health, the General Hospital Reykjavik, the Joseph Levy Charitable Foundation, the Wellcome Trust Grant 055379, The Priory Hospital, the Neuroscience Research Charitable Trust, the University of Iceland and the Icelandic Science Council (HMDG [UCL]); Warner-Lambert, Parke-Davis Pharmaceuticals Company and NIMH Grant R01-MH44245 (LEL [US/International]); the Deutsche Forschungsgemeinschaft (HWM [Kiel]; MA WM, SGS, DBW [Indonesia]); the German Israeli Foundation for Scientific Research (BL; DBW); Mammalian Genotyping Service HV48141 (DBW [Indonesia]; CREST of JST (Japan Science and Technology Agency) TA [Japan]; Pfizer, Inc. and the SANE Foundation (LEL [Costa Rica]); the Israel Science Foundation, US. Israel Binational Science Foundation, the National Alliance for Research on Schizophrenia and Depression, and the Harry Stern Family Foundation (BL and YK [Israel]); the VA Merit Review Program (AF); recruitment of the NIMH Genetics Initiative sample was supported by NIMH Grants 5 UO1MH46318, UO1MH46289 and UO1MH46276; the Taiwan Schizophrenia Linkage Study was supported by NIMH Grant 1R01 MH59624-01 and Grant NHRI-90-8825PP, NHRI -EX91,92-9113PP from the National Health Research Institute, Taiwan, and support from the Genomic Medicine Research Program of Psychiatric Disorders, National Taiwan University Hospital; the VA Linkage Study was supported by funds from the Department of Veterans Affairs Cooperative Studies Program; the US/Mexico/Central America study was supported by a collaborative NIMH grant ('Genetics of Schizophrenia in Latino Populations') (MH60881 and MH60875) to ME [University of Texas Health Science Center at San Antonio], R Mendoza [University of California at Los Angeles-Harbor], HR [University of Costa Rica, San Jose, Costa Rica], A Ontiveros [Instituto de Informacion de Investigacion en Salud Mental, Monterrey, Mexico], HN [Medical and Family Research Group, Carracci SC, Mexico City, Mexico], and R Munoz [Family Health Centers of San Diego, CA].

References

1 Levinson DF, Levinson MD, Segurado R, Lewis CM. Genome scan meta-analysis of schizophrenia and bipolar disorder, part I: Methods and power analysis. Am J Hum Genet 2003; 73: 17–33.
2 Pardi F, Levinson DF, Lewis CM. GSMA: software implementation of the genome search meta-analysis method. Bioinformatics 2005; 21: 4430–4431.
3 Wise LH, Lanchbury JS, Lewis CM. Meta-analysis of genome searches. Ann Hum Genet 1999; 63(Part 3): 263–272.
4 Lewis CM, Levinson DF, Wise LH, DeLisi LE, Straub RE, Hovatta I et al. Genome scan meta-analysis of schizophrenia and bipolar disorder, part II: Schizophrenia. Am J Hum Genet 2003; 73: 34–48.
5 Risch N, Merikangas K. The future of genetic studies of complex human diseases. Science 1996; 273: 1516–1517.
6 McCarroll SA, Altshuler DM. Copy-number variation and association studies of human disease. Nat Genet 2007; 39(7 Suppl): S37–S42.
7 Manolio TA, Brooks LD, Collins FS. A HapMap harvest of insights into the genetics of common disease. J Clin Invest 2008; 118: 1590–1605.
8 McMillan I, Roberson A. Power of detection of major genes affecting quantitative characters. Heredity 1974; 32: 349–356.
9 Risch N. Genetic linkage and complex diseases, with special reference to psychiatric disorders. Genet Epidemiol 1990; 7: 3–16; discussion 17–45.
10 Jablensky A, Sartorius N, Ernberg G, Anker M, Korten A, Cooper JE et al. Schizophrenia: manifestations, incidence and course in different cultures. A World Health Organization ten-country study. Psychol Med Monogr Suppl 1992; 20: 1–97.
11 Cardno AG, Marshall EJ, Coid B, Macdonald AM, Ribchester TR, Davies NJ et al. Heritability estimates for psychotic disorders: the

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Maudsley twin psychosis series. Arch Gen Psychiatry 1999; 56: 162–168.
12 Farmer AE, McGuffin P, Gottesman, II. Twin concordance for DSM-III schizophrenia. Scrutinizing the validity of the definition. Arch Gen Psychiatry 1987; 44: 634–641.
13 Faraone SV, Blehar M, Pepple J, Moldin SO, Norton J, Nurnberger JI et al. Diagnostic accuracy and confusability analyses: an application to the Diagnostic Interview for Genetic Studies. Psychol Med 1996; 26: 401–410.
14 Kendler KS, Neale MC, Walsh D. Evaluating the spectrum concept of schizophrenia in the Roscommon Family Study. Am J Psychiatry 1995; 152: 749–754.
15 Moises HW, Yang L, Kristbjarnarson H, Wiese C, Byerley W, Macciardi F et al. An international two-stage genome-wide search for schizophrenia susceptibility genes. Nat Genet 1995; 11: 321–324.
16 Devlin B, Bacanu SA, Roeder K, Reimherr F, Wender P, Galke B et al. Genome-wide multipoint linkage analyses of multiplex schizophrenia pedigrees from the oceanic nation of Palau. Mol Psychiatry 2002; 7: 689–694.
17 Brzustowicz LM, Hodgkinson KA, Chow EW, Honer WG, Bassett AS. Location of a major susceptibility locus for familial schizophrenia on chromosome 1q21-q22. Science 2000; 288: 678–682.
18 Paunio T, Ekelund J, Varilo T, Parker A, Hovatta I, Turunen JA et al. Genome-wide scan in a nationwide study sample of schizophrenia families in Finland reveals susceptibility loci on chromosomes 2q and 5q. Hum Mol Genet 2001; 10: 3037–30348.
19 Garver DL, Holcomb J, Mapua FM, Wilson R, Barnes B. Schizophrenia spectrum disorders: an autosomal-wide scan in multiplex pedigrees. Schizophr Res 2001; 52: 145–160.
20 Gurling HM, Kalsi G, Brynjolfson J, Sigmundsson T, Sherrington R, Mankoo BS et al. Genomewide genetic linkage analysis confirms the presence of susceptibility loci for schizophrenia, on chromosomes 1q32. 2, 5q33. 2, and 8p21–22 and provides support for linkage to schizophrenia, on chromosomes 11q23. 3–24 and 20q12. 1–11. 23. Am J Hum Genet 2001; 68: 661–673.
21 Lindholm E, Ekholm B, Shaw S, Jalonen P, Johansson G, Pettersson U et al. A schizophrenia-susceptibility locus at 6q25, in one of the world's largest reported pedigrees. Am J Hum Genet 2001; 69: 96–105.
22 DeLisi LE, Shaw SH, Crow TJ, Shields G, Smith AB, Larach VW et al. A genome-wide scan for linkage to chromosomal regions in 382 sibling pairs with schizophrenia or schizoaffective disorder. Am J Psychiatry 2002; 159: 803–812.
23 DeLisi LE, Mesen A, Rodriguez C, Bertheau A, LaPrade B, Llach M et al. Genome-wide scan for linkage to schizophrenia in a Spanish-origin cohort from Costa Rica. Am J Med Genet 2002; 114: 497–508.
24 Wijsman EM, Rosenthal EA, Hall D, Blundell ML, Sobin C, Heath SC et al. Genome-wide scan in a large complex pedigree with predominantly male schizophrenics from the island of Kosrae: evidence for linkage to chromosome 2q. Mol Psychiatry 2003; 8: 695–705. 643.
25 Lerer B, Segman RH, Hamdan A, Kanyas K, Karni O, Kohn Y et al. Genome scan of Arab Israeli families maps a schizophrenia susceptibility gene to chromosome 6q23 and supports a locus at chromosome 10q24. Mol Psychiatry 2003; 8: 488–498.
26 Arinami T, Ohtsuki T, Ishiguro H, Ujike H, Tanaka Y, Morita Y et al. Genomewide high-density SNP linkage analysis of 236 Japanese families supports the existence of schizophrenia susceptibility loci on chromosomes 1p, 14q, and 20p. Am J Hum Genet 2005; 77: 937–944.
27 Fallin MD, Lasseter VK, Wolyniec PS, McGrath JA, Nestadt G, Valle D et al. Genomewide linkage scan for schizophrenia susceptibility loci among Ashkenazi Jewish families shows evidence of linkage on chromosome 10q22. Am J Hum Genet 2003; 73: 601–611.
28 Williams NM, Norton N, Williams H, Ekholm B, Hamshere ML, Lindblom Y et al. A systematic genomewide linkage study in 353 sib pairs with schizophrenia. Am J Hum Genet 2003; 73: 1355–1367.
29 Sklar P, Pato MT, Kirby A, Petryshen TL, Medeiros H, Carvalho C et al. Genome-wide scan in Portuguese Island families identifies 5q31–5q35 as a susceptibility locus for schizophrenia and psychosis. Mol Psychiatry 2004; 9: 213–218.
30 Abecasis GR, Burt RA, Hall D, Bochum S, Doheny KF, Lundy SL et al. Genomewide scan in families with schizophrenia from the founder population of Afrikaners reveals evidence for linkage and uniparental disomy on chromosome 1. Am J Hum Genet 2004; 74: 403–417.
31 Maziade M, Roy MA, Chagnon YC, Cliche D, Fournier JP, Montgrain N et al. Shared and specific susceptibility loci for schizophrenia and bipolar disorder: a dense genome scan in Eastern Quebec families. Mol Psychiatry 2005; 10: 486–499.
32 Faraone SV, Skol AD, Tsuang DW, Young KA, Haverstock SL, Prabhudesai S et al. Genome scan of schizophrenia families in a large Veterans Affairs Cooperative Study sample: evidence for linkage to 18p11. 32 and for racial heterogeneity on chromosomes 6 and 14. Am J Med Genet B Neuropsychiatr Genet 2005; 139: 91–100.
33 Suarez BK, Duan J, Sanders AR, Hinrichs AL, Jin CH, Hou C et al. Genomewide linkage scan of 409 European-ancestry and African American families with schizophrenia: suggestive evidence of linkage at 8p23. 3-p21. 2 and 11p13. 1-q14. 1 in the combined sample. Am J Hum Genet 2006; 78: 315–333.
34 Escamilla MA, Ontiveros A, Nicolini H, Raventos H, Mendoza R, Medina R et al. A genome-wide scan for schizophrenia and psychosis susceptibility loci in families of Mexican and Central American ancestry. Am J Med Genet B Neuropsychiatr Genet 2007; 144: 193–199.
35 Faraone SV, Hwu HG, Liu CM, Chen WJ, Tsuang MM, Liu SK et al. Genome scan of Han Chinese schizophrenia families from Taiwan: confirmation of linkage to 10q22. 3. Am J Psychiatry 2006; 163: 1760–1766.
36 Levinson DF, Mahtani MM, Nancarrow DJ, Brown DM, Kruglyak L, Kirby A et al. Genome scan of schizophrenia. Am J Psychiatry 1998; 155: 741–750.
37 Mowry BJ, Ewen KR, Nancarrow DJ, Lennon DP, Nertney DA, Jones HL et al. Second stage of a genome scan of schizophrenia: study of five positive regions in an expanded sample. Am J Med Genet 2000; 96: 864–869.
38 Holmans P, Riley B, Pulver AE, Owen MJ, Wildenauer DB, Gejman PV. Genomewide linkage scan of schizophrenia in a large multicenter pedigree sample using single nucleotide polymorphisms. Mol Psychiatry (in press).
39 Schwab SG, Hallmayer J, Albus M, Lerer B, Eckstein GN, Borrmann M et al. A genome-wide autosomal screen for schizophrenia susceptibility loci in 71 families with affected siblings: support for loci on chromosome 10p and 6. Mol Psychiatry 2000; 5: 638–649.
40 Bonnet-Brilhault F, Laurent C, Campion D, Thibaut F, Lafargue C, Charbonnier F et al. No evidence for involvement of KCNN3 (hSKCa3) potassium channel gene in familial and isolated cases of schizophrenia. Eur J Hum Genet 1999; 7: 247–250.
41 Campion D, d'Amato T, Bastard C, Laurent C, Guedj F, Jay M et al. Genetic study of dopamine D1, D2, and D4 receptors in schizophrenia. Psychiatry Res 1994; 51: 215–230.
42 Cao Q, Martinez M, Zhang J, Sanders AR, Badner JA, Cravchik A et al. Suggestive evidence for a schizophrenia susceptibility locus on chromosome 6q and a confirmation in an independent series of pedigrees. Genomics 1997; 43: 1–8.
43 Blouin JL, Dombroski BA, Nath SK, Lasseter VK, Wolyniec PS, Nestadt G et al. Schizophrenia susceptibility loci on chromosomes 13q32 and 8p21. Nat Genet 1998; 20: 70–73.
44 Faraone SV, Matise T, Svrakic D, Pepple J, Malaspina D, Suarez B et al. Genome scan of European-American schizophrenia pedigrees: results of the NIMH Genetics Initiative and Millennium Consortium. Am J Med Genet 1998; 81: 290–295.
45 Kaufmann CA, Suarez B, Malaspina D, Pepple J, Svrakic D, Markel PD et al. NIMH Genetics Initiative Millenium Schizophrenia Consortium: linkage analysis of African-American pedigrees. Am J Med Genet 1998; 81: 282–289.
46 Straub RE, MacLean CJ, Ma Y, Webb BT, Myakishev MV, Harris-Kerr C et al. Genome-wide scans of three independent sets of 90 Irish multiplex schizophrenia families and follow-up of selected regions in all families provides evidence for multiple susceptibility genes. Mol Psychiatry 2002; 7: 542–559.

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

47 Irmansyah, Schwab SG, Heriani, Handoko HY, Kusumawardhani I, Widyawat I et al. Genome-wide scan in 124 Indonesian sib-pair families with schizophrenia reveals genome-wide significant linkage to a locus on chromosome 3p26–21. Am J Med Genet B Neuropsychiatr Genet 2008; 147B: 1245–1252.
48 Coon H, Jensen S, Holik J, Hoff M, Myles-Worsley M, Reimherr F et al. Genomic scan for genes predisposing to schizophrenia. Am J Med Genet 1994; 54: 59–71.
49 Cooper-Casey K, Mesen-Fainardi A, Galke-Rollins B, Llach M, Laprade B, Rodriguez C et al. Suggestive linkage of schizophrenia to 5p13 in Costa Rica. Mol Psychiatry 2005; 10: 651–656.
50 Vazza G, Bertolin C, Scudellaro E, Vettori A, Boaretto F, Rampinelli S et al. Genome-wide scan supports the existence of a susceptibility locus for schizophrenia and bipolar disorder on chromosome 15q26. Mol Psychiatry 2007; 12: 87–93.
51 Bulayeva KB, Glatt SJ, Bulayev OA, Pavlova TA, Tsuang MT. Genome-wide linkage scan of schizophrenia: a cross-isolate study. Genomics 2007; 89: 167–177.
52 Stefansson H, Sigurdsson E, Steinthorsdottir V, Bjornsdottir S, Sigmundsson T, Ghosh S et al. Neuregulin 1 and susceptibility to schizophrenia. Am J Hum Genet 2002; 71: 877–892.
53 Matise TC, Chen F, Chen W, De La Vega FM, Hansen M, He C et al. A second-generation combined linkage physical map of the human genome. Genome Res 2007; 17: 1783–1786.
54 Lander E, Kruglyak L. Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. Nat Genet 1995; 11: 241–247.
55 Zintzaras E, Ioannidis JP. Heterogeneity testing in meta-analysis of genome searches. Genet Epidemiol 2005; 28: 123–137.
56 Lewis CM, Levinson DF. Testing for genetic heterogeneity in the genome search meta-analysis method. Genet Epidemiol 2006; 30: 348–355.
57 Risch N. Linkage strategies for genetically complex traits. I. Multilocus models. Am J Hum Genet 1990; 46: 222–228.
58 Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 2007; 447: 661–678.
59 Frayling TM. Genome-wide association studies provide new insights into type 2 diabetes aetiology. Nat Rev Genet 2007; 8: 657–662.
60 Mathew CG. New links to the pathogenesis of Crohn disease provided by genome-wide association scans. Nat Rev Genet 2008; 9: 9–14.
61 Lewis CM, Whitwell SC, Forbes A, Sanderson J, Mathew CG, Marteau TM. Estimating risks of common complex diseases across genetic and environmental factors: the example of Crohn disease. J Med Genet 2007; 44: 689–694.
62 Brzustowicz LM, Simone J, Mohseni P, Hayter JE, Hodgkinson KA, Chow EW et al. Linkage disequilibrium mapping of schizophrenia susceptibility to the CAPON region of chromosome 1q22. Am J Hum Genet 2004; 74: 1057–1063.
63 Xu B, Wratten N, Charych EI, Buyske S, Firestein BL, Brzustowicz LM. Increased expression in dorsolateral prefrontal cortex of CAPON in schizophrenia and bipolar disorder. PLoS Med 2005; 2: e263.
64 Guo S, Tang W, Shi Y, Huang K, Xi Z, Xu Y et al. RGS4 polymorphisms and risk of schizophrenia: an association study in Han Chinese plus meta-analysis. Neurosci Lett 2006; 406: 122–127.
65 Levitt P, Ebert P, Mirnics K, Nimgaonkar VL, Lewis DA. Making the case for a candidate vulnerability gene in schizophrenia: convergent evidence for regulator of G-protein signaling 4 (RGS4). Biol Psychiatry 2006; 60: 534–537.
66 Puri V. Fine mapping by genetic association implicates the chromosome 1q23.3 gene UHMK1, encoding a serine/threonine protein kinase, as a novel schizophrenia susceptibility gene. Biol Psychiatry 2007; 61: 873–879.
67 Stefansson H, Rujescu D, Cichon S, Pietilainen OP, Ingason A, Steinberg S et al. Large recurrent microdeletions associated with schizophrenia. Nature 2008; 455: 232–236.
68 Stone JL, O'Donovan MC, Gurling H, Kirov GK, Blackwood DH, Corvin A et al. Rare chromosomal deletions and duplications increase risk of schizophrenia. Nature 2008; 455: 237–241.
69 Gerber DJ, Hall D, Miyakawa T, Demars S, Gogos JA, Karayiorgou M et al. Evidence for association of schizophrenia with genetic variation in the 8p21. 3 gene, PPP3CC, encoding the calcineurin gamma subunit. Proc Natl Acad Sci USA 2003; 100: 8993–8998.
70 Stefansson H, Sarginson J, Kong A, Yates P, Steinthorsdottir V, Gudfinnsson E et al. Association of neuregulin 1 with schizophrenia confirmed in a Scottish population. Am J Hum Genet 2003; 72: 83–87.
71 O'Donovan MC, Craddock N, Norton N, Williams H, Peirce T, Moskvina V et al. Identification of loci associated with schizophrenia by genome-wide association and follow-up. Nat Genet 2008; 40: 1053–1055.

Supplementary Information accompanies the paper on the Molecular Psychiatry website (http://www.nature.com/mp)

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Elucidating the genetic architecture of familial schizophrenia using rare copy number variant and linkage scans

Bin Xu[a,b,1], Abigail Woodroffe[c,1], Laura Rodriguez-Murillo[a], J. Louw Roos[d], Elizabeth J. van Rensburg[e], Gonçalo R. Abecasis[f], Joseph A. Gogos[b,g,2], and Maria Karayiorgou[a,2]

Departments of [a]Psychiatry, [b]Physiology and Cellular Biophysics, and [g]Neuroscience, Columbia University Medical Center, New York, NY 10032; Departments of [c]Epidemiology and [f]Biostatistics, University of Michigan, Ann Arbor, MI 48109; [d]Department of Psychiatry, Weskoppies Hospital, Pretoria, Republic of South Africa 0001; and [e]Department of Genetics, University of Pretoria, Pretoria, Republic of South Africa 0001

Communicated by David E. Housman, Massachusetts Institute of Technology, Cambridge, MA, August 8, 2009 (received for review June 16, 2009)

To elucidate the genetic architecture of familial schizophrenia we combine linkage analysis with studies of fine-level chromosomal variation in families recruited from the Afrikaner population in South Africa. We demonstrate that individually rare inherited copy number variants (CNVs) are more frequent in cases with familial schizophrenia as compared to unaffected controls and affect almost exclusively genic regions. Interestingly, we find that while the prevalence of rare structural variants is similar in familial and sporadic cases, the type of variants is markedly different. In addition, using a high-density linkage scan with a panel of nearly 2,000 markers, we identify a region on chromosome 13q34 that shows genome-wide significant linkage to schizophrenia and show that in the families not linked to this locus, there is evidence for linkage to chromosome 1p36. No causative CNVs were identified in either locus. Overall, our results from approaches designed to detect risk variants with relatively low frequency and high penetrance in a well-defined and relatively homogeneous population, provide strong empirical evidence supporting the notion that multiple genetic variants, including individually rare ones, that affect many different genes contribute to the genetic risk of familial schizophrenia. They also highlight differences in the genetic architecture of the familial and sporadic forms of the disease.

rare mutations | chromosome 13q34 | chromosome 1p36 | RAPGEF gene family

Schizophrenia (SCZ) is a chronic, psychiatric disorder that has an estimated worldwide prevalence of ≈1%. The genetic architecture of the disease remains largely unknown. The strongest predictor of SCZ is having an affected first-degree relative (1). In addition to the familial forms, nonfamilial (sporadic) forms of the disease also exist (1, 2). The exact proportion of each form is largely unknown, but it is thought that at least 60% of cases are sporadic (3, 4). Genome-wide linkage scans have been conducted to identify loci harboring rare mutations/variants that increase susceptibility to familial SCZ. Loci have been identified on almost every chromosome, but only a few regions have been replicated across studies. One such region is near the telomere of chromosome 13q (5–12). Across studies the region of peak linkage is broad, from 13q12 to 13q34. This region has also been linked to bipolar disorder (13). In addition to linkage studies, a number of earlier (14) as well as more recent studies (15–18) have provided strong evidence supporting the importance of rare structural mutations/variants in SCZ vulnerability. Rare inherited structural lesions, in particular, are expected to be prominent in familial SCZ, but their full contribution to transmitted liability in familial SCZ cases has not been examined so far in a systematic manner (19).

To understand the genetic architecture of familial SCZ and the pattern of transmission of rare risk alleles in affected families, we combine high-resolution linkage analysis with studies of fine-level chromosomal variation in families recruited from the Afrikaner population in South Africa, a genetically and environmentally homogeneous population who have descended from mostly Dutch immigrants who settled in South Africa beginning in 1652 (20). In addition to the genetic homogeneity, the Afrikaners are valuable for genetic studies because they present a close-knit family structure and offer the potential to perform detailed genealogical analysis, which affords reliable discrimination of familial and nonfamilial forms of the disease and facilitates family-based genetic studies. Here, using approaches designed to detect risk variants with relatively low-frequency and high penetrance, we provide strong empirical evidence supporting the notion that multiple genetic variants, including individually rare ones that affect many different genes, contribute to the genetic risk of familial SCZ.

Results

Patient Cohorts. We performed a genome-wide survey of rare inherited copy number variants (CNVs) in a total of 182 individuals, consisting of 48 probands with familial SCZ [positive disease history in a first-degree ($n = 33$) or second-degree ($n = 15$) relative; Fig. S1] and both of their biological parents, as well as all additional affected relatives that were available for genotyping. Of the 48 probands, 40 are diagnosed as affected in the narrow category and eight in the broad category (see *Methods*). The familial cases cohort was compared to a control cohort ($n = 159$ triad families) as well as to a cohort enriched in sporadic cases ($n = 152$ triad families), defined as cases with negative family history of SCZ in a first- or second-degree relatives, also recruited from the Afrikaner community as previously described (15). In that respect, it should be noted that there were no significant differences in the average number of first- or second-degree relatives among families with and without family history. Specifically, in the 48 families with positive family history of SCZ in first- or second-degree relatives reported here, the average proband sibship was comprised of 3.4, the average maternal sibship of 4.3, and the average paternal sibship of 4.2 individuals. In the cohort enriched in sporadic cases (15), these numbers are 3.3, 4.3, and 4.6, respectively. Negative or positive family history or availability of additional affecteds was not a screening criterion (see *Methods*).

For our linkage studies, we genotyped 479 subjects from 130 families. Sixty-nine families are informative. In these, 112 indi-

Author contributions: B.X., A.W., G.R.A., J.A.G., and M.K. designed research; B.X. and A.W. performed research; J.L.R. and E.J.v.R. contributed new reagents/analytic tools; B.X., A.W., L.R.-M., and G.R.A. analyzed data; and B.X., A.W., G.R.A., J.A.G., and M.K. wrote the paper.

The authors declare no conflict of interest.

[1]B.X. and A.W. contributed equally to this work.

[2]To whom correspondence may be addressed. E-mail: mk2758@columbia.edu or jag90@columbia.edu.

This article contains supporting information online at www.pnas.org/cgi/content/full/0908584106/DCSupplemental.

www.pnas.org/cgi/doi/10.1073/pnas.0908584106

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Table 1. Increased frequency of rare inherited genic CNVs in familial SCZ

		Total rare inherited CN mutation carriers				Genic inherited CN mutation carriers				Non-genic inherited CN mutation carriers			
				P				*P*				*P*	
Cohort	No. of families	*n*	%	vs. controls	vs. sporadic	*n*	%	vs. controls	vs. sporadic	*n*	%	vs. controls	vs. sporadic
Familial	48	19	39.6	0.012	NS	17	35.4	0.001	0.007	2	4.2	NS	NS
Sporadic	152	46	30.3			24	15.8			22	14.5		
Controls	159	32	20.1			21	13.2			11	6.9		

Statistical significance was determined using Fisher's exact probability test; NS = non-significant.

viduals are diagnosed as affected in the narrow categor y, and 128 individuals are classified as broadly affected. In the 54 informative families with at least two affected members, there are 60 and 79 affected relative pairs for the narrow and broad affection categories, respectively (Table S1); this is 43% more affected relative pairs than in our previous, 9-cM, linkage study (12). A subset of the families (67%) used in the CNV studies ($n = 32$) was also included in the linkage scan. The appropriate Institutional Review Boards and Ethics Committees at University of Pretoria and Columbia University have approved all procedures for this study.

Genome-Wide Survey of Rare Inherited CNVs. We sur veyed single nucleotide polymorphisms (SNPs) and CNVs using the Affymetrix Genome-Wide Human SNP 5.0 arrays and used intensity and genotype data from both SNP and CN probes to identify autosomal deletions and duplications as described previously (15). The estimated rare inherited mutation rate was compared to the collective rate of inherited CNVs among sporadic cases and unaffected individuals from the same population (15). Rare inherited CNVs detected in familial cases and their parents were considered only if they involved at least 10 consecutive probe sets (average resolution of ≈30 kb) and did not show ≥50% overlap with a CNV detected in any parental chromosome (other than those of the biological parents) in the familial, sporadic, or control cohorts ($n = 1,432$ chromosomes). Using these criteria, we identified 24 rare inherited CNVs in 19 familial cases affecting 52 genes (Tables S2 and S3). The frequency of carriers of rare inherited structural lesions is ≈40% (19 out of 48) in our cohort of familial cases as compared to the ≈20% (32 out of 159) collective rate of inherited CNVs among unaffected individuals from the same population (15) (relative enrichment 1.97, Fisher's Exact Test $P = 0.01$) (Table 1). Cases and controls carr y on average 0.5 (24 CNVs in 48 cases) and 0.2 (32 in 159 controls) rare CNVs per person, respectively, a ≈2-fold difference in rare CNV burden. It should be noted that our population-specific filtering process is preferable to the one based on the diverse set of CNVs present in the database of genomic variants (DGV) (16) because DGV includes samples that have not been screened for psychiatric phenotypes and likely includes several pathogenic variants, and in addition, CNV frequenc y and disease-penetrance may var y across human populations (21). Nevertheless, essentially identical enrichment (relative enrichment 1.85, 27% vs. 14.5%, Fisher's Exact Test $P = 0.05$) was obtained upon further filtering that removed rare inherited variants overlapping with DGV (hg18 version 4). Familial clustering offers an important validation measure for all identified CNV regions (22). Nevertheless, we also confirmed, in four out of four tested families, the obser ved patterns of inheritance using an independent approach, the multiplex ligation-dependent probe amplification (MLPA) assay (23) (see *SI Methods* and Fig. S2).

Two lines of evidence suggest that the obser ved ≈2-fold enrichment of rare inherited CNVs in familial SCZ has a bone fide pathogenic basis. First, the obser ved enrichment in inherited structural mutations is highly specific, since we did not find any enrichment of de novo structural lesions in the same familial cases cohort (15). Second, as would be expected for pathogenic lesions, the enrichment in inherited CNVs among familial cases is confined exclusively to CNVs overlapping at least one gene, either partly or in its entirety (herein referred to as genic CNVs). Specifically, when we analyzed CNVs separately according to their gene composition, we found that individuals with familial SCZ were ≈2.7 times as likely as controls to harbor rare genic CNVs (relative enrichment 2.7, Fisher's Exact Test $P = 10^{-3}$, Table 1). Essentially identical enrichment in genic CNVs (rel- ative enrichment 2.6, 25% vs. 9.5%, Fisher's Exact Test $P =$ 0.012) was obtained upon further filtering that removed rare inherited variants overlapping with DGV (hg18 version 4). In contrast, there was no significant difference in the proportion of cases versus controls carr ying rare nongenic CNVs (Table 1). Moreover, no such enrichment in inherited genic CNVs was obser ved in the cohort enriched for sporadic cases (relative enrichment 1.2, $P = 0.52$), where the contribution of inherited CNVs appears to be relatively minor. In that respect, it is noteworthy that while the overall frequency of carriers of all rare structural variants is the same between the familial and sporadic cases (15) (≈40%), the type of variants is markedly different (Fig. 1*A*). Sporadic SCZ is characterized by a marked enrichment of rare de novo mutations and only a modest increase in the rate of rare inherited CNVs, which do not appear to preferentially affect genes. By contrast, familial SCZ is characterized by enrichment in rare inherited genic CNVs (predicted to have higher penetrance), while de novo mutations are less prominent.

None of the rare CNVs found in familial cases are present in control chromosomes. To identify which of these CNVs are most likely to be pathogenic, we investigated the relationship between rare inherited CNVs and disease status within each family. For 12 of the 19 CNV carriers, DNA samples and genotypic information were available for at least one more affected relative. In nine of these 12 families, the CNV segregated to all genotyped affected members (only two of the 23 affected members of these nine families were not available for genotyping). Thus, CNVs in these nine families showed clear co-segregation with the clinical diagnosis, in a manner consistent with incomplete penetrance models (i.e., present also in some unaffected family members) (Fig. 1*B*). The remaining three CNVs segregated only to one of the genotyped affected members in each family. Given the ≈20% basal rate of inherited CNVs in unaffected controls (15), these are likely to be neutral variants. Alternatively, these cases may be indicative of more complex modes of inheritance, such as bilineal transmission of risk alleles. To exclude that the obser ved pattern of co-inheritance of CNVs and diagnosis in the informative families is due to chance alone, we conducted a simulation study where we disrupted the relationship between CNVs and diagnosis by permuting the diagnosis, while keeping constant the pedigree structure, the inherited pattern of CNVs

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

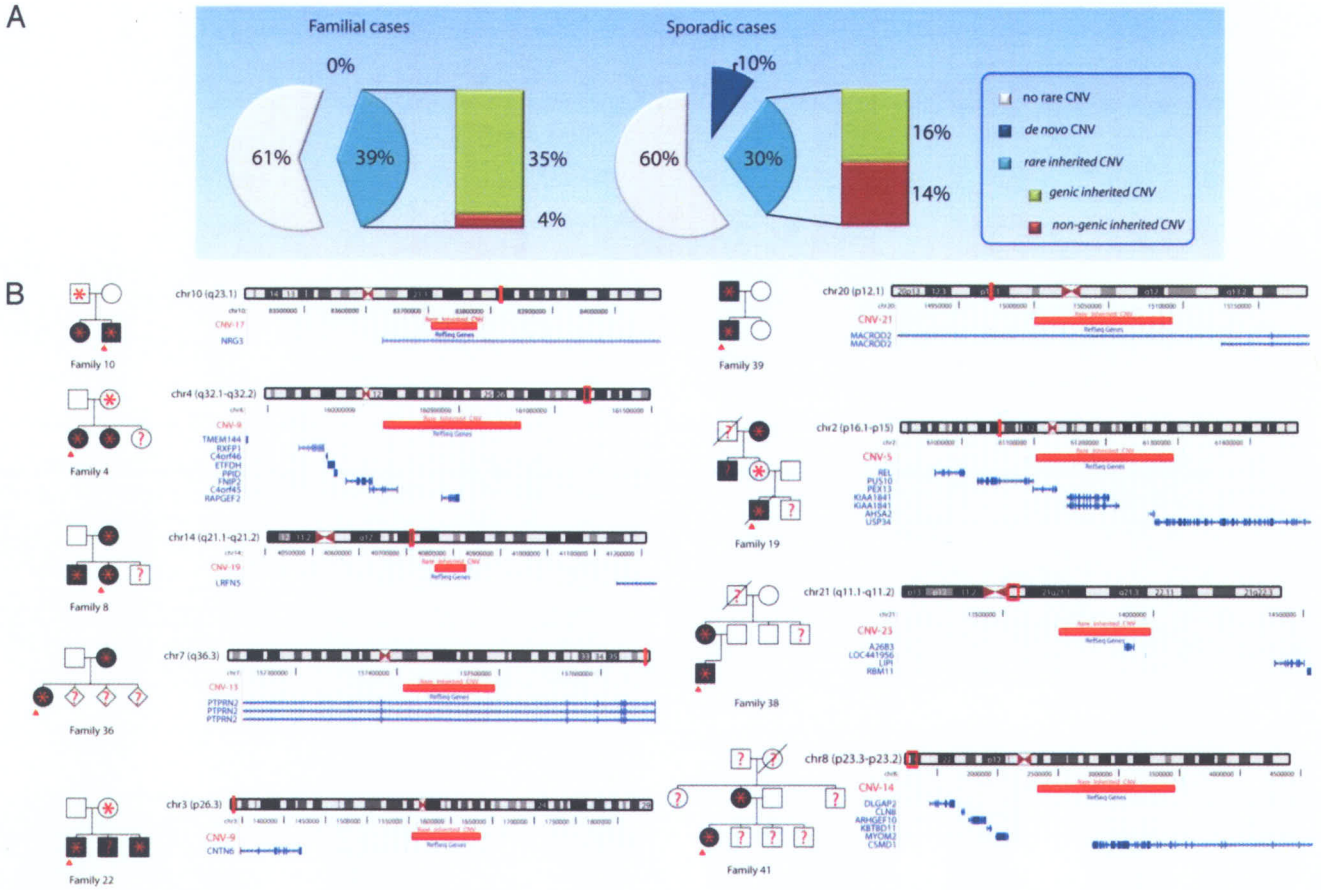


Fig. 1. Inherited CNVs in families with SCZ (*A*) Frequency distribution of rare CNVs identified in familial and sporadic cases of SCZ at the resolution afforded by the Affymetrix Genome-Wide Human SNP 5.0 arrays. There is a =20% basal rate of inherited CNVs in unaffected controls, while the overall frequency of carriers of all rare CNVs is the same between familial and sporadic cases (=40%). (*B*) Rare inherited CNVs showing co-segregation with the clinical diagnosis in the respective affected families. For each CNV, the structure of the affected family, as well as the genomic position of the CNV, are indicated. Affected individuals are marked in black. Probands are indicated by red arrows. Individuals who carry rare CNVs are indicated by an asterisk. Individuals where no genotype information is available are indicated by question marks.

and the number of affected individuals in each family. Co-inheritance of CNVs and diagnosis in nine out of 12 families was observed only once in 10,000 permutation runs (empirical *P* value = 0.0001) and, therefore, it is unlikely to be due to chance.

The nine CNVs segregating to all genotyped affected members within the respective families, alter 12 known genes: *PEX13*, *KIAA1841*, *AHSA2*, *USP34*, *C4orf45*, *RAPGEF2*, *PTPRN2*, *CSMD1*, *NRG3*, *MACROD2*, *A26B3*, and *LOC441956*, all of which are candidates for follow-up studies. Convergence with previous studies highlights at least two of these genes as particularly worthy of follow-up. Specifically, the neuregulin-3 (*NRG3*) gene is implicated by a 73.6-kb-long duplication in the first intron, which may cause regulatory deficits. An overlapping CNV has been reported previously in one case with SCZ (24) (Table S2). Three SNPs in the same intron were recently associated with SCZ-related quantitative traits (24). The homologous gene, *NRG1*, is also a well-known candidate gene (25). In addition, the *RAPGEF2* gene, encoding for a GTP exchange factor, is implicated by a 716.4-kb duplication that encompasses this gene. Mutation of the *RAPGEF2* ortholog in mice affects the formation of the cerebral cortex, reduces the threshold for the induction of epileptic seizures and results in commissural fiber defects (26). Previously, we have identified in a patient with SCZ a de novo exonic microdeletion affecting another member of the RAPGEF family (*RAPGEF6*) located within a SCZ susceptibility locus at chromosome 5q (15). Mutations in another member of this family (*RAPGEF4*) have been described in autism (27).

We examined the familial cases for differences in clinical and phenotypic variables (history of developmental delays or learning disabilities, mental retardation, age at onset, and disease severity) that may discriminate inherited CN mutation carriers. We focused our analysis on probands from families where CNVs show apparent co-segregation with the illness and are therefore more likely to be pathogenic. Among them, the male to female ratio is identical to that of the entire familial cases cohort, and there is no statistically significant evidence for parental origin effects. None of these probands had a history of developmental delays or learning disabilities, or presence of mental retardation. In addition, there was no difference in the age at disease onset between these probands and noncarriers of rare CNVs, but we found some suggestive differences in indices of disease severity, including co-morbid substance abuse (67% versus 23%, $P = 0.03$, uncorrected), duration of illness (141.6 months versus 92.2 months, $P = 0.19$), and number of hospitalizations (4.9 compared to 2.9, $P = 0.13$), indicating a more debilitating or treatment-resistant form of illness among CNV carriers.

High-Density Linkage Analysis. Previously, a low-density (9-cM) linkage scan in our Afrikaner sample identified suggestive

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

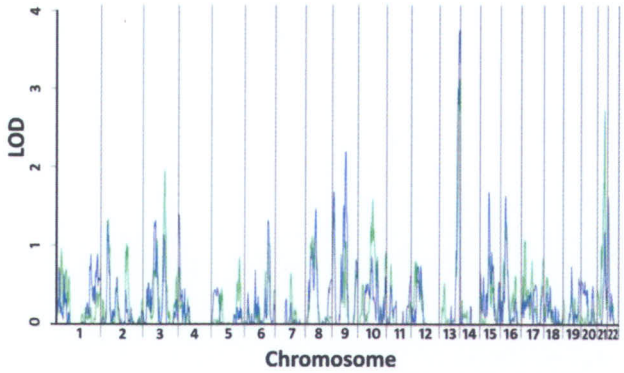


Fig. 2. MOD score analysis. Green line shows MOD scores for the narrow classification. Blue line shows MOD scores for the broad classification.

evidence for linkage at 13q34 and 1p36 (maximum nonparametric LOD scores 2.99 and 2.23, respectively) (12). Here we genotyped 2005 di-, tri-, and tetra-nucleotide repeat microsatellite markers, in one of a few linkage scans of psychiatric disorders to attain this level of genome-wide coverage. The average inter-marker distance was 1.9 cM (±1.4), and the average heterozygosity for the autosomal markers was 0.71 (±0.12) resulting in an information content of 0.84 (±0.046) (Figs. S3, S4). We conducted both parametric and nonparametric analyses. For our parametric analyses, we used the algorithm implemented in LAMP to estimate the disease allele frequency and genotype penetrances by maximum likelihood. We used the optimized parameters to calculate model maximized LOD (MOD) scores, which are more powerful than LOD scores when the disease parameters are unknown (28).

Singlepoint Parametric Results. We found three markers with a MOD score of at least 3.0 in either affection category: D13S285 on 13q34, D9S50 on 9p13, and D21S270 on 21q22. The highest MOD scores for both affection statuses are for D13S285, located at 127 cM. For the narrow affection status the MOD score is 3.30 and for the broad classification is 3.67. D9S50 at 60 cM has a MOD score of 3.56 for the broad classification and 2.51 for the narrow. For D21S270 at 46 cM, the MOD score for the narrow classification is 3.0 and for the broad is 1.88 (Table S4).

Multipoint Parametric Results. For our multipoint parametric analysis (Fig. 2), the maximum MOD scores for both affection classifications are also on 13q34. The maximum MOD score for the narrow affection status is 3.13 at 126 cM. When we repeated the linkage analyses on 1,000 simulated data sets, we calculated an empirical P value of 0.093. For the broad affection status, the highest MOD score is 3.76 at 131 cM, near D13S293, and the 1-MOD region spans from 115 cM to the q terminus. The empirical P value for the broad affection status alone is 0.025. When we include both the narrow and broad classifications in our simulations, 42 of the data sets resulted in a MOD score greater than or equal to 3.76. This empirical genome-wide P value of 0.042 meets the criteria for a significant linkage result (29). For the broad phenotype, the maximum likelihood estimate for the disease penetrance for an individual with two copies of the disease allele is 1.0, for an individual with one copy is 0.073, and for an individual with no copies is 0.005. Although the disease allele frequency is estimated to be fairly rare ($f_d = 0.030$), the relative risk is very high (RR = 13.77). In addition to the 13q locus, we identified linkage peaks on chromosomes 21q22 at 46 cM, near D21S1900, with a 1-MOD interval from 41 to 48 cM and on 9q21 at 85 cM, near D9S1877, with a 1-MOD interval from 80 to 96 cM (Table 2).

Singlepoint Nonparametric Results. The highest LOD score is for marker D1S2885 at 46 cM on 1p36. The LOD score for the narrow classification is 2.30, and is the genome-wide maximum at 2.87 for the broad classification. D13S285, at 127 cM on 13q34, has elevated LOD scores for both the narrow and broad classifications; the LOD scores are 2.80 and 2.70, respectively. Also noteworthy are three adjacent markers on 21q22 with LOD scores greater than 2.0 using the narrow category: D21S1900 (LOD = 2.30), D21S1919 (LOD = 2.55), and D21S270 (LOD = 2.31). For the broad category these three markers have LOD scores ranging from 1.20 to 1.53 (Table S5).

Multipoint Nonparametric Results. The maximum multipoint LOD score is located about 2 cM away from D13S285 at 125 cM on

Table 2. Parametric multipoint MOD scores > 1.5 for either SCZ status (multiplicative model)

Chr	Position (1-MOD)	Nearest Marker	Aff	MOD	Freq of A	RR	Pen of AA	Pen of Aa	Pen of aa
3q21	137 cM (128–147)	D3S1589	N	1.95	0.39	99.84	0.065	0.00065	6.5×10^{-6}
			B	1.13	0.53	99.97	0.035	0.00035	3.5×10^{-6}
9p24	12 cM (3–19)	D9S1686	N	1.40	0.01	9.86	0.77	0.078	0.0079
			B	1.68	0.01	11.03	>.99	0.091	0.0082
9q21	85 cM (80–96)	D9S1877	N	1.05	<.01	10.44	>.99	0.096	0.0092
			B	2.20	<.01	10.72	>.99	0.093	0.0087
10q22	92 cM (77–112)	D10S537	N	1.58	0.01	11.32	>.99	0.088	0.0078
			B	0.71	0.17	4.04	0.072	0.018	0.0044
13q34	131 cM (115-qter)	D13S293	N	3.13	0.03	14.76	>.99	0.068	0.0046
			B	3.76	0.03	13.77	>.99	0.073	0.0053
15q21	56 cM (54–65)	D15S1022	N	0.87	<.01	10.35	>.99	0.097	0.0094
			B	1.69	<.01	10.39	>.99	0.096	0.0093
16p13	31 cM (27–45)	D16S3047	N	1.31	0.17	5.60	0.10	0.018	0.0032
			B	1.64	0.10	5.64	0.14	0.026	0.0046
21q22	46 cM (41–48)	D21S1900	N	2.72	0.36	99.95	0.073	0.00073	7.4×10^{-6}
			B	1.18	0.52	99.86	0.037	0.00037	3.7×10^{-6}
22q11	3 cM (2–9)	D22S420	N	1.47	0.44	99.99	0.049	0.00050	5.0×10^{-6}
			B	1.62	0.45	99.99	0.048	0.00048	4.8×10^{-6}

'A' is the disease allele, 'a' is the non-disease allele. 1-MOD, region in which the MOD score is within 1 MOD score of the highest MOD score; Aff, affection classification; N, narrowly affected; B, broadly affected; Freq, allele frequency; RR, relative risk based on a 1% prevalence of SCZ; Pen, penetrance of SCZ for the given genotype. Therefore, Pen of AA is the probability of having SCZ, given two copies of the disease allele.

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Table 3. Parametric linkage to chromosome 1p36 (dominant model)

Family group	Number of families	Position	Nearest Marker	Narrow MOD	Broad MOD
All*	69	35cM	D1S2826	1.28	0.20
Unlinked to 13q	25	35cM	D1S2826	3.21	1.74
Linked to 13q	43	35cM	D1S2826	0.12	0.00

*One family had a MOD score of 0.000 and was not used in the subset analysis.

13q34. At this location, the LOD score for the narrow classification is 2.65; for the broad classification the LOD score is slightly higher at 2.66. The 1-LOD interval around this peak extends from 119 cM to the q terminus. However, the empirical significance based on 1,000 simulations is not significant ($P = 0.25$). The nonparametric multipoint analysis also provides evidence for linkage at 21q22. The LOD scores are 2.16 for the narrow and 0.83 for the broad classification (Table S6). This is the region where three markers showed evidence for linkage in the singlepoint analysis.

Targeted Analyses in Families Nonlinked to Chromosome 13. Twenty-five of our families in our sample exhibited a MOD score of <0.0 at the location of our strongest linkage peak, at 131 cM on chromosome 13q34 near marker D13S293. To examine evidence for additional susceptibility loci, we carried out a series of analysis targeted at these 25 families (Table 3). Notably, the maximum MOD score in these 25 families occurs on chromosome 1p36 at 35 cM (=1 cM from D1S2644), near the peak that was identified in our previous 9-cM scan. The MOD score for the narrow classification at that position is 3.21 (empirical P value = 0.15); the MOD score for the broad classification is 1.74. The nonparametric LOD scores in the region (both at =29 cM, near D1S2697) were only 1.32 and 0.54 for the narrow and broad classifications, respectively. The MOD score analysis results suggest a dominant mode of inheritance, with the estimated disease allele frequency of 0.003 and penetrances/genotype relative risks of 1.0, 0.93, and 0.004.

Exploration of the Relation Between Linkage Signals and CNVs. We first assessed the contribution of copy number mutations to the two primary and consistent among studies linkage signals at 13q34 and 1p36. Whole genome CNV annotation, conducted using dCHIP program, identified 788 putative CNVs in the 241 cases included in our scan. In the 13q34 region, there was only one CNV identified, which did not overlap with any gene (Table S7). This CNV was found in two out of 224 cases (0.89%) and three out of 361 parents of unaffected controls (0.83%). In the 1p36 region, there were seven CNVs including six genomic gains and one loss (Table S7). There was no statistical difference in the frequencies of these CNVs between cases and controls. None of the identified CNVs was present in families linked to the respective linkage loci. Thus, at the level of resolution of our scan, we could not identify any CNV that accounts for the linkage signals identified.

Because only a subset of the families used in the linkage studies ($n = 32$) were also included in the CNV scan and additional, yet unidentified rare CNV carrying families are likely to be part of the linkage cohort, it is not possible at this point to conclusively evaluate the effect that removal of families, which carry at least one rare CNV, has on linkage. Nevertheless, preliminary analysis shows that, despite the decrease in the sample size, after removing CNV carrying families ($n = 14$, eight of them with CNVs co-segregating with the clinical diagnosis), the evidence for linkage in 13q34 remained unchanged (MOD = 3.77). To estimate the expected change, under the assumption that the families removed from the analysis are contributing to the 13q34 signal, we randomly removed 14 families from among the ones showing linkage to 13q34. In 100 trials, the MOD scores varied from 0.53 to 2.76 for the narrow definition, with an average of 1.50, and from 0.63 to 2.99 for the broad definition, with an average of 1.74. This notable reduction in the MOD score suggests that at least compared to the families linked to 13q34, the rare CNV containing families may represent a largely distinct subset of genetic liability. Interestingly, our exploratory analysis shows that in a number of loci, weak linkage signals are amplified following removal of rare CNV families despite decrease in power (for example, at 3q21 MOD increases from 1.95 to 2.3 for the narrow classification, at 9p24 MOD increases from 1.68 to 2.13 for the broad classification, and at 16p13 MOD increases from 1.64 to 2.3 for the broad classification). These may reflect true linkage signals, which are masked by the heterogeneity introduced by the rare CNV families.

Discussion

Our results offer a comprehensive picture of the genetic architecture of familial SCZ in a relatively homogeneous population. Our chromosomal variation analysis provides evidence for a role of inherited structural lesions in familial SCZ and highlights some important differences in the genetic architecture of familial and nonfamilial forms of the disease. The majority of the identified inherited CNVs, co-segregate with disease in a manner consistent with necessary but not always sufficient genetic "hits," which lead to a disease state only in combination with additional, inherited structural or sequence variation or environmental factors. Patient stratification suggests that inherited CNVs do not correlate with history of developmental delays, learning disabilities, or presence of mental retardation, but appear to be enriched in more debilitating or treatment-resistant form of illness. Although our study is statistically underpowered to prove the involvement of any specific CNV, analysis of co-segregation with the clinical diagnosis, as well as convergence with previous studies, highlights a number of genes, gene families, and related pathways (such as the contactin family, the *CSMD1*, *ADARB1*, *RXFP2*, *LRFN5*, and *NRG3* genes) as particularly worthy of follow-up (see *SI Text*). In particular, we provide additional evidence strongly supporting a previously unknown role of a family of Rap1 guanine nucleotide exchange factors (*RAPGEF* family) and Rap1-mediated processes (30) in psychiatric disorders.

Our linkage analyses indicated one or more genes that increases susceptibility to SCZ on chromosome 13q34 in our broadly affected individuals. It is noteworthy that three of the four other SCZ linkage scans that identified a LOD score greater than 2.0 at 13q34 included schizoaffective disorder, bipolar type in the affection category (7–10). That our broad classification also includes this diagnosis supports the hypothesis that one or more genes on 13q34 increases susceptibility to SCZ and bipolar spectrum disorders. Based on the results of this study and previous ones, there is considerable evidence for linkage to 13q. When we analyzed families that do not show evidence for linkage to 13q34, we identified another linkage peak on 1p36. This result was much stronger when we used the narrow definition of SCZ that did not include schizoaffective disorder, bipolar subtype. Compared to the 13q34 locus, this could indicate different causal alleles and mechanisms in subjects with and without symptoms

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

of bipolar disorder. Notably, at the level of resolution of our scan, our analysis indicates that CNVs within both linkage signal regions are likely to be neutral and unlikely to account for the linkage signals identified. Finally, although the sample sizes are relatively small, our findings suggest that compared to the families linked to 13q34, the rare CNV carrying families represent a largely distinct subset of genetic liability. CNV carrying families may be a source of heterogeneity and future studies will focus on identifying all rare CNVs as putative risk loci and as a tool to stratify the samples to reduce genetic heterogeneity for linkage analyses and improve detection of weak signals.

Irrespective of the pathogenic potential and the precise mode of action of each risk locus, our results highlight the difference in the genetic architecture of the familial and sporadic forms of the disease and support the notion that multiple genetic variants, including individually rare ones (often unique to a single patient) that affect many different genes contribute to the genetic risk of familial SCZ. This heterogeneity (present to some degree even in founder populations) is consistent with the hypothesis that there are many genes that contribute to SCZ and may account for past and present difficulties in finding bone fide genetic variants. Because there are significant clinical similarities of SCZ cases diagnosed in Afrikaners and those diagnosed in more heterogeneous populations (such as the U.S.) (20), our results are likely to have general implications regarding the genetic architecture of SCZ.

Samples and Methods

Cohorts. Both affected and control families were recruited and diagnosed as part of our ongoing, large-scale genetic study of SCZ in the Afrikaner population in South Africa, as previously described (12, 15, 20) (see also *SI Methods*). For our linkage study, affected subjects were classified as either narrowly or broadly affected. The narrow diagnosis includes subjects with SCZ or schizoaffective disorder-depressive type, as previously described (12). The broad diagnosis includes all individuals classified as affected under the narrow definition as well as individuals with schizoaffective disorder-bipolar type. Compared to our previous classification (12), it is more encompassing than LCI, but not as broad as LCII. For our CNV studies, the criteria for inclusion in the affected cohort are: (*i*) Afrikaner heritage; (*ii*) proband meeting full diagnostic criteria for SCZ or schizoaffective disorder; and (*iii*) both biological parents alive and willing to participate. It should be noted that presence of negative or positive family history or availability of additional affected relatives is not a screening criterion. Nevertheless, for all recruited subjects, detailed information about family history of any psychiatric or medical illness was solicited from at least three sources (proband and each participating parent) by two independent raters [the nursing sister, who completes the Medical and Personal History form with each study participant and also draws a detailed pedigree for at least three to four generations for each family, as well as the psychiatrist who administers the Diagnostic Instrument for Genetic Studies (DIGS) to the proband]. In addition, since we also trace the ancestry of all recruited families, we routinely use several informants in each family to inquire about all relatives' names, date and place of birth, and death and psychiatric status. Because of the close-knit family structure of the Afrikaner families and the availability of detailed psychiatric records over several generations due to the large catchment area and long-term care provided by the local recruiting hospital, we are typically able to obtain information about psychiatric status for at least three to four generations removed from the proband. For any relative identified as possibly having symptoms of SCZ or schizoaffective disorder, or a history of treatment or hospitalization for a psychiatric condition, every effort was made to include that relative in the study, if alive and willing to participate, by obtaining a blood sample and administering an in-person diagnostic interview. In a few instances where the exact nature of a reported psychiatric diagnosis in a first- or second-degree relative could not be substantiated (i.e., because the person was not alive or access to records was not possible), the family history status was left unknown. Such families are not considered in the present study or in the Xu et al. study (15). Finally, in addition to being matched by ancestry, a subset of the control families (three informants per family inquiring for up to three generations) completed a detailed self-report questionnaire that inquired about several psychiatric conditions, including psychosis, phobias, anxiety, and depression (see also *SI Methods*).

Genotyping procedures, linkage analysis and CNV identification, and verification are outlined in detail in the *SI Methods*.

ACKNOWLEDGMENTS. We thank Alexandra Abrams-Downey and Yan Sun for expert technical assistance. This work was supported in part by National Institute of Mental Health Grant MH061399 (to M.K.) and MH077235 (to J.A.G.) and the Lieber Center for Schizophrenia (SCZ) Research at Columbia University Medical Center (CUMC).

1. Gottesman II (1991) *Schizophrenia Genesis* (W.H. Freeman and Company, New York, NY).
2. Griffiths TD, et al. (1998) Minor physical anomalies in familial and sporadic schizophrenia: The Maudsley family study. *J Neurol Neurosurg Psychiatry* 64:56–60.
3. Kendler KS, Diehl SR (1993) The genetics of schizophrenia: A current, genetic-epidemiologic perspective. *Schizophr Bull* 19:261–285.
4. Gottesman II, Erlenmeyer-Kimling L (2001) Family and twin strategies as a head start in defining prodromes and endophenotypes for hypothetical early-interventions in schizophrenia. *Schizophr Res* 51:93–102.
5. Lin MW, et al. (1995) Suggestive evidence for linkage of schizophrenia to markers on chromosome 13q14.1-q32. *Psychiatr Genet* 5:117–126.
6. Lin MW, et al. (1997) Suggestive evidence for linkage of schizophrenia to markers on chromosome 13 in Caucasian but not Oriental populations. *Hum Genet* 99:417–420.
7. Blouin JL, et al. (1998) Schizophrenia susceptibility loci on chromosomes 13q32 and 8p21. *Nat Genet* 20:70–73.
8. Shaw SH, et al. (1998) A genome-wide search for schizophrenia susceptibility genes. *Am J Med Genet* 81:364–376.
9. Brzustowicz LM, et al. (1999) Linkage of familial schizophrenia to chromosome 13q32. *Am J Hum Genet* 65:1096–1103.
10. Camp NJ, et al. (2001) Genomewide multipoint linkage analysis of seven extended Palauan pedigrees with schizophrenia, by a Markov-chain Monte Carlo method. *Am J Hum Genet* 69:1278–1289.
11. Faraone SV, et al. (2002) Linkage of chromosome 13q32 to schizophrenia in a large veterans affairs cooperative study sample. *Am J Med Genet* 114:598–604.
12. Abecasis GR, et al. (2004) Genomewide scan in families with schizophrenia from the founder population of Afrikaners reveals evidence for linkage and uniparental disomy on chromosome 1. *Am J Hum Genet* 74:403–417.
13. Badner JA, Gershon ES (2002) Meta-analysis of whole-genome linkage scans of bipolar disorder and schizophrenia. *Mol Psychiatry* 7:405–411.
14. Karayiorgou M, et al. (1995) Schizophrenia susceptibility associated with interstitial deletions of chromosome 22q11. *Proc Natl Acad Sci USA* 92:7612–7616.
15. Xu B, et al. (2008) Strong association of de novo copy number mutations with sporadic schizophrenia. *Nat Genet* 40:880–885.
16. Walsh T, et al. (2008) Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia. *Science* 320:539–543.
17. Stefansson H, et al. (2008) Large recurrent microdeletions associated with schizophrenia. *Nature* 455:232–236.
18. International Schizophrenia Consortium (2008) Rare chromosomal deletions and duplications increase risk of schizophrenia. *Nature* 455:237–241.
19. Maher BS, Riley BP, Kendler KS (2008) Psychiatric genetics gets a boost. *Nat Genet* 40:1042–1044.
20. Karayiorgou M, et al. (2004) Phenotypic characterization and genealogical tracing in an Afrikaner schizophrenia database. *Am J Med Genet B* 124:20–28.
21. Jakobsson M, et al. (2008) Genotype, haplotype and copy-number variation in worldwide human populations. *Nature* 451:998–1003.
22. Ionita-Laza I, Laird NM, Raby BA, Weiss ST, Lange C (2008) On the frequency of copy number variants. *Bioinformatics* 24:2350–2355.
23. Vorstman JA, et al. (2006) MLPA: A rapid, reliable, and sensitive method for detection and analysis of abnormalities of 22q. *Hum Mutat* 27:814–821.
24. Chen PL, et al. (2009) Fine mapping on chromosome 10q22–q23 implicates Neuregulin 3 in schizophrenia. *Am J Hum Genet* 84:21–34.
25. Stefansson H, et al. (2002) Neuregulin 1 and susceptibility to schizophrenia. *Am J Hum Genet* 71:877–892.
26. Bilasy SE, et al. (2009) Dorsal telencephalon-specific RA-GEF-1 knockout mice develop heterotopic cortical mass and commissural fiber defect. *Eur J Neurosci* 29:1994–2008.
27. Bacchelli E, et al. (2003) Screening of nine candidate genes for autism on chromosome 2q reveals rare nonsynonymous variants in the cAMP-GEFII gene. *Mol Psychiatry* 8:916–924.
28. Greenberg DA, Abreu P, Hodge SE (1998) The power to detect linkage in complex disease by means of simple LOD-score analyses. *Am J Hum Genet* 63:870–879.
29. Lander E, Kruglyak L (1995) Genetic dissection of complex traits: Guidelines for interpreting and reporting linkage results. *Nat Genet* 11:241–247.
30. Kawasaki H, et al. (1998) A family of cAMP-binding proteins that directly activate Rap1. *Science* 282:2275–2279.

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Exome sequencing supports a *de novo* mutational paradigm for schizophrenia

Bin Xu[1,2], J Louw Roos[3], Phillip Dexheimer[4], Braden Boone[4], Brooks Plummer[4], Shawn Levy[4], Joseph A Gogos[2,5] & Maria Karayiorgou[1]

Despite its high heritability, a large fraction of individuals with schizophrenia do not have a family history of the disease (sporadic cases). Here we examined the possibility that rare *de novo* protein-altering mutations contribute to the genetic component of schizophrenia by sequencing the exomes of 53 sporadic cases, 22 unaffected controls and their parents. We identified 40 *de novo* mutations in 27 cases affecting 40 genes, including a potentially disruptive mutation in *DGCR2*, a gene located in the schizophrenia-predisposing 22q11.2 microdeletion region. A comparison to rare inherited variants indicated that the identified *de novo* mutations show a large excess of non-synonymous changes in schizophrenia cases, as well as a greater potential to affect protein structure and function. Our analyses suggest a major role for *de novo* mutations in schizophrenia as well as a large mutational target, which together provide a plausible explanation for the high global incidence and persistence of the disease.

© 2011 Nature America, Inc. All rights reserved.

Schizophrenia has a strong genetic component[1,2]. However, despite its high heritability, a large fraction of individuals with schizophrenia do not have a family history of the disease[3]. Although largely ignored in earlier efforts to model disease risk, *de novo* germline mutations may account for a substantial fraction of sporadic schizophrenia cases. In agreement with this hypothesis, rare *de novo* copy number variants (CNVs)[4] are emerging as an important genomic cause of schizophrenia and other psychiatric diseases, and the variant with the strongest statistical support for association with schizophrenia, namely the 22q11.2 microdeletion, is a *de novo* and recurrent mutation[5,6].

Availability of next-generation whole-genome or whole-exome sequencing[7] now permits the study of *de novo* mutations (point substitutions or single nucleotide variants (SNVs) and small insertions or deletions (indels)) in a systematic genome-wide manner[8,9]. Pilot studies focusing on specific synaptic genes identified a small number of putative *de novo* mutations in individuals with schizophrenia[10]. However, the full contribution of rare *de novo* SNVs and indels to schizophrenia remains unknown.

To test the hypothesis that *de novo* protein-altering mutations contribute substantially to the genetic component of schizophrenia, we sequenced the exomes of 53 family trios of subjects diagnosed with schizophrenia or schizoaffective disorder with no history of the disease in a first- or second-degree relative ('sporadic cases cohort') as well as family trios of 22 unrelated healthy controls, all recruited from the genetically homogeneous Afrikaner population of European descent in South Africa[11,12]. Presence or absence of family history in cases was not a screening criterion during recruitment but could be reliably determined because of the close-knit family structure of the families and the availability of detailed psychiatric records over several generations because of the large catchment area and long-term care provided by the local recruiting hospital[6,12,13]. Control families completed a detailed self-report questionnaire that inquired about several psychiatric conditions, including phobias, anxiety, depression and history of treatment for any of these conditions. Also, mental illness in first- or second-degree relatives was excluded. Based on previous results[6,13], we excluded carriers of rare *de novo* CNVs. Identities were coded and analyses were performed blind to affected status while maintaining knowledge of the parent-child relationships. From all 225 individuals, we extracted DNA samples from whole blood.

We enriched exonic sequences using the Agilent SureSelect technology for targeted exon capture and performed Illumina paired-end sequencing (one lane of flow cell per sample; see Online Methods). On average, we obtained 7.3 Gb of mappable sequence data per individual after exome enrichment, targeting 37 Mb from exons and their flanking regions. Overall, we covered 1.22% of the genome, a fraction corresponding to the NCBI Consensus Coding Sequences database (CCDS). We cross matched the paired-end reads to the reference genome (hg19 build) using the Burrows-Wheeler Aligner (BWA v0.5.81536)[14]. Ninety-seven point nine percent of the reads were properly aligned to the reference genome. Our median read depth was 65.2×, which is higher than the estimated average depth (33×) required for highly accurate downstream heterozygous variant detection. In addition, 92.4% of the captured target exons were covered by high quality genotype calls at least eight times to ensure good detection sensitivity[15] (**Table 1**).

Our *de novo* mutation detection pipeline is shown in **Supplementary Figure 1**. We implemented a series of filters, including final validation by standard Sanger sequencing (**Supplementary Figs. 2,3**), to eliminate

[1]Department of Psychiatry, Columbia University, New York, New York, USA. [2]Department of Physiology & Cellular Biophysics, Columbia University, New York, New York, USA. [3]Weskoppies Hospital & Department of Psychiatry, University of Pretoria, Pretoria, South Africa. [4]HudsonAlpha Institute for Biotechnology, Huntsville, Alabama, USA. [5]Department of Neuroscience, Columbia University, New York, New York, USA. Correspondence should be addressed to M.K. (mk2758@columbia.edu) or J.A.G. (jag90@columbia.edu).

Received 26 May; accepted 12 July; published online 7 August 2011; doi:10.1038/ng.902

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

table 1 Overview of exome sequencing data

	Average	Percent
Total sequence (bp)	7,475,900,117	100.00
Aligned sequence (bp)	7,319,357,510	98.34
Aligned paired reads	69,747,339	97.55
Aligned singleton reads	566,300	0.35
Median read depth	65.2×	–
1× coverage	37,191,631	98.84
4× coverage	36,190,388	96.18
8× coverage	34,769,095	92.40
20× coverage	30,428,158	80.87
30× coverage	27,171,231	72.21

variants that would appear *de novo* either from undercalling in the parents or from systematic false positive calls in the subjects (Online Methods). In total, in the affected trios, we observed 34 *de novo* point mutations (33 SNVs and 1 dinucleotide substitution) and 4 *de novo* indel candidates (**Table 2**). Among the 34 *de novo* point mutations, 32 are predicted to be non-synonymous missense mutations and 2 are predicted to be synonymous. Of the 32 non-synonymous mutations, 19 affect evolutionarily conserved residues and are predicted to affect protein function by PolyPhen-2 analysis. Three of the indels result in protein truncations and one causes a single amino acid deletion. Additional analyses of *de novo* SNVs located within the flanking intronic regions identified two SNVs located within predicted donor or acceptor splice sites (**Table 2**). Notably, among the identified exonic SNVs, one synonymous and three non-synonymous SNVs were also located within predicted donor or acceptor splice sites (**Table 2**). Further analysis using the Human Splicing Finder tool (see URLs) showed that two out of the six mutations directly alter splice signals and may interfere with splicing (**Table 2**). By our filtering

table 2 *De novo* mutations identified in 53 schizophrenia trios

Gene	Mutation type	NS or S	PolyPhen-2	Grantham score	phyloP score	Chr. (position)	Nucleotide change	Amino acid change	Diagnosis	Sex	Trio ID
PLCL2	SNV	NS	Probably damaging	112	4.99	3 (17,051,253)	TGT-aGT	p.Cys30Ser	SCZ	M	trio_002
WDR11	SNV	NS	Probably damaging	29	6.30	10 (122,664,879)	CGC-CaC	p.Arg1081His	SCZ	M	trio_011
DPYD	SNV	NS	Probably damaging	125	3.89	1 (97,981,407)	GGA-aGA	p.Gly539Arg	SCZ	M	trio_016
OR4C46	SNV	NS	Probably damaging	125	1.61	11 (51,515,885)	GGA-aGA	p.Gly202Arg	SCZ	F	trio_019
UGT1A3	SNV	NS	Probably damaging	15	0.53	2 (234,637,866)	TTG-aTG	p.Leu32Met	SCZ	F	trio_023
FAM3D	SNV	NS	Probably damaging	194	1.18	3 (58,622,886)	TAC-TgC	p.Tyr147Cys	SCZ	M	trio_024
KLF12	SNV	NS	Probably damaging	112	6.02	13 (74,289,537)	TCT-TgT	p.Ser45Cys	SCZ	M	trio_033
ADCY7	SNV	NS	Probably damaging	56	4.93	16 (50,349,011)	AGC-gGC	p.Ser1020Gly	SCZ	M	trio_038
GPR153	SNV	NS	Probably damaging	89	5.80	1 (6,314,661)	ACC-AtC	p.Thr102Ile	SCZAFF-dpr	M	trio_040
PML	SNV	NS	Probably damaging	81	2.93	15 (74,290,439)	ACG-AtG	p.Thr75Met	SCZAFF-dpr	M	trio_044
SLC26A8	SNV	NS	Probably damaging	56	1.87	6 (35,927,251)	GAG-aAG	p.Glu512Lys	SCZAFF-bp	F	trio_077
CCDC108	SNV	NS	Probably damaging	46	2.43	2 (219,900,235)	AAT-AgT	p.Asn105Ser	SCZ	F	trio_080
TRAK1	SNV	NS	Probably damaging	29	4.80	3 (42,261,055)	CAT-CgT	p.His678Arg	SCZ	F	trio_083
FASTKD5	SNV	NS	Probably damaging	60	5.61	20 (3,128,479)	GCA-GgA	p.Ala413Gly	SCZ	M	trio_089
DGCR2	SNV	NS	Probably damaging	103	6.24	22 (19,028,681)	CCT-CgT	p.Pro429Arg	SCZ	M	trio_091
ACOT6	SNV	NS	Possibly damaging	194	–0.08	14 (74,086,428)	TAT-TgT	p.Tyr170Cys	SCZ	F	trio_047
PITPNM1	SNV	NS/splice[a]	Probably damaging	101	0.30	11 (67,267,884)	CGG-tGG	p.Arg217Trp	SCZ	M	trio_039
NPRL2	SNV	NS/splice[b]	Possibly damaging	56	6.11	3 (50,385,987)	GGC-aGC	p.Gly231Ser	SCZ	F	trio_023
MAGEC1	SNV	NS	Unknown	58	0.37	X (140,993,957)	ACT-AgT	p.Thr256Ser	SCZ	M	trio_003
TRRAP	SNV	NS	Unknown	21	5.12	7 (98,498,329)	ATC-tTC	p.Ile295Phe	SCZ	M	trio_033
COL3A1	SNV	NS/splice[c]	Unknown	155	2.54	2 (189,851,792)	TCT-TtT	p.Ser152Phe	SCZ	M	trio_089
GIF	SNV	NS	Benign	29	0.83	11 (59,603,474)	GTA-aTA	p.Val294Ile	SCZAFF-dpr	F	trio_001
TEKT5	SNV	NS	Benign	89	0.25	16 (10,783,119)	ATC-AcC	p.Ile237Thr	SCZ	M	trio_011
THBS1	SNV	NS	Benign	56	5.66	15 (39,881,442)	GAG-aAG	p.Glu605Lys	SCZ	M	trio_015
PAG1	SNV	NS	Benign	29	0.15	8 (81,905,378)	GTC-aTC	p.Val29Ile	SCZ	M	trio_020
RGS12	SNV	NS	Benign	98	0.06	4 (3,429,844)	CCA-CtA	p.Pro518Leu	SCZAFF-dpr	M	trio_040
SAP30BP	SNV	NS	Benign	56	5.37	17 (73,702,542)	GGC-aGC	p.Gly274Ser	SCZ	F	trio_047
ZNF530	SNV	NS	Benign	46	–0.04	19 (58,118,122)	AGT-AaT	p.Ser410Asn	SCZAFF-bp	F	trio_077
MTOR	SNV	NS	Benign	46	2.82	1 (11,293,489)	AAT-AgT	p.Asn796Ser	SCZ	F	trio_080
INPP5A	SNV	NS	Benign	64	2.62	10 (134,463,942)	GCG-GtG	p.Ala80Val	SCZAFF-bp	F	trio_093
EDEM2	SNV	NS	Benign	83	3.39	20 (33,703,457)	TAC-cAC	p.Tyr469His	SCZ	M	trio_095
CELF2	SNV	S/splice[d]	Coding-synon	–	3.06	10 (11,356,223)	GGT-GGc	p.Gly345Gly	SCZ	M	trio_016
SLC26A7	SNV	S	Coding-synon	–	0.03	8 (92,346,630)	CAG-CAa	p.Gln250Gln	SCZ	M	trio_094
VPS35	SNV	Splice[e]	–	–	–	16 (46,705,610)	C/T	–	SCZ	M	trio_002
ADAMTS3	SNV	Splice[f]	–	–	–	4 (73,185,683)	G/A	–	SCZ	F	trio_023
GPR115	DNV	NS	Probably damaging	99	3.35	6 (47,682,855)	CTC-aaC	p.Leu625Asn	SCZ	F	trio_080
SPATA5	Indel	Amino acid deletion	Damaging	215[g]	–	4 (123,855,728)	TTCTT-caa-CAACA	–	SCZ	F	trio_023
RB1CC1	Indel	Frameshift deletion	Damaging	215	–	8 (53,568,705)	ACTGT-tc-TCTGT	–	SCZ	M	trio_026
LAMA2	Indel	Frameshift deletion	Damaging	215	–	6 (129,835,668)	GGTGG-aagccca-AAGCC	–	SCZ	M	trio_092
ESAM	Indel	Frameshift deletion	Damaging	215	–	11 (124,626,163)	tggac-AGCG-agcgg	–	SCZ	M	trio_042

NS, non-synonymous; S, synonymous; SNV, single nucleotide variant; DNV, dinucleotide variant; chr., chromosome; SCZ, schizophrenia; SCZAFF-dpr, schizoaffective disorder depressed subtype; SCZAFF-bp, schizoaffective disorder bipolar subtype; M, male; F, female.

[a–f]The difference (%) between the mutant and reference sequence in the HSF algorithm-derived consensus values at donor or acceptor splice sites: a = –2.48; b = 56.53; c = –0.68; d = –34.92; e = 0; f = –8.14. [g]We used the maximum Grantham score (215) for the indels.

© 2011 Nature America, Inc. All rights reserved.

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Table 3 Ns:s ratio comparison between *de novo* and rare inherited mutations in schizophrenia trios

	Cases					Controls				
Class	Total number	NS	S	NS:S	P^a	Total number	NS	S	NS:S	P^a
De novo mutations	34	32	2	16.0		7	4	3	1.33	
Novel inherited variants	14,378	8,867	5,511	1.61	0.0002	6,213	3,825	2,388	1.60	0.81
Private inherited variants	6,727	4,223	2,504	1.69	0.0003	3,079	1,956	1,123	1.74	0.73

S, synonymous variants; NS, non-synonymous variants.
[a]*P* value for *de novo* compared to inherited mutations.

criteria, all identified *de novo* mutations were absent in a total of 1,658 control chromosomes (the exomes of the 679 individuals from the 1000 Genomes Project[16] included in dbSNP132, as well as of all 150 unaffected parents in our two cohorts). Overall, we found 27 out of 53 cases (~51%) to carry at least one *de novo* mutational event. This rate is comparable to that reported for 20 parent-child trios with autism spectrum disorders (51%)[8] but is somewhat lower than the rate reported for 10 parent-child trios with intellectual disability (90%)[9]. Ten of the 27 cases carried more than one *de novo* mutation, and the rest each carried a single mutation or indel.

Using the same pipeline and filtering criteria, we identified seven exonic *de novo* SNVs but no indel candidates in 7 out of 22 control subjects. Among these seven *de novo* point mutations, four are non-synonymous missense mutations and three are synonymous mutations. In addition, we identified one *de novo* mutation within a predicted intronic splice site (**Supplementary Table 1**). Overall, 7 out of 22 controls carry at least one *de novo* event, with one control having more than one *de novo* mutation. The fraction is lower when compared to cases, but the difference between the fraction of cases and controls carrying at least one *de novo* event is not statistically significant (Fisher's exact test, $P = 0.2$). There was no difference in the coverage between cases and controls or between trios with and without *de novo* events (**Supplementary Fig. 4**).

The overall *de novo* rate in affected families (0.75 events per family) is comparable to several empirical estimates of the background *de novo* mutation rate[8,16], suggesting that we identified most of the *de novo* events in these trios. Several lines of evidence suggest that the identified mutations have a high likelihood of causation with respect to schizophrenia. First, our screen yielded a ratio of non-synonymous missense ($n = 32$) to synonymous ($n = 2$) *de novo* changes (NS:S ratio) of 16:1, which is considerably higher than the 2.85:1 ratio expected based on the probability of causing an amino acid change under a random model (71.25% non-synonymous missense substitutions and 25% synonymous substitutions[17]). By contrast, the ratio of non-synonymous missense ($n = 4$) to synonymous ($n = 3$) *de novo* changes in the control cohort (NS:S ratio of 1.33:1) is consistent with neutral expectation and is very close to the NS:S ratio reported by the 1000 Genomes Project (1.14–1.45)[16]. Second, we found non-synonymous *de novo* point mutations in large excess compared to neutral ones relative to rare inherited variants, which are less likely to contribute to the pathogenesis in the sporadic cases (**Table 3**). Specifically, we first compared the relative enrichment of non-synonymous *de novo* point mutations to that observed among all novel (that is, not observed in dbSNP132) inherited variants segregating in cases. Our analysis identified a NS:S substitution ratio of 1.61:1, consistent with previous analyses of normal genetic variation[16,18]. Thus, in sporadic schizophrenia cases, rare *de novo* variants are approximately ten times more likely than inherited rare variants to harbor non-synonymous changes (χ^2 test $P = 0.0002$). A similar analysis in the control cohort did not show significant differences in the number of non-synonymous changes between the two types of variants. The NS:S ratios were 1.33 and ~1.60 for rare *de novo* and inherited variants, respectively (relative enrichment 0.83; χ^2 test $P = 0.81$). We obtained similar results when we limited the analysis to private inherited variants (that is, present only in one affected family), which serve as proxies for evolutionarily young mutation events. This analysis yielded a NS:S ratio of ~1.69:1 in cases (relative enrichment 9.5; χ^2 test $P = 0.0003$) and ~1.74 in controls (relative enrichment 0.76; χ^2 test $P = 0.73$). Consistent with expectations that disease mutations have a greater effect on protein function[19], we observed a more striking enrichment when we restricted our analysis to non-synonymous SNVs predicted by PolyPhen-2 to affect protein function. For such changes, the NS:S ratios in schizophrenia cases were 9.5 and ~0.79 for rare *de novo* and private inherited variants, respectively (relative enrichment 12.1; χ^2 test $P < 0.0001$).

