Genome-wide association study identifies five new schizophrenia loci

Citation for published version:

Digital Object Identifier (DOI):
10.1038/ng.940

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Peer reviewed version

Published In:
Nature Genetics

Publisher Rights Statement:

General rights
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.
Genome-wide association study identifies five new schizophrenia loci

© 2011 Nature America, Inc. All rights reserved.

A full list of authors and affiliations appears at the end of the paper.

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

AUTHOR CONTRIBUTIONS

COMPETING FINANCIAL INTERESTS
The authors declare competing financial interests: details accompany the full-text HTML version of the paper at http://www.nature.com/naturegenetics/.

Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

Institute of Neuroscience and Medicine (INM-1), Research Center Juelich, Juelich, Germany.

Center for Psychiatric Neuroscience, The Feinstein Institute for Medical Research, Manhasset, New York, USA.

Department of Psychiatry, University of Erlangen-Nuremberg, Erlangen, Germany.

Published in final edited form as:

Nat Genet. ; 43(10): 969–976. doi:10.1038/ng.940.
The Schizophrenia Psychiatric Genome-Wide Association Study (GWAS) Consortium

Abstract

We examined the role of common genetic variation in schizophrenia in a genome-wide association study of substantial size: a stage 1 discovery sample of 21,856 individuals of European ancestry and a stage 2 replication sample of 29,839 independent subjects. The combined stage 1 and 2 analysis yielded genome-wide significant associations with schizophrenia for seven loci, five of which are new (1p21.3, 2q32.3, 8p23.2, 8q21.3 and 10q24.32-q24.33) and two of which have been previously implicated (6p21.32-p22.1 and 18q21.2). The strongest new finding \( P = 1.6 \times 10^{-11} \) was with rs1625579 within an intron of a putative primary transcript for \( MIR137 \) (microRNA 137), a known regulator of neuronal development. Four other schizophrenia loci achieving genome-wide significance contain predicted targets of \( MIR137 \), suggesting \( MIR137 \)-mediated dysregulation as a previously unknown etiologic mechanism in schizophrenia. In a joint analysis with a bipolar disorder sample (16,374 affected individuals and 14,044 controls), three loci reached genome-wide significance: \( CACNA1C \) (rs4765905, \( P = 7.0 \times 10^{-9} \)), \( ANK3 \) (rs10994359, \( P = 2.5 \times 10^{-8} \)) and the \( ITIH3-ITIH4 \) region (rs2239547, \( P = 7.8 \times 10^{-9} \)).

In stage 1, we conducted a mega-analysis combining genome-wide association study (GWAS) data from 17 separate studies (with a total of 9,394 cases and 12,462 controls; Table 1 and Supplementary Tables 1,2). We imputed allelic dosages for 1,252,901 autosomal SNPs (Table 1, Supplementary Table 3 and Supplementary Note) using HapMap3 as the reference panel. We tested for association using logistic regression of imputed dosages with sample identifiers and three principal components as covariates to minimize inflation in significance testing caused by population stratification. The quantile-quantile plot (Supplementary Fig. 1) deviated from the null distribution with a population stratification inflation factor of \( \lambda = 1.23 \). However, \( \lambda_{1000} \), a metric that standardizes the degree of inflation by sample size, was only 1.02, similar to that observed in other GWAS meta-analyses. This deviation persisted despite comprehensive quality control and inclusion of up to 20 principal components (Supplementary Fig. 1). Thus, we interpret this deviation as indicative of a large number of weakly associated SNPs consistent with polygenic inheritance. We also examined 298 ancestry-informative markers (AIMs) that reflect European-ancestry population substructure. Unadjusted analyses showed greater inflation in the test statistics than we saw for all markers (AIMs \( \lambda = 2.26 \) compared to all markers \( \lambda = 1.56 \)). After inclusion of principal components, the distributions of the test statistics did not differ between AIMs (\( \lambda = 1.18 \)) and all markers (\( \lambda = 1.23 \)), a result inconsistent with population stratification explaining the residual deviation seen in Supplementary Figure 1. Moreover, the results of a meta-analysis using summary results generated using study specific principal components (Supplementary Note) were highly correlated with those from the mega-analysis (Pearson correlation = 0.94, with a similar \( \lambda = 1.20 \); Supplementary Fig. 2). Of the ten SNPs in Table 2, four increased and six decreased in significance, suggesting that the most extreme values did not result from systematic inflation artifacts. Therefore, our primary analysis used unadjusted \( P \) values (nevertheless, see Table 2 for stage 1 \( P \) values adjusted for \( \lambda \) ref. 6).