We explored further the possibility that *de novo* mutations in cases have a greater potential to affect protein structure and function than private inherited variants by examining the evolutionary conservation of affected nucleotides (using the phyloP score[20]) as well as the potential of the *de novo* protein-altering mutations to affect the structure or function of the resulting proteins (using the Grantham score[21]) (**Table 2**). When we compared the cumulative distribution of these scores between *de novo* and private inherited variants in the sporadic cases cohort (**Fig. 1**), we observed that the distribution of the *de novo* variants was clearly shifted to the right (phyloP $P = 0.0005$ and Grantham $P = 0.14$). Overall, our analysis shows an enrichment of highly conserved and disruptive amino acid changes among *de novo* events and suggests a high likelihood for pathogenicity. Notably, carriers of one or more *de novo* mutations appear to be indistinguishable from other individuals with schizophrenia in terms of sex distribution, clinical presentation and developmental course (**Supplementary Note** and **Supplementary Table 2**).

All the mutations occurred in different genes, precluding a statistical assessment for any specific locus. Identification of recurrent mutations will provide proof of disease causality. With one exception, none of the affected genes has been previously associated with genetic loci or biological pathways unequivocally associated with schizophrenia. We therefore used phyloP and Grantham scores as a guide to prioritize

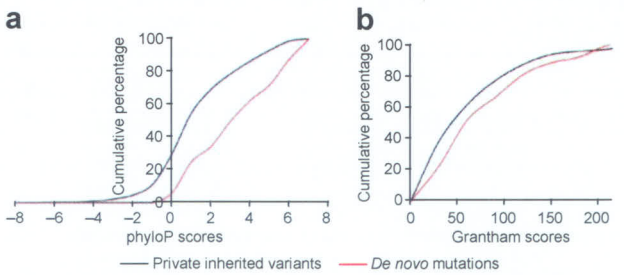


Figure 1 Assessment of the predicted pathogenicity of *de novo* SNVs identified in schizophrenia cases with respect to protein function. Comparison of the distribution of phyloP scores (which depend on the evolutionary conservation of affected nucleotides) (**a**) and Grantham scores (which depend on the properties of the changed residue) (**b**) for *de novo* mutations and private inherited variants found in schizophrenia cases.

© 2011 Nature America, Inc. All rights reserved.

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

for further discussion events that are more likely to be causal. In addition to protein truncating indels (in *LAMA2*, *SPATA5* and *RB1CC1*), there are 12 SNVs with phyloP scores ≥4 and 9 SNVs with Grantham scores ≥100, whereas 3 SNVs (in *DGCR2*, *KLF12* and *PLCL2*) show high values for both scores (**Table 2**). Most notable among the putative pathogenic events is a p.Pro429Arg substitution in *DGCR2* in a male with schizophrenia (**Supplementary Note**). *DGCR2* is located in the 22q11.2 locus and is hemizygously deleted by recurrent *de novo* microdeletions at this locus, which have high penetrance (~30%) and account for up to 2% of sporadic schizophrenia cases. The gene encodes a putative transmembrane adhesion receptor of unknown function[22]. The p.Pro429Arg substitution is located within a conserved domain of the protein (**Supplementary Fig. 2**) and shows one of the highest Grantham and phyloP scores among all the identified changes. Identification of a disruptive *de novo* SNV in *DGCR2* in a case with structurally intact 22q11.2 chromosomes suggests that disruption of this gene may contribute to the elevated schizophrenia risk associated with the 22q11.2 locus. Whether heterozygous deletions or point mutations in the *DGCR2* are sufficient to render the susceptibility to schizophrenia observed in 22q11.2 microdeletion carriers or whether additional genetic interactions are required[23] cannot be resolved until more *DGCR2* mutations are identified and their penetrance is determined. Additional putative pathogenic events were identified in three G-protein–coupled receptor (GPR) genes (*GPR153*, *GPR115* and *OR4C46*)[24,25] as well as in genes encoding proteins thought to either modulate (*RGS12*) or mediate aspects of GPR signaling, such as regulation of cAMP levels (*ADCY7*)[26]. Notably, we recently reported an association between schizophrenia and structural *de novo* mutations in another gene encoding a GPR (*VIPR2*)[27] and showed that these mutations alter cAMP levels. For other genes with high phyloP or Grantham scores, such as *WDR11*, *PLCL2*, *TRAK1*, *KLF12* and *LAMA2*, a potential causal link with schizophrenia is suggested by literature on previously described mutations, model organisms and other functional studies (**Supplementary Note**).

Our work shows that *de novo* protein-altering mutations contribute substantially to the genetic component of schizophrenia and, taken together with previous estimates of the *de novo* CNV rate in the same population[6], indicates that *de novo* mutations account for more than half of the sporadic cases of schizophrenia. Our findings are also in line with results from genome-wide scans for *de novo* CNVs[6,28], or CNVs in general[29,30], supporting the notion that multiple *de novo* genetic variants that affect many different genes contribute to the genetic risk of schizophrenia. The complexity of the neural substrates affected in schizophrenia and other psychiatric disorders offers a large mutational target comprised of many genes. We propose that this large number of targets that, when mutated, can give rise to schizophrenia, along with the relatively high rate of protein-altering mutations empirically shown in this study, provides a plausible explanation for both the high global incidence and the persistence of schizophrenia despite extremely variable environmental factors, severely reduced fecundity and increased mortality. Our findings are an important step toward understanding the pathogenesis of the disease and emphasize the challenge in determining the neural substrates that these diverse genetic risk factors converge upon to generate a common pattern of clinical dysfunction and symptoms[23,31–34].

URLs. Picard, http://picard.sourceforge.net/; SAMtools, http://samtools.sourceforge.net/; PolyPhen-2, http://genetics.bwh.harvard.edu/pph2/; UCSC Table Browser, http://genome.ucsc.edu/cgi-bin/hgTables; The Human Splicing Finder (HSF, Version 2.4.1) software, http://www.umd.be/HSF/; R, http://www.r-project.org/; dbSNP v132, ftp://ftp.ncbi.nlm.nih.gov/snp/organisms/human_9606/VCF/v4.0/00-All.vcf.gz, GATK VCF annotation file for hg19, ftp://gatk-ftp:PH5UH7Pa@ftp.broadinstitute.org/refseq/ ; Dindel, http://sites.google.com/site/keesalbers/soft/dindel.

MetHODs

Methods and any associated references are available in the online version of the paper at http://www.nature.com/naturegenetics/.

Accession codes. Reference sequences are available from NCBI under the following accession codes: *PLCL2*, NM_001144382; *WDR11*, NM_018117; *DPYD*, NM_000110; *OR4C46*, NM_001004703; *UGT1A3*, NM_019093; *FAM3D*, NM_138805; *KLF12*, NM_007249; *ADCY7*, NM_001114; *GPR153*, NM_207370; *PML*, NM_002675; *SLC26A8*, NM_052961; *CCDC108*, NM_152389; *TRAK1*, NM_001042646; *FASTKD5*, NM_021826; *DGCR2*, NM_005137; *ACOT6*, NM_001037162; *PITPNM1*, NM_001130848; *NPRL2*, NM_006545; *MAGEC1*, NM_005462; *TRRAP*, NM_003496; *COL3A1*, NM_000090; *GIF*, NM_005142; *TEKT5*, NM_144674; *THBS1*, NM_003246; *PAG1*, NM_018440; *RGS12*, NM_002926; *SAP30BP*, NM_013260; *ZNF530*, NM_020880; *MTOR*, NM_004958; *INPP5A*, NM_005539; *EDEM2*, NM_001145025; *CELF2*, NM_001083591; *SLC26A7*, NM_134266; *VPS35*, NM_018206; *ADAMTS3*, NM_014243; *GPR115*, NM_153838; *SPATA5*, NM_145207; *RB1CC1*, NM_014781; *LAMA2*, NM_000426; *ESAM*, NM_138961.

Note: Supplementary information is available on the Nature Genetics website.

AcKnowLeDGMentS

We thank all the families who participated in this research. We also thank H. Pretorius and nursing sisters R. van Wyk, C. Botha and H. van den Berg for their assistance with subject recruitment, family history assessments and diagnostic evaluations. We thank Y. Sun for technical assistance with DNA extractions and sample preparations and J. Grun for information technology support. We also thank E. Fledderman and S. Thomas for support of the sequencing studies and M. Robinson for critical project support. This work was supported in part by National Institute of Mental Health (NIMH) grants MH061399 (to M.K.) and MH077235 (to J.A.G.) and the Lieber Center for Schizophrenia Research at Columbia University. B.X. was partially supported by a National Alliance for Research on Schizophrenia and Depression (NARSAD) Young Investigator Award.

AUtHoR contRIBUtIonS

B.X., J.A.G. and M.K. designed the study, interpreted the data and prepared the manuscript. B.X. developed the analysis pipeline and had the primary role in analysis and validation of sequence data. J.L.R. collected the samples and was the primary clinician on the project. S.L. and B.P. performed exome library construction, capture and sequencing. P.D. contributed to the analysis of the data. B.B. contributed to the primary sequence data analysis. S.L. supervised the sequencing project at HudsonAlpha Institute and contributed to the manuscript.

coMPetInG FInAncIAL InteReStS

The authors declare no competing financial interests.

Published online at http://www.nature.com/naturegenetics/.
Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Gottesman, I.I. & Shields, J. A polygenic theory of schizophrenia. *Proc. Natl. Acad. Sci. USA* **58**, 199–205 (1967).
2. Sullivan, P.F., Kendler, K.S. & Neale, M.C. Schizophrenia as a complex trait: evidence from a meta-analysis of twin studies. *Arch. Gen. Psychiatry* **60**, 1187–1192 (2003).
3. Lichtenstein, P. *et al.* Common genetic determinants of schizophrenia and bipolar disorder in Swedish families: a population-based study. *Lancet* **373**, 234–239 (2009).
4. Lupski, J.R. Genomic rearrangements and sporadic disease. *Nat. Genet.* **39**, S43–S47 (2007).
5. Karayiorgou, M. *et al.* Schizophrenia susceptibility associated with interstitial deletions of chromosome 22q11. *Proc. Natl. Acad. Sci. USA* **92**, 7612–7616 (1995).

© 2011 Nature America, Inc. All rights reserved.

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

6. Xu, B. *et al.* Strong association of *de novo* copy number mutations with sporadic schizophrenia. *Nat. Genet.* **40**, 880–885 (2008).
7. Cirulli, E.T. & Goldstein, D.B. Uncovering the roles of rare variants in common disease through whole-genome sequencing. *Nat. Rev. Genet.* **11**, 415–425 (2010).
8. O'Roak, B.J. *et al.* Exome sequencing in sporadic autism spectrum disorders identifies severe *de novo* mutations. *Nat. Genet.* **43**, 585–589 (2011).
9. Vissers, L.E. *et al.* A *de novo* paradigm for mental retardation. *Nat. Genet.* **42**, 1109–1112 (2010).
10. Awadalla, P. *et al.* Direct measure of the *de novo* mutation rate in autism and schizophrenia cohorts. *Am. J. Hum. Genet.* **87**, 316–324 (2010).
11. Abecasis, G.R. *et al.* Genomewide scan in families with schizophrenia from the founder population of Afrikaners reveals evidence for linkage and uniparental disomy on chromosome 1. *Am. J. Hum. Genet.* **74**, 403–417 (2004).
12. Karayiorgou, M. *et al.* Phenotypic characterization and genealogical tracing in an Afrikaner schizophrenia database. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* **124B**, 20–28 (2004).
13. Xu, B. *et al.* Elucidating the genetic architecture of familial schizophrenia using rare copy number variant and linkage scans. *Proc. Natl. Acad. Sci. USA* **106**, 16746–16751 (2009).
14. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
15. Bentley, D.R. *et al.* Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* **456**, 53–59 (2008).
16. 1000 Genomes Project Consortium. *et al.* A map of human genome variation from population-scale sequencing. *Nature* **467**, 1061–1073 (2010).
17. Lynch, M. Rate, molecular spectrum, and consequences of human mutation. *Proc. Natl. Acad. Sci. USA* **107**, 961–968 (2010).
18. Li, Y. *et al.* Resequencing of 200 human exomes identifies an excess of low-frequency non-synonymous coding variants. *Nat. Genet.* **42**, 969–972 (2010).
19. Botstein, D. & Risch, N. Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease. *Nat. Genet.* **33** (suppl.), 228–237 (2003).
20. Pollard, K.S., Hubisz, M.J., Rosenbloom, K.R. & Siepel, A. Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Res.* **20**, 110–121 (2010).
21. Grantham, R. Amino acid difference formula to help explain protein evolution. *Science* **185**, 862–864 (1974).
22. Kajiwara, K. *et al.* Cloning of *SEZ-12* encoding seizure-related and membrane-bound adhesion protein. *Biochem. Biophys. Res. Commun.* **222**, 144–148 (1996).
23. Karayiorgou, M., Simon, T.J. & Gogos, J.A. 22q11.2 microdeletions: linking DNA structural variation to brain dysfunction and schizophrenia. *Nat. Rev. Neurosci.* **11**, 402–416 (2010).
24. Bjarnadóttir, T.K. *et al.* The human and mouse repertoire of the adhesion family of G-protein-coupled receptors. *Genomics* **84**, 23–33 (2004).
25. Gloriam, D.E., Schioth, H.B. & Fredriksson, R. Nine new human Rhodopsin family G-protein coupled receptors: identification, sequence characterisation and evolutionary relationship. *Biochim. Biophys. Acta* **1722**, 235–246 (2005).
26. Cruz, M.T. *et al.* Type 7 adenylyl cyclase is involved in the ethanol and CRF sensitivity of GABAergic synapses in mouse central amygdala. *Front. Neurosci.* **4**, 207 (2011).
27. Vacic, V. *et al.* Duplications of the neuropeptide receptor gene *VIPR2* confer significant risk for schizophrenia. *Nature* **471**, 499–503 (2011).
28. Stefansson, H. *et al.* Large recurrent microdeletions associated with schizophrenia. *Nature* **455**, 232–236 (2008).
29. International Schizophrenia Consortium. Rare chromosomal deletions and duplications increase risk of schizophrenia. *Nature* **455**, 237–241 (2008).
30. Walsh, T. *et al.* Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia. *Science* **320**, 539–543 (2008).
31. Fénelon, K. *et al.* Deficiency of *Dgcr8*, a gene disrupted by the 22q11.2 microdeletion, results in altered short-term plasticity in the prefrontal cortex. *Proc. Natl. Acad. Sci. USA* **108**, 4447–4452 (2011).
32. Sigurdsson, T., Stark, K.L., Karayiorgou, M., Gogos, J.A. & Gordon, J.A. Impaired hippocampal-prefrontal synchrony in a genetic mouse model of schizophrenia. *Nature* **464**, 763–767 (2010).
33. Arguello, P.A. & Gogos, J.A. Cognition in mouse models of schizophrenia susceptibility genes. *Schizophr. Bull.* **36**, 289–300 (2010).
34. Arguello, P.A. & Gogos, J.A. Modeling madness in mice: one piece at a time. *Neuron* **52**, 179–196 (2006).

© 2011 Nature America, Inc. All rights reserved.

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

ONLINe MetHODs

Cohorts. All 53 schizophrenia families were recruited from the Afrikaner population in South Africa. Heritage was established by surname and by having four Afrikaans-speaking grandparents. Informed consent was obtained from all participants. The institutional review committees of Columbia University and the University of Pretoria approved all procedures. Diagnostic evaluations were done in person as previously described[6,13]. Family history was obtained from the proband, each participating parent and additional relatives as needed, by two independent raters, a nursing sister, who recorded pedigree information, and by the clinical interviewer, who inquired in detail about family history during the clinical interview[6,13]. For additional cohort characteristics, see the **Supplementary Note**. The control cohort consisted of 22 families (trios) with established Afrikaner heritage recruited from the Afrikaner community. Paternity and maternity were confirmed before sequencing for all case and control families using the Affymetrix Genome-Wide Human SNP Array 5.0 (refs. 6,13) as well as with a panel of microsatellite markers.

Exome library construction. Exome enrichment was conducted using the SureSelect Human All Exon Target Enrichment System (Agilent Technologies) as described[35]. Briefly, 3 μg of genomic DNA was fragmented by sonication using the Covaris S2 to achieve a uniform distribution of fragments with a mean size of 300 bp. The sonicated DNA was purified using Agencourt's AMPure XP Solid Phase Reversible Immobilization paramagnetic bead (SPRI) followed by polishing of the DNA ends by removing the 3′ overhangs and filling in the 5′ overhangs resulting from sonication using T4 DNA polymerase and Klenow fragment (New England Biolabs). Following end polishing, a single A base was added to the 3′ end of the DNA fragments using Klenow fragment (3′ to 5′ exo minus). This prepared the DNA fragments for ligation to specialized adaptors that have a T-base overhang at their 3′ ends. The end-repaired DNA with a single A-base overhang was ligated to the Illumina paired-end adaptors in a standard ligation reaction using T4 DNA ligase and 2–4 μM final adaptor concentration, depending on the DNA yield following purification after the addition of the A base. Following ligation, the samples were purified using SPRI beads, quality controlled by assessment on the Agilent Bioanalyzer and then amplified by six cycles of PCR to maintain complexity and avoid bias caused by amplification.

Library capture and sequencing. We prepared 500 ng of amplified, purified DNA (DNA library) for hybridization by adding the DNA library to Agilent blocking reagents, denaturing at 95 °C and incubating at 65 °C. All subsequent steps were performed at 65 °C. Hybridization buffer was added to the prepared library, and the entire mix was then added to an aliquot of the Agilent SureSelect Capture Library and mixed. The DNA library and biotin-labeled capture library were hybridized by incubation at 65 °C for 24 h. Following hybridization, streptavidin-coated magnetic beads were used to purify the RNA:DNA hybrids formed during hybridization. The RNA capture material was digested by acid hydrolysis following elution from the purification beads. The neutralized captured DNA was purified, desalted and amplified by 12 cycles of PCR using Herculase II Fusion DNA polymerase. The libraries were purified following amplification, and the library was assessed using the Agilent Bioanalyzer. A single peak between 300–400 bp indicates a properly constructed and amplified library ready for sequencing. Final quantitation of the library was performed using the Kapa Biosciences Real-time PCR assay, and appropriate amounts of the library were loaded onto the Illumina flowcell for sequencing by paired-end 50 nt sequencing on the Illumina HiSequation (2000) instrument. Sequencing was performed largely as described[16]. Following dilution to 10 nM final concentration based on the real-time PCR and bioanalyzer results, the final library stock was then used in paired-end cluster generation at a final concentration of 6–8 pM to achieve a cluster density of 600,000/mm^2 (on the Illumina HiSequation (2000) instrument). Following cluster generation, 50 nt paired-end sequencing was performed using the standard Illumina protocols.

Exome data analysis for *de novo* coding point mutations, indels and splice site mutations. Raw sequencing data for each individual were mapped to the human reference genome (build hg19) using the Burrows-Wheeler Aligner (BWA v0.5.81536)[14]. The BWA aligned sequencing reads were processed by Picard (see URLs) to label the PCR duplicates. The Genome Analysis Toolkit (GATK, version 5091) was then used to remove duplicates, perform local realignment and map quality score recalibration to produce a 'cleaned' BAM file for each individual. SNP calls were made by the Unified Genotyper module in GATK using the 'cleaned' BAM files in batch fashion (90 samples per batch). The resulting Variant Call Format (VCF, version 4.0) files were annotated using the GenomicAnnotator module in GATK to identify and label the called variants that are within the targeted coding regions and overlap with known and likely benign SNPs reported in dbSNP v132 (see URLs). The annotated VCF files were then filtered using the GATK variant filter module with a hard filter setting and a custom script for initial filtering. Variant calls that failed to pass the following filters were eliminated from the call set: (i) MQ0 >= 4 && ((MQ0 / (1.0 * DP)) > 0.1); (ii) QUAL < 30.0 || QD < 5.0 || HRun > 5 || SB > 0.00; (iii) cluster size 10; (iv) contain dbSNP id; and (v) outside the targeted regions. Combined VCF files were then split into individual files, and variants in each offspring were compared to variants present in their parents using a custom script pipeline in order to determine the inheritance pattern and annotate *de novo* mutations.

Because the GATK Unified Genotyper is set to maximize the sensitivity of variant calls, it allows for a substantial portion of false positives among candidate variants even following the initial filtering process. To address this issue and eliminate potential false positive calls in the offspring and false negative calls in the parents, we took advantage of the inheritance information provided by our family design and revalidated all variants identified using the mpileup module in the SAM tools (see URLs) according to the following rules: (i) the forward reference (fr) count (the number of forward reads that match the reference base at this locus), the reverse reference (rr) count (the number of reverse reads that match the reference base at this locus), the forward non-reference (fnr) count (the number of forward reads that do not match the reference base at this locus) and the reverse non-reference (rnf) count (the number of reverse reads that do not match the reference base at this locus) in the offspring must be two or greater; (ii) total read depth in both parents must be ten or greater; (iii) both fr and rr count in both parents must be two or greater; (iv) either fnr or rnr count in both parents must be zero; (v) the fnr and rnr count to total count ratio in the parental population (defined as all 150 parental samples sequenced) must be less than $1/2n$, where n is the population size; and (vi) if any of rules i–v was violated, the sequence information was considered insufficient to make a *de novo* call at this locus.

Indel calls were made by the Dindel software using one 'cleaned' BAM file per run. The resulting VCF files were used to determine inheritance patterns using the same procedure described above for the point mutations. To determine potential mutations at splice-donor or acceptor sites, GATK variant calls were made in a batch fashion (with 90 samples per batch) that covered each target coding region and 50-bp flanking segments in each direction. The variants in the resulting VCF files were annotated according to the GATK annotation file for hg19 (see URLs).

The PolyPhen-2 (see URLs) online server was used to determine the non-synonymous and synonymous nature of the mutations and to predict their functional impact by further classifying them as non-tolerated (damaging) or benign at a given site. The Grantham score for each coding variant was determined by the Grantham matrix table[21]. The phyloP score for each coding variant was extracted from the 'phyloP46wayAll' table in the UCSC Table Browser (see URLs). The Human Splicing Finder (HSF, Version 2.4.1) software (see URLs) was used to predict potential functional impact of the mutations at splice sites.

Statistics. The Kolmogorov-Smirnov test was used to compare the distribution of phyloP and Grantham scores among *de novo* or private inherited mutations in cases. The Kolmogorov-Smirnov test was conducted using R (see URLs). Fisher's exact test or χ^2 test with Yates' correction was used for the analysis of contingency tables, depending on the sample sizes.

***De novo* mutation validation.** Candidate *de novo* variants were tested using standard Sanger sequencing on an ABI 3730xl DNA Analyzer to validate presence of each mutation in the subjects and the absence of each in the parental genomes, by designing custom primers (Sigma) based on ~500 bp of genomic sequence flanking each variant. *De novo* occurrence of mutations was not confirmed for 6 out of 46 candidate alterations in cases and 2 out of 9 candidate alterations in controls.

35. Gnirke, A. *et al.* Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nat. Biotechnol.* **27**, 182–189 (2009).

© 2011 Nature America, Inc. All rights reserved.

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

nature genetics

De novo gene mutations highlight patterns of genetic and neural complexity in schizophrenia

Bin Xu[1], Iuliana Ionita-Laza[2], J Louw Roos[3,4], Braden Boone[5], Scarlet Woodrick[1,6], Yan Sun[1], Shawn Levy[5], Joseph A Gogos[6,7] & Maria Karayiorgou[1]

To evaluate evidence for *de novo* etiologies in schizophrenia, we sequenced at high coverage the exomes of families recruited from two populations with distinct demographic structures and history. We sequenced a total of 795 exomes from 231 parent-proband trios enriched for sporadic schizophrenia cases, as well as 34 unaffected trios. We observed in cases an excess of *de novo* nonsynonymous single-nucleotide variants as well as a higher prevalence of gene-disruptive *de novo* mutations relative to controls. We found four genes (*LAMA2*, *DPYD*, *TRRAP* and *VPS39*) affected by recurrent *de novo* events within or across the two populations, which is unlikely to have occurred by chance. We show that *de novo* mutations affect genes with diverse functions and developmental profiles, but we also find a substantial contribution of mutations in genes with higher expression in early fetal life. Our results help define the genomic and neural architecture of schizophrenia.

© 2012 Nature America, Inc. All rights reserved.

Schizophrenia is a severe psychiatric disorder with a strong genetic component[1]. Whereas the contribution of rare *de novo* copy-number variants (CNVs) to schizophrenia risk is well established[2–5], the contribution of *de novo* nucleotide-level variants has not yet been probed extensively[6,7]. We completed exome sequencing of 146 Afrikaner proband-parent family trios of subjects with a diagnosis of schizophrenia or schizoaffective disorder (SCZAFF)[3,8] and 34 unaffected control trios (53 and 22 trios, respectively, have been described previously[7]). We also sequenced the exomes of 85 US family trios of subjects with schizophrenia or SCZAFF (Online Methods and **Supplementary Table 1**). None of the subjects carry rare *de novo* CNVs (of ≥30 kb in size), as determined in previous CNV scans of these cohorts[3,4]. We used a previously described analytical pipeline[7] and a series of filters, including final validation by Sanger sequencing of all family members (Online Methods and **Supplementary Fig. 1**). More than 90% of single-nucleotide variants (SNVs) and 20% of insertions and/or deletions (indels) were validated (**Supplementary Table 2**).

In our control cohort, we identified 16 exonic *de novo* SNVs and 1 protein-truncating indel in 34 subjects (0.50 events per sample) (**Table 1**). The point mutation rate in the captured coding sequence was 1.28×10^{-8} mutations per base per generation. Among the 16 *de novo* SNVs, 11 were predicted to be nonsynonymous, missense changes, and 5 were predicted to be synonymous changes. The observed nonsynonymous-to-synonymous ratio in our control cohort of 2.20 is consistent with neutral expectation (2.23)[9,10] and with those reported for control samples (unaffected siblings) in three studies of independent subsets of the Simons Simplex Collection (SSC) (2.23, $n = 200$ (ref. 11); 2.11, $n = 31$ (ref. 12); 2.99, $n = 343$ (ref. 13); average nonsynonymous-to synonymous ratio = 2.65, $n = 574$ (combined SSC control group)).

In the 146 Afrikaner probands, we observed 93 confirmed *de novo* exonic point mutations (92 SNVs and 1 dinucleotide substitution) and 9 confirmed *de novo* indel events (**Table 1**). Six of the indels resulted in protein truncation, and three resulted in deletion of a single amino acid. Additional query of *de novo* SNVs located within the flanking intronic regions identified three SNVs that altered canonical splice sites (**Table 1**) and seven that altered the consensus sequences flanking canonical splice sites (Online Methods). Overall, 73 of 146 cases (50%) carried at least 1 likely functional *de novo* event (nonsynonymous, indel or splice-site mutation). The point mutation rate in the captured coding sequence was 1.73×10^{-8} mutations per base per generation, not significantly different than the one observed in our control sample. Moreover, we found no differences in the distribution or frequency of multiple *de novo* point mutations in cases versus controls (**Supplementary Fig. 2**). The 93 identified *de novo* point mutations included 80 nonsynonymous and 13 synonymous changes. The nonsynonymous-to-synonymous ratio of 6.15 was higher than neutral expectation (2.23)[9,10] ($P = 1.92 \times 10^{-4}$, two-sided exact binomial test). To further assess the statistical significance of the observed enrichment of nonsynonymous variants in cases, we performed permutation testing by randomly permuting the case-control labels of the trios in our data set (**Supplementary Note**). On the basis of 100,000 such permuted data sets, we obtained permutation one-sided $P = 0.033$. In contrast, analysis of the nonsynonymous-to-synonymous ratios among private inherited variants (variants present only in one family and serving as proxies for evolutionarily young mutation events) did not show any significant difference in cases versus controls (**Supplementary Table 3**). Indeed, *de novo* variants were

[1]Department of Psychiatry, Columbia University, New York, New York, USA. [2]Department of Biostatistics, Columbia University, New York, New York, USA. [3]Weskoppies Hospital, Pretoria, South Africa. [4]Department of Psychiatry, University of Pretoria, Pretoria, South Africa. [5]HudsonAlpha Institute for Biotechnology, Huntsville, Alabama, USA. [6]Department of Physiology & Cellular Biophysics, Columbia University, New York, New York, USA. [7]Department of Neuroscience, Columbia University, New York, New York, USA. Correspondence should be addressed to M.K. (mk2758@columbia.edu) or J.A.G. (jag90@columbia.edu).

Received 17 September; accepted 26 September; published online 3 October 2012; doi:10.1038/ng.2446

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

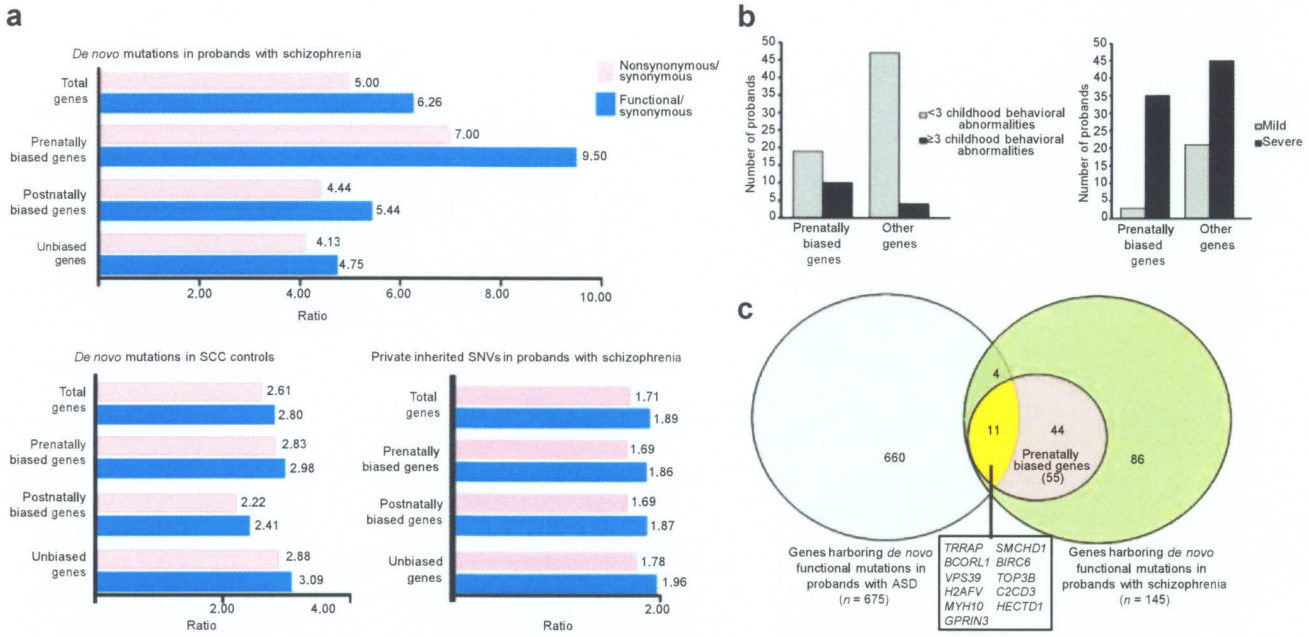


Figure 1 Enrichment of nonsynonymous or functional *de novo* variants according to temporal expression profiles of genes mutated in schizophrenia. (**a**) We grouped the genes determined to have *de novo* mutations in the combined schizophrenia cohorts into three classes—prenatally biased, prenatal brain expression bias; postnatally biased, postnatal brain expression bias; and unbiased—according to their temporal expression trajectory in reference to a previously described global expression switch occurring before birth. We determined the relative enrichment of nonsynonymous or, more generally, functional *de novo* variants over synonymous ones. The relative enrichment in differentially regulated mutated genes is shown for *de novo* mutations. Nonsynonymous variants include missense and nonsense SNVs; functional variants include nonsynonymous, indel and splice-site mutations. Indel data were not available for transmitted variants. (**b**) Left, an elevated frequency of cases who had multiple (≥3) childhood behavioral abnormalities (**supplementary Note**) was found among carriers of *de novo* functional mutations in prenatally biased genes (10/29 compared to 4/51 cases carrying mutations in genes with no bias toward fetal brain expression; 4.4-fold enrichment, $P = 0.0047$, Fisher's exact test). Right, an elevated frequency of cases with severe disease functional outcome (**supplementary Note**) was found among carriers of *de novo* functional mutations in prenatally biased genes (35/38 compared to 45/66 cases carrying mutations in genes with no fetal brain expression bias; 1.35-fold enrichment, $P = 0.007$, Fisher's exact test). (**c**) Venn diagrams depicting the overlap between either all genes with *de novo* mutations in schizophrenia ($n = 145$) or only genes with prenatal expression bias harboring functional *de novo* mutations identified in the Afrikaner and US probands ($n = 55$) and genes with *de novo* mutations identified in ASD exome scans[10–13] ($n = 675$).

~3.7 times more likely than rare inherited variants to harbor nonsynonymous changes in cases ($P < 0.0001$, χ-squared test) but not in controls ($P < 0.77$). In addition to enrichment in nonsynonymous events, we observed a 2.6-fold enrichment in the rate of *de novo* events that almost certainly lead to loss of function (nonsense, frameshift indels and canonical splice-site mutations) in cases (11/146; 7.5%) versus controls (1/34; 2.9%). Overall, using permutation testing, we found a significant enrichment of likely functional *de novo* variants over synonymous ones in Afrikaner cases versus controls (permutation one-sided $P = 0.017$) (**Table 1** and **Supplementary Note**).

Analysis of the US cohort revealed a point mutation rate of 1.73×10^{-8} mutations per base per generation, similar to that observed in the Afrikaner cases (**Table 1**). The nonsynonymous-to-synonymous ratio of 3.42 (**Table 1**) is higher than neutral expectation (2.23)[9,10], but the difference is not statistically significant. In addition to the smaller size of the US cohort, this is likely due to greater uncertainty about family history of schizophrenia in that cohort, meaning that it may not be entirely depleted of familial cases, consistent with a lower *de novo* CNV rate[4] compared to the Afrikaner cohort[3]. By contrast, when comparing the rate of loss-of-function *de novo* events per case (8/85; 9.4%), we observed a 3.2-fold enrichment relative to controls, similar to the enrichment observed in the Afrikaner sample.

In the combined sample of 231 affected families, the nonsynonymous-to-synonymous ratio for *de novo* variants (4.84; **Table 1**) remained higher compared to neutral expectation[9] ($P = 2.1 \times 10^{-4}$, two-sided exact binomial test). There was a differential enrichment of nonsynonymous events between *de novo* and rare, private inherited events ($P < 0.0001$). Notably, as in both individual cohorts, the number of loss-of-function events per affected trio (20/231; 8.7%) was 2.8 times higher than in control trios (1/34, 2.9%), strongly supporting a role for gene-disrupting mutations in disease (**Tables 1** and **2**). Permutation testing showed a significant enrichment of likely functional *de novo* variants over synonymous ones (permutation one-sided $P = 0.026$; **Supplementary Note**). We estimate that 46% of all likely functional *de novo* mutations identified represent genuine risk variants (**Supplementary Note**).

Considering phenotypic correlates, we observed a correlation between paternal age at the proband's birth and the number of *de novo* events per offspring (**Supplementary Fig. 3**) but did not find any other significant differences between proband carriers and non-carriers of *de novo* functional mutations (**Supplementary Tables 4** and **5** and **Supplementary Note**).

Genes affected by *de novo* variants in our study were not significantly over-represented in two previously established comprehensive lists of synaptic genes that encode proteins localized to the presynaptic terminus or postsynaptic density of neurons (**Supplementary Note**). In addition, pathway analyses using the Database for Annotation, Visualization and Integrated Discovery (DAVID) Annotation Tool did not identify any significantly enriched functional clusters. Evaluation of protein-protein

© 2012 Nature America, Inc. All rights reserved.

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

table 1 Distribution of *de novo* events in family cohorts

	Total number of *de novo* events (number of events/subject)			
Variant type	Afrikaner cases *n* = 146	US cases *n* = 85	Total cases *n* = 231	Afrikaner controls *n* = 34
SNVs	93 (0.64)	53 (0.62)	146 (0.63)	16 (0.47)
Nonsynonymous	80 (0.55)	41 (0.48)	121 (0.52)	11 (0.32)
Synonymous	13 (0.09)	12 (0.14)	25 (0.11)	5 (0.15)
Nonsynonymous-to-synonymous ratio	6.15	3.42	4.84	2.20
Nonsense	2 (0.01)	4 (0.05)	6 (0.03)	0 (0.00)
Canonical splice site	3 (0.02)	2 (0.02)	5 (0.02)	0 (0.00)
Splice consensus	7 (0.05)	4 (0.05)	11 (0.05)	1 (0.03)
All *de novo* indels	9 (0.06)	4 (0.05)	13 (0.06)	1 (0.03)
No frameshift	3 (0.02)	1 (0.01)	4 (0.02)	0 (0.00)
Frameshift	6 (0.04)	3 (0.04)	9 (0.04)	1 (0.03)
All loss of function	11 (0.08)	9 (0.11)	20 (0.09)	1 (0.03)
All likely functional	99 (0.68)	51 (0.6)	150 (0.65)	13 (0.38)
Functional-to-synonymous ratio	7.62	4.25	6	2.6

interactions showed a significantly greater connectivity among mutational targets than would be expected by chance ($P < 0.05$). Multigene protein-protein interaction clusters included one centered on *MTOR* and one centered on *CANX*, which includes genes encoding extracellular matrix and cell adhesion proteins (**Supplementary Fig. 4**), suggesting that diverse genes associated with schizophrenia risk may converge on a shorter list of functional modules.

We then examined to what extent the enrichment in functional *de novo* events is determined by the developmental pattern of brain expression of the mutated genes[14,15] (**Supplementary Table 6** and **Supplementary Note**). In both the hippocampus (**Fig. 1a** and **Supplementary Fig. 5**) and dorsolateral prefrontal cortex (DLPFC) (**Supplementary Fig. 6**), two brain areas implicated in schizophrenia[16], the highest effect size was observed for genes showing the highest expression during the prenatal period (**Fig. 1a** and **Supplementary Fig. 6**). The nonsynonymous-to-synonymous ratio for genes with prenatal expression bias was significantly higher than expected for neutral changes (hippocampus, $P = 0.004$; DLPFC, $P = 0.004$). In comparison, the corresponding P values for genes with postnatal expression bias were 0.06 and 0.04, whereas genes with unbiased expression did not show significant differences ($P = 0.12$ and 0.26) (**Fig. 1a** and **Supplementary Fig. 6**). Similar analysis of *de novo* variants in the SSC control group or of private transmitted variants in Afrikaner and US probands showed no difference in the distribution of effect sizes in any of the temporal trajectories tested (**Fig. 1a**).

Notably, there was a functional correlation between the prenatal expression bias of the mutated genes and the neurodevelopmental impact of the corresponding mutations. Specifically, among cases carrying *de novo* mutations, those with mutations in genes with a bias toward prenatal expression were more likely to have had multiple (≥3) behavioral abnormalities in childhood (before the age of 10)[17] compared to the cases carrying *de novo* mutations in genes with no bias toward fetal brain expression, as well as worse functional outcome following disease onset (**Fig. 1b** and **Supplementary Note**). In addition, comparison of all genes with functional *de novo* events identified in our families with schizophrenia ($n = 145$) with those identified in families with autism spectrum disorder (ASD) ($n = 675$)[10–13] identified 15 shared genes (**Fig. 1c**), an overlap within the range expected by chance ($P = 0.29$; **Supplementary Note**). However, 11 of the 15 shared genes (73%) are included in our list of genes that show a bias toward prenatal expression (**Supplementary Table 7**). The probability that this overlap arose by chance is very low ($P = 0.004$).

The majority of mutated genes with prenatal expression bias are highly expressed during the first and second trimesters of pregnancy (**Supplementary Fig. 7**) and show an over-representation of genes encoding nuclear factors involved in chromatin remodeling, nuclear transport and transcriptional control, as well as in protein translation and degradation (**Supplementary Tables 6** and **8**). They also include genes that code for factors involved in cell-cell and cell-matrix interactions, a subset of which interact with two key adhesive proteins, THBS1 (ref. 18) and ITGA6 (ref. 19), involved in synaptogenesis, axonal growth and cortical layering (**Supplementary Fig. 4**). We confirmed that genes with prenatal expression bias were highly enriched for genes that are targeted by one or more microRNAs (miRNAs)[15] (**Supplementary Note**) and also found a nominally significant enrichment of hsa-mir-367 and hsa-mir-1244 targets (**Supplementary Tables 9** and **10**). By comparison, the risk conferred by mutated genes with bias toward postnatal expression was related to

table 2 Loss-of-function mutations in schizophrenia probands

ID	Schizophrenia cohort	Mutation type	Gene symbol	Loss-of-function mutations in SA/SSC controls
trio_090	US	Frameshift	*XPR1*	–
trio_107	US	Frameshift	*CCDC39*	+[a]
trio_121	US	Frameshift	*KDM5C*	+[a]
trio_005	Afrikaner	Frameshift	*KIAA0467*	–
trio_020	Afrikaner	Frameshift	*HIST1H1E*	–
trio_026	Afrikaner	Frameshift	*RB1CC1*	–
trio_042	Afrikaner	Frameshift	*ESAM*	–
trio_092	Afrikaner	Frameshift	*LAMA2*	–
trio_027	Afrikaner	Frameshift	*DDHD2*	–
trio_101	US	Nonsense	*SSBP3*	–
trio_118	US	Nonsense	*NUP54*	–
trio_124	US	Nonsense	*DPYD*	–
trio_128	US	Nonsense	*STAP2*	–
trio_053	Afrikaner	Nonsense	*URB2*	–
trio_085	Afrikaner	Nonsense	*RARG*	–
trio_018	Afrikaner	Splice site[b]	*SYNGAP1*	–
trio_072	Afrikaner	Splice site[b]	*BRPF1*	–
trio_111	US	Splice site[b]	*PRDX6*	–
trio_103	US	Splice site[b]	*NLRC5*	–
trio_016	Afrikaner	Splice site[b]	*CUGBP2*	–

[a]From ref. 13. [b]Canonical splice site.

© 2012 Nature America, Inc. All rights reserved.

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

their involvement in intracellular signaling processes (GTPase, diacylglycerol (DAG) or calcium signaling; **Supplementary Table 8**), which regulate diverse aspects of neuronal connectivity.

We next set out to evaluate which of the individual genes that are mutated in schizophrenia are more likely to confer disease risk. We found four genes altered by two *de novo* events each in unrelated probands (three of these genes were affected in cases across the two different populations tested) (**Table 3** and **Supplementary Note**). These genes were mutated twice by a combination of a nonsense and a missense *de novo* SNV (*DPYD*) or a combination of a splice-site mutation (**Supplementary Tables 11** and **12**) with either a *de novo* missense SNV (*TRRAP* and *VPS39*) or an indel (*LAMA2*). None of these genes were affected in the Afrikaner or the SSC control group, and no such mutational combinations were reported for any gene among the 488 with *de novo* events in the SSC control group. Given the number of *de novo* mutations in our data set, observation of four such recurrent events is associated with a *P* value of 0.002 (**Supplementary Note**), with the occurrence of one of the four combinations (in *LAMA2*) being individually statistically significant ($P = 0.017$).

LAMA2 encodes the laminin a2 chain, which constitutes one of the subunits of laminin 2 and 4 and binds to ITGA6, which is encoded by a gene with prenatal expression bias that was mutated in our cases (**Supplementary Fig. 4**). The indel in *LAMA2* disrupts a key C-terminal domain, and the splice-site mutation affects a highly conserved nucleotide at position −1 relative to the canonical splice acceptor AG motif and is expected to disrupt splicing (**Supplementary Table 12**). Homozygous mutations in *LAMA2* lead, with variable penetrance, to congenital muscular dystrophy characterized by central nervous system (CNS) involvement, including white matter abnormalities, cognitive impairment, seizures and neuronal migration defects[20]. In addition, a *de novo* mutation in another member of the laminin gene family, *LAMA1* was described in another schizophrenia cohort[6].

DPYD encodes dihydropyrimidine dehydrogenase, the initial and rate-limiting factor in the pathway of pyrimidine catabolism[21] that also modulates production of b-alanine, a neuromodulator of inhibitory transmission in the brain[22]. We identified one missense and one nonsense SNV in the Afrikaner and US cohorts, respectively. Abnormal urinary excretion of thymine and uracil confirmed DPYD deficiency in the carrier of the missense mutation (**Supplementary Fig. 8**). Heterozygous deletions, either encompassing or within *DPYD*, as well as altered expression[23], have been described in ASD[24] and intellectual disability[25]. Neither autistic features nor intellectual disability were present in carriers of the *DPYD* mutation in our cohorts, suggesting variable expressivity. A genome-wide association study (GWAS) mega-analysis in schizophrenia identified the strongest association ($P = 1.6 \times 10^{-11}$) at rs1625579, a variant located at 1p21.3 in the intron of *MIR137* (ref. 26) and within a haplotype block ($D' > 0.9$) that extends to the 5′ exons of *DPYD* (**Supplementary Fig. 9**). We did not find mutations in *MIR137*; therefore, association with rs1625579 may reflect the contribution of *DPYD* variants to disease.

The splice-site mutation in *TRRAP* affects position +6 relative to the splice junction site and is predicted to disrupt splicing and binding of the SRP55 splicing factor within a splicing enhancer. Notably, *TRRAP* and *VPS39* are also mutated in ASD cases[12].

We also compared the identified functional *de novo* mutations to the *de novo* CNVs identified previously in our two cohorts (22 CNVs affecting 156 genes)[3,4]. Five genes (*DGCR2, TOP3B, CIT, STAG1* and *SMAP2*) were altered by both *de novo* SNVs and CNVs (**Table 3**), two of them in affected individuals across the two different populations tested. Two of these genes are within the 22q11.2 schizophrenia susceptibility locus.

Our findings implicate contribution from a diverse set of *de novo* mutations to the genomic architecture of schizophrenia in the context of a mutation-selection balance model and highlight the importance of using family samples where disease history has been thoroughly ascertained. Focusing on our comprehensively ascertained Afrikaner cohort, we estimate that at least 17.6% of sporadic cases carry a *de novo* pathogenic exonic mutation (**Supplementary Note**), and at least 9.9% carry a *de novo* CNV[3]. Thus, such mutations may contribute to risk in approximately one-fourth to one-third of all sporadic cases. Given that results from scans of non-exonic

table 3 Genes altered by recurrent *de novo* events

Sample ID	Gene symbol	Locus	Mutation type	DNA change	RNA or amino-acid change	Cohort	Gene name
Genes altered by recurrent *de novo* sNVs and/or indels							
trio_124	*DPYD*		Nonsense	c.1863G>A	p.Trp621*	US	Dihydropyrimidine dehydrogenase
trio_016	*DPYD*	1p21.3	Missense	c.1615G>A	p.Gly539Arg	Afrikaner	
trio_092	*LAMA2*		Frameshift	c.9139_9146del7	p.Ser3050Thrfs*27	Afrikaner	
trio_049	*LAMA2*	6q22.13	Splice site[a]	c.4718–3	Splice-site mutation	Afrikaner	Laminin, a2
trio_033	*TRRAP*		Missense	c.883A>T	p.Ile295Phe	Afrikaner	Transformation/transcription domain–associated protein
trio_099	*TRRAP*	7q22.1	Splice site[a]	c.7223.+6	Splice-site mutation	US	
trio_120	*VPS39*		Missense	c.2368C>T	p.Arg790Cys	Afrikaner	Vacuolar protein sorting 39 homolog (*S. cerevisiae*)
trio_125	*VPS39*	15q15.1	Splice site[a]	c.441+8	Splice-site mutation	US	
Genes altered by both *de novo* sNVs and *de novo* CNVs							
trio_091	*DGCR2*		Missense	c.1163C>G	p.Pro388Arg	Afrikaner	DiGeorge syndrome critical region gene 2
	DGCR2	22q11.2	CNV (del)[b]			Afrikaner	
trio_064	*TOP3B*		Missense	c.1415G>A	p.Arg472Gln	Afrikaner	
	TOP3B	22q11.2	CNV (del)			Afrikaner	Topoisomerase (DNA) III b
trio_121	*CIT*		Missense	c.238T>C	p.Tyr80His	US	Citron (rho-interacting, serine/threonine kinase 21)
	CIT	12q24.23	CNV (dup[c])[b]			Afrikaner	
trio_111	*STAG1*		Missense	c.667A>T	p.Thr223Ser	US	
	STAG1	3q22.3	CNV (del)[b]			Afrikaner	Stromal antigen 1
trio_078	*SMAP2*		Missense	c.896G>A	p.Ser299Asn	Afrikaner	
	SMAP2	1p34.2	CNV (dup)[b]			Afrikaner	small ArfGAP2

[a]Consensus splice-site mutation. Mutations in *LAMA2* and *TRRAP* are predicted to be damaging (**supplementary tables 11** and **12**). [b]From ref. 3. [c]Intragenic duplication. Del, deletion; dup, duplication.

© 2012 Nature America, Inc. All rights reserved.

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

regions are still forthcoming, this is likely an underestimate. Our findings also contribute to the understanding of the neural architecture of schizophrenia risk. Given that we estimate the number of schizophrenia risk loci at more than 850 (**Supplementary Note**), our findings imply an exquisite sensitivity of the neural circuits underlying susceptibility to schizophrenia to precise levels or activity of many diverse proteins and signaling modules and suggest that focusing on circuits may be more commensurate with the heterogeneity of schizophrenia than other proposed mechanisms that concentrate on specific neurotransmitters or cell types[27]. In addition, we show that, in determining disease risk, not only the function of the mutated gene but also the timing of the genetic insult may be of critical importance. Specifically, although *de novo* mutations affect genes with diverse functions and developmental profiles, we describe a substantial contribution of mutations in developmentally regulated genes with higher expression during early and mid-stage fetal life and show that such mutations are enriched among adult cases with prominent early prepsychotic, deviant behaviors. Notably, a bias towards prenatal expression was also demonstrated for genes affected by multiple types of schizophrenia-associated genetic variation[28]. Our findings provide a mechanistic context to interpret epidemiological correlations among various prenatal environmental insults during the first and second trimesters of pregnancy and risk for schizophrenia[29]. Moreover, the fact that expression of many prenatally biased genes is under strict miRNA control may explain emerging links between miRNA dysregulation and psychiatric disorders[30]. The challenge remains to identify the affected biological processes and neural circuits and to determine how they are affected. Unbiased network-based approaches as well as animal and cellular models of recurrent mutations will be invaluable in reaching this goal[31].

URLs. dbSNP132, http://www.ncbi.nlm.nih.gov/projects/SNP/; SAMtools, http://samtools.sourceforge.net/; refGene-big-table-hg19.txt, http://www.broadinstitute.org/gatk/; PolyPhen-2, http://genetics.bwh.harvard.edu/pph2/; UCSC Table Browser, http://genome.ucsc.edu/cgi-bin/hgTables; HSF, Version 2.4.1, http://www.umd.be/HSF/; DAVID Functional Annotation Bioinformatics Microarray Analysis, http://david.abcc.ncifcrf.gov/; DAPPLE, http://www.broadinstitute.org/mpg/dapple/; HBT, http://hbatlas.org/.

METHODS

Methods and any associated references are available in the online version of the paper.

Accession codes. Exome sequencing data are available upon request.

Note: Supplementary information is available in the online version of the paper.

ACKNOWLEDGMENTS

We are enormously grateful to all the families who participated in this research. We thank H. Pretorius and nursing sisters R. van Wyk, C. Botha and H. van den Berg for their assistance with subject recruitment, family history assessments and diagnostic evaluations. We thank S.L. Lundy for valuable assistance with clinical database maintenance and L. Rodriguez-Murillo for help with **Supplementary Figure 9**. We also thank B. Plummer and M. Robinson and the HudsonAlpha Genomics Services Laboratory for experimental support. Finally, we thank L.J. Mienie for the thymine loading test. This work was partially supported by National Institute of Mental Health (NIMH) grants MH061399 (to M.K.) and MH077235 (to J.A.G.) and the Lieber Center for Schizophrenia Research at Columbia University. B.X. was partially supported by a National Alliance for Research in Schizophrenia and Depression (NARSAD) Young Investigator Award.

AUTHOR CONTRIBUTIONS

B.X., J.A.G. and M.K. designed the study, interpreted the data and prepared the manuscript. B.X. developed the analysis pipeline and had the primary role in the analysis and validation of sequence data. I.I.-L. performed statistical analysis of the sequence data. J.L.R. contributed to sample collection and clinical characterization. S.W. and Y.S. contributed to sample preparation and *de novo* mutation validation. B.B. performed exome library construction, capture and sequencing and initial analysis of SNV genotyping and indel variant calls. S.L. supervised the sequencing project at the HudsonAlpha Institute.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Published online at http://www.nature.com/doifinder/10.1038/ng.2446.
Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Rodriguez-Murillo, L., Gogos, J.A. & Karayiorgou, M. The genetic architecture of schizophrenia: new mutations and emerging paradigms. *Annu. Rev. Med.* **63**, 63–80 (2012).
2. Karayiorgou, M. *et al.* Schizophrenia susceptibility associated with interstitial deletions of chromosome 22q11. *Proc. Natl. Acad. Sci. USA* **92**, 7612–7616 (1995).
3. Xu, B. *et al.* Strong association of *de novo* copy number mutations with sporadic schizophrenia. *Nat. Genet.* **40**, 880–885 (2008).
4. Malhotra, D. *et al.* High frequencies of *de novo* CNVs in bipolar disorder and schizophrenia. *Neuron* **72**, 951–963 (2011).
5. Kirov, G. *et al. De novo* CNV analysis implicates specific abnormalities of postsynaptic signalling complexes in the pathogenesis of schizophrenia. *Mol. Psychiatry* **17**, 142–153 (2012).
6. Girard, S.L. *et al.* Increased exonic *de novo* mutation rate in individuals with schizophrenia. *Nat. Genet.* **43**, 860–863 (2011).
7. Xu, B. *et al.* Exome sequencing supports a *de novo* mutational paradigm for schizophrenia. *Nat. Genet.* **43**, 864–868 (2011).
8. Xu, B. *et al.* Elucidating the genetic architecture of familial schizophrenia using rare copy number variant and linkage scans. *Proc. Natl. Acad. Sci. USA* **106**, 16746–16751 (2009).
9. Kryukov, G.V., Pennacchio, L.A. & Sunyaev, S.R. Most rare missense alleles are deleterious in humans: implications for complex disease and association studies. *Am. J. Hum. Genet.* **80**, 727–739 (2007).
10. Neale, B.M. *et al.* Patterns and rates of exonic *de novo* mutations in autism spectrum disorders. *Nature* **485**, 242–245 (2012).
11. Sanders, S.J. *et al. De novo* mutations revealed by whole-exome sequencing are strongly associated with autism. *Nature* **485**, 237–241 (2012).
12. O'Roak, B.J. *et al.* Sporadic autism exomes reveal a highly interconnected protein network of *de novo* mutations. *Nature* **485**, 246–250 (2012).
13. Iossifov, I. *et al. De novo* gene disruptions in children on the autistic spectrum. *Neuron* **74**, 285–299 (2012).
14. Kang, H.J. *et al.* Spatio-temporal transcriptome of the human brain. *Nature* **478**, 483–489 (2011).
15. Colantuoni, C. *et al.* Temporal dynamics and genetic control of transcription in the human prefrontal cortex. *Nature* **478**, 519–523 (2011).
16. Barch, D.M. & Ceaser, A. Cognition in schizophrenia: core psychological and neural mechanisms. *Trends Cogn. Sci.* **16**, 27–34 (2012).
17. Sobin, C., Roos, J.L., Pretorius, H., Lundy, L.S. & Karayiorgou, M. A comparison study of early non-psychotic deviant behavior in Afrikaner and US patients with schizophrenia or schizoaffective disorder. *Psychiatry Res.* **117**, 113–125 (2003).
18. Christopherson, K.S. *et al.* Thrombospondins are astrocyte-secreted proteins that promote CNS synaptogenesis. *Cell* **120**, 421–433 (2005).
19. Georges-Labouesse, E., Mark, M., Messaddeq, N. & Gansmuller, A. Essential role of α6 integrins in cortical and retinal lamination. *Curr. Biol.* **8**, 983–986 (1998).
20. Jones, K.J. *et al.* The expanding phenotype of laminin α2 chain (merosin) abnormalities: case series and review. *J. Med. Genet.* **38**, 649–657 (2001).
21. Van Kuilenburg, A.B. *et al.* Genotype and phenotype in patients with dihydropyrimidine dehydrogenase deficiency. *Hum. Genet.* **104**, 1–9 (1999).
22. Tiedje, K.E., Stevens, K., Barnes, S. & Weaver, D.F. β-alanine as a small molecule neurotransmitter. *Neurochem. Int.* **57**, 177–188 (2010).
23. Ben-David, E. *et al.* Identification of a functional rare variant in autism using genome-wide screen for monoallelic expression. *Hum. Mol. Genet.* **20**, 3632–3641 (2011).
24. Carter, M.T. *et al.* Hemizygous deletions on chromosome 1p21.3 involving the *DPYD* gene in individuals with autism spectrum disorder. *Clin. Genet.* **80**, 435–443 (2011).
25. Willemsen, M.H. *et al.* Chromosome 1p21.3 microdeletions comprising *DPYD* and *MIR137* are associated with intellectual disability. *J. Med. Genet.* **48**, 810–818 (2011).
26. Ripke, S. *et al.* Genome-wide association study identifies five new schizophrenia loci. *Nat. Genet.* **43**, 969–976 (2011).
27. Arguello, P.A. & Gogos, J.A. Genetic and cognitive windows into circuit mechanisms of psychiatric disease. *Trends Neurosci.* **35**, 3–13 (2012).
28. McGrath, J.J. & Susser, E.S. New directions in the epidemiology of schizophrenia. *Med. J. Aust.* **190**, S7–S9 (2009).
29. Gilman, S.R. *et al.* Diverse types of genetic variation converge on functional gene networks involved in schizophrenia. *Nat. Neurosci.* (in the press).
30. Stark, K.L. *et al.* Altered brain microRNA biogenesis contributes to phenotypic deficits in a 22q11-deletion mouse model. *Nat. Genet.* **40**, 751–760 (2008).
31. Karayiorgou, M., Flint, J., Gogos, J.A. & Malenka, R.C. The best of times, the worst of times for psychiatric disease. *Nat. Neurosci.* **15**, 811–812 (2012).

© 2012 Nature America, Inc. All rights reserved.

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

ONLINE METHODS

Cohorts. The samples analyzed here comprise trios collected from two distinct populations: the Afrikaner population from South Africa (European, mostly Dutch descent) (146 trios with schizophrenia) and the US population (Northern European descent) (85 trios with schizophrenia). Of the 146 Afrikaner probands, 122 (83.6%) had a diagnosis of schizophrenia, and 24 (16.4%) were diagnosed with SCZAFF disorder. Of the 85 US probands, 46 (54.1%) had a diagnosis of schizophrenia, and 39 (45.9%) were diagnosed with SCZAFF disorder. The control cohort consisted of 34 trios with established Afrikaner heritage. Control families included unaffected subjects screened against presence and history of treatment for any psychiatric condition, as well as history of mental illness in first- or second-degree relatives. Both affected and control trios were recruited and characterized in the context of our ongoing large-scale genetic studies of schizophrenia and have been described previously[3,7,8]. Because *de novo* mutations are more likely to account for sporadic forms of the disease, we took great care to determine reliably and in depth the family history status and generate cohorts enriched in sporadic cases (**Supplementary Note**). However, the presence or absence of a family history was not used as a screening criterion.

In the Afrikaner cohort, it was possible to determine absence of disease in first- or second-degree relatives because of the cohesive family structures, the large catchment area and the fact that long-term care was provided by the local recruiting hospital, thereby affording detailed hospital records over several generations[3,8]. In the geographically fragmented and ancestrally diverse US cohort, we were able to determine absence of disease only in first-degree relatives (**Supplementary Note**). Additional cohort characteristics are given in the **Supplementary Note**. Informed consent was obtained from all participants, and the Institutional Review Committees of Columbia University and the University of Pretoria approved all procedures. Paternity and maternity were confirmed before sequencing via the Affymetrix Genome-Wide Human SNP Array 5.0 as well as via a panel of microsatellite markers. DNA for all study subjects was extracted from whole blood, and analysis was performed blind to affected status while maintaining knowledge of the parent-child relationships.

Exome library construction. For exome capture and sequencing, genomic DNA (~3 µg) was sheared to 200–300 bp in size using a Covaris Acoustic Adaptor. Fragments were then end repaired, polyA tailed and ligated to sequencing adaptor oligonucleotides using reagents from New England BioLabs. Libraries were barcoded using the Illumina index read strategy, which uses 6-base sequences within the adaptor that are sequenced separately from the genomic DNA insert. Ligated products were size selected during purification steps. The DNA library was subsequently enriched for sequences with 5′ and 3′ adaptors by PCR amplification using primers complementary to the adaptor sequences (ligation-mediated PCR, LM-PCR). Exonic DNA was captured using two hybridization systems: Agilent SureSelect v2 (n = 85 trios) and NimbleGen SeqCap EZ v2 (n = 180 trios). After capture, another round of LM-PCR was performed to generate the final libraries. Each library was quantified by fluorescent methods (PicoGreen, Invitrogen), and fragment size was measured with the Agilent Bioanalyzer. Finally, the molar concentration of each library was measured using the size information from the Agilent Bioanalyzer and DNA quantitation information from an RT-PCR assay (Kapa Biosystems; performed according to the manufacturer's protocol). Each library was normalized to a 10 nM concentration and sequenced using an Illumina HiSeq 2000.

Exome data analysis for *de novo* SNVs and indels. The exome data analysis pipeline has been described previously[7]. Briefly, raw sequencing data were mapped to the human reference genome (hg19) using the Burrows-Wheeler Aligner (BWA v0.5.81536). The Genome Analysis Toolkit (GATK, version 5091) was used to remove duplicates, perform local realignment and map quality score recalibration to produce a cleaned BAM file and then make genotype calls for all trios jointly. The resulting variant call format (VCF, version 4.0) files were annotated using the GenomicAnnotator module in GATK to identify and label the called variants that were within the targeted coding regions and overlap with known and likely benign SNPs reported in dbSNP v132 (see URLs). The filtered genotype calls were further validated using the mpileup module in SAMtools (see URLs) as described previously[7]. Indel calls were made by Dindel software using one cleaned BAM file per run. The resulting VCF files were further revalidated using the same SAMtools procedure as for the point mutations. To determine potential mutations at splice donor or acceptor sites, GATK variant calls were made in a batch fashion (90 samples per batch) that covered each target coding region and the 50-bp flanking segments on each side of it. The variants in the resulting VCF files were annotated according to refGene-big-table-hg19.txt (see URLs). A variant was annotated as a 'canonical splice-site mutation' if it disrupted the largely invariable core canonical 2-bp acceptor (AG) or donor (GU) sites. *De novo* variants within 10 bp of the exon-intron boundary, included in the consensus sequence flanking core canonical splice sites and therefore likely to modulate splicing efficiency, were annotated as 'consensus splice-site mutations'. Candidate *de novo* variants were tested using standard Sanger sequencing on an ABI 3730xl DNA Analyzer (Applied Biosystems) to validate the presence of each mutation in the proband and its absence in the parental genomes, using custom primers that were designed using Primer3 software on the basis of the ~500 bp of sequence flanking each variant. The total number of *de novo* SNVs found and validated in a given cohort was divided by the total number of bases analyzed to calculate the per-base rate of point mutation in the captured coding sequence.

Variant detection pipeline and quality control. Because the entire capture and sequencing procedure was conducted without knowledge of the affected status in all three cohorts, we expected no bias between cohorts. To further demonstrate that variant detection and quality control were consistent across all samples and all experimental conditions, we compared the percentage of average reads at 1×, 8×, 20× and 30× coverage under each condition (**Supplementary Fig. 1**). There were no differences for any of the parameters analyzed.

Statistics. The two-sided exact binomial test was conducted using R. Fisher's exact test or the χ-squared test with Yates' correction was used for the analysis of contingency tables, depending on the sample sizes, using R.

Annotation of the functional impact of the *de novo* mutations. The functional impact of the *de novo* mutations was annotated from several different resources. The PolyPhen-2 online batch query server (see URLs)[32] was used with full annotation settings to determine the nonsynonymous or synonymous nature of the mutations and predict their functional impact by further classifying them as non-tolerated (damaging) or benign at a given site. The Grantham score for each coding variant was determined using the Grantham matrix table[33]. The phyloP score for each coding variant was extracted from the phyloP46wayAll table in the UCSC Table Browser (see URLs). For splice-site variants, we considered mutations directly disrupting canonical splice sites as severe disruptive events without further analysis. For mutations in consensus splice sites, we used a mutation analysis module in the Human Splice Finder (HSF) program (Version 2.4.1; see URLs)[34] to predict their functional impact. Briefly, 100 nucleotides of sequence surrounding the exon-intron boundary were extracted from the UCSC browser and the wild-type and mutated sequences were imported into the HSF mutation analysis module to detect potential disruption of splicing signals. The HSF-derived results for the identified consensus splice-site mutations are shown in **Supplementary Tables 11** and **12**.

Gene set enrichment and protein-protein interaction network analyses. The DAVID Functional Annotation Chart[35] (see URLs) was used to determine whether a given gene set with *de novo* mutations was enriched in particular gene ontology (GO) terms or functional keywords defined in Swiss-Prot and the Protein Information Resource. Genes with mutations in schizophrenia were mapped in the database, and functional annotation chart analysis was conducted with the default settings. We used the Disease Association Protein-protein Link Evaluator (DAPPLE; see URLs)[36] to determine whether there was higher protein-protein interaction than expected by chance among the genes altered by likely functional *de novo* variants. A list of all genes with likely functional mutations was submitted to the DAPPLE server with default settings.

Temporal expression profile analysis of the genes carrying *de novo* mutations. To investigate developmental expression of genes mutated in

© 2012 Nature America, Inc. All rights reserved.

doi:10.1038/ng.2446

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

schizophrenia, we took advantage of the Human Brain Transcriptome (HBT) database (see URLs), a compendium of exon-level expression profiles across developmental stages from embryonic to late adulthood[14]. Genes harboring *de novo* events were grouped into three classes (with prenatal brain-biased, postnatal brain-biased and unbiased expression) according to their temporal expression trajectory in reference to a global expression turning point occurring between the mid-late and late fetal stages[14]. For each class, the ratio of nonsynonymous or likely functional variants to neutral variants was calculated.

32. Adzhubel, I.A. *et al.* A method and server for pred1ct1ngdamagmg m1ssense mutations. *Nat. Methods*7, 248-249 (2010).
33. Grantham, R. Amino acid difference formula to help explain protein evolution. *Science* 185, 862–864 (1974).
34. Desmet, F.O. *et al.* Human Spl1c1ng Finder, an online b1o1nformat1cs tool to predict spl1c1ngSignals. *Nucleic Acids Res.* 37, e67 (2009)
35 Huang DW, Sherman, B.T. & Lemp1cki, R.A. Systematic and 1ntegrat1ve analySIS of large gene lists us1ng DAVID b101nformat1cs resources. *Nat. Protoc.* 4, 44-57 (2009).
36. Ross1n, E.J. *et at.* Prote1ns encoded 1n genom1c reglons associated w1th Immune-mediated disease physically Interact and suggest underlying biology. *PLoS Genet.* 7, e1001273 (2011).

doi:10.1038/ng.2446

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

ELSEVIER

Available online at www.sciencedirect.com
SCIENCE DIRECT™

Psychiatry Research 117 (2003) 113–125

PSYCHIATRY RESEARCH

www.elsevier.com/locate/psychres

A comparison study of early non-psychotic deviant behavior in Afrikaner and US patients with schizophrenia or schizoaffective disorder

Christina Sobin[a,*], J. Louw Roos[b], Herman Pretorius[b], Laura S. Lundy[a], Maria Karayiorgou[a]

[a]*Laboratory of Human Neurogenetics, Rockefeller University, 1230 York Avenue, Box 313, New York, NY 10021, USA*
[b]*University of Pretoria and Weskoppies Hospital, Pretoria, South Africa*

Received 20 June 2002; received in revised form 21 November 2002; accepted 26 November 2002

Abstract

In a previous study early non-psychotic deviant behaviors in US adult schizophrenic patients recruited for a large-scale genetic study were examined (*Psychiatry Research*, 101, 101). Early deviance characterized a distinct subgroup of patients at rates that were consistent with earlier reports. In addition, specific early non-psychotic deviant behaviors were meaningfully associated with later disease outcomes. In the present study, we examined the demographic, syndrome course, symptom and early deviant behavior history of 109 Afrikaner probands who met criteria for DSM schizophrenia or schizoaffective disorder, and compared them to 109 age- and gender-matched US probands. Consistent with past findings, 68% of Afrikaner probands, as compared to 67% of age- and gender-matched US probands, reported one or more forms of early non-psychotic deviance, including poor socialization, extreme fearsy chronic sadness, andyor attentionylearning impairment. The remaining 32 and 33% of probands, respectively, were without behavioral deviance until the onset of schizophrenia or schizoaffective disorder. The frequency and distribution of individual deviant behaviors were strikingly consistent between the samples. However, logistic regression analyses revealed different patterns of associations between the early deviant behaviors manifested and disease outcome. Afrikaner participants with early fearsychronic sadness were 3 times more likely to attempt suicide, while among US participants, this form of early deviance conferred 3.5 times more risk for later schizoaffective disorder, and 3 times greater likelihood of later sensory (tactile andyor olfactory) hallucinations. Afrikaner participants with attentiony learning impairment were 2.5 times more likely to experience later auditory hallucinations, while US participants with these early difficulties were 3 times more likely to experience thought disorder. We concluded that early non-psychotic childhood deviance in this independently collected Afrikaner population distinguished a distinct subtype of patients and that the forms of early deviance manifested were meaningfully linked to later disease outcome. Possible reasons for the association pattern differences in these two populations are considered.
© 2002 Elsevier Science Ireland Ltd. All rights reserved.

Keywords: Schizophrenia; Schizoaffective disorder; Early onset

*Corresponding author. Tel.: q 1-212-327-7807; fax: q 1-212-327-7329.
E-mail address: sobinc@rockvax.rockefeller.edu (C. Sobin).

0165-1781/03/$ - see front matter © 2002 Elsevier Science Ireland Ltd. All rights reserved.
PII: S0165-1781Ž02.00321-9

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

1. Introduction

A convergence of evidence over the past decade from the medical (Geddes and Lawrie, 1995; Verdoux et al., 1997, reviews), anatomical (Green et al., 1994; O'Callaghan et al., 1991; Guy et al., 1983; Cannon et al. 1994) and neuroanatomical (e.g. Lewis, 1997; Waddington and Buckley, 1996) literature has suggested that schizophrenia is a neurodevelopmental disorder with roots in early childhood (e.g. Lieberman, 1999). Moreover, early deviant behavior patterns in children who go on to develop schizophrenia have been consistently reported in high-risk, prospective and retrospective behavioral studies of schizophrenia (Cannon and Jones, 1996, review). These aberrant behaviors have been linked to early neurodevelopmental anomalies (Cannon et al., 1997; Foerster et al., 1991; Guy et al., 1983) and likely reflect abnormalities in specific neural pathways. For example, prospective longitudinal studies of large cohorts have suggested that during infancy motor development is delayed and speech problems are more frequent among children who later develop schizophrenia (e.g. Jones et al., 1994). During the preschool and early school years, studies have reported preference for solitary play (Jones et al., 1994) and behavioral maladjustment in the classroom (Done et al., 1994), and cognitive impairment in verbal, non-verbal and math skills were reported at ages 8, 11 and 15 (Jones et al., 1994). Retrospective studies have provided findings consistent with these. An analysis of taped recordings of children revealed that abnormal social reactions, odd movement and anomalous posture reliably indicated to clinicians which children would later become schizophrenic (Walker and Lewine, 1990). Social and academic impairments during the first school-age years were more frequent among 45 future schizophrenic patients as compared to a psychiatric control sample (Foerster et al., 1991). In a clinic-referred population of 61 patients with schizophrenia onset between the ages of 7 and 17, social impairment, motor deficits and language learning disabilities were increased as compared to findings in matched non-psychotic psychiatric controls.

However, early behavioral deviance appears to occur in only a subgroup of patients (Sobin et al., 2001; Done et al., 1994; Torrey et al., 1994; Offord and Cross, 1969)—a distinction that has often gone unrecognized among those interested in describing a neurodevelopmental model of schizophrenia. That early behavioral deviance, and perhaps also the various neurodevelopmental anomalies that underlie these behaviors, may in fact be both a subtype and disease-onset marker, has important implications. Understanding the patterns and distribution of early deviant behaviors that in some patients long precede DSM schizophrenia may be fundamental to accurately characterizing its broad phenotypic heterogeneity. Moreover, for those patients whose histories include early deviant behavior, the behaviors themselves may provide important clues about brain development and, by extension, possible neurobiologic and neurogenetic underpinnings of schizophrenia.

In a previous study we described early non-psychotic deviance in a population of US participants who met lifetime criteria for schizophrenia or schizoaffective disorder. We found 67% of 205 patients reported marked behavioral abnormalities before age 10 and that the specific early behavioral abnormalities predicted later aspects of the disease course that were clinically consistent with the early behaviors, and also predicted family history of schizophrenia (Sobin et al., 2001). However, by necessity, the data were based on retrospective report of patients. School records were generally unavailable. (Even when available, school records only reflected the readily observable classroom behaviors spontaneously provided by teachers or aides.) Obtaining reports from parents andyor siblings was also problematic. Families of schizophrenic patients are often fragmented. At the same time, the perceptions of involved families are understandably influenced by lingering emotional reactions that can include guilt, disappointment, anger or fear. Parents are especially unwilling to remember or reveal painful details of a sick child's early experiences. Thus, while the approach is not without limitations, it seemed useful to further explore patient-reported non-psychotic early deviant childhood behaviors.

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

In this study, we compared childhood information obtained from 109 Afrikaner patients with schizophrenia or schizoaffective disorder to an age- and gender-matched subsample of patients from our US sample. We reasoned that if the early childhood deviance data were largely due to reporting error and reflected patients' fluctuating internal states, severity of illness or cognitive decline, the specific results would not replicate. Additionally, in our previous study, and consistent with other investigators' findings, gender was associated with specific symptom manifestation, and age at time of interview was associated with the report of early learning disabilitiesyattentional impairment (likely attributable to increased societal awareness of these difficulties). To control for these effects, an age- and gender-matched subsample of our US patients was selected for comparison to the Afrikaner population. We predicted that the gender differences found in the original US study would replicate in the Afrikaner population. Thus, specifically, we predicted that females would have a higher rate of marriage, and higher rates of schizoaffective disorder, and auditory, visual and sensory (tactiley olfactory) hallucinations. We predicted that the previous rates of early non-psychotic deviant childhood behaviors would replicate in the independently collected Afrikaner population. We also predicted that the associations previously reported between early deviant behaviors and later illness features would replicate. Specifically, we predicted that early poor socialization would predict earlier age of syndrome onset, early extreme fearsychronic sadness would predict later schizoaffective disorder, and attentional impairmentylearning disability would predict increased likelihood of schizophrenia among probands' family members.

2. Methods

This study included participants from independently collected samples in South Africa and the US. All participants provided written informed consent before beginning the initial interview. Afrikaner participants were evaluated as part of a collaborative genetic project that began in November 1997 between the Laboratory of Human Neurogenetics at Rockefeller University in New York City and the Department of Psychiatry at the University of Pretoria. A total of 168 probands and their family members were recruited and diagnosed, and enrollment for the genetic study is ongoing. The subsample used in this study included 109 participants (78 males, 31 females) who completed a diagnostic interview, who met DSM-IV criteria for schizophrenia or schizoaffective disorder and were without co-morbid substance dependence or other confounding medical diagnosis, and who provided information regarding early deviant childhood behavior occurring before age 10. Fifty-nine unrelated proband participants did not meet these criteria. Participants were recruited from in patient, outpatient and follow-up clinics at a large psychiatric hospital in Pretoria, South Africa, or from its greater catchment area. Approximately 10% of participants were referred by local private psychiatrists or were self-referred in response to radio recruitment and notices by local support group leaders. The remaining participants were offered enrollment in the study by their treating psychiatrists.

A full description of the US sample was given in Sobin et al. (2001). Approximately 85% of the sample were recruited through advertisements in advocate newspapers, and the remaining 15% were recruited from area day treatment clinics. Potential differences in selection biases will be considered in Section 4.1. The same enrollment, assessment and interviewer-training procedures were used in both sites. Information on family history of mental disorders was obtained from parents, subjects andy or other family members just following enrollment and before the diagnostic interview. In the case of the Afrikaner participants, when previous admissions of family members to the psychiatric hospital where the study was based were reported, clinical records were obtained and diagnoses recorded. For approximately 75% of the US participants, report of family history of mental illness was obtained from participants' parents. The remaining information was obtained from the participants themselves.

All participants were interviewed using the Diagnostic Interview for Genetic Studies (DIGS) by a specially trained clinician with a minimum of 3 years of clinical experience in diagnosis

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Table 1
Demographic and clinical characteristics of 109 Afrikaners and 109 age- and gender-matched US participants with a lifetime history of DSM schizophrenia or schizoaffective disorder

	Afrikaner population (*n*=109)		US population (*n*=109)		Afrikaner males (*n*=78)		US males (*n*=78)		Afrikaner females (*n*=31)		US females (*n*=31)	
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Demographics												
Interview age	30.94	8.49	30.74	8.09	30.01	7.81	30.03	7.67	33.26	9.78	32.55	8.94
No. years education	**11.90**	**2.10**	**13.46**	**1.99****	**11.68**	**1.98**	**13.50**	**1.95****	12.45	2.29	13.36	2.14
	%	(*N*)	%	(*N*)	%	(*N*)	%	(*N*)	%	(*N*)	%	(*N*)
Left-handed	13.76	(15)	12.84	(14)	14.10	(11)	12.82	(10)	12.90	(4)	12.90	(4)
Single (never married)	79.82	(87)	85.32	(93)	83.33	(65)	*92.31*	(*72*)	70.97	(22)	*67.74*	(*21*)††
Fam hxyschiz	31.19	(34)	34.91	(37)	29.49	(23)	33.33	(25)	35.48	(11)	38.71	(12)
Fam hxyother psych dx	68.81	(75)	61.32	(65)	67.95	(53)	62.69	(42)	70.97	(22)	58.06	(18)
Syndrome features	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Age DSM schiz onset	**22.55**	**5.47**	**19.84**	**4.37****	**22.77**	**5.60**	**19.54**	**3.88****	22.00	5.16	20.58	5.42
	%	(*N*)	%	(*N*)	%	(*N*)	%	(*N*)	%	(*N*)	%	(*N*)
Schizoaffective	**24.77**	**(27)**	**48.62**	**(53)****	**24.36**	**(19)**	**46.15**	**(36)****	**25.81**	**(8)**	**54.84**	**(17)***
Attempted suicide	**27.52**	**(30)**	**50.46**	**(55)****	**25.64**	**(20)**	**48.72**	**(38)****	32.26	(10)	54.84	(17)
Symptom history	%	(*N*)	%	(*N*)	%	(*N*)	%	(*N*)	%	(*N*)	%	(*N*)
Delusions	98.17	(107)	95.28	(101)	97.44	(76)	92.31	(72)	100	(31)	93.55	(29)
Auditory hallucinations	70.64	(77)	64.76	(68)	70.51	(55)	57.69	(45)	70.97	(22)	74.19	(23)
Visual hallucinations	36.70	(40)	43.40	(46)	37.18	(29)	38.46	(30)	35.48	(11)	51.61	(16)
Sensory hallucinations	22.94	(25)	33.94	(37)	23.08	(18)	*26.92*	(*21*)	**22.58**	**(7)**	***51.61***	***(16)***††
Disordered behavior	69.72	(76)	65.14	(71)	70.51	(55)	65.38	(51)	67.74	(21)	64.52	(20)
Thought disorder	**77.06**	**(84)**	**50.46**	**(55)****	**78.20**	**(61)**	**48.72**	**(38)****	74.19	(23)	54.84	(17)
Alogiayflat affect	70.64	(77)	75.00	(81)	71.80	(56)	*80.77*	(*63*)	67.74	(21)	*58.06*	(*18*)†
Deviant childhood behav —age 10	%	(*N*)	%	(*N*)	%	(*N*)	%	(*N*)	%	(*N*)	%	(*N*)
Social withdrawaly aggressionyodd behav	30.28	(33)	40.37	(44)	28.20	(22)	42.31	(33)	35.48	(11)	35.48	(11)
Extreme fearsy chronic sadness	**34.86**	**(38)**	**20.18**	**(22)***	32.05	(25)	17.95	(14)	41.94	(13)	25.81	(8)
Attentional impairmenty learning disability	39.45	(43)	37.62	(41)	41.03	(32)	39.74	(31)	35.48	(11)	32.26	(10)

Comparisons between Afrikaners and US participants are indicated in bold and: *Fisher's exact test (two-tailed) $P \leq 0.05$ or *t*-test $P \leq 0.05$; **Fisher's exact test (two-tailed) $P \leq 0.01$ or *t*-test $P \leq 0.01$. Comparisons between Afrikaner males and females or US males and females are indicated in italics and: † Fisher's exact test (two-tailed) $P \leq 0.05$ or *t*-test $P \leq 0.05$; †† Fisher's exact test (two-tailed) $P \leq 0.01$ or *t*-test $P \leq 0.01$.

(Nurnberger et al., 1994). In multi-site test–retest analyses, the DIGS was shown to be reliable for the major DSM diagnostic categories. In ongoing reliability studies between the two Afrikaner clinicians, the two US clinicians and the US investigator who trained all the clinicians, diagnostic agreement ›90% for DSM-IV diagnoses was obtained. In cases where patients were unable to provide symptom or syndrome information, with the patients' permission, information was obtained from a family informant, typically the mother.

The clinical variables analyzed (Table 1) are largely self-evident. The variable 'family history of schizophrenia' indicated whether a first, second, andyor third degree relative of a subject had been diagnosed with andyor hospitalized for schizophrenia or schizoaffective disorder. The variables 'family history of other psychiatric disorder' indicated

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

whether a first, second, andyor third degree relative of a participant had been diagnosed with, hospitalized for, andyor treated for any other major psychiatric disorder excluding schizophrenia or schizoaffective disorder.

History of early childhood deviance was obtained using a semi-structured questionnaire probing early deviant non-psychotic childhood behaviours (see Sobin et al., 2001, for a complete description). For approximately half of the subjects, the inquiry was conducted separately from the DIGS and by different interviewers who were blind to the results of the DIGS. For the other half, inquiry regarding childhood behaviors was completed before the diagnostic interview to prevent interviewer bias that might occur as a result of information obtained during the DIGS. The early childhood behavior questionnaire probed seven areas of possible deviance including social dysfunction (avoidance of other children, inability to have friends, isolated play), extreme odd behaviors (unprovoked screaming fits, disorganized or irrational behavior, inappropriate affect), unprovoked aggression, extreme anxiety, chronic sadness, attentional impairment and learning disabilities. Yesyno responses were recorded. In the case of a 'yes' response, examples were obtained and the interviewer determined whether the behavior reported was chronic or transient, whether there was a likely environmental trigger for the reported behavior (e.g. marital discord, parental abuse, death of a loved one), and whether the behavior began before age 10. Age 10 was used as a cut-off to exclude behaviors that might be attributable to pre-pubertal hormonal changes. Behaviors were considered present only if the behavior was both chronic and severe, if no environmental precipitant was present and if the behavior began before age 10. Learning disabilities andy or attentional impairment were considered present only if the child received a formal diagnosis, was provided with specific remediation, or if either of these were reported by teachers or school personnel or noted in records or school reports.

Data were entered and analyzed using Statview version 5.0 for PC and SAS version 6.1. *T*-tests for continuous variables and Fisher's exact tests for dichotomous variables were used to first determine possible associations between gender and the demographic, syndrome feature, and symptom history variables and to examine possible associations between early deviant variables. Logistic regression analyses were then used to determine the level of risk conferred for later disease outcomes by specific early deviant behaviors found to significantly differ in *t*-test and x^2 analyses.

3. Results

The sample included a total of 218 patients with a lifetime history of either schizophrenia or schizoaffective disorder and for whom information regarding early deviant childhood behavior had been obtained. One hundred and nine participants were of Afrikaner descent, were unrelated at least up to the fifth degree, and included 78 males and 31 females. The remaining 109 were also unrelated and were age- and gender-matched to the Afrikaner sample, selected from our larger US database of schizophrenic patients. Preliminary to our main analyses, we compared the demographic, syndrome and symptom features of 109 Afrikaner participants included in these analyses with the remaining 59 probands for whom childhood information was not available to ensure that this subsample was in fact representative of the group as a whole.

Afrikaner participants with and without early childhood information did not differ with regard to current marital status, family history of schizophrenia, family history of psychiatric disorder not including schizophrenia, age of DSM schizophreniayschizoaffective disorder onset, frequency of schizoaffective disorder, frequency of attempted suicide, or frequency of delusions, visual hallucinations, sensory hallucinations, thought disorder or alogiayflat affect. The subsamples differed with regard to three variables. The subsample with completed early deviant childhood behavior questionnaires had a higher frequency of left-handed participants (15.46% vs. 2.38%, $x^2 \leq 6.26$, Fisher's exact $P \leq 0.039$); a lower frequency of the auditory hallucinations (71.96 vs. 87.5%, $x^2 \leq 4.49$, Fisher's exact $P \leq 0.040$); and a lower frequency of the symptom disordered behavior (70.27% vs. 85.42%, $x^2 \leq 4.01$, Fisher's exact $P \leq 0.048$). We did not expect that the difference with regard to handed-

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

B1: 169

ness would influence these findings. Differences with regard to the frequency of two isolated symptoms in the absence of differences on all of the remaining course and symptom variables suggested to us that the selected group of Afrikaner participants was largely representative of the entire sample recruited to date.

We compared the clinical profiles of the Afrikaner and US populations to determine their clinical similarity (Table 1). With regard to demographics, there were no differences between the samples in handedness, marital status, and family history of schizophrenia or family history of other psychiatric illness. The samples did differ

achieved.

With regard to syndromal features, the Afrikaner population reported a later onset of DSM schizophrenia and a substantially lower rate of schizoaffective disorder and attempted suicide than in the US population. In terms of symptom history, the samples differed only with regard to lifetime history of thought disorder, which Afrikaner participants experienced at a substantially higher frequency than that reported by the US participants. Thus, the schizophrenia syndrome was found to be essentially equivalent in these two populations with greater heterogeneity indicated in the US sample by the substantially higher frequency of schizoaffective disorder.

In order to determine whether the same gender differences found among the US population also existed among the Afrikaner population, males and females were compared. If such differences were found, they would need to be statistically controlled in the subsequent analyses. Contrary to our predictions there were no differences between male and female Afrikaner patients on any of the demographic, syndromal feature or symptom variables assessed. (Among the US population, more males than females had never married, substantially more females than males reported a history of sensory hallucinations, and more males than females reported experiencing alogiayflat affect.)

Preliminary to the analysis of early childhood deviant behaviors and their association with later syndromal features, we compared the samples with regard to syndromal features alone. The difference

Table 2
Distribution and frequency of individual early deviant behaviors

	Afrikaner population (n=109)		US population (n=109)	
	Freq.	%	Freq.	%
No. of deviant behaviors				
None	34	31.19	36	33.03
Any	75	68.81	73	66.97
1	39	52.00	36	43.32
2	19	25.33	25	34.25
3	11	14.67	12	16.44
4	1	1.33	0	0.00
5	5	6.67	0	0.00
Social withdrawal	16	14.68	26	23.85
Extreme aggression	13	11.93	13	11.93
Odd behavior	12	11.01	13	11.93
Extreme fears	**32**	**29.36**	**12**	**11.01****
Chronic sadness	12	11.01	11	10.09
Learning disability	16	14.68	23	21.10
Attentional impairment	**38**	**34.86**	**24**	**22.02***

* Fisher's exact test (two-tailed) P=0.05.
** Fisher's exact test (two-tailed) P=0.01.

between samples in the number of years of education obtained was attributable to the lower educational achievement of male rather than female Afrikaner participants. The differences between the Afrikaner and US males and not the females again accounted for the difference between samples with regard to age of syndrome onset. The frequency of reported schizoaffective illness was evident in both males and females, while only males differed significantly in the rates of attempted suicide, with substantially more US males than Afrikaner males attempting suicide. The frequency of sensory hallucinations reported by the US female participants was markedly higher than that reported by the Afrikaner females, but this difference was not significant when the populations were compared without regard to gender.

Table 2 compares the distribution of individual early deviant childhood behaviors in the two populations. Seventy-five (68.8%) Afrikaner participants reported experiencing one or more non-psychotic deviant behaviors beginning before 10 years of age. These included social withdrawal (n=16, 14.68%), odd behavior (n=12, 11.0%),

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

extreme aggression (*n*=13, 11.9%), excessive fears (*n*=32, 29.4%), chronic sadness (*n*=12, 11.0%), attentional impairment (*n*=38, 34.9%), and learning disabilities (*n*=16, 14.7%). The populations were strikingly consistent in the number and type of early deviant behaviors endorsed. Our hypothesis regarding the similarity of these samples in the early deviant childhood behaviors reported was largely supported. The frequencies of any early deviant childhood behaviors and the distribution of the numbers of behaviors reported did not differ. An examination of the frequency of individual behaviors queried revealed that the Afrikaner participants reported a higher frequency of early extreme fears. In addition, the Afrikaner participants reported more attentional impairment.

In order to increase and balance cell sizes for statistical analyses, these seven behaviors were clustered into three logical categories. Variables reflecting socially abnormal behaviors, including social withdrawal, aggression against others, and socially odd behavior, were grouped (*n*=33, 30.3%); mood-based variables, including extreme anxiety and chronic sadness, were combined (*n*=38, 34.9%); and variables reflecting cognitive dysfunction, including attentional impairment and learning disability, were combined (*n*=43, 39.5%). There were no differences in the proportions of males and females reporting childhood behavior problems including poor socialization, extreme fearsychronic sadness, or attentional impairmentylearning disability (Table 1). The differences in the frequency of early fears and attentional impairment were not evident once the behaviors were clustered.

In order to assess possible associations between early deviant childhood behaviors and later syndromal characteristics single order analyses were completed (Tables 3–5) for all variables listed. Early social withdrawal, aggression andyor odd behavior were not related to any later syndrome outcomes in either population. Early extreme fearsychronic sadness in the Afrikaner population was associated with an increased frequency of attempted suicide and also with a lower frequency of later alogiayflat affect. In the US population this form of early deviant behavior was associated with a higher frequency of schizoaffective illness and an increased frequency of sensory hallucinations. Early attentionalylearning impairment in the Afrikaner population was associated with a lower level of academic achievement and an increased frequency of auditory hallucinations. Among the US patients, early attentional impairmentylearning disability was associated with lower academic achievement and an increased risk of later thought disorder. An unexpected finding was a lower frequency of family members with manifest schizophrenia among Afrikaner participants who reported early attentional andyor learning impairment. This will be discussed further below.

The results of these single-order association tests were used to determine which early deviant childhood behaviors might pose a significant risk for particular syndrome outcomes, and thus which variables to include in a logistic regression model. As suggested by the above analyses, early social withdrawal, aggression andyor odd behavior was not associated with an increased risk of any syndromal features in either subsample. Among the Afrikaner population, patients with extreme fearsy chronic sadness in early childhood were nearly 3 times more likely to attempt suicide (weight coefficient=1.00; S.E.=0.45; $x^2_{d.f.=1}$=3.32; *P*=0.03; odds ratio=2.72; 95% CI=1.12–6.60). Among the US population, patients with this type of early deviance were almost 3.5 times more likely to report lifetime symptoms consistent with a diagnosis of schizoaffective disorder (weight coefficient=1.21; S.E.=0.54; $x^2_{d.f.=1}$=3.99; *P*=0.02; odds ratio=3.36; 95% CI=1.18–9.58). In addition, these US patients were nearly 3 times more likely to experience sensory hallucinations (weight coefficient=1.00; S.E.=0.50; $x^2_{d.f.=1}$=3.99; *P*=0.05; odds ratio=2.73; 95% CI=1.02–7.33).

With regard to attentional impairmentylearning disabilities before age 10, Afrikaner patients with this early cognitive impairment were 2.5 times more likely to experience later auditory hallucinations (weight coefficient=0.895; S.E.=0.47; $x^2_{d.f.=1}$=3.62; *P*=0.05; odds ratio=2.72; 95% CI=0.96–6.16). US patients with early attentional impairmentylearning disabilities were nearly 3 times more likely to experience thought disorder (weight coefficient=1.01; S.E.=0.41; $x^2_{d.f.=1}$=

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

B1: 171

Table 3
Single-order associations of social withdrawalyaggressionyodd behavior and demographic, syndrome and symptom features

	Social withdrawalyaggressionyodd behavior							
	Afrikaner probands (*N*=109)				US probands (*N*=109)			
	No (*n*=76)		Yes (*n*=33)		No (*n*=65)		Yes (*n*=44)	
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Demographics								
Age at interview	31.50	8.58	29.64	8.28	31.34	8.20	29.86	7.94
No. of years of education	11.84	2.07	12.03	2.20	13.66	1.98	13.16	2.00
	%	(*N*)	%	(*N*)	%	(*N*)	%	(*N*)
% Female	26.32	(20)	33.33	(11)	30.77	(20)	25.00	(11)
Single (never married)	77.63	(59)	84.85	(28)	80.00	(52)	93.18	(41)
Left-handed	11.84	(9)	18.18	(6)	10.77	(7)	15.91	(7)
	(*n*=76)		(*n*=33)		(*n*=69)		(*n*=37)	
Family hx of schizophrenia	30.26	(23)	33.33	(11)	33.33	(21)	37.21	(16)
Family hx of other psychia dx	65.79	(50)	75.76	(25)	61.90	(39)	60.46	(26)
Syndrome features	(*n*=76)		(*n*=33)		(*n*=65)		(*n*=44)	
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Schizophrenia onset age	23.14	5.59	21.18	5.00	20.46	4.01	18.87	4.76
	%	(*N*)	%	(*N*)	%	(*N*)	%	(*N*)
Attempted suicide	26.32	(20)	30.30	(10)	55.38	(36)	43.18	(19)
Schizoaffective	25.00	(19)	24.24	(8)	47.69	(31)	50.00	(22)
Symptom history								
Delusions	97.37	(74)	100	(33)	**96.92**	**(63)**	**86.36**	**(38)***
Auditory hallucinations	68.42	(52)	75.76	(25)	60.00	(39)	65.91	(29)
Visual hallucinations	32.90	(25)	45.46	(15)	40.00	(26)	45.46	(20)
Sensory hallucinations	22.37	(17)	24.24	(8)	35.38	(23)	31.82	(14)
Disordered behavior	68.42	(52)	72.73	(24)	70.77	(46)	56.82	(25)
Thought disorder	77.63	(59)	75.76	(25)	44.62	(29)	59.09	(26)
Alogiayflat affect	67.10	(51)	78.79	(26)	73.85	(48)	75.00	(33)

*Fisher's exact test (2-tailed) $P \leq 0.05$ or *t*-test $P \leq 0.05$. **Fisher's exact test (2-tailed) $P \leq 0.01$ or *t*-test $P \leq 0.01$.

6.07; $P \leq 0.01$; odds ratio=2.76; 95% CI=1.23–6.17).

Similar to our previous findings, early deviant childhood behaviors occurred in approximately 70% of participants with schizophreniayschizoaffective disorder, suggesting that these early behaviors distinguish a meaningful subgroup of patients. The frequency distribution and rates of these early problems were remarkably consistent across the two samples. In addition, specific early deviant childhood behaviors were statistically associated with later clinically similar disease characteristics. However, the associations of these early deviant behaviors to later syndromal features was different in the two samples. The possible source of these differences are considered in Section 4.

4. Discussion

We attempted to replicate our previous findings with regard to early deviant non-psychotic behaviors in an independently collected sample of adult patients meeting lifetime diagnostic criteria for DSM schizophreniayschizoaffective disorder. The compared samples were matched by gender and age, and they were evaluated using identical instruments and methods. In order to determine the clinical comparability of the samples, we first compared their demographic, syndrome and symptom characteristics. The one demographic difference found was in the number of years of education obtained by Afrikaner males, who completed fewer years of education than the US

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Table 4
Single-order associations of extreme fearsychronic sadness and demographic, syndrome and symptom features

	Extreme fearsychronic sadness							
	Afrikaner probands (N=109)				US probands (N=109)			
	No (n=71)		Yes (n=38)		No (n=87)		Yes (n=22)	
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Demographics								
Age at interview	31.61	8.20	29.68	9.00	30.66	7.99	31.09	8.65
No. of years of education	12.10	2.37	11.53	1.45	13.48	2.14	13.36	1.26
	%	(N)	%	(N)	%	(N)	%	(N)
% Female	25.35	(18)	34.21	(13)	26.44	(23)	36.36	(8)
Single (never married)	76.06	(54)	86.84	(33)	83.91	(73)	90.91	(20)
Left-handed	84.51	(60)	89.47	(34)	10.34	(9)	22.73	(5)
	(n=71)		(n=38)		(n=69)		(n=37)	
Family hx of schizophrenia	30.99	(22)	31.58	(12)	35.71	(30)	31.82	(7)
Family hx of other Psychia dx	70.42	(50)	65.79	(25)	57.14	(48)	77.27	(17)
Syndrome features	(n=71)		(n=38)		(n=87)		(n=22)	
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Schizophrenia onset age	23.01	5.41	21.68	5.54	19.92	4.43	19.46	4.22
	%	(N)	%	(N)	%	(N)	%	(N)
Attempted suicide	**19.72**	**(14)**	**42.10**	**(16)***	50.58	(44)	50.00	(11)
Schizoaffective	25.35	(18)	23.68	(9)	**42.53**	**(37)**	**72.73**	**(16)****
Symptom history								
Delusions	97.18	(69)	100	(38)	93.10	(81)	90.91	(20)
Auditory hallucinations	71.83	(51)	68.42	(26)	63.22	(55)	59.09	(13)
Visual hallucinations	39.48	(28)	31.58	(12)	37.93	(33)	59.09	(13)
Sensory hallucinations	21.13	(15)	26.32	(10)	**28.74**	**(25)**	**54.54**	**(12)***
Disordered behavior	69.01	(49)	71.05	(27)	68.97	(60)	50.00	(11)
Thought disorder	73.24	(52)	84.21	(32)	51.72	(45)	45.46	(10)
Alogiayflat affect	**77.46**	**(55)**	**57.90**	**(22)***	71.26	(62)	86.36	(19)

* Fisher's exact test (2-tailed) $P \leq 0.05$ or t-test $P \leq 0.05$.
** Fisher's exact test (2-tailed) $P \leq 0.01$ or t-test $P \leq 0.01$.

patients. While this suggested an earlier and more severe disturbance of cognitive functioning than among the US patients, it was not reflected in an earlier age of syndrome onset among Afrikaner males. Just the opposite was true; age of syndrome onset for Afrikaner participants was later than that of the US participants. This seemingly contradictory finding in a more genetically homogeneous sample may suggest that the recognition of symptoms and formal diagnosis of disorders is delayed in the Afrikaner culture. Moreover, compulsory military service in South Africa may pre-empt the completion of advanced education following high school.

The difference in the frequency of suicide attemptors was striking, with the US sample double that of the Afrikaner population. In fact many of the patients in this study had access to or were housed in a long-term care facility that provided a protected environment for those with unremitting illness. Such facilities no longer exist in the US, regardless of need. These supervised living circumstances for the chronically ill in this study perhaps provided greater protection from suicide attempts.

Our analyses of early non-psychotic deviant behaviors broadly replicated our previous findings. The overall frequencies of behaviors and their distributions were strikingly consistent in the two

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Table 5
Single-order associations of attentionalylearning impairment and demographic, syndrome and symptom features

	Attentional impairmentylearning disability							
	Afrikaner probands (*N*=109)				US probands (*N*=109)			
	No (*n*=66)		Yes (*n*=43)		No (*n*=68)		Yes (*n*=41)	
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Demographics								
Age at interview	30.88	8.81	31.02	8.09	31.07	8.25	30.20	7.89
No. of years of education	**12.23**	**2.22**	**11.40**	**1.83***	**13.75**	**1.96**	**12.98**	**1.97***
	%	(*N*)	%	(*N*)	%	(*N*)	%	(*N*)
% Female	30.30	(20)	25.58	(11)	30.88	(21)	24.39	(10)
Single (never married)	80.30	(53)	79.07	(34)	82.35	(56)	90.24	(37)
Left-handed	12.12 (*n*=66)	(8)	16.28 (*n*=43)	(7)	10.29 (*n*=69)	(7)	17.07 (*n*=37)	(7)
Family hx of schizophrenia	**39.39**	**(26)**	**18.60**	**(8)***	34.85	(23)	35.00	(14)
Family hx of other psychia dx	71.21	(47)	65.12	(28)	63.64	(42)	57.50	(23)
Syndrome features	(*n*=66)		(*n*=43)		(*n*=68)		(*n*=41)	
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Schizophrenia onset age	23.09	5.70	21.72	5.05	20.02	4.70	19.51	3.82
	%	(*N*)	%	(*N*)	%	(*N*)	%	(*N*)
Attempted suicide	25.76	(17)	30.23	(13)	54.41	(37)	43.90	(18)
Schizoaffective	22.73	(15)	27.91	(12)	50.00	(34)	46.34	(19)
Symptom history								
Delusions	98.48	(65)	97.67	(42)	95.59	(65)	87.80	(36)
Auditory hallucinations	**63.64**	**(42)**	**81.40**	**(35)***	57.35	(39)	70.73	(29)
Visual hallucinations	40.91	(27)	30.23	(13)	36.77	(25)	51.22	(21)
Sensory hallucinations	27.27	(18)	16.28	(7)	32.35	(22)	36.58	(15)
Disordered behavior	72.73	(48)	65.12	(28)	61.76	(42)	70.73	(29)
Thought disorder	78.79	(52)	74.42	(32)	**41.18**	**(28)**	**65.85**	**(27)****
Alogiayflat affect	68.18	(45)	74.42	(32)	76.47	(52)	70.73	(29)

* Fisher's exact test (2-tailed) $P \leq 0.05$ or *t*-test $P \leq 0.05$.
** Fisher's exact test (2-tailed) $P \leq 0.01$ or *t*-test $P \leq 0.01$.

samples, with 68.8% of Afrikaners compared to 67.0% of US participants reporting early deviance. With regard to the frequencies of individual early deviant behaviors reported, Afrikaners reported more early extreme fears and more attentional impairment, but the frequencies reported did not differ when individual symptoms were clustered. Moreover, early deviant behaviors in both populations were meaningfully linked to later disease outcome. However, the populations differed with regard to the specific associations found. Among Afrikaner participants, early extreme fearsychronic sadness were associated with an increased risk of later suicide attempts in the absence of schizoaffective disorder, suggesting greater severity of the schizophrenia symptoms manifested. Among US participants early, extreme fearsychronic sadness were associated with an increased risk of schizoaffective disorder, and not suicidality, as well as sensory hallucinations, suggesting greater clinical heterogeneity in the US sample. It is interesting to note that in the Afrikaner population, early extreme fearsychronic sadness were associated with a reduced frequency of later alogiayflat affect. We could speculate that in this population early emotional reactivity might represent a relatively stable trait that becomes beneficial in the course of later disorder. Also interesting to consider, though not possible to interpret on the basis of these data, is the finding among Afrikaner participants that early

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

attention and learning problems were associated with later risk of auditory hallucinations—a severe symptom of schizophrenia related to a specific perceptual system (McGuire et al., 1995). Among US participants, this form of early deviance was associated only with thought disorder—a more cognitively diffuse form of intellectual impairment. Whether this reflects a concordance between genetic endowment (more vs. less homogeneous) and clinical syndrome manifestation awaits further investigation.

One seemingly contradictory finding was that early attentionalylearning impairment was associated with a significantly decreased frequency of family members with manifest schizophrenia. In the original US sample, this form of early deviance was strongly associated with an increased frequency of family schizophrenia. (The overall rates of family history of schizophrenia did not differ in the two subsamples; Table 1.) In a genetically heterogeneous population, the association between early attentionalylearning impairment and family history of schizophrenia broadly suggested that those with this form of early deviance might carry a more familial subtype. However, the absence of this association does not necessarily indicate the reverse.

It is important to note that the findings in this subsample of US patients differed in some regards from our previous full sample analyses (Sobin et al., 2001), which included a more balanced gender distribution (59% males as opposed to 72% males in the current study). In the original study, patients with a history of early social impairment were 2.5 times more likely to experience schizophrenia onset before age 17. In the US subsample examined for the purposes of this study, early social impairment was not significantly associated with age of onset in either the US or Afrikaner subsamples. However, differences were all in the expected direction and the absolute values closely approximated our original findings. In the first report, the (statistically significant) onset age difference in US patients with early social impairment was 1.76 years earlier. In this smaller subsample, the absolute difference was 1.59. In Afrikaner patients with early social impairment, onset of schizophrenia was 1.96 years earlier than, Afrikaner patients without this form of early deviance. Thus, in a larger sample with a more even gender distribution, the difference here reported would likely have reached statistical significance.

Also in our earlier report, patients with attentionalylearning impairment were 2 times more likely to have 18, 28 or 38 relatives with schizophrenia. In the subsample here considered, this difference was minimal. This is likely attributable to the reduced number of females in the matched subsample whose family history of schizophrenia was greater than that of the males in the original sample. (In all three groups—the original US sample, $N=205$; the Afrikaner subsample, $N=109$; and the US subsample, $N=109$—family history of schizophrenia among females was higher than that among males, though not substantially so.)

A finding that was consistent across the full sample and the subsample of US patients but was not found in the Afrikaner subsample was the association between early extreme fearsychronic sadness and later sensory hallucinations. The substantially higher rate of sensory hallucinations among US participants accounted for by females, and not males, may suggest that this symptom is associated with more heterogeneous forms of schizophrenia.

To summarize, as suggested in previous reports (Done et al., 1994; Torrey et al., 1994; Offord and Cross, 1969), early childhood deviance in the subsamples here examined were reported by approximately 68% of patients who met lifetime criteria for schizophreniayschizoaffective disorder, and thus differentiated a large but distinct subgroup of patients whose disorder may begin decades before the onset of full syndromal illness, as indicated by behaviors that are not included in the current nosology. These forms of early deviance are meaningfully linked to later logically related syndrome features, perhaps substantiating their validity as precursors of the full disease syndrome. Indeed, initial molecular findings support a genetic basis for this subtype (Liu et al., 2002). At the same time, the value and the significance of these symptoms appear to differ in these two populations. These differences may be related to the different genetic background of a given patient

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

population. In our large heterogeneous US sample, these behaviors indicated earlier onset age and increased family history of schizophrenia (Sobin et al., 2001). In the more genetically homogeneous Afrikaner population, these behaviors seemed to indicate increased syndrome severity as suggested by the association with markedly increased risk of suicide, poor school achievement and increased auditory hallucinations. While the specific associations differed, they were all clinically consistent and meaningful. Early fears predicted later syndromal features associated with extreme mood disturbance (suicide or schizoaffective disorder), and early attentionalylearning impairment was associated with increased risk of specific types of cognitive disturbance (auditory hallucinations or thought disorder).

4.1. Limitations

The participants in this study were all Euro-Caucasian, albeit of varying descents. While these analyses help extend our previous findings from the US to a cross-national population, the findings may not apply to other races. In order to maximize comparability and reduce confounding effects of age and gender on symptoms and early deviance behaviors reported, we matched the US patients to the Afrikaner patients by age and gender. However, in the Afrikaner sample, the genders were not equally represented. Unequal cell sizes (males 78, females 31) diminished the statistical likelihood of detecting differences between the genders. While we compared the genders in both the main (Afrikaner) and comparison (US) sample and found no differences among the Afrikaners and only symptom differences among the US participants, in samples with large balanced cell sizes, differences between the genders might be more readily identified and subsequently controlled. It is also important to note that matching in this manner altered some of the results obtained in our original study of 205 US patients where the distribution of males to females was more even (120 males, 85 females).

The participant-selection processes in these two samples differed. In the US population, a large majority of participants enrolled at the suggestion of an unaffected family member who had contacted the study recruiter for details regarding participation. In the Afrikaner population, the study enrollment was explained and offered by the treating physician. While one might expect that the method used in South Africa might have yielded a more severe patient population, the marked similarity in frequency of symptoms reported does not support this notion. Nonetheless, it is a potential limitation and deserves further consideration in future studies.

As in the previous study, the preponderance of information regarding early deviant childhood behaviors was obtained from the patients themselves and the validity of this approach can be questioned. In particular patients, current states may significantly influence their report of past experience, and thus this issue is most relevant to the finding of an association between early anxietyychronic sadness and schizoaffective disorder, but may also apply to current perception of social relatedness and early childhood social difficulties. The current states of patients, who met lifetime criteria for schizoaffective disorder were not analyzed for the purposes of this study, but such analyses would be valuable in future studies. The striking similarity of findings in these two independently collected samples may provide support for this approach. Early records are likely to be superior to retrospective report for some of the behaviors queried, but only to the extent that teachers or school personnel recognized and had the opportunity to record information regarding a child's deviant behavior andyor school problems. It is likely that teachers and even parents would be unaware of the inner experiences of their children. We are currently in the process of comparing and contrasting parent and patient reports of early childhood deviance and later clinical outcome. These results should help to suggest which sources are likely to be most informative with regard to specific behavioral experiences.

Acknowledgments

The authors extend special thanks to all of the families for their ongoing participation in our work. They also thank Sister Ria van Wyk at the

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

B1: 176

Weskoppies Hospital for family recruitment. This research was supported by RO1 MH61399 (to MK) and also by a General Clinical Research Center grant (M01-RR00102) from the National Center for Research Resources, National Institutes of Health.

References

Cannon, M., Jones, P., 1996. The epidemiology of schizophrenia. Journal of Neurology, Neurosurgery and Psychiatry 60, 604–612.

Cannon, M., Byrne, M., Cotler, D., Sham, P., Larkin, C., O'Callaghan, E., 1994. Further evidence for anomalies in the handprints of patients with schizophrenia. Schizophrenia Research 13, 179–184.

Cannon, M., Jones, P., Gilvarry, C., Rikin, L., McKenzie, K., Loerster, A., Murray, R.M., 1997. Premorbid social functioning in schizophrenia and bipolar disorder: similarities and differences. American Journal of Psychiatry 154, 1544–1550.

Done, D.J., Crow, T.J., Johnstone, E.C., Sacker, A., 1994. Childhood antecedents of schizophrenia and affective illness: social adjustment at ages 7 and 11. British Medical Journal 310, 57–58.

Foerster, A., Lewis, S., Owen, M., Murray, R., 1991. Premorbid adjustment and personality in psychosis: effects of sex and diagnosis. British Journal of Psychiatry 155, 623–627.

Geddes, J.R., Lawrie, S.M., 1995. Obstetric complications and schizophrenia: a meta-analysis. British Journal of Psychiatry 167, 786–793.

Green, M.F., Bracha, H.S., Satz, P., Christenson, C.D., 1994. Preliminary evidence for an association between minor physical anomalies and second trimester neurodevelopment in schizophrenia. Psychiatry Research 53, 119–127.

Guy, J.D., Majorski, L.V., Wallace, C.J., Guy, M.P., 1983. The incidence of minor physical anomalies in adult male schizophrenics. Schizophrenia Bulletin 9, 571–582.

Jones, P., Rodgers, B., Murray, R., Marmot, M., 1994. Child development risk factors for adult schizophrenia in the British 1946 birth cohort. Lancet 344, 1398–1402.

Lewis, S., 1997. Psychopathology and brain dysfunction: structural imaging studies. In: Keshavan, M.S., Murray, R.M. (Eds.), Neurodevelopment and Adult Psychopathology. Cambridge University Press, Cambridge, England.

Lieberman, J., 1999. Is schizophrenia a neurodegenerative disorder? A clinical and neurobiological perspective. Biological Psychiatry 46, 729–739.

Liu, H., Heath, S.C., Sobin, C., Roos, J.L., Galke, B.L., Blundell, M.L., Lenane, M., Robertson, B., Wijsman, E.M., Rapoport, J.L., Gogos, J.A., Karayiorgou, M., 2002. Genetic variation at the 22q11 PRODH2yDGCR6 locus presents an unusual pattern and increases susceptibility to schizophrenia. Proceedings of the National Academy of Sciences, USA 99, 3717–3722.

McGuire, P.K., Silbersweig, D.A., Wright, I., Murray, R.M., David, A.S., Frackowiak, R.S.J., 1995. Abnormal monitoring of inner speech: a physiological basis for auditory hallucinations. Lancet 346, 596–600.

Nurnberger, J.I., Blehar, M.C., Kaufmann, C.A., York-Cooler, C., Simpson, S.G., Harkavy-Friedman, J., Severe, J.B., Malaspina, D., Reich, T. and collaborators from the NIMH Genetics Initiative, 1994. Diagnostic Interview for Genetic Studies: rationale, unique features, and training. Archives of General Psychiatry 51, 859–949.

O'Callaghan, E., Larkin, C., Kinsella, A., Waddington, J.L., 1991. Familial, obstetric, and other clinical correlates of minor physical anomalies in schizophrenia. American Journal of Psychiatry 148, 479–483.

Offord, D.R., Cross, L.A., 1969. Behavioral antecedents of adult schizophrenia. Archives of General Psychiatry 21, 267–283.

Sobin, C., Blundel, M., Weiller, F., Gavigan, C., Haiman, C., Karayiorgou, M., 2001. Early non-psychotic deviant behavior in schizophrenia: a possible endophenotypic marker for genetic studies. Psychiatry Research 101, 101–113.

Torrey, E.F., Bowler, A.E., Taylor, E.H., Gottesman, I.I., 1994. Schizophrenia and Manic-Depressive Disorder. Basic Books, New York.

Verdoux, H., Geddes, J.R., Takei, N., Lawrie, S.M., Bovet, P., Eagles, J.M., Heun, R., McCreadie, R.G., McNeil, T.R., O'Callaghan, E., Stober, G., Willinger, V.M., Wright, P., Murray, R.M., 1997. Obstetric complications and age at onset in schizophrenia: an international collaborative meta-analysis of individual patient data. American Journal of Psychiatry 154, 1220–1227.

Waddington, J.L., Buckley, P.F., 1996. The neurodevelopmental basis of schizophrenia. Landes, Austin TX.

Walker, E.F., Lewine, R., 1990. The prediction of adult-onset schizophrenia from childhood home movies. American Journal of Psychiatry 147, 1052–1056.

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Case Report

Velocardiofacial syndrome case report:
Is this a homogeneous genetic subtype of Schizophrenia?

H. Roos, Department of Psychiatry, University of Pretoria
EM Hoal-, Department of Human Genetics, University of Pretoria
HW Pretorius, Department of Psychiatry, University of Pretoria
C Sobin, The Rockefeller University, Human Neurogenetics Laboratory, New York
M Karayiorgou, The Rockefeller University, Human Neurogenetics Laboratory, New York

Summary

In this case report of a 21-year-old Afrikaner male, a diagnosis of velocardiofacial syndrome was made on clinical grounds. Later a FISH analysis test for 22 q 11 microdeletion confirmed this clinical diagnosis.

Specifically two independent studies have reported that 25 - 31% of patients with the 22 q 11 microdeletion met diagnostic criteria for schizophrenia or schizo-affective disorders, while the microdeletion occurs in the population at a rate of 0.025%, it has been found in 2% of adult schizophrenic patients and in 6% of cases with childhood onset schizophrenia.

The complete history, physical and psychiatric examination of the 21-year-old male schizophrenic with velocardiofacial syndrome is described.

It is likely that the 22 q 11 region harbours genes that alone, or in combinations, are causally implicated in schizophrenia in certain proportions of patients. Uncovering of one or more genes that predispose to schizophrenia in this region will lead to a better understanding and management of this homogeneous subtype of schizophrenia.
S Afr Fam Pract 2003; 45(2):20-23

INTRODUCTION

Schizophrenia has been called "arguably the worst disease affecting mankind". The worldwide lifetime prevalence of the disorder is ~ 1%[1]. Studies carried out over the last 50 years have demonstrated conclusively that the genetic constitution of an individual plays a large role in determining whether he or she will develop schizophrenia[2,3]. Attempts to dissect out the individual genes that may play a role in the onset of this disease have met substantial difficulties, primarily because the mode of inheritance of the disease is complex and likely involves interaction among multiple genes and environmental factors[4]. The nature of such interaction, the number of susceptibility loci, the risk ratio conferred by each, or the extent of locus heterogeneity, all remain unknown. Thus, conclusive identification and replication of a chromosomal region that harbours a susceptibility locus has not been forthcoming.

Technological advances as well as progress in genetic analytic methods will facilitate the discovery of genes influencing susceptibility to mental disorders. It seems clear that ultimate success will require large-scale data collection efforts to obtain samples of sufficient power for the detection of susceptibility loci of relatively small effect.

It is likely that the requirement for large sample-sizes can be relaxed if the target population presents less heterogeneity at the susceptibility loci, as may be the case for populations founded by a relatively small number of individuals (genetic isolates), such as the Afrikaner population. Several diseases occur at unusually high frequencies among the South African Afrikaners (as much as 5-10 times higher than in most other population groups), such as variegate porphyria, Huntington's chorea, Familial Hypercholesterolemia, Fanconi Anaemia and myotonic dystrophy[5-9].

Moreover, when a founder effect occurs and a mutant allele is introduced within a population with limited admixture, the frequency of a mutant allele can increase rapidly during a few generations. For this reason a collaborative project between The Rockefeller University in New York and the University of Pretoria / Weskoppies Hospital on genetic studies of schizophrenia was initiated in 1997. More than 175 families have been recruited up to now. The ongoing collection of schizophrenic patients of Afrikaner origin in Genetic Studies of Schizophrenia will be described in detail elsewhere (C Sobin et al, manuscript in preparation).

The notable increase in the number of published studies on velocardiofacial syndrome over the past 3 years is likely attributable to data suggesting that the syndrome may represent a homogeneous genetic subtype of schizophrenia[10]. The concurrent breakthrough in genetic research of the complete sequencing of chromosome 22 will undoubtedly fuel additional investigations of the syndrome and its genetic underpinner[10].

A 7 microdeletion of chromosome 22 q 11 is associated with variable phenotypic expression that often

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

includes learning disabilities, palatal abnormalities, congenital heart defects, and mildly dysmorphic facial features[17]. One of the first clues that psychiatric symptomatology may also be associated with this microdeletion was a report of expressionless face, monotonous speech, and flattened affect among children with the microdeletion[15]. In light of the evidence for suggestive linkage for schizophrenia on chromosome 22[14], patients with the 22 q 11 microdeletion were evaluated for psychiatric symptoms or disorders, and a relatively high frequency of severe mental illness has since been reported. Specifically two independent studies have reported that 25 - 31% of patients with the 22 q 11 microdeletion met diagnostic criteria for schizophrenia or schizo-affective disorders[15,16], while the microdeletion occurs in the population at a rate of 0.025%, it has been found in 2% of adult schizophrenic patients[1] and in 6% of cases with childhood onset schizophrenia[18]. In addition, several studies have described 22 q 11 microdeletions among schizophrenic patients of various ethnic origins[19]. These studies collectively suggest that the morbid risk of schizophrenia for a patient with a 22 q 11 microdeletion may be approximately 20-30 times the general population risk of 1% and that the rate of 22 q 11 microdeletions in schizophrenia, although relatively low, may be approximately 80 times the estimated general population rate. It therefore seems likely that the 22 q 11 region harbours genes that alone, or in combinations, are causally implicated in schizophrenia in certain proportions of patients.

In this case study a diagnosis of velocardiofacial syndrome was made on clinical grounds (see history and physical examination). Later a FISH analysis test for 22 q 11 microdeletion confirmed this clinical diagnosis. As approximately 10% of individuals with velocardiofacial syndrome apparently lack the deletion, negative results do not exclude a clinical diagnosis of velocardiofacial syndrome. Of the more than 200 subjects and affected family members so far recruited in this genetic study, only 4 comply with a possible clinical diagnosis of velocardiofacial syndrome. This case study is presented here because the clinical features are typical of a velocardiofacial syndrome.

Case study:

Mr J is a 21-year-old unmarried student who is living with his parents. He was referred to the genetic study by a private psychiatrist.

Birth to 6 years of age:

He is the oldest of 4 siblings. His mother was healthy during pregnancy. There was an induction of labour and he was born by forceps delivery. At birth he had a ventricular septal heart defect. This defect was corrected with cardio thoracic surgery at 13 months of age. (His second brother and youngest sister were born with craniostenosis defects). He had chronic urinary problems, orthodontic problems (two permanent incisor teeth instead of four) and a umbilical hernia that was surgically corrected. His early neuro-development was within normal limits but according to his mother he experienced feeding problems (regurgitated his food).

He attended a pre-school playgroup where he experienced socialization and learning problems. The socialization problems started at 2 years of age. He was not capable of socializing with friends and he is still experiencing this problem. He was very sensitive to loud noise e.g. vacuum cleaner, from one year of age and this continued to 6 years of age. He was afraid of the dark and experienced concentration and day-dreaming problems from a pre-school age and this continued into high-school years. The playgroup teacher said that he was not school ready because of an emotional and motoric delay in his development. He could not speak clearly.

Primary school (7 - 13 years of age)

He did not make good academic progress. He needed his mother's support to cope with his schoolwork. According to his mother he had a low self-image and lacked confidence. A neurologist evaluated him at the age of 7 years. A diagnosis of attention deficit disorder without perceptual disturbance was made. Methylphenidate was prescribed and he attended occupational therapy classes. He attended these classes for one year and received methylphenidate for four years. From grade one he experienced periods of extreme sadness. His mother says she had the impression that this sadness got worse after methylphenidate was prescribed and it improved mildly after the Ritalin was discontinued. At the age of 12 years the mathematics teacher reported that he was behaving inappropriately in class and pupils were laughing at him. According to Mr J he could not control this inappropriate behaviour. During his grade 7 year (13 years of age) he could not think clearly. He says it felt as if a curtain was covering his eyes and he also experienced a metal taste in his mouth. He was evaluated for epilepsy but an EEG was within normal limits. At the end of his primary school years he became very suspicious and anxious towards fellow pupils. He consulted a psychiatrist who diagnosed an anxiety disorder and an anxiolytic and antidepressant were prescribed.

Primary School Reports:

Two school reports at the age of 6 years and 10 years were obtained.

Grade 1 (first term - 6 years of age):

General observations

- Self confidence/independence:
  - He behaves in a insecure manner.
  - Is in constant need of support.
  - His self image can improve.
- Participation in school life:
  - Kind towards teachers.
  - Very obedient but in need of support.
- Fine motoric development:
  - Not on par.
- Major motoric development
  - Has problems with most of these performances.
  - He has no self confidence.
- Attention and concentration:
  - He is starting to listen with attention.
- Speech and language use:
  - Average vocabulary.
  - Speech is unclear to a mild degree.
  - Lack of confidence.
- General remarks
  - There is an improvement in his self-confidence and self-image. We must continue to work on this.

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

- He is anxious and wants to please his teacher.
- Very loveable and obedient.

Grade 4 (10 years of age):
General remarks:
At times he does not participate in classroom conversations. I must talk to him directly to get his attention. He only mixes with the pupil who sits directly next to him.

High school (13 years - 19 years):
He did not have friends and he struggled on academically. He travelled by bicycle to school but had the fear that he would die on his way to school. He says it was as if he could not judge his environment e.g. deciding when to cross a busy street. From grade 10 (15 years) he attended a private school where his mother was teaching. This environment was more supportive and good for him e.g. he was not forced to do oral examination that caused him to become extremely anxious. During the second half of his grade 10 year he told a teacher that demons were causing him problems and he experienced auditory hallucinations. These hallucinations did not continue for long periods. He repeated his grade 11 year because his results were not on a university entrance level. During his grade 12 year (18 - 19 years) he became very depressed. This started after the breaking up of a relationship with a girlfriend. At this stage obsessional thoughts that he had experienced from 13 years of age became worse. He consulted a psychiatrist who prescribed bensodiazepines and hypnotherapy. His grade 12-year results did not allow him to study further at university. His mother says that at this stage he was more than just depressed; he was staring at people and could not think for himself.

Post school (20 years to present state):
He did in service training for 2 months in view of obtaining work on a farm. After his training he worked on several farms for a few months at a time. He left these jobs because of interpersonal problems. He was forgetful and did not progress in his work. During this period he also tried to study for higher grades in order to obtain university entrance. He was not able to get university entrance exemption.

He then enrolled at a private academy for training as a teacher. For the first few months he coped well as this was a small and protective student environment. Then he experienced the following symptoms:

- Derealization.
- Vague somatic complaints.
- Auditory hallucinations.
- Somatic and paranoid delusions.
- Thought process disturbance.
- Bizarre behaviour and rituals.
- Obsessional thoughts (controlled these thoughts better).

The above group of symptoms were present for less than 6 months, but he has been experiencing psychotic symptoms from 16 years of age, not always to the same extent. He was admitted to a private psychiatric hospital and treated with an atypical antipsychotic and mood stabilizer. He could no longer function as a student. When the antipsychotic was reduced his psychotic symptoms got worse.

Physical examination:

Length	182.3 cm
Mass	79 kg
Skull circumference	59 cm

He shows facial dysmorphic features which include a prominent nose, without the broad base typical of velocardiofacial syndrome (VCF), as well as maxillary hypoplasia and a maxillary excess causing a long face. This facial profile fits the classical VCF features. Additionally synophyris (eyebrows meeting in the midline), prominent supraorbital ridges, small ears with prominent helixes, a prominent filtrum and a high palate are seen. His neck and shoulders are broad and he has a minor pectus excavatum. His hands are normal. A thoracic surgical scar is present and his heart is normal on auscultation.

DISCUSSION:

Three contemporary prospective cohort studies specify the types of non-psychotic deficits that can occur in pre-schizophrenic patients from infancy onwards.

In the British 1946 birth cohort study early motor development was delayed and speech problems were more frequent among the pre-schizophrenic group. Preference for solitary play at 4 and 6 years was more frequent among pre-schizophrenic children and cognitive performance including verbal, non-verbal and mathematic skills, measured at ages 8, 11 and 15, were impaired. Furthermore, indicators of deviance increased with age and suggested a progressively worsening developmental trend.[20]

In a similar study, teacher ratings of social development by age 7 of children who later developed schizophrenia, were more behaviourally maladjusted than children in any of the other control groups.[21]

In the other cohort study including 9236 participants, school performance ratings between the ages of 7 and 11 of 338 pre-schizophrenic children were compared to those 338 randomly selected as control. Poor classroom conduct, absenteeism and poor performance in sports and handicrafts (activities dependent upon motor coordination) predicted later schizophrenia.[22]

Retrospective studies of childhood behaviour among schizophrenic patients yielded similar findings. Investigators using home movies made during the first 8 years of life found that abnormal social behaviours and reactions, as well as odd movements and postures, reliably cued clinicians as to which children would go on to develop schizophrenia in later years.

In another study, early social and academic deficits were more frequent among 45 schizophrenic patients as compared to psychiatric control with affective disorder with psychotic features.[24]

Other work investigated the association between schizophrenia onset and early deviant behaviours. Among 61 clinic referrals with schizophrenia onset between ages 7 and 17, and their age- and gender - matched non-psychotic controls, early pre-schizophrenic social impairment, motor deficits and language disturbance occurred in a larger proportion of the juvenile onset schizophrenic patients.[25]

An early study had reported similar findings namely that non-psychotic deviance before age 10 predicted an earlier age of first clinical contact among adult schizophrenic patients.[26]

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Early deviant behaviour as reported in Mr J's case report include:

- Socialization difficulties (onset 2 years) and learning problems (onset pre-school playgroup).
- Concentration and daydreaming problems.
- Emotional and motoric delay (pre school).
- Speech not clear.
- Low self-image and lack of confidence.
- Inappropriate behaviour (causing children to laugh at him).

School reports confirmed some of the above behaviour as follows:

- 6 years (Gr. 1):
  - Fine and major motoric behaviour problems.
  - Speech unclear.
  - Attention problems.
  - Lacks self confidence and is anxious.
- 10 years (Gr. 4):
  - Attention problems.
  - Only mixes with pupil seated next to him.

Whether these behaviours indicate syndrome onset in some patients has yet to be determined.

The researchers in the present genetic study mostly choose to question patients rather than parents on early childhood behaviour by using a simple semi-structured interview format that precedes the DIGS interview (Diagnostic Interview for Genetic Studies).[27] Brief initial interviews with parents suggested a number of different biases:

- No memory of childhood behaviours.
- Reporting a global impression that the ill adult was impaired in every way - but unable to provide behaviour samples.
- Did not want to provide information that would reflect badly on the child.

In this case report the mother's information on childhood behaviour was objective and more reliable than the information obtained from the patient. The school reports were objective and added value to the existing information. The reporting of learning disabilities/attention impairment is likely attributable to the increased awareness and detection of learning disabilities over the past 20 years.

The early deviant behaviour became more severe over time in Mr J's case report, as DSM IV-defined schizophrenia approaches. Early non-psychotic deviance does not occur in 100% of schizophrenic patients, however, summarized results from eight independent investigators concluded that 60% of schizophrenic patients functioned normally until the onset of cardinal symptoms.[28]

The onset of Mr J's psychotic symptoms was earlier than the full-blown psychotic breakdown he had at 20 years of age. During his grade 10 school year (16 years old), he experienced auditory hallucinations, delusions, mood symptoms, and strange behaviour with a definite decline in his overall functioning. This sets the onset of Mr J's illness at 16 years and perhaps even earlier (± 13 years).

Age of onset of illness has been an avenue to the understanding of disease across all of medicine, with earlier onset cases often having more striking genetic and environmental influence and/or differing pathophysiologies.[29]

In some cases, the early onset disorder may be physiologically and etiologically different from the later onset illness. For example, the study of early onset disease led to a description of different and important mechanisms of illness. Insulin-dependent diabetes mellitus[30] have been tied to an earlier age of onset of illness, attesting to the importance of the topic. While most cases of schizophrenia have their onset in later adolescence and early adulthood,[31] the disorder has been identified in children since early in the 20th century.

The predominant clinical signs of velocardiofacial syndrome include cleft palate or velopharyngeal insufficiency, cardiac abnormalities, a specific set of facial features, and learning disabilities[34].

Mr J has most of these clinical signs.

Obsessions and compulsions and schizophrenic symptoms may assume various relationships. In some cases the symptoms remain distinct; insight into the irrationality of the obsessions and compulsions is preserved, and they are experienced as ego-alien. In other cases, the obsessions and compulsions become the subject of delusional elaboration during periods of active psychosis. In still other cases, the OCD symptoms are inseparable from delusional beliefs[35].

Some patients with OCD lose insight into the irrationality of their obsessions and compulsions in the course of their illness and become delusional. They can be distinguished from schizophrenic patients with comorbid OCD by the absence of the other signs and symptoms of schizophrenia[35].

The presence of obsessive-compulsive symptoms in patients with schizophrenia is associated with a much poorer long-term outcome[36].

In Mr J's case the obsessions remained distinct initially during the course of the illness. As the active psychosis became more prominent, the obsessions became the subject of delusional elaboration. As the psychotic illness progressed, the obsessions became less distinct.

Initially this family was not keen on becoming involved in the genetic study. After the psychiatrist's diagnosis of schizophrenia was discussed with the patient and the family they reconsidered the option of participating in this study. To them the diagnosis of schizophrenia is still far removed from a medical illness and is stigmatised. The fact that a genetic abnormality may be linked to the development of schizophrenia made it more acceptable as a medical illness for this patient and his family.

CONCLUSIONS:

Because velocardiofacial syndrome and the concomitant symptomatology result from a known genetic region (micro deletions on chromosome 22 q 11), the biological and behavioural characteristics of the syndrome provide an optimal framework for conceptualising the associations among genes, brain development and behaviour.

In the near future, with the uncovering of one or more genes that predispose to schizophrenia on the 22 q 11 chromosome, we will start to understand, and possibly manage, this homogeneous genetic subtype of schizophrenia in a better way.

Hopefully these advances will also help to destigmatise "arguably the worst disease affecting mankind". ❐

References available on page 28.

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

3. Antibiotics:

Antibiotics are given for severe and moderately severe acute exacerbations.

4. Oxygentherapy:

Proper care needs to be taken not to worsen respiratory failure, but hypoxaemia needs to be relieved.

5. Non-invasive Positive-Pressure Ventilation (NPPV):

NPPV might improve the survival of patients with acute exacerbations of COPD. The following treatment options are not recommended and some may be harmful in the treatment of acute exacerbations:

i. Mucolytic medications.
ii. Chest physiotherapy.
iii. Methylxanthines.

V. Alpha-1-Antitrypsin deficiency replacement:

Exogenous α-1-AT derived from pooled human plasma administered intravenously weekly is an option for severe deficiency, but it is inconvenient and expensive.

SUMMARY

Relentless dyspnoea is a constant feature of COPD and these measures are currently available to relieve dyspnoea. The mainstay of current treatment is described with cessation of smoking of paramount importance.❐

Please refer to CPD Questionnaire on pg 51

References:

1. Snow V et al. Joint expert panel on chronic obstructive pulmonary disease of American College of Chest Physicians. *Ann Intern Med* 2001; 134: 595-599.
2. Ferguson GT et al. *N Engl J Med* 1993; 328: 1017-1022.
3. Barnes P J. *Thorax* 1998; 53: 137-147.
4. Mannino DM. *Chest* 2002; 121 [Suppl]: 121S-126S.
5. Tobin MJ. *Am J Respir Crit Care Med* 2002; 165: 642-662.
6. Fabbri L et al. *Clin Exp All Rev* 2002; 2: 129-136.

(Continued from page 23)

References:

1. Gottesman TI, Shields J. *Schizophrenia: The Epigenetic Puzzle*. Cambridge University Press: Cambridge, UK, 1982.
2. Kendler KS, Diehl SR. The genetics of schizophrenia: a current, genetic epidemiologic perspective. *Schiz Bulletin*, 1993;19:261-285.
3. Kety SS, Wender P, Jacobsen B, et al. Mental Illness in the biological and adoptive relatives of schizophrenic adoptee: replication of the Copenhagen study in the rest of Denmark. *Arch Gen Psychiatry*, 1994;51:442-455.
4. Tsuang MT, Faraone SV. The case for heterogeneity in the etiology of schizophrenia. *Schiz Research*, 1995;7:161-175.
5. Warnich L, et al. Identification of three mutations and associated haplotypes in the protoporphytinogen oxidase gene in South Africa. *Hum Mol Genet*, 1996;5:981-984.
6. Hayden MR, MacGregor JM, and Beighton PH. The origin of Huntington's chorea in the Afrikaner population. S *Afr Med J*, 1980;58: 197-200.
7. Brink PA, Steyn LT, Coetzee GA, and Van der Westhuizen DR. Familial hypercholesterolemia in South African Afrikaners. Pvu II and Stu I DNA polymorphisms in the LDL-receptor gene consistent with a predominating founder gene effect. *Hum Genet*, 1987;77:32-35.
8. Rosendorff J, Bernstein R, Macdougall L, and Jenkins T. Fanconi anaemia: another disease of high prevalence in the Afrikaans population of South Africa. *Am J Genet*, 1987;27:793-797.
9. Goldman A, Krause A, Ramsay M, and Jenkins T. Founder effect and the prevalence of myotonic dystrophy in South Africans: Molecular studies. *Am J Hum Genet*, 1996;59:445-452.
10. Basset AS, Chow EW. 22 q 11 Deletion Syndrome: a genetic subtype of schizophrenia. *Biol Psychiatry*, 1999;46:882-891.
11. Dunham I, et al. The DNA sequence of human chromosome 22. *Nature*, 1999;402:489-495.
12. Scambler PJ. The 22 q 11 deletion syndrome. *Hum Mol Genet*, 2000;9:2421-2426.
13. Golding-Kushner KJ, Weller G, Shprintzen RJ. Velocardiofacial syndrome: language and psychological profiles. *J Cranio-Facial Genet Devel Biolog*, 1985;5:259-266.
14. Baron M. Genetics of Schizophrenia and the New Millenium: Progress and Pitfalls. *Am J Hum Genet*, 2001;68:299-312.
15. Pulver AE, et al. Psychotic Illness in Patients diagnosed with Velocardiofacial Syndrome and the Relatives. *J Nerv Mental Disease*, 1994;182:476-478.
16. Murphy KC, Jones LA, and Owen MJ. High rates of schizophrenia in adults with velocardiofacial syndrome. *Arch Gen Psychiatry*, 1999;56: 940-945.
17. Karayiorgou M, et al. Schizophrenia Susceptibility Associated with Interstitial Deletions of Chromosome 22 q 11. *Proc Natl Acad Sci*, USA. 1995;92:7612-7616.
18. Usiskin SI et al. Velocardiofacial syndrome in childhood-onset schizophrenia. *J Am Acad Child Adolesc Psychiatry*, 1999;38: 1536-1543.
19. Sugama S, Namihira T, Matsuoka R, et al. Psychiatric inpatients and chromosome deletions within 22 q 11.2. *J Neurol Neurosurg Psychiatry*, 1999;67:803-806.
20. Jones P, Rodgers B, Murray R, and Marmot M. Child development risk factors for adult schizophrenia in the British 1946 birth cohort. *Lancet*, 1994;334:1398-1402.
21. Done DJ, Crow TJ, Johnstone EC, and Sacker A. Childhood antecedents of schizophrenia and affective illness: social adjustment at ages 7 and 11. *British Medical Journal*, 1994;310:57-58.
22. Cannon TD, Rosso IM, Bearden CE, Sanchez IE, and Hadley T. A prospective cohort study of neurodevelopmental processes in the genesis and epigenesis of schizophrenia. *Developmental and Psychopathology*, 1999;11:467-485.
23. Walker EF, Lewine R. The prediction of adult onset schizophrenia from childhood home movies. *American Journal of Psychiatry*, 1990; 147:1052-1056.
24. Foerster A, Lewis S, Owen M, and Murray R. Premorbid adjustment and personality in psychosis: effects of sex and diagnosis. *British Journal of Psychiatry*, 1991;155: 623-627.
25. Hollis C. Child and adolescent (juvenile onset) schizophrenia: a case control study of premorbid developmental impairments. *British Journal of Psychiatry*, 1995;166: 489-495.
26. Belmont L. Perceptual evidence of CNS dysfunction in schizophrenia. *Archives of General Psychiatry*, 1964;10:395-408.
27. Nurnberger J, et al. Diagnostic Interview for Genetic Studies: Rationale, unique features, and training. *Archives of General Psychiatry*, 1994;51:949.
28. Offord Dr, Cross LA. Behavioural antecedent of adult schizophrenia. *Archives of General Psychiatry*, 1969;21:267-283.
29. Childs B, Scriver CR. Age at onset and causes of disease. *Perspect BioMed*, 1986;29:437-460.
30. Weiss KM. Genetic Variation and Human Disease: Principles and Evolutionary Approaches, 354. Cambridge University Press, New York, 1993.
31. Fitzgerald MG, et al. Germ-line BRCA1 mutations in Jewish and non-Jewish women with early onset breast cancer. *N Engl J Med*, 1996; 334:143-149.
32. St George-Hyslop PH, et al. Two homologous genes causing early-onset familial Alzheimer's disease. *Cold Spring Harb Symp Quant Biol*, 1996;61:559-564.
33. Hafner H, Maurer K, Loffler W, and Riecher-Rossler A. The influence of age and sex on the onset and early course of schizophrenia. *Br J Psychiatry*, 1993;162:80-86.
34. Goldberg R, Motzkin G, Marion R, Scambler PJ, and Shprintzen RJ. Velocardiofacial syndrome: a review of 120 patients. *Am J Med Genet*, 1993;45:313-319.
35. Koran LM. Obsessive-compulsive and related disorders in adults. *A comprehensive clinical guide*, 115, Cambridge University Press: UK, 1999.
36. Fenton WS, and McGlashan TH. The prognostic significance of obsessive-compulsive symptoms in schizophrenia. *Am J Psychiatry*, 1986;143:437-441.

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

American Journal of Medical Genetics Part B (Neuropsychiatric Genetics) 129B:20–22 (2004)

Brief Research Communication

Assessment of the Frequency of the 22q11 Deletion in Afrikaner Schizophrenic Patients

G.J. Wiehahn,[1] G.P. Bosch,[1] R.R. du Preez,[1] H.W. Pretorius,[1] M. Karayiorgou,[2] and J.L. Roos[1]*

[1]University of Pretoria Department of Psychiatry & Weskoppies Hospital, Pretoria, Republic of South Africa
[2]The Rockefeller University, Human Neurogenetics Laboratory, New York, New York

A hemizygous deletion of the q11 band on chromosome 22 occurs in 1 of every 5,950 live births (0.017%). The deletion is mediated by low copy repeats (LCRs) flanking this locus. Presence of the deletion is associated with variable phenotypic expression, which can include distinctive facial dysmorphologies, congenital heart disease and learning disabilities. An unusually high percentage of individuals with this deletion (25–30%) have been described to develop schizophrenia or schizoaffective disorder. In previous studies, the prevalence of the 22q11 deletion in patients with schizophrenia was found to be approximately 2% in Caucasian adults and 6% in childhood-onset cases. Both these frequencies represent a dramatic increase from the prevalence of the deletion in the general population. In this study, we investigate the occurrence of the 22q11 deletion in an independent sample of schizophrenic patients of Afrikaner origin. We first ascertained a sample of 85 patients who meet full diagnostic criteria for schizophrenia for presence of two or more of the clinical features associated with presence of the 22q11 deletion. A group of six patients (7%) met these criteria. This group was subjected to fluorescent in situ hybridization (FISH) and presence of the 22q11 deletion was confirmed for two subjects. Our study therefore confirms the previously reported rate of 2% frequency of the 22q11 deletion in adult schizophrenic patients and provides a two-stage screening protocol to identify these patients. © 2004 Wiley-Liss, Inc.

KEY WORDS: 22q11 deletion; schizophrenia; Afrikaners

The etiology of schizophrenia is unknown. Genetic epidemiologic studies have identified a strong genetic component in the etiology of the disease [Gottesman and Shields, 1982], but the high degree of genetic and environmental heterogeneity that seems to exist has hindered gene identification. One chromosomal region under intense investigation is located on the long arm of chromosome 22. In addition to evidence from linkage studies, it has been shown that patients with a hemizygous deletion of the q11 band of their chromosome 22 (otherwise diagnosed as 22q11 deletion syndrome, 22q11DS) are at very high risk for serious psychiatric illness. Specifically, in a study of 14 adults with 22q11DS, the most common psychiatric disorder found on direct interview using DSM-IV criteria was schizophrenia (29%) [Pulver et al., 1994]. This finding was replicated in a second, larger study that included 40 adults with 22q11DS [Murphy et al., 1999]. These two studies together suggest that the morbid risk of schizophrenia for a patient with 22q11DS may be approximately 29 times the general population risk and is three times the risk of a first-degree relative of an individual with schizophrenia [Gottesman and Shields, 1982].

22q11 microdeletions are mediated by low copy repeats (LCRs) [Edelmann et al., 1999] and are not detected by routine karyotype analysis. The microdeletions are detected using fluorescence in situ hybridization (FISH) technique [Scambler, 2000]. Their estimated prevalence is 1 in 5,950 live births [Botto et al., 2003]. This makes 22q11DS the second most common genetic syndrome, Down syndrome being the most prevalent. The 22q11DS occurs most frequently as a sporadic mutation, but inheritance plays a role in 10% of cases. The 22q11DS can have highly variable phenotypic expression that can include facial dysmorphologies, cardiovascular anomalies, oral clefts, thymus and parathyroid anomalies, and learning disabilities. The deletions occur in up to 85% of individuals that have the phenotype of the velo-cardio-facial syndrome [Scambler, 2000].

The dysmorphic features associated with the 22q11DS are distinct and include a long narrow face with flat cheeks, notched alae nasi, a prominent nose, small low set ears, a small mouth, and a retruded chin. Specific dysmorphologies of the velum include cleft palate (usually submucosal), velo-pharyngeal incompetence, and therefore hypernasal speech.

Karayiorgou et al. [1995] examined a sample of schizophrenic patients for presence of the 22q11 deletion and identified it in a frequency of 2%. This is a substantial increase from the 1:5,950 (0.017%) frequency with which this deletion is found in the general population. Furthermore, Usiskin et al. [1999] estimated the frequency of the 22q11 deletion in a sample of childhood-onset schizophrenia cases (onset before the age of 13) to 6%. One other study screened a large cohort of Japanese adult schizophrenic patients and identified the deletion in 0.33% [Arinami et al., 2001], while two independent screens of schizophrenic patients who had been pre-selected for presence of one or more of the physical features associated with 22q11DS identified the deletion in 20 and 59%, respectively [Gothelf et al., 1997; Bassett et al., 1998]. These studies together, along with the clinical studies mentioned above [Pulver et al., 1994; Murphy et al., 1999] have made a strong case for 22q11DS to be a genetic subtype of schizophrenia [Bassett and Chow, 1999].

Here, we estimate the frequency of the 22q11 deletion in an independent sample of adult schizophrenic patients. We contacted patients who have already participated in an ongoing genetic study of schizophrenia at the University of Pretoria, Department of Psychiatry and the Human Neurogenetics Laboratory of the Rockefeller University in New York. All patients have undergone in-person diagnostic interviews

*Correspondence to: Professor J. Louw Roos, Chairman, Department of Psychiatry, Weskoppies Hospital, Ketjen Street, Pretoria West 0002, Republic of South Africa.
E-mail: inaroos@cis.co.za

Received 17 March 2003; Accepted 21 October 2003
DOI 10.1002/ajmg.b.20168

© 2004 Wiley-Liss, Inc.

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

TABLE I. Characteristics of the Six Patients Identified Through Our Screening as Candidate for Presence of the 22q11 Deletion

Sex	Age	Education (# years of school completed)	Age of SCZ onset	MR	History of learning disabilities	22q11 del	Clinical features
Female	54	15	35	No	No	No	Facial features (low-seated small ears); kyphosis
Male	31	12	22	No	Yes	No	Facial features (retruded chin); high palate; ventricular septal defect
Male	27	10	24	No	Yes	No	Facial features (long narrow face, prominent nose); high palate
Female	21	12	19	No	Yes	Yes	Facial features (long narrow face, prominent nose, retruded chin, long slender fingers); high palate
Male	18	9	17	Yes	Yes	No	Facial features (long narrow face, prominent nose); high palate
Male	21	12	16	No	Yes	Yes	Facial features (narrow palpebral fissures, low-seated small ears, prominent nose, flat cheeks); ventricular septal defect

All six patients met at least two of the groups of screening criteria listed in Bassett and Chow [1999].

using the Diagnostic Instrument for Genetic Studies (DIGS) [Nurnberger et al., 1994], which was translated and back translated into Afrikaans for use in this study. The clinical interviewers are senior psychiatrists from the Weskoppies Hospital in Pretoria with a minimum of 10 years clinical experience who were trained in the use of the DIGS and the research application of DSM-IV. On the basis of information gathered in the DIGS, the clinical interviewers assign appropriate DSM-IV diagnoses. Following each interview, they are responsible for the completion of DSM-IV checklists for each category in which positive symptoms are identified in the DIGS; a narrative chronologic summary of prodromal traits, symptom onset, and functional impairment; and an English-language data summary form of selected DIGS items. Upon completion, these materials are entered in a computerized database. More detailed description of this sample can be found in Karayiorgou et al. [2004].

Approval for this study was obtained from the Ethical Committee of the Faculty of Health Sciences of the University of Pretoria. After the patients were fully informed of all aspects of the study, a written consent was obtained.

The patients who had already participated in the genetic study were contacted by telephone to be asked if they were willing to participate in the 22q11DS study. Patients willing to participate were evaluated at Weskoppies Hospital, Pretoria [Tshwane], or a home visit was done. A total of 85 subjects were recruited and included in the study [52 males and 33 females, 17–48 years of age (mean age ± SD, 31.8 ± 10 years), mean age of onset of schizophrenia ± SD, 22.9 ± 6.7]. Examination focused on the physical anomalies associated with 22q11DS, as well as evidence for mental retardation (MR), history of learning disabilities, or history of special education. Presence or absence of each feature was noted and the screening criteria suggested by Bassett and Chow [1999] were applied to identify patients at increased risk for 22q11DS. According to Bassett and Chow [1999], patients with two or more of the following groups of screening criteria are at increased risk for 22q11DS: (1) hypernasal speech, history of speech therapy, velopharyngeal incompetence, cleft palate (usually submucosal); (2) characteristic facial features: e.g., long narrow face, narrow palpebral fissures, flat cheeks, prominent nose, small ears, small mouth, retruded chin; (3) learning difficulties, history of special education, MR (borderline to mild); (4) congenital heart defects: e.g., ventricle septum defect, tetralogy of Fallot, right Aortic arch, double aortic arch; (5) other significant congenital anomalies: e.g., club feet, polydactily, renal anomalies, kyphosis/scoliosis, hypospadia; (6) history of hypocalcaemia, and/or hypoparathyroidism; (7) history of athymia or severe immunodeficiency in infancy.

Of the 85 subjects examined, 27 had one or more of the physical anomalies associated with 22q11DS. Of these, six patients (four males, two females) had two or more of the positive screening criteria (Table I). A FISH analysis was performed on these six patients and two (one male, one female) were found to carry hemizygous deletions on chromosome 22q11. The FISH-based identified rate of 22q11DS in this sample (2.4%, 2 of 85) is clearly a minimal estimate, since we did not test by FISH all 85 patients. However, in an independent set of studies described in Liu et al. [2002a,b], the entire sample was genotyped using a panel of single nucleotide polymorphisms (SNPs) from the 22q11 locus. In the group of 85 schizophrenic patients described here, the same two individuals identified by FISH as carrying the 22q11 deletion were apparently "homozygous" for all tested markers. The rest of the patients were heterozygous for at least one marker locus, thus excluding the presence of a typical 22q11 microdeletion. Therefore, we conclude that the rate of 2 in 85 represents an accurate estimate and supports our conclusion that the accuracy of the clinical screening for 22q11DS, as performed in this study, is 33%.

One of the subjects carrying the deletion had typical facial dysmorphisms and a ventricular septal defect. The second subject also had typical facial dysmorphisms and high palate/hypernasal speech (Table I). Both subjects were born by forceps delivery and experienced feeding problems as infants. Motor development was delayed and they experienced socialization and learning problems since preschool age. The onset of the psychosis was insidious and they experienced paranoid delusions, auditory hallucinations, and thought process disturbance during the acute phase of their illness. Interestingly, depression and obsessions in their late teens were prominent features of both subjects' illness, and for one of them full diagnostic criteria for obsessive compulsive disorder (OCD) were met. Both patients were resistant to typical antipsychotic (haloperidol and trifluoperazine) treatment. Atypical antipsychotics (risperidone) and combination treatment modalities (risperidone and SSRIs, e.g., citalopram) were necessary to control the symptoms.

Our study confirms in an independent sample that the frequency of the 22q11DS in adult Caucasian schizophrenic

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

patients is approximately 2% as previously suggested [Karayiorgou et al., 1995]. The current study adds to the existing literature that strongly supports an important contribution of the 22q11 locus in schizophrenia.

We also demonstrate that the screening of samples for clinical features compatible with the 22q11DS is relatively simple to perform, identifies accurately patients with the 22q11DS (in this study, one third of the "flagged" cases were true positives and there were no false negatives) and therefore narrows considerably the sample size in need of confirmation by FISH. The criteria we used, as suggested by Bassett and Chow [1999], are useful despite differences in their specificity and predictive value regarding 22q11DS. Some of the criteria include conditions that are rare in the population, but have high predictive value (such as interrupted aortic arch type B, where one of every two cases carries the 22q11 deletion). Other criteria include conditions that are common in the population but have less specificity or predictive value regarding 22q11DS, such as scoliosis, kyphosis, or learning disabilities.

Our sample of schizophrenic patients with 22q11DS is very small (N¼2) to allow any kind of further analysis in terms of symptom quality, symptom severity, drug response, age at onset, etc. Studies on larger cohorts of schizophrenic patients with the 22q11 deletion will be important to define more accurately the clinical phenotype of this subset of Afrikaner patients and possibly differentiate it from other schizophrenia cases.

REFERENCES

Arinami T, Ohtsuki T, Takase K, Shimizu H, Yoshikawa T, Horigome H, Nakayama J, Toru M. 2001. Screening for 22q11 deletions in a schizophrenia population. Schizophr Res 52:167–170.

Bassett AS, Chow EWC. 1999. 22q11 deletion syndrome: A genetic subtype of schizophrenia. Biol Psychiatry 46:882–891.

Bassett AS, Hodgkinson K, Chow EW, Correia S, Scutt LE, Weksberg R. 1998. 22q11 deletion syndrome in adults with schizophrenia. Am J Med Genet 81:328–337.

Botto LD, May K, Fernhoff PM, Correa A, Coleman K, Rasmussen SA, Merritt RK, O'Leary LA, Wong LY, Elixson EM, Mahle WT, Campbell RM. 2003. A population-based study of the 22q11.2 deletion: Phenotype, incidence, and contribution to major birth defects in the population. Pediatrics 112:101–107.

Edelmann L, Pandita RK, Morrow BE. 1999. Low-copy repeats mediate the common 3-Mb deletion in patients with velo-cardio-facial syndrome. Am J Hum Genet 64:1076–1086.

Gothelf D, Frisch A, Munitz H, Rockah R, Aviram A, Mozes T, Birger M, Weizman A. 1997. Velocardiofacial manifestations and microdeletions in schizophrenic inpatients. Am J Med Genet 72:455–461.

Gottesman II, Shields J. 1982. Schizophrenia: The epigenetic puzzle. Cambridge, UK: Cambridge University Press.

Karayiorgou M, Morris MA, Morrow B, Shprintzen RJ, Goldberg R, Borrow J, Gos A, Nestadt G, Wolyniec PS, Lasseter VK, Eisen H, Childs B, Kazazian HH, Kucherlapati R, Antonarakis SE, Pulver AE, Housman DE. 1995. Schizophrenia susceptibility associated with interstitial deletions of chromosome 22q11. Proc Natl Acad Sci USA 92:7612–7616.