In stage 1 (Table 2, Supplementary Table 4 and Supplementary Figs. 3 and 4), 136 associations reached genome-wide significance (\( P < 5 \times 10^{-8} \)). The majority of these associations (\( N = 129 \)) mapped to 5.5 Mb in the extended major histocompatibility complex (MHC, 6p21.32-p22.1), a region of high linkage disequilibrium (LD) previously implicated in schizophrenia in a subset of the samples used here. The other stage 1 regions included new regions (10q24.33 and 8q21.3) and previously reported regions (18q21.2 at \( TCF4 \) (encoding transcription factor 4) and 11q24.2 (ref. 8)). The signal at 11q24.2 is \( \sim 0.85 \) Mb.
from *NRGN* (encoding neurogranin) and is uncorrelated with the previously associated variant near this gene.

In Table 2 and Supplementary Table 4, we denote regions of association by the most significant marker. Associated SNPs with $r^2 \geq 0.2$ in HapMap3 (CEU+TSI populations) were not considered independent. However, we noticed instances where multiple SNPs within 250 kb of each other yielded evidence for association ($P < 10^{-5}$) despite weak LD ($r^2 < 0.2$) between them. For regions with $P < 10^{-6}$, we performed a conditional analysis using as covariates the dosages of the strongest associated SNP, principal components 1–4 and 6 and study indicator. We observed multiple statistically independent signals at the MHC. Although a number of SNPs within the MHC were potentially independent per HapMap $r^2$ values, only rs9272105 withstood formal conditional analysis, showing $P = 1.8 \times 10^{-6}$ conditional on association to the best SNP, rs2021722 (stage 1 $P = 4.3 \times 10^{-11}$, inter-SNP distance = 2.4 Mb, $r^2 = 0.01$ in HapMap). Excluding the MHC region, we identified six regions with at least one SNP associated at $P < 10^{-3}$ and a second SNP with a conditionally independent $P < 10^{-3}$ (Supplementary Table 5). We performed 100 simulations after permuting case-control status randomly within each study. In contrast to the six regions in the real dataset, we never observed more than a single region with co-localized statistically independent signals in any simulated genome-wide scan, indicating our observation is highly unlikely to have occurred by chance.

Noteworthy co-localizing independent signals occurred at three regions (Supplementary Table 5): one region with a genome-wide significant association at 10q24.32-q24.33 (Table 2), a second region that nearly met this threshold at *MADIL1* (encoding mitotic arrest deficient-like 1; rs10226475, $P = 5.06 \times 10^{-8}$; Supplementary Table 4) and a third region at *CACNA1C* (encoding calcium channel, voltage-dependent, L type, α1C subunit), the latter of which has previously been associated with bipolar disorder and other psychiatric phenotypes including schizophrenia. The conditionally independent signal at *CACNA1C* was more significant than any observation made in 100 permutations of the entire experiment (both conditional $P < 10^{-5}$) and supports *CACNA1C* in schizophrenia after genome-wide correction ($P < 0.01$), even without considering these prior reports.