Karayiorgou M, Torrington M, Abecasis GR, Pretorius H, Robertson B, Kaliski S, Lay S, Sobin C, Möller N, Lundy LS, Blundell ML, Gogos JA, Roos JL. 2004. Phenotypic characterization and genealogical tracing in an Afrikaner schizophrenia database. Am J Med Genet 124B:20–28.

Liu H, Heath SC, Sobin C, Roos JL, Galke BL, Blundell ML, Lenane M, Robertson B, Wijsman EM, Rapoport JL, Gogos JA, Karayiorgou M. 2002a. Genetic variation at the 22q11 PRODH2/DGCR6 locus presents an unusual pattern and increases susceptibility to schizophrenia. Proc Natl Acad Sci USA 99:3717–3722.

Liu H, Abecasis GR, Heath SC, Knowles A, Demars S, Chen Y-J, Roos JL, Rapoport JL, Gogos JA, Karayiorgou M. 2002b. Genetic variation in the 22q11 locus and susceptibility to schizophrenia. Proc Natl Acad Sci USA 99:16859–16864.

Murphy KC, Jones LA, Owen MJ. 1999. High rates of schizophrenia in adults with velo-cardio-facial syndrome. Arch Gen Psychiatr 56:940–945.

Nurnberger JI, Blehar MC, Kaufmann CA, York-Cooler C, Simpson SG, Harkavy-Friedman J, Severe JB, Malaspina D, Reich T, and collaborators from the NIMH Genetics Initiative. 1994. Diagnostic interview for genetic studies, rationale, unique features, and training. Arch Gen Psychiatry 51:949.

Pulver AE, Nestadt G, Goldberg R, Shprintzen RJ, Lamacz M, Wolyniec PS, Morrow B, Karayiorgou M, Antonarakis SE, Housman DE, Kucherlapati R. 1994. Psychotic illness in patients diagnosed with velo-cardio-facial syndrome and their relatives. J Nerv Mental Disease 182:476–478.

Scambler PJ. 2000. The 22q11 deletion syndromes. Hum Mol Genet 9:2421–2426.

Usiskin SI, Nicolson R, Krasnewich DM, Yan W, Lenane M, Wudarsky M, Hamburger SD, Rapoport JL. 1999. Velocardiofacial syndrome in childhood-onset schizophrenia. J Am Acad Child Adolesc Psychiatry 38:1536–1543.

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

ORIGINAL

S Afr Psychiatry Rev 2004;7:10-14

Multiple affected Afrikaner families in a schizophrenia genetic study: environmental risk factors in interaction with genotypes

Johannes L Roos[1], **Herman W Pretorius**[1], **Maria Karayiorgou**[2]

[1] *Department of Psychiatry, University of Pretoria, Pretoria, South Africa*

[2] *Human Neurogenetic Laboratory, The Rockefeller University, New York, USA*

Abstract

The authors report on six multiple affected Afrikaner families suffering from schizophrenia or schizoaffective disorders. These families form part of an ongoing study on genetics of schizophrenia. Three or more first degree relatives were affected in these families. In each family, the following will be reported on: a family tree, sociodemographic data, diagnostic features, substance abuse, early insults, early deviant behaviour, and longitudinal course of illness. Environmental risk factors in interaction with genotypes are discussed and it is emphasised that diseases will tend to cluster in families not because of a direct genetic effect, but because relatives are more vulnerable to the risk-increasing effect of a prevalent environmental risk factor. Progress in the study of environmental factors that interact with genes needs to go hand in hand with developments in molecular genetics.

Key words: *Multiple affected founder families; genetics; schizophrenia; environmental risk factors*

Received: 01.04.2004
Accepted: 29.04.2004

Schizophrenia has a heritability in the 60% to 90% range.[1] The search for chromosomal loci and genes has been slow and frustrating. The reason for this may be because there is multiple susceptibility genes involved, each of small effect. In many cases the development of the disorder may depend on the synergistic interaction among particular patterns of genes (epistasis) and not just on the cumulative additive effect of multiple susceptibility genes.[2] The epigenetic processes and environmental factors act in conjunction with the multiple susceptibility genes, each of small effect. The empirical evidence shows that environmental influences are strong and pervasive but rarely determinative.[1]

In the case of genotype-environment interaction, diseases will tend to cluster in families, not because of a direct genetic effect, but because relatives are more vulnerable to the risk-increasing effect of a prevalent environmental risk factor. Some of the affected probands to be discussed in this paper were raised by a schizophrenic mother or father, which could have exposed these probands to a dysfunctional family rearing environment. The Finnish Adoptive Family Study of Schizophrenia showed that the risk of developing schizophrenia spectrum disorders in high-risk adoptees was greater than for control adoptees, but only for those high-risk adoptees who were additionally exposed to a dysfunctional family rearing environment.[3]

There are certain expectations that *founder* populations may be useful to circumvent some of these obstacles in schizophrenia genetic research. Founder populations may offer reduced genetic heterogeneity, a more uniform environment, and the potential for genealogical research. These populations have proven useful for mapping genes that underlie Mendelian disorders and should also facilitate the search for genes for schizophrenia.[4]

In a recent report of an ongoing study of the genetics of schizophrenia it was concluded that the Afrikaner population is likely an appropriate founder population to map genes for schizophrenia using both linkage and linkage disequilibrium (LD) approaches.[4] The authors are involved in this ongoing study.

Most previous quantitative genetic studies of schizophrenia have employed a phenotypic syndromal definition of this illness based on a single main lifetime diagnosis. Another approach requires a diagnostic hierarchy, with schizophrenia at the top, followed by schizoaffective disorder, then bipolar disorder. If, for example, in the early stages of an illness an individual shows symptoms that fulfill the criteria for a manic or depressive episode but later the clinical presentation becomes typical of schizophrenia, the main lifetime diagnosis is regarded as schizophre-

Correspondence:
Professor JL Roos, Department of Psychiatry, Auditorium
PO Box X113, Pretoria, 0001, South Africa
email: avgreun@postillion.up.ac.za

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

nia and the earlier mood symptoms are regarded as non-specific and not considered in further analysis. A hierarchical diagnosis should take all these diagnostic aspects of the psychotic equation into account. Supplementing the traditional approach of assigning a single main lifetime diagnosis, with information on within person co morbidity of psychotic syndromes, may provide valuable information about the familial aggregation of psychotic symptoms.[5]

We report on six multiple affected families (where three or more first degree relatives were affected) with schizophrenia or schizoaffective disorders. Certain sociodemographics, diagnostic features, substance abuse, early deviant behaviour, early insults and long-term outcome will be reported on. Aspects of the nature-nurture interplay in these families will be discussed.

Methodology

Since 1997 a collaborative study between the Human Neurogenetic Laboratory of The Rockefeller University, New York and the Department of Psychiatry, University of Pretoria, South Africa has been ongoing. Approval for the protocol was obtained from the IRB committees at both sites. All participants gave written informed consent.

Our current sample includes more than 300 probands of Afrikaner descent who meet diagnostic criteria for schizophrenia or schizoaffective disorder, according to the American Psychiatric Association 1994.[6] Of the probands ± 20% are members of families with two or more affected individuals, while the rest are members of triads (affected individual and both biological parents). Families where three or more first-degree relatives are affected are scarcer and six such families were identified. In this paper certain data obtained from these six families will be discussed.

For the full detail of the sample collection, clinical evaluation and Genealogical Research the reader is referred to a paper by M Karayiorgou et al, 2004.[4] In the multiple affected families an affected subject was recruited and then the other affected family members were approached for inclusion in the study. In other instances the researchers were notified from the start that it is a multiple affected family, and the members were then approached individually for inclusion in the study.

The clinical evaluation consisted of the following:-

1. The diagnostic interview for genetic studies (DIGS)[7] was used for all diagnostic evaluations. Continued interviewer reliability and reduction in interviewer drift were ensured. Reliability of final diagnosis was found to be excellent. The diagnostic interviews were conducted in Afrikaans.
2. Following each interview, diagnosticians completed DSM-IV checklists for each category in which positive symptoms were identified in the DIGS.
3. A narrative chronological summary of prodromal traits, symptom onset, and functional impairment.

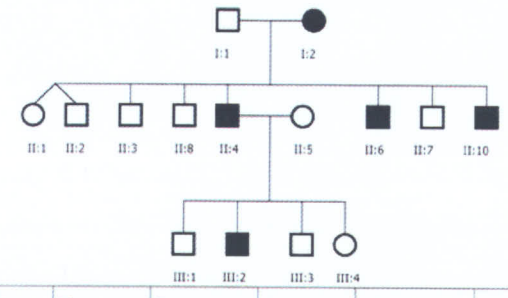
The following aspects were extracted from the above mentioned data in each multiple affected family: main psychiatric diagnosis and age of onset of illness, comorbid psychiatric conditions and age of onset, age at recruitment, substance abuse history, early life insults, early deviant behaviour (1-10 yrs) and long-term course and pattern of severity. In the results section a familytree will be presented for each multiple affected family as well as the above mentioned data for each numbered affected proband.

Results

The reason for not including certain data in family III & VI was the following:

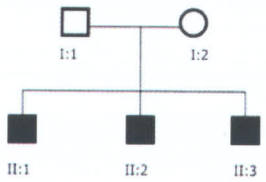
In family III, subject 11:2 had passed away at the time of the initial interview and retrospective data was not available. In family VI, subject 1:1 was not prepared to participate in the study.

Family I



Proband identification	I:2	II:4	II:6	II:10	III:2
Diagnosis (age of onset) and comorbid diagnosis	Schizoaffective disorder (bipolar) (46 yrs) Post partem psychotic disorder (Psychosis NOS) (25 yrs)	Schizophreniform (46 yrs) ?Psychosis secondary to physical brain condition (calcification basal ganglia) Major depressive disorder (28+33 yrs)	Schizoaffective disorder (bipolar) (36 yrs) Bipolar disorder (26 yrs)	Schizoaffective disorder(bipolar) (24 yrs) Bipolar disorder (16 yrs)	Schizophrenia (18 yrs)
Age at recruitment	72 years	46 years	44 years	32 years	18 years
Substance abuse	Absent	Absent	Absent	Absent	Absent
Early deviant behaviour (1-10 yr)	Unknown	Absent	Absent	Absent	Social isolation
Early insults	Unknown	Home delivery by father	Absent	Absent	Unknown
Longitudinal course and pattern of severity of illness	Severe and continuous course	Lost to follow up	Moderate deterioration Episodic with inter episode residual symptoms	Moderate deterioration. Episodic with inter episode residual symptoms	Lost to follow up

Family II

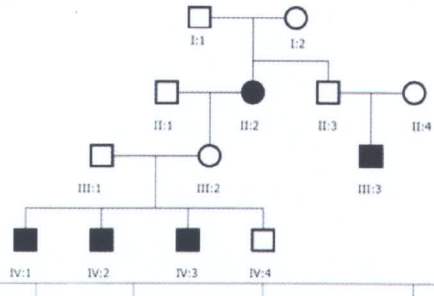


Proband identification	II:1	II:2	II:3
Diagnosis (age of onset) and comorbid diagnosis	Paranoid schizophrenia (23 years)	Paranoid schizophrenia (16 yrs) Major depressive episode (17 yrs) Cannabis abuse (19 yrs)	Paranoid schizophrenia (19 yrs) Obsessive compulsive disorder (15 yrs) Major depressive episode (15 yrs)
Age at recruitment	25 years	21 years	20 years
Substance abuse	Absent	Cannabis	Absent
Early deviant behaviour (1-10 yr)	Absent	Absent	Absent
Early insults	Premature birth 36 weeks gestation – mother had PET. Lung problems after birth	Absent	Immunoglobulin deficiency disease with lung problems
Longitudinal course and pattern of severity of illness	Episodic with inter-episode residual symptoms Episodic shift	Episodic with inter-episode residual symptoms Episodic shift	Episodic with inter-episode residual symptoms Episodic shift

© University of Pretoria

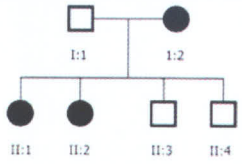
UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Family III



Proband identification	IV:1	IV:2	IV:3	III:3
Diagnosis (age of onset) comorbid diagnosis	Paranoid schizophrenia (15 years)	Schizoaffective disorder (bipolar) (19 yrs) Alcohol dependency (15 – 23 yrs) Polisubstance + cannabis dependency (17 – 23 years)	Schizoaffective disorder(bipolar) (22 yrs) Polisubstance dependency and psychosis (20-21 yrs)	Schizoaffective disorder (14 yrs)
Age at recruitment	32 years	28 years	26 years	58 years
Substance abuse	Absent	Alcohol, polisubstance and cannabis	Cannabis (18 – 20 yrs) LSD (20 – 21 yrs)	Absent
Early deviant behaviour (1-10 yr)	Shy, submissive and asocial	Shy, loner and extreme fear of the dark	Absent	Learning disability
Early insults	Forceps delivery Prolonged labour	Allergy during small pox inoculation	Absent	Absent
Longitudinal course and pattern of severity of illness	Moderate to severe deterioration with continuous course	Mild deterioration Episodic with inter-episodic residual symptoms	Episodic shift with inter- episode residual symptoms	Severe deterioration and continuous course

Family IV



Proband identification	I:2	II:1	II:2
Diagnosis (age of onset) and comorbid diagnosis	Schizophrenia (55yrs)	Schizophrenia (21yrs)	Schizophrenia (18yrs)
Age at recruitment	58 years	38 years	36 years
Substance abuse	Absent	Absent	Absent
Early deviant behaviour (1-10 yr)	Absent	Absent	Absent
Early insults	Absent	Absent	Absent
Longitudinal course and pattern of severity of illness	Severe deteriorating course, continuous	Severe deteriorating course, continuous	Severe deteriorating course, continuous

Discussion

Psychiatric diagnosis and age at onset of illness

In families II, IV and VI, diagnoses of schizophrenia were made in all of the probands. The diagnoses that were made in the other families were schizoaffective disorder and, in isolated affected probands, diagnoses of schizophrenia. There is evidence of familial co-aggregation between schizoaffective disorder and schizophrenia but no familial relationship was generally found between schizophrenia and bipolar disorder.

The diagnosis of schizoaffective disorder in this study was made according to the DSM IV criteria.[6] The diagnostic criteria leave much to interpretation. Determining the exact length of each mood episode is not always easy or even possible. The term "substantial portion" referring to the time period of the mood episode in the total duration of the active and residual period of the illness, is not defined.

A reasonable conclusion from the available data is that patients suffering from schizoaffective disorder make out a heterogeneous group: some have schizophrenia with prominent mood symptoms, others have a mood disorder with prominent schizophrenic symptoms and still others have a distinct clinical syndrome. The hypothesis that schizoaffective disorder patients have both schizophrenia and a mood disorder, is untenable, as the calculated co-occurrence of the two disorders is much lower than the incidence of schizoaffective disorder.[8]

In this study of molecular genetics of schizophrenia in Afrikaners, schizoaffective disorder has been included among affected relatives for linkage analysis. When gene loci are eventually identified, decisive evidence may be obtained for the genetic etiology and transmission of schizoaffective disorders. This will allow the determination of whether they are subtypes or genetic interforms of schizophrenia or mood disorders, or perhaps a combination of these possibilities.[9]

In family I, IV, V and VI, where more than one generation had affected probands, the age of onset of illness of the youngest generation probands was at a younger age than that of the older generation. In the case of families V and VI, the younger generation probands with an earlier age of onset of illness, all abused cannabis.

Cannabis abuse

In a genetic study as reported here, the investigators cannot exclude underreporting of drug use because it was assessed by using self reported data in a high percentage of cases. We often see that where cannabis use data is reported in a social background report, data is not always reliable. Cannabis use/abuse was not confirmed with toxicological screening.

In the vulnerability model for schizophrenia an individual might be vulnerable to schizophrenia but not get the disease unless some life event stressor triggers it. The findings of the Swedish Conscripts Longitudinal Study suggest that cannabis may be such a stressor.[10] According to this stress-vulnerability model, individuals differ in their sensitivity to adverse environmental circumstances, and genetically sensitive individual are more likely to develop symptoms when exposed to these environments than others.[11]

In most of these families, where cannabis abuse was a positive factor (family II, V and VI) the onset of psychosis was at a younger age or may have triggered the psychosis as in family III. The earlier the onset of the psychosis the worse the outcome of the illness.

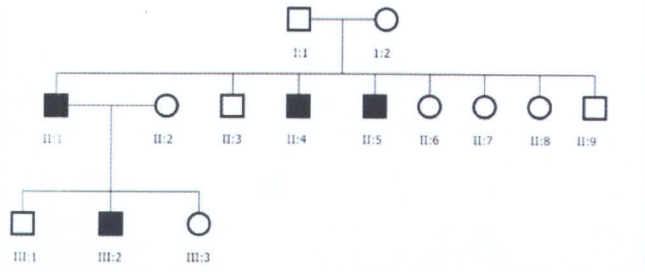
Psychosis outcome is more strongly predicted by a baseline lifetime history of use of cannabis than by recent use. This finding suggests that this association is not fully explained by the short term effects of cannabis leading to acute occurrence of psychotic experience.[12]

Van Os hypothesized that the neurobiological changes in-

© University of Pretoria

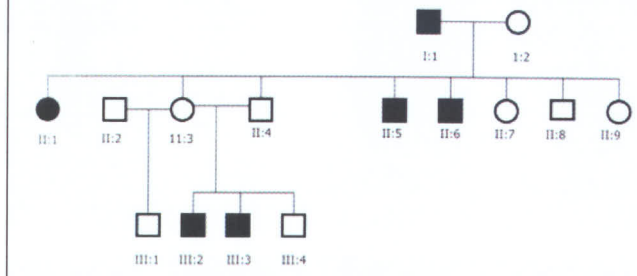
UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Family V



Proband identification	II:1	II:4	II:5	III:2
Diagnosis (age of onset) and comorbid diagnosis	Schizoaffective disorder(bipolar) (36 yrs)	Schizophrenia (40 yrs)	Psychosis secondary porphyria variegata (mania) (40 yrs) ? SAD (bipolar)	Schizoaffective disorder (bipolar) (19 yrs) Major depressive episode (14 yrs) Alcohol + cannabis abuse (20 yrs)
Age at recruitment	58 years	52 years	50 years	23 years
Substance abuse	Absent	Absent	Alcohol dependence	Alcohol Cannabis
Early deviant behaviour (1-10 yr)	Absent	Absent	Absent	Absent
Early insults	Absent	Absent	Absent	Absent
Longitudinal course and pattern of severity of illness	Continuous with severe deterioration	Continuous with severe deterioration	Continuous with severe deterioration	Continuous with severe deterioration

Family VI



Proband identification	II:1	II:5	II:6	III:2	III:3
Diagnosis (age of onset) and comorbid diagnosis	Schizophrenia (18 years)	Schizophrenia (16 yrs) Atypical Major depressive episode (18 yrs)	Schizophrenia (30 yrs) Major depressive episode Panic disorder	Schizophrenia (16 yrs) Major depressive episode (16 yrs)	Schizophrenia (16 yrs) OCD (7-16 yrs)
Age at recruitment	52 years	48 years	46 years	19 years	18 years
Substance abuse	Absent	Absent	Alcohol dependency	Cannabis abuse (from 16 years)	Cannabis abuse (from 13 years)
Early deviant behaviour (1-10 yr)	Social isolation and strained interpersonal relationships	No information	No information	Concentration and attention difficulties Daydreamer	OCD symptoms
Early insults	Prolonged difficult labour	Absent	Absent	Umbilical cord around neck (jaundiced at birth) Incubator for 2 days after birth	Absent
Longitudinal course and pattern of severity of illness	Episodic with inter-episode residual symptoms Mild to moderate deterioration	Episodic with inter-episode residual symptoms Mild to moderate deterioration	Episodic with inter-episode residual symptoms Mild to moderate deterioration	Continuous moderate to severe deterioration	Continuous moderate to severe deterioration

duced by tetrahydro-cannabinol may interact with a preexisting vulnerability to dysregulation of the cannabinoid system or to other neurotransmission systems interacting with the cannabinoid system. In accordance with this hypothesis, their findings demonstrate that the impact of cannabis use on psychosis outcome is especially marked in subjects with an established vulnerability to psychosis.[12] These multiple affected families will have such an established genetic vulnerability to psychosis.

The DSM IV exclusion criteria for schizophrenia and schizoaffective disorder states that the disturbance is not due to the direct physiological effects of a substance (e.g. drug of abuse, a medication) or a general medical condition. If one looks at the above mentioned dysregulation of the cannabinoid/neurotransmission systems hypothesis then the use of the DSM IV classification in genetic research becomes problematic and a hierarchical diagnostic system will be more valuable.[5]

Early deviant behaviour (one to ten years of age)

In a previous study, early non-psychotic deviant behaviours, the demographics, syndrome course and symptoms of 109 Afrikaner probands who met criteria for DSM IV schizophrenia or schizoaffective disorder were compared to 109 age- and gender-matched US probands.[13] Consistent with past findings, 68% of Afrikaner probands, as compared to 67% of age- and gender-matched US probands, reported one or more form of early non-psychotic deviance. These behaviours included poor socialization, extreme fears/chronic sadness, and/or attention learning impairment. The early non-psychotic childhood deviance distinguished a distinct subtype of patients and the forms of early deviance manifested were meaningfully linked to later disease outcome.

In families III and VI early deviant non-psychotic behaviours featured very prominently. An notable feature of all the probands who had this behaviour is that their onset of illness was before 20 years of age (ranging from 14 to 19 years). It has been found before that premorbid social impairments are more marked in child and adolescent onset schizophrenia than in other psychoses. There appeared to be developmental continuity from premorbid impairment to negative symptoms in this study.[14] Most of the probands in this discussion who had early non-psychotic deviant behaviour also had a chronic deteriorating course of illness.

Pregnancy and delivery complications

Previous literature has linked abnormalities of pregnancy and complications of delivery with increased risk of adult schizophrenia.[15] Examples of specific factors that have been recorded include delayed fetal growth, infections, poor nutrition, and hypoxic-ischemic damage. Many previous studies are compromised by recall and selection effects. Some cohort studies have failed to confirm this effect.[16] It remains uncertain whether or not obstetric complications are the result of pre-existing fetal abnormalities. A recent study found no support for suggestions that pre-existent fetal abnormalities precede obstetric complications in the histories of individuals with schizophrenia.[17]

Pregnancy and delivery complications as well as early insults were reported in family II, III and VI to a limited extent. Five of the six probands where these complications were reported, were males. Information in this regard was obtained from mothers and/or probands. Information obtained in this way is susceptible to bias. A study on maternal recall bias, obstetric history and schizophrenia suggested that studies that rely on maternal recall alone, are susceptible to bias. The excess of obstetric complications recalled by the mother, could be related to abnormal behaviour in their child rather than maternal illness, family history or psychotic symptoms.[18] Stud-

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

ies have suggested that males exhibit more or different risk markers, more serious symptomatology, as well as poor prognosis, than do females.[19]

The psychopharmacological treatment and longitudinal course of illness in these multiple affected families was discussed in a poster presentation at the CINP Conference June 2004 in Paris.[20]

Conclusion

Recruitment of multiple affected families in psychiatric genetic research remains important. We reported on six multiple affected Afrikaner families where three or more first-degree relatives were affected with schizophrenia or schizoaffective disorders.

The number of multiple affected families reported on are small and no definite conclusions can be drawn from the results.

Although the heritability of schizophrenia is substantial, its estimates do vary, and its contribution to the schizophrenia phenotype cannot be understood without taking in account the importance of environmental factors.[21] Furthermore, disease will tend to cluster in families, not because of a direct genetic effect, but because relatives are more vulnerable to the risk-increasing effect of a prevalent environmental risk factor.

There is evidence of familial co-aggregation between schizoaffective disorder and schizophrenia but no familial relationship was generally found between schizophrenia and bipolar disorder. It is emphasized that a hierarchial diagnostic approach would take all diagnostic aspects of the psychotic equation into account when doing genetic research in schizophrenia.

The exposure to cannabis in these genetic sensitive probands triggered an earlier age of onset of illness with a poor prognosis in some probands.

Developments in molecular genetics need to go hand in hand with progress in the study of environmental factors that interact with genes. Prevention of phenotypes by targeting environmental risk factors that interact with genotypes is both feasible and acceptable.[11]

References

1. *Rutter M. The interplay of nature, nurture, and developmental influences. The challenge ahead for mental health. Arch Gen Psychiatry 2002; 59:996-1000.*
2. *Harrison PJ, Owen MJ. Genes for schizophrenia? Recent findings and their pathophysiological implications. Lancet 2003; 361.*
3. *Tienari P, Wynne LC, Moring J, Lhti I, Naarala M, Sorri A, et al. The Finnish Adoptive Family Study of Schizophrenia : implications for family research. Br J Psychiatry 1997; 164:20-26.*
4. *Karayiorgou M, Torrington M, Abecasis GR, Pretorius HW, Robertson B, Kaliski S, et al. Phenotypic characterization and genealogical tracing in an Afrikaner schizophrenia database. Am J Med Gen Part B (Neuropsychiatric Genetics) 2004; 124B:20-28.*
5. *Cardno AG, Rijsdijk FV, Sham PC, Murray RM, McGuffin PMB. A Twin study of genetic relationships between psychotic symptoms. Am J Psychiatry 2002; 159(4):539-545.*
6. *American Psychiatric Association. Diagnostic and statistical manual. 4th Edition. Washington, DC: American Psychiatric Association, 1994.*
7. *Nurnberger JI, Blehar MC, Kaufmann CA, York-Cooler C, Simpson SG, Harkavy-Friedman J, et al. NIMH Genetics Initiative 1994. Diagnostic interview for genetic studies, rationale, unique features, and training. Arch Gen Psychiatry 1994;51:949.*
8. *Lauriello J, Erickson BR, Keith SJ. Schizoaffective disorder, schizophreniform disorder, and brief psychotic disorder. In: Sadock BJ, Sadock VA, eds. Kaplan & Sadock's Comprehensive Textbook of Psychiatry 7th ed, Vol 1. Baltimore: Lippincott Williams and Wilkens;2000:1232.*
9. *Bertelson A, Gottesman I I: Schizoaffective psychoses: Genetical clues to classification. Am J Med Genet 1995;60:7-11.*
10. *Andreasson S, Allebeck P, Engstrom A, Rydberg U. Cannabis and schizophrenia. A longitudinal study of Swedish conscripts. Lancet 1987; 2(8574):1483-1486.*
11. *Van Os J, Marcelis M. The ecogenetics of schizophrenia: a review. Schiz Res 1998; 32: 127-135.*
12. *Van Os J, Bak M, Hanssen M, Bijl R.V., de Graaf R, Verdoux H. Cannabis use and psychosis: A longitudinal population-based study. Am J Epidemial 2002; 156, (4): 319-327.*
13. *Sobin C, Roos JL, Pretorius HW, Lundy LS, Karayiorgou M. A comparison study of early non-psychotic deviant behaviour in Afrikaner and US patients with schizophrenia or schizoaffective disorder. Psychiatric Res 2003; 117:113-125.*
14. *Hollis C. Developmental precursors of child- and adolescent-onset schizophrenia and affective psychoses: diagnostic specificity and continuity with symptom dimensions. Br J Psychiatry 2003; 182:37-44.*
15. *Isohanni M, Jones P, Kemppainen L, Croudace T, Isohanni I, Veijola J, et al. Childhood and adolescent predictors of schizophrenia in the Northern Finland 1966 Birth Cohort - a descriptive life-span model. Eur Arch Psychiatry Clin Neurosci 2000; 250:311-319.*
16. *Geddes JR, Lawrie SM. Obstetric complications and schizophrenia: a meta-analysis. Br J Psychiatry 1995; 167:786-793.*
17. *McNeil TF, Cantor-Graae E. Does preexisting abnormality cause labor-delivery complications in fetuses who will develop schizophrenia? Schiz Bull 1999; 25:425-435.*
18. *McIntosh AM, Holmes S, Gleeson S, Burns JK, Hodges AK, Byrne MM, et al. Maternal recall bias, obstetric history and schizophrenia. Brit J Psychiatry 2001; 181:520-525.*
19. *Räsänen S, Pakaslahti A, Syvälahti E, Jones P, and Isohanni M. Gender differences in schizophrenia. A review. Nord J Psychiatry 2000; 54:37-45.*
20. *Roos JL, Pretorius HW, Karayiorgou M. Treatment profile and pattern of severity of illness in multiple affected founder families in a Schizophrenia Genetic Study. International J Neuropsychopharmacol 2004; 7(Suppl 1), S391.*
21. *Tsuang MT, Stone WS, Faraone SV. Genes, environment and schizophrenia. Br J Psychiatry 2001; 178(Suppl. 40) 518-524.*

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

ORIGINAL ARTICLE S Afr Psychiatry Rev 2005;8:153-159

Early non-psychotic deviant behaviour as an endophenotypic marker in bipolar disorder, schizo-affective disorder and schizophrenia

Martin C Scholtz[1], **Melissa S Janse van Rensburg**[1], **J Louw Roos**[1], **Herman W Pretorius**[1], **Maria Karayiorgou**[2], **Jonathan B Levin**[3]
Department of Psychiatry, University of Pretoria, South Africa[1], Human Neurogenetic Laboratory, Rockefeller University, New York, U.S.A.[2], Medical Research Council, Pretoria, South Africa.[3]

ABSTRACT

Objective: To determine and compare the incidence of early non-psychotic deviant behaviour (i.e. under the age of ten) in Afrikaner patients with bipolar disorder, schizo-affective disorder and schizophrenia. **Methods:** Patients with bipolar disorder, schizo-affective disorder and schizophrenia were interviewed using a structured questionnaire probing for early deviant childhood behaviour starting before the age of 10 years. Information from close family members was also obtained where possible. Seven areas of possible deviance were probed into: social dysfunction, unprovoked aggression, extreme anxiety, chronic sadness, extreme odd behaviours, attention impairment and learning difficulties. Demographic data included: age, marital status, gender, and years of formal education. The following clinical features were also recorded: age of onset of illness and suicide attempts. **Results:** A total of 74 patients diagnosed with bipolar disorder, 43 patients diagnosed with schizo-affective disorder and 80 patients diagnosed with schizophrenia were interviewed. Early deviant behaviour was statistically more prevalent in schizophrenia (65%) and schizo-affective disorder (60,5%), than in the bipolar group (21,6%). Deviant childhood behaviour was grouped into 3 clusters: social functioning impairment cluster (social isolation, aggression, extreme odd behavior), mood/anxiety cluster (extreme fears, chronic sadness) and a cognitive impairment cluster (attention impairment, learning disability). Bipolar patients showed significantly less social functioning and cognitive impairment compared to patients with schizo-affective disorder and schizophrenia. **Conclusion:** Our findings suggest that early deviant behaviour may be a possible endophenotypic marker in schizophrenia and schizo-affective disorder.

Keywords: Early non-psychotic deviant behaviour, Endophenotype, Bipolar disorder, Schizo-affective disorder, Schizophrenia

Received: 29.11.04
Accepted: 16.04.05

Introduction

Schizophrenia genetic research has advanced greatly with the new tools that are currently available in molecular genetics. Gene identification, however, remains a task of extraordinary difficulty in part because of the heterogenous and imprecise definition of the clinical phenotype. A great deal of heterogeneity exists among the symptoms of individuals sharing a common clinical diagnosis, and many are not readily classifiable using standard criteria. Refinement of phenotypes by using endophenotypes should be extremely valuable and should increase the chances of discovering mechanisms of inheritance as well as susceptible genes. Endophenotypes provide a means for identifying the "downstream" traits or facets of clinical phenotypes, as well as the "upstream" consequences of genes. The endophenotype could mark the path between the

Correspondence:
Dr MC Scholtz
Department of Psychiatry, University of Pretoria
PBag X113, Pretoria 0001, South Africa
email: mcscholtz@absamail.co.za

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

ORIGINAL ARTICLE S Afr Psychiatry Rev 2005;8:153-159

genotype and the behavior of interest, in this case schizophrenia.[1]

Early deviant patterns of behaviour in children who go on to develop schizophrenia have been consistently reported in high-risk, prospective and retrospective behavioural studies of schizophrenia.[2] These aberrant behaviours could constitute valid endophenotypes for the disease indicating early neurodevelopmental abnormalities.

Founder populations, such as the Afrikaner population (Appendix 1), hold tremendous promise for mapping genes for complex traits, as they offer less genetic and environmental heterogeneity and greater potential for genealogical research.[3]

A previous comparison study examined the demographic, syndrome course, symptom and early deviant behavior history of 109 Afrikaner probands who met criteria for DSM schizophrenia and schizo-affective disorder, and compared them to 109 age- and gender-matched US probands. Approximately two thirds of both the Afrikaner probands (68%) and the US probands (67%) reported one or more forms of early non-psychotic deviance, including poor socialization, extreme fears/chronic sadness, and/or attention/learning impairment.[2] The question was asked whether this early deviant behaviour was unique to subjects diagnosed with schizophrenia or schizo-affective disorder or if it also occurred in bipolar disorder subjects.

There have been several studies investigating premorbid intellectual and social functioning in bipolar, schizo-affective and schizophrenia patients. However, the majority of these studies have included the adolescent period in their assessment of premorbid functioning. None of the studies specified deviant behaviour occurring under the age of ten.

One comparison study of schizophrenia (N=536) , schizo-affective disorder (N=31) and non-psychotic bipolar disorder (N=68), demonstrated that impaired premorbid intellectual, language, and behavioural functioning were specific to future schizophrenia patients compared to future non-psychotic bipolar disorders.[4] In a large longitudinal population-based study of 50 087 male subjects over 27 years, lower premorbid IQ score was associated with increased risk for schizophrenia, severe depression, and other non-affective psychoses, but not for bipolar disorder. This finding indicated that at least some aspects of the neurodevelopmental etiology of bipolar disorder may differ from the other disorders that were investigated.[5]

Our aim with this study was to determine and compare the incidence of early non-psychotic deviant behaviour (i.e. under the age of ten) in Afrikaner patients with bipolar disorder, schizo-affective disorder and schizophrenia. We hypothesized that the prevalence of early non-psychotic deviant behaviour was less prevalent in bipolar patients when compared to schizophrenia and schizo-affective disorder patients.

Methods

Approval for conducting this study was obtained from the Research Ethics Committee of the Faculty of Health Sciences of the University of Pretoria and from the Research Ethics Committee of 1-Military Hospital, Thaba Thswane. Afrikaner patients with schizophrenia or schizo-affective disorder who had multiple affected (two or more) family members had been recruited previously for an ongoing genetic project being conducted collaboratively by the Department of Psychiatry, University of Pretoria and the Laboratory for Human Genetics, Rockefeller University. These subjects were randomly recruited via the outpatient and inpatient departments of Weskoppies Hospital, Pretoria and interviewed using the Diagnostic Interview for Genetic Studies (DIGS) Modified for Genetics in Mental Function Initiative Laboratory of Neurogenetics Rockefeller University by two senior psychiatrists with more than 20 years clinical experience each. Afrikaner bipolar patients were also recruited randomly from the outpatient and inpatients departments of Weskoppies Hospital, Pretoria, as well as from the outpatient and inpatient departments of 1-Military Hospital, Thaba Thswane. The interviewers of the bipolar disorder patients were two registrars in Psychiatry who were supervised by one of the senior clinicians mentioned before. All clinicians are staff members of the Department of Psychiatry at the University of Pretoria.

All of the subjects had been diagnosed formally under supervision of an experienced consultant psychiatrist using the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM IV). Their history, collateral history, clinical findings and diagnosis were recorded in their clinical files.

The criteria used for being an Afrikaner were : Afrikaans language, typical Afrikaans surnames of both parents and grandparents on the paternal and maternal side, and geneaological tracings by a geneaologist.[3]

Written informed consent was obtained from all patients taking part in the study. All participants were interviewed using a semi-structured questionnaire probing for early deviant childhood behavior beginning before the age of ten (Appendix 2). The age of ten was used as a cut-off to exclude behaviors that might be attributable to pre-pubertal hormonal changes.[2] Where possible, additional information was obtained from close family members with permission of the participants.

Regarding the mental state at the time of interview, all patients were stable and apsychotic during the interview and were, according to the judgement of the interviewer, a reliable source of relevant information at that time.

The early deviant behaviour questionnaire probed seven areas of possible deviance, including social isolation (avoidance of other children, inability to have friends, isolated play), unprovoked aggression, extreme anxiety, chronic sadness, extreme odd behaviours (unprovoked screaming fits, disorganized or irrational behaviour, inappropriate affect), attention impairment and learning disabilities. Yes/no responses were recorded. In the case of a yes response, examples were obtained and the interviewer determined whether the reported behaviour was a permanent feature or transient, and whether there was a likely environmental trigger for the reported behaviour (e.g. marital discord,

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

parental abuse or death of a loved one). Behaviours were considered present only if the behaviour was both of a more permanent nature and severe. If an environmental precipitant or other social circumstances could explain the deviant behaviour, it was coded as absent. The deviant behaviour had to be present before ten years of age. Learning disabilities and/or attention impairment were considered present only if the child received a formal diagnosis, was provided with specific remediation or if either of these were reported by teachers or noted in school reports.

Demographic data (age at interview, marital status, gender, number of years of education) and certain clinical features (age of onset of illness, suicide attempts) were recorded for every participant. Suicide attempts were only recorded as present or not present and were not quantified.

Data was entered and analyzed using the statistical program Stata (Release 8.0). Sociodemographic data and syndrome features were compared between the three diagnostic groups : bipolar disorder, schizo-affective disorder and schizophrenia. Continuous variables (age, age of onset of illness and years of education) were compared using a one way analysis of variance. Categorical variables (gender, marital status and suicide attempts) were compared using chi squared tests for contingency tables. Suicide attempts were explored further using the logistic regression model adjusting for age and gender, and the schizo-affective group were used as the baseline reference group.

Behaviours were compared between the three groups using chi squared tests. The 7 types of deviant behaviour were joined into three clinically logical clusters: social functioning impairment cluster (included social isolation, aggression and extreme odd behaviour); mood/anxiety cluster (included extreme fears and chronic sadness); and cognitive impairment cluster (included attention impairment and learning disability). The number of deviant behaviours exhibited in each cluster was compared between the 3 groups using chi squared tests. The relationship of the number of deviant behaviours to disease type was further investigated using ordinal logistic regression models which could also adjust for the effects of age and gender.

Results

The patient groups participating in this study were : patients with bipolar disorder (n=74); patients with schizo-affective disorder (n=43); and patients with schizophrenia (n=80).

Sociodemographic characteristics

The average age at the time of the interview was 44,5 years for the bipolar group, markedly higher than the age of 33,7 years for the schizo-affective group and 31,3 years for the schizophrenia group. The results of marital status were as follows: In the bipolar group 27% of patients were divorced, 43,2 % were married, 24,3% were single and 5,4 % were widowed. In the schizo-affective group 23,3% of patients were divorced, only 16,3% were married, 55,8% were single and 4,6% were widowed. In the schizophrenia group 15% of patients were divorced, only 8,7% were married, 75% were single and 1,3% were widowed. The gender distribution in the bipolar group was 51,3 % male versus 48,6 % female, in the schizo-affective group 58,1 % male and 41,9 % female, while in the schizophrenia group 70% were males and 30% females. The bipolar group had on average 13,5 years of education, the schizo-affective group 12,2 years of education, and the schizophrenia group 11,5 years of education (Table I).

Clinical features

The average age of onset of illness was 28,4 years for the bipolar group, 23,5 years for the schizo-affective group and 22,9 years for the schizophrenia group. Suicide attempts occurred in 47,3 % of bipolar patients, 27,9 % of schizo-affective disorder patients and 12,5 % of schizophrenia patients. When the schizo-affective group was used as a reference group, it was found that the bipolar patients were about 3 times more likely to have attempted suicide than the schizo-affective group. The

Table I: Demographic characteristics of patients with Bipolar disorder, Schizo-affective disorder and Schizophrenia.

Characteristic		*Group*			
		Bipolar disorder n=74	*Schizo-affective disorder n=43*	*Schizophrenia n=80*	*(P-value)*
Age at interview (years)	Mean S.D.	44.5 13.6	33.7 12.1	31.3 11.4	F=22.9 P<0.001
Marital status n (%)	Divorced Married Single Widowed	20 (27%) 32 (43.2%) 18 (24.3%) 4 (5.4%)	10 (23.3%) 7 (16.3%) 24 (55.8%) 2 (4.6%)	12 (15%) 7 (8.7%) 60 (75%) 1 (1.3%)	x2=44.8 P<0.001
Gender n (%)	Male Female	38 (51.3%) 36 (48.6%)	25 (58.1%) 18 (41.9%)	56 (70%) 24 (30%)	x2=5.7 P=0.058
Numbers of years of education	Mean S.D	13.5 2.6	12.2 2.5	11.5 2.8	F=11.14 P<0.001

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Table II: Clinical features and early deviant childhood behaviour of patients with Bipolar disorder, Schizo-affective disorder and Schizophrenia

Clinical Features		*Group*			
		Bipolar disorder n=74	*Schizo-affective disorder n=43*	*Schizophrenia n=80*	*P-value*
Selected clinical features					
Age of onset of illness	Mean S.D.	28.4 14.3	23.5 6.3	22.9 7.6	F=5.97 P=0.003
Suicide attempts	No attempts Attempts	39 (52.7%) 35 (47.3%)	31 (72%) 12 (27.9%)	70 (87.5%) 10 (12.5%)	x2=22.7 P<0.001

Table III: Clusters of deviant childhood behaviour < age 10

Clinical Features		*Group*			
		Bipolar disorder n=74	*Schizo-affective disorder n=43*	*Schizophrenia n=80*	*P-value*
One or more deviant behaviour	n(%) P<0.001	16 (21.6%)	26 (60.5%)	52 (65%)	x2=32.6
Social functioning impairment cluster (Social withdrawal/ aggression/ odd behaviour)	n(%) P<0.001	6 (8.1%)	14 (32.6%)	31 (38.7%)	x2=20.1
Mood/Anxiety cluster (Extreme fears/ chronic sadness)	n(%) P=0.195	8 (10.8%)	10 (23.3%)	12 (15%)	x2=3.3
Cognitive impairment cluster (Attention impairment/ learning disability)	n(%) P<0.001	12 (16.2%)	20 (46.5%)	41 (51.2%)	x2=22.3

schizophrenia group was less likely to have attempted suicide than the schizo-affective group (Table II).

One or more forms of deviant behaviour were present in 21,6 % of the bipolar group, 60,5 % of the schizo-affective group and 65 % of the schizophrenia group (Table III, Figure 1).

Specifically, the distribution of the 7 different categories of deviant behaviour that we probed for was as follows (Table IV) : Social isolation: 6,8 % in the bipolar

Figure 1: The frequency of deviant behaviour in bipolar disorder, schizo-affective disorder and schizophrenia patients.

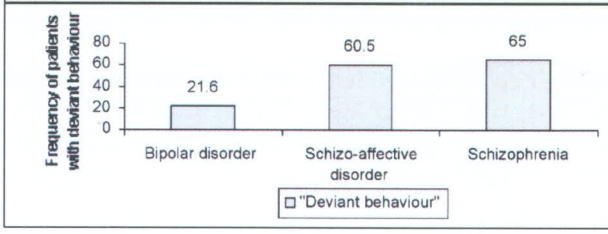


Table IV: Distribution and frequency of individual early deviant behaviours

	Bipolar disorder (n=74)		*Schizo-affective disorder (n=43)*		*Schizophrenia (n=80)*	
	Freq.	*%*	*Freq.*	*%*	*Freq.*	*%*
Type of deviant behaviour						
Social isolation	5	6.8	7	16.3	16	20
Extreme agression	3	4	2	4.6	10	12.5
Odd behaviour	1	1.3	6	13.9	16	20
Extreme fears	7	9.5	7	16.3	7	8.7
Chronic sadness	3	4	8	18.6	6	7.5
Learning disability	4	5.4	17	39.5	24	30
Attention impairment	11	14.9	17	39.5	37	46.2

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

ORIGINAL ARTICLE S Afr Psychiatry Rev 2005;8:153-159

group, 16,3 % in the schizo-affective group and 20 % in the schizophrenia group; Aggression: 4 % in the bipolar group, 4,6 % in the schizo-affective group and 12,5 % in the schizophrenia group; Extreme odd behaviour: 1,3 % in the bipolar group, 13,9 % in the schizo-affective group and 20 % in the schizophrenia group; Extreme fears: 9,5 % in the bipolar group, 16,3 % in the schizo-affective group and 8,7 % in the schizophrenia group; Chronic sadness: 4 % in the bipolar group, 18,6 % in the schizo-affective group and 7,5 % in the schizophrenia group; Learning disability: 5,4 % in the bipolar group, 39,5 % in the schizo-affective group and 30 % in the schizophrenia group; Attention impairment: 14,9 % in the bipolar group, 39,5 % in the schizo-affective group and 46,2 % in the schizophrenia group.

The 7 domains of early deviant behaviour were then grouped into 3 clusters, namely social functioning impairment cluster, mood/anxiety cluster and cognitive impairment cluster. The social functioning impairment cluster included social isolation, aggression and extreme odd behavior. The mood/anxiety cluster included extreme fears and chronic sadness, while the cognitive impairment cluster combined attention impairment and learning disability. While in the bipolar group 8,1 % showed social functioning impairment, in the schizo-affective group this figure was 32,6 % and in the schizophrenia group it was 38,7 %. Mood and anxiety problems were present in 10,8 % of the bipolar group, in 23,3 % of the schizo-affective group and in 15 % of the schizophrenia group. Cognitive dysfunction was reported in 16,2 % of bipolar patients, but in 46,5 % of schizo-affective patients and 51,2 % of schizophrenia patients.

Discussion

Our study comparing 7 domains of early deviant behaviour before the age of ten across 3 different diagnostic entities produced results which strongly support the possible use of early deviant behaviour as an endophenotypic marker in schizophrenia/schizo-affective disorder. The frequency of these behaviours in 65 % of schizophrenic patients and 60 % of schizo-affective patients is statistically significantly higher than the frequency of 21% found in bipolar patients. This result supports our hypothesis that early non-psychotic deviant behaviour is specific to schizophrenia and schizo-affective disorder and therefore may be a valid and useful endophenotype for genetic studies of these disorders. Of note are the findings that bipolar patients were much less likely to show social functioning impairment and also much less likely to show cognitive impairment than the other two diagnostic groups. As mentioned earlier, there have been several previous studies that compared childhood/premorbid intellectual and social functioning in bipolar, schizo-affective and schizophrenia patients. However, none of these studies had excluded the adolescent period. Our study specifically probed for early deviant behaviour occurring before the age of ten across 3 diagnostic groups. Our findings are nevertheless in line with those of some previous studies with regard to childhood premorbid intellectual and behavioral functioning: One study, comparing subjects with schizophrenia (N=536), schizo-affective disorder (N=31) and non-psychotic bipolar disorder (N=68) demonstrated that impaired intellectual, language, and behavioral functioning were specific to future schizophrenia patients relative to future non-psychotic bipolar patients.[4] In a large longitudinal study of 50 087 male subjects, lower premorbid IQ score was associated with increased risk for schizophrenia, severe depression, and other non-affective psychoses, but not bipolar disorder.[5]

Data regarding a history of psychosis was not collected at the initial interview. In retrospect however, psychosis is an important part of the history in the context of this study, since the argument could be built that early deviant behaviour may be related to psychosis per se and not necessarily to schizophrenia or schizo-affective disorder. Going back to the clinical files, we were only able to obtain data about psychosis in 61 of the 74 bipolar disorder subjects who participated in the study. Of these 61 subjects, 39 (63,9%) suffered from bipolar I disorder and 22 (36,0 %) from bipolar II disorder. Of these bipolar subjects (n=61), 22 subjects (36,0%) had a history of psychosis. These 22 subjects with a history of psychosis all suffered from bipolar I disorder and they comprised 56,4% of the bipolar I subjects. With just over one third of bipolar patients having had a history of psychosis, one could argue that the higher prevalence of early non-psychotic deviance found in schizophrenia and schizo-affective disorder, may be related to psychosis per se, and not to the disorders themselves. Future comparison studies with larger samples of bipolar patients, using only bipolar I patients with histories of psychosis, will be needed to clarify this issue.

Our analysis of the demographics of the three patient groups revealed some noteworthy differences. Patients with bipolar disorder were significantly older (more than 10 years) at the time of interview compared to those in the other two groups. This finding may be a reflection of the later onset of illness found in the bipolar group. Three quarters of our schizophrenia patients had remained unmarried during their lifetime, possibly reflecting the devastating effect of schizophrenia on social functioning.

The number of years of education was significantly higher in bipolar than in schizophrenia and schizo-affective patients. Bipolar patients had on average completed approximately one and a half years tertiary education and schizo-affective patients on average less than three months tertiary education. It is, however, important to note that the average schizophrenia patient had not completed grade 12 at school, reflecting schizophrenia's adverse impact on school performance. This finding supports those of previous studies. Patients with schizophrenia have been shown to have poor school adjustment, even when differences in premorbid IQ are taken into account.[6] Bipolar patients, on the other hand, have demonstrated relative preservation of premorbid school functioning despite deterioration of social functioning in adolescence[6], and significantly better academic performance than schizophrenia patients in the

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

ORIGINAL ARTICLE S Afr Psychiatry Rev 2005;8:153-159

premorbid period.[7]

The only two clinical features that we investigated were age of onset of illness and the presence or absence of suicide attempts. Our findings suggest that bipolar disorder tends to have its onset at a significantly later age than schizophrenia and schizo-affective disorder. Although this study was based mostly on subjective reporting by the patients themselves, their estimate of the age of onset of illness might still give us an indication of when their symptoms started, or at least at what age their symptoms started to have a significant impact on their lives.

Suicide attempts was significantly more frequent in the bipolar group (47,3%) compared to the schizo-affective group (27,9%) and the schizophrenia group (12,5%). Bipolar subjects were at least three times more likely to attempt suicide than schizophrenia or schizo-affective subjects. These findings are slightly inconsistent with those of a previous comparison study which showed that suicide attempts occurred in 28% of bipolar patients, 42,8% of schizo-affective patients and 27,3% of schizophrenia patients.[8]

Previous studies estimated that 25-50 % of patients with bipolar disorder attempt suicide during their lifetimes.[9] The bipolar subjects in our study fell near the upper end of this range.

Limitations

It could be argued that the higher prevalence of early non-psychotic deviance found in schizophrenia and schizo-affective disorder may be related to psychosis per se, and not the disorders themselves. We did not collect this data at interview. As explained earlier, we suggest that future studies have larger samples of bipolar subjects, all of them with a history of psychosis.

Information regarding early deviant behavior was obtained from the patients themselves. The validity of this method may be questioned. The current mental state of the patients may have influenced the accuracy of their accounts of their early lives. Early school records and interviews with close family members are likely to have improved the accuracy and reliability of information. However, school records were unavailable in the vast majority of cases, and parents and other close family members were seldom available for interviews, especially in the older patients. One can also argue that teachers and even parents were not necessarily aware of the inner experiences of the children, and that subjective accounts of behaviors may be as accurate as objective accounts.

The average age at interview was more than ten years higher in the bipolar group compared to the schizophrenia and schizo-affective groups. One can only speculate about the influence of this difference on our results. One might expect that older patients would have a poorer recall of the details of their early life, but this would be difficult to verify.

Sample sizes were relatively small and we would suggest larger samples for future studies.

The study was limited to one ethnic group and the generalizability of the results is unknown. However, the schizophrenic and schizo-affective patients from this study have been previously compared with schizophrenic and schizo-affective patients from another ethnic group (US) with remarkably similar results.[2]

Our study investigated early deviant behaviour in three major mental illnesses only, namely bipolar disorder, schizo-affective disorder and schizophrenia. Similar comparison studies which include other mental disorders, as well as a control group of healthy subjects, are suggested.

Conclusion

This comparison study supported our hypothesis that early non-psychotic deviant behaviour was less prevalent in bipolar disorder patients when compared to schizophrenia and schizo-affective disorder patients. Early non-psychotic deviant behavior may therefore be a possible endophenotypic marker in schizophrenia and schizo-affective disorder. Furthermore, bipolar patients had significantly less impairment in social and cognitive functioning compared to schizophrenia and schizo-affective patients. In the interpretation of our demographic data, we found bipolar patients to be much older at the time of interview, a finding which may represent the later onset of illness also found in bipolar patients in this study. The high prevalence of unmarried subjects in schizophrenia emphasized the devastating effect of schizophrenia on social functioning. The adverse effect of schizophrenia relative to schizo-affective disorder and bipolar disorder on school functioning was also reflected in our results, with the average schizophrenia patient having never completed school. Bipolar patients in our study were significantly likelier to attempt suicide in their lifetimes than those with schizophrenia and schizo-affective disorder. We suggest that future studies have larger sample sizes, include other mental disorders and healthy control groups, and make use of school records and family interviews in addition to subjective patient reports.

References

1. *Gottesman I I, Gould TD. The Endophenotype concept in psychiatry : Etymology and strategic intentions. Am J Psychiatry 2003; 160:636-645.*
2. *Sobin C, Roos JL, Pretorius HW, Lundy LS, Karayiorgou M. A comparison study of early non-psychotic deviant behavior in Afrikaner and US patients with schizophrenia or schizo-affective disorder. Psychiatry Res 2003; 117:113-125.*
3. *Karayiorgou M, Torrington M, Abecasis GR, Pretorius HW, Robertson B, Kaliski S et al. Phenotypic characterization and genealogical tracing in an Afrikaner schizophrenia database. Am J Med Genet 2004; 124B:20-28.*
4. *Reichenberg A, Weiser M, Rabinowitz J, Caspi A, Schmeidler J, Mark M et al. A population-based cohort study of premorbid intellectual, language, and behavioral functioning in patients with schizophrenia, schizoaffective disorder, and nonpsychotic bipolar disorder. Am J Psychiatry 2002; 159:2027-2035.*
5. *Zammit S, Allebeck P, David AS, Dalman C, Hemingsson T,*

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Lundberg I et al. A longitudinal study of premorbid IQ score and risk of developing schizophrenia, bipolar disorder, severe depression, and other nonaffective psychoses. Arch Gen Psychiatry 2004; 61:354-360.

6. *Cannon M, Jones P, Gilvarry C, Rifkin L, McKenzie K, Foerster A, Murray RM. Premorbid social functioning in schizophrenia and bipolar disorder: Similarities and differences. Am J Psychiatry 1997; 154:1544-1550.*
7. *McClellan J, Breiger D, McCurry C, Hlastala SA. Premorbid functioning in early-onset psychotic disorders. J Am Acad Child Adolesc Psychiatry 2003; 42(6):666-672.*
8. *Radomsky ED, Haas GL, Mann JJ, Sweeny JA. Suicidal behaviour in patients with Schizophrenia and other psychotic disorders. Am J Psychiatry 1999; 156(10):1590-1595.*
9. *Lopez P, Mosquera F de LJ, Gutierrez M, Ezcurra J, Ramirez F, Gonzalez-Pinto A. Suicide attempts in bipolar patients. J Clin Psychiatry 2001; 62(12):963-966.*

APPENDIX 1 :

Definition of an Afrikaner as used in the study : Early non-psychotic deviant behavior as an endophenotypic marker in bipolar disorder, schizo-affective disorder and schizophrenia.

The Afrikaner population owe their origin to the first European Settlement led by the Dutchman Jan van Riebeeck in 1652. These people were servants of the Dutch East India Company and were Protestants. Afrikaners also originate from settlers who arrived in later years : nine free burghers of Dutch and German origins in 1657 (pioneer farmers of which only two remained) and less than 200 French Huguenots who arrived during the years from 1688 to 1689.

The language spoken by this group is Afrikaans. The slow emergence of Afrikaans as a cultural language developed late in the 19th century. Afrikaner terminology was used for the first time in the 17th century to differentiate between children who, born from parents of European origin, were born in Africa as opposed to those who were born in Europe.

Lists are available of the most important Dutch, German and immigrant surnames from other countries. Important French Huguenot surnames are also listed.

Probands in our study met the following criteria for an Afrikaner :

1. Afrikaans language
2. Typical surnames of both parents and grandparents (paternal and maternal)
3. Genealogical tracings by a geneaologist.[1]

REFERENCES :

1. Karayiorgou M et al. Phenotypic Characterization and Genealogical Tracing in an Afrikaner Schizophrenia Database. Am J Med Genet 2004; 124B:20-28.
2. Walker EA. A history of South Africa. Longman Group Ltd. London 1967.
3. Rosenthal E. Encyclopaedia of South Africa. Frederick Warne and Co Ltd. London 1964.

APPENDIX 2: Questionnaire used to probe for early non-psychotic deviant behavior:

1. How did you get along with other children? Did you have friends? Did you enjoy playing with other children? (IF SUBJECT INDICATES ISOLATION OR PROBLEMS). What problems did you have? When (at what age) did these difficulties begin?
2. Did you have a lot of fears before the age of 10? Do you think you had more fears than other children? (IF YES) Why do you say that? What were you afraid of? (IF FEARS ARE REPORTED) When (at what age) did these fears begin?
3. Did you have any problems with attention or daydreaming before the age of 10? (IF YES) Could you describe the problems? Did anyone else notice and comment on this? Did you have problems in school because of attention? When (at what age) did you or others notice the problems?
4. As a child before age 10, were you ever diagnosed with a learning disability or did you have trouble completing schoolwork? (IF YES) What kind of problems did you have e.g. reading delay, speech impediment, and poor concentration? Were you ever placed in a special class or were you ever given special tutoring to help with your schoolwork? (IF YES) Could you describe the kind of help you received? When (at what age) did these problems begin?
5. Were you aggressive as a child, compared to others? (IF YES) Did you have any problems related to this? Could you describe the problems to me? Did anyone ever comment on this? (IF YES) When (at what age) did these problems begin?
6. Did you ever go through periods of extreme sadness? (IF YES) More so than other children? Could you describe this? When (at what age) did this begin?
7. As a child before age 10, did you do things that you or others thought were odd or unusual? (IF YES) What were they? When (at what age) did this begin?

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

ORIGINAL S Afr Psychiatry Rev 2006;9:99-103

Cannabis and other variables affecting age at onset in a schizophrenia founder population

JL Roos[1], HW Pretorius[2], M Karayiorgou[3], H Boraine[3]
[1]Department of Psychiatry, University of Pretoria, Pretoria, South Africa
[2]Human Neurogenetics Laboratory, Rockefeller University, New York, USA
[3]Department of Statistics, University of Pretoria, Pretoria, South Africa

Abstract
Objective: An ongoing collaborative study between the Human Neurogenetic Laboratory of Rockefeller University, New York and the Department of Psychiatry at the University of Pretoria, has been taking place since 1997, to map genes for schizophrenia. Aspects of cannabis use/abuse will be reported. **Method:** The Diagnostic Interview for Genetic Studies [DIGS] is used for all diagnosis. A narrative report is completed. From this database we determined: gender, diagnostic subtype, age of illness onset, early insults, early deviant behaviour, cannabis use/abuse and family history of schizophrenia. **Results:** From 341 subjects we found: 61% [n=209] males and 39% [n=132] females. In males, 75% [n=156] had schizophrenia and 25% [n=53] schizoaffective disorder[SADJ]. In females 74% [n=97] had schizophrenia and 26% [n=35] SAD . 35% of subjects, [n=118], 44% [n=91] males and 20% [n=27] female's used/abused cannabis. The mean age at onset of illness for the males was 20.5 years. This is significantly earlier than that of males with no cannabis use/abuse group as well as for both these groups in females. The long-term course of illness was similar in both cannabis use/abuse and non-use/abuse groups. An analysis of variance was used to determine the contribution of different factors in determining the age of criteria onset. According to the results obtained, early deviant behaviour was the most important determining age of criteria onset. The group with the lowest mean age of criteria were male users of cannabis with severe early deviant behaviour [18,4 years]. **Conclusion:** Cannabis use/abuse is common amongst male schizophrenia subjects, and affects age at onset of illness. Early deviant behaviour in the first ten years of life in these subjects is more important in this regard, and may be seen as a possible endophenotypical marker.

Keywords: Cannabis; Psychosis; Schizophrenia; Founder population

Received: 10.05.05
Accepted: 13.07.05

Introduction

Community surveys reveal that cannabis is the illicit drug of choice in South Africa. Lifetime rates of 32% in males and 13,1% in females were reported for cannabis use in a representative sample of grade 11 students in Cape Town. Cannabis has the second most common lifetime prevalence rate after alcohol.[1] The predominance of cannabis abuse in South Africa is also reflected in statistics gathered from specialist substance treatment centres. Between July and December 2003, 20% to 40% of patients in various substance treatment centres scattered across South Africa reported cannabis to be their primary drug of abuse. Their average age was 20 to 21 years.[2]

People who suffer from psychotic illnesses are far more likely to consume cannabis than the general population. The question arises as to whether this is because the psychotic subjects use cannabis to alleviate their symptoms or whether cannabis use is a risk factor in precipitating their condition.

The nature of this link is not yet fully understood. The findings drawn from clinical samples of subjects identified with schizophrenia have limited value in shedding light on the mechanism underlying this association. In these cases the potential confounding factors linked to the clinical status of the subjects are difficult to control.[3]

However, data from clinical samples of subjects suffering from schizophrenia may shed some light on how cannabis use/abuse may affect factors such as age at onset of illness and the

Correspondence:
Professor JL Roos, Department of Psychiatry, University of Pretoria, PO Box 667, Pretoria, 0001,South Africa
email: avgreun@postillion.up.ac.za

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

ORIGINAL S Afr Psychiatry Rev 2006;9:99-103

longitudinal course of the illness. Furthermore, the interaction of cannabis use/abuse with factors such as gender, early deviant behaviour in the first ten years of life, family history of schizophrenia in first and second-degree relatives and early insults (prenatal, perinatal and postnatal) can be investigated.

We report on some of this information as collected in a genetic database of Afrikaners with schizophrenia.

Method

A collaborative study on the genetics of schizophrenia in an Afrikaner founder population has been conducted since 1997. Subjects are recruited by the Department of Psychiatry at the University of Pretoria, South Africa and genetic analyses are undertaken by the Human Neurogenetics Laboratory at Rockefeller University in New York. The findings of this research have thus far has been published elsewhere.[4,5] The clinical database of this ongoing study contains information obtained from the Diagnostic Interview for Genetic Studies (DIGS), a chronological summary report of each subject and other sources of information.[4]

This database provided the following information for 341 subjects included in the study thus far:

- Subject identification number and initials
- Gender
- Age at the time of initial recruitment
- Age at which the subject meets the criteria for the onset of schizophrenia
- Previous hospitalization
- Longitudinal course of illness, including:
  - Episodic course of illness with interepisodic residual symptoms
  - Episodic course of illness with no interepisodic residual symptoms
  - Continuous course of illness
  - Single episode psychosis with partial remission
  - Single episode of psychosis with full remission
- Cannabis use:
  - Cannabis use [< 21 times a year]
  - Cannabis heavy use [> 21 times a year]
  - Use/abuse prior to onset of schizophrenia
- Schizophrenia diagnostic entities according to the DSM-IV, including:
  - Schizophrenia
  - Schizoaffective disorder (depressed type)
  - Schizoaffective disorder (bipolar type)
- Early deviant behaviour in the first 10 years of life(poor socialization, extreme fears/chronic sadness, and/or attention/learning impairment). – absent, poorly present, strongly present
- Family history of schizophrenia – absent, first-degree relatives, second-degree relatives
- Early insults (birth and perinatal)

The above data were coded and prepared on a spreadsheet. The statistical methods of analysis included basic descriptive statistics to summarize the data, chi-squared tests of independence between different factors and an analysis of the subject meeting the criteria for the onset of the disease/ schizophrenia.

Ethical approval of this study was obtained from the Ethical Review Board at the University of Pretoria, South Africa and the Rockefeller University, New York, USA.

Results

The mean age at the time of the DIGS interview was 34,09 years for all subjects. The mean criteria onset age was 23,5 years and the mean age of onset of symptoms was 21,5 years. There were 209 males (61%) and 132 females (39%) in the total study group. Of the total sample 60,12% (n = 205) of the total sample did not use/abuse cannabis, whereas 34,6% (n = 118) did use/abuse cannabis. In 5,28% (n = 18) of subjects this information was unknown. Twenty seven percent (n = 91) of the subjects reported use of cannabis prior to onset of illness. Eighty patients (24%) reported current use or heavy use as well as prior use. Seven patients (2%) have once used cannabis, but are not using it any more. The age of first use of cannabis was not reported in this study. Early deviant behaviour was reported by 60,71% (n = 207) of all patients, and was partly present (mild) in 29,33% (n = 100) of cases and strongly present (severe) in 31,38% (n = 107) of cases. The study group consisted primary patients (n=260) and family members (n = 81) who had also been diagnosed with schizophrenia. Of all the primary patients 25,38% (n=66) had a first-degree and 13,08% (n=34) a second-degree relative with schizophrenia.

The total sample comprised 74,19% (n=253), 11.14% (n=38) and 14,66% (n=50) of the three diagnostic groups, namely DSM-IV schizophrenia, SAD (depressed type) and SAD (bipolar type) respectively [SAD - Schizo Affective Disorder].

Early insults were reported by 35,2% (n = 120) of all patients, 51,9% (n = 177) reported that they did not have any early insults and in 12,9% (n = 44) of cases no information was available.

The number of previous hospitalizations ranged between 0 and 20. A percentage of 11,48 of patients did not have any previous hospitalizations while 70,7% (n = 234) of patients had been hospitalized three times or less.

A summary of the longitudinal course of the illness for the 281 patients for whom the information is known is given in Table I.

Table I: Longitudinal course of illness for 281 patients

Longitudinal course	*Percentage*
Episodic course of illness with interepisodic residual symptoms	9,6% (n = 27)
Episodic course of illness with no interepisodic residual symptoms	21,4% (n = 60)
Continuous course of illness	42,4% (n = 119)
Single episode psychosis with partial remission	23,1% (n = 65)
Single episode of psychosis with full remission	3,6% (n = 10)

In the total sample of the males, 43,54% (n = 91) used/abused cannabis while the corresponding figure for females was only 20,45% (n = 27). In a small percentage of cases the information was not known. A statistically significant larger proportion of males than females used/abused cannabis (p = < 0,0001). This trend was also discernible in the group with heavy cannabis use. A statistically significant larger proportion of males (30,14%, n = 63) than females (8,33%, n = 16) reported heavy use (more than 21 times per year) (p < 0,0001).

Although the study group revealed no significant relation between cannabis use/abuse and the number of previous hospitalizations, it was found that of those subjects who did not

© University of Pretoria

use/abuse cannabis 14,07% (n = 28) had never been hospitalized while the corresponding figure for the cannabis use/abuse group was 8,77% (n = 10).

Table II gives a comparison of the longitudinal course of illness for the cannabis use/abuse and non-use/abuse group. The percentages are similar for the two groups, although a slightly higher percentage of cannabis users (45%) were classified as having had a continuous course of illness than non-users (38%) and there was a slightly lower percentage of cannabis users with a single episode of psychosis with full remission (1%) versus non-users (5,4%). These differences were not significant.

Table II: Longitudinal course of illness for no cannabis vs. cannabis

Longitudinal course	*Cannabis group*	
	Non-use/abuse group	*Use/abuse group*
Episodic course of illness with interepisodic residual symptoms	10,24% (n = 17)	9% (n = 9)
Episodic course of illness with no interepisodic residual symptoms	21,69% (n = 36)	22% (n = 22)
Continuous course of illness	37,95% (n = 63)	45% (n = 45)
Single episode psychosis with partial remission	24,7% (n = 41)	23% (n = 23)
Single episode of psychosis with full remission	5,42% (n = 9)	1% (n = 1)

An analysis of variance was used to determine the contribution of different factors in determining the age of criteria onset. Factors considered were cannabis use, gender, family history of schizophrenia, early deviant behaviour, the occurrence of early insults and interactions between cannabis use on the one hand and gender, family history, early deviant behaviour and early insults on the other. A gender early deviant behaviour interaction was also included in the model. We included various factors in the model in view of the complex correlation structure of the different factors. A total of 256 patients' data was used for the analysis of variance. Only one representative per family (the primary patient) was used and subjects with incomplete data were excluded. Least squares means are reported throughout, since they are the best estimates of the true marginal population means when the design is unbalanced (unequal number of respondents per subgroup) and when there are interactive terms in the model.

The estimated mean age of cannabis users (22,7 years) was lower than that of non-users (24,4 years; p = 0,1191). The age of criteria onset was significantly different for males and females in this study. The mean age of criteria onset for males was estimated to be 21,62 years while that for females was 24,7 years (p = 0,0208). The interactive effect of gender and cannabis use was also significant in explaining the age of criteria onset (p = 0,0197). In Table III it can be seen that the effect of cannabis use on age of criteria onset was not the same for the two genders. The mean age of criteria onset of males who did not use cannabis (24,6 years) was significantly older (p=0.0124) than that of males who did (20,4 years), while the mean ages of females who used and did not use cannabis were similar (25,0 years and 24,3 years respectively). The mean age of criteria onset of male users was significantly younger than that of both female users (p = 0,025) and female non-users (p = 0,0317).

Table III: Mean age of criteria onset of gender and Cannabis groups

Gender	*Cannabis group*	
	Non-use/abuse group	*Use/abuse group*
Male	24,6 years	20,4 years
Female	24,3 years	25,0 years

According to the results obtained, early deviant behaviour was the most important factor determining age of criteria onset (p = 0,0037). The mean age of the group with no early deviant behaviour (25,7 years) was significantly older than that of both those with less severe (22,9 years; p = 0,0219) and more severe (22,1 years; p = 0,0024) early deviant behaviour.

Although the interactive effect of substance use and early deviant behaviour was not significant in the model, there was a difference in the effect of early deviant behaviour between cannabis users and non-users. In the group using cannabis there was no significant difference in age of criteria onset between those without and with early deviant behaviour, while in the non-user group the age of criteria onset of those with no early deviant behaviour (27 years) was significantly older than that of those with severe early deviant behaviour (22,8 years; p = 0,0169). It can be seen in Table IV that the mean age of criteria onset decreases with the severity of the deviant behaviour in both groups, as expected, and that the non-users' mean age of criteria onset was consistently older than that of the cannabis users in all three behavioural groups.

Table IV: Mean age of criteria onset of early deviant behaviour and cannabis groups

Early deviant behaviour	*Cannabis group*	
	Non-use/abuse group	*Use/abuse group*
None	27,0 years	24,5 years
Mild	23,5 years	22,3 years
Severe	22,8 years	21,3 years

Gender and early deviant behaviour interaction was a significant, albeit marginally so, source of variation of criteria onset age (p = 0,0543). This implies that the effect of early deviant behaviour is not the same for males and females. The mean age of criteria onset of males with no (23,8 years; p = 0,0339) and males with less severe early deviant behaviour (23,5 years; p = 0,0572) was significantly higher than that of males with severe early deviant behaviour (20,1 years; p-values given). The mean age of females with no early deviant behaviour (27,6 years) was significantly higher than that of females with less severe (22,3 years; p = 0,0299) early deviant behaviour, while the difference was not significant between females with no and females with severe (24,1 years; p = 0,1905) early deviant behaviour (Table V).

Males with severe early deviant behaviour comprise the group with the youngest estimated mean age of criteria onset (Table V). The difference is significant at least at the 10% level

Table V: Mean age of criteria onset of early deviant behaviour and gender groups

Early deviant behaviour	*Gender*	
	Male	*Female*
No	23,8 years	27,6 years
Mild	23,5 years	22,3 years
Severe	20,1 years	24,1 years

© University of Pretoria

in all cases except from females with mild early deviant behaviour.

The mean age of males with severe early deviant behaviour (edb) was significantly younger than that of males with no edb (p=0,0339), males with mild edb (p=0,0572), females with no edb (p<0,0001) and females with severe edb (p=0,0934).

Table VI gives the mean age of criteria onset for the three characteristics: gender, cannabis use and early deviant behaviour. There is no statistically significant interaction between the three characteristics, but the results are interesting and consistent with the expectations of the researchers.

Table VI: Mean age of criteria onset of cannabis, early deviant behaviour and gender groups

	Gender			
	Male		*Female*	
	Cannabis group		*Cannabis group*	
Early deviant behaviour	*Non-users*	*Users*	*Non-users*	*Users*
None	26,8 years (n=24)	20,8 years (n=26)	27,2 years (n=33)	28,8 years (n=10)
Mild	25,2 years (n=27)	21,7 years (n=26)	21,8 years (n=23)	22,1 years (n=4)
Severe	21,7 years (n=33)	18,4 years (n=23)	23,9 years (n=20)	23,2 years (n=7)

It can be seen that the group with the lowest mean age of criteria onset were male users of cannabis with severe early deviant behaviour (18,4 years). In males the difference between the mean age of criteria onset of users and non-users was more pronounced for all three behavioural groups than the corresponding differences in females. In the case of females no clear picture emerges, possibly due to the relatively small sample sizes.

Discussion

The cut-off point used in the Diagnostic Interview for Genetic Studies[6] to determine cannabis abuse or dependency is more than 21 times a year. This is an arbitrary figure and the information obtained from subjects is not reliable. Who would be keeping an accurate record of the number of days they were using cannabis? On the other hand, clinicians treating schizophrenia subjects usually have a fair idea of who the subjects are that abuse cannabis. Of the male subjects in this database, 43,54% reported cannabis use/abuse as against 20.45% of females. This figure may be higher because cannabis use/abuse is usually underreported and the information collected from the subjects was not confirmed by urine analysis for cannibinoids. More males than females were heavy users. Looking at the figures in this database one can assume that approximately half of the male schizophrenia subjects and a quarter of the female schizophrenia subjects' used or abused cannabis. These figures are high and make cannabis use/abuse part of the management strategy of these subjects. Cannabis use/abuse must have an effect on factors such as the criteria age onset and the longitudinal course of the illness.

The longitudinal course of illness as classified by the DIGS[6] can be applied only after at least one year has elapsed since the initial onset of the active-phase symptoms. If we refer to Table II of the results, the percentages are similar for the non-cannabis use/abuse group and the cannabis use/abuse group. The exception is the slightly higher percentage of cannabis users (45%) being classified as having a continuous course of illness than non-users (38%) and a slightly lower percentage of cannabis users with a single episode of psychosis with full remission (1%) versus non-users (5,4%). These differences are not significant. A higher percentage (14,1%) of non-cannabis use/abusers were never hospitalized compared to 8,8% of the cannabis use/abuse group. Although these findings are not statistically significant, they do point towards the fact that cannabis use/abuse may have a negative effect on the longitudinal course of the illness.

Information for early non-psychotic deviant behaviour was available for 207 subjects in the database. 60,7% (n = 207) reported one or more forms of early non-psychotic deviance. This was partly present (at least one form of early non-psychotic deviance) in 29,3% (n = 100), and strongly present (2 or more forms of early non-psychotic deviance) in 31,4% (n = 107). In a previous report the early non-psychotic deviant behaviour in Afrikaner and US patients with schizophrenia or schizoaffective disorder were compared and reported on.[7] It was concluded that early non-psychotic childhood deviance in this independently collected Afrikaner population distinguished a distinct sub-type of patients and that the forms of early deviance manifested were meaningfully linked to the later outcome of the disease.[7] The approach was based on retrospective reports of patients. This approach was not without its limitations but seemed useful to further explore patient-reported non-psychotic early deviant childhood behaviours.

In another comparative study of early non-psychotic deviant behaviour in the first ten years of life in Afrikaner patients with schizophrenia, schizoaffective disorder and bipolar disorder, it was found that the prevalence of early non-psychotic deviant behaviour was not significant in Afrikaner patients with bipolar disorder when compared to the other two diagnostic entities. The authors concluded that early non-psychotic deviant behaviour may therefore be used as a possible endophenotypic marker in schizophrenia and schizoaffective disorder, but not in bipolar disorder.[8]

The age at which a subject with schizophrenia first becomes psychotic is a variable trait relating to prognosis. An earlier onset of psychosis is associated with a more severe course, irrespective of the duration of the illness.[9]

The age of criteria onset being significantly different for males and females in this study corresponds to previous findings that the peak ages of onset are 15 to 25 years for men and 25 to 35 years for women.[10]

The results indicating a strong association between the use of cannabis and age at onset of psychosis in male schizophrenia patients match those of a study in Dutch male schizophrenia patients.[9]

Early deviant behaviour emerged as the most important factor in determining age of criteria onset in the present study, with the early deviant behaviour group having the oldest age of onset and the more severe group the youngest.

Although the interactive effect of substance use and early deviant behaviour is not significant in the model, there was a difference in the effect of early deviant behaviour between

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

cannabis users and non-users. In the group using cannabis there was no significant difference in age of criteria onset between those without and with early deviant behaviour, while in the non-user group the age of criteria onset of those with no early deviant behaviour was significantly older than that of those with severe early deviant behaviour. The mean age of criteria onset decreased with the severity of the deviant behaviour in both cannabis groups, with the non-users' mean age of criteria onset being consistently higher by at least one year compared to that of the cannabis users in all three behavioural groups.

Males with severe early deviant behaviour comprised the group with the youngest estimated mean age of criteria onset. These findings would add on to previous findings indicating that early deviance was linked to the later outcome of the disease.[7] In the present study early deviance affected the age of onset of schizophrenia which is known to be associated with a more severe course of illness. These findings are significant in males with mild and severe, and females with severe early deviant behaviour.

Male users of cannabis with severe early deviant behaviour had the lowest mean age of criteria onset, namely 18,4 years. This can be interpreted that in males with a distinct subtype of schizophrenia (where early deviance is present) cannabis triggers the onset of a psychosis with a poor prognosis. These findings emphasize the importance of vulnerable male subjects abstaining from cannabis use.

Whilst previous findings from a study of Finnish families that the age at onset decreased as the family risk of schizophrenia increased[11], in this database family history of schizophrenia was less important in explaining the age of criteria onset than early deviant behaviour; the interaction of early deviant behaviour and male gender; and the interaction of cannabis use/abuse and male gender.

Limitations of study

Information on the use/abuse of cannabis obtained from the subjects may be unreliable, as it was not verified by urine analysis for cannabis. Had the subjects been admitted to the state psychiatric hospital, the investigators could have scrutinized clinical files where this information was available.

Information on early deviant behaviour was obtained from the subjects. This information was verified in a small percentage of subjects, and additional information was obtained from parents, as well as nursery school and early school records. This information may be inaccurate and biased. Information on early insults e.g. pregnancy history, perinatal and postnatal history, was obtained from subjects and parents, if possible. This information was not verified by obstetrician or hospital records.

The mean age at the time of DIGS interview was 34 years. This may have influenced the patients' recall of early deviant behaviour. Patients' current mental state may significantly have affected their report of past experiences. The mean age was verified with collateral information, wherever possible.

Conclusion

Although the figures are not directly comparable, it was found that the percentage of male and female Afrikaner schizophrenia subjects who used/abused cannabis was higher than the lifetime rates for males and females in a representative sample of grade 11 students in SA[1], which corresponds more or less to the figure reported by various substance treatment centres scattered across South Africa, for cannabis as the primary drug of abuse.[2] More emphasis should be placed on the psycho-education and rehabilitation of young vulnerable individuals with schizophrenia who use and abuse cannabis.

Early deviant behaviour was the most important factor determining age of criteria onset and in male patients with schizophrenia may be a possible endophenotypic marker. The interactive effect of gender and cannabis use was also significant in explaining the age of criteria onset in males. Yet it seems that when cannabis is used, the effect of early deviant behaviour becomes less important in determining the age of criteria onset.

In our search for more sophisticated endophenotypes to aid the identification of susceptibility and modifying genes, early deviant behaviour, male gender and cannabis use/abuse in interaction may be factors to consider in future research.

Acknowledgement

The authors thank all of the families involved in the study for their ongoing participation, and are also grateful to Sisters Ria van Wyk, Cheryl Botha and Liezl Bouwer for collecting family data.

References

1. *Parry CD, Myers B, Morojele NK, Flisher AJ, Bhana A, Donson H, Pludderman A. Trends in adolescent alcohol and other drug use: finding from three sentinel sites in South Africa. J Adolesc. 2001 27(4): 429-440.*
2. *South African Community Epidemiology Network on Drug use (SACENDU). Monitoring Alcohol and Drug Abuse Trends in South Africa, SACENDU Research Brief,2004; 7(1) Available at http:// www.sahealthinfo.org/admodule/sacendu.htm.*
3. *Verdoux, H. Cannabis and psychosis proneness. In: Castle D, Murray R eds, Marijuana and Madness. Cambridge: Cambridge University Press, 2004:75.*
4. *Karayiorgou M, Torrington M, Abecasis GR, Pretorius HW, Robertson B, Kaliski S, et al. Phenotypic characterization and genealogical tracing in an Afrikaner schizophrenia database. Am J Med Genetics Part B (Neuropsychiatr Genet) 2004: 124B: 20-28.*
5. *Abecasis GR, Burt RA, Hall D, Bochum S, Doheny KF, Lundy SL, et al. Genomewide scan in families with schizophrenia from the founder population of Afrikaners reveals evidence for linkage and uniparental disomy on chromosome 1. Am J Human Genetics. 2004: 74: 403-417.*
6. *Nurnberger JI, Blehar MC, Kaufmann CA, York-Cooler C, Simpson SG, Harkavy-Friedman J, et al. NIMH genetics initiative. Diagnostic interview for genetic studies, rationale, unique features, and training. Arch Gen Psychiatry 1994; 51: 949.*
7. *Sobin C, Roos JL, Pretorius HW, Lundy LS, Karayiorgou M. A comparison study of early non-psychotic deviant behaviour in Afrikaner and US patients with schizophrenia and schizoaffective disorder. Psychiatry Res 2003:117: 113-125.*
8. *Scholtz MC, Janse van Rensburg M, Roos JL, Pretorius HW, Levin JB. Early non-psychotic deviant behaviour as an endophenotypic marker in bipolar disorder, schizoaffective disorder and schizophrenia. S Afr Psychiatry Rev 2005:153-159.*
9. *Veen ND, Selten JP, Van der Tweel I, Feller WG, Hoek HW, Kahn RS. Cannabis use and age of onset of schizophrenia. Am J Psychiatry 2004:501-506.*
10. *Kaplan HI, Sadock BJ. Synopsis of Psychiatry, 8th edition, Baltimore: Lippincott, Williams and Wilkins 1998:475.*
11. *Suvisaari JM, Haukka J, Tanskanen A. Age at onset and outcome in schizophrenia are related to the degree of familial loading. Br J Psychiatry 1998:173 : 494-500.*

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

ORIGINAL Afr J Psychiatry 2007;10:219-224

Prevalence and clinical characteristics of obsessive-compulsive disorder and obsessive compulsive symptoms in Afrikaner schizophrenia and schizoaffective disorder patients

F Seedat[1], JL Roos[1], HW Pretorius[1], M Karayiorgou[2], B Nel[3]
[1]Department of Psychiatry, Weskoppies Hospital, University of Pretoria, South Africa
[2]Department of Psychiatry, Columbia University Medical Center, New York, USA
[3]Department of Statistics, University of Pretoria, South Africa

Abstract

Objective: There is evidence of variation in the prevalence of co-morbid obsessive-compulsive disorder in schizophrenia amongst ethnic groups. This study evaluated the lifetime prevalence and clinical characteristics of obsessive-compulsive disorder (OCD)/ obsessive-compulsive symptoms (OCS) in Afrikaner schizophrenic and schizoaffective disorder patients. **Method:** An ongoing genetic study of schizophrenia is currently being conducted on the Afrikaner founder population. In this cohort of 400 subjects from the original genetic study, we identified 53 subjects with schizophrenia or schizoaffective disorder and co-morbid OCD/OCS (study group). They were matched for gender and age of onset of illness with 59 subjects who do not have OCD/OCS (control group). The diagnostic instrument used in this cohort is the Diagnostic Interview for Genetic Studies (DIGS) version 2, which has been translated into Afrikaans. In addition to the DIGS, information for the relevant clinical characteristics reported in this study was also drawn from a detailed narrative chronological summary report and clinical files. A checklist was completed. **Results:** The prevalence of co-morbid OCD/OCS amongst 400 subjects with schizophrenia or schizoaffective disorder was 13.2% [n=53] of which 40 were male and 13 female patients. The prevalence of OCD was 10.7% and OCS was 2.5%. Contamination obsessions [n=17] were the most common type of obsession reported, followed by religious obsessions [n=8]. The most prevalent compulsions were repetitive rituals [n=32] followed by checking behaviour [n=22]. Onset of psychotic symptoms was found to be insidious in 86.8% of the study group compared to 24.6% of the control group ($p<0.0001$). Second-generation antipsychotic use was found to be statistically more prevalent in the study group (77.4%), compared to the control group (45.8%) ($p=0.0008$). 73% of the study group experienced depressive symptoms compared to 50.8% of the control group. Both groups were found to have a similar incidence of suicidal thoughts and suicide attempts. Substance abuse amongst the control group was significantly higher (35.9%) compared to the study group (19.2%) ($p <0.05$). Cannabis was most commonly abused in both groups, followed by alcohol. **Conclusion:** The prevalence rate of 13.2% of co-morbid OCD/OCS in Afrikaner schizophrenia and schizoaffective disorder patients differs from findings in other ethnic groups, suggesting the possible role of genetic and cultural factors in the prevalence of co-morbid OCD/OCS. Second-generation antipsychotic use amongst schizophrenia and schizoaffective disorder patients with co-morbid OCD/OCS was found to be significantly higher than in those without co-morbid OCD/OCS. Clinical characteristics of Afrikaner schizophrenics and schizoaffective disorder patients with and without co-morbid OCD/OCS are the same, both groups were associated with significant psychopathology and a poor prognosis.

Key words: Obsessive-compulsive disorder (OCD), Obsessive-compulsive symptoms (OCS), Schizophrenia, Schizoaffective disorder, Afrikaner population

Received: 21-12-2006
Accepted: 26-04-2007

Correspondence:
Dr F Seedat
Department of Psychiatry, University of Pretoria,
PBag X113, Pretoria, 0001, South Africa
email: seedatf@webmail.co.za

Introduction

Obsessive-compulsive disorder (OCD) and obsessive-compulsive symptoms (OCS) occur in a substantial percentage of patients with schizophrenia or schizoaffective disorder.[1,2,3]

Epidemiological and clinical studies estimate the rate of OC

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

phenomena in schizophrenia at between 3-59.2%.[4] This is considerably higher than the rate of 2-3% of OCD/OCS in the general population.[5] Higher than expected co-morbidity rates raise the possibility of a common underlying pathology for both disorders. Recent literature suggests some overlap between schizophrenia and OCD in structural and functional brain abnormalities, and in the dopamine and serotonin neurotransmitter systems.[6,7]

Distinguishing between delusions, obsessions, ruminations and preoccupations can be challenging in patients with thought disorder. Delusions are described most often as ego-syntonic and are actively embraced by the patient, whereas obsessions are typically ego-dystonic and are recognised as pathological intrusions.[8] Apart from difficulties in identifying OC symptoms in schizophrenia and schizoaffective disorder, methodological differences result in varying frequencies of reporting of co-morbid OCD in schizophrenia. Direct patient interviews lead to higher frequencies of reporting whereas chart reviews result in lower frequencies.[7] Despite these difficulties, the majority of contemporary studies reveal a considerable base rate of OC features in schizophrenia patients, supporting the validity of the schizophrenia-OCS/OCD association.[9]

Studies have shown evidence of variation in the prevalence of schizophrenia across ethnic groups.[10] It has also been reported that OCD may be less common in certain communities.[11]

Niehaus et al[12] found the prevalence of co-morbid OCD in a Xhosa-speaking schizophrenia group to be significantly lower compared to that found in Caucasian schizophrenia patients, raising the question as to whether genetic and cultural factors play a role in the development of co-morbid OCD/OCS in different ethnic groups.

Hwang et al[13] found that schizophrenic patients with OCD have a worse clinical course than schizophrenic patients without OCD. Such co-morbidity is also associated with poor treatment response, higher levels of negative symptoms, and greater impairment in levels of functioning.

More recent studies agree with Fenton and McGlashan[2] regarding poor outcome in schizophrenia-OCD patients. This group of patients has a less favourable employment record, lower age-appropriate functioning and more severe impairment of social behaviour.[3,13]

Previous studies[14] have described a history of early deviant behaviour in subjects with schizophrenia. Early deviance has also been shown to affect the age of onset of schizophrenia[15], with an earlier onset being associated with a more severe course of illness.[16] Early non-psychotic deviance included poor socialization, extreme fears/chronic sadness, and/or attention/learning impairment.[14] No reports of the incidence of early deviant non-psychotic behaviour in a sub-group of Afrikaner schizophrenia and schizoaffective disorder patients and it's relation to those with and without co-morbid OCD/OCS were found. This relationship was assessed in our study.

One of the objectives of this study was to evaluate the lifetime prevalence and clinical characteristics of OCD and OCS in Afrikaner schizophrenia and schizoaffective disorder patients and to compare clinical characteristics in schizophrenia and schizoaffective disorder patients with co-morbid OCD/OCS and those without co-morbid OCD/OCS. We hypothesized that the prevalence of OCD/OCS in Afrikaner patients with schizophrenia and schizoaffective disorder would be similar to that described in other studies of Caucasian samples.

Methods

Approval for conducting this study was obtained from the Research Ethics Committee of the Faculty of Health Sciences of the University of Pretoria. Written informed consent was obtained from all subjects participating in the study.

Genetic Study

A collaborative ongoing study on the genetics of schizophrenia in the Afrikaner founder population is being conducted in the Department of Psychiatry, University of Pretoria in collaboration with Dr Karayiorgou (Columbia University, New York).[17]

Four hundred subjects with schizophrenia and schizoaffective disorder have been recruited since November 1997. Subjects were recruited from inpatient and outpatient departments of Weskoppies Hospital, Pretoria. The criteria used for being an Afrikaner were: Afrikaans as a first language; having typical Afrikaans surnames of both parents and grandparents on the paternal and maternal side; and geneaological tracings by a geneaologist.

Study subjects and controls

The original genetic study group consisted of four hundred subjects, two hundred and forty eight males (62%), and one hundred and fifty two females (38%). Diagnoses were assigned according to the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV[18] with schizophrenia or schizoaffective disorder and other co-morbid diagnoses.

Fifty-three of these subjects between the ages of 12 and 59 years were identified to have co-morbid OCD/OCS. This subgroup, comprised of 40 males and 13 females, was matched for age of onset of illness and gender with a subgroup of fifty-nine subjects, comprised of 39 male and 20 female patients who do not have OCD/OCS, from the original genetic group.

Age of onset of illness (schizophrenia) was defined as emergence of first psychotic symptoms, according to subjective reports and information from clinical files and family members. Age of onset of schizophrenia or schizoafffective disorder was used as a criterion for matching the study to control group due to the effect it could have on the course and prognosis of the illness.

Patient Assessment

The Diagnostic Interview for Genetic Studies (DIGS version 2.0[19] was utilised after being translated into Afrikaans for use in the original study.[13] The DIGS is a validated clinical assessment tool designed for diagnosing psychotic spectrum disorders, mood disorders as well as anxiety disorders and includes assessment for the lifetime presence of OCD.

The clinical interviewers administering the DIGS were senior psychiatrists at Weskoppies Hospital in Pretoria. Each interviewer had more than twenty years clinical experience and was trained in the use of the DIGS. A chronological summary report was compiled for each subject.

Additional information, including demographic data, medical and treatment history, family history and collateral information, was obtained by a trained research nursing sister from patients, clinical files and family.

A checklist was compiled and completed from information obtained from the narrative chronological summary reports and the DIGS. Information on checklist included: demographic

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Table 1: Demographic characteristics

Variables		Study Group (n=53)		Control Group (n=59)	
Mean Age at time of DIGS interview (years)		30.6		33.6	
		Frequency (n=)	%	Frequency (n=)	%
Gender	Male	40	75.5	39	66.1
	Female	13	24.5	20	33.9
Marital status	Single	38	71.7	41	69.5
	Married	6	11.3	3	5.1
	Divorced	9	17.0	15	25.4
Highest level of education	Primary	3	5.8	5	8.5
	Secondary	27	51.9	47	79.7
	Tertiary	22	42.3	7	11.8
Employed		17	32.7	13	22.0
Unemployed		35	67.3	46	78.0

data, age at interview, marital status, gender, number of years of education and employment status. Clinical characteristics included age of onset of illness, co-morbid diagnosis, mode of onset of symptoms (acute versus insidious onset), types of obsessions and compulsions, depressive symptoms, presence or absence of suicide attempts, substance abuse, the nature of psychotic symptoms (prominent positive or negative symptoms), pharmacological treatment, number of hospitalisations, and present level of functioning.

The checklist also probed seven areas of early deviant behaviour before age of ten, including social isolation, unprovoked aggression, extreme anxiety, chronic sadness, extreme odd behaviours, attention impairment and learning disabilities.

Statistical analysis

SAS 9.1 and the BMDP Statistical Software package 7.1 were used for statistical analysis. Analysis of descriptive statistics was conducted on the relevant variables within the two groups. Descriptive statistics were computed as mean and frequencies (count and percentages). Chi-squared tests were used to compare and find relationships between the two groups and the other categorical variables. The Student's t – test was used to test for differences between the age of onset of schizophrenia or schizoaffective disorders and the age of onset of OCD/OCS.

A value of $p < 0.05$ was considered significant.

Results

Socio-demographic characteristics

The study group consisted of 53 patients, of whom 40 were males and 13 females diagnosed, according to the DSM-IV criteria, with schizophrenia or schizoaffective disorder with co-morbid OCD or OCS. Within the study group only 2 male patients were siblings. The non-sibship groups consisted of the remaining subjects of both the study (n=51) and control groups (n=59). Mean age at time of DIGS interview for the study group was 30.6 years (range of 12-59 years).

The control group consisted of 59 patients, of which 39 were males and 20 females.

Mean age at time of DIGS interview for the control group was 33.6 years (range of 14-60 years).

Age of onset of OCD/OCS in the study group varied between 7-31 years. Age of onset of schizophrenia and schizoaffective disorder in the study group varied between 11-33 years.

Mean age of onset of OCD/OCS in the study group was found to be earlier (at 18.5 years) than age of onset of schizophrenia or schizoaffective disorder (at 22 years) ($p<0.0001$). [Table 1]

Clinical features

The prevalence of co-morbid OCD/OCS was 13.2% [n=53]. The prevalence of OCD was 10.7% [n=43] and OCS was 2.5% [n=10]. Of the study subjects, 75.5% [n=40] were diagnosed with schizophrenia and 24.5% [n=13] with schizoaffective disorder. Schizoaffective disorder (depressive type) occurred in 69.2% [n=9] of patients, and schizoaffective disorder (bipolar type) in 30.8% [n=4] in the study group. Paranoid schizophrenia was diagnosed most often in both the study group (82.9%) [n=34], and the control group (50%) [n=23], followed by undifferentiated schizophrenia in the study group (14.6%) [n=6], and (26.1%) [n=12] in the control group. Other lifetime co-morbid psychiatric disorders diagnosed in both groups included major depressive disorder, other mood disorders, panic disorder, phobias, anxiety disorder and eating disorders. [Table 2]

Table 2: Other lifetime co-morbid psychiatric disorders in study and control groups

Co-morbid diagnosis	Study Group (n=53)		Control Group (n=59)	
	N	%	N	%
Major Depressive Disorder	15	28.3	3	5.1
Other Mood Disorders	0	0	1	1.7
Panic Disorder	2	3.8	0	0
Phobias	3	5.7	0	0
Anxiety Disorder	5	9.4	0	0
Eating Disorder	1	1. 9	1	1.7

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

The majority of study patients 94.3% [n=50] had both obsessions and compulsions. More than one type of OCS was observed per patient. Obsessions alone were reported in 5.7% [n=3].

Contamination obsessions (32.1%) [n=17] were commonly reported, followed by religious obsessions (15.1%) [n=8]. Compulsions most often found were repetitive rituals (64%) [n=32], followed by checking behaviour (44%) [n=22]. (Table 3).

Table 3: Obsessions and compulsions in study group of patients with schizophrenia or schizoaffective disorder

Content of obsessions (n=53)	(n=)	%	Types of compulsions (n=50)	(n=)	%
Contamination	17	(32)	Repeating rituals	32	(64)
Religious	8	(15)	Checking	22	(44)
Pathological doubt	6	(11.3)	Cleaning-washing	9	(18)
Sexual	4	(7.5)	Counting	8	(16)
Aggressive	3	(5.6)	Symmetry-precision	4	(8)
Somatic	2	(3.7)	Hoarding-collecting	3	(6)
Symmetry	2	(3.7)	Ordering-arranging	3	(6)
Multiple obsessions	29	(54.7)	Need to ask/confess	1	(2)
Other	13	(24.5)	Other	9	(18)

There was no significant difference between the two groups with regard to family history of psychiatric illness ($p>0.05$). No family history of OCD/OCS was found in the two groups.

The onset of psychotic symptoms was found to be insidious in 86.8% of the study group compared to 24.6% in the control group ($p<0.0001$). Positive symptoms were found to be more frequent in both the study group (48.1%) and control group (42.4%) of patients, compared to negative symptoms being experienced by only 3.8% of the study group. Both groups had a similar incidence of combined positive and negative symptoms of which the frequency in the study group was 48.1% and in 57.6% in the control group. Depressive symptoms were experienced by 73% of the study group compared to 50.1% of the control group. Patients in both groups showed a similar incidence of suicide attempts, namely 36% in the study group and 35.6%in the control group.

Substance abuse in the control group was significantly higher (35.6%) as compared to the study group (19.2%) ($p<0.05$). Cannabis was the most commonly abused substance in the control group (41.2%) as well as the study group (20.6%). This was followed by alcohol abuse in both groups of patients (36.4% of the control group compared to 12.1% of the study group.)

Second–generation antipsychotic use was found to be statistically more common in the study group (77.4%), compared to the control group of patients (45.8%) ($p=0.0008$). In the study sample, 77.4% [n=41] of subjects were treated with second-generation antipsychotics, compared to 45.8% [n=27] of subjects in the control group. Conventional antipsychotics were used by 22.6% [n= 12] of the study subjects and 54.2% [n=32] of the control subjects. There was no significant difference between the two groups as far as number of hospitalisations was concerned.

Both groups displayed a similar incidence of early (before the age of 10) non-psychotic deviant behaviour, with the exception of learning disabilities: 32.1% of the study group experienced learning disabilities compared to 54.2% of the control group of patients. ($p<0.01$). [Table 4]

Table 4: Frequency of early non-psychotic deviant behaviour in the study and control groups

Type of deviant behaviour	Study Group (n=53)		Control Group (n=59)	
	Frequency (n=)	%	Frequency	%
Social isolation	21	39.6	28	47.5
Unprovoked aggression	7	13.2	7	11.9
Extreme anxiety	10	18.9	9	15.2
Chronic sadness Extreme	4	7.5	11	18.6
odd behaviour Attention	9	17	19	32.2
impairment Learning	21	39.6	33	55.9
disabilities ($p<0.01$)	17	32.1	32	54.2

Only 5.1% [n=3] of the control group of patients had a high level of global and social functioning whereas none of the study group met this criterion.

Discussion

The prevalence of co-morbid OCD/OCS amongst 400 subjects with schizophrenia and schizoaffective disorder was 13.2%, which is in keeping with the prevalence of co-morbid OCD and schizophrenia estimated at 12.2% by the US National Institute for Mental Health Epidemiologic Catchment Area Study.[5] 10.7% of the study group met the criteria for OCD, whereas 2.5% of patients had OCS that were not severe enough to fulfil the DSM-IV[18] criteria for OCD. Two subjects from the study group were the only siblings included in both study and control groups. Since they formed the only sibship, their inclusion did not significantly affect the prevalence rates of OCD/OCS.

Patients had insight into their OCS and were able to distinguish between psychotic symptoms and OCS at time of interview, thereby avoiding over identification of OCS within our study group. This result supports our hypothesis that the prevalence of OCD/OCS in founder Afrikaner patients with schizophrenia or schizoaffective disorder is significant. The higher prevalence rate in this study is in contrast to the 0.5% co-morbidity of OCD in South African Xhosa-speaking schizophrenia patients.[12]

Dominquez[20] et al also found ethnic differences in their study on a US population group with more OCD-schizophrenic patients in the Hispanic than the Caucasian group with a 32.7% (17/52) prevalence rate of OCS/OCD in schizophrenia.

Weissman et al[11] found variability in symptom presentation of co-morbid OCD across diverse countries suggesting that cultural factors may affect the symptom expression.

Thus, the possibility of genetic and cultural factors interacting with environmental factors could influence the expression of OCD/OCS in patients with schizophrenia or schizoaffective disorder. Further studies are necessary to explore this association. The majority of patients enrolled for the original genetic study were male, with a male to female

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

ratio of 248/152 (62% males compared to 38% females). This trend is also evident in our study sample whereby the male to female ratio was found to be 40/13, (75.5% males compared to 24.5% females) with a preponderance of male patients diagnosed with schizophrenia or schizoaffective disorder and co-morbid OCD/OCS.

This finding is in keeping with results of a Turkish study[21] of schizophrenia subjects with co-morbid OCD, were the male to female ratio was 15/5. The association between male gender and schizophrenia with co-morbid OCD needs to be explored further.

Recently accumulated data by Tibbo and Warneke[22] and Gross-Isseroff et al[23] suggest that schizophrenia and OCD have a similar age-at onset distribution, with a trend towards earlier age of onset for OCD. With regards to the mean age of onset of OCD/OCS in the study group, we found an earlier age of onset at 18.5 years, compared to the age of onset of schizophrenia or schizoaffective disorder at 22 years. This is in keeping with data found in other studies.[22,23,24]

Age of onset of schizophrenia/schizoaffective disorder is usually in the second to third decade of life, whilst for OCD this figure is in the first to second decade. OCD/OCS may be regarded as a feature of the prodromal phase of schizophrenia/schizoaffective disorder.[4] This finding may be an explanation for the earlier age of onset of OCD/OCS in comparison to the age of onset of schizophrenia.

The analysis of the socio-demographics of the patients suggests that both the schizophrenia-OCD/OCS group and schizophrenia group without OCD/OCS have similarly high rates of unemployment and were less often married. Both groups of patients had a less than favourable employment record. These findings are suggestive of the impact that schizophrenia and schizoaffective disorder have on the social functioning of the patients with and without co-morbid OCD/OCS. The impact of these disorders on school functioning is reflected in the results, with higher percentages of individuals not completing high school.

There were no significant differences between the groups with regards to number of hospitalisations, with the majority of patients reporting more than three admissions since being diagnosed. The frequent hospitalisations further emphasise the impact of the illness on the level of functioning and severity of psychopathology of patients in both groups.

None of the subjects with co-morbid OCD/OCS met the criteria for high functioning, whilst both groups displayed equal prevalence of low and medium level of global and social functioning.

Contamination obsessions, religious obsessions, repetitive rituals and checking behaviour were the most common obsessive-compulsive symptoms reported in the study group. This finding is similar to that of Fenton and McGlashan.[2] In the majority of their sample, the obsessions and compulsions were typical (eg. washing, contamination and rearrangement rituals).

The relationship between OCD/OCS and subtypes of schizophrenia in our study showed an increased prevalence in paranoid schizophrenia at 82.9%, followed by undifferentiated schizophrenia at 14.6% in the study group. This corresponds to findings in other studies,[1,14]

however, the association needs to be explored in future studies. Our findings were that positive symptoms, as well as depressive symptoms occurred more frequently in patients with schizophrenia or schizoaffective disorder and OCD/OCS as compared to patients without co-morbid OCD/OCS, supporting findings of other studies.[2,3]

Substance abuse was significantly higher in the control group without co-morbid OCD/OCS compared to the study group diagnosed with schizophrenia or schizoaffective disorder and co-morbid OCD/OCS. Cannabis was the most common substance of abuse in both groups, followed by alcohol. These findings were in agreement with those found in a study by Poyurovsky et al[24], where the incidence of substance abuse in the group diagnosed with schizophrenia and OCD were one-half of those found in the non-OCD schizophrenia group. This suggests the possibility of OCD in schizophrenia serving as a protective factor against substance abuse.[5]

The onset of psychotic symptoms was found to be insidious in the study group with co-morbid OCD/OCS compared to the control group. These findings, together with co-morbid depressive symptoms occurring in 73.1% [n=38] of patients, are of clinical significance as they point to severity of psychopathology and subsequent psychosocial dysfunction.

The main focus of the study was not on the identification of antipsychotic induced OCD/OCS. However, we found that more than three quarters of the study sample (77.4%) used second-generation antipsychotics as compared to first generation antipsychotics (22.6%).

Reports have indicated that clozapine, olanzapine and risperidone may induce de novo or aggravate pre-existing OCS in schizophrenia patients.[25,26,27] Previous studies have found anti-psychotic induced OCS to affect predominantly males with more compulsive behaviour occurring compared to obsessions. Symptoms usually appeared within 2-3 weeks after initiation of treatment with second-generation antipsychotics and with higher dosages of risperidone use (>4 mg per day.)[4] Future studies on this distinct subgroup of patient, examining the relationship between age of onset of OCD/OCS and the start of treatment with second-generation antipsychotics, type and dosage of treatment, as well as determining severity of OCS, might provide important evidence as to whether second-generation antipsychotics could induce or aggravate pre-existing OCD/OCS amongst study subjects.

Early non-psychotic deviant behaviour has been identified in a study to have some value as a possible endophenotypic marker in schizophrenia and schizoaffective disorder.[27]

Both study and control groups displayed a similar incidence of early non-psychotic deviant behaviour in the first ten years of life, with a significantly higher incidence of learning disabilities being experienced by the group of patients without co-morbid OCD/OCS. The frequency of these symptoms in all domains, encompassing social functioning impairment, mood/anxiety and cognitive impairment in both study and control groups is significant. The presence of early non-psychotic deviant behaviour in Afrikaner schizophrenia or schizoaffective disorder patients is in keeping with findings in previous studies[15] and does not seem to be influenced by a co-morbid diagnosis of OCD/OCS. The presence of early deviance is shown to affect the outcome and prognosis of the disease[15] and our findings may be seen as an indicator of poor prognosis in both the study and control groups.

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Limitations

This study has several methodological limitations. Although a structured clinical interview (DIGS) was used, one limitation of the study was that other assessment instruments were not utilised. When OC symptoms or OCD was diagnosed, these symptoms were not quantified by means of the Yale-Brown Obsessive-Compulsive Checklist and Severity Scale (Y-BOCS).The mean age at the time of the DIGS interview was 30.6 years for the study group and 33.6 years for the control group, and this could reflect the risk of a subject developing OCD/OCS at a later stage.The preponderance of male patients included in the study sample compared to female patients could be argued to be due to earlier age of onset of schizophrenia in males. More male patients could have been diagnosed and subsequently enrolled in the study. The difference in age of onset of illness between male and female probands in the study group was not taken into account.

Information regarding early deviant behaviour was obtained from the patients. The information may have been more reliable if access to school records and collateral information from family members were available.The current mental state of the patients may have influenced the accuracy of their account of their early lives and past psychiatric history.

Conclusion

The prevalence rate of 13.2% of co-morbid OCD/OCS in the Afrikaner schizophrenia and schizoaffective disorder patients supported our hypothesis that the prevalence is similar to studies in other Caucasian populations. It is however, in contrast to findings in a Xhosa-speaking schizophrenia population, suggesting the possible role of genetic and environmental factors in the prevalence of co-morbid OCD/OCS in schizophrenia and schizoaffective disorder in different ethnic groups.Second-generation antipsychotic use amongst schizophrenia and schizoaffective disorder patients with co-morbid OCD/OCS was found to be significantly higher than in those without co-morbid OCD/OCS.Clinical characteristics of Afrikaner schizophrenia and schizoaffective disorder patients with and without co-morbid OCD/OCS are the same, both groups being associated with significant psychopathology and a poor prognosis.

References

1. Rosen I. The clinical significance of obsessions in schizophrenia. J Ment S 1957; 103:773-785.
2. Fenton WS, McGlashan TN. The prognostic significance of obsessive-compulsive symptoms in schizophrenia. Am J Psychiatry 1986; 143:437-441.
3. Berman I, Kalinowski A, Berman SM, et al. Obsessive and compulsive symptoms in chronic schizophrenia. Compr Psychiatry 1995; 36:6-10.
4. Poyurovsky M, Weizman A, Weizman R. Obsessive-compulsive disorder in schizophrenia:clinical characteristics and treatment. CNS Drugs 2004;18(14):990-1010.
5. Karno M, Golding JM, Sorenson SB, et al. The epidemiology of obsessive-compulsive disorder in five US communities. Arch Gen Psychiatry 1998 Dec; 45(12): 1094-1099.
6. Yaryura-Tobias JA, Campisi TA, McKay D, et al. Schizophrenia and obsessive-compulsive disorder: shared aspects of pathology. Neurol Psychiatry Brain Res 1995;3:143-148.
7. Tibbo P, Warneke L. Obsessive-compulsive disorder in schizophrenia: epidemiologic and biologic overlap. J Psychiatry Neuroscience 1999;24:15-24.
8. Hwang MY, Opler LA. Management of schizophrenia with obsessive-compulsive disorder. Psychiatr Ann 2000;30:23-28.
9. M Poyurovsky, LM Koran. Obsessive-compulsive disorder (OCD) with schizotypy vs. schizophrenia with OCD: diagnostic dilemmas and therapeutic implications. Journal of Psychiatric Research 2005;39:399-408.
10. Emsley RA, Roberts MC, Rataemane S, et al. Ethnicity and treatment response in schizophrenia: a comparison of 3 ethnic groups. J Clin Psychiatry 2000; 63:9-14.
11. Weismann MM, Bland RC, Canino GJ, et al. The cross national epidemiology of obsessive-compulsive disorder. The Cross National Collaborative Group. J Clin Psychiatry 1994;55:suppl,5-10.
12. Niehaus DJH, Koen L, Muller J et al. Obsessive compulsive disorder-prevalence in Xhosa-speaking schizophrenia patients. SAMJ Feb 2005;vol 95, No 2:120-122.
13. Hwang MY, Morgan JE, Losconzco MF. Clinical and neuropsychological profiles of obsessive-compulsive schizophrenia: a pilot study. J Neuropsychiatry Clin Neuroci 2000;12:91-94.
14. Sobin C, Roos JL, Pretorius HW, et al. A comparison study of early non-psychotic deviant behaviour in Afrikaner and US patients with schizophrenia and schizoaffective disorder. Psychiatry Res 2003;117:113-125.
15. Roos JL, HW Pretorius, M Karayiorgou et al. Cannabis and other variables affecting age at onset in a schizophrenia founder population. S Afr. Psychiatry Rev 2006;9:99-103.
16. Veen ND, Selten JP, Van der Tweel I, et al. Cannabis use and age of onset of schizophrenia. Am J Psychiatry 2004:501-506.
17. Karayiorgou M, Torrington M, Abecasis GR, et al. Phenotypic characterization and genealogical tracing in an Afrikaner schizophrenia database. Am J Med Genet 2004;124B:20-28.
18. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders. 4th ed. Washington DC: American Psychiatric Association, 1994.
19. Nurnberger J, Blehar MC, Kaufmann CA, et al. Diagnostic Interview for Genetic Studies: Rationale, unique features and training. Arch Gen Psychiatry 1994;51:849-859.
20. Dominquez RA, Backman KE, Lugo SC. Demographics, prevalence and clinical features of the schizo-obsessive subtype of schizophrenia. CNS Spectrums 1999;12:50-56.
21. Ozdemir O, Turkel R, Turksoy N, et al. Clinical characteristics in obsessive-compulsive disorder with schizophrenia. Comprehensive Psychiatry 2003;vol 44(4):311-316.
22. Tibbo P, Warnke L. Obsessive-compulsive disorder in schizophrenia: epidemiologic and biologic overlap. J Psychiatry Neuroscience 1999,Jan;24(1):15-24.
23. Gross-Isseroff R, Hermesh H, Zohar J, et al. Neuroimaging communality between schizophrenia and obsessive compulsive disorder: a putative basis for schizo-obsessive disorder. World J Biol Psychiatry 2003,Jul 4(3):129-134.
24. Poyurovsky M, Fuchs C, Faragian S, et al. Aggregation of obsessive-compulsive spectrum disorders in schizophrenia patients with obsessive-compulsive disorder. Canadian J Psychiatry 2006;51(12):746-753.
25. Poyurovsky M, Hermesh H, Weizman A. Fluvoxamine treatment in clozapine induced obsessive-compulsive symptoms in schizophrenia patients. Clinical Neuropharmacology 1996;19:305-313.
26. Lykouras L, Zervas IM, Gournelius R, et al. Olanzapine and obsessive-compulsive symptoms. European Neuropsychopharmacology 2000;10:385-387.
27. Alevizos B, Lykouras L, Zervas IM, et al. Risperidone induced obsessive-compulsive symptoms: a series of 61 cases. Journal of Clinical Psychopharmacology 2002;22:461-467.

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Navorsings- en oorsigartikels / Research and review articles

Die Afrikaner en skisofrenie: 'n Genetiese studie

The Afrikaner and schizophrenia: A Genetic study

J L Roos
Hoof: Departement Psigiatrie, Universiteit van Pretoria,
Weskoppies Hospitaal

Louw Roos

Louw Roos het die grade MBChB en MMed(Psych) in onderskeidelik 1975 en 1983 (laasgenoemde met lof) behaal aan die Universiteit van Pretoria. In Oktober 1989 behaal hy die MD(Psigiatrie) aan die Universiteit van Pretoria met 'n proefskrif getiteld: "'n Ondersoek na oorsake van selfmoord by pasiënte met skisofrenie". Sedert 1976 is hy verbonde aan Weskoppies Hospitaal en die Universiteit van Pretoria en in 2001 word hy Professor en Hoof van die Departement Psigiatrie, Universiteit van Pretoria. Hy is die outeur of mede-outeur van 80 publikasies in spesialis en akademiese joernale. Hy is geïnteresseerd in navorsing oor gesinsmoord, selfmoord in pasiënte met skisofrenie en die genetika van skisofrenie.

Louw Roos obtained the degrees of MBChB and MMed(Psych) (the latter with distinction) at the University of Pretoria in 1975 and 1983 respectively. In October 1989 he obtained an MD(Psychiatry) at the University of Pretoria with a theses entitled: "An investigation into the causes of suicide by schizophrenic patients". He has been working at Weskoppies Hospital and the University of Pretoria since 1976 and in 2001 he became Professor and Head of the Department of Psychiatry, University of Pretoria. He is the author or co-author of 80 publications in specialist and academic journals. He is interested in research on family murder, suicide in schizophrenic patients and the genetics of schizophrenia.

ABSTRACT

The Afrikaner and schizophrenia: A Genetic study

An overview is given on the genetics of schizophrenia in an Afrikaner founder population. This research has been done over a period of ten years. Founder populations hold tremendous promise for mapping genes for complex traits, as they offer less genetic and environmental heterogeneity and greater potential for genealogical research. Not all founder populations are equally valuable, however. The Afrikaner population meets several criteria that make it an ideal population for mapping complex traits. These include founding by a small number of initial founders that likely allowed for a relatively restricted set of mutations and a large current population size that allows identification of a sufficient number of cases. Since a large number of patients was traced to an original founder couple, the genealogical data of the research support the hypothesis that a substantial portion of the schizophrenia cases the researchers have identified will share the same or a small number of susceptibility alleles. Results further show that genetic

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

variation at the 22q11 locus enhances the risk to develop schizophrenia. More than one gene in this region can independently or synergistically increase the risk of schizophrenia in non-deleted patients. In three independent samples, evidence was provided for a contribution of the PRODH/DGCR6 locus in 22q11 associated schizophrenia. One of these samples was from South Africa as part of an ongoing collection of schizophrenia patients of Afrikaner origin. An unusual pattern of PRODH2 gene variation that mimics the sequence of a linked pseudogene was uncovered. Several of the pseudogene-like variants identified, resulted in missense changes at conserved residues and may prevent synthesis of a fully functional enzyme. The results have implications for understanding the genetic basis of the 22q11-associated psychiatric phenotype and provide further insight into the genomic instability of this region. It was shown in an independent sample of Afrikaner schizophrenia patients that the frequency of 22q11 deletion syndrome was 2%, as suggested in earlier studies. Early deviant behaviour before 10 years of age was the most important factor related to the onset of illness. Early deviant childhood behaviour was grouped into three clusters: social functioning impairment cluster, mood/anxiety cluster and cognitive impairment cluster. Some of these factors are important in male schizophrenia patients and may be used as endophenotypic markers. The LRRTM1 gene is implicated in the paternal transmission of schizophrenia. This gene may be implicated in the development of specific forebrain structures. It influences neuronal differentiation and connections. Male users of cannabis suffering from schizophrenia; with prominent early deviant behaviour have the lowest age of onset of illness, namely 18.4 years. Approximately half of the male schizophrenia patients and a quarter of the female schizophrenia patients used or abused cannabis. These figures are high and make cannabis use/abuse an important part of the management strategy of these patients. It was stated that in male patients with a specific subtype of schizophrenia (where early deviant behaviour was present) the use of cannabis triggers an illness with earlier onset and a poor prognosis. The prevalence of obsessive-compulsive disorder and obsessive-compulsive symptoms in Afrikaner patients suffering from schizophrenia was 13.25%. This finding differs from those in another ethnic group in South Africa suggesting the possible role of genetic and cultural factors in the prevalence of co-morbid OCD/OCS in this patient population. The most common obsession reported was contamination followed by religious obsession. The most common compulsion was checking behaviour.

Clinical characteristics of Afrikaner schizophrenia and schizo-affective disorder patients with and without co-morbid OCD/OCS are the same; both groups were associated with significant psychopathology and a poor prognosis.

KEY CONCEPTS: Afrikaner, schizophrenia, schizo-affective disorder, founder population, endophenotype, genetics, linkage disequilibrium, 22q11, genomewide, cannabis, lefthandedness, single-nucleotide polimorphism, deletion

TREFWOORDE: Afrikaner, skisofrenie, skiso-affektiewe versteuring, stigtersbevolking, endofenotipe, genetiese, koppelingsonewewigtigheid, 22q11, genoomwye, cannabis, linkshandigheid, enkel-nukleotied-polimorfismes, uitwissing.

OPSOMMING

Die artikel gee 'n oorsig oor navorsingbevindings die afgelope tien jaar oor die genetika van skisofrenie in die Afrikaner as 'n stigtersbevolking. Die genealogiese data in die studie steun die hipotese dat 'n aansienlike deel van die skisofrenie-pasiënte wat geïdentifiseer is, dieselfde of 'n klein aantal vatbaarheid-allele deel, omdat 'n groot aantal pasiënte teruggespoor kan word na 'n

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

stigtersegpaar. Die resultate toon verder dat genetiese variase op die 22q11-lokus die risiko om die siekte te ontwikkel verhoog. Meer as een geen op hierdie gebied kan onafhanklik of sinergisties die risiko in pasiënte, waar die geen nie uitgewis is nie, verhoog. Die studie het ook, in 'n onafhanklike steekproef, bevestig dat die frekwensie van die 22q11-uitwissingsindroom in volwasse Kaukasiese skisofreniese pasiënte ongeveer 2% is, soos voorheen voorgestel. Vroeë afwykende gedrag is die belangrikste faktor wat die ouderdom van aanvang van siekte bepaal veral in manlike pasiënte met skisofrenie en is 'n moontlike endofenotipe merker. Die geen LRHTM1 (leusienryk-herhaal-transmembraan-neuronaal-1) word geïmpliseer by vaderlike oordrag van skisofrenie en skiso-affektiewe versteuring. Waarskynlik speel die geen 'n rol gedurende die ontwikkeling van spesifieke voorbreinstrukture, deur neuronale differensiasie en verbinding te beïnvloed. Manlike verbruikers van cannabis met ernstige vroeë afwykende gedrag het die laagste gemiddelde ouderdom van aanvang van siekte gehad, naamlik 18.4 jaar. Dit word geïnterpreteer dat in mans met 'n spesifieke subtipe van skisofrenie (waar vroeë afwykende gedrag teenwoordig is) cannabis die aanvang van siekte sal ontlont op 'n vroeër ouderdom en die psigose het gevolglik 'n swak prognose. Die voorkoms van obsessief-kompulsiewe versteuring en obsessief-kompulsiewe simptome in Afrikaner pasiënte met skisofrenie was 13.25%. Die algemeenste gerapporteerde obsessie was kontaminasie, gevolg deur geloofsobsessies. Die algemeenste kompulsie was nagaan-gedrag.

1. INLEIDING

Die artikel gee 'n oorsig oor navorsingstudies die afgelope tien jaar oor die genetika van skisofrenie in die Afrikaner as 'n stigersbevolking. Navorsing is ook onder meer gedoen oor die verwantskap tussen skisofrenie onder Afrikanerpasiënte en cannabisverbruik, linkshandigheid, vroeë afwykende gedrag as endofenotipe merker en obsessief-kompulsiewe gedrag.

As 'n nuut gekwalifiseerde psigiater (met baie lewenslus en drome), het die outeur 'n 78-jarige vrou vir bipolêre gemoedsversteuring by 'n plattelandse gemeenskaps psigiatriese kliniek behandel. Sy was die ma van agt sibbe, van wie drie manlike afstammelinge ook psigiatriese behandeling ontvang het. Hierdie drie mans se diagnoses was onder meer skisofrenie en skiso-affektiewe versteuring: bipolêre tipe.

Die vraag het ontstaan oor hoe genetiese kennis psigiaters in die toekoms nader aan antwoorde rakende veroorsakende faktore van ernstige psigiatriese versteurings sou kon bring. Die outeur het 'n baie eenvoudige kennis van die genetika van mediese siektes gehad. Dit het beswaarlik meer as 'n basiese kennis oor trisomie 21 (Down se sindroom) en porfirie ingesluit. In daardie stadium het die literatuur oor die genetika van psigiatriese versteurings hoofsaaklik familie- en aannamestudies beklemtoon.

Fondse was, en is, steeds beperk vir die genetiese navorsing in die veld van psigiatrie in Suid-Afrika. Wat wel vir die outeur duidelik was, is die feit dat daar in die land volop families was met lede wat op een of ander manier psigiatries aangetas was.

Toe die Menslike Neurogenetika-laboratorium aan die Rockefeller-universiteit in Amerika die Departement van Psigiatrie aan die Universiteit van Pretoria (UP) nader rakende 'n samewerkingsooreenkoms oor die genetika van skisofrenie in 'n stigtersbevolking, het die departement die kans aangegryp. Hier was 'n kans om meer oor die genetika van 'n ernstige psigiatriese toestand te leer en om sodoende betrokke te raak by die ontrafeling van veroorsakende faktore van 'n algemeen voorkomende toestand, naamlik skisofrenie. Nadat etiese goedkeuring by die twee universiteite verkry is, is daar in November 1997 met die werwing van pasiënte begin.

In dié artikel word 'n oorsig gegee van die bevindings van die bogenoemde studie. Die afgelope agt jaar is die bevindings daaroor in nasionale en internasionale joernale gepubliseer. Die genetiese

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

metodologie kan in die oorspronklik gepubliseerde artikels gevind word. Hierdie studie gaan steeds voort en tans is bykans 500 families en 670 persone met skisofrenie in die studie ingesluit.

Die oorsig begin met die rasionaal waarom navorsing in hierdie verband op die Afrikaners gedoen is. Hiermee word nie gesuggereer dat die insidensie van psigose hoër in die Afrikaner as in ander bevolkings is nie. Die klem val op bevindings in 'n chronologiese formaat soos gepubliseer in die verskillende joernale.

2. WAAROM IS 'N AFRIKANER-STIGTERSBEVOLKING GEBRUIK VIR 'N GENETIESE STUDIE VAN SKISOFRENIE?

Afrikaners verteenwoordig ongeveer drie miljoen van die Suid-Afrikaanse bevolking. Hulle is afstammelinge van 'n klein getal oorspronklike Setlaars. In 1652 het die eerste immigrante hulle in die Kaap gevestig. Daar was net meer as 'n duisend, wat primêr van Nederlandse afkoms was.

Hulle het later die binneland ingetrek, deur sekere spesifieke roetes te volg. Sommige van hierdie Afrikanerfamilies het uiteindelik in die huidige Tshwane-area en wat tans bekend staan as die Limpopo- en Mpumalanga-provinsies beland. Die Afrikanerfamilies vir die betrokke genetiese skisofreniestudie is versamel uit bogenoemde areas, asook omliggende gebiede.

Die gemeenskappe wat uiteindelik gestig is, was geografies geïsoleer. Faktore wat verder tot hulle isolasie bygedra het sluit in:

a. Kulturele oorwegings
b. Taalverskille (Afrikaans afkomstig van Nederlands)
c. Geloofsoorwegings (die meeste Afrikaners was lidmate van die Hollandse Gereformeerde Kerk)

In die vroeëre generasies was dit algemeen gebruiklik om te ondertrou. Die bevolkingsgroei oor 13 – 15 generasies was feitlik uitsluitlik die gevolg van voortplanting, omdat immigrasie ná die aanvanklike stigting minimaal was. In meer onlangse jare het vermenging begin plaasvind. Die demografiese geskiedenis van hierdie bevolking word gereflekteer in die ongewone hoë frekwensie van sekere unieke mendeliese toestande, die ongewone lae diversiteit van die geassosieerde alleliese variante en die ongewone groot omvang (8 – 11 cM) van gespaarde haplotipes rondom siekte-gene.[1]

2.1 Uitgebreide intermerker-koppelingsonewewigtigheid in die Afrikaners

Wanneer ingewikkelde oorerwingspatrone in 'n siekte teenwoordig is, toon die tradisionele analise van parallelle verdeling van siekte en genetiese merkers binne stambome (koppelingsanalise) 'n beperkte krag. Om dié probleem te omseil, kan 'n aantal benaderings wat koppelingsonewewigtigheid gebruik, voorgestel word.[2]

Hierdie benaderings sluit bevolkingsgebaseerde alleel-assosiasiemetodes of gesamentlike analise van koppelings- en alleel-assosiasies in. Deur gebruik te maak van families wat afkomstig is van stigtersbevolkings, kan albei hierdie faktore gefasiliteer word.[2]

Verskeie elemente bepaal die gemete koppelingsonewewigtigheid (KO) tussen 'n merker en 'n siekte-alleel:

- Die afstand tussen die merker en die siekte-geen
- Die ouderdom van die siekte-alleel
- Die frekwensie van die gekoppelde merker-alleel
- Genetiese strekking/neiging

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

- Tempo van bevolkingsuitbreiding
- Genetiese bottelnekke
- Vermenging

(Die laaste vier is voorspelbare kragte)

Koppelingsonewewigtigheid (KO) vertraag soos die afstand toeneem. Verskeie onlangse studies toon aan dat met 'n kort afstandreeks die bydrae van 'n rekombinasie van die vlak van KO weglaatbaar klein sal wees vergeleke met ander kragte. Soos die afstand egter toeneem, sal die toename in rekombinasie-frekwensie enige assosiasie degradeer. Gevolglik sal rekombinasie en afstand die primêre element word wat die vlak van KO bepaal.[2]

Pasiënte uit stigtersbevolkings sal waarskynlik segmente van chromosome, wat die siektegeen bevat, oorerf, sowel as omringende merker-loci waarvoor 'n gedeelde haplotipe geïdentifiseer kan word. Hierdie segmente van chromosome word van 'n gedeelde voorouer (gemeenskaplike voorsaat) oorgeërf. Die relatief onlangse ontstaan van 'n bepaalde stigtersbevolking maak dit moontlik dat chromosoomgebiede wat identies is, as gevolg van herkoms (descent) wat die siekte alleel omring, groter sal wees as by uitgeteelde (outbred) bevolkings. Daarom sal 'n optimaal ontwerpte genoomwye skandering – wat voordeel trek uit KO, en wat ook stigtersbevolkings betrek van onlangse voorouers – beduidend minder merkers benodig, of 'n beduidende kleiner steekproef as soortgelyke skanderings in uitgeteelde bevolkings vereis.[2]

Baie vrae bly onopgelos rakende die optimale ontwerp van sulke genoomwye karteringstudies. Om hierdie vraagstukke aan te spreek, moet die kragte wat die verspreiding van agtergrond-KO in die genoom bestuur, verstaan word. Die outeurs rapporteer in *Genome Research* (2002) 'n ondersoek oor die agtergrondsvlak van KO in die Afrikanerbevolking. In die proses evalueer hulle die toepaslikheid van hierdie geneties-geïsoleerde groep vir die kartering van komplekse kenmerke.[2]

Die navorsers analiseer intermerker-KO in 62 kernfamilies deur gebruik te maak van mikro-satellietmerkers wat uitgebreide chromosomale gebiede dek. Die merkers is so geselekteer dat die eerste direkte vergelyking van lang afstand KO in die Afrikaner met die KO in ander demografiese groepe kon plaasvind. Deur gebruik te maak van verskeie statistiese metodes, is voldoende bewys gevind vir KO in die Afrikaner wat verby die 6 cM reeks strek.[2]

Hierdie studie het die toepaslikheid van die Afrikanerbevolking vir genoomwye skanderings bevestig. Die uitgangspunt was dat KO gebruik sou word om algemene, multigenetiese versteurings te karteer.

'n Algemene kommer by die gebruik van stigtersbevolkings is die relevansie van die studie se bevindings vergeleke met ander kosmopolitaanse bevolkings. Tans is daar geen empiriese data vir of teen só 'n argument nie. Indien hierdie aspek van kritieke belang sou word, sou geredeneer kon word dat die Noord-Europese/Hollandse oorsprong van die Afrikaner die studieresultate in die Afrikanerbevolking vir ten minste baie ander Westerse bevolkings relevant maak.

Dit is moontlik dat slegs 'n subgroep van vatbaarheidsgene vir enige komplekse versteuring sou segregeer in 'n stigtersbevolking. Desnieteenstaande sou identifikasie van hierdie gene belangrike inligting verskaf rakende die geaffekteerde molekulêre en sellulêre seinbane en sou dit nuwe geneesmiddelontwikkelings en -benaderings inspireer.[2]

2.2 Omvang, verspreiding en hoeveelheid van koppelingsonewewigtigheid (KO) in elf geïsoleerde bevolkings

Die genoomwye verspreiding van KO bepaal die strategie van hoe merkers vir assosiasiestudies gekies sal word, en dit varieer tussen bevolkings.

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Met die vergelyking van 11 bevolkings het die ondersoekers gedemonstreer dat geïsoleerde bevolkings afkomstig van 'n klein stigtersbevolking en wat onlangs dramatiese bevolkingsgroei getoon het, in die algemeen hoër vlakke van KO het as dié van uitgeteelde bevolkings. Die geïsoleerde bevolkings het minder gebiede van baie lae KO.[3]

Die volgende geïsoleerde bevolkings is in die studie ingesluit:

1. Antioquia Columbië
2. Asjkenasi (Joodse bevolking)
3. Azores
4. Costa Ricaans: Sentraal-vallei
5. Suidwestelike Nederland
6. Vroeë Finlandse Setlaars
7. Laat Finlandse Setlaars: sub-isolaat Kuusamo
8. Finland (hele nasie)
9. Newfoundland
10. Afrikaner – Suid-Afrika
11. Sardinië, provinsie van Nuoro

Hierdie geïsoleerde bevolkings is vergelyk met uitgeteelde Kaukasiërs. Sommige gebiede van lae KO in hierdie studie is voldoende gedek met 'n hoër merkerdigtheid. Die ondersoek van KO-gapings in twee weergawes van die HapMap-data wys dat 'n substansiële aantal gebiede lae KO-areas bly selfs ná die byvoeging van etlike duisende merkers.

Gebiede van lae KO, aangetoon deur KO-gapings, veroorsaak afname in die krag van assosiasiestudies en dui op gebiede waar KO-kartering sal faal om assosiasies te identifiseer. Skynbaar is daar lae KO-gebiede wat moeilik in assosiasie-analise gedek sal word.

Bevolkings met minder sulke gebiede sal voordeliger wees vir assosiasiestudies. Die navorsers beklemtoon dat daar min gebiede in geïsoleerde bevolkings sal wees waar KO-gebaseerde assosiasiekartering onuitvoerbaar is, terwyl dit wel die geval kan wees in uitgeteelde bevolkings.[3]

2.3 Fenotipe-kenmerke en genealogiese nasporing in 'n databasis vir Afrikaners met skisofrenie

Soos reeds genoem, voldoen die Afrikanerbevolking aan verskeie kriteria wat dit 'n ideale bevolking maak vir die kartering van komplekse trekke, omdat:[1]

- dit gestig is deur 'n klein aantal oorspronklike stigters
- dit aanleding gee tot 'n relatief beperkte groep mutasies
- dit tans 'n groot bevolkingsgrootte het, wat sal toelaat dat 'n voldoende aantal gevalle geïdentifiseer kan word

Die vrae wat beantwoord moes word, sluit in:

- Wat is die potensiaal van hierdie bevolking om genealogies nagevors te word?
- Wat is die kliniese ooreenkomste van gevalle van skisofrenie gediagnoseer in Suid-Afrika en dié gediagnoseer in ander heterogene bevolkings, spesifiek gevalle in die VSA?[1]

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

3. GENEALOGIESE NAVORSING IN DIE AFRIKANER

Dit is moontlik om genealogiese verbande met akkuraatheid in die Afrikanerbevolking na te spoor. Die genealogiese data in hierdie studie ondersteun die hipotese dat 'n substansiële deel van skisofrenie-pasiënte wat geïdentifiseer is dieselfde of 'n klein aantal vatbaarheid-allele sal deel. Dít is omdat 87 pasiënte wat op daardie stadium geïdentifiseer is, na 'n enkele stigtersegpaar teruggespoor kon word.[1]

Deur die vasstelling van gedetailleerde genealogiese verbande tussen aangetaste pasiënte, kan dié wat na alle waarskynlikheid die genetiese basis van die siekte deel, saamgegroepeer word. Sodoende kan segmente DNS wat – identies-weens-herkoms – gedeel word, op 'n genoomwye vlak geïdentifiseer word.[1]

3.1 Kliniese vergelyking tussen enersyds skisofreniegevalle geïdentifiseer en gediagnoseer in die Afrikanerbevolking en andersyds 'n steekproef van pasiënte in die VSA

Hierdie vergelyking dui daarop dat die genetiese bevindings in die Afrikanerstudie toepaslik sal wees in ander bevolkings en ten opsigte van skisofrenie in die algemeen.

Die diagnostiese evaluasie by die twee sentra is uitgevoer deur onafhanklike groepe klinici wat spesiale opleiding ondergaan het om die Diagnostiese Onderhoud vir Genetiese Studies (DOGS) toe te pas. Elf demografiese en kliniese diagnostiese kenmerke was stabiel regoor beide bevolkings en met behulp van basiese steekproefomskrywers en kardinale simptome van die siekte, is daar geoordeel dat die twee bevolkings dieselfde is.[1]

Demografiese en kliniese veranderlikes was verskillend. Dit is nie bekend of hierdie verskille dui op genetiese of kulturele kenmerke van die Afrikanergroep nie.[1]

3.2 Cannabis en ander veranderlikes wat die aanvangsouderdom in 'n skisofrenie-stigtersbevolking beïnvloed

Gemeenskapsopnames toon aan dat cannabis die onwettige middel van keuse in Suid-Afrika is. In gerapporteerde leeftydsyfers van cannabisgebruik, in 'n verteenwoordigende groep graad 11-studente in Kaapstad,[4] is bevind dat 32% in manlike leerders en 13.1% in vroulike leerders voorkom.

Cannabis het die tweede mees algemene leeftyd prevalensie ná alkohol. Die belang van cannabismisbruik in Suid-Afrika word ook aangetoon in statistiek afkomstig van spesialis middelbehandelingsentrums. In die tydperk Julie tot Desember 2003, rapporteer 20%–40% van pasiënte in verskeie middelbehandelingsentrums, wat versprei is oor Suid-Afrika, dat cannabis hulle primêre middel van verbruik is. Hierdie groep respondente se gemiddelde ouderdom was 20 tot 21 jaar.[4]

Persone wat aan 'n psigotiese versteuring ly se kans om cannabis te gebruik, is waarskynliker as die algemene bevolking s'n. Die vraag ontstaan of psigotiese persone cannabis gebruik om hulle simptome te verlig en of cannabisgebruik 'n risiko-faktor is om hulle toestand te presipiteer. Die aard van hierdie verband is nog nie bepaal nie.

Bevindings afkomstig van kliniese groepe met skisofrenie het beperkte waarde om lig te werp op die meganisme wat hierdie assosiasie onderlê. In hierdie gevalle is die potensieel verwarrende faktore wat verband hou met die kliniese status van die individu moeilik om te kontroleer.

Data van kliniese groepe wat navorsing doen oor skisofrenie pasiënte mag lig werp op hoe cannabisgebruik/misbruik sekere faktore beïnvloed; en interakteer met ander soos: ouderdom van aanvang van siekte, die langtermynverloop van die siekte, geslag, vroeë afwykende gedrag in die

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

eerste tien jaar van lewe, familiegeskiedenis van skisofrenie in eerste- en tweederangse families en vroeë inmenging rondom geboorte. Vroeë afwykende gedrag in die eerste tien jaar van lewe het aspekte soos: ingekorte sosialisering, erge vrese/chroniese hartseer, en/of aandag- en leerinkorting ingesluit. Hierdie faktore is ondersoek in die genetiese databank van Afrikaners met skisofrenie.[4]

In 2005 was daar inligting beskikbaar oor 341 pasiënte wat in daardie stadium by die studie ingesluit was.

Resultate van hierdie studie toon 61% (n=209) van dié groep is mans en 39% (n=132) vroue. Altesame 35% van die pasiënte het cannabis gebruik/misbruik. Dit was 44% (n=91) mans en 20% (n=27) vroue.[4]

Mans wat cannabis gebruik het se gemiddelde ouderdom vir die aanvang van die siekte was 20.5 jaar. Dit is beduidend vroeër as die ooreenstemmende ouderdom in mans sonder cannabis-gebruik/misbruik, sowel as in beide groepe vroue. Die langtermyn verloop van die siekte was dieselfde in albei groepe.[4]

'n Analise van variasie toon dat vroeë afwykende gedrag die belangrikste faktor is wat die aanvangsouderdom van die siekte aangetoon het. Die groep met die laagste gemiddelde ouderdom vir aanvang van die siekte, is manlike verbruikers van cannabis met ernstige afwykende gedrag in die eerste 10 jaar van hulle lewe (18.4 jaar). Die syfers vir cannabisgebruik/misbruik in hierdie groep pasiënte is waarskynlik hoër aangesien hierdie feite gewoonlik onder-rapporteer word.[4]

Die inligting verkry van pasiënte is nie deur 'n urine-analise vir cannabinoïede bevestig nie. Gevolglik kan aangeneem word dat ongeveer die helfte van manlike skisofrenie-pasiënte en 'n kwart van vroulike pasiënte cannabis gebruik/misbruik. Hierdie syfers is beduidend en beklemtoon die belang van cannabis-gebruik/misbruik as 'n sentrale tema in die algehele behandeling van hierdie groep pasiënte.[4]

Die langtermyn siekteverloop van verbruikers en nie-verbruikers was nie statisties beduidend verskillend in die twee groepe nie. Daar is egter aanduidings dat cannabis-gebruik/misbruik 'n negatiewe effek mag hê op die langtermynverloop van die siekte.[4]

Die ouderdom waarop 'n persoon met skisofrenie die eerste maal psigoties raak, is 'n veranderlike faktor wat verband hou met die prognose. 'n Vroeër aanvang van psigose word geassosieer met 'n ernstiger verloop van die siekte, ongeag die duur daarvan.

Mans met erge vroeë afwykende gedrag was die groep met die jongste gemiddelde aanvangsouderdom van die siekte. Hierdie bevindings bevestig vorige bevindings wat daarop dui dat vroeë afwykende gedrag verband hou met die latere verloop in die siekte. In hierdie studie het vroeë afwykende gedrag die aanvangsouderdom van die siekte, wat geassosieer word met 'n ernstige verloop daarvan, beïnvloed. Hierdie bevindings was beduidend in mans met geringe en ernstige afwykende gedrag, en in vroue met erge afwykende gedrag in die eerste tien jaar van lewe.[4]

Manlike verbruikers van cannabis met ernstige vroeë afwykende gedrag het die laagste gemiddelde ouderdom van aanvang van siekte gehad, naamlik 18.4 jaar. Dit word geïnterpreteer dat in mans met 'n spesifieke subtipe van skisofrenie (waar vroeë afwykende gedrag teenwoordig is) cannabis die aanvang van siekte sal ontlont op 'n vroeër ouderdom en in sodanige geval het die psigose 'n swak prognose. Hierdie bevindings beklemtoon die belang vir vatbare manlike pasiënte wat cannabisgebruik moet vermy.[4]

Vorige bevindings in 'n studie van Finse families het getoon dat die aanvangsouderdom afgeneem het namate die familierisiko vir skisofrenie toegeneem het. Onder die Afrikaners was 'n familiegeskiedenis van skisofrenie minder belangrik om die aanvangsouderdom te verklaar as vroeë afwykende gedrag; die interaksie van vroeë afwykende gedrag en manlike geslag; en die interaksie van cannabis-gebruik/misbruik en manlike geslag.[4]

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Gevolgtrekkings van hierdie studie het die volgende ingesluit:

Daar is bevind dat die persentasie manlike en vroulike Afrikanerpasiënte met skisofrenie wat cannabis gebruik/misbruik hoër is as die leeftydsyfers in manlike en vroulike leerders in graad 11 in Suid-Afrika. Die syfers in hierdie studie stem ooreen met syfers gerapporteer in verskeie middelbehandelingsentrums wat versprei is oor Suid-Afrika — met cannabis as die primêre middel van misbruik. Meer klem behoort gelê te word op die psigo-opvoeding en rehabilitasie van jong, vatbare individue met skisofrenie wat cannabis gebruik en misbruik.[4]

Vroeë afwykende gedrag is die belangrikste faktor wat die aanvangsouderdom van siekte bepaal het in manlike pasiënte met skisofrenie en is 'n moontlike endofenotipe merker. Die interaktiewe effek van geslag en cannabisgebruik was 'n beduidende faktor om die aanvangsouderdom in mans te verklaar.[4]

Dit het egter geblyk dat wanneer cannabis gebruik word, die effek van vroeë afwykende gedrag minder belangrik sal word in die bepaling van aanvangsouderdom.

In ons soeke na meer gesofistikeerde endofenotipes om ons te help met die identifikasie van vatbare en modifiserende gene, behoort vroeë afwykende gedrag, manlike geslag en cannabis gebruik/misbruik in interaksie in toekomstige navorsing nagegaan te word.[4]

3.3 Prevalensie en kliniese kenmerke van obsessief-kompulsiewe versteuring en obsessief-kompulsiewe simptome in Afrikanerpasiënte met skisofrenie en skiso-affektiewe versteuring

Epidemiologiese en kliniese studies skat die voorkoms van obsessief-kompulsiewe verskynsels (OKV) in skisofrenie tussen 3%–59%. Hierdie syfer is aansienlik hoër as die 2%–3% voorkoms van OKV en obsessief-kompulsiewe simptome (OKS) in die algemene bevolking. Daar is ook aanduidings van variasie in die voorkoms van komorbiede OKV tussen etniese groeperings. Die huidige studie evalueer die voorkoms en kliniese kenmerke van OKV/OKS in Afrikanerpasiënte met skisofrenie en skiso-affektiewe versteuring.[5]

Die voorkoms van komorbiede OKV/OKS by 400 pasiënte was 13.25% (n = 53). Van hierdie groep pasiënte was 40 mans en 13 vroue. Die gemiddelde aanvangsouderdom van OKV/OKS was vroeër ongeveer 18.5 jaar teenoor die aanvangsouderdom van skisofrenie en skiso-affektiewe versteuring van 22 jaar ($p<0.0001$).[5]

Altesame 78.72% van die studiegroep het voldoen aan die kriteria vir OKV en 21.28% vir obsessief-kompulsiewe simptome. Die algemeenste obsessie gerapporteer was kontaminasie (n = 17), gevolg deur geloofsobsessies (n = 8). Die algemeenste kompulsie was nagaan-gedrag (n = 22). Daar was geen familiegeskiedenis van OKV/OKS in die twee groepe nie.[5]

Die aanvang van psigose en OKS was geleidelik in 86.79% van die studiegroep vergeleke met 24.56% van die kontrolegroep ($p<0.0001$). Tweede generasie antipsigotika-gebruik was statisties meer prevalent in die studiegroep (77.55%) vergeleke met die kontrolegroep (45.76%)($p=0.0008$). 73% van die studiegroep het depressiewe simptome beleef vergeleke met 50.85% van die kontrolegroep. Beide groepe het 'n soortgelyke voorkoms van selfmoordgedagtes en -pogings gereflekteer.[5]

Middelmisbruik was beduidend hoër in die kontrolegroep (35.59%) vergeleke met die studiegroep (19.23%) ($p<0.05$), met cannabisgebruik die algemeenste middel van misbruik gevolg deur alkohol.[5]

Beide groepe het 'n soortgelyke voorkoms van vroeë nie-psigotiese afwykende gedrag gehad, maar 'n beduidende hoër voorkoms van leerprobleme onder 10-jarige ouderdom het voorgekom in die kontrolegroep (54.24%) vergeleke met die studiegroep (32.08%) ($p <0.01$)[5]

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Die voorkoms van 13.25% van komorbiede OKV/OKS in Afrikanerpasiënte met skisofrenie en skiso-affektiewe versteuring kontrasteer met die lae voorkoms van OKV in 'n Xhosa-sprekende skisofrenie-studiebevolking wat 0.5% was. Hierdie bevindings dui op die moontlike rol van genetiese en kulturele faktore in die voorkoms van komorbiede OKV/OKS.[5]

Die hoër voorkoms van tweede generasie antipsigotika-gebruik dui op die moontlikheid van behandelingsgeïnduseerde OKV/OKS of die verergering van voorafbestaande OKV/OKS.[5]

3.4 Vroeë nie-psigotiese afwykende gedrag as 'n endofenotipe merker in bipolêre gemoedsversteuring, skiso-affektiewe versteuring en skisofrenie

In 'n vorige vergelykende studie het ons groep die demografie, sindroomverloop, simptome en vroeë afwykende gedraggeskiedenis van 109 Afrikanerpasiënte met skisofrenie en skiso-affektiewe versteuring vergelyk met 109 ouderdom- en geslagsverwante Amerikaanse pasiënte. Ongeveer $^{2}/_{3}$ van sowel Afrikanerpasiënte (68%) as 67% van Amerikaanse pasiënte het een of meer vorms van vroeë nie-psigotiese afwykende gedrag getoon.[6]

Hierdie vroeë nie-psigotiese gedrag sluit in aspekte soos: swak sosialisering, erge vrese/chroniese hartseer, en/of aandaggebrek en leerinkorting. In die huidige studie is die vraag gestel of vroeë afwykende gedrag uniek was vir persone wat gediagnoseer is met skisofrenie en skiso-affektiewe versteuring en of dit ook voorgekom het by bipolêre gemoedsversteuring.[7]

Skisofrenie-genetiese navorsing toon vooruitgang met nuwe tegnologie tans beskikbaar in molekulêre genetika. Geen-identifikasie word bemoeilik deur die heterogene en onsensitiewe definisie van die kliniese fenotipe. 'n Groot mate van heterogenisiteit bestaan in die simptome van individue wat 'n gemeenskaplike kliniese diagnose deel. Baie van die individue is nie klassifiseerbaar in kategorieë wanneer daar gebruik gemaak word van standaard kriteria nie.

Verfyning van die fenotipes deur gebruik te maak van endofenotipes is waardevol en behoort die kanse te vergroot om meganismes van oorerwing, sowel as vatbaarheidsgene, te ontdek. Endofenotipes verskaf 'n meganisme vir identifikasie van "afstroom"-kenmerke of fasette van kliniese fenotipes, sowel as "opstroom"-gevolge van gene. Die endofenotipe kan 'n aanwysing wees tussen die genotipe en die gedrag wat ondersoek word, in hierdie geval skisofrenie.[7]

In die huidige studie is 74 pasiënte met bipolêre gemoedversteuring, 43 met skiso-affektiewe versteuring en 80 pasiënte met skisofrenie ondersoek. 'n Gestruktureerde vraelys is gebruik om vas te stel of vroeë afwykende kindergedrag teenwoordig was voor die ouderdom van 10 jaar. Inligting van nabye familie is ook waar moontlik ingewin.[7]

Sewe areas van moontlike afwykende gedrag is ondersoek: sosiale disfunksie, ongeprovokeerde aggressie, erge angs, chroniese hartseer, erge vreemde gedrag, aandaginkorting en leerprobleme.

Resultate het aangedui dat vroeë afwykende gedrag as 'n endofenotipe-merker in skisofrenie en skiso-affektiewe versteuring gebruik kan word. Die frekwensie van hierdie gedrag in 65% van skisofrenie-pasiënte en 60% met skiso-affektiewe versteuring, is statisties beduidend hoër as die frekwensie van 21% in pasiënte met bipolêre gemoedversteuring. Hierdie resultate ondersteun ons hipotese dat vroeë nie-psigotiese afwykende gedrag spesifiek op skisofrenie en skiso-affektiewe versteuring van toepassing is. Dit mag 'n handige en geldige endofenotipe merker in genetiese studies vir hierdie toestande wees. Dit is aangetoon dat pasiënte met bipolêre gemoedsversteuring minder geneig tot sosiale en kognitiewe inkorting as die ander twee diagnostiese groepe was.[7]

Inligting oor 'n geskiedenis van psigose is nie verkry in die oorspronklike onderhoud van die pasiënte met bipolêre gemoedsversteuring nie. In retrospek is psigose 'n belangrike deel van die neem van geskiedenis in die konteks van hierdie studie, aangesien daar geredeneer kan word dat vroeë afwykende gedrag verband mag hou met psigose per se, en nie noodwendig met skisofrenie en skiso-affektiewe versteuring nie.

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Met die naspoor van inligting uit kliniese lêers was dit moontlik om inligting oor psigose te kry by 61 van die 74 pasiënte met bipolêre gemoedsversteuring wat aan die studie deelgeneem het. Van hierdie 61 pasiënte het 39 (63.9%) gely aan bipolêre gemoedsversteuring tipe I en 22 (36.0%) aan bipolêre gemoedsversteuring tipe II. Van hierdie bipolêre pasiënte (n=61) het 22 pasiënte (36.0%) 'n geskiedenis van psigose gehad. Hierdie 22 pasiënte met 'n geskiedenis van psigose het almal aan bipolêre gemoedsversteuring tipe I gely, en het 56.4% van die totale groep van bipolêre pasiënte uitgemaak.[7]

Met net meer as $^{1}/_{3}$ van bipolêre pasiënte met 'n geskiedenis van psigose sou mens kon redeneer dat die hoër voorkoms van vroeë nie-psigotiese afwykende gedrag wat gevind is in skisofrenie en skiso-affektiewe versteuring, verband hou met die psigose per se, en nie met die versteuring opsigself nie. Toekomstige vergelykende studies met groter groepe pasiënte met bipolêre gemoedsversteuring, waar gebruik gemaak word slegs van pasiënte met tipe I van die siekte, en wat 'n geskiedenis van psigose het, sal nodig wees om lig te werp op hierdie aspek.[7]

In die interpretasie van die demografiese data is bevind dat die bipolêre pasiënte ouer was ten tye van die onderhoud. Dit mag verklaar word deur die latere aanvang van siekte, soos bevind in bipolêre pasiënte in hierdie betrokke studie.

Die hoë voorkoms van ongetroude pasiënte met skisofrenie, beklemtoon die verlammende effek van skisofrenie op sosiale funksionering. Die negatiewe effek van skisofrenie, relatief tot skiso-affektiewe en bipolêre gemoedsversteuring ten opsigte van skoolfunksionering, word ook in ons resultate weerspieël. Die gemiddelde skisofrenie-pasiënt het nooit sy skoolopleiding voltooi nie.

Bipolêre pasiënte was meer geneig tot selfmoordpogings in hulle leeftyd as pasiënte wat gediagnoseer is met skisofrenie of skiso-affektiewe versteuring. Daar word voorgestel dat toekomstige studies groter groepe sal insluit. Verder behoort ander diagnostiese versteurings en 'n gesonde kontrolegroep gebruik te word. Voorts behoort skoolverslae/rapporte en familie-onderhoude addisioneel tot die subjektiewe pasiëntonderhoude gebruik te word.[7]

4. GENETIESE VARIASIE BY DIE 22Q11-LOKUS DUI OP 'N ABNORMALE PATROON EN VERHOOG DIE RISIKO VIR SKISOFRENIE

4.1 Mikro-uitwissing

Mikro-uitwissing van chromosoom 22q11 word met 'n verskeidenheid van fenotipiese uitdrukkings geassosieer wat leerabnormaliteite, abnormaliteite van die verhemelte, kongenitale hartdefekte en geringe dismorfiese kenmerke insluit.

Daar is aanduidings vir moontlike koppeling met skisofrenie op chromosoom 22. Die evaluasie van psigiatriese simptome in persone met die 22q11-mikro-uitwissing toon 'n relatief hoë frekwensie van ernstige psigiatriese versteurings. Twee onafhanklike studies rapporteer dat 25%–31% van pasiënte met die 22q11 mikro-uitwissing aan die diagnostiese kriteria vir skisofrenie of skiso-affektiewe versteuring voldoen.[8]

Die mikro-uitwissing op chromosoom 22q11 kom voor in die samelewing met 'n voorkoms van 0.025%, in 2% van volwasse skisofreniese pasiënte en 6% van kinder-aanvangsskisofreniegevalle. Die siekterisiko vir skisofrenie in 'n pasiënt met 'n 22q11-mikro-uitwissing is ongeveer 20–30 keer hoër as die algemene bevolking se risiko van 1%. Die voorkoms van die 22q11-mikro-uitwissing in skisofrenie, hoewel dit relatief laag is, is dus ongeveer 80 keer die geskatte voorkoms in die algemene bevolking.

Dit is waarskynlik dat die 22q11-gebied gene op sy eie, of in kombinasie huisves, wat betrek

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

word by die oorsaak van skisofrenie in 'n sekere proporsie van pasiënte. Die meerderheid van die 22q11-uitwissing is 3Mb in grootte, terwyl 8% 'n kleiner 1.5 Mb uitwissing betrek. Die meeste gene in hierdie gebied is bekend.[8]

Deur gebruik te maak van koppelingsonewewigtigheidanalise (KO) in drie onafhanklike pasiëntgroepe, waarin daar nie-uitwissing was nie, het die navorsers sterk bewys gelewer van 'n assosiasie tussen 'n variasie by die 22q11-lokus en skisofrenie met vroeë aanvangskenmerke. Die drie onafhanklike nie-uitgewissings-pasiëntgroepe het ingesluit: volwasse skisofreniegroep van die VSA, volwasse skisofreniegroep van Suid-Afrika en kinderaanvangskisofreniegroep van die VSA (aanvang voor 13 jarige ouderdom).[8]

Die navorsers het bewys gelewer van 'n bydrae van die PRODH2/DGCR6-lokus in 22q11-geassosieerde skisofrenie.