In stage 2, we evaluated in 29,839 independent subjects (8,442 cases and 21,397 controls) the most significant SNPs ($N = 81$) in each LD region where at least one SNP had surpassed $P < 2 \times 10^{-5}$ (Supplementary Table 6) in the mega-analysis. Of 22 SNPs from the MHC, 5 surpassed the genome-wide significant threshold in stages 1 and 2 combined (minimum $P = 2.2 \times 10^{-12}$ at rs2021722; Supplementary Table 6). Excluding the MHC region, a sign test for consistency between stages 1 and 2 was highly significant ($P < 10^{-6}$), with the same direction of effect as observed stage 1 also being observed in stage 2 for 49 of 59 SNPs. A Fisher’s combined test revealed the distribution of stage 2 $P$ values was unlikely to have occurred by chance ($P < 10^{-15}$). We also performed a transmission analysis using the family based Multicenter Pedigree replication sample in conjunction with a GWAS of 622 parent-offspring schizophrenia trios from Bulgaria, and the stage 1 associated allele was over-transmitted to cases for 44 of the 59 SNPs (one-sided $P = 1.0 \times 10^{-4}$). Thus, the stage 2 replication results are highly consistent with the stage 1 discovery results.

In the combined dataset (stages 1 and 2), five new (1p21.3, 2q32.3, 8p23.2, 8q21.3 and 10q24.32-q24.33) and two previously reported (6p21.32-p22.1 and 18q21.2) loci met genome-wide significance (Figs. 1, 2, Table 2, Supplementary Tables 6,7 and Supplementary Fig. 4). After adjusting for λ (ref. 6), four loci (1p21.3, 6p21.32-p22.1, 10q24.32-q24.33 and 18q21.2) remained significant at $P \leq 5 \times 10^{-8}$. For the primary analyses (unadjusted for λ), the strongest new association was at 1p21.3 (rs1625579; $P = 1.6 \times 10^{-11}$), which is over 100 kb from any RefSeq protein-coding gene but is within intron 3 of AK094607, which...
observations suggest
The proportion of variance (Nagelkerke’s training set), and the testing set consisted of 6,428 individuals independent of the ISC report. (Supplementary Note). The training set had 15,429 subjects (over twice the size of the ISC polygenic model, dividing stage 1 samples into independent training and testing sets greatly minimizing the possibility of population stratification artifact contribution to schizophrenia. The International Schizophrenia Consortium (ISC) reported evidence for a polygenic schizophrenia. Because SNP effects are estimated with error estimate is much lower than the true total variation in liability that is tagged by all SNPs in the ISC to around 6% here (Supplementary Table 9 and Supplementary Fig. 6). This sign test (loci that did not reach genome-wide significance in the combined stage 1 and 2 analysis, a most highly significant results of our stage 1 analysis. Supporting this hypothesis, of the top other complex traits explain a substantial fraction of the heritability of schizophrenia.

MIR137 has been implicated in regulating adult neurogenesis, and neuronal maturation, mechanisms through which variation at this locus could contribute to brain development abnormalities in schizophrenia. Of relevance, two independent schizophrenia imaging studies found MIR137 to be one of three microRNAs with targets significantly enriched for association. In stage 1, SNPs in or near 301 high-confidence predicted MIR137 targets (with a TargetScan probability of conserved targeting ≥0.9) were enriched for association compared with genes matched for size and marker density: 17 predicted MIR137 targets (Supplementary Table 8) had at least one SNP with $P < 10^{-4}$, which is more than twice as many as the control gene sets ($P < 0.01$). Excluding the MHC and MIR137, of the nine loci with genome-wide significant support either in stage 1 or in the combined set (six loci, 2q32.3, 8p23.2, 8q21.3, 10q24.32-q24.33, 11q24.2 and 18q21.2; Table 2 and Supplementary Tables 6,7) or in a joint analysis with bipolar disorder (three genes, CACNA1C, ANK3 and ITIH3-ITIH4, described below), four genes (TCF4, CACNA1C, CSMD1 and C10orf26) have predicted MIR137 target sites according to analyses using three different prediction programs (TargetScan, PicTar and miRanda). In vitro overexpression and locked nucleic acid-mediated knockdown of MIR137 in neuronal cell line N2a leads to changes in expression levels of TCF4 protein, strongly supporting the prediction that TCF4 is a target of MIR137 (L.-H. Tsai, personal communication). Our observations suggest MIR137-mediated dysregulation as a new etiologic mechanism in schizophrenia.