Navorsers het ook 'n ongewone patroon van die PRODH2-geenvariasie blootgelê. Hierdie variasie boots die sekwensie van 'n gekoppelde pseudogeen na. Verskeie pseudogeen-tipe variante is geïdentifiseer wat lei tot wanwaarneming veranderings by opgegaarde oorblyfsels. Hierdie variante voorkom sintese van 'n volle funksionele ensiem.[8]

Hoe beïnvloed genetiese variasie by die PRODH2/DGCR6-lokus die vatbaarheid vir skisofrenie?

Aanduidend is onlangse studies van 'n diermodel met 'n PRODH-tekort. In 'n vergelyking van 42 muise (homosigoties) vir 'n afgekapte vorm van PRODH met 'n 7 fout verminderde aktiwiteit, en 26 wilde-tipe werpselmaats is 'n tekort in pre-pulsinhibisie aangetoon. Dit is 'n maatstaf vir sensorimotoriese beheer en 'n sentrale inhibisiefiltermeganisme wat aangetas is in pasiënte met psigiatriese versteurings, insluitende skisofrenie.[9]

Prolien is opsigself 'n direkte modulator van glutamergiese transmissie in die brein. Dit is 'n rol wat primêr voorgestel word deur die selektiewe uitdrukking van 'n brein-spesifieke, hoë affiniteit prolienvervoerder in 'n subgroep van glutamergiese sinapse. 'n Alternatiewe model word voorgestel deur onlangse observasies wat PRODH2 impliseer in apoptose.[8]

4.2 Genetiese variasie in die 22q11-lokus en die vatbaarheid vir skisofrenie

Daar is aanduidings van die moontlike koppeling vir skisofrenie op die 22q11 lokus. Die navorsers se hipotese was, terwyl uitwissings van chromosoom 22q11 verantwoordelik is vir slegs 'n klein deel van die skisofrenie-gevalle in die algemene bevolking (tot omtrent 2%), die nie-uitwissingsvariante van individuele gene binne die 22q11 gebied 'n groter bydrae kan lewer tot die vatbaarheid van 'n digte versameling van merkers (gemiddeld 1 enkel-nukleotied-polimorfisme (SNP) / 20 kb oor 1.5 Mb) in die 22q11 lokus gebied. In familiegebaseerde en bevolkingsgebaseerde monsters is sekere resultate opgelewer.[8,10]

Die resultate dui op meer as een geen, wat die siekterisiko wat geassosieer word met hierdie lokus, verhoog. Fyner skaalhaplotipe kartering het twee subgebiede geïdentifiseer, binne die 1.5 Mb-lokus, wat moontlik kandidaat-vatbaarheidsgene vir skisofrenie huisves.[8,10]

Die studie van Liu et al, 2002, in kombinasie met hierdie studie dui op die eerste omvattende toepassing van KO-kartering regoor die 22q11-vatbaarheidslokus rakende skisofrenie. Die implikasies van hierdie studie is dat meer as een geen van hierdie gebied, onafhanklik of sinergisties, die siekte-risiko in nie-uitwissingspasiënte mag verhoog.[8,10]

Die huidige analise dui op 'n beperkte aantal kandidaatgene. Dit sluit in 2-membraan-geassosieerde proteïene K1AA292 en NOGO-R. Beide toon 'n hoë voorkoms in breinareas wat geïmpliseer word in skisofrenie. Hierdie bevindings moet herhaal word in addisionele monsters, en die biologiese funksie van hierdie proteïene moet verder in detail ondersoek word in generasies van dieremodelle.[8,10]

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

4.3 Assessering van die frekwensie van die 22q11-uitwissing in Afrikaner-pasiënte met skisofrenie

Die navorsers het die frekwensie van die 22q11-uitwissing in 'n onafhanklike steekproef van volwasse Afrikanerpasiënte met skisofrenie bepaal. Pasiënte wat in die aanvanklike studie ingesluit was, is gekontak.

Altesaam is daar 85 pasiënte in die studie (52 mans en 33 vroue) ingesluit. Die ondersoek het op fisiese anomalieë, geassosieer met die 22q11-uitwissingsindroom, gefokus. Verder is ook ondersoek ingestel na geestesvertraging, geskiedenis van leerprobleme, of 'n geskiedenis van spesiale skoolopleiding soos voorgestel in die siftingskriteria deur Bassett & Chow, om pasiënte te identifiseer wat 'n verhoogde risiko vir 22q11-uitwissingsindroom het.[11]

Van hierdie 85 pasiënte wat ondersoek is, het 6 pasiënte, 4 mans en 2 vroue, aan twee of meer positiewe siftingskriteria voldoen. 'n FISH-analise is uitgevoer op hierdie 6 pasiënte en twee (een man, een vrou) was draers van homosigotieseuitwissings op chromosoom 22q11.[11]

Die FISH-gebaseerde identifikasie-ratio van 22q11-uitwissingsindroom in hierdie steekproef was 2.4% (2 in 85). Dit is duidelik 'n minimum-skatting, aangesien nie al 85 pasiënte met die FISH-toets/analise ondersoek is nie. In vorige studies soos beskryf in Liu et al (2002 a,b) was die totale groep pasiënte gegenotipeer deur gebruik te maak van 'n paneel van enkel-nukleotied-polimorfismes (SNP's) van die 22q11-lokus.[8,10,11]

In die groep van 85 skisofrenie-pasiënte wat hier beskryf word, het dieselfde twee individue soos geïdentifiseer deur die FISH-toets, wat die 22q11-uitwissing gedra het, homopsigoties getoets vir al die getoetste merkers. Die res van die pasiënte was heteropsigoties vir ten minste een merker-lokus, en gevolglik het dit die teenwoordigheid van 'n tipiese 22q11-mikro-uitwissing uitgeskakel.[11]

Die navorsers het tot die gevolgtrekking gekom dat die ratio van 2 in 85 'n akkurate skatting verteenwoordig en dit ondersteun ons gevolgtrekking dat die akkuraatheid van kliniese sifting vir 22q11-uitwissingsindroom, soos uitgevoer in hierdie studie, 33% is.[11]

Die studie het verder, in 'n onafhanklike steekproef, bevestig dat die frekwensie van die 22q11-uitwissingsindroom in volwasse Kaukasiese skisofreniese pasiënte ongeveer 2% is, soos voorheen voorgestel. Die huidige studie is 'n bevestiging van bestaande literatuur wat die belangrike bydrae van die 22q11-lokus in skisofrenie beklemtoon.[11]

Verder is aangetoon dat die sifting van groepe pasiënte vir kliniese kenmerke, wat sou saamhang met die 22q11-uitwissingsindroom, relatief eenvoudig is om uit te voer. Die sifting identifiseer pasiënte met 22q11-uitwissingsindroom akkuraat. Een derde van geïdentifiseerde gevalle was ware positiewes en daar was geen vals negatiewes nie. Hierdie metode beperk die aantal pasiënte wat 'n FISH-analise kan ondergaan ter bevestiging van hulle status.[11]

4.4 LRHTM1 op chromosoom 2p12 is 'n geen wat aan moederskant onderdruk word en aan vaderskant geassosieer word met handigheid en skisofrenie

Die links-regs-asimmetriese breinfunksie onderlê grootliks menslike kognisie, gedrag en emosie. Abnormaliteite van serebrale asimmetrie word geassosieer met skisofrenie en ander neuropsigiatriese versteurings. Die molekulêre, ontwikkelings, en evolusionêre oorsprong van menslike breinasimmetrie is onbekend.[12]

Ongeveer negentig persent van mense is regshandig. Hierdie is die sterkste bevolkings-vlakverskil in handigheid vir enige primaat. Linkshandigheid in mense word geassosieer met inkorting of omkeer van normale breinasimmetrie, veral van die serebrale kortikale areas wat verband hou met taalpersepsie en -produksie. Hierdie assosiasie is waarskynlik gedeeltelik geneties in etiologie.

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Handigheid en komplekse kognisie in mense mag op 'n ontwikkelings- en 'n evolusionêre vlak verband hou. Die aard en omvang van hierdie verband bly onseker.

Voorheen het genoomwye koppelingsskandering vir 'n kwantitatiewe maatstaf van menslike handigheid 'n koppeling op chromosoom 2p12q11 gevind. Hierdie analise is uitgevoer op 222 leesgestremde sibbe. Daar is geen verband opgespoor tussen leesgestremdheid en handvaardigheid in hierdie groep nie.[12]

Navorsers het hierdie 2p12q11-koppeling, met dieselfde mate van relatiewe handvaardigheid, in 105 pare linkshandige broers ($p=0.0009$) herhaal. Daar is bevind die koppeling tot relatiewe handvaardigheid in die leesgestremde sibbe was in die geheel afkomstig van vaderlike oorerwing van 2p12q11 ($p=0.0000037$). Die moederlike oorgeërfde lokus was nie gekoppel nie. Daar is voorgestel dat die onderliggende geen ingeprent en onaktief of afgereguleer is op die moederlike oorgeërfde chromosoom.

Skisofrenie word geassosieer met inkorting of omkering van normale serebrale asimmetrie. Dit sluit in die mediale temporale lob, boonste temporale girus, planum temporale en die hele brein voorste-agterste embriologiese draaiing. Skisofrenie word geassosieer met 'n verhoogde kans van gemengde- of linkshandigheid.[12]

In 'n onlangse meta-analise van 20 koppelingsskanderings van skisofrenie, was die 2p12q11-lokalisasie die enigste wat beduidend was. Dit is nadat aanpassings gemaak is vir genoomwye toetsing. Hierdie koppeling was hoofsaaklik (maar nie uitsluitlik nie) afkomstig van 'n groep van geaffekteerde sibpare van Europese afkoms. Hulle is versamel in New York en Oxford. Dit is voorheen aangetoon dat die 2p12q11-koppeling van hierdie toetsgroep hoofsaaklik afkomstig is van vaderlike deling en nie van moederskant nie.[12]

In die huidige studie is 'n familiegebaseerde-SNP (enkel-nukleotied-polimorfisme) assosiasiesifting van menslike handigheid gedoen op 222 leesgestremde sibbe. Dit was afkomstig van 'n ouer-van-oorsprong-model, waarin kandidaatgene naby aan die piek van koppeling op chromosoom 2p12 was.[12]

Die assosiasie kartering het 'n haplotipe geïdentifiseer wat die geen LRRTM1 (leucine-rich-repeat-transmembrane neuronal 1) impliseer. Afrikaans: (leusienryk-herhaal-transmembraan neuronaal 1 = LRHTM1).[12]

Verdere genotipering is gebruik om die gebied van assosiasie meer akkuraat te definieer. In 'n studie van 1 002 aangetaste families van Europese afkoms is aangetoon dat dieselfde haplotipe te veel oorgedra word deur die vader na pasiënte met skisofrenie/skiso-affektiewe versteuring. Die 1002 aangetaste families is soos volg saamgestel:[12]

'n Deel van die New York/Oxford-groep (226 families), 'n Ierse "hoë digtheid" -groep (236 families), 'n groep van Montreal (124 families) en 'n Afrikanergroep (416 families).[12]

Uitgebreide funksionele analise dui daarop dat LRRTM1 (LRHTM1) 'n ingeprente, moederlik afgereguleerde geen is. Waarskynlik speel die geen 'n rol gedurende die ontwikkeling van spesifieke voorbreinstrukture, deur neuronale differensiasie en verbinding te beïnvloed. LRHTM1 is 'n ideale kandidaatgeen wat betrokke is by subtiele ontwikkelingsabnormaliteite van die sentrale senuweestelsel.

LRHTM1 is die eerste potensiële genetiese invloed wat met menslike handigheid te doen het, wat identifiseer is. Dit is naas die transkripsie faktor CMO4[12] die tweede geen wat geïmpliseer word in die menslike sentrale senuweestelsel se asimmetriese ontwikkeling, Handigheid, brein-asimmetrie en skisofrenie is etiologie-komplekse kenmerke met baie genetiese en omgewingsinvloede. Gedetailleerde studies van die rol van LRHTM1 in die soogdierbrein se ontwikkeling mag ons tot nuwe insigte bring met betrekking tot die vasstelling en/of onderhoud van normale en abnormale brein-asimmetrie.[12]

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

5. SLOTOPMERKING

Skisofrenie is een van die mees aftakelende en komplekse psigiatriese versteurings. Die ontrafeling van die genetiese aspekte van skisofrenie is besig om 'n keerpunt te bereik. Die posisionele vatbaarheidsgene wat tot dusver geïdentifiseer is, is betrokke by 'n aantal biologiese prosesse, insluitend sinaptiese plastisiteit, neuronale ontwikkeling, neuro- en seinoordrag.[13] Hopelik sal die ondersoek na die genetiese aspekte van skisofrenie in die Afrikaner 'n bydrae lewer tot die beter verstaan en behandeling van hierdie ernstige psigiatriese toestand.

BIBLIOGRAFIE

1. Karayiorgou M., Torrington M., Abecasis GR., Pretorius HW., Robertson B., Kaliski S., Lay S., Sobin C., Maller N., Lundy SL., Blundell ML., Gogos JA., Roos JL. (2004). Phenotypic characterization and genealogical tracing in an Afrikaner schizophrenia database. *American Journal of Medical Genetics*. 124B(1):20-8.
2. Hall D., Wijsman EM., Roos JL., Gogos JA., Karayiorgou M., Joseph A. (2002). Extended intermarker linkage disequilibrium in the Afrikaners. *Genome Research*. 12(6):956-61.
3. Servier S., De Young J., Karayiorgou M., Roos JL., Pretorius HW., Bedoya G et al. (2006). Magnitute and distribution of linkage disequilibrium in population isolates and implication for genome wide association studies. *Nature Genetics*. 38(5):556-60.
4. Roos JL., Pretorius HW., Karayiorgou M., Boraine H. (2006). Cannabis and other variables affecting age at onset in a schizophrenia founder population. *SA Psychiatry Review*. 9:99-103.
5. Seedat F., Roos JL., Pretorius HW., Karayiorgou M., Nel B. (2007). Prevalence and clinical characteristics of obsessive-compulsive disorder and obsessive-compulsive symptoms in Afrikaner schizophrenia and schizoaffective disorder patients. *African Journal of Psychiatry*. 10(10):219-224.
6. Sobin C., Roos JL., Pretorius HW., Lundy LS., Karayiorgou M. (2003). A comparison study of early non-psychotic deviant behavior in Afrikaner and US patients with schizophrenia or schizo-affective disorder. *Psychiatry Res*. 117:113-125.
7. Scholtz MC., Janse van Rensburg MS., Roos JL., Pretorius HW., Karayiorgou M., Levin JB. (2005). Early non-psychotic deviant behaviour as an endophenotypic marker in bipolar disorder, schizoaffective disorder and schizophrenia. *S A Psychiatry Review*. 8:153-9.
8. Liu H., Heath SC., Sobin C., Roos JL., Galke BL., Blundell ML., Lenane M., Robertson B., Sijsman EM., Rapoport JL., Gogos JA., Karayiorgou M. (2002). Genetic variation at the 22q11 PRODH2/DGCR6 locus presents an unusual pattern and increases susceptibility in schizophrenia. *Proc Natl Acad Sci USA*. 99(6):3171-22.
9. Gogos JA., Santha M., Takaes Z., Beck KD., Luine V., Lucas LR., Nadler JV., Karayiorgou M. (1999). The gene encoding praline dehydrogenase modulates sensorimotor gating in mice. *Nat Genet*. 21:434-9.
10. Liu H., Abecasis GR., Heath SC., Knowles A., Demars S., Chen YJ., Roos JL., Rapoport JL., Gogos JA., Karayiorgou M. (2002). Genetic variation in the 22q11 locus and susceptibility to schizophrenia. *Proc Natl Acad Sci USA*. 99(26):16859-64.
11. Wiehahn GJ., Bosch GP., Du Preez RR., Pretorius HW., Karayiorgou M., Roos JL. (2004). Assessment of the frequency of the 22q11 deletion in Afrikaner schizophrenic patients. *The American Journal of Medical Genetics*. 124B(1):1-3.
12. Francks C., Maegawa S., Lauren J., Abrahams BS., Velayos-Baeza A., Medland SE., Colella S., Groszer M., McAuley EZ., Caffrey TM., Timmusk T., Pruunsild P., Koppel I., Lind PA., Matsumoto-Itaba N., Nicod J., Xiong L., Joober R., Enard W., Krinsky B., Roos JL. (2007). LRRTM1 on chromosome 2p12 is a maternally suppressed gene that is associated paternally with handedness and schizophrenia. *Molecular Psychiatry*. 12(12):1129-39.
13. Karoktzou G., Emrich HM., Dietrich DE. (2008). The myelin-pathogenesis puzzle in schizophrenia: aliterature review. *Molecular Psychiatry*. 13:245.

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Clinical Characteristics of an Afrikaner Founder Population Recruited for a Schizophrenia Genetic Study

Johannes Lodewikus Roos,[a] Herman Walter Pretorius,[a] and Maria Karayiorgou[b]

[a] *Department of Psychiatry, University of Pretoria, Weskoppies Hospital, Pretoria, South Africa*

[b] *Department of Psychiatry, Columbia University, New York, New York, USA*

The clinical characteristics of an Afrikaner founder population sample recruited for a schizophrenia genetic study are described. Comparisons on several clinical characteristics between this sample and a U.S. sample of schizophrenia patients show that generalization of findings in a founder population to the population at large is applicable. The assessment of the frequency of the 22q11 deletion in Afrikaner schizophrenia patients is approximately 2%, similar to findings in a U.S. sample. Results of analysis of early non-psychotic deviant behavior in subjects under the age of 10 years in the Afrikaner population broadly replicated findings in a U.S. sample. Approximately half of male schizophrenia patients and a quarter of female patients in the Afrikaner schizophrenia database used or abused cannabis. Male users of cannabis with severe early deviant behavior had the lowest mean age of criteria onset, namely 18.4 years. These findings confirm previous findings, indicating that early deviance is linked to later outcome of disease. The clinical characteristics and premorbid variables in 12 childhood-onset Afrikaner schizophrenia patients thus far recruited in this study compare favorably with what is known about childhood-onset schizophrenia in a U.S. sample. The prevalence of co-morbid OCD/OCS in this Afrikaner schizophrenia founder sample was 13.2% which is in keeping with that of co-morbid OCD in schizophrenia, estimated at 12.2% by the U.S. National Institute of Mental Health. These findings confirm that the clinical characteristics of a schizophrenia sample drawn from the Afrikaner founder population can be generalized to the schizophrenia population at large when compared to findings reported in the literature.

Key words: **clinical characteristics; schizophrenia; Afrikaner founder population; genetic study; 22q11 deletion frequency; early non-psychotic deviant behavior; cannabis use; age of onset of schizophrenia; childhood-onset schizophrenia**

Introduction

Schizophrenia genetic research has advanced greatly with the new tools that are currently available in molecular genetics. Still, the identification of susceptibility genes for common, complex disorders has been a task of unprecedented difficulty. At the core of the difficulty is the clinical heterogeneity in a disease like schizophrenia, as well as the high degree of genetic heterogeneity. There are certain expectations that founder populations may be useful in the circumvention of some of these difficulties. Founder populations offer reduced genetic heterogeneity, more uniform environment, and the potential for genealogical research. Such populations have proven useful for mapping genes that underlie Mendelian disorders and multigenic disorders that segregate

Address for correspondence: J.L. Roos, Professor, Head of Department, Department of Psychiatry, University of Pretoria, Weskoppies Hospital, Private Bag x113, Pretoria, 0001. Voice: +27-12-3199741; fax: +27-12-3199617. erna.fourie@up.ac.za

The Year in Human and Medical Genetics 2009: Ann. N.Y. Acad. Sci. 1151: 85–101 (2009).
doi: 10.1111/j.1749-6632.2008.03453.x © 2009 New York Academy of Sciences.

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

as Mendelian conditions in specific populations.[1–4] Particular features of these populations should also facilitate the search for genes that influence susceptibility to complex traits.

Founder populations originate from a relatively small number of individuals and expand over several generations in relative isolation. As a result, it may be reasonable to expect that fewer independent susceptibility loci segregate in such a population than in an outbred population. At any susceptibility locus, all affected individuals in the population may carry the same or, more likely, a limited set of alleles derived identical-by-descent from a few common ancestors. The reduced locus and allelic heterogeneity in founder populations means that genetic mapping strategies may require smaller sample sizes for success. Furthermore, the relatively recent origin of this population ensures that the chromosomal region surrounding the disease allele is larger than in outbred populations, which allows for relatively sparse genetic maps in these populations to be as informative as denser maps in outbred populations.[5]

As a result, founder populations are more amenable to linkage disequilibrium (LD) approaches, which may be more powerful than traditional linkage analysis for complex traits. It is important to note that the degree of resolution of mapping strategies based on LD between loci will vary between founder populations and depend on the number of generations since the founding.

In one study, the extent, distribution, and magnitude of linkage disequilibrium in 11 population isolates were compared to an outbred Caucasian sample.[6] One of these 11 population isolates was the Afrikaner. The findings of this study demonstrated the following: that population isolates have recently experienced dramatic population growth; and that individuals from a small founding population have overall higher levels of LD than do outbred populations. Conversely, these isolates have far fewer regions of very low LD than do outbred populations. The authors suggested that genome-wide association studies may be conducted with fewer markers in isolates. Furthermore, the authors pointed out that in many isolates there will be few areas in which LD-based association mapping is intractable, while in outbred populations there may be many such areas. In a previous study, supporting the appropriateness of the Afrikaner population for mapping complex traits, the authors have shown that the level of background LD in the Afrikaner sample is detectable at 6 cM.[7]

The Afrikaner population of South Africa is an ideal genetic isolate for the gene mapping studies in schizophrenia. The Afrikaner population meets several criteria that make it an ideal population for mapping complex traits, including: a) founding by a small number of initial individuals that likely allowed for a relatively restricted set of mutations to have been introduced in the population; b) large current population size to allow identification of a sufficient number of cases; and c) strong infrastructure, which allows genealogical research as well as collection of good quality clinical data.[5]

The Afrikaners were founded by approximately 2000 initial settlers who came from the Netherlands and other parts of northern Europe in consecutive waves, starting in 1652. The first immigrants settled in the greater Cape Town area of South Africa. They later spread inland, founding geographically isolated communities. Besides geography, other factors such as cultural considerations, language differences, and religious practices contributed to the isolation of the Afrikaner. Most Afrikaners were members of the Dutch Reformed Church. They lived in large families scattered over the country, so consanguinity was common, especially in early generations. Population growth was almost entirely through reproduction, and immigration subsequent to the founding was minimal. In more recent years, some admixture has occurred. The current size of the Afrikaner population is approximately 3 million and their age of ancestry is an average 12.5 generations.[5]

The demographic history of this population is reflected in the unusually high frequency of certain rare Mendelian disorders (as much as

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

5–10 times higher than in most other population groups). The disorders include porphyria variagata (PV), Huntington's chorea, familial hypercholesterolemia (FH), Fanconi anaemia, and myotonic dystrophy.[8–14] Moreover, unusually low allelic diversity at the associated disease loci has been observed in the Afrikaner. Such a result of genetic drift is expected in a population of this size. Indeed, studies on FH identified only three founder mutations responsible for more of the 95% of FH cases in the Afrikaner population.[15] These Afrikaner mutations were also identified in FH patients in the Netherlands,[16] and one mutation was shown to have originated there and been introduced to South Africa by a single individual of Dutch origin.[17] An unusually large range (8–11 cM) of conserved haplotypes around rare disease genes has also been described in the Afrikaner.[18,19]

Because of these attributes, M. Karayiorgou selected the population for a large-scale genetic study of schizophrenia. This study has been ongoing in close collaboration with the other two authors. The clinical comparability between Afrikaner schizophrenia patients and schizophrenia patient groups in the U.S. was monitored.

Schizophrenia is an etiologically heterogeneous psychiatric disease that exists in familial and non-familial (sporadic) forms. Both forms were recruited in the Afrikaner founder population. Some of the families recruited where three or more first-degree relatives were affected are shown in Fig. 1.[20]

While there is general agreement that using founder populations to map genes for complex traits may be meritorious, concern exists about generalization of findings to the population at large, given the uncertainty about accuracy of diagnosis in disorders that are diagnosed exclusively as clinical syndromes; such is the case for all psychiatric disorders at present.

In this overview we address the clinical comparability and refer to scientific reports published during the collaboration period. Aspects addressed include the following:

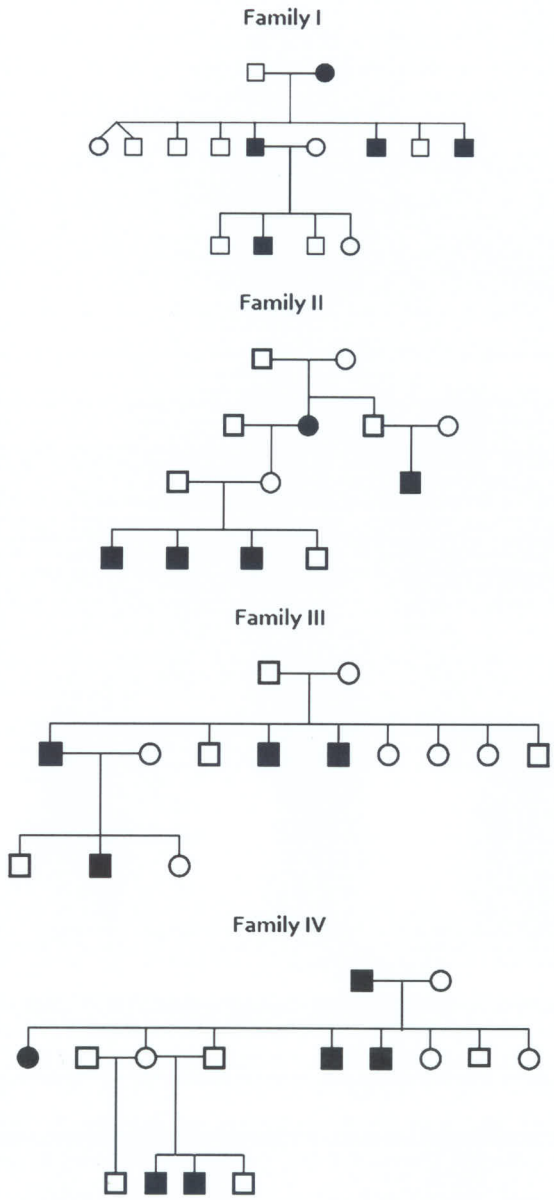


Figure 1. Afrikaner founder population families recruited where three or more first degree relatives were affected.

- Clinical comparability between Afrikaner schizophrenia patients and U.S. schizophrenia patients.
- A comparison of early non-psychotic deviant behavior in Afrikaner and U.S. patients with schizophrenic and schizoaffective disorder.

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

- Cannabis and other variables affecting the age at onset in a schizophrenia founder population.
- Childhood-onset schizophrenia cases identified during this collaboration, and findings in childhood-onset schizophrenia from U.S. studies.
- Prevalence and clinical characteristics of obsessive-compulsive disorder (OCD) and obsessive-compulsive symptoms (OCS) in Afrikaner schizophrenia patients
- Assessment of the frequency of the 22q11 deletion in Afrikaner schizophrenia patients.

Clinical Comparison of Afrikaner to U.S. Schizophrenia Sample

M. Karayiorgou and colleagues reported phenotype characterization and genealogical tracing in an Afrikaner schizophrenia database in 2004.[5] The researchers characterized the demographic and clinical features of a subset of Afrikaner patients (N = 260). These patients were clinically compared to a U.S. schizophrenia sample (N = 238). In both samples only one affected individual per family (proband) was included. Other affected family members were excluded from these analyses, as were probands who were unavailable for interview and thus diagnosed solely on the basis of medical records and/or family reports. For both samples the full Diagnostic Interview for Genetic Studies (DIGS) interview was completed.

Eleven demographic and clinical diagnostic characteristics were indistinguishable across both populations. These included mean age at the time of interview, gender distribution, percentage of left-handed patients, percentage of patients with family history of schizophrenia or other psychiatric diagnosis, percentage of patients with co-morbid substance abuse or dependence, and reported rates of delusions, auditory hallucinations, visual hallucinations, other sensory hallucinations, disorganized behavior, alogia, and flat affect. With regard to the basic sample descriptors and cardinal symptoms of disease, the two populations were comparable.

Three demographic and four clinical variables were different between the Afrikaner and U.S. samples:

a. Afrikaner patients completed fewer years of education (12% vs 4%, $P \leq 0.01$).
b. Fewer Afrikaner patients were single (i.e., never married 74% vs 83%, $P \leq 0.01$).
c. More Afrikaner patients were living with parents or care-taking family members ("living dependently") (80% vs 59%, $P \leq 0.01$).
d. Mean onset age of DSM-IV schizophrenia or schizoaffective disorder was reported to be approximately 3 years later in the Afrikaner sample (23 yrs vs 20 yrs, $P \leq 0.01$).
e. Rates of reported schizoaffective disorder were lower in the Afrikaner sample (26% vs 46%, $P \leq 0.01$).
f. Suicide attempts were reported less frequently among Afrikaner patients with schizophrenia or schizoaffective disorder (28% vs 52%, $P \leq 0.01$).
g. Rates of reported thought disorder were higher in the Afrikaner sample (68% vs 50%, $P \leq 0.01$)

The differences in level of education, marital status, and dependent living arrangements can probably reflect cultural variation. For example, the overall rate of high school completion among the Afrikaner population is lower than in the U.S. The higher percentages of patients living at home as adults may grossly reflect the greater proportion of nuclear intact families in South Africa.

The differences in clinical variables may reflect a higher allelic homogeneity in this population that predisposes to a subtype of the disease with fewer mood symptoms, higher rates of thought disorder, and somewhat later onset. For example, recent analysis of families from the Irish study of high-density schizophrenia

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

families (a more heterogeneous dataset) indicated that affected individuals from families with evidence for linkage to chromosome 8p had significantly more affective deterioration, poorer outcome, more thought disorder, and fewer depressive symptoms than affected individuals from other families in the study.[21] However, the researchers currently could not exclude the possibility that the observed differences in clinical variables were culturally influenced. For example, the later diagnosable onset age may reflect the close-knit family structure in the Afrikaner population, which tends to support probands longer before the family seeks medical attention for the afflicted family member. The difference in rates of schizoaffective disorder between the Afrikaner and U.S. sample could be attributable to more frequent establishment of this diagnosis in the U.S. than in South Africa. Unlike the U.S. patients, relatively few Afrikaner patients carry the schizoaffective diagnosis prior to the research interview. If the possibility of a schizoaffective diagnosis is raised during a research interview, clinicians in South Africa can refer to a complete historical file to confirm the diagnosis, since they have access to extensive clinical data that has been accumulated throughout the patient's lifespan at one hospital. Full records are rarely available in the U.S., where patients tend to move from home and their care over decades often occurs at different hospitals and clinicians. The decreased rates of suicide may be the result of the closer supervised care provided to Afrikaner patients living with family members or in long-term institutionalized hospital settings, an option that still exists in South Africa.[5]

A Comparison of Early Non-psychotic Deviant Behavior

Different sources in the literature have suggested that schizophrenia is a neurodevelopmental disorder with roots in early childhood.[22–24,36] Early deviant behavior patterns in children who go on to develop schizophrenia have been consistently reported in high-risk, prospective, and retrospective behavioral studies of schizophrenia.[26] Prospective longitudinal studies of large cohorts have suggested that during infancy motor development is delayed, and speech problems are more frequent among children who later develop schizophrenia.[27] During the preschool and early school years, studies have reported preference for solitary play, behavioral maladjustment in the classroom, and cooperative impairment in verbal, nonverbal, and mathematic skills at ages 8, 11, and 15.[28,29] Retrospective studies analyzing taped recordings of children revealed that abnormal social reactions, odd movements, and anomalous posture reliably indicated which children would later develop schizophrenia.[29]

Early behavioral deviance appears to occur in only a subgroup of patients.[28,30] Early behavior deviance and the various neurodevelopmental anomalies may in fact be both a subtype and disease-onset marker. These findings have important implications.[30] Understanding the patterns and distribution of early deviant behavior that precede DSM schizophrenia in some patients may be fundamental to accurate characterization of the illness's broad phenotypic heterogeneity. For those patients whose histories include early deviant behavior, the behaviors themselves may provide important clues about brain development and, by extension, possible neurobiologic and neurogenetic underpinnings of schizophrenia.[30]

The U.S. subsample with which the Afrikaner sample was compared was selected from U.S. participants who met lifetime criteria for schizophrenia and schizoaffective disorder. In a previous study, early non-psychotic deviance was described in the U.S. patients.[30] In this sample it was found that 67% of 205 U.S. patients with schizophrenia or schizoaffective disorder reported marked behavioral abnormalities before the age of 10 years. The specific early behavioral abnormalities predicted later aspects of the disease course that were clinically consistent with the early behaviors and also predicted family history of schizophrenia.[30] In

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

spite of the fact that this study was not without limitations (e.g., retrospective report of patients), it seemed useful to further explore non-psychotic early deviant childhood behaviors.

In the comparison study the researchers then compared childhood information, obtained from 109 Afrikaner patients with schizophrenia or schizoaffective disorder, to an age- and gender-matched subsample of patients from the U.S. sample. It was reasoned that if the early childhood deviance data were largely due to reporting error and reflected patients' fluctuating internal states, severity of illness, or cognitive decline, the specific results would not replicate.[30]

In 2001 Sobin and colleagues gave a full description of the U.S. sample,[42] and in 2003 Sobin and colleagues gave a full description of the comparison study and methods followed was given in Sobin *et al.*[31]

The early childhood behavior questionnaire probed seven areas of possible deviance including social dysfunction (avoidance of other children, inability to have friends, isolated play), extreme odd behaviors (unprovoked screaming fits, disorganized or irrational behavior, inappropriate affect), unprovoked aggression, extreme anxiety, chronic sadness, attentional impairment, and learning disabilities.

In order to increase and balance cell sizes' statistical analyses, these seven behaviors were clustered into three logical categories. Variables reflecting socially abnormal behaviors, including social withdrawal, aggression against others, and socially odd behavior, were grouped (n = 33, 30.3%); mood based variables, including extreme anxiety and chronic sadness were combined (n = 38, 34.9%); and variables reflecting cognitive dysfunction, including attentional impairment and learning disability, were combined (n = 43, 39.5%).[31]

The analysis of early non-psychotic deviant behaviors before the age of 10 years broadly replicated the previous findings in the U.S. study. The overall frequencies of behaviors and their distributions were strikingly consistent in the two samples, with 68.8% of Afrikaners compared to 67.8% of U.S. participants reporting early deviance. With regard to the frequencies of individual early deviant behaviors reported, Afrikaners reported more early extreme fears and more attentional impairment, but the frequencies reported did not differ when individual symptoms were clustered. Early deviant behaviors in both populations were meaningfully listed to later disease outcome. However, the populations differed with regard to the specific associations found. Among Afrikaner participants, early extreme fears/chronic sadness were associated with an increased risk of later suicide attempts in the absence of schizoaffective disorder, suggesting more severity of the schizophrenia symptoms manifested. Among U.S. participants, early, extreme fears/chronic sadness were associated with an increased risk of schizoaffective disorder, and not suicidality, as well as sensory hallucinations, suggesting greater clinical heterogeneity in the U.S. sample. It is interesting to note that in the Afrikaner population, early extreme fears/chronic sadness were associated with a reduced frequency of later alogia/flat affect.[31]

One seemingly contradictory finding was that early attentional learning impairment was associated with a significantly lower frequency of family members who manifested schizophrenia. In the original U.S. sample this form of early deviance was strongly associated with an increased frequency of family schizophrenia (the overall rates of family history of schizophrenia did not differ in the two subsamples). In a genetically heterogeneous population the association between early attentional/learning impairment and family history of schizophrenia broadly suggested that those with this form of early deviance might carry a more familial subtype. However, the absence of this association does not necessarily indicate the reverse.[31]

A finding that was consistent across the full sample and the subsample of U.S. patients but not found in the Afrikaner subsample was the association between early extreme fears/chronic sadness and later sensory hallucinations. The substantially higher rate of sensory

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

hallucinations among U.S. participants accounted for by females, and not males, may suggest that this symptom is associated with more heterogeneous forms of schizophrenia.[31]

Sixty-eight percent of patients who met lifetime criteria for schizophrenia/schizoaffective disorder had a history of early childhood deviance. They differentiated a large but distinct subgroup of patients whose disorder may begin decades before the onset of full syndromal illness, as indicated by behaviors that are not included in the current nosology. This form of early deviance is meaningfully linked to later logically related syndrome features, perhaps substantiating such features' validity as precursors to the full disease syndrome. Indeed, initial molecular findings support a genetic basis for this subtype.[78] The value and significance of these symptoms appear to differ in these two populations. These differences may be related to the different genetic background of a given patient population. In the large heterogeneous U.S. sample these indicated earlier onset age and increased family history of schizophrenia.[30] In the more genetically homogeneous Afrikaner population, these behaviors seemed to indicate increased syndrome severity as suggested by the association with markedly increased risk of suicide, poor school achievement, and increased auditory hallucinations. While the specific associations differed, they were all clinically consistent and meaningful. Early fears predicted later syndromal features associated with extreme mood disturbance (suicide of schizoaffective disorder), and early attentional/learning impairment was associated with increased risk of specific types of cognitive disturbance.

In the following section it will be discussed how *Cannabis* and early deviant behavior affected the age at onset in a group of Afrikaner schizophrenia patients.

Cannabis and Other Variables Affecting Age at Onset

People who suffer from psychotic illnesses are far more likely to consume cannabis than the general population.[32] Cannabis has the second most common lifetime prevalence rate as a substance of abuse after alcohol in South Africa.[32] The predominance of cannabis abuse in South Africa is also reflected in statistics gathered from specialist substance treatment centers. Between July and December 2003, 20 to 40% of patients in various substance treatment centers scattered across South Africa reported cannabis to be their primary drug of abuse. The average age of this group was 20 to 21 years.[33]

The question was asked, What percentage of patients included in the ongoing collaborative study on the genetics of schizophrenia in the Afrikaner consume cannabis? Do they consume it to alleviate their symptoms or does cannabis use precipitate their condition?

The nature of this link is not yet fully understood. The findings drawn from clinical samples of subjects identified with schizophrenia have limited value in shedding light on the mechanism underlying this association. In these cases the potential confounding factors linked to the clinical status of the patients are difficult to control.[34]

However, data from clinical samples of subjects suffering from schizophrenia may shed some light on how cannabis use/abuse may affect factors such as age at onset of illness and the longitudinal course of the illness. Furthermore, the interaction of cannabis use/abuse with factors such as gender, early deviant behavior in the first 10 years of life, family history of schizophrenia in first and second degree family members, and early insults (pre-, peri-, and postnatal) can be investigated. The researchers reported some of this information as gathered from 341 patients collected in a genetic database of Afrikaners with schizophrenia.[35]

Of the male patients in this database, 43.54% reported cannabis use/abuse as opposed to 20.45% females. This figure may be higher because cannabis use/abuse is usually underreported and the information collected from the subjects was not confirmed by urine analysis for cannibinoids. More males than females were heavy users of cannabis. Looking at the

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

figures in this database one can assume that approximately half of the male schizophrenia patients and a quarter of the female schizophrenia patients used or abused cannabis. These figures are high and point to consideration of cannabis use/abuse as an important part of the management strategy of the subjects.

Cannabis use/abuse could have an effect on factors such as age of onset and the longitudinal course of the illness. These factors were investigated in this study. The longitudinal course of illness as designated by the DIGS can be applied only after at least one year has elapsed since the initial onset of the active phase symptoms.[36] Keeping this fact in mind, the findings on the longitudinal course of illness in cannabis users versus non-users were not statistically significantly different. They did point towards the fact that cannabis use/abuse may have a negative effect on the longitudinal course of the illness.

It was previously reported that early non-psychotic childhood deviance in an independently collected Afrikaner population distinguished a distinct subtype of patients and that the forms of early deviance manifested were meaningfully linked to the later outcome of the illness.[31] In another comparative study of early non-psychotic deviant behavior in the first 10 years of life in Afrikaner patients with schizophrenia, schizoaffective disorder, and bipolar disorder, it was found that the prevalence of early non-psychotic deviant behavior was not significant in Afrikaner patients with bipolar disorder when compared to the other two diagnostic entities. The authors concluded that early non-psychotic deviant behavior may therefore be used as a possible endophenotypic marker in schizophrenia and schizoaffective disorder, but not in bipolar disorder.[37]

The age at which a subject with schizophrenia first becomes psychotic is a variable trait relating to prognosis. An earlier onset of psychosis is associated with a more severe course, irrespective of the duration of illness.[38] In the Afrikaner cannabis use in schizophrenia study, males with severe early deviant behavior comprised the group with the youngest estimated mean age of criteria onset.[35] Early deviance affected the age of onset of schizophrenia, which is known to be associated with a more severe course of illness. These findings are significant in males with mild and severe early deviant behavior and females with severe early deviant behavior. Male users of cannabis with severe early deviant behavior had the lowest mean age of criteria onset, namely 18.4 years. This can be interpreted as meaning that in males with a distinct subtype of schizophrenia (where early deviance is present) cannabis triggers the onset of a psychosis with a poor prognosis. These findings merit the importance of vulnerable male patients abstaining from cannabis use.[35] These findings reflect previous findings indicating that early deviance was linked to the later outcome of disease.[31]

While previous findings from a study of Finnish families showed that age of onset decreased as family risk of schizophrenia increased, in the Afrikaner database, family history of schizophrenia was less important in predicting the age of criteria onset than: early deviant behavior, the interaction of early deviant behavior and male gender, and the interaction of cannabis use/abuse and male gender.

Twelve Childhood-Onset Afrikaner Schizophrenia Patients—Variables and Characteristics

The study of phenotypic extremes with an early onset has been an important approach to the understanding of disease in all areas of medicine.[39] In different medical illnesses the combination of greater severity, relative treatment resistance, greater heritability, and earlier age of onset has been observed.[40] In Alzheimer's disease and breast cancer, important molecular genetic discoveries have been observed in subpopulations with an earlier age of illness onset.[41,42]

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Most cases of schizophrenia have their onset in late adolescence and early adulthood, but the disorder has been identified in children since early in the 20th century.[43] The nosological status of schizophrenia in children was controversial for many years; however, the landmark studies by Kolvin and colleagues clearly differentiated schizophrenia with onset in childhood from autistic disorder.[44]

The National Institute of Mental Health study of childhood-onset schizophrenia (onset of schizophrenia at the age of 13 years or younger) has provided evidence for clinical and biological continuity between childhood- and adult-onset schizophrenia.[45] This treatment refractory group resembled refractory chronically ill adult patients with schizophrenia.[46]

The current Afrikaner sample includes more than 500 probands of Afrikaner descent who meet the American Psychiatric Association's 1994 diagnostic criteria for schizophrenia or schizoaffective disorder.[47] From this database all those participants having an onset of schizophrenia or schizoaffective disorder at the age of 12 years or younger were selected and their DIGS summaries were obtained. Twelve subjects that had been recruited in this database with childhood-onset schizophrenia were selected. Selected parameters were then tabulated for comparison. The results were reported in an upcoming paper.[48] There were findings unique to this study, and other findings corroborate with what is known about childhood-onset schizophrenia.

Some of the main findings include the following: The male to female ratio is 1:1. This is in contrast to other studies, where there is reported a predominance of males compared to females in childhood-onset schizophrenia.[49,50] It must be kept in mind that the number of childhood-onset schizophrenia patients in this sample is at present very small (n = 12), so it is hard to draw meaningful conclusions. Nevertheless, in this subsample, females had worse symptomatology as they had earlier onset of psychoses, higher incidences of ADHD symptoms, more aggression, and more school achievement difficulties than males. The fact that female puberty occurs earlier than male puberty may be an additional factor to consider for the earlier onset of psychosis in females. All psychosis onset was insidious. This is consistent with other studies.[49]

The treatment of ADHD was always unsuccessful. The rate of ADHD symptoms observed in our study (33%) is similar to another study (40%).[50] The incidence of mental retardation, including borderline mental retardation, was 33%, similar to other studies.[49] When comparing the family history of the childhood onset subgroup to the reported adult Afrikaner schizophrenia subgroup (n = 260), we found a higher incidence of family history of psychotic disease (primary, secondary, and tertiary relatives) in the childhood onset cohort (42%, 5 out of 12 patients) compared to the adult cohort (33.33%). This suggests that the childhood onset cohort has a greater genetic predisposition for schizophrenia than the adult cohort. Another study involving patients with onset of schizophrenia before 18 years of age demonstrated that 12.2% of patients had siblings affected with schizophrenia while the present study demonstrated 8% (1 out of 12 patients).[51]

The non-psychotic deviant behavior before the age of 10 years had an overall rate of 68.8% in the adult Afrikaner schizophrenia cohort compared to the childhood-onset schizophrenia study's 100%.[31] When this early non-psychotic deviant behavior was analyzed further, the study showed 30.28% social functioning impairment (social isolation, aggression, or extreme and odd behavior) compared to the present study's 100%; 34.86% mood and anxiety problems (extreme fears or chronic sadness) compared to the present study's 58% (7 out of 12); 39.45% cognition and behavior problems (attention impairment or learning disability) compared to the present study's 58% (7 out of 12). This shows that the population of the present study (childhood-onset)

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

has much higher rates of non-psychotic deviant behavior as well as social functioning impairment than the adult Afrikaner schizophrenia database. This would fit in with the authors' suggestion that early non-psychotic deviant behavior may be used as an endophenotypic marker for schizophrenia genetic studies.[31]

The attempted suicide rate in the present study of 17% (2 out of 12) is less than the attempted suicide rate of the adult Afrikaner schizophrenia database (n = 260), which was 27.76%.[5]

None of the patients of the childhood-onset study used cannabis at any stage, which differs from the cannabis usage rate of 34.6% of the adult Afrikaner schizophrenia database.[35]

The delusion and hallucination pattern is similar to the adult Afrikaner schizophrenia database (n = 260), 97.26% of adults had delusions compared to 92% of the present study; 79.17% of adults had auditory hallucinations compared to 83% in the present study; 38.89% of adults had visual hallucinations compared to 58% in the present study; and 27.32% of adults had other sensory hallucinations compared to 0% in the present study.[5]

The low rates of school dropout, social difficulty, and treatment resistance found in multidimensional impairment (MDI) patients of the present study is consistent with previous research that pointed out that the psychotic symptoms of MDI patients got better over time and that neuroleptic treatment may not be necessary.[53]

Second generation antipsychotics were prescribed more often to these childhood-onset schizophrenia patients than typical antipsychotics. There is a consensus among previous studies in childhood and adolescent onset schizophrenia that second generation antipsychotics may be more successful in the treatment.[49,50] Risperidone was more often prescribed than Clozapine. Although these patients seemed to be more treatment-resistant, the use of Clozapine remains problematic because of the side-effect profile.

Prevalence and Clinical Characteristics of OCD and OCS

Obsessive–compulsive disorder and obsessive–compulsive symptoms occur in a substantial percentage of patients with schizophrenia or schizoaffective disorder.[54,55] Epidemiological and clinical studies estimate the rate of OC phenomena in schizophrenia at between 3 and 59.2%.[60] This is considerably higher than the rate of 2–3% of OCD/OCS in the general population.[56] The higher than expected co-morbidity rates raise the possibility of a common underlying pathology for both disorders. Recent literature suggests some overlap between schizophrenia and OCD in structural and functional brain abnormalities and the dopamine and serotonin neurotransmitter systems.[57,58]

Distinguishing between delusions, obsessions, ruminations, and preoccupations can be a challenge to those who treat patients with thought disorder. Delusions are described most often as ego-syntonic and are actively embraced intrusion by the patient, whereas obsessions are typically ego-dystonic and are recognized as pathological intrusions.[59] Apart from difficulties in identifying OC symptoms in schizophrenia and schizoaffective disorder, methodological differences result in varying frequencies of reporting co-morbid OCD in schizophrenia. Direct patient interviews lead to higher frequencies of reporting, whereas chart reviews result in lower frequencies.[58] Despite these difficulties the majority of contemporary studies reveal a considerable base rate of OC feature in schizophrenia patients, which supports the validity of the schizophrenia–OCS/OCD association.[60]

A study in South Africa has shown evidence of variation in the prevalence of schizophrenia across ethnic groups.[61] It has also been reported that OCD may be less common in certain communities.[62] In a South African study it was found that the prevalence of co-morbid OCD in a Xhosa-speaking schizophrenia group was significantly lower compared to that found in

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Caucasian schizophrenia patients, raising the question as to whether genetic and cultural factors play a role in the development of co-morbid OCD/OCS in different ethnic groups.[63]

It was found that schizophrenia patients with OCD have a more deteriorating clinical course than schizophrenia patients without OCD.[59] Such co-morbidity is associated with poor treatment response, higher levels of negative symptoms, and greater impairment in levels of functioning. More recent studies agree with previous findings regarding poor outcome in schizophrenia–OCD patients.

With this literature information in mind researchers evaluated the lifetime prevalence and clinical characteristics of OCD and OCS in an Afrikaner schizophrenia and schizoaffective disorder database and compared the clinical patients with co-morbid OCD/OCS to those without co-morbid OCD/OCS. It was hypothesized that the prevalence of OCD/OCS in Afrikaner patients with schizophrenia and schizoaffective disorder would be similar to that described in other studies of Caucasian samples. Furthermore, as no reports of the relationship between the incidence of early deviant non-psychotic behavior in Afrikaner schizophrenia and schizoaffective disorder patients to those with and without co-morbid OCD (OCS were found, this relationship was also assessed in this study).[64]

Four hundred patients with schizophrenia and schizoaffective disorder were assessed. Fifty-three of these subjects were identified to have co-morbid OCD/OCS (study group). This subgroup comprised 40 males and 13 females and was matched for age of onset of illness and gender with a subgroup of 59 subjects—39 male and 20 female patients who did not have OCD/OCS—from the original genetic group (control group). The findings were published in the African Journal of Psychiatry.[64] Significant findings will be discussed here.

The prevalence of co-morbid OCD/OCS among 400 patients with schizophrenia and schizoaffective disorder was 13.2%, which is in keeping with the prevalence of co-morbid OCD and schizophrenia, estimated at 12.2% by the U.S. National Institute for Mental Health, Epidemiologic Catchment Area Study.[56] 10.7% of the study group met the criteria for OCD, whereas 2.5% of patients had OCS not severe enough to fulfill the DSM-IV criteria for OCD. The results supported researchers' hypothesis that the prevalence of OCD/OCS in founder Afrikaner patients with schizophrenia and schizoaffective disorder is significant. The higher prevalence rate in this study is in contrast to the 0.5% co-morbidity of OCD in South African Xhosa-speaking schizophrenia patients.[63] Thus, the possibility of genetic and cultural factors interacting with environmental factors could influence the expression of OCD/OCS in patients with schizophrenia or schizoaffective disorder. Further studies are necessary to explore this association.

With regard to the mean age of onset of OCD/OCS in the study group, the researchers found an earlier age of onset at 18.5 years, compared to the age of onset of schizophrenia or schizoaffective disorder at 22 years. This is in keeping with data found in other studies.[58,65,66]

The analysis of the sociodemographics of the patients suggests that both the schizophrenia OCD/OCS group and schizophrenia group without OCD/OCS have similarly high rates of unemployment and were less often married. Both groups of patients had a less than favorable employment record. These findings are suggestive of the impact that schizophrenia and schizoaffective disorder have on the social functioning of the patients with and without co-morbid OCD/OCS. The impact of these disorders on school functioning is also reflected in the results, with higher percentages of individuals not completing high school.

There were no significant differences between the groups with regard to number of hospitalizations, with the majority of patients reporting more than three admissions since being diagnosed. The frequent hospitalizations further emphasize the impact of the illness on

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

the level of functioning and severity of psychopathology of patients in both groups.

Contamination obsessions, religious obsessions, repetitive rituals, and checking were the most common obsessive-compulsive symptoms reported in the study group. This finding is similar to a previous study.[54]

Substance abuse was significantly higher in the control group without co-morbid OCD/OCS compared to the study group diagnosed with schizophrenia of schizoaffective and co-morbid OCD/OCS. Cannabis was the most common substance of abuse in both groups, followed by alcohol. These findings were in agreement with those found in Pyurovsky *et al.*,[60] where the incidence of substance abuse in the group diagnosed with schizophrenia and OCD was one-half of that found in the non-OCD schizophrenia group. This suggests the possibility of OCD in schizophrenia serving as a protective factor against substance abuse.[56]

The onset of psychotic symptoms was found to be insidious in the study group with co-morbid OCD/OCS compared to the control group. These findings, together with co-morbid depressive symptoms occurring in 73.2% (n = 38) of patients, are of clinical significance as they point to severity of psychopathology and subsequent psychosocial dysfunction.

The main focus of the study was not on the identification of antipsychotic-induced OCD/OCS. However it was found that more than 3/4 of the study group sample (77.4%) used second-generation antipsychotic as compared to first generation antipsychotic (22.6%).

Both study and control groups displayed a similar incidence of early non-psychotic deviance in the first 10 years of life, with a significantly higher incidence of learning disabilities being experienced by the group of patients without co-morbid OCD/OCS. The frequency of these symptoms in all domains, encompassing social function impairment, mood/anxiety, and cognitive impairment in both study and control groups, is significant. The presence of early non-psychotic deviant behavior in Afrikaner schizophrenia or schizoaffective disorder patients is in keeping with findings in previous studies[31] and does not seem to be influenced by a co-morbid diagnosis of OCD/OCS. The presence of early deviance is shown to affect the outcome and prognosis of the disease and the findings may be seen as an indicator of poor prognosis in both the study and control groups.

Assessment of the Frequency of the 22q11 Deletion

One chromosomal region under intense investigation of genes causing schizophrenia is located on the long arm of chromosome 22. A hemizygous deletion of the q11 band on chromosome 22 occurs in 1 of every 5950 live births (0.017%). It has been shown that patients with a hemizygous deletion of the q11 band of chromosome 22 (otherwise diagnosed as 22q11 deletion syndrome, or 22q11DS) are at very high risk for serious psychiatric illness. Studies have shown that schizophrenia was the most common psychiatric disorder found on direct interview using DSM-IV criteria.[67,68] These two studies together suggest that the morbid risk of schizophrenia for a patient with 22q11DS may be approximately 29 times the general population risk and 3 times the risk of a first-degree relative of an individual with schizophrenia.[69]

The 22q11 microdeletions are not detected by routine karyotype analysis, but by the fluorescence *in situ* hybridization (FISH) technique.[70] The 22q11DS occurs most frequently as a sporadic mutation, while inheritance plays a role in 10% of cases. The 22q11DS can have highly variable phenotypic expression. According to Basset and Chow,[71] patients with two or more of the following groups of screening criteria are at increased risk for 22q11DS:

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

1. Hypernasal speech, history of speech therapy, velopharyngeal incompetence, cleft palate (usually submucosal).
2. Characteristic facial features, e.g., long narrow face, narrow palpebral fissures, flat cheeks, prominent nose, small ears, small mouth, retruded chin.
3. Learning difficulties, history of special education, borderline to mild mental retardation.
4. Congenital heart defects, e.g., ventricle septum defect, tetralogy of Fallot, right aortic arch, double aortic arch.
5. Other significant congenital anomalies, e.g., clubfeet, polydactyly, renal anomalies, kyphosis/scoliosis, hypospadia.
6. History of hypocalcaemia and/or hypoparathyroidism.
7. History of arrhythmia or severe immunodeficiency in infancy.

Karayiorgou and colleagues[72] examined a sample of schizophrenia patients for the presence of the 22q11 deletion and identified it at a frequency of 2% in a U.S. sample. This is a substantial increase over the 1:5.950 (0.017%) frequency with which this deletion is found in the general population. Usisken and colleagues[73] estimated the frequency of the 22q11 deletion in a sample of childhood-onset schizophrenia cases (onset before the age of 13) to 6%. Another study screened a large cohort of Japanese adult schizophrenia patients and identified the deletion in 0.33%.[74] Two independent screens of schizophrenia patients who had been pre-selected for presence of one or more of the physical features associated with 22q11DS identified the deletion at rates of 20 and 59%, respectively.[75,76] These studies together have made a strong case for 22q11DS to be a genetic subtype of schizophrenia.[71]

With the above information in mind researchers investigated the occurrence of the 22q11 deletion in an independent sample of schizophrenia patients of Afrikaner origin, as this had not been investigated before.[77]

Patients who had already participated in the ongoing genetic study of the genetics of schizophrenia in the Afrikaner were contacted by telephone and if they would be willing to participate in the 22q11DS study. A total of 85 subjects were recruited and included in the study. The patients recruited included 52 males and 33 females, 17–48 years of age (mean age ± SD, 31.8 ± 10 years). The above mentioned screening criteria suggested by Basset and Chow[71] were applied to identify patients at increased risk for 22q11DS.

Twenty-seven of the 85 subjects examined had one or more of the physical anomalies associated with 22q11DS. Of the 27, six patients (4 males and 2 females) had two or more of the positive screening criteria. Then a FISH analysis was performed on these 6 patients. Two patients (1 male, 1 female) were found to carry hemizygous deletions on chromosome 22q11. The FISH-based identified rate of 22q11DS in this sample (2.4% of 85) is clearly a minimal estimate, since researchers did not test all 85 patients by FISH. However in an independent set of studies described by Liu and colleagues[78] the entire sample was genotyped using a panel of single nucleotide polymorphisms (SNPs) from the 22q11 locus. In the group of 85 schizophrenia patients described here, the same two individuals were apparently "hemizygous" for all tested markers. The rest of the patients were heterozygous for at least one marker, thus excluding the presence of a typical 22q11 microdeletion. It was concluded that the frequency of 22q11 deletion is present in approximately 2% of schizophrenia patients as previously reported[72] and that the accuracy of the clinical screening for 22q11DS, as performed in this study, is 33%. The two patients carrying the deletion had the following positive screening criteria: The male patient had typical dysmorphisms and a ventricular septal defect. The female patient also had typical facial dysmorphisms and high palate/hypernasal speech. Both subjects were born by forceps delivery and experienced feeding problems

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

as infants. Motor development was delayed and they experienced socialization and learning problems since preschool age. The onset of psychosis was insidious and they experienced paranoid delusions, auditory hallucinations, and thought-process disturbance during the acute phase of their illness. Interestingly, depression and obsessions in their late teens were prominent features of both patients' illness. For the male patient, full diagnostic criteria for OCD were met. Both patients were resistant to typical antipsychotic (Haloperidol and trifluoperazine) treatment. Second generation antipsychotic (Risperidone) and combination treatment modalities (Risperidone and SSRIs, e.g., Citalopram) were necessary to control the symptoms.

This study confirmed in an independent sample that the frequency of the 22q11DS in adult Afrikaner schizophrenia patients is approximately 2%. This was previously found by Karayiorgou and colleagues[72] (1995) in a U.S. sample. The Afrikaner 22q11DS study added to the existing literature that strongly supports an important contribution of the 22q11 locus in schizophrenia.

Furthermore, this study also demonstrates that the screening of samples for clinical features compatible with the 22q11DS is relatively simple to perform; it identifies patients with the 22q11DS accurately. In this study 1/3 of the "flagged" cases were true positives, and there were no false negatives. It therefore considerably narrows the sample size in need of confirmation by FISH, which may be a costly investigation if done on all suspected patients with 22q11DS. The criteria used, as suggested by Bassett and Chow,[71] are useful, despite differences in their specificity and predictive value regarding 22q11DS. Some screening criteria include conditions that are rare in the population, but have high predictive value. This includes interrupted aortic arch type B, where one of every two cases carries the 22q11 deletion. Other criteria include conditions that are common in the population but have less specificity or predictive value regarding 22q11DS, such as scoliosis, kyphosis, or learning disabilities.

The sample of schizophrenia patients with 22q11DS in the Afrikaner 22q11DS study is too small (N = 2) to allow any kind of further analysis in terms of symptom quality, symptom severity drug response, age at onset of illness, etc. Studies on larger cohorts of schizophrenia patients with the 22q11 deletion will be important for researchers to more accurately distinguish the clinical phenotype of these patients from other schizophrenia cases.

Conclusion

In this overview the clinical characteristics of an Afrikaner founder population recruited for a schizophrenia genetic study were described. Clinical heterogeneity in a disease like schizophrenia is well known, and there is uncertainty about the accuracy of the diagnosis of psychiatric disorders that are exclusively diagnosed as clinical syndromes. Concerns exist about the generalization of findings in a founder population to the population at large.

The clinical characteristics of Afrikaner schizophrenia patients recruited for an ongoing genetic study on the genetics of schizophrenia compared favorably with a U.S. sample regarding 11 demographic and clinical characteristics. The minor differences in demographics and clinical variables were discussed and possible reasons for these differences were mentioned.

The assessment of the frequency of the 22q11 deletion in Afrikaner schizophrenia patients was approximately 2%, as was found previously in a U.S. sample.

The analysis of early non-psychotic deviant behavior present before the age of 10 years in the Afrikaner population broadly replicated the previous findings in a U.S. sample.

In this Afrikaner schizophrenia database approximately 1/2 of the male schizophrenia patients and a quarter of the female schizophrenia patients used or abused cannabis. Furthermore,

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

it was shown that male users of cannabis with severe early deviant behavior had the lowest mean age of criteria onset, namely 18.4 years. These findings added to previous findings indicating that early deviance was linked to later outcome of disease.

The clinical characteristics and pre-morbid variables in 12 childhood-onset Afrikaner schizophrenia patients thus far recruited in this study compared accurately with present knowledge about childhood-onset schizophrenia in a U.S. sample. The specific findings made in the Afrikaner sample were discussed.

The prevalence of co-morbid OCD/OCS in this schizophrenia Afrikaner founder sample was 13.2%, which is in keeping with the prevalence of co-morbid OCD in schizophrenia, estimated at 12.2% by the U.S. National Institute of Mental Health.

We conclude that the clinical characteristics of an Afrikaner schizophrenia founder population may be generalized to the schizophrenia population at large, as compared to findings in the literature.

References

1. Hastbacka, J. *et al.* 1994. The diastrophic dysplasia gene encodes a novel sulphate transporter, positional cloning by fine structure linkage disequilibrium mapping. *Cell* **78:** 1073–1087.
2. Kalaydjieva, L. *et al.* 1996. Gene mapping in Gypsies identifies a novel demyelinating neuropathy on chromosome 8q24. *Nat. Genet.* **14:** 214–217.
3. Nuystuen, A. *et al.* 1996. A cerebellar ataxia locus identified by DNA pooling to search for linkage disequilibrium in an isolated population from the Cayman Islands. *Hum. Mol. Genet.* **5:** 525–531.
4. Visapaa, I. *et al.* 2002. GRACILE syndrome, a lethal metabolic disorder with iron overload, is caused by a point mutation in BCS1L. *Am. J. Hum. Genet.* **71:** 863–876.
5. Karayiorgou, M. *et al.* 2004. Phenotypic characterization and genealogical tracing in an Afrikaner schizophrenia database. *Am. J. Med. Genet. Part B (Neuropsychiatr. Genet.)* **124B:** 20–28.
6. Service, S. *et al.* 2006. Magnitude and distribution of linkage disequilibrium in population isolates and implications for genome-wide association studies. *Nat. Genet.* **38:** 556–560.
7. Hall, D. *et al.* 2002. Extended intermarker linkage disequilibrium in the Afrikaners. *Genome. Res.* **12:** 956–961.
8. Warnich, L. *et al.* 1996. Identification of three mutations and associated haplotypes in the protoporphyrinogen oxidase gene in South Africa. *Hum. Mol. Genet.* **5:** 981–984.
9. Groenewald, J.Z. *et al.* 1998. Linkage disequilibrium analysis in a recently founded population, evaluation of the variegate porphyria founder in South African Afrikaners. *Am. J. Hum. Genet.* **62:** 1254–1258.
10. Hayden, M.R. *et al.* 1980a. The prevalence of Huntington's chorea in South Africa. *S. Afr. Med. J.* **58:** 193–196.
11. Hayden, M.R. *et al.* 1980b. The origin of Huntington's chorea in the Afrikaner population of South Africa. *S. Afr. Med. J.* **58:** 197–200.
12. Brink, P.A. *et al.* 1987. Familial hypercholesterolemia in South African Afrikaners. Pvu II and Stu I DNA polymorphisms in the LDL-receptor gene consistent with a predominating founder gene effect. *Hum. Genet.* **77:** 32–35.
13. Rosendorff, J. *et al.* 1987. Fanconi anaemia, another disease of high prevalence in the Afrikaans populations of South Africa. *Am. J. Med. Genet.* **27:** 793–797.
14. Goldman, A. *et al.* 1996. Founder effect and the prevalence of myotonic dystrophy in South Africans, molecular studies. *Am. J. Hum. Genet.* **59:** 445–452.
15. Leitersdorf, E. *et al.* 1989. Two common low density lipoprotein gene mutations cause familial hypercholesterolemia in Afrikaners. *J. Clin. Invest.* **84:** 954–961.
16. Defesche, J.C. *et al.* 1993. South African founder mutations in the low-density lipoprotein receptor gene causing familial hypercholesterolemia in the Dutch population. *Hum. Genet.* **92:** 567–570.
17. Defesche, J.C. 1996. Origin and migration of an Afrikaner founder mutation FHAfrikaner-2. V408M. causing familial hypercholesterolemia. *Gene. Georg.* **10:** 1–10.
18. Roby, P. *et al.* 1999. Autosomal dominant. Beukes. Premature degenerative osteoarthropathy of the hip joint maps to an 11cM region on chromosome 4q35. *Am. J. Hum. Genet.* **64:** 904–908.
19. Pronk, J.C. *et al.* 1995. Localisation of the Fanconi anemia complementation group A gene to chromosome 16q23.3. *Nat. Genet.* **11:** 338–340.
20. Roos, J.L. *et al.* 2004. Multiple affected Afrikaner families in a schizophrenia genetic study: environmental risk factors in interaction with genotypes. *S. Afr. Psychiat. Rev.* **7:** 10–14.
21. Kendler, K.S. *et al.* 2000. Clinical features of schizophrenia and linkage to chromosome 5q, 6p, 8p, and 10p in the Irish study of high-density schizophrenia families. *Am. J. Psychiatry* **157:** 402–408.

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

22. Verdoux, H. *et al.* 1997. Obstetric complications and age at onset in schizophrenia: an international collaborative meta-analysis of individual patient data. *Am. J. Psychiatry* **154:** 1220–1227.
23. Cannon, M. *et al.* 1994. Further evidence for anomalies in the handprints of patients with schizophrenia. *Schizophrenia Res.* **13:** 179–184.
24. Lewis, S. 1997. Psychopathology and brain dysfunction: structural imaging studies. *In* Neurodevelopment and Adult Psychopathology. M.S. Kehavan & R.M. Murray, Eds.: Cambridge University Press. Cambridge, England.
25. Lieberman, J. 1999. Is schizophrenia a neurodegenerative disorder? A clinical and neurobiological perspective. *Biol. Psychiatry* **46:** 729–739.
26. Cannon, M. *et al.* 1996. The epidemiology of schizophrenia. *J. Neurol. Neurosurg. Psychiatry* **60:** 604–612.
27. Jones, P. *et al.* 1994. Child development risk factors for adult schizophrenia in the British 1946 birth cohort. *Lancet* **334:** 1398–1402.
28. Done, D.J. *et al.* 1994. Childhood antecedents of schizophrenia and affective illness: social adjustment at ages 7 and 11. *Br. Med. J.* **310:** 57–58.
29. Walker, E.F. *et al.* 1990. The prediction of adult-onset schizophrenia from childhood home movies. *Am. J. Psychiatry* **147:** 1052–1056.
30. Sobin, C. *et al.* 2001. Early non-psychotic deviant behaviour in schizophrenia: a possible endophenotypic marker for genetic studies. *Psychiatry Res.* **101:** 101–113.
31. Sobin, C. *et al.* 2003. A comparison study of early non-psychotic deviant behaviour in Afrikaner and US patients with schizophrenia or schizoaffective disorder. *Psychiatry Res.* **117:** 113–125.
32. Parry, C.D. *et al.* 2001. Trends in adolescent alcohol and other drug use: findings from three sentinel sites in South Africa. *J. Adolesc.* **27:** 429–440.
33. South African Community Epidemiology Network on Drug Use (SACENDU). 2004. Monitoring alcohol and drug abuse trends in South Africa. *SACENDU Res. Brief* **7**. Available at http://www.sahealthinfor.org/admodule/sacendu.htm
34. Verdoux, H. 2004. Cannabis and psychosis proneness. *In* Marijuana and Madness. D. Castle & R. Murray, Eds.: 77. Cambridge University Press. Cambridge.
35. Roos, J.L. *et al.* 2006. Cannabis and other variables affecting age at onset in a schizophrenia founder population. *S. Afr. Psychiatry Res.* **9:** 99–103.
36. Nurnberger, J.I. *et al.* 1994. NIMH genetics initiative. Diagnostic interview for genetic studies, rationale, unique features and training. *Arch. Gen. Psychiatry* **51:** 949.
37. Scholtz, M.C. *et al.* 2005. Early non-psychotic deviant behaviour as an endophenotypic marker in bipolar disorder, schizoaffective disorder and schizophrenia. *S. Afr. Psychiatry Rev.* **8:** 153–157.
38. Veen, N.D. *et al.* 2004. Cannabis use and age of onset of schizophrenia. *Am. J. Psychiatry* **161:** 501–506.
39. Tsuang, M.T. *et al.* 2001. Genes, environment and schizophrenia. *Br. J. Psychiatry* **40**(Suppl): 185–245.
40. Childs, B. *et al.* 1986. Age at onset and cause of disease. *Perspect. Biol. Med.* **29:** 437–460.
41. FitzGerald, M.G. *et al.* 1996. Germ-line BRCA mutations in Jewish and non-Jewish women with early-onset breast cancer. *N. Engl. J. Med.* **334:** 143–149.
42. St George-Hyslop, P.H. *et al.* 1996. Two homologous genes causing early onset familial Alzheimer's disease. *Cold Spring Harb. Symp. Quant. Biol.* **61:** 559–564.
43. Arsarnow, J.R. *et al.* 1994. Annotation: childhood-onset schizophrenia. *J. Child. Psychol. Psychiatry* **35:** 1345–1371.
44. Kolvin, I. 1971. Studies in the childhood psychoses I diagnostic criteria and classification. *Br. J. Psychiatry* **118:** 381–384.
45. Gordon, C.T. *et al.* 1994. Childhood-onset schizophrenia: a NIMH study in progress. *Schizoph. Bull.* **20:** 677–712.
46. Nicolson, R. *et al.* 2000. Interactive report. Lessons from childhood schizophrenia. *Brain Res. Rev.* **31:** 147–156.
47. American Psychiatric Association. 1994. Diagnostic and Statistical Manual. 4th Edition. American Psychiatric Association. Washington, DC.
48. Maydell, R.J. *et al.* 2009. Clinical characteristics and premorbid variables in childhood-onset schizophrenia: a descriptive study of twelve cases from a schizophrenia founder population. *Af. J. Psychiatry* (in press).
49. McClellan, J. *et al.* 2001. Practice parameter from the assessment and treatment of children and adolescents with schizophrenia. *J. Am. Acad. Child. Adolesc. Psychiatry* **40**(Suppl): 45–235.
50. Cosenza, A. *et al.* 2000. CEOS in a six year old girl: Premorbid state, onset, course and therapy. *Psychiatry* **63:** 385–398.
51. McGuffin, P. *et al.* 2002. Psychiatric Genetics and Genomics. Oxford University Press. Oxford.
52. Jacobson, L.K. *et al.* 1998. Childhood-onset schizophrenia, implications for clinical and neurobiological research. *J. Child Psychol. Psychiatr.* **38:** 697–712.
53. Asarnow, J.R. *et al.* 2004. Annotation: childhood-onset schizophrenia: clinical and treatment issues. *J. Child Psychol. Psychiatry* **42:** 180–194.
54. Fenton, W.S. *et al.* 1986. The prognostic significance of obsessive-compulsive symptoms in schizophrenia. *Am. J. Psychiatry* **143:** 437–441.

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

55. Bernman, I. *et al.* 1995. Obsessive and compulsive symptoms in chronic schizophrenia. *Compr. Psychiatry* **36:** 6–10.
56. Karno, M. *et al.* 1998. The epidemiology of obsessive-compulsive disorder in five US communities. *Arch. Gen. Psychiatry* **45:** 1094–1099.
57. Yaryura-Tobias, J.A. *et al.* 1995. Schizophrenia and obsessive-compulsive disorder: shared aspects of pathology. *Neurol. Psychiatry Brain Res.* **3:** 143–148.
58. Tibbo, P. *et al.* 1999. Obsessive-compulsive disorder in schizophrenia: epidemiologic and biologic overlap. *J. Psychiatry Neurosci.* **24:** 15–24.
59. Hwang, M.Y. *et al.* 2000. Management of schizophrenia with obsessive-compulsive disorder. *Psychiatr. Ann.* **30:** 23–28.
60. Poyurovsky, M. *et al.* 2005. Obsessive-compulsive disorder with schizotypy vs schizophrenia with OCD: diagnostic dilemmas and therapeutic implications. *J. Psychiatry Res.* **39:** 399–408.
61. Emsley, R.A. *et al.* 2000. Ethnicity and treatment response in schizophrenia: a comparison of 3 ethnic groups. *J. Clin. Psychiatry* **63:** 9–14.
62. Weismann, M.M. *et al.* 1994. The cross national epidemiology of obsessive-compulsive disorder. The cross national collaborative group. *J. Clin. Psychiatry* **55**(Suppl): 5–10.
63. Niehaus, D.J. *et al.* 2005. Obsessive compulsive disorder prevalence in Zhosa-speaking schizophrenia patients. *SAMJ* **2:** 120–122.
64. Seedat, F. *et al.* 2007. Prevalence and clinical characteristics of obsessive-compulsive disorder and obsessive-compulsive symptoms in Afrikaner schizophrenia and schizoaffective disorder patients. *Afr. J. Psychiatry* **4:** 219–224.
65. Gross-Isseroff, R. *et al.* 2003. Neuroimaging communality between schizophrenia and obsessive compulsive disorder: a putative basis for schizo-obsessive disorder. *World J. Biol. Psychiatry* **4:** 129–134.
66. Poyurovsky, M. *et al.* 2006. Aggregation of obsessive-compulsive spectrum disorder in schizophrenia patients with obsessive-compulsive disorder. *Can. J. Psychiatry* **51:** 746–753.
67. Pulver, A.E. *et al.* 1994. Psychotic illness in patients diagnosed with velo-cardio-facial syndrome and their relatives. *J. Nerv. Mental. Disease* **182:** 476–478.
68. Murphy, K.C. *et al.* 1999. High rates of schizophrenia in adults with velo-cardio-facial syndrome. *Arch. Gen. Psychiatr.* **56:** 940–945.
69. Gottesman, I.I. 1982. Schizophrenia: The Epigenetic Puzzle. Cambridge University Press. Cambridge, UK.
70. Scambler, P.J. 2000. The 22q11 deletion syndrome. *Hum. Mol. Genet.* **9:** 2421–2426.
71. Bassett, A.S. *et al.* 1999. 22q11 Deletion syndrome: a genetic subtype of schizophrenia. *Biol. Psychiatry* **46:** 882–891.
72. Karayiorgou, M. *et al.* 1995. Schizophrenia susceptibility associated with interstitial deletions of chromosome 22q11. *Proc. Nat. Acad. Sci. USA* **92:** 7612–7616.
73. Usiskin, S.I. *et al.* 1999. Velocardiofacial syndrome in childhood-onset schizophrenia. *J. Am. Acad. Child Adolesc. Psychiatry* **38:** 1536–1543.
74. Arinami, T. *et al.* 2001. Screening for 22q11 deletions in a schizophrenia population. *Schizophr. Res.* **52:** 167–170.
75. Gothelf, D. *et al.* 1997. Velocardiofacial manifestations and microdeletions in schizophrenic inpatients. *Am. J. Med. Genet.* **72:** 455–461.
76. Bassett, A.S. *et al.* 1998. 22q11 deletion syndrome in adults with schizophrenia. *Am. J. Med. Genet.* **81:** 328–337.
77. Wiehahn, G.J. *et al.* 2004. Assessment of frequency of the 22q11 deletion in Afrikaner schizophrenia patients. *Am. J. Med. Genet. Part B (Neuropsychiatric Genetics)* **129B:** 20–22.
78. Liu, H. *et al.* 2002a. Genetic variation at the 22q11 PRODH/DGCR6 locus presents an unusual pattern and increases susceptibility to schizophrenia. *Proc. Natl. Acad. Sci. USA* **99:** 3717–3722.

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

ORIGINAL Afr J Psychiatry 2009;12:144-148

Clinical characteristics and premorbid variables in childhood-onset schizophrenia: a descriptive study of twelve cases from a schizophrenia founder population

RJ Maydell, C van der Walt, JL Roos, L Scribante, A Ladikos
Department of Psychiatry, University of Pretoria, Pretoria, South Africa

Abstract

Objective: To analyze clinical and demographic data of childhood-onset (12 years and younger) schizophrenia patients collected for a genetic study in schizophrenia, undertaken nationally in South Africa, using multiple parameters. **Method:** Patients with an onset of schizophrenia at 12 years or younger, were included. From the Diagnostic Interview for Genetic Studies (DIGS), patients' information and summary report data was tabulated and analyzed. Specific subgroups were further compared. This sub-population of 12 subjects was further compared with a group of the adult sample. **Results:** Of the 12 patients recruited, prominent results were: male to female ratio of 1:1; all had insidious onset of psychosis; a third had all 3 multidimensional impairment (MDI) symptoms; all patients that received ADHD treatment had ADHD treatment failure; two thirds had milestone delay; 58% had birth complications; a third were predominantly bottle fed; 42% had family history of schizophrenia; a third had family history of other major psychiatric conditions; all patients had at least one non-psychotic deviant behaviour (NPDB); no patient used cannabis; all delusions were paranoid; 92% had school achievement difficulty and a third had treatment resistance. Gender comparison included: earlier onset of psychosis in females; all females had aggression versus a third of males; more females had school achievement difficulty than males; males had more treatment resistance. Patients with MDI, compared to the sample average had: earlier onset of non-psychotic deviant behaviour; lower school drop-out rate; less social difficulty and no treatment resistance. **Conclusion:** The results compare well to previous research on this topic. The new concepts introduced by the present study require further investigation.

Keywords: Schizophrenia; Childhood; South Africa; Genetics; Population

Received: 03-02-2008
Accepted: 27-06-2008

Introduction

The study of phenotypic extremes with an early-onset has been an important approach to the understanding of disease across all medical disciplines.[1] In different medical illnesses, the combination of greater severity, relative treatment resistance, greater heritability and earlier age of onset has been observed.[2] In Alzheimer's disease and breast cancer, important molecular genetic discoveries have been observed in subpopulations with an earlier age of illness onset.[3,4]

Most cases of schizophrenia have their onset in late adolescence and early adulthood, but the disorder has been identified in children since early in the 20th century.[5] The nosological status of schizophrenia in children was controversial for many years, however the landmark studies by Kolvin and colleagues clearly differentiated schizophrenia with onset in childhood from autistic disorder.[6]

Since 1990, a study of childhood-onset schizophrenia (onset of schizophrenia at the age of 12 or younger) has been ongoing at the National Institute of Mental Health (NIMH).[7] This study shows clear and consistent clinical and biological continuity between childhood-and adult-onset schizophrenia.

Correspondence:
Professor JL Roos
Department of Psychiatry, University of Pretoria,
PO Box 667, Pretoria, 0001, South Africa
email: erna.fourie@up.ac.za

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

As expected, this treatment refractory group resembled refractory, chronically ill adult patients with schizophrenia.[8]

The treatment resistance of patients with schizophrenia is a frequent and important clinical problem. The definition of treatment resistance remains controversial in spite of its importance. Definitions of treatment resistance differ in several domains: dichotomy (response or non-response) versus continuum of response, the type and duration of previous treatment, the psychopathologic symptoms assessed, and whether psychosocial functioning has been evaluated. The approach of Kane et al[9] to operationalize treatment resistance has been followed in many trials exploring new therapeutic strategies for treatment-resistant patients. However, in clinical practice there is probably no need for 3 previous trials with anti-psychotics, usually 2 trials or even 1 can be considered sufficient to show treatment resistance. The duration of level of good functioning and duration of trial periods are not practical for use in childhood-onset schizophrenia.

While obstetric complications may play a role in the development of schizophrenia in some patients, they are not as previously hypothesized, more salient in childhood-onset cases. Other environmental factors, such as socioeconomic status and unusual psychological trauma do not appear clinically to account for the earlier age of onset in this cohort. Similarly, there is no evidence that the earlier age of onset is due to early puberty. Thus, investigation of non-genetic factors for this very early onset cohort has not revealed more salient features than in adult-onset schizophrenia. Indirectly, this suggests that genetic factors may be related to the earlier age of onset.[8]

Patients with childhood-onset schizophrenia have more significant premorbid abnormalities than adult-onset schizophrenics and may also have a greater history of schizophrenia spectrum disorders in their first degree relatives. While eye movement abnormalities are also seen in the families, it is not clear whether the rates are higher than those for relatives of adult-onset schizophrenics. Additionally, there may be a greater rate of cytogenetic abnormalities.[8]

A collaborative study on the genetics of schizophrenia in an Afrikaner founder population has been ongoing since November 1997. Participants are recruited by the Department of Psychiatry at the University of Pretoria, South Africa and the necessary genetic analyses are done by the Human Neurogenetic Laboratory at the Rockefeller University in New York and now at the Columbia University in New York. The clinical database of this study contains information obtained from the Diagnostic Interview for Genetic Studies (DIGS), a chronological summary report of each subject inducing other sources of information pertaining to the subject.[10,11] From this database 12 subjects, thus far recruited with childhood-onset schizophrenia were selected, and data was obtained from the above mentioned sources.

While there is general agreement that using founder populations to map genes for complex traits may be meritorious, concern exists about the generalization of findings to the population at large, given the uncertainty about the accuracy of the diagnosis in disorders diagnosed exclusively as clinical syndromes as all psychiatric disorders are at present.

In this article the authors compare what is known about childhood-onset schizophrenia with the information obtained from these 12 childhood-onset schizophrenia patients from a founder population. Furthermore, these patient's data can be compared to the adult schizophrenia patients recruited for this study.

Certain specific findings were made in this study, whilst other findings compare accurately with what is known about childhood-onset schizophrenia. The implication of the results is debated and further recommendations for further research on this topic is made.

Method

The participants were selected from independently collected samples in South Africa. The current sample includes more than 500 probands of Afrikaner descent who meet diagnostic criteria for schizophrenia or schizoaffective disorder, according to the American Psychiatric Association 1994.[13] All participants provided written informed consent before the initial interview. Approval for the protocol was obtained from both the IRB committees at the University of Pretoria, South Africa and at the Rockefeller University in New York. For full detail of the sample collection, clinical evaluation and genealogical research, please refer to Karayiorgou M et al (2004).[10] The 260 patients discussed in this paper will be the adult reference group for comparison with the childhood onset schizophrenia.

From this database, all the participants that had an onset of schizophrenia or schizoaffective disorder at the age of 12 or younger were selected and their DIGS summaries were obtained. Selected parameters were then tabulated for comparison. The tabulated information was scrutinized and subgroups where chosen to be compared to the sample reported on. The following variables were selected and compared: gender, DSM-IV diagnoses including schizoaffective disorder and schizophrenia, psychosis onset, delusions, hallucinations, multidimensional impairment (MDI) symptomatology (refers to poor affect regulation, attention difficulty and poor impulse control)[12], mental retardation (according to Intelligence Quotient tests), attention deficit hyperactive disorder (ADHD), epilepsy, milestone delays, birth complications, bottle feeding, serious stressors(divorce, moving house, death of a parent, abuse), family history of schizophrenia and other major psychiatric disorders, non-psychotic deviant behaviour (NPDB)[14], learning impairment, language impairment, coordination impairment, suicide, cannabis use, school achievement difficulty, school dropout, social difficulty and treatment resistance. "Extreme odd behaviour" includes unprovoked screaming fits, disorganized or irrational behaviour or inappropriate affect.[15] "Non-psychotic deviant behaviour" (NPDB) includes social functioning impairment (social isolation, aggression or extreme odd behaviour), mood and anxiety problems (extreme fears or chronic sadness) and cognition and behaviour problems (attention impairment or learning disability). "Aggressive patients" in the present study only includes unprovoked aggression.[15] Prolonged labour, forceps delivery, Caesarean section, neonatal jaundice and low Apgar scores were all included in assessing birth complications. Patients were referred to as having treatment resistance if it was stated so by the psychiatrist or child and adolescent psychiatrist treating the patient.

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Results

Twelve early onset schizophrenia patients were recruited. The following clinical and sociodemographic data was obtained from the 12 patients in addition to the key findings noted in Table I. Out of the eleven who had hallucinations: ten had auditory hallucinations; seven had visual hallucinations. No other forms of hallucinations were reported. All delusions were of the paranoid type. When delusions and hallucinations occurred concurrently, hallucinations always started earlier than delusions in each patient, with a median age of onset of 11,09 years versus 14,36 years respectively. Eleven patients had at least one multidimensional impairment disorder (MDI) symptom. All four patients treated for ADHD had ADHD treatment failure, two of which had their ADHD treatment stopped due to adverse effects. It must be appreciated that because of the small sample size (n=12) in this study, the comparisons made were not subjected to statistical analysis.

When looking at the total sample, patients with at least one serious stressor (divorce, moving house, death of a parent or abuse) (6 out of 12) had earlier onsets of psychosis (median age of 10.67 years versus 11.17 years of the total sample).

When further analyses of patients who had non-psychotic deviant behaviour (NPDB) was undertaken: all these patients had social functioning impairment; seven patients had mood and anxiety problems (extreme fears or chronic sadness) and seven patients had cognition and behaviour problems (attention impairment or learning disability).

When further analyses of patients with milestone impairment were undertaken: six patients had delayed walking; two had delayed sitting; four had delayed talking and two had enuresis. A third were predominantly bottle fed. Half of the patients had serious stressors in life (divorce, moving house, death of a parent, abuse).

Regarding treatment, a third were treated with typical (older) antipsychotics (clopenthixol deconate, haloperidol or trifluperazine) of which half (2 out of the 4) were successfully treated. Seven patients were treated with second generation antipsychotics (risperidone, clozapine or olanzepine) of which five of the seven were successfully treated. Two patients were treated with clozapine of which none were successfully treated and a third (4 out of 12) were treated with risperidone of which all four were successfully treated. Two of the four patients with mental retardation had treatment resistance while a third of the total sample had treatment resistance.

Table I: General Findings of the 12 Cases Studied

Males	6
Females	6
Diagnosed with Schizophrenia	11
Diagnosed with Schizoaffective Disorder	1
Insidious onset of psychosis	12
Subjects with delusions	11
Subjects with hallucinations	11
Fulfilled criteria for Multidimensional impairment disorder (MDI)	12
Mental retardation	4
- Borderline mental retardation	1
- Mild mental retardation	2
- Moderate mental retardation	1
- Severe mental retardation	0
Received ADHD Treatment	4
ADHD treatment failed	4
Had non-psychotic deviant behaviour	12
Petit mal epilepsy	3
Other epilepsies	0
Milestone impairment	8
Birth complications	7
Family history of Schizophrenia (primary, secondary or tertiary relatives)	5
Family history of other major psychiatric illness (primary, secondary or tertiary relatives)	4
Attempted suicide	2
Used cannabis	0
School achievement difficulty	11
Dropped out of school	6
Social difficulty	4
Learning impairment	7
Language impairment	2
Coordination impairment	1
Poor affect regulation	8
Poor impulse control	7
Extreme odd behaviours	7
Aggression	8
Treatment resistance	4

Males versus Females

When comparing males (6 out of 12) to females (6 out of 12), the following results were obtained. There was an earlier onset of psychosis in females (median age of 10,67 years for females versus 11,67 years for males). Earlier onset of delusions was seen in females (median age of 11,83 years versus 17,40 years for males). Earlier onset of hallucinations was seen in females (median age of 10,4 years versus 11,67 years for males). Five males had milestone impairment versus three females. Five females had serious stressors versus one male. All females had aggression versus a third of males. Half the females had ADHD symptoms and received treatment for it, versus one male. All the females had school achievement difficulty versus five males and half the males had treatment resistance versus one female.

MDI symptoms

Four patients had all 3 symptoms of MDI and can therefore be called MDI patients. They had the following signs and symptoms. They had earlier onset of delusions (median age of 11,75 years versus 14,36 years of the total sample). They had later onset of hallucinations (median age of 11,75 years versus 11,09 years of the total sample). Delusions and hallucinations started in the same year in each of these four patients. All four patients had extreme odd behaviour versus seven of the total sample. All four patients had learning impairment versus seven of the total sample. None of these patients had language impairment versus two of the total sample. They had earlier onsets of non-psychotic deviant behaviour (NPDB) (median age of 3,00 years versus 4,75 years of the total sample). A quarter dropped out of school versus half of the total sample. A quarter had social difficulty versus a third of the total sample. None had treatment resistance versus a third of the total sample. Half of the total sample dropped out of school. All the patients with treatment resistance dropped out of school.

NPDB at 1 year of age

The patients with non-psychotic deviant behaviour (NPDB) starting at 1 year of age (3 out of 12) had the following: they had an earlier onset of psychosis (median age of 9.67 years versus

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

11.17 years of the total sample); they had earlier onset of delusions (median age of 12,50 years versus 14,36 years of the total sample) and earlier onset of hallucinations (median age of 9,67 years versus 11,09 of the total sample); all (3 out of 3) of these patients had poor affect regulation versus two thirds of the total sample; all (3 out of 3) of these patients had birth complications versus seven of the total sample; all (3 out of 3) of these patients had milestone impairment versus two thirds of the total sample; all (3 out of 3) of these patients had learning impairment versus seven of the total sample; one third (1 out of 3) had language impairment versus two of the total sample; one third (1 out of 3) had coordination impairment versus one patient of the total sample; none had social difficulty versus one third of the total sample.

NPDB at 3 years of age

The patients with NPDB starting at 3 years of age (3 out of 12) had the following. They all had attention difficulty versus two thirds of the total sample. They had poor impulse control versus seven of the total sample. All these patients had ADHD symptoms and received treatment for it versus one third of the total sample. They had earlier onset of delusions (median age of 11,00 years versus 14,36 years of the total sample). These patients had slightly earlier onset of hallucinations (median age of 11,00 years versus 11,09 of the total sample). Delusions and hallucinations started in the same year in each of these patients. Two thirds had social difficulty versus one third of the total sample. None had treatment resistance versus one third of the total sample.

Discussion

It seems that the 1:1, female to male ratio of the present study, is an isolated finding in the childhood-onset-schizophrenia patients of the Afrikaner founder population. Other studies show a predominance of males compared to females, with a ratio of 2:1, in childhood onset schizophrenia.[16,17] One must keep in mind that the number of childhood onset schizophrenia patients in the sample is small which might contribute to the finding. Specifically given that from the total Afrikaner schizophrenia database (this included adults), 31,54% were female (n=260).[10]

In the present study, females had worse symptomatology as they had earlier onset of psychosis, higher incidences of ADHD symptoms, more aggression and more school achievement difficulty than males. The fact that female puberty occurs earlier than male puberty may be an additional factor to consider for the earlier onset of psychosis in females.

A higher incidence of schizophrenia compared to schizoaffective disorder was found when compared to results of an adult sample (N=260).[10] The adult study had 25.58% patients with schizoaffective disorder compared to only 8% (1 out of 12) that was found in the present study. The present study showed that all psychoses onset was insidious. This is consistent with other studies.[16]

The incidence of mental retardation in the present study is similar to other studies that have rates of 10-20%, including borderline mental retardation.[16] These results come from studies of patients with an onset of schizophrenia before 18 years of age. The present study has an incidence of 33%, including borderline mental retardation. Population studies normally show lower rates of mild mental retardation than borderline mental retardation, therefore the finding of higher rates of mild mental retardation in this study may deserve further attention.

In this group of patients, the treatment of ADHD was always unsuccessful. The rate of ADHD symptoms in the present study is similar to another study by Russel et al., 1989 (quoted in Cosenza A, Bruni G, Muratori F, 2000).[17] Their rate was 40% compared to a rate of 33% in the present study. There is a higher incidence of petit mal epilepsy in the present study suggesting that it may be an additional factor to consider in the ethiopathogenesis of childhood onset schizophrenia.

Previous studies have described milestone impairment with predominance of delayed walking just like in the present study.[19] A study involving 61 patients with schizophrenia onsets before 18 years of age, showed a rate of enuresis of 36%.20 This approximates the present study's enuresis rate of 17% (2 out of the 12). In the National Institute of Mental Health (NIMH) study there was no evidence of increased obstetric complications or environmental stress.[21]

When comparing the family history of the present study to the reported adult Afrikaner schizophrenia database (N=260), there was a higher incidence of family history of psychotic disease (primary, secondary and tertiary relatives) in the childhood onset cohort (42%, 5 out of 12) compared to the adult cohort (33,33%). There was a lower incidence of family history of other major psychiatric disease in the childhood onset cohort (33%, 4 out of 12) compared to the adult cohort (68,99%). This suggests that the childhood onset cohort has a greater genetic predisposition for schizophrenia but not for other major psychiatric diseases than the adult cohort. Another study involving patients with an onset of schizophrenia before 18 years of age, had a 12,2% rate of patients with siblings affected with schizophrenia while the present study has 8% (1 out of 12).[22]

The Afrikaner schizophrenia database which reported on non-psychotic deviant behaviour before the age of 10 years, had an overall rate of 68,8% compared to the present study's 100%.[15] When this early non-psychotic deviant behaviour was analyzed further, the study had: 30,28% social functioning impairment (social isolation, aggression or extreme odd behaviour) compared to the present study's 100% (12 out of 12); 34,86% mood and anxiety problems (extreme fears or chronic sadness) compared to the present study's 58% (7 out of 12); 39,45% cognition and behaviour problems (attention impairment or learning disability) compared to the present study's 58% (7 out of 12). This shows that the present study has much higher rates of non-psychotic deviant behaviour as well as social functioning impairment, when compared to the adult Afrikaner schizophrenia database. This would fit in with the authors' suggestion that early non-psychotic deviant behaviour may be used as an endophenotypic marker for schizophrenia genetic studies.[15] NPDB starting at 1 year of age and NPDB starting at 3 years of age appear to have different clinical presentations and may be different entities that should be studied individually.

The attempted suicide rate of 17% (2 out of 12) is less than the attempted suicide rate of the adult Afrikaner schizophrenia database (N=260), which is 27,76%.[10] None of the patients of the present study used cannabis at any stage which greatly differs to the cannabis usage rate of 34.6% of the adult Afrikaner schizophrenia database (N=260).[10]

The delusion and hallucination pattern is similar to the adult Afrikaner schizophrenia database (N=260). 97,26% of adults had delusions compared to 92% of the present study. 79,17% of adults had auditory hallucinations compared to 83% of the

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

present study. 38,89% of adults had visual hallucinations compared to 58% of the present study. 27,32% of adults had sensory hallucinations compared to 0% of the present study.[10] Another study involving patients with a schizophrenia onset of before 18 years (N=110) had 84% delusions and 69% hallucinations. These results are less than the present study's 92% (11 out of 12) delusions and 92% (11 out of 12) hallucinations.[20] The present study confirms the findings of the NIMH study that provided strong evidence that childhood-onset schizophrenia is clinically continuous with later onset forms of the disorder.[21]

The low rates of school dropout, social difficulty and treatment resistance found in MDI patients of the present study is consistent with previous research that pointed out that the psychotic symptoms of MDI patients get better over time and that neuroleptic treatment may not be necessary.[12] Second generation antipsychotics were prescribed more often to these patients than typical antipsychotics. There is a consensus among previous studies in childhood and adolescent onset schizophrenia that second generation antipsychotics may be more successful in the treatment.[16,17] Risperidone was more often prescribed than clozapine. Although these patients seem to be more treatment resistant, the use of clozapine remains problematic because of the side-effect profile.

Limitations

The present study involves a small number of patients, however childhood-onset schizophrenia is a rare entity. Furthermore, the findings uncover certain trends that may be investigated further. Since the information gathered was mainly from patients, recall bias may be a problem. This was dealt with by interviewing family members and using information from patient files (including school reports, medical specialist reports, psychological evaluations and social background information reports) to complement the information from the patient. Trained child and adolescent psychiatrists, with many years of experience, were used to interview the patients and family members to limit the recall bias as much as possible.

Conclusion

The present study analyzed sociodemographic and clinical data of childhood onset (12 years and younger) schizophrenia patients from an Afrikaner schizophrenia database, in South Africa, using multiple parameters. These patients were compared to information obtained from the literature on childhood-onset schizophrenia. Certain groups from the sample were then compared with each other. The main conclusion is that the Afrikaner childhood-onset schizophrenia group is very similar to other groups described in the literature. The results also give some new information, confirmed information from previous studies and allowed for the comparison of adult schizophrenia patients, with childhood onset schizophrenia patients of the same (Afrikaner) founder population. The results are useful in seeking out trends and initiating further research, possibly comparing other psychiatric disease entities like bipolar, ADHD and autism to schizophrenia.

References

1. Tusang MT, Stone WS, Faraone SV. Genes, environment and schizophrenia. Br J Psychiatry 2001; 40(suppl):185-245.
2. Childs B, Scriver CR. Age at onset and causes of disease. Perspect Biol Med 1986; 29:437-60.
3. FitzGerald MG, MacDonald DJ, Kariner M, Hoover I, O'Neil E, Unsal H et al. Germ-line BRCA mutations in Jewish and Non-Jewish women with early-onset breast cancer. N Engl J Med 1996; 334:143-9.
4. St George-Hyslop PH, Levesque G, Levesque L, Rommens J, Westaway D, Fraser PE. Two homologous genes causing early onset familial Alzheimer's disease. Cold Spring Harb Symp Quant Biol 1996; 61:559-64.
5. Arsarnow JR. Annotation: Childhood-onset schizophrenia. J Child Psychol. Psychiat 1994; 35:1345-71.
6. Kolvin I. Studies in the childhood psychoses I Diagnostic criteria and classification. Br J Psychiatry 1971; 118:381-4.
7. Gordon CT, Frazier JA, McKenna K, Giedd J, Zametkin A, Zahn T, et al. Childhood-onset schizophrenia: a NIMH study in progress. Schizoph Bull 1994; 20:677-712.
8. Nicolson R, Lenane M, Hamburger SD, Fernandez T, Bedwell J, Rapoport JL. Interactive report. Lessons from childhood schizophrenia. Brain Research Reviews 2000; 31:147-56.
9. Kane J, Honigfeld G, Singer Jet al. Clozapine for the treatment-resistant schizophrenic. Arch Gen Psychiatry 1988; 45:789-96.
10. Karayiorgou M, Torrington M, Abecasis GR, Pretorius HW, Robertson B, Kaliski S, et al. Phenotypic characterization and genealogical tracing in an Afrikaner schizophrenia database. Am J Med Genetics Part B (Neuropsychiatr Genet) 2004;124B:20-28.
11. Abecasis GR, Burt RA, Hall D, Bochum S, Doheny KF, Lundy SL, et al. Genomewide scan in families with schizophrenia from the founder population of Afrikaners reveals evidence for linkage and uniparental disomy on chromosome 1. Am J Human Genetics. 2004;74:403-417.
12. Asarnow JR, Tompson MC, McGrath EP. Annotation: Childhood-onset schizophrenia: clinical and treatment issues. J Child Psychology and Psychiatry. 2004;42:180-194.
13. American Psychiatric Association. Diagnostic and statistic manual. 4th Edition. Washington, DC: American Psychiatric Association, 1994.
14. Cannon M, Jones P. The epidemiology of schizophrenia. J of Neurology, Neurosurgery and Psychiatry. 1996;60:604-612.
15. Sobin C, Roos JL, Pretorius H, Lundy LS, Karayiorgou M. A comparison study of early non-psychotic deviant behaviour in Afrikaner and US patients with schizophrenia or schizoaffective disorder. Psychiatry Research. 2003;117:113-125.
16. McClellan J, Werry J. Practice Parameter for the Assessment and Treatment of Children and Adolescents With Schizophrenia. J Am Acad Child Adolesc Psychiatry. 2001;40(7 Supplement):4S-23S.
17. Cosenza A, Bruni G, Muratori F. VEOS in a Six Year Old Girl: Premorbod State, Onset, Course and Therapy. Psychiatry. 2000;63:385-398.
18. Roos JL, Pretorius HW, Karayiorgou M, Boraine H. Cannabis and other variables affecting age at onset in a schizophrenia founder population. SAfr Psychiatry Review 2006; 9:99-103.
19. Isohanni M, Jones P, Kemppainen L, Croudace T, Isohanni I, Veijola J, et al. Childhood and adolescent predictors of schizophrenia in the Northern Finland 1966 Birth Cohort – a descriptive life-span model. Eur Arch Psychiatry Clin Neurosci. 2000;250:311-319.
20. Hollis C. Developmental precursors of child- and adolescent-onset schizophrenia and affective psychosis: diagnostic specificity and continuity with symptom dimensions. British Journal of Psychiatry. 2003;182:37-44.
21. Jacobsen LK, Rapaport JL. Childhood-onset schizophrenia; implications of clinical and neurobiological research. J Child Psychol Psychiatr. 1998;38:697-712.
22. McGuffin P, Owen MJ, Gottesman II. Psychiatric Genetics and Genomics. Oxford: Oxford University Press; 2002.

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

articles

Velocardiofacial syndrome – a syndrome with many faces

M Lombard, MB ChB, BA, MFamMed
JL Roos, MB ChB, MMed (Psych), MD, FCPsych (SA)
Department of Psychiatry, University of Pretoria, and Weskoppies Hospital, Pretoria

Among the many features with which velocardiofacial syndrome (VCFS) may present, we review in this paper the literature on working memory deficits, dysmorphic features, co-morbid obsessive-compulsive disorder, early deviant behaviour, functional impairment, and early age of onset of schizophrenia. These features are considered in a case series of 5 patients, and in addition we report on uncommon features. We highlight the importance of appropriate genetic counselling and need for preventive treatments in the future. Considering that covert symptoms in VCFS may easily be overlooked, we call for clinical vigilance in detecting VCFS among patients suffering from schizophrenia using the Bassett and Chow criteria, because the diagnosis of VCFS is of importance in determining prognosis and providing appropriate genetic counselling.

MERGEFIELD backgroundVelocardiofacial syndrome (VCFS), also known as DiGeorge or Shprintzen syndrome, is associated with small deletions of chromosome 22q11. It was first described by Shprintzen *et al.* and has an estimated prevalence of 1 in 5 950 births, making it the second most common genetic syndrome after Down syndrome.[1,2] It has been shown that patients with a hemizygous deletion of the q11 band of chromosome 22 are at a high risk of serious psychiatric illnesses, schizophrenia being the most common of these.[3-8] They have a 29 times higher risk than the general population of developing schizophrenia and 3 times the risk of having a first-degree relative with schizophrenia.[8,9] The familial pattern of schizophrenia has been recognised for a long time.[1,10-12] The 22q11 deletion syndrome (DS) occurs most frequently as a sporadic mutation in schizophrenia, but inheritance plays a role in 10% of cases.

According to Bassett and Chow, patients with two or more of the following groups of screening criteria are at an increased risk for 22q11 DS:

- hypernasal speech, history of speech therapy, velopharyngeal incompetence, cleft palate
- characteristic facial features – long narrow face, small low-set ears, a small mouth and retruded chin
- learning difficulties, history of special education
- congenital heart disease – e.g. ventrical septal defect (VSD) and tetralogy of Fallot
- other significant congenital abnormalities – e.g. club feet, polydactyly, and hypospadias
- history of hypocalcaemia or hypoparathyroidism
- history of athymia.[1,2,8]

Individuals with VCFS also display a distinct neuropsychological phenotype with deficits in executive functioning, visuospatial processing and working memory, which is a system for the short-term storage and manipulation of information.[13-17] In a recent study on DF (16) A1+ mice, which uses a microdeletion on the human chromosome 22 (22q11.2) model, it was suggested that impaired long-range synchrony of neural activity was one consequence of the 22q11 deletion and may be a fundamental component of the pathophysiology underlying schizophrenia.[18,19]

Genetic research on schizophrenia now focuses on genomic copy number variations (CNVs).[18] The 22q11.2, 1q21.1 and 15q13.3 deletions are the only major genetic risk factors identified to date for schizophrenia, and for which clinical genetic assessment and testing are widely available. The relatively modest prevalence of these anomalies suggests to us that at present expensive genomic investigations are not warranted for all individuals with schizophrenia.[18,19,20]

Since 1997, 5 patients with VCFS-type schizophrenia have been identified in an ongoing collaborative study with Columbia University in the USA on schizophrenia in the Afrikaner population.[8] Clinical data were collected from the Diagnostic Interview for Genetic Studies (DIGS) and DIGS Summary Report information. For each of the 5 patients who forms part of this review, we present a short case study describing each patient's unique features. The case studies are followed by short reviews of the literature on a particular feature. Where findings in the literature have a particular bearing on or reflect the unique features of any of these patients, this association is highlighted.

Early deviant behaviour in VCFS

Case 1

John is the oldest of four siblings. His mother was healthy during pregnancy and he was born by forceps delivery. He was born with a VSD, which was later operated on, and defective valves in his urethra. At the age of 4 his frenulum had to be slit because of his indistinct speech. John has experienced problems with socialisation and has been afraid of the dark since the age of 2. He had academic difficulties from the start of primary school, including problems of daydreaming and concentration that remained throughout his school career. He was diagnosed with attention deficit hyperactivity disorder and treated with methylphenidate. He also had periods of extreme sadness and anxiety.

At age 15 he reported that demons were causing him problems, and he experienced auditory hallucinations. After leaving school his functioning gradually deteriorated further and he was unable to finish his tertiary studies. At age 20 he became acutely psychotic and

© University of Pretoria

could not function in the open labour market. He had few friends and his working memory was found to be very poor. Features that point towards a poor prognosis included early onset of illness, obsessive-compulsive disorder (OCD), working memory deficits, and prominent negative symptoms. He fulfilled all the Bassett criteria for VCFS except history of hypocalcaemia and athymia.

Genetic testing confirmed the diagnosis of VCFS. A *de novo* mutation had caused the 22q11 deletion. Two of his siblings were born with craniostenosis, which has been associated with VCFS.

Literature

Early deviant behaviour in children who eventually develop schizophrenia has been reported consistently.[20,21] The domains of early deviant behaviour may be grouped into three clusters: a social functioning impairment cluster, a mood/anxiety cluster, and a cognitive impairment cluster. The social impairment cluster includes social isolation, aggression and extremely odd behaviour. The mood/anxiety cluster includes extreme fear and chronic sadness and the cognitive impairment cluster includes attention impairment and learning disability.[21] All 5 of our patients whose cases are reviewed in this paper displayed early deviant behaviour.

Craniostenosis is a premature fusion of one or more of the cranial sutures. It may be seen as an isolated case or as part of the syndrome. In a study by McDonald-McGinn *et al.*,[22] an association between craniostenosis and 22q11 DS was found.[23-24] It was concluded that craniostenosis should be included in the diagnostic criteria of VCFS. Two of John's siblings had craniostenosis and should be further investigated for 22q11 deletions. The results may have significant implications for providing genetic counselling to them.

Dysmorphic features in VCFS

Case 2

Peter is a 47-year-old man who lives in a chronic care facility. He is the youngest of 4 siblings, and was born with a congenital heart defect. He was socially isolated and his school progress was poor. He completed 2 years of military service and worked as a storeman for 10 years. At age 33, Peter became psychotic and suffered auditory hallucinations and paranoid delusions. After this episode he never worked again. He abused alcohol periodically and tried to commit suicide twice. Since the onset of psychosis he has not been able to function in society.

Peter has a poor prognosis as a result of his prominent negative features as well as the long duration of untreated psychosis. His diagnosis of VCFS was confirmed through genetic analysis, prompted by his meeting several of the Bassett criteria, including the facial dysmorphic features of a prominent nose, small chin, high hard palate, small malformed teeth and low-set ears, and a congenital heart defect. Other features were learning disability, long, tapering fingers, psychosis and nasal speech.

Literature

Dysmorphic features in VCFS include mild auricular abnormalities, hooded eyelids, a tubular-shaped nose, a broad nose tip, a small mouth, and slender and/or tapering fingers. Velopharyngeal insufficiency is one of the most consistent features of this syndrome and is usually confirmed in conjunction with a speech examination. Hypernasal resonance, weak pressure consonants and compensatory articulation are the typical speech symptoms related to velopharyngeal insufficiency.[25] Other features are included in the Bassett criteria.[1,8,19,25]

Dysmorphic features were present in all 5 patients. Although Peter's dysmorphia was clearly observable, in some patients dysmorphia may require a careful examination before they will be recognised.

Pathological laughter in VCFS

Case 3

Julie is 27 years old and was initially referred because of an early onset of illness (she was 16 years old), as well as a treatment-resistant illness that suggested a poor prognosis. Julie is the youngest of 4 children and her siblings are all high achievers academically. She was born at 36 weeks' gestation and was kept in an incubator for 4 days after birth. Her developmental milestones were reached later than expected, and she received occupational therapy as a child. In primary school she had attention and concentration problems and was put in a special class. By age 16, children had reportedly teased her, she had become very anxious, and she was diagnosed with inappropriate laughter syndrome. Magnetic resonance imaging revealed a small lesion in the frontal area, which was thought to have been caused by birth trauma. She objectively hallucinated and suffered from the delusion that she was eating worms, snakes and frogs. She stopped attending school and did not participate in social activities.

Her parents report that at present Julie does not talk, her self-care and functioning are very poor, and she finds changes in her environment or routine distressing. She suffers from fits of uncontrollable laughter that last up to 6 hours, and occur as often as 3 times a day and 3 times a week. During interviews she makes poor eye contact, talks softly to herself, and smiles inappropriately.

Literature

Pathological laughter is a syndrome of recurrent, episodic and involuntary laughter that occurs in the context of various neurological disorders.[26-31] The exact mechanism by which this condition develops is currently unknown; however, there is general agreement that injury to pathways controlling emotional expression is involved. Emotional experience is the subjective feeling during an emotional event, whereas emotional expression is the objective behaviour that is experienced during such event.[27,28] Both the experience and the expression of an emotion are dependent in part upon the cognitive appraisal of the emotional

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

stimuli that are triggering them.[27] Many psychiatric conditions lead to a problem with emotional experience (e.g. mood disorders), and many neurological disorders are associated with dysregulation of emotional expression.[27,28] Pathological laughter and crying are often indistinguishable from normal laughter and crying.[27] Since the outbursts may occur at socially inappropriate times, this problem causes social handicap and suffering.[26,27,31]

The neuro-anatomical basis for pathological laughter is poorly defined. Parvizi *et al.* propose that pathological laughter and crying (PLC) result from a disruption in the tracts originating from the primary motor cortex, which project topographically through the internal capsule and continue through the cerebral peduncles to the basis pontis and cerebellum (the cerebro-ponto-cerebellar pathways).[27,28] They hypothesise that an insult to this pathway results in incomplete information being relayed to the cerebellar structures responsible for the execution of laughter and crying associated with the cognitive and situational context of a given stimulus.

Examples of neurological causes of PLC are pseudo-bulbar palsy (for example due to multiple sclerosis and motor neuron disease), gelastic seizures, Alzheimer's disease, multi-infarct dementia, Parkinson's disease and Down syndrome.[27,29-31] PLC responds to the selective serotonin reuptake inhibitors (SSRIs) that are used for mood disorders.[28,30,31]

Julie's MRI scan showed none of the structural abnormalities associated with neurological causes of PLC. The changes that were evident – namely atrophy and enlargement of the ventricles – are concordant with changes found in schizophrenia. A video electro-encephalogram revealed no abnormality. PLC has not been described in schizophrenia, but we hypothesise that the PLC may be attributed to VCFS and that the same pathways implicated in neurological causes of PLC may be affected in VCFS.

OCD and working memory deficits in VCFS

Case 4

Haily is a 32-year-old divorced woman suffering from schizophrenia and obsessive-compulsive disorder (OCD). She is the oldest of 3 children, and her siblings are all good achievers. Haily's problems began at the age of 18, when she experienced visual hallucinations. After school she attended college and in her first year started hearing voices and thought people where talking about her. She found socialising very difficult and thought that people could read her mind. She had obsessive thoughts about her molesting other people. She attempted suicide impulsively at the age of 30.

Haily is currently a resident in an independent living unit. She functions at a low level and during her most recent interview had a flat affect and laughed inappropriately at times. A Letter-Number Span Test revealed very low scores, suggesting poor working memory.

Literature

OCD is estimated to occur in 13.2% of patients with schizophrenia, which is considerably higher than in the general population (2 - 3%).[32] The higher co-morbidity suggests the possibility of a common underlying pathogenesis for the diseases. There are structural and functional brain abnormalities common to schizophrenia and OCD, including common serotonin and dopamine neurotransmitter pathways. In recent studies it has been found that the clinical course of schizophrenia patients with OCD was more deteriorating than that of schizophrenia patients without OCD.

The schizophrenia-OCD co-morbidity is associated with poor treatment response, higher levels of negative symptoms, and more impairment in functioning.[8] Studies by Gothelf *et al.* suggest that there is a strong association between VCFS and early-onset OCD.[33] The COMT-gene is a candidate gene for schizophrenia because of its role in the breakdown of dopamine in the prefrontal cortex. CThe OMT high activity Val allele is associated with more obsessive compulsive symptoms in young patients with schizophrenia. The genetic contribution to the expression of OCD in patients with schizophrenia lends support to the validity of a putative schizo-obsessive diagnostic entity. The current literature suggests that this co-morbid subgroup has a larger overall illness burden and these patients suffer worse outcomes than patients with schizophrenia without OCD.[32,34]

Regarding working memory in VCFS, measurements of brain activity in patients with schizophrenia suggest that functional connectivity between the frontal and temporal lobe, including the hippocampus, may be altered.[18]

Working memory impairment is increasingly recognised as an important feature in the neuropsychology of schizophrenia. Dysfunction of working memory, a system for the short-term storage and manipulation of information, may relate to a number of core symptoms in schizophrenia.[18] Working memory refers to short-term memory that is a flexible and dynamic memory resource where sensory information can be held briefly and manipulated. A study by Baddeley described working memory as a system for the temporary holding and manipulation of information during a range of neurocognitive tasks such as comprehension, learning and reasoning. This implies an active process rather than simply a static, temporary store.[35]

The dorsolateral prefrontal cortex (DLPC) is thought to play a major role in experience, thought, memory and emotion. The DLPC builds much of its circuitry during adolescence and childhood. Two cell types that play a major role in memory are the pyramidal cells and the chandelier cells.[36] Multiple autopsy studies have shown that in people with schizophrenia, the pyramidal cells have smaller cell bodies and fewer dendritic spines. These cells control especially

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

working memory. Chandelier cells can either exitate or inhibit pyramidal cells and it seems that during childhood these chandelier cells fail to cultivate pyramidal cells. This failure leads to the defect in thought, emotions and memory in schizophrenia.[36]

Cognitive problem deficits play a major role in the functional impairment in schizophrenia, especially the impairment in working memory. This is particularly the case in VCFS.[35] Haily has severe functional impairment, which may be attributed to her working memory deficits as well as her co-morbid OCD. Her functioning has declined in the course of the 10 years since the diagnosis of schizophrenia.

Later age of onset of schizophrenia

Case 5

Jake presented with onset of schizophrenia at the age of 53. He had passivity phenomena, auditory hallucinations, paranoid ideations, and grandiose delusions. He was born with a heart defect and his milestones were retarded. Jake has had difficulty socialising since a young age. In school he had concentration problems and he was moody and irritable. His final year at school was Grade 8. He was described as a 'loner' and his self-care was poor.

Jake married at the age of 35 even though he was homosexual. His wife came from a different cultural group and he lived with her in her cultural group. He was sent to prison for having sexual relationships with minors. He used alcohol socially, but never abused any substances. His mother had a psychiatric illness. Jake met all of the Bassett criteria except a history of athymia, hypocalcaemia and hypoparathyroidism. The diagnosis of VCFS was confirmed by fluorescent *in situ* hybridisation (FISH) analysis.

Literature

Schizophrenia typically has its onset during late adolescence or early adulthood whereas VCFS is associated with the even earlier onset of symptoms.[1,12] Shprintzen *et al.* reported that some of their patients diagnosed with VCFS developed schizophrenia between the ages of 10 and 21.[37] In adult-onset schizophrenia the incidence of VCFS is 2% and in child-onset schizophrenia (onset before the age of 13) it is 6%.[1,8,12]

The late onset of psychosis in VCFS has not previously been reported in the literature, as far as we could establish. Some patients with schizophrenia, however, develop symptoms only in middle age or old age. These patients with late-onset (older than 40) or very late-onset (older than 60) schizophrenia show similar characteristics to those with early-onset schizophrenia with regard to psychopathology, family history, cognitive deficits , nonspecific brain imaging abnormalities and treatment response.38 However, late- and very late-onset schizophrenia psychosis is more often associated with paranoid symptoms, less severe cognitive impairment and a need for lower dosages of antipsychotic medication.[39-44] These patients also tend to have higher levels of education and relatively more of them are married.[41-45] The proportion of patients with schizophrenia who present with late-onset schizophrenia has been estimated at 23%. It is conceivable that factors associated with ageing such as retirement, financial difficulties, bereavement, the death of peers or physical disability may contribute to the precipitation of psychosis in later life.[46]

The mean age of onset of schizophrenia is 10 years later in women than in men. It is proposed that women may have protective factors early in life that result in a later onset of schizophrenia.[47] It is also suggested that adults from migrant populations may be at higher risk.[3,46] In Jake's case, the move to living in a different culture and the stress caused by the move may have precipitated the onset of psychosis.

Conclusion

In future, treatment of schizophrenia should ideally focus more preventing psychosis from developing in vulnerable individuals such as patients with VCFS. New developments in neuro-imaging of infants and children have the potential to improve our understanding of the critical periods in brain development.[48] According to Gothelf *et al.*, early recognition of 22q11DS patients at particularly high risk for psychosis will permit early intervention and decrease disability.[2,3,33] Also, if patients are given a genetic diagnosis for their condition, they may be relieved of guilt or inappropriate blame for their behavioural manifestations of the condition.[2]

With regard to genetic counselling in VCFS it is important to note that there is a 50% chance of transmitting the deletion at each pregnancy, but the caveat is that the severity of the phenotype cannot be predicted.[2] These individuals with VCFS have a 25% chance of developing psychosis.[2] There is prenatal test available for the 22q11 deletion.[2,3,19,33] Patients with VCFS should be seen at regular intervals to screen for early deviant behaviour and prodromal symptoms, which may then be treated appropriately.

Detection of VCFS significantly changes genetic counselling and anticipatory care in comparison with the counselling and care given to other patients with schizophrenia. With respect to psychiatric diagnosis, these findings of copy number variations suggest that, as for the dementias, the schizophrenias may begin to yield to classification based on major causal factors. Comparable to an Alzheimer-type dementia associated with a beta-amyloid precursor protein (APP) gene mutation, a schizophrenia that is related to a 22q11.2 deletion (and perhaps a 1q21.1 or 15q13.3 deletion) may be considered a subtype with distinct management implications.[19]

In our findings, patients with VCFS can have varying clinical pictures and do not always have the same clinical presentation. Almost all of our patients suffered from an early onset of psychosis with poor prognostic features. All of them are functioning poorly – those who are not in hospital are living with their families. Early deviant behaviour as well as working memory deficits were prominent

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

symptoms in all our patients. Nonetheless, VCFS can additionally present with many different faces – pathological laughter, family history of craniostenosis, and late onset of psychosis.

Clinical vigilance should be maintained for the presence of two or more of the Bassett criteria, which indicate a FISH analysis for confirmation of the VCFS diagnosis. It is necessary to do so because this disease has a different prognosis from that for other types of schizophrenia and appropriate genetic counselling is extremely important.

Acknowledgement. This publication was partially made possible by an incentive grant from the NRF to Professor J L Roos.

References

1. Wiehann GJ, Bosch GP, du Preez RR, Pretorius HW, Karayiorgou M, Roos JL. Assessment of the frequency of the 22q11 deletion in Afrikaner schizophrenic patients. Am J Med Genet B Neuropsychiatr Genet 2004;129B (1):20-22.
2. Basset A, Chow EW. 22q11 deletion syndrome: a genetic subtype of schizophrenia. Biol Psychiatry 1999;46(7):882-891.
3. Smith M. The year in human and medical genetics: highlights of 2007-2008. Ann NY Acad Sci 2009;151:1-21.
4. Rockefeller University. Schizophrenia predisposition linked to two genes on chromosome 22 [press release on the Internet]. 19 March 2002. http://runews.rockefeller.edu/index.php?page=engine&id=14Q (accessed March 2010).
5. Bassett AS, Hodgkinson K, Chow EW, et al. 22q11 deletion syndrome in adults with schizophrenia. Am J Med Genet 1998;81:328-337.
6. Scambier PJ. The 22q11 deletion syndromes. Hum Mol Genet 2000;9(16):2421-2426.
7. Murphy KG, Jones LA, Owen MJ. High rates of schizophrenia in adults with velo-cardio-facial syndrome. Arch Gen Psychiatry 1999;56(10):940-945.
8. Roos JL, Pretorius HW, Karayiorgou M. Clinical characteristics of an Afrikaner founder population recruited for a schizophrenia genetic study. Ann NY Acad Sci 2009;1151: 85-101.
9. Liu H, Abecasis GR, Heath SC, et al. Genetic variation in the 22q11 locus and susceptibility to schizophrenia. Proc Nat Acad Sci 2002;99(26):6859-6864.
10. Chackravarti A. A compelling genetic hypothesis for a complex disease: PRODH2/DGCR6 variation leads to schizophrenia susceptibility. Proc Natl Acad Sci USA 2002;99(8):4755-4756.
11. Usiskin SI, Nicolson R, Krasnewich DM, et al. Velocardiofacial syndrome in childhood-onset schizophrenia. J Am Acad Child Adolesc Psychiatry 1999;38(12):1536-1543.
12. Liu H, Heath SC, Sobin C, et al. Genetic variation at the 22q11 PRODH2/DGCR6 locus presents an unusual pattern and increases susceptibility to schizophrenia. Proc Natl Acad Sci USA 2002;99(6):3717-3722.
13. Kates WR, Krauss BR, AbdulSabur N, et al. The neural correlates of non-spatial working memory in velocardiofacial syndrome (22q11.2 deletion syndrome). Neuropsychologia 2007;45(12):2863-2873.
14. Lewandowski KE, Shashi V, Berry PM, Kwapil TR. Schizophrenic-like neurocognitive deficits in children and adolescents with 22q11 deletion syndrome. Am J Med Genet B Neurospychiatr Genet 2007;144B(1):27-36.
15. Sobin C, Kiley-Brabeck K, Daniels S, et al. Neuropsychological characteristics of children with the 22q11 deletion syndrome: a descriptive analysis. Child Neuropsychol 2005;11(1):39-53.
16. Meyer-Lindberg AS, Olsen RK, Kohn PD, et al. Regionally specific disturbance of dorsolateral prefrontal hippocampal functional connectivity in schizophrenia. Arch Gen Psychiatry 2005;62(4):379-386.
17. Karayiorgou M, Simon TJ, Gogos JA. 22q11.2microdeletions: linking DNA structural variation to brain dysfunction and schizophrenia. Nature 2010;11:402-416.
18. Sigurdsson T, Stark KL,Karayiourgou M, Gogos JA, Gordon JA. Impaired hippocampal-prefrontal synchrony in a genetic mouse model of schizophrenia. Nature 2010; 464:736-737.
19. Basset AS, Scherer SW, Brzustowicz LM. Copy number variations in schizophrenia: critical review and new perspectives on concepts of genetics and disease. Am J Psychiatry 2010;167:899-914.
20. Sobin C, Roos JL, Pretorius H, Lundy LS, Karayiorgou M. A comparison study of early non-psychotic behavior in Afrikaner and US patients with schizophrenia or schizoaffective disorder. Psychol Res 2003;117:113-125.
21. Scholtz MC, Janse van Rensburg M, Roos JL, et al. Early non-psychotic deviant behavior as an endophenotypic marker in bipolar disorder, schizoaffective disorder and schizophrenia. South African Psychiatry Review 2002;8:153-159.
22. McDonald-McGinn DM, Gripp KW, Kirscner E, et al. Craniostenosis: Another feature of the 22q11Deletion Syndrome. Am J Med Genet A 2005;136:358-362.
23. Ryan AK, Goodship JA, Wilson DI, et al. Spectrum of clinical features associated with interstitial chromosome 22q11 deletions: a European collaborative study. J Med Genet 1997;34:798-804.
24. Dean JC, De Silva DC, Reardon W. Craniostenosis and chromosome 22q11 deletion. J Med Genet 1998;35:346.
25. Oskarsdottir S, Persson C, Eriksson BO, Fasth A. Presenting phenotype in 100 children with the 22q11 deletion syndrome. Eur J Pediatr 2005;164:146-153.
26. Okuda DT, Chyung ASC, Chin CT, Waubant W. Acute pathological laughter. Mov Disord 2005;20:1389-1390.
27. Parvizi J, Coburn KL, Shillcutt SD, et al. Neuroanatomy of pathological laughing and crying: a report of the American Neuropsychiatry Association Committee on Research. J Neuropsychiatry Clin Neurosci 2009;21(1):75-87.
28. Parvizi J, Arciniegas DB, Bernardini G, et al. Diagnosis and management of pathological laughter and crying. Mayo Clin Proc 2006;81(11):1482-1486.
29. Nadkarni TD, Goel A. Trochlear nerve neurinoma presenting as pathological laughter. Br J Neurosurgery 1999;13(2):212-213.
30. Mendez MF, Nakawatase TV, Brown CV. Involuntary laughter and inappropriate hilarity. J Neuropsychiatry Clin Neurosci 1999;11(2):253-258.
31. Møller M, Andersen G. Inhibition of selective noradrenergic reuptake as treatment of pathological laughter. J Clin Psychopharmacol 2007;27(1):108-110.
32. Seedat F, Roos JL, Pretorius HW, Karayiorgou M, Nel B. Prevalence and clinical characteristics of obsessive-compulsive disorder and obsessive compulsive symptoms in Afrikaner schizophrenia and schizoaffective disorder patients. African Journal of Psychiatry 2007;10(4):219-224.
33. Gothelf D, Presburger G, Zohar AH, et al. Obsessive-compulsive disorder in patients with velocardiofacial (22q11 deletion) syndrome. Am J Med Genet 2004;1268(1):99-105.
34. Bottas A. Comorbidity: schizophrenia with obsessive compulsive disorder. Psychiatric Times [serial on the Internet]. 15 April 2009. http://www.psychiatrictimes.eom/display/articie/10168/1402540 (accessed March 2010).
35. Baddeley A. Working memory. Science 1992;255:556-559.
36. Dobbs D. The making of a troubled mind. Nature 2010;468:154-157.
37. Shprintzen RJ, Goldberg R, Golding-Kushner KJ, Marion RW. Late onset psychosis in the velo cardio facial syndrome .Am J Med Genet 1992;42:141-142.
38. Cohen CL, Vahia I, Reyes P, et al. Focus on geriatric psychiatry: Schizophrenia in later life: clinical symptoms and social wellbeing. Psychiatr Serv 2008;59(3):232-234.
39. Psarros C, Theleritis CG, Paparrigopoulos TJ, Politis AM, Papadimitriou GN. Amisulpiride for the treatment of very late onset schizophrenia like psychosis. Int J Geriatr Psychiatry 2009;24:518-522.
40. Girard C, Simard M. Clinical characteristics of late and very late onset first psychotic episode in psychiatric inpatients. Am J Geriatr Psychiatry 2008;16:478-487.
41. Howard R, Rabins PV, Seeman MV, Jeste DV. Late onset and very late onset schizophrenia-like psychosis: An international consensus. Am J Psychiatry 2000;157:172-178.
42. Schurhoff F, Golmard JL, Szoke A, et al. Admixture analysis of age at onset in schizophrenia. Schizophr Res 2004;71:35-41.
43. Alici-Evcimen Y, Ertan T, Eker E. Case series with late onset psychosis hospitalized in a geriatric psychiatry unit in Turkey. Int Psychogeriatrics.2003;15:69-72.
44. Mazeb D, Zemisblani C, Aizenberg D, Barak Y. Patients with very late onset schizophrenia like psychosis. Am J Geriatr Psychiatry 2005;13:417-419.
45. Sato T, Bottlender R, Schroter A, Moller HJ. Psychopathology of early onset versus late onset schizophrenia revisited: an observation of 473 neuroleptic naïve patients before and after first admission treatments. Schizophr Res 2004;67:175-183.
46. Rabins PV, Lavrisha M. Long term follow-up and phenemologic differences distinguish among late onset schizophrenia, late life depression and progressive dementia. Am J Geriatr Psychiatry 2003;11:589-594.
47. Mitter P, Reeves S, Romero-Rubiales F, et al. Migrant status, age, gender and social isolation in very late onset schizophrenia like psychosis. Int J Geriatr Psychiatry 2005;20:1046-1051.
48. Ross RG. Neuroimaging the infant: the application of modern neurobiological methods to the neurodevelopmental hypothesis of schizophrenia. Am J Psychiatry 2010;167:1017-1018.

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

SCIENTIFIC LETTER | http://dx.doi.org/10.4314/ajpsy.v14i3.9 Afr J Psychiatry 2011;14:242-244

Can we close the barn door before the horses get out? A case study of high genetic loading and subsequent development of psychosis

We would like to report on a 10 year old girl who was referred to a genetic study of schizophrenia in the Afrikaner population because of a high familial genetic loading. This patient subsequently at the age of 16 years was diagnosed with a psychotic disorder. The initial presentation at 10 years of age as well as the course of illness until the diagnosis of psychosis was made will be reported on. Possible predictors of the ultimate development of psychosis will be highlighted and the continuous follow-up of a high-risk case is emphasized.

"A" was a 10 year old grade 4 pupil at a private school when she was referred to us. The mother was diagnosed with schizoaffective disorder, bipolar type and the father with paranoid schizophrenia. Anne's birth and milestones of development were normal. She was raised by a family member. Between 2 – 3 years of age she regressed emotionally, was not sleeping and eating well.

She experienced the following early deviant behaviour during the first 10 years of her life:

- Fear of the dark,
- Struggled to socialize,
- Experienced periods of extreme sadness,
- Daydreaming, and
- Odd behaviour.

She was extremely fussy about the food she would eat. At one stage she was complaining of a terrible taste in her mouth and said she tasted her mother's illness. At school she struggled with reading, spelling and concentration. She failed her grade 2 school year. After the death of her mother and a close friend she became depressed and her school functioning declined even more. A child psychiatrist diagnosed a major depressive disorder and ADHD. She was prescribed citalopram 10 mg daily and methylphenidate. She became hyperactive on the antidepressant, which was subsequently discontinued. The methylphenidate was continued as the teacher could not cope with her in the classroom. She was lost for follow up for 5 years.

Correspondence
Prof. JL Roos
email: Ema.fourie@up.ac.za

At 16 years of age she presented to the state psychiatric services with further information: her highest level of school education was grade 7. She was staying at home with her caregiver. She could not function as a helper at a nursery school. Her menarche was at 11 years of age and she had a bisexual sexual orientation. She started smoking marijuana at 13 years of age and was involved in poly-substance abuse. This was discontinued 5 months prior to this consultation. She was raped at 14 years of age by an older male. Since she started abusing marijuana she was experiencing psychotic symptoms (persecutory delusions & auditory hallucinations). These symptoms worsened over a three year period of time. Her mood was depressed and never had mania for more than a few days. Her self care was poor. She just wanted to sleep.

The following working diagnosis was made at that stage:

- Schizoaffective disorder – bipolar type
- Poly-substance abuse in a patient with borderline personality traits

The duration of untreated psychosis (DUP) was noted as 2+ years. When this patient was evaluated at 10 years of age she was not psychotic. It was suggested that she be observed on a regular basis for the further development of a psychiatric disorder. She was seen as a high risk for the development of psychosis in the future for the following reasons:

- There was a strong family history of psychosis
- Presence of early deviant non psychotic behaviours
- There were odd psychotic like thoughts e.g. tasting her mother's illness
- A possible manic switch on citalopram

Each of these factors led to the subsequent development of psychosis: empiric risk estimates are used for multi-factorial disorders, in which the inheritance patterns are unknown and do not appear to follow the typical mendelian pedigree patterns. Currently, the major genes contributing to mental illness remain to be conclusively established and genetic testing is not yet available. For this reason, recurrence risk estimates are based on empiric risks. (typically, risks based on data from available family studies).[1] By their very nature, empiric risks are summary estimates and thus would not be expected to apply directly to

© University of Pretoria

individual families. Genes may confer risk for a disorder, but the expression of the phenotype and the probability of the disorder occurring may be difficult to predict.[1] When providing recurrence risk for schizophrenia one should provide a range of risks, gathered from multiple sources. Comparisons with general population risks should be offered.[2] Empiric risks are usually available for simple family relationships.[3,4] If two affected parents suffer from schizophrenia a child will have an approximate empirical risk of 45% to develop schizophrenia.[3,4] Early deviant patterns of non-psychotic behaviour in children in their first ten years of life, who go on to develop schizophrenia have been constantly reported in high-risk, prospective and retrospective behavioural studies of schizophrenia.[5] These aberrant behaviours constitute valid endophenotypes for the disease indicating early neurodevelopmental abnormalities.

Deviant childhood behaviour can be grouped into 3 clusters:

- Social functioning impairment cluster
- Mood/anxiety cluster and
- Cognitive impairment cluster

"A" scored high on most of these clusters. The odd thoughts she experienced (e.g. tasting her mother's illness) may have been part of the prodromal phase of schizophrenia. Currently available screening tools have been able to identify a group of patients who are at ultra-high risk of developing schizophrenia, but such tools are nonspecific. Only 25% to 50% of the individuals identified by these tools as being at risk will go on to develop a psychotic disorder within 1 or 2 years.[6] Of the prodromal symptoms depressive symptoms occur on average 5 years before first admission. Depression can be seen as an integral part of the disease process leading to psychosis. A significant decrease in psychosocial functioning especially in young people may be the first sign of severe mental illness.[7]

What about her possible manic switch whilst being treated with citalopram? The STEP-BD study has confirmed the association of rapid cycling with antidepressant use, supporting the viewpoint that these agents can worsen overall illness, causing more mood episodes (including depression), in patients with a rapid-cycling course.[8] This switching had a definite influence on this patient's diagnosis and further management.

Patients using marijuana should be warned about the risk for developing schizophrenia and monitored closely. This use is associated with a two fold increase in the odds of developing schizophrenia and related disorders.[9,10] Recent reviews suggest that marijuana is a partial causal factor.[11] The association between marijuana use and the onset of psychosis is stronger for people with a predisposition for psychotic disorders. Onset of use before the age of 16 has a much stronger effect than if the cannabis use started at a later age.[11] It can be concluded that early marijuana use is a risk modifying factor for psychosis related outcomes.[9]

There is an emerging body of evidence that suggests that the early stage of schizophrenia might provide an opportunity for interventions, even very simple ones that may help to prevent the development of full blown schizophrenia. Studies suggest that delayed treatment leads to poorer long-term outcomes for patients with schizophrenia. The DUP here was more than 2 years.

The goal of prevention of first episode of psychosis is to develop a model of care similar to that used in cardiovascular disease, in which patients with risk factors that increase the likelihood that they will experience cardiac events are identified early and treated to prevent such occurrences.[6] What could have been done for this patient prior to the development of frank psychotic symptoms? Physicians should engage such patients, monitor their condition and provide support. Family targeted therapies can help protect these at-risk individuals who appear to be very vulnerable to stress. The family's role is to buffer the individual from stresses in the community. Simple and safe treatments must be the first priority. Psycho-education on substance use in vulnerable patients is of paramount importance. Treat what is already a problem and do not use antipsychotics before the onset of psychosis.[6]

It has been hypothesized that schizophrenia is a neurodevelopmental disorder in which brain development begins to go awry at around 12 or 13 years of age. Earlier interventions may help to stop progression and improve outcomes which had been described as: "trying to close the barn door before the horses get out".[12] We were accurate with our clinical predictions of this patient at 10 years of age, but the question may be asked: "Did we do enough to close the barn door before the horses got out?"

JL Roos, M Lombard, A Ladikos
Department of Psychiatry, University of Pretoria, Pretoria, South Africa

References

1. Finn CT, Smoller JW. Genetic counselling in psychiatry. Harv Rev Psychiatry. 2006; 14:109-121.
2. Tsuang MT, Stone WS, Faraone SV. Genes, environment and schizophrenia. Br J Psychiatry. 2001; 40(suppl):185-425.
3. Harper PS. Practical genetic counselling. Oxford, Boston: Butterworth-Heinemann. 1998.
4. McGuffin P., Owen MJ, Gottesman II. Psychiatric genetics and genomics. Oxford, New York: Oxford University Press. 2002.
5. Sobin C, Roos JL, Pretorius HW, Lundy LS, Karayiorgou M. A comparison study of early non-psychotic deviant behaviour in Afrikaner and US patients with schizophrenia or schizoaffective disorder. Psychiatry Res. 2003; 117:113-125.
6. Kuehn BM. Early interventions for schizophrenia. Aim to Improve Treatment Outcomes. JAMA. 2010; 304:139-141.
7. Salokangas RK, McGlashan TH. Early detection and intervention of psychosis. A Review. Nord J Psychiatry. 2008; 62:92.
8. Ghaemi NS. Treatment of rapid cycling bipolar disorders: are antidepressant mood destabilisers? AM J Psych. 2008; 165:300-302.
9. McGrath J, Welham J, Scott J, Varghese D, Degenhardt V. Association between cannabis use and psychosis related outcomes using sibling pair analysis in a cohort of young adults. Arch Gen Psych. 2010; 67(5):440-447.
10. Foti DJ, Kotov R, Guey LT, Bromet EJ. Cannabis use and the course of schizophrenia: 10 year follow up after first hospitalization. Am J Psych. 2010; 167:987-993.
11. Korver N, Nieman DH, Becker HE, Van de Fliert FR, Dingemans PH et al. Symptomatology and neuropsychological functioning in cannabis using subjects at ultra high risk for developing psychosis and healthy controls. Australian & New Zealand J Psych. 2010; 44:230-236.
12. Scholtz M, Janse van Rensburg MS, Roos JL, Pretorius HW, Karayiorgou M, Levin JB. Early non-psychotic deviant behaviouir as an endophenotypic marker in bipolar disorder, schizo-affective disorder and schizophrenia. SAfr Psychiatry Rev. 2005; 8:153-159.

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

B13:252

REVIEW Afr J Psychiatry 2011;14:105-111

Genetics of schizophrenia: communicating scientific findings in the clinical setting

JL Roos
Department of Psychiatry, University of Pretoria, Pretoria, South Africa

Abstract
The expected identification of susceptibility genes for psychiatric disorders may bring new opportunities and expectations from patients and families for the clinical translation of research findings in psychiatric genetics. In this article information is provided about familial risk of schizophrenia with the theory behind individualizing risk of recurrence highlighted. Recent new findings regarding the new genetic frontier, Copy Number Variations (CNV), are summarized and the genetic architecture of familial and sporadic schizophrenia applicable to the clinical situation is reviewed. A scenario in which genetic testing could be applied in velocardiofacial syndrome (VCFS) type schizophrenia is debated. Referring to genetic discrimination in mental disorders, reference is made to the implementation of the Federal Genetic Information non-discrimination Act (GINA) of 2008 in the USA and the Mental Health Care Act of 2002 in SA.

Keywords: Schizophrenia; Copy number variations; Familial risk; Individual risk; Velocardiofacial syndrome

Received: 20-05-2010
Accepted: 25-06-2010
doi: 10.4314/ajpsy.v14i2.1

Introduction

In the coming years, the expected identification of susceptibility genes for psychiatric disorders may bring new opportunities and expectations from patients and families for the clinical translation of research findings in psychiatric genetics. There is evidence for possible increasing demand for genetic counselling, particularly if specific genes related to psychiatric disorders are identified.

One area in which familial / genetic research has already been clinically relevant is in providing information about familial risks of illness. The relative inattention to genetic counselling for psychiatric disorders is undoubtly due, in part, to the limitation of available information about genetic risks for common disorders and the inability to individualize risks (due to the absence of genetic testing). Nevertheless, perhaps related to the recent "high profile" of genetics, clinicians are increasingly confronted with questions from patients and families, regarding risks of disorders for themselves or their children. The coming years will likely see significant advances in the clinical application of this knowledge in psychiatric practice.[1]

As scientific advances allow more accurate identification of individual genetic risks, increasing numbers of people will be seeking this kind of help. But professional with genetic expertise often have a poor understanding of psychiatric diagnoses. Given that genetic counselling for complex disorders is among the most challenging topics encountered by genetics professionals, gaps in genetic knowledge may become a major limiting factor in psychiatrists' ability to provide counselling. Hence, better professional education will be needed to improve patient and family education about the genetics of psychiatric disorders.[1]

Having a family history of the disorder is currently the best predictor for the development of several disorders, including schizophrenia and bipolar disorder. Does our current state of knowledge regarding the genetics of schizophrenia support the clinical application of this knowledge in the form of genetic counselling?

Being involved with the genetic study of schizophrenia in the Afrikaner for more than 10 years, I have been approached by family members and patients alike for diagnostic, preventive and reproduction purposes for genetic testing and counselling for schizophrenia. This paper will address

Correspondence
Prof. JL Roos
Weskoppies Hospital, Keljen St, Pretoria West 0183, Pretoria, South Africa
email: c/o erna.fourie@up.ac.za

© University of Pretoria

aspects of the current status of knowledge regarding the genetics of schizophrenia and some of the issues involved in conveying genetic information in the clinical setting.

Intersection of Genetics and Clinical Psychiatry

For most psychiatrists, genetic counselling and the role of genetic contributions to psychiatric illness are unlikely to be a frequent clinical concern. However the internet provides fast, easy and sometimes unreliable medical information e.g. on the genetics of schizophrenia. Patients might engage their psychiatrists in this regard. An awareness of genetic contributions to psychiatric symptoms can at times, have important implications for diagnosis and patient care. For example, the diagnosis and treatment of a patient with first-episode psychosis could be crucially affected by the determination that she has a family history of acute intermittent porphyria. This genetic disorder can present with psychosis and be one where certain psychotropic medications may precipitate acute symptoms. Less dramatically, a strong family history of bipolar disorder might influence drug selection for a patient who present with a first episode psychosis. The two scenarios also highlight a crucial distinction between mendelian, single-gene disorders and complex (polygenic and multifactorial) disorders. Single-gene disorders (such as Huntington's disease) typically follow classical Mendelian patterns of inheritance (e.g. recessive, dominant) and are often highly penetrant (that is, the risk of illness in those carrying the genetic liability is high). In contrast, most cases of psychiatric illness are believed to be genetically "complex". For these disorders, the onset of illness may require the additive or interactive effect of multiple genes; each individual gene contributing a small amount to risk. In addition, additive or interactive effects of environmental risk factors (e.g. stressful life events, smoking of marijuana) may also contribute to the risk. For most types of psychiatric illness, genetic variants may represent risk factors for onset of disease, conferring susceptibility to a disorder without being sufficient to cause it. In addition to making the search for susceptibility genes very difficult, the small individual effects of these genes means that, even if we can identify the genes, the feasibility and value of genetic testing may be limited.[1]

What is the potential role of genetic counselling in psychiatry?

Genetic counselling for psychiatric disorders would be primarily concerned with informing individuals about the role of genetic factors in psychiatric illness, estimating recurrence risks based on family history data, and counselling based on these considerations.[1]

Genetic counselling offers an opportunity to reduce damaging misattributions of etiology. Such interventions must be handled with care to avoid engendering other ideas that may be equally misguided and distressing – for example a sense of guilt at having "transmitted" an illness, or fatalistic ideas about genetic determinism.[2] Genetic counselling may also clarify misperceptions of recurrence risk for affected individuals and their family. If they have overestimated the risk they may be reassured. If they have underestimated the risk, they will suitably informed.[1]

Counselling may have other benefits, including the opportunity to talk about the experience of having a psychiatrically ill family member. The counselling session may be the first opportunity, for unaffected family members, to meet with a medical professional to discuss the disorder of interest. Affected individuals may also wish to discuss the implications of familial psychiatric illness. This education may alter perceptions of disease by dispelling inaccurate beliefs about the course, prognosis and treatment options available for mental illness. The potential risks associated with conveying risk information also need to be considered, such as discrimination (by health or life insurance companies or by employers) increased stigmatization and raising anxiety regarding test results that indicate elevated risk for disease in oneself or family members.[3]

Providing recurrence risk for schizophrenia

Emperic risk estimates are used for multifactorial disorders, in which the inheritance patterns are unknown and do not appear to follow the typical mendelion pedigree patterns. Currently, the major genes contributing to mental illness remain to be conclusively established and genetic testing is not yet available. For this reason, recurrence risk estimates are based on empiric risks. (typically, risks based on data from available family studies).[1]

By their very nature, empiric risks are summary estimates and thus would not be expected to apply directly to individual families. These risks are based on particular ascertained study populations and are not specific to the family in question. In addition, the use of our current diagnostic system may result in the inclusion of a heterogeneous group of disorders that are combined under shared diagnostic labels. This again makes it unlikely that recurrence risks would apply directly to particular families. The complexity of risk estimation is compounded by the common presence of comorbidity in probands or family members since empiric risks accounting for comorbidity may not be available.

An important component of discussing risks involves education about the distinction between classic single-gene inheritance, and complex, multi-factorial inheritance. Psychiatric illness is likely due to the effect of multiple genes in concert with environmental risk factors i.e. genes may confer risk for a disorder, but the expression of the phenotype and the probability of the disorder occurring may be difficult to predict.[1]

When providing recurrence risk for schizophrenia one should provide a range of risks, gathered from multiple sources. Comparisons with general population risks should be offered (Table I).[4,5,6] Empiric risks are usually available for

Table I: Estimated Recurrence Risks and General Population Lifetime risks for Schizophrenia and Bipolar Disorder

Disorder	*General population lifetime risks*	*Estimated empiric FDR recurrence risks*
Schizophrenia	~ 1%	5 ~ 16%
Bipolar disorder	~ 1 – 5%	4 ~ 18% 9 ~ 25%

Source: Reference [5,6]

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Table II: Recurrence Risks for Schizophrenia based on Relationship to Affected Individual

[Empirical risks of developing schizophrenia for relatives of a person with schizophrenia]	
General population	1%
First-degree Relatives	
Two affected parents	~ 45%
MZ twins	40 – 48%
DZ twins	10 – 17%
Children	~ 13%
Parents	6 – 13%
Siblings	6 – 9%
Half-siblings	3 – 6%
Second-degree Relatives	
Grandchildren	3 – 5%
Aunts / Uncles	2 – 3%
Nephews / nieces	2 – 4%
First cousin	2 – 3%

Source: Reference [7,8]

simple family relationships as illustrated in Table II.[7,8] Empiric risks are likely most helpful in families with relatively few affected family members and without comorbid psychiatric illnesses. As an alternative to empiric risks, theoretical recurrence risks can be derived using estimates of certain epidemiologic parameters. These theoretical risks can serve as a guide but do not represent individualized risk estimates. Another important factor in providing recurrence risks for psychiatric disorders is the issue of the accuracy of diagnoses derived from family history and the likelihood that psychiatric symptoms exist on a continuum.

Recent new findings in schizophrenia genetic research

Genetic research in schizophrenia has introduced copy number variation to the field. The content which follows includes clinically applicable findings with reference to a paper on the genetic architecture of familial schizophrenia within a subset of Afrikaner schizophrenia patients. These findings may help to elucidate familial and sporadic cases of schizophrenia during genetic counselling.

Copy Number Variation (CNV), the new genetic frontier

Rare structural changes in the genome have been identified, known as copy number variations. These changes even occur in healthy people. It has been possible to identify these rare structural changes because improved technology has been developed. CNV's occur when genetic material is lost, inserted in the wrong place, or duplicated inappropriately. Most of these variations are not inherited; instead they result from spontaneous mutations in DNA, which happens by chance at conception or during development.[9]

Traditional cytogenetics using light microscopy, and going back over 40 years ago, first showed that variations in chromosome copy number and other structural changes could cause disease in humans. The earliest psychiatric disorder implicated was Downs syndrome caused by the presence of an additional copy of chromosome 21. Many other visible deletions and duplications were later identified associated with disease. There are also a good number of apparently benign variations present in normal individuals. The classic examples in schizophrenia are VCFS, due to a deletion at 22q11 associated with psychosis in 30% of cases. In a large Scottish pedigree a disruption of the DISC1 gene in a balanced 1q43:11q14 translocation associated with schizophrenia and other forms of major mental illness were identified.[9]

An emerging hypothesis is that CNVs may contribute in a more significant way than previously realized to schizophrenia and other psychiatric disorders. CNVs are often unique to an individual. The fact that these structural changes often vary greatly from one person to the next may help explain why it has been so difficult to find genetic patterns for psychiatric disease.[9]

CNV and Schizophrenia

In recent times a series of major articles have reported associations with schizophrenia of CNV at 1q21, 15q11.2, 15q13.3, 16p11.2, 22q12 and neurexin1 loci.[10] What are the implications of these new findings? All the studies have strengths and weaknesses, but when combined - and in contrast to the often contradictory findings in schizophrenia DNA linkage and SNP allelic association studies - they are internally consistent and consistent with each other:

- These studies show beyond doubt that carriers of some de novo and inherited CNVs, especially when they can be shown to disrupt gene function, are at high risk of developing schizophrenia. In the great majority of cases, the schizophrenia is phenotypically indistinguishable from schizophrenia as it presents to psychiatric clinics the world over. In a small proportion there may be associated features such as learning disability or VCFS, and in these cases the schizophrenia can be considered syndromic.[10]
- Deletions and/or duplications of CNVs at 1q21, 15q11.2, 15q13.3, 16p11.2 and neurexin1 loci substantially increase the risk of a broad range of major psychiatric disorders in addition to schizophrenia i.e. mental retardation, autism, bipolar disorder and attention deficit/hyperactivity disorder.
- The findings help to explain why some cases of schizophrenia appear familial and other cases appear sporadic. Sporadic cases are much more likely to be caused by de novo events. The high rate of de novo events probably also explains the well-known observation of higher risk of schizophrenia in the children than in the siblings of schizophrenia probands.[10]
- These new discoveries have important implications for genetic counselling. Risk of schizophrenia in siblings of individuals with schizophrenia is far less if the proband carries de novo CNV mutation. Many CNVs disrupt several genes, only one of which may predispose to

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

schizophrenia. Others may however impact on the physical health of individuals with schizophrenia and could in part explain the poorer health outcomes associated with schizophrenia.[10]

- CNVs account for 2 – 4% of schizophrenia's overall genetic burden. As many more novel loci are identified by newer high-resolution platforms, they may account for 10 – 20%.[10]
- Determining the precise nature of the CNVs and what the genetic/epigenetic and environmental factors are that influence their penetrance and expressivity and cause them to increase risk for a broad range of neurodevelopmental disorders will be an exciting challenge for neuropsychiatric research in coming years.[10]

Genetic architecture of Familial schizophrenia

The strongest predictor of schizophrenia is having an affected first degree relative. In addition to the familial forms, nonfamilial (sporadic) forms of the disease also exist. The exact proportion of each form is largely unknown, but it is thought that at least 60% of cases are sporadic.[11]

Genome-wide linkage scans have been conducted to identify loci harbouring relatively rare mutations / variants that increase susceptibility to familial schizophrenia. Loci have been identified on almost every chromosome, but only a few regions have been replicated across studies. One such region is near the telomere of chromosome 13q. This region has also been linked to bipolar disorder. In addition to linkage studies, a number of earlier as well as more recent studies have provided strong evidence supporting the importance of rare structural mutations / variants in schizophrenia vulnerability. Rare inherited structural lesions are expected to be prominent in familial schizophrenia. The full contribution of these structural lesions to transmitted liability in familial schizophrenia cases was not examined in a systematic manner until the study by Xu et al undertaken amongst the Afrikaner population in South Africa (SA).[11]

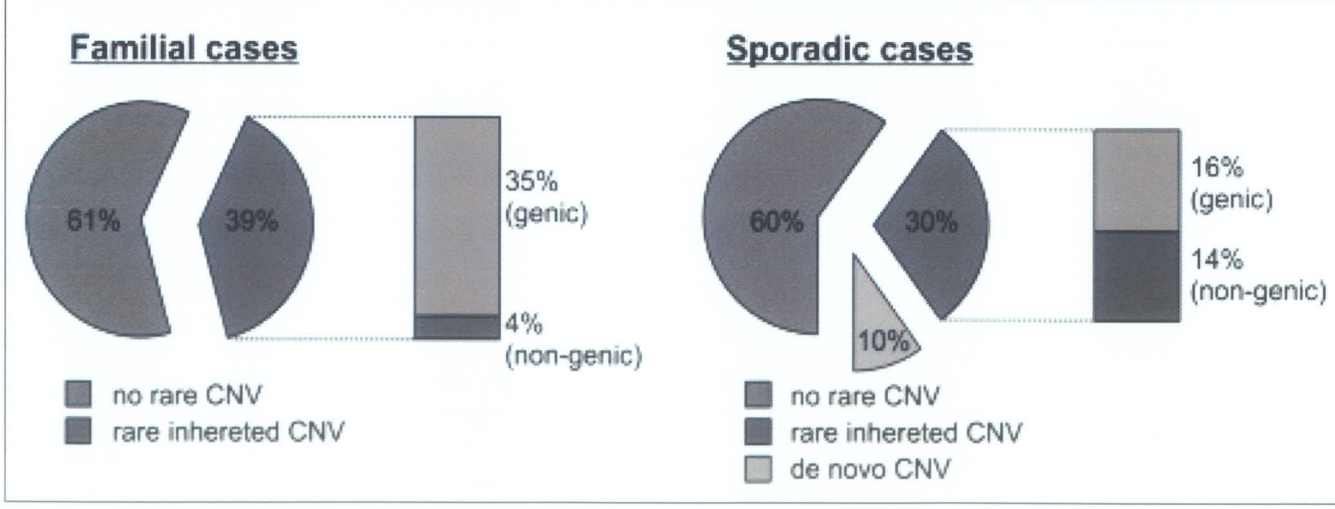
In addition to the genetic homogeneity, the Afrikaners are valuable for genetic studies because they present a close-knit family structure and offer the potential to perform detailed genealogical analysis. This aspect affords reliable discrimination of familial and nonfamilial forms of the disease and facilitates family-based genetic studies.

The results highlight the difference in the genetic architecture of the familial and sporadic forms of the disease.[11] [See Fig 1] Based on the results of this study and previous ones, there is considerable evidence for linkage to 13q. The findings support the notion that multiple genetic variants, including individually rare ones (often unique to a single patient), that affect many different genes contribute to the genetic risk of familial schizophrenia. This heterogeneity (present to some degree even in founder populations) is consistent with the hypothesis that there are many genes that contribute to schizophrenia and may account for past and present difficulties in finding bona fide genetic variants. These results are likely to have general implications regarding the genetic architecture of schizophrenia as there are significant clinical similarities of schizophrenia cases diagnosed in the Afrikaners and those diagnosed in more heterogeneous populations (such as in the United States of America- USA).[11]

Those with schizophrenia from families with no history of the illness were found to harbour eight times more spontaneous mutations than in healthy controls. These mutations were mostly in pathways affecting brain development.[12] The findings strongly suggest that rare, spontaneous mutations likely contribute to vulnerability in case of schizophrenia from previously unaffected families. This may also shed light on why the illness has frustrated efforts to implicate gene variants with major effects, and seems to defy natural selection by persisting in the population even though relatively few of those affected have children. There now exists a dramatic demonstration that genetic vulnerabilities for these illnesses may not be inherited from parents, at least in the sense that these vulnerabilities were not present in the parenteral genome.[12]

Although it is known that genetics play a major role in the transmission of both autism and schizophrenia, most cases are sporadic rather than familial.[12]

Figure 1: Genetic architecture of schizophrenia



© University of Pretoria

Individualizing recurrence risks

Recurrence risk assessment is more challenging to provide in the context of complex psychiatric disorders. In these cases, empiric recurrence risk (of varying quality and comprehensiveness) are often available and can be used as a basis for individualizing recurrence risks.

Limitations of empiric risks for psychiatric disorders: the theory behind individualized assessment

Empiric recurrence risk figures are the foundations on which individualized risk assessment can be developed however they do have a number of limitations. There are a limited number of family relationships for which empiric risks are available. There are limited data when there are multiple disorders within a family, or when relatives on both maternal and paternal sides are affected. Because risks for psychiatric disorders vary according to a number of factors that are typically not taken into account during the generation of empiric risks, they should generally not be applied without critical assessment.[13]

As with any complex disorder individualized recurrence risks should be produced by adjustment of empiric data based on characteristics including: the individual for whom risk is being calculated, the affected family members and the aggregation pattern in the family. It is important to recognize that a process of inference is used to produce individualized risks and that making a diagnosis of risk based on family history and other factors is not equivalent to making a diagnosis of disease based on the presence of certain symptoms. The distinction should be reflected in the language used by genetic counsellors. The primary reason why this distinction is so important is to resist the growing trend of identifying unaffected individuals as "patients without symptoms".

The paper by Austin and Peay[14] provides a full and detailed discussion of the limitations of empiric risks, and of the theory behind individualization of these risks.

The characteristics of the individual for whom risk is being calculated that may be pertinent to individualized risk assessment include age, gender and psychiatric history. Characteristics of the affected family members that may be pertinent to risk assessment include gender, age at illness onset, and severity of illness. Factors relating to the aggregation pattern in the family that may be pertinent include bilineality, number and proportion of affected family members, and types of psychiatric diagnosis.

Individualizing risks

Unfortunately, empiric risks from available studies may not reflect the risks in individual families; those risks may be far higher or lower depending on the mode or transmission in a given family and the particular configuration and comorbidities of affected family members. Several factors may help determine whether particular families may be at the higher or lower range of risk:

- the presence of a greater number of affected family members is often thought to represent a more highly genetic form of the disorder and thus a higher risk of disease recurrence
- early onset of a disorder may reflect greater genetic loading
- gender of affected and at-risk family members

Given the variability and complexity of factors that may influence risk for a given family, it may not be possible to provide individualized risks in most instances. It is important to highlight the limitations of the information and to discuss how risks are derived, the spectrum of the diagnoses assessed in the supporting family studies, and the differences between empiric and individualized risks. It is most helpful to provide a range of risk, accompanied by the above qualifications. It is also important to remember that the perceived magnitude and impact of risks and probabilities may vary greatly from person to person.[14]

After genetic counselling there should be an improved understanding of the following issues:

- the complex nature of the genetic and environmental basis of psychiatric disorders
- the importance of accurate diagnoses for recurrence risk counselling
- an understanding of the limitations of empiric risks
- specific information about psychiatric disorder of interest (recurrence risks, symptoms, course of illness, treatment options)
- opportunities for decision making based on risk information
- sources of additional information and support related to being at risk or passing on risks to other family members

A comprehensive discussion of genetic risks for psychiatric disorders may be time-consuming and beyond the scope of practice for many psychiatrists. For patients who have questions beyond the knowledge base of their psychiatrists referral for consultation by genetics professional is appropriate.[14]

The prospects of genetic testing

Will genetic testing be feasible for psychiatric illness? In theory there is a scenario in which testing could be applied in schizophrenia. It involves velocardiofacial syndrome (VCFS) caused by a microdeletion of chromosome 22q11 that occurs in 1/4000 live births. In addition to a variety of physical features individuals with VCFS have elevated risks of psychiatric symptoms. In particular, rates of psychotic disorders have been reported to be as high as 25 – 30%.

Deletion of 22q11 and schizophrenia

Hemizygous deletion in the chromosome 22q11 region is associated with VCFS. These 1.5 – 3-Mb deletions are usually de novo events. Approximately 70% of 22q11 deletions are de novo; of the remaining 30%, more than three quarters are maternally inherited.[15]

Hemizygosity of the 22q11.2 region seems to confer the major CNV risk for the expression of schizophrenia in 22q11.2DS. While further investigations are required to determine molecular risk factors that mediate de novo 22q11.2 events during gameto genesis, the research results to date reinforce the need for genetic counselling of individuals with 22q11.2DS regarding the risk for schizophrenia. Further efforts are required to identify the 1% of patients with schizophrenia who carry 22q11.2 deletions.[10]

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Deletion in the chromosome 22q11 region is also a risk factor for schizophrenia. Patients with schizophrenia have been observed to have increased rates of 22q11 deletion. Estimates range from 0.2% to 5.3%. The estimated rate in Afrikaner schizophrenia patients was 2%. In the general population this rate is 0.025%. Estimates of psychosis in adult VCFS patients with 22q11 deletion range from 6% to 16%. A recent study estimated the rate of DSM-IV schizophrenia diagnosis in 78 adults with 22q11 deletion syndrome to be 22.6%. This is a more than 20-fold increase over the rate of schizophrenia in the general population. The mechanism by which this deletion contributes to the risk for schizophrenia and psychosis in VCFS patients is not yet understood. Proline dehydrogenase, a gene in the 22q11 deletion interval, was reported to be associated with schizophrenia. Other reports suggested this gene is not associated with schizophrenia. The deletions identified in this region – 3Mb and 1.5Mb – are large and can be identified by standard cytology techniques (e.g. fluorescence in situ hybridibation – FISH). Until recently, technology to detect smaller deletions was unavailable. With newly available technology and decreasing resolution of detectable CNV size, it is possible that other structural variants in this region, such as smaller deletions or even duplications, will be identified. If additional variants are identified, they may serve to ultimately define a critical region for schizophrenia susceptibility with the 22q11 locus. This region is the most likely candidate region to harbor CNVs that contribute to schizophrenia susceptibility.[15]

Genetic counsellors can help to recognize when psychiatric disorders within a particular family may in fact be syndromic and due to an identifiable cause such as 22q11 deletion.

In a study to assess the frequency of 22q11 deletion in Afrikaner schizophrenic patients the authors provided a two-stage screening protocol to identify schizophrenia patients with this deletion.[17] The screening of this sample for clinical features compatible with the 22q11DS was relatively simple to perform, accurately identified patients with the 22q11DS and therefore considerably narrowed the sample size in need of confirmation by FISH. One third of the "flagged" cases were true positives and there were no false negatives. The criteria used, as suggested by Bassett and Chow are useful despite differences in their specificity and predictive value regarding 22q11DS.[18]

According to Bassett and Chow (1999)[18] patients with 2 or more of the following groups of screening criteria are at increased risk for 22q11DS:

1. Hypernasal speech, history of speech therapy, velopharyngeal incompetence, cleft palate (usually submucosal)
2. Characteristic facial features e.g. long narrow face, narrow palpebral fissures, flat cheeks, prominent nose, small ears, small mouth, retruded chin
3. Learning difficulties, history of special education, MR (borderline to mild)
4. Congenital heart defects e.g. ventricle septum defect, tetralogy of Fallot, right Aortic arch, double aortic arch
5. Other significant congenital anomalies e.g. club feet, polydactily, renal anomalies, kyphosis/scoliosis, hypospadia
6. History of hypocalcaemia, and/or hypoparathyroidism
7. History of athymia or severe immunodeficiency in infancy

Of the 85 subjects with schizophrenia examined, 27 had one or more of the physical anomalies associated with the 22q11DS.[17] Of these, 6 patients (four males, 2 females) had two or more of the positive screening criteria. A FISH analysis was performed on these six patients and two (one male, one female) were found to carry hemizygous deletions on chromosome 22q11. In an independent set of studies the entire sample was genotyped using a panel of single nucleotide polymorphismo (SNPs) from the 22q11 locus. In the group of 85 schizophrenic patients described in this study, the same two individuals identified by FISH as carrying the 22q11 deletion were apparently "homozygous" for all tested markers. The rest of the patients were heterozygous for at least one marker locus, thus excluding the presence of a typical 22q11 microdeletion. Therefore it was concluded that the rate of 2 in 85 represents an accurate estimate and supports the conclusion that the accuracy of the clinical screening for 22q11DS, as performed in this study, is 33%.[17]

Physical and intellectual phenotypic features will remain the primary means of identifying a 22q11DS subtype of schizophrenia. It is apparent that many individuals with 22qDS-schizophrenia will not have obvious congenital anomalies and that the majority of such individuals will not have mental retardation, although learning difficulties may be prevalent. This is consistent with the fact that individuals with 22qDS schizophrenia have been recruited into research samples despite intensive pre-screening, as illustrated by the National Institute of Mental health study of childhood-onset schizophrenia. Lower rates of comorbid substance use disorders may further facilitate the potential inclusion of subjects with 22qDS-schizophrenia in research studies of schizophrenia. A high index of suspicion, careful medical history taking and assessment for the more subtle physical and cognitive features of the syndrome will often be necessary to identify subjects with 22qDS. Separate studies of these individuals will be important to determine which features distinguish this etiologic subtype from other forms of schizophrenia.[19]

Genetic discrimination in mental disorders

As discussed, genetic tests related to psychiatric disorders and their treatments are becoming more prominent in research and in the near future will be more so in clinical settings too. The possibility becomes more troubling that such information will be used for purposes other than those for which it was collected. For these reasons, the stages implementation of the federal Genetic Information non-discrimination Act (GINA) of 2008 in the USA is of substantial importance to persons with mental disorders, persons at risk for the conditions, and family members of both groups.[20] The final provisions of the bill are due to have taken effect in 2010. GINA's major sections, Titles I and II, address health insurance and employment contexts, respectively. Title I bars health insurers from requesting genetic information from clients, including requiring clients to undergo genetic tests, and from using genetic information obtained from other sources to make coverage decisions or to set premiums. In comparable fashion, Title II prohibits employers from

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

mandating that workers be tested for genetic propensities to illness and from using genetic information that may otherwise come to their attention for a range of employment-related decisions and termination. Violations of either are subject to substantial monetary penalties.[21]

This type of legislation impacts on psychiatry and might be more limited than in other areas of medicine for several reasons. But as has been discussed earlier, family history information on many mental disorders can be quite useful in identifying risk categories for close relatives. In the USA when questions about genetic testing arise, clinicians should make every effort to ensure that their patients are aware of GINA and related state laws and that they have an accurate understanding of the protections that these regulations provide. In the SA context the Mental Health Care Act states in article 10 that "users" may not be unfairly discriminated against on grounds of mental illness or that users must receive the same health services as any health user.[22] Further development of GINA related legislation in the SA context will be necessary in the future.

Conclusion

The recent advances in the field of genetic research and their wide publicity in the scientific and clinical journals as well as on the internet have created an ever rising demand for genetic counselling regarding schizophrenia. Quite often it is the psychiatrist who must convey this information as genetic counsellors feel uncertain when it comes to accurately making psychiatric diagnosis as well as the course and prognosis of illness. Furthermore, genetic counselling for complex disorders is among the most challenging areas encountered by genetics professionals and the role that environmental factors play are not clear cut either. The newest developments in understanding hereditary and other factors believed to contribute to the etiology, clinical presentation, course and prognosis of the disease should be taken into account when genetic counselling is requested. Psychiatrists and allied mental health and other professionals need to be vigilant in preventing genetic discrimination in mental disorders.

References

1. *Finn CT, Smoller JW. Genetic Counseling in Psychiatry. Harv Rev Psychiatry 2006; 14:109-121.*
2. *Austin JC, Honer WG. The potential impact of genetic counselling for mental illness. Clin Genet 2005; 67:134-42.*
3. *Smith LB, Sapers B, Reus VI, Freimer NB. Attitudes towards bipolar disorder and predictive genetic testing among patients and providers. J Med Genet 1996; 33:544-9.*
4. *Smoller JW, Finn CT. Family, twin, and adoption studies of bipolar disorder. Am J Med Genet C Semm Med Genet. 2003; 123:48-58.*
5. *Tsuang MT, Stone WS, Faraone SV. Genes, environment and schizophrenia. Br J Psychiatry 2001; 178:518-24.*
6. *Tsuang M. Schizophrenia: genes and environment. Biol Psychiatry 2000; 47:210-20.*
7. *Harper PS. Practical genetic counselling. Oxford, Boston: Butterworth-Heinemann; 1998.*
8. *McGuffin P, Owen MJ, Gottesman II. Psychiatric genetics and genomics. Oxford, New York: Oxford University Press, 2002.*
9. *"Copy number variation: The new genetic frontier (Report)". Harvard Mental Health letter (Oct 1, 2008): NA Academic OneFile.*
10. *St Clair D. Copy Number Variation and Schizophrenia. Schizophrenia Bulletin 2009, 35(1):9-12.*
11. *Xu B, Woodroffe A, Rodriquez-Murillo L, Roos JL, J v Rensburg L, Abecasis GR, Gogos JA, Karayiorgou M. Elucidating the genetic architecture of familial schizophrenia using rare copy number variant and linkage scans. PNAS 2009:106(39):16746-16751.*
12. *Xu B, Roos JL, Van Rensburg EJ, Gogos JA, Karayiorgou M. Strong association of de novo copy number mutations with sporadic schizophrenia. Nat Genet 2008; 40 (7): 880-885.*
13. *Austin JC, Palmer CGS, Rosen-Sheidly B, Veach PM, Gettig E, Peay HL. Psychiatric Disorders in Clinical Genetics II: individualizeing Recurrence Risks. J Genet Cousel 2008; 17:18-29.*
14. *Austin JC, Peay HL. Applications and limitations of empiric data in provision of recurrence risks for schizophrenia: A practical review for healthcare professionals providing clinical psychiatric genetics consultations. Clinical Genetics 2006; 70(3):177-187.*
15. *Mulle J. Genomic Structural Variation and Schizophrenia. Current Psychiatry Reports 2008;10:171-177.*
16. *Bassett AS, Marshall CR, Lionel AC, Chow EWC, Scherer SE. Copy number variations and risk for schizophrenia in 22q11 deletion syndrome. Human Molecular Genetics 2008; 17(24):4045-4053.*
17. *Wiehahn GJ, Bosch GP, Du Preez RR, Pretorius HW, Karayiorgou M, Roos JL. Assessment of the Frequency of the 22q11 deletion in Afrikaner schizophrenia patients. American Journal of Medical Genetics Part B (Neuropsychiatric Genetics) 2004; 129B: 20-22.*
18. *Bassett AS, Chow EWC. 22q11 deletion syndrome: A genetic subtype of schizophrenia. Biol Psychiatry 1999; 46:882-891.*
19. *Bassett AS, Chow EWC, AbdelMalik P, Gheorghiu M, Husted J, Weksberg R. The Schizophrenia Phenotype in 22q11 Deletion Syndrome. Am J Psychiatry 2003; 160:1580-1586.*
20. *Genetic Information Nondiscrimination Act of 2008, Public Law 110-233, May 21, 2008.*
21. *Appelbaum PS. Genetic Discrimination in Mental Disorders: The impact of the Genetic Information non-discrimination Act. 2010. Psychiatric Services 2010; 61:338-340.*
22. *New Mental Health Care Act (Act 17 of 2002) Government Gazette No 27117:1-76.*

© University of Pretoria

Nuwe bevindings oor die genetiese argitektuur van skisofrenie: Die bydrae van studies onder Afrikaners

Authors:
Johannes L. Roos[1]

Affiliations:
[1]Department of Psychiatry, University of Pretoria, South Africa

Correspondence to:
Johannes Roos

Email:
erna.fourie@up.ac.za

Postal address:
Private Bag X113, Pretoria 0001, South Africa

Dates:
Received: 16 Jan. 2013
Accepted: 15 Oct. 2013
Published: 24 Feb. 2014

How to cite this article: Roos, J.L., 2014, 'Nuwe bevindings oor die genetiese argitektuur van skisofrenie: Die bydrae van studies onder Afrikaners', *Suid-Afrikaanse Tydskrif vir Natuurwetenskap en Tegnologie* 33(1), Art. #396, 9 pages. http://dx.doi.org/10.4102/satnt.v33i1.396

Copyright:
© 2014. The Authors. Licensee: AOSIS OpenJournals. This work is licensed under the Creative Commons Attribution License.

Read online:
Scan this QR code with your smart phone or mobile device to read online.

Die genetiese komponent in die etiologie van skisofrenie word reeds 'n geruime tyd al aanvaar. Die onderliggende argitektuur van die genetiese risiko bly 'n punt van bespreking. Vroeë studies oor koppeling en kandidaatverbintenis het grootliks gelei tot onoortuigende of onbesliste resultate. Meer onlangs het kragtige tegnologieë beskikbaar geword. Dít, tesame met groot genoeg toetsgroepe en genoomwye panele van genetiese merkers, het dit moontlik gemaak om sistematiese en agnostiese skandering te doen regdeur die genoom. In agnostiese skandering word tegnologie aangewend om die probleem op te los, sonder enige voorveronderstelling oor waar die genoomrelevante allele gevind sal word. Daar word gesoek na algemene en unieke siektevariante met onderskeidelik klein of groot effek. Die soektog na variante wat nie 'n groot effek het nie, het beperkte sukses gehad. Aan die ander kant het die rol van unieke genetiese gebeure, soos kopienommervariasie (KNV) en uniekepunt-mutasies al hoe belangriker geraak by geenontdekking in geval van skisofrenie. Onlangse navorsing onder Afrikanerpasiënte met skisofrenie, wat bou op vroeëre bevindings van spontane KNV by die 22q11 lokus, het lig gewerp op die spontane mutasieparadigma as 'n groot komponent in die argitektuur van skisofrenie. 'n Oorsig oor onlangse bevindings van hierdie studies word gegee.

Latest findings in the genetic architecture of schizophrenia: The contribution of genetic studies among Afrikaners. A genetic component of schizophrenia has been acknowledged for a long time. The underlying architecture of the genetic risk remains a contentious issue. Early linkage and candidate association studies led to largely inconclusive results. More recently powerful technologies became available. This aspect coupled with samples of sufficient sizes, and genome-wide panels of genetic markers facilitated systematic and agnostic scans throughout the genome for either common or rare disease risk variants of small or large effect size, respectively. Although the former had limited success, the role of rare genetic events, such as copy-number variants (CNVs) or rare point mutations, has become increasingly important in gene discovery for schizophrenia. Recent research done among Afrikaner patients with schizophrenia, building upon earlier findings of de novo recurrent CNVs at the 22q11.2 locus, has highlighted a de novo mutational paradigm as a major component of the genetic architecture of schizophrenia. Recent progress in this regard will be reviewed.

Inleiding

In 'n vorige artikel (Roos 2008) het die outeur 'n oorsig gegee oor bevindings in navorsing die voorafgaande 10 jaar oor die genetika van skisofrenie by die Afrikaner as 'n stigtersbevolking. Dit was in 2007 toe hierdie samewerkingooreenskoms betreffende die genetika van skisofrenie in 'n stigtersbevolking tien jaar aan die gang was. Nou, 15 jaar later, het dié samewerkingooreenkoms tussen die Columbia University in New York en die Departement van Psigiatrie aan die Universiteit van Pretoria (UP) interessante resultate begin oplewer.

Waar aanvanklike resultate in dié studie baie klem gelê het op die stigtersbevolking en genealogie van die Afrikaner, het onlangse bevindings juis begin klem lê op die ander pool, naamlik sporadiese gevalle van skisofrenie, met ander woorde waar daar geen geskiedenis van skisofrenie by familielede in die eerste en tweede generasie teenwoordig is nie. In die slotopmerkings van die vorige artikel (Roos 2008) sluit die outeur af met: 'Hopelik sal die ondersoek na die genetiese aspekte van skisofrenie in die Afrikaner 'n bydrae lewer tot beter begrip en behandeling van hierdie ernstige psigiatriese toestand.' Moontlik is hierdie groep navorsers besig om in daardie rigting te begin beweeg.

Dié oorsig som aspekte van die genetiese navorsing oor die etiologie van skisofrenie op. Daar sal klem gelê word op bevindings in 'n stigtersbevolking van Afrikaners. Sleutelbesprekingspunte en uitdagings in die navorsing sal toegelig word. Toekomstige rigtingwysers in hierdie tipe

© University of Pretoria

navorsing word beklemtoon. Daar is min rigtingwysers beskikbaar oor die patogeniese meganismes wat skisofrenie onderlê. Skisofrenie word gedefinieer as 'n kliniese sindroom gebaseer op die teenwoordigheid van operasionele kriteria. Hierdie kriteria sluit in waanidees, hallusinasie, apatie en kognitiewe probleme wat lei tot disfunksie en progressiewe agteruitgang. (American Psychiatric Association 1994) Die prevalensie van skisofrenie in die algemene bevolking word geskat op ongeveer 1%. Klassieke studies van families, tweelinge en aangenome kinders, wat die herhalingsrisiko vir familielede bepaal, het regstreekse bewys gelewer vir 'n genetiese etiologie. (Sullivan, Kendler & Neale 2003) Die risiko om die siekte te ontwikkel neem eksponensieel toe met die graad van genetiese verwantskap tot 'n individu wat aan die siekte ly. Dit bereik 'n punt van ongeveer 50% vir monosigotiese tweelinge. (Sullivan *et al.* 2003)

'n Onlangse Sweedse studie wat 'n totaal van 9 009 202 unieke individue insluit, wat afkomstig was van meer as 2 miljoen kernfamilies, het die risiko vir skisofrenie in biologiese en aanneemouers, afstammelinge, regstreekse sibbe en halfsibbe van individue met skisofrenie nagegaan. Inligting van 'n nasionale multigeneratiewe databasis is vergelyk met gegewens by hospitaalontslag. Bevindings van hierdie studie bevestig dat die oorerflikheid van skisofrenie hoog is. Bepalings dui op 64.3% (95.0% sekerheidsinterval = 61.7% – 67.5%) (Lichtenstein *et al.* 2009). Ander risikofaktore sluit in 'n gedeelde familie-omgewing, manlike geslag, hoë ouderdom van vader by konsepsie, perinatale gebeure en substansmisbruik. Dit moet beklemtoon word dat hoewel 'n groot genetiese komponent vir skisofrenie onomwonde aangetoon is, het slegs 'n klein proporsie van gevalle 'n familiegeskiedenis van skisofrenie. (Sullivan *et al.* 2003; Lichtenstein *et al.* 2009; Tandon, Keshaven & Nasrallah 2008.) Die resultate van hierdie artikel sal aantoon dat die grootliks geïgnoreerde aspek in vroeëre pogings om die siekterisiko te modelleer, moontlik 'n belangrike begrip kan wees in die opklaring van die genetiese argitektuur van skisofrenie.

Die belangrike rol wat genetika in die etiologie van skisofrenie speel, word nie meer bevraagteken nie. Die fundamentele onsekerheid in die genetika van skisofrenie was nog altyd die aard van die onderliggende alleliese argitektuur. Histories kan probleme toegeskryf word aan 'n kombinasie van alleliese lokus en fenotipiese heterogenisiteit wat gelei het tot: waninterpretasie van of oorinterpretasie van negatiewe data, of nie oortuigende data. Ongeag hierdie uitdagings, is daar 'n oorvloed van onlangse bevindings, aangedryf deur tegnologiese vooruitgang wat die kragte wat die genetiese risiko vir skisofrenie vorm, begin ontsyfer. Dié bevindings lewer 'n lang verwagte kans op om hoogs betroubare resultate te produseer en kan groot praktiese implikasies inhou (Rodriguez-Murillo, Gogos & Karayiorgou 2012).

Die genetiese argitektuur van skisofrenie

Vier faktore word betrek by die genetiese argitektuur van komplekse genetiese versteurings:

- Die getal risiko-allele wat bydra tot die siekte in die bevolking. Dit kan veelvuldige risiko-allele in een of meer gene insluit.
- Die frekwensie van elk van die risiko-allele in die bevolking.
- Die grootte van die effek (penetrasie) van 'n risiko-alleel.
- Die manier hoe risiko-allele saamwerk, byvoorbeeld op 'n byvoegende of sinergistiese manier (Laursen & Munk-Olsen 2010).

Pogings om die genetiese argitektuur van skisofrenie op te klaar moet 'n bevredigende beskrywing van al hierdie faktore gee. Verder moet dit 'n verklaring gee vir die hoë wêreldwye voorkoms, die voortbestaan van hierdie versteuring ondanks omgewingsfaktore wat voorkom, die verhoogde mortaliteit en die ernstig verminderde voortplantingspotensiaal van aangetaste individue (Laursen & Munk-Olsen 2010).

In hierdie konteks word die genetiese argitektuur wat skisofreniese vatbaarheid onderlê, geformuleer deur twee groot hipoteses, wat fundamenteel verskil in hul konseptualisasie van die etiologie:

Die hipotese van algemene siekte/algemene alleel (ASAA), en die hipotese van algemene siekte/unieke alleel (ASUA) (Goldstein 2009; Manolio *et al.* 2009).

Twee hipotetiese modelle van skisofrenie: Algemene siekte/ algemene alleel versus algemene siekte/unieke alleel

Die ASAA beklemtoon die samewerking tussen verskeie relatief algemeen oorgeërfde allele wat elk dui op 'n klein of matige risiko. Dit werk saam om die voortdurende hoë voorkoms van die siekte te verklaar. Hierdie hipotese neem aan dat daar verminderde sigbaarheid van elke kleineffek-alleel tot negatiewe seleksie is. Verder neem die hipotese aan dat hierdie allele by uitsondering in sporadiese gevalle herkombineer in hoërisiko-kombinasies (Yang, Visscher & Wray 2010).

Algemene siekte/unieke alleel

Algemene siekte/unieke alleel (ASUA) beklemtoon op sy beurt 'individualiteit' en hoë genetiese heterogenisiteit, veral die betrokkenheid van individueel unieke, maar hoogs deurdringende allele. Hier word elke geval veroorsaak deur 'n enkele unieke variant, en hierdie variante kan voorkom in verskillende gene by verskillende individue. 'n Krities deurslaggewende aanname van die ASUA-model is dat nuwe (of spontane) mutasies in sporadiese gevalle voortdurend hoogs penetrerende allele genereer en aanvul. Hierdie allele word vinnig teengeselekteer in die bevolking (Goldstein 2009; Manolio *et al.* 2009)

Die bruikbaarheid van die splitsing lê primêr in die feit dat groot duidelik metodologiese benaderings gebruik word om

© University of Pretoria

siekte-assosiasie op te spoor met unieke of algemene variante. Gevolglik help dit om verskillende geen-opsporingstrategieë te evalueer. Vanuit 'n biologiese oogpunt, aan die ander kant, in die konteks van die komplekse neurale substraat wat die vatbaarheid van skisofrenie onderlê, is hierdie genetiese konstrukte moeiliker om te ontrafel. Pasiënte se genoom bevat allele oor 'n wye reeks van frekwensies. Sommige allele dryf en ander wysig die siekte-risiko en -uitdrukking. In hul samewerking beïnvloed hulle die struktuur en funksie van neurale netwerke.

Oor die algemeen het die poligeniese ASAA-paradigma van skisofrenie baie aandag geniet. In kontras hiermee, is daar slegs 'n paar voorbeelde waar enkele hoogs penetrerende mutasies geïdentifiseer is wat iemand ontvanklik sou maak vir skisofrenie. Hierdie studies is beskou as op die buitegrens van bevindings en van beperkte belang in die genetiese argitektuur van die siekte. Onlangse empiriese data onderskryf die ASUA-model. Dit het veroorsaak dat daar 'n drastiese herwaardering van die teoretiese basis onderliggend aan die ASUA-model was (Sullivan *et al.* 2003).

Oorweging van die algemene siekte/unieke alleel-model

Die sterkste formulering van die ASUA-model stel voor dat die meeste, indien nie alle, gevalle van skisofrenie die gevolg is van die teenwoordigheid van sommige unieke maar penetrerende mutasies. Unieke variante (0.5% tot 5.0% frekwensies) oorskadu algemene variante by verre (The 100 Genomes Project Consortium 2010). Dit is dus nie verbasend dat hulle moontlik beduidend kan bydra tot die genetiese argitektuur van komplekse siektes nie. Relevante bewys is eers gelewer deur studies wat gesoek het na 'n unieke variant in die ekstreem fenotipiese distribusie van komplekse kenmerke (Cohen *et al.* 2004). Hierdie unieke mutasies is meer geneig om proteïenfunksie te affekteer en veroorsaak dan ook skadelike fenotipes, wat 'n impak het op voortplantingsfiksheid (Flanagan, Patch & Ellard 2010; Valdmanis, Verlaan & Rouleau 2009). As sodanig, is hulle meer geneig om 'n faktor te wees wat seleksie suiwer, met 'n kort halflewe in geselekteerde stambome en is hulle beperk tot lae alleliese frekwensie. In dié konteks, sou die hoë voorkoms van die versteuring ten minste 'n hoë genoeg verhouding van spontane mutasies vereis. Die tempo van spontane plaasvervanger-mutasies op kiemlynbasis is relatief hoog en is bepaal op ~10^{-8} per basispaar per generasie (Crow 2000). Dit sou 0.5–1.0 spontane eksoniese mutasies per nuut gebore baba voorspel (The 100 Genomes Project 2010).

Vanweë tegniese beperkings was dit nie moontlik om die wêreldwye bydrae van unieke variante tot psigiatriese versteurings op 'n sistematiese manier na te gaan nie. Dit het slegs 'n paar jaar tevore moontlik geword, danksy tegniese vooruitgang, om kopienommer-analise met hoë resolusie by individuele genome na te gaan. Dit was dan ook die eerste kans om onbevooroordeelde, genoomwye opnames van unieke variante in psigiatriese versteurings te doen. Onlangs is 'n groot hoeveelheid variasies in die menslike genoom aangetoon. Een van die mees belangwekkende tipes van variasie is kopienommervariasie (KNV), die weglating of dupliserings wat tipies groter as 1 kb is (Iafrate *et al.* 2004; Sebat *et al.* 2004). KNV'e is óf uniek óf algemeen en kollektief sou dit ~12% van die genoom in die algemene bevolking dek (Redon *et al.* 2006). Dit het gelei tot die voorstel dat addisioneel tot hul bekende rol in 'n getal ongewone sindrome, KNV'e ook dalk kan bydra tot meer algemene siektefenomene. In der waarheid het uniek oorgeërfde en spontane KNV'e na vore gekom as 'n belangrike genomiese oorsaak van siekte. Heel onlangs het die wyd verspreide beskikbaarheid van volgendegenerasie-tegnologie vir geenreeksbepalings dit moontlik gemaak dat individuele genome se volgorde bepaal kon word. Gevolglik kan daar sifting gedoen word van unieke variante op die hoogste vlak van genetiese resolusie (The 100 Genomes Project 2010). Dit het die gaping gevul wat deur KNV-studies gelaat is.

Mikro-uitwissing by 22q11.2 en die spontane paradigma

'n Mikro-uitwissing by die 22q11.2-lokus was die eerste spontane KNV beskryf in skisofrenie deur Karayiorgou (Karayiorgou *et al.* 1995). Karayiorgou is die studieleier van die navorsingsprojek oor skisofrenie in die Afrikaner. Sedert die ontdekking is daar 'n sterk en spesifieke tweerigtingverband vasgestel tussen die teenwoordigheid van 22q11.2-mikro-uitwissing en psigose. Tot 'n derde van alle laat adolessente of vroeë volwassenes wat die mikro-uitwissings het, ontwikkel skisofrenie of skiso-affektiewe versteuring volgens die diagnostiese DSM IV-kriteria (Gothelf *et al.* 2007; Murphy, Jones & Owen 1999; Pulver 1994). Verder het 'n getal studies aangedui dat 22q11.2-miko-uitwissings bydra tot 1% – 2% van skisofrenie-gevalle (Karayiorgou *et al.* 1995; Wiehahn *et al.* 2004;). Hierdie sterk tweerigtingverband tussen skisofrenie en die 22q11.2-mikro-uitwissing maak dié weglating een van die sterkste risikofaktore vir skisofrenie tans bekend. Verder is pasiënte met skisofrenie wat die 22q11.2-mikro-uitwissing het, en dié daarsonder klinies nie van mekaar onderskeibaar wat die skisofrenie-fenotipe aanbetref nie (Bassett *et al.* 2006). Die ontdekking van so 'n herhalende strukturele mutasie wat verantwoordelik is vir sporadiese gevalle van skisofrenie in die bevolking het die vraag laat ontstaan of spontane KNV's, of spontane mutasies in die algemeen, 'n selfs groter rol speel in die genetiese samestelling van mense met psigiatriese versteurings. Die vinnig ontwikkelende genomiese en reeksvolgorde-tegnologie het dit moontlik gemaak om hierdie vraag te ondersoek. Daar is nou bewys gelewer dat unieke spontane mutasies bydra tot die genetiese etiologie van skisofrenie.

Genoomwye skanderings vir spontane kopie-nommervariante

In een van die eerste KNV-skanderings om die rol van spontane strukturele mutasies in skisofrenie te ondersoek, is pasiënte met skisofrenie uit die Afrikaner-stigterbevolking gebruik. Xu *et al.* (2009) het 'n enkelnukleotied-polimorfisme (ENP), met 'n geen-reeksresolusie van 30 kb bekyk by 359

© University of Pretoria

mense met skisofrenie, asook kontrolefamilie-triades. Die navorsers toon aan dat spontane KNV's beduidend meer algemeen teenwoordig is in sporadiese gevalle (met geen geskiedenis van siekte by 'n familielid van die eerste of tweede generasie nie) as by die kontroles (10.0% teenoor 1.3%). Sporadiese gevalle bevat slegs 1.5 keer meer oorgeërfde KNV's as kontroles, wat 'n kleiner toename verteenwoordig, van 20.0% tot 30.0% (Xu *et al.* 2009). Daar word aangetoon dat daar versterking tussen die gene, geaffekteer deur spontane KNV's, was wat geïmpliseer word in neurale ontwikkeling en RNA-prosessering. In sommige gevalle het spontane KNV's in een generasie oorgeërfde variante geword wanneer hulle oorgedra is na 'n volgende generasie. In der waarheid is dit in 'n komplementêre skandering aangetoon dat uniek oorgeërfde KNV's geassosieer word met oorerflike skisofrenie. Oorgeërfde KNV's was amper twee maal so algemeen in familiële gevalle as in sporadiese gevalle of kontroles (Xu *et al.* 2009). Hierdie bevindings het die wyd verspreide patogeniese rol van spontane KNV's uitgelig, asook die verskille in die genetiese argitektuur van familiële versus sporadiese gevalle van skisofrenie. Sporadiese skisofrenie is gekenmerk deur 'n duidelike verryking van uniek spontane KNV's en slegs 'n matige toename in die tempo van uniek oorgeërfde KNV's. Laasgenoemde blyk gene nie by voorkeur aan te tas nie. Skisofrenie in die familie daarenteen, word gekenmerk deur die verryking in uniek oorgeërfde KNV's wat gene aantas (het 'n hoër penetrasie), terwyl spontane mutasies minder prominent is (Xu *et al.* 2008).

Nog 'n genoomwye soektog vir KNV's uitgevoer deur die SGENE + konsortium het bykomstige bewys gelewer van die rol van spontane KNV's (Stefansson *et al.* 2008). Hierdie studie het aanvaar dat weens negatiewe seleksie, die weglating van KNV's in skisofrenie uniek sal wees; gevolglik sal KNV's, wat voorkom in verskeie onverwante individue, deur onafhanklike spontane mutasies plaasvind. Hierdie navorsers vind drie spontane KNV's wat beduidend verband hou met psigose met 'n hoë waarskynlikheidverhouding. Gevolglik, KNV's met 'n beduidende effek op siekterisiko kan miskien oorheersend wees tussen dié wat hul oorsprong het in spontane gebeure. Spontane strukturele mutasie kom met 'n redelike frekwensie voor en daardeur verskaf dit 'n konstante bron van unieke en hoogs penetrerende variante wat 'n poel van mutasie vorm wat weer geïmpliseer word in skisofrenie (Lupski 2007).

Uniek spontane variasie-ontdekking deur grootskaalse reeksbepaling

Hoewel KNV-studies goeie aanvanklike resultate opgelewer het, is dit eers onlangs dat die volle bydrae van uniek spontane mutasies nagevors is. Xu *et al.* (2011) toets die hipotese dat spontane, proteïenveranderende mutasies substansieel bydra tot die genetiese komponent van skisofrenie. Die navorsers gebruik reeksbepaling van die volgende generasie om spontane mutasies of enkelnukleotied-variante en klein insluitings of weglatings op 'n sistematiese, genoomwye manier te bestudeer. Die eksoom van 53 Afrikaner familie-trio's met 'n diagnose van skisofrenie of skiso-affektiewe versteuring is geskandeer. ('n Eksoom is deel van die genoom gevorm deur eksone waarvan die geenreekse bly in volwassens). Hierdie persone het geen geskiedenis van siekte by familielede in die eerste of tweede generasie nie. Die kontrolegroep was 22 gesonde Afrikanerfamilie-trio's wat nie verwant is nie. In die aangetaste sporadiese gevalle vind Xu *et al.* (2011) 34 kandidate met spontane puntmutasies en vier spontane indels (insluiting en weglating). Onder die 34 spontane puntmutasies was 32 enkelnukleotied-mutasies nie-sinoniem en slegs twee sinoniem.

Onder die 32 nie-sinonieme mutasies het 19 evolusionêr behoue oorskot betrek. Drie van hierdie indels het gelei tot proteïenvertakkings en een het gelei tot 'n enkele aminosuurweglating. Hulle vind ook twee enkelnukleotied-variante, gelokaliseer binne voorspelbare skenker- of ontvanger- splitsingsetels, by die introniese gebied wat dit omsluit.

In die geheel bevind die navorsers dat 27 van die 53 pasiënte (~51%) minstens een spontane gebeurtenis dra. Dit was opvallend dat hierdie skandering 'n verhouding van nie-sinoniem (n = 32) tot sinoniem (n = 2) spontane mutasies van 16:1 toon. (Nie-sinonieme mutasie: indien 'n mutasie lei tot 'n verandering in die geënkodeerde polipeptied, is dit bekend as 'n nie-sinonieme mutasie; sinonieme mutasies daarenteen: indien 'n mutasie nie die polipeptiedproduk van 'n geen verander nie, word dit 'n sinonieme mutasie genoem). Dit is duidelik hoër as die 2.85:1 wat verwag sou word in 'n lukrake model. Verder ook is nie-sinonieme spontane puntmutasies in 'n groot oormaat gevind (~10-voudig), vergeleke met die neutrale mutasies relatief tot uniek oorgeërfde variante. Laasgenoemde sal waarskynlik minder bydra tot die patogenese van sporadiese gevalle. Die mutasies is ondersoek deur evolusionêre behoud na te gaan, sowel as die potensiaal van die spontane, proteïenveranderde mutasies om die struktuur of funksie van die geproduseerde proteïene te beïnvloed. Hierdie ondersoek toon die geïdentifiseerde mutasies het 'n groter potensiaal om proteïenstruktuur-funksie te beïnvloed as die uniek oorgeërfde variante. In die kontrolegroep identifiseer Xu *et al.* (2011) sewe eksotiese, spontane enkelnukleotied-variante, maar geen indelkandidate in sewe van die 22 kontrolepersone nie. Tussen hulle was vier nie-sinonieme wanmutasies en drie sinoniemes, 'n verhouding soos wat verwag sou word in neutrale omstandighede en baie na aan die verhouding vir normale funksionele variasie (1.14–1.45). Die geobserveerde verryking van nie-sinonieme, hoogs behoue, en ontwrigtende aminosuurmutasies tussen spontane gebeure by pasiënte, was sterk beduidend op 'n hoë moontlikheid van patogenisiteit. Dit is opmerklik dat al die mutasies gevind in sporadiese gevalle, voorgekom het in verskillende gene. Dit was voor statistiese evaluering vir enige spesifieke lokus en met een uitsondering was geeneen van die aangetaste gene voorheen onvoorwaardelik verbind aan skisofrenie nie. Die identifisering van herhalende mutasies in hierdie gene bewys dat dit siekte veroorsaak. Dit kan moontlik vereis dat duisende aangetaste individue ondersoek moet word.

© University of Pretoria

Tussen die moontlike patogene gebeure was daar 'n wanmutasie binne-in 'n gebied wat evolusionêr behoue gebly het, genaamd die DGCR2-geen (DiGeorge Critical Region 2-geen). Hierdie bevinding is veral interessant wanneer gekyk word na vorige werk oor die 22q11.2-lokus.

Die DGCR2 enkodeer vir 'n voorgestelde transmembraan-aanhegtingreseptor en is hemisigoties weggelaat vanweë die herhalende spontane mikro-uitwissings by hierdie lokus. Die identifisering van dié ontwrigtende, spontane puntmutasie is gedoen in 'n pasiënt met 'n struktureel intakte 22q11.2-lokus.

In 'n onafhanklike, kleiner studie rapporteer Girard *et al.* (2011) 15 spontane mutasies in 8 van die 14 sporadiese gevalle. Dié verhouding is vergelykbaar met dié gerapporteer in die groter studie onder Afrikaners (Xu *et al.* 2011). Addisioneel en soortgelyk aan die studie van Xu *et al.* (2011), is al die mutasies wat geïdentifiseer is, nie sinoniem nie (Girard *et al.* 2011). Hoewel hulle geen proteïenvertakkende indels gevind het nie, rapporteer hulle dat vier van 15 mutasies niekoderend was. Dit is 'n syfer wat hoër is as wat ewekansig verwag sou word.

Die voorafgaande resultate saamgevat, dui daarop dat spontane mutasies 'n groot rol speel by skisofrenie. Hierdie resultate is verkry van die reeksbepalingskanderings van die aanvang-eksoom. Die bevinding oor spontane KNV, enkelnukleotied-variasies, asook insluiting- en weglatingverhouding van 'n groep van sporadiese gevalle by Afrikaners, lewer sterk empiriese bewys dat spontane mutasies baie patogenies teenwoordig is in meer as die helfte van sporadiese gevalle van skisofrenie (Xu *et al.* 2011; Xu *et al.* 2012).

Verhoogde las van unieke mutasies

'n Samevatting van onlangse bevindings in die literatuur dui aan dat KNV's, enkelnukleotied-variasies, insluiting en weglating affekteer by 'n verskeidenheid van gene (Gothelf *et al.* 2007). Dit onderskryf die hipotese dat veelvuldige, uniek spontane (herhalende én unieke), asook oorgedraagde genetiese variante in moontlik honderde duidelike loki almal bydra tot die genetiese risiko van skisofrenie. Dit is moontlik weens die kompleksiteit van die neurale substrate wat aangetas word deur hierdie siekte. Dit betrek 'n groot mutasieteiken wat baie gene insluit. Die presiese getal van hierdie gene, wat moontlik honderde kan beloop, moet nog bepaal word. Dit sal die sifting van etlike duisende families vereis. Dit is ook moontlik dat die nuut ontdekte loki kan uitloop op 'n kleiner getal algemene biologiese paaie of neurale netwerke om 'n algemene patroon van kliniese disfunksie of simptome te genereer. Hierdie verklaring is nog nie bevestig nie. Die neurale kompleksiteit wat skisofrenie onderlê, tesame met die relatief hoë mutasieverhouding, gee 'n moontlike verklaring vir die bestaan en voortbestaan van skisofrenie onder verskeie menslike bevolkings, ongeag die uiteenlopende omgewingsfaktore en ernstige inkorting van voortplantingspotensiaal (Girard *et al.* 2011).

Spontane geenmutasies dui op 'n patroon van genetiese en neurale kompleksiteit

In 'n onlangse publikasie van ons groep is die bewys van spontane oorsake in skisofrenie geëvalueer (Xu *et al.* 2012).

Die bydrae van uniek spontane kopienommer-variante (KNV) tot skisofrenie is goed bevestig (Karayiorgio *et al.* 1995; Xu *et al.* 2011). Die bydrae van spontane nukleotiedvariante is nog nie omvattend ondersoek nie. Eksoom-reeksbepaling van 146 Afrikaner-familietrio's van pasiënte met 'n diagnose van skisofrenie of skiso-affektiewe versteuring is voltooi (Xu *et al.* 2012).

Vier-en-dertig (34) onaangetaste kontrole-trio's is nagegaan. Die eksoom van 85 pasiënte, met skisofrenie of skiso-affektiewe versteuring, uit die Verenigde State van Amerika is ook nagegaan. Draers van uniek spontane KNV's (> 30kb) is uitgesluit. Die uitsluiting is gebaseer op vorige KNV-skanderings van hierdie groepe pasiënte. Die navorsers gebruik 'n analitiese pyplyn wat voorheen beskryf is en 'n reeks filters, insluitend die finale validasie deur Sanger se reeksbepaling van alle familielede (Xu *et al.* 2011). Meer as 90% van die enkelnukleotied-variante (ENV'e) en 20% van insluiting of weglatings is geëvalueer.

Wanneer ons kyk na fenotipiese ooreenstemmings in die bevindings, sien ons 'n ooreenkoms tussen die ouderdom van die vader by die pasiënt se geboorte en die getal spontane gebeure per afstammeling, maar geen ander beduidende verskille is gevind nie. Individue met jonger vaders (22–30 jaar oud tydens pasiënt se geboorte) neig om 'n laer gemiddelde spontane mutasietempo te hê as diegene met ouer vaders (Xu *et al.* 2012).

Gene, aangetas deur spontane variante, is in ons studie nie beduidend verteenwoordigend in twee voorheen vasgestelde, uitgebreide lyste van sinaptiese gene nie. Verder het padanalise, deur gebruik te maak van die DAVID-annotasie-instrument nie enige beduidende verrykte funksionele groepe gevind nie (Xu *et al.* 2012). Evaluasie van proteïen met proteïen-interaksies (PPI) toon 'n beduidend groter verbintenis tussen mutasieteikens as wat verwag sou word as wanneer dit toevallig was. Multigene PPI-groepe sluit een in wat sentreer op MTOR en een wat sentreer op CANX, wat weer insluit 'n ekstrasellulêre matriks- en selaanhegtingsproteïene. Dit suggereer dat diverse skisofrenie-risikogene moontlik kan uitloop op 'n korter lys van funksionele modules.

Die navorsers ondersoek verder tot watter mate verryking in funksioneel spontane gebeure bepaal word deur die ontwikkelingspatroon van breinuitdrukking van die gemuteerde gene (Kang *et al.* 2011; Colantuoni *et al.* 2011). In beide die hippokampus en dorsolaterale prefrontale korteks, twee breinareas geïmpliseer in skisofrenie, is die grootste effek waargeneem by gene wat die hoogste uitgedruk

© University of Pretoria

word in die prenatale periode (Barch & Ceaser 2012). Wat belangrik is, is dat daar 'n funksionele ooreenkoms is tussen die prenatale uitdrukking van teikengene en die neuro-ontwikkelingsimpak van die ooreenkomstige mutasies. Spesifiek by pasiënte wat spontane mutasies dra, was dié met mutasies in prenataal gebaseerde gene meer geneig tot veelvoudige (> 3) gedragsabnormaliteite in die eerste 10 jaar van hul lewe, sowel as 'n erger funksionele uitkoms ná die aanvang van die siekte (Sobin *et al.* 2003). Verder het die vergelyking van al die gene met funksioneel spontane gebeure in ons families met skisofrenie, saam met die identifisering van die hele spektrum outistiese versteurings (ASV), 15 gedeelde gene opgelewer. (Neale *et al.* 2012; Sanders *et al.* 2012; O'Roak *et al.* 2012; Iossifov *et al.* 2012). Dit is 'n oorvleueling binne kansverwagting ($p = 0.29$). Daar moet opgelet word dat 11 van die 15 gedeelde gene (73%) ingesluit is in ons lys van teikens vir prenatale uitdrukkingsydigheid. Die kans dat die oorvleueling 'n toevallige bevinding is, is baie laag ($p = 0.004$).

Die groot meerderheid van teikens vir prenatale uitdrukkingsydigheid kom sterk na vore gedurende die eerste en tweede trimester van swangerskap. Hulle toon 'n oorverteenwoordiging van kerngene betrokke by chromatienhermodellering, kernvervoer en transkripsiebeheer, sowel as in proteïenoordrag- en afbou. Die teikens sluit ook gene in betrokke by interaksies van sel tot sel en interaksies van sel tot matriks, 'n onderafdeling wat interaksie toon met twee sleutel-aanhegtingsproteïene: THBSI en ITGA6 (Christopherson *et al.* 2005; Georges-Labouesse *et al.* 1998). Laasgenoemde twee is ook betrokke by sinaptogenese, aksionale groei en kortikale laagvorming. Navorsers bevestig dat gene met sydigheid vir prenatale uitdrukking hoogs verryk is met mikro-RNA-teikens en daar word ook bevind dat 'n nominaal beduidende verryking van hsa-mir 367- en hsa-mir 1244-teikens teenwoordig is (Colantuoni *et al.* 2011). In vergelyking hiermee sal die risiko bygedra deur teikens van postnatale uitdrukkingsydigheid verband hou met hul betrokkenheid in intrasellulêre seinprosesse (GTPase, DAG- of die kalsium-modelleringsein), wat diverse aspekte van neuronale konnektiwiteit reguleer.

Verder het die navorsers probeer vasstel watter individuele gene meer geneig sal wees om tot die risiko van siekte by te dra. Hulle vind vier gene wat verander is deur twee spontane gebeure elk in nieverwante pasiënte (drie van hierdie gene was geaffekteer in pasiënte uit twee verskillend getoetste bevolkings). Hierdie gene is twee maal getref deur 'n kombinasie van niekoderende mutering en 'n spontane wanmutasie ENV (DPYD), of 'n kombinasie van 'n splitslassingsetel. Daar was óf 'n spontane wanmutasie ENV (TRRAP en VPS39) óf indel (invoeging of weglating) (LAMA2). Geeneen van hierdie gene is aangetas in die Afrikaner- of die SSC-kontrolegroep nie en geen so 'n wanmutasie is gerapporteer vir enige geen in die 488 spontane gebeure in die SSC-kontrole-groep nie. Gegewe die getal spontane mutasies in ons datastel, sal die observasie van vier sulke herhalende gebeure 'n p-waarde van 0.002 hê. Daar is een van die vier kombinasies (in LAMA2) wat individueel statisties beduidend is ($p = 0.017$).

LAMA2 enkodeer die laminin alpha-tweeketting, wat een van die twee subeenhede van laminin twee en vier verteenwoordig en bind aan ITGA6, 'n teiken vir prenatale uitdrukkingsydigheid. Homosigotiese mutasies in LAMA2 lei met variërende penetrasie tot oorerflike spierdistrofie, gekenmerk deur SSS-betrokkenheid, insluitend witstofabnormaliteite, kognitiewe inkorting, konvulsies en neuronale migrasiedefekte (Jones *et al.* 2001). 'n Spontane mutasie in die isoform LAMA1 is ook deur 'n ander studiegroep oor skisofrenie beskryf.

DPID (Dihidropiramidien dehidrogenase) is die aanvanklike en tydbepalende faktor in die pad van pirimidien-katabolisme (Van Kuilenburg *et al.* 1999). Dit moduleer die produksie van beta-alanien, 'n neuromodulator van transmissie-inhibering in die brein (Tiedje *et al.* 2010). Ons identifiseer een wanmutasie en een niekoderende mutasie ENV by beide die Afrikaner- en die Amerikaanse groep. Abnormale urinêre ekskresie van timien en urasiel bevestig DPID-gebrek of tekort by die wanmutasiedraer. Heterosigotiese weglatings wat óf DPID insluit óf daarin is, sowel as gewysigde uitdrukking is beskryf in die spektrum outistiese versteurings en intellektuele inkorting (Ben-David *et al.* 2011; Carter *et al.* 2011; Willemsen *et al.* 2011). Geen outistiese kenmerke of intellektuele inkorting was teenwoordig in ons draer van die DPID-mutasie nie, wat dui op 'n variasie in uitdrukkingsvermoë.

Die splitslassingmutasie in TRRAP2 beïnvloed posisie +6 tot dié se aansluitingposisie en voorspel verbreking van splitslassing en binding van die SRPSS-faktor binne-in die splitsbevorderaar. Dit is opmerklik dat TRRAP en VPS39 ook mutasies ondergaan in die spektrum outistiese versteurings (O'Roak *et al.* 2012).

Die geïdentifiseerde funksioneel spontane mutasies is vergelyk met die spontane KNV's, voorheen geïdentifiseer in ons twee groepe (22 KNV's wat 156 gene beïnvloed) (Xu *et al.* 2008; Malhotra *et al.* 2011). Vyf gene (DGCR2, TOP3B, CIT, STAG1 en SMAP2) is verander deur beide spontane ENV en KNV's, twee van hulle in pasiënte uit die twee verskillende bevolkings wat getoets is. Twee van hierdie gene val binne die 22q11.2-lokus wat vatbaarheid vir skisofrenie aandui (Xu *et al.* 2012).

Ons bevindings impliseer die bydrae van 'n diverse stel spontane mutasies met 'n relatief hoë, maar onvolledige penetrasie tot die genomiese argitektuur van skisofrenie in die konteks van 'n balansmodel vir mutasieseleksie. Dit beklemtoon die belang van die gebruik van 'n familiesteekproef, waar die siektegeskiedenis versigtig vasgestel is, om hul rol uit te lig. In dié verband, as ons fokus op ons fyn uitgewerkte Afrikanergroep, voorspel ons dat ten minste 17.6% van sporadiese gevalle 'n spontaan patogeniese, eksoniese mutasie dra en ten minste 9.9% 'n spontane KNV dra (Xu *et al* 2008; Xu *et al.* 2012). Met ander woorde, sulke mutasies verklaar 'n kwart tot 'n derde van alle sporadiese gevalle. Gegewe dat die resultate van skanderings van nie-eksoniese gebiede, nog na vore moet

© University of Pretoria

kom, verteenwoordig hierdie syfers onderrapportering. Net so belangrik is die bydrae van ons bevindings tot die verstaan van die neurale argitektuur van 'n risiko vir skisofrenie. Daar word bespiegel dat die getal loki vir skisofrenie meer as 850 is. Ons bevindings ontmasker die beduidende sensitiwiteit van die neurale netwerke wat die vatbaarheid tot skisofrenie onderlê, tot presiese vlakke of aktiwiteit van baie diverse proteïene en seinmodules. Daar is voorgestel dat die fokus op netwerke moontlik meer gelykmatig kan wees, met die heterogenisiteit van skisofrenie in ag geneem, eerder as om te fokus op ander voorgestelde meganismes wat weer toegespits is op spesifieke neurosenders of seltipes (Arguello & Gogos 2012).

Verder word aangetoon dat in die bepaling van siekterisiko, moet nie slegs die funksie van die teikengeen in ag geneem word nie. Die tydsberekening van die genetiese aanvalle is van kritiese belang. Spontane mutasies affekteer gene met diverse funksies en ontwikkelingsprofiele, maar ons omskryf 'n substansiële bydrae van mutasies in ontwikkelingsregulerende gene met hoër uitdrukkings gedurende die vroeë en midfetale leeftyd. Verder is aangetoon dat sulke mutasies verryk word in die geval van volwasse pasiënte met prominente, vroeë prepsigoties versteurde gedrag. Ons bevindings verskaf 'n meganistiese konteks waarin epidemiologiese korrelasies tussen verskeie prenatale omgewingsaanvalle gedurende die eerste en tweede trimester van swangerskap en die risiko vir skisofrenie oorweeg kan word (Gilman *et al.* 2012). Die feit dat die uitdrukking van baie gene vir prenatale sydigheid onder streng mikro-RNA-kontrole is, kan die skakeling tussen mikro-RNA-disregulasie en psigiatriese versteurings verklaar (Start *et al.* 2008). Die uitdaging wat oorbly, is om die biologiese prosesse en neurale netwerke te identifiseer en te bepaal hoe hulle geaffekteer word. Onbevooroordeelde netwerkbasis-benaderings sowel as dier- en sellulêre modelle van herhalende mutasies sal waardevol wees vir dié doel (Karayiorgou *et al.* 2012).

Slot

Baie vordering is gemaak in die genetika van skisofrenie. Die reaksie van kliniese navorsers, befondsers, families en pasiënte wat ly aan skisofrenie is dikwels dié van teleurstelling. Hierdie persone voer aan die ontdekkings verklaar slegs 'n klein gedeelte van die genetiese risiko vir skisofrenie en vertel ons eintlik min van die aard van skisofrenie.

Die *Genetic and Neural Complexity in Psychiatry 2011 Working Group* (Karayiorgou *et al.* 2012) maak 'n paar aanbevelings oor die vraag van wat die beste pad vorentoe is om verklarende brûe te bou tussen menslike genetika, diermodelle en neurale netwerke:

- Navorsingspogings en befondsing moet gerig wees op die identifikasie van unieke, hoogs penetrerende mutasies en die genereer van 'n paar dier- en sellulêre modelle, geanker in mutasies.
- Baie van die huidige teorieë oor psigiatriese siekte-etiologie moet hersien of uitgeskakel word, byvoorbeeld disregulasie van neurosendersisteme en afgeleides daarvan.
- Meer gesofistikeerde siektehipoteses soos onbevooroordeelde, netwerkgesentreerde benaderings behoort gevolg te word.
- Die beste manier om samelopende patofisiologiese meganismes vas te stel is om te begin met genetiese ontdekkings in mense, want dit gooi hipoteses oor meganismes eenkant en toets geenfunksie sonder vooroordeel.
- Vir die uiteindelike eksperimente, wat die basis sal vorm van gevolglike, funksionele en verklarende analise, benodig navorsers oorsaaklike mutasies eerder as geassosieerde variante (dit betrek polarisasie tussen aanhangers van GWAS van algemene variante en die voorstanders van unieke variantontdekkings).
- Steekproewe wat gebruik word, moet familiegebaseerd wees, waar die spontane aard van mutasies of die medeskeiding met siekte bepaal kan word – en dít bly 'n veilige hawe.
- Nuwe genetiese bevindings kan moontlik lei tot die herdefiniëring van diagnostiese kriteria, wat moontlik spesifieke terapeutiese implikasies kan hê.
- Die doelstelling moet wees om die oorsaak van siekte te verstaan, en gevolglik moet ons met die siekte self begin, eerder as die fenotipes wat 'n onbekende afstand van die siekte af is en wat geneties net so kompleks kan wees soos die siekte waarmee dit verband hou.
- Die genereer van 'n kern van sowat 'n dosyn dieremodelle vir elke siekte-entiteit, geanker op 'n werklik herhalende mutasie wat geassosieer is met goed gedefinieerde en noue kliniese presentering sal netwerkgebaseerde analise fasiliteer.
- Multivlak-analise sal nodig wees vir die toets van oorsaaklike verbindings in bevindings, op verskeie hiërargiese vlakke van die aangetaste netwerk.
- Op dieselfde manier sal vooruitgang in die herprogrammering van velselle van pasiënte se neurone die moontlikheid van sellulêre siektemodelle daarstel.

As ons die genetiese kompleksiteit in ag neem, sal hierdie benadering meer produktief en betroubaar wees as dit fokus op pasiënte wat herhalende, hoogs penetrerende mutasies dra eerder as om agnosties te wees ten opsigte van die onderliggende genotipe.

Die oorkoepelende doel is om nuwe en effektiewe behandeling te ontwikkel en om die onttrekking van farmaseutiese maatskappye uit psigiatriese siektenavorsing om te keer. Op die lang duur sal die biologiese rekonstruksie van psigiatriese diagnose die oorweging van mediese behandeling bevorder en sal die beleggingsrisiko aansienlik verminder. In die pogings om geneesmiddels te ontwikkel sal gefokus word op kleiner, biologies gestratifiseerde subgroepe van pasiënte wat gerig word deur genetiese bevindings of op spesifieke netwerke en sinaptiese prosesse. Op die kort termyn behoort bestaande geneesmiddels (met bekende veiligheidsprofiele)

© University of Pretoria

vir 'n nuwe doel geherposisioneer en aangewend te word, deur gebruik te maak van 'n toenemende getal betroubare genetiese dier- en sellulêre modelle. Daar behoort belê te word in basiese wetenskap van hoë kwaliteit om vooruitgang te maak en om verklarende brûe te bou van die meganisme tot siekte oor die klinies belangrike probleme van psigiatriese siekte (Karayiorgou *et al.* 2012).

Baie vordering is gemaak in psigiatriese genetika. Hierdie vordering beloof om ons veilig te stuur in die rigting van verbeterde psigiatriese siektenavorsing. Tans onttrek baie farmaseutiese maatskappye hulle van geneesmiddelontwikkeling. Laasgenoemde lei ons tot die slegste tye in hierdie navorsingsveld. Daar is 'n duidelike gevaar van verspeelde kanse en die sin van 'n duidelike definiëring van die pad vorentoe (Karayiorgou *et al.* 2012).

Erkenning

Die outeur het 'n samewerkingsooreenkoms met die laboratorium vir neurogenetika van Columbia University, New York, onder leiding van Prof. M. Karayiorgou. Die outeur het 'n gradering van die Nasionale Navorsingstigting (NS) en ontvang sedert 2010 'n jaarlikse aansporingsbeurs.

Mededingende belange

Die outeur verklaar hiermee dat hy geen finansiële of persoonlike verbintenis het met enige party wat hom nadelig of voordelig kon beïnvloed het in die skryf van hierdie artikel nie.

Literatuurverwysings

American Psychiatric Association, 1994, *Diagnostic and Statistical Manual of Mental disorders: DSM-IV*, American Psychiatric Association, Washington D.C.

Arguello, P.A. & Gogos, J.A., 2012, 'Genetic and cognitive windows into circuit mechanisms of psychiatric disease', *Trends in Neurosciences* 35, 3–13. http://dx.doi.org/10.1016/j.tins.2011.11.007

Barch, D.M. & Ceaser, A., 2012, 'Cognition in schizophrenia: Core psychological and neural mechanisms', *Trends in Cognitive Sciences* 16, 27–34. http://dx.doi.org/10.1016/j.tics.2011.11.015

Bassett, A.S., Chow, E.W., AbdelMalik, P., Gheorghiu, M., Husted, J. & Weksberg, R., 2003, 'The schizophrenia phenotype in 22q11 deletion syndrome', *American Journal of Psychiatry* 160, 1580–1586. http://dx.doi.org/10.1176/appi.ajp.160.9.1580

Ben-David, E., Granot-Hershkovitz, E., Monderer-Rothkoff, G., Lerer, E., Levi, S. & Yaari, M., 2011, 'Identification of a functional rare event in autism using genome-wide screen for mono allelic expressions', *Human Molecuar Genetics* 20, 3632–3641. http://dx.doi.org/10.1093/hmg/ddr283

Carter, M.T., Nikkel, S.M., Fernandez, B.A., Marshall, C.R., Noor, A. & Lionel, A.C., 2011, 'Hemizygous deletions on chromosome 1p21.3 involving the DPYD gene in individuals with autism spectrum disorder', *Clinical Genetics* 80, 435–443. http://dx.doi.org/10.1111/j.1399-0004.2010.01578.x

Christopherson, K.S., Ullian, E.M., Stokes, C.C.A., Mullowney, C.E., Hell, J.W. & Agah, A., 2005, 'Thrombospondins are astrocyte-secreted proteins that promote CNS synaptogenesis', *Cell* 120, 421–433. http://dx.doi.org/10.1016/j.cell.2004.12.020

Cohen, J.C., Kiss, R.S., Pertsemlidis, A., Marcel, Y.L., McPherson, R. & Hobbs, H,H., 2004, 'Multiple rare alleles contribute to low plasma levels of HDL cholesterol', *Science* 305, 869–872. *http://dx.doi.org/10.1126/science.1099870*

Colantuoni, C., Lipska, B.K., Ye, T., Hyde, T.M., Tao, R. & Leek, J.T., 2011, 'Temporal dynamics and genetic control of transcription in the human prefrontal cortex', *Nature* 478, 519–523. http://dx.doi.org/10.1038/nature10524

Crow, J.F., 2000, 'The origins, patterns and implications of human spontaneuous mutation', *Nature Reviews Genetics* 1, 40–47. http://dx.doi.org/10.1038/35049558

Flanagan, S.E., Patch, A.M. & Ellard, S., (2010), 'Using SIFI and PolyPhento predict loss-of-function and gain-of-function mutations', *GeneticTesting and Molecular Biomarkers* 14, 533–537. http://dx.doi.org/10.1089/gtmb.2010.0036

Georges-Labouesse, E., Mark, M., Messaddeq, N. & Gansmuller, A., 1998, 'Essential role of a6 integrins in cortical and retinal lamination', *Current Biology* 8, 983–986. http://dx.doi.org/10.1016/S0960-9822(98)70402-6

Gilman, S.R., Chang, J., Xu, B., Bawa, T.S., Gogos, J.A. & Karayiorgou M., 2012, 'Diverse types of genetic variation converge on functional gene networks involved in schizophrenia', *Nature Neuroscience* 12, 1723–1728. http://dx.doi.org/10.1038/nn.3261

Girard, S.L., Gauthier, J., Noreau, A., Xiong, L., Zhou, S. & Jouan, L., 2011, 'Increased exonic de novo mutation rate in individuals with schizophrenia', *Nature Genetics* 43, 860–863. http://dx.doi.org/10.1038/ng.886

Goldstein, D.B., 2009, 'Common genetic variation and human traits', *New England Journal of Medicine* 360, 1696–1698. http://dx.doi.org/10.1056/NEJMp0806284

Gothelf, D., Feinstein, C., Thompson, T., Gu, E., Penniman, L. & Van Stone, E., 2007, 'Risk factors for the emergence of psychotic disorders in adolescents with 22q11.2 deletion syndrome', *American Journal of Psychiatry* 164, 663–669. http://dx.doi.org/10.1176/appi.ajp.164.4.663

Iafrate, AJ., Feuk, L., Rivera, M.N., Listewnik, M.L., Donahoe, P.K. & Qi, Y. 2004, 'Detection of large-scale variation in the human genome', *Nature Genetics* 36, 949–951. http://dx.doi.org/10.1038/ng1416

Iossifov, I., Ronemus, M., Levy, D., Wang, Z., Hakker, I. & Rosenbaum, J., 2012, 'De novo gene disruptions in children on the autistic spectrum', *Neuron* 74, 285–299. http://dx.doi.org/10.1016/j.neuron.2012.04.009

Karayiorgou, M., Flint, J., Gogos, J. & Malenka, R.C., 2012, 'The best of times, the worst of times for psychiatric disease', *Nature Neuroscience* 15, 811–812. http://dx.doi.org/10.1038/nn.3115

Karayiorgou, M., Morris, M.A., Morrow, B., Shprintzen, R.J., Goldberg, R. & Borrow, J., 1995, 'Schizophrenia succeptability associated with interstitial deletions of chromosome 22q11', *Proceedings of the National Academy of Sciences* 92, 7612–7616. http://dx.doi.org/10.1073/pnas.92.17.7612

Jones K.J., Morgan, G., Johnston, H., Tobias, V., Ouvrier, R.A., Wilkinson, I. & North K.N., 2001, 'The expanding phenotype of laminin a2 cahin (merosin) abnormalities: Case series and review', *Journal of Medical Genetics* 38, 649–657. http://dx.doi.org/10.1136/jmg.38.10.649

Kang, H.J., Kawasawa Y.I. Cheng, F., Zhu, Y., Xu, X. & Li, M., 2011, 'Spatio-temporal transcriptome of the human brain', *Nature* 478, 483–489. http://dx.doi.org/10.1038/nature10523

Laursen, T.M. & Munk-Olsen T., 2010, 'Reproductive patterns in psychotic patients, *Schizophrenia Research'*, 121, 234–240. http://dx.doi.org/10.1016/j.schres.2010.05.018

Lichtenstein, P., Yip, B.H., Bjork, C., Pawitan, Y., Cannon, T.D.,& Sullivan, P.F., 2009, 'Common genetic determinants of schizophrenia and bipolar disorder in Swedish families: A population-based study', *Lancet* 373, 234–239. http://dx.doi.org/10.1016/S0140-6736(09)60072-6

Lupski, J.R., 2007, 'Structural variation in the human genome', *New England Journal of Medicine* 356, 1169–1171. http://dx.doi.org/10.1056/NEJMcibr067658

Malhotra D., McCarthy, S., Michaelson, J.J., Vacic, V., Burdick, KE. & Yoon, S., 2011, 'High frequency of de novo CNVs in bipolar disorder and schizophrenia', *Neuron* 72, 951–963. http://dx.doi.org/10.1016/j.neuron.2011.11.007

Manolio T.A., Collins, F.S., Cox, N.J., Goldstein, D.B., Hindorff, L.A., Hunter, D.J., 2009, 'Finding the missing heritability of complex diseases', *Nature* 461, 747–753. http://dx.doi.org/10.1038/nature08494

Murphy, K.C., Jones, L.A. & Owen, MJ., 1999, 'High rates of schizophrenia in adults with velo-cardio-facial syndrome', *Archives of General Psychiatry* 53, 940–945. http://dx.doi.org/10.1001/archpsyc.56.10.940

Neale, B.M., Kou, Y., Liu, L., Ma'ayan, A., Samocha, K.E. & Sabo, A., 2012, 'Patterns and rates of exonic de novo mutations in autism spectrum disorders', *Nature* 485, 242–245. http://dx.doi.org/10.1038/nature11011

O'Roak, B.J., Vives, L., Girirajan, S., Karakoc, E., Krumm, N. & Coe, B.P., 2012, 'Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations', *Nature* 485, 246–250. http://dx.doi.org/10.1038/nature10989

Pulver, A.E., Nestadt, G., Goldbert, R., Shprintzen, R.J., Lamacz, M. & Wolyniec, P.S., 1994, 'Psychotic illness in patients diagnosed with velo-cardio-facial syndrome and their relatives', *Journal of Nervous and Mental Disease* 182, 476–478. http://dx.doi.org/10.1097/00005053-199408000-00010

Redon, R., Ishikawa, S., Fitch, K.R., Feuk, L., Perry, G.H. & Andrews, T.D., 2006, 'Global variation in copy number in the human genome', *Nature* 444, 444–454. http://dx.doi.org/10.1038/nature05329

Rodriguez-Murillo, L., Gogos, J.A. & Karayiorgou M., 2012 'The genetic architecture of schizophrenia: New mutations and emerging paradigms', *Annual Review of Medicine* 63, 63–80. http://dx.doi.org/10.1146/annurev-med-072010-091100

Roos, J.L., 2008, 'Die Afrikaner en skisofrenie: 'n Genetiese studie', *Suid-Afrikaanse Tydskrif vir Natuurwetenskap en Tegnologie* 27(2), 73–87. http://dx.doi.org/10.4102/satnt.v27i2.83

Sanders, S.J., Murtha, M.T., Gupta, A.R., Murdoch, J.D., Raubeson, M.J. & Willsey, A.J., 2012, 'De novo mutations revealed by whole-exome sequencing are strongly associated with autism', *Nature* 485, 237–241. http://dx.doi.org/10.1038/nature10945

Sebat, J., Lakshmi , B., Troge, J., Alexander, J., Young, J. & Lundin, P., 2004, 'Large-scale copy number polymorphism in the human genome', *Science* 305, 525–528. http://dx.doi.org/10.1126/science.1098918

Sobin, C., Roos, J.L., Pretorius, H., Lundy, L.S. & Karayiorgou, M., 2003, 'A comparison study of early non-psychotic deviant behaviour in Afrikaner and US patients with schizophrenia or schizoaffective disorder', *Psychiatry Research* 117, 113–125. http://dx.doi.org/10.1016/S0165-1781(02)00321-9

Start, K.L., Xu, B., Bagchi, A., Lai, W., Liu, H. & Hsu, R., 2008, 'Altered brain micro RNA brogenesis contributes to phenotypic deficits in a 22q11-deletion mouse model', *Nature Genetics* 40, 751–760. http://dx.doi.org/10.1038/ng.138

© University of Pretoria

Stefansson, H., Rujescu, D., Cichon, S., Pietiläinen, OPH., Ingason, A. & Steinberg, S., 2008, 'Large recurrent microdeletions associated with schizophrenia', *Nature* 445, 232–236. http://dx.doi.org/10.1038/nature07229

Sullivan, P.F. Kendler, K.S. & Neale,MC., 2003, 'Schizophrenia as a complex trait: Evidence from a meta-analysis of twin studies', *Archives of General Psychiatry* 60, 1187–1192. http://dx.doi.org/10.1001/archpsyc.60.12.1187

Tandon, R., Keshaven M.S. & Nasrallah H.A., 2008, 'Schizophrenia, "just the facts": What we know in 2008:2. Epidemiology and etiology', *Schizophrenia Research* 102, 1–18. http://dx.doi.org/10.1016/j.schres.2008.04.011

The 1000 Genomes Project Consortium, 2010, 'A map of human genome variation from population-scale sequencing', *Nature* 467, 1061–1073.

Tiedje, K.E., Stevens, K., Barnes, S. & Waever, D.F., 2010, 'b alanine as a small molucule neurotransmitter', *Neurochemistry International* 57, 177–188. http://dx.doi.org/10.1016/j.neuint.2010.06.001

Valdmanis, P.N., Verlaan, D.J. & Rouleau, G.A., 2009, 'The proportion of mutations predicted to have a deleterious effect differs between gain and loss of function genes in neurodegenerative disease', *Human Mutatation* 30, 481–489. http://dx.doi.org/10.1002/humu.20939

Van Kuilenburg, A.B., Vreken, P., Abeling, N.G.G.M., Bakker, H.D., Meinsma, R. & Van Lenthe, H.,1999, 'Genotype and phenotype in patients with dihydropyrimidine dehydrogenase deficiency', *Human Genetics* 104, 1–9. http://dx.doi.org/10.1007/PL00008711

Wiehahn, G.J., Bosch, G.P., Du Preez, R.R., Pretorius, H.W., Karayiorgou, M. & Roos, J.L., 2004, 'Assessment of the frequency of the 22q11 deletion in Afrikaner schizophrenic patients', *The American Journal of Medical Genetics* 124B(1), 1–3. http://dx.doi.org/10.1136/jmedgenet-2011-100294

Willemsen, M.H., Vallès, A., Kirkels, L.A.M.H., Mastebroek, M., Loohuis, N.O. & Kos, A., 2011, 'Chromosome 1p22.3 microdeletions comprising DPYD and MIR 137 are associated with intellectual disability', *Journal of Medical Genetics* 48, 810–818.

Xu, B., Roos, J.L., Levy, S., Van Rensburg, E.J., Gogos, J.A., Karayiorgou, M., 2008, 'Strong association of de novo copy number mutations with sporadic schizophrenia', *Nature Genetics* 40, 880–885. http://dx.doi.org/10.1038/ng.162

Xu, B., Woodroffe, A., Rodriquez-Murillo, L., Roos, J.L., Van Rensburg, E.J. & Abecasis, G.R., 2009, 'Elucidating the genetic architecture of familial schizophrenia using rare copy number variant and linkage scans', *Proceedings of the NationalAcademy of Sciences* 106, 16746–16751. http://dx.doi.org/10.1073/pnas.0908584106

Xu, B., Roos, J.L., Dexheimer, P., Boone, B., Plummer, B. & Levy, S., 2011, 'Exome sequencing supports a de novo mutational paradigm for schizophrenia', *Nature Genetics* 43, 864–868. http://dx.doi.org/10.1038/ng.902

Xu, B., Ionita-Laza, I., Roos, J.L., Boone, B., Woodrick, S. & Sun, Y., 2012, 'De novo gene mutations highlight patterns of genetic and neural complexity in schizophrenia', *Nature Genetics* 42, 1365–1371. http://dx.doi.org/10.1038/ng.2446

Yang, J., Visscher, PM. & Wray, N.R., 2010, 'Sporadic cases are the norm for complex disease', *European Journal of Human Genetics* 18, 1039–1043. http://dx.doi.org/10.1038/ejhg.2009.177

© University of Pretoria