The International Schizophrenia Consortium (ISC) reported evidence for a polygenic contribution to schizophrenia. An independent family based study confirmed these results, greatly minimizing the possibility of population stratification artifact. We reevaluated the polygenic model, dividing stage 1 samples into independent training and testing sets (Supplementary Note). The training set had 15,429 subjects (over twice the size of the ISC training set), and the testing set consisted of 6,428 individuals independent of the ISC report. The proportion of variance (Nagelkerke’s $r^2$) explained in the testing set increased from 3% in the ISC to around 6% here (Supplementary Table 9 and Supplementary Fig. 6). This estimate is much lower than the true total variation in liability that is tagged by all SNPs because SNP effects are estimated with error. The polygenic model appears to explain a substantial fraction of the heritability of schizophrenia, as has been shown for other complex traits. Some of these additional risk loci are likely contained near the most highly significant results of our stage 1 analysis. Supporting this hypothesis, of the top loci that did not reach genome-wide significance in the combined stage 1 and 2 analysis, a sign test ($P < 10^{-5}$) and a Fisher’s combined test ($P < 10^{-5}$) both showed an excess of same-direction allelic association (41 of 51 non-MHC SNPs) in the discovery and replication datasets.

Clinical, epidemiological and genetic findings suggest shared risk factors between bipolar disorder and schizophrenia. In stage 1, three genes with strong support had prior genome-wide significant associations with bipolar disorder: CACNA1C, the region containing ITIH3-ITIH4 (encoding inter-α (globulin) inhibitors H3 and H4) and ANK3 (encoding ankyrin 3, 

Nat Genet. Author manuscript; available in PMC 2012 October 1.
We performed a joint analysis with the Schizophrenia Psychiatric Genome-Wide Association Study (GWAS) Consortium (PGC) for bipolar disorder applying identical analytical methods. After removing duplicate subjects, we analyzed 16,374 cases with schizophrenia, schizoaffective disorder or bipolar disorder and 14,044 controls. Support for shared susceptibility was strengthened (Supplementary Table 11) at *CACNA1C* (rs4765905, \( P = 7.0 \times 10^{-9} \)), *ANK3* (rs10994359, \( P = 2.5 \times 10^{-8} \)) and the *ITIH3-ITIH4* region (rs2239547, \( P = 7.8 \times 10^{-9} \)), each of which reached genome-wide significance. A coding variant in *ITIH4* (p.Pro698Thr; rs4687657) is in perfect LD with the most associated SNP. Although we included all subjects from an earlier report\(^{10}\), the increased support found with additional independent cases (\( N = 11,987 \)) and controls (\( N = 7,835 \)) provides further evidence for shared risk effects of schizophrenia and bipolar disorder.

The risk variants implicated here confer small risks (odds ratios \( \sim 1.10 \)), but the polygenic analysis shows many more susceptibility variants with effects for which our sample is underpowered (Supplementary Table 12). At every stage where samples were added, we found an increase in the number of genome-wide significant loci and enhancement of signals at *CACNA1C*, *ANK3* and *ITIH3-ITIH4* when schizophrenia and bipolar disorder were jointly analyzed. Thus, gains in power offset any penalty for increased heterogeneity.

In summary, we report seven genome-wide significant schizophrenia associations (five of which are new) in a two-stage analysis of 51,695 individuals. We also report loci that confer susceptibility to both bipolar disorder and schizophrenia. The association near *MIR137*, associations in multiple predicted *MIR137* targets and the known role of *MIR137* in neuronal maturation and function together suggest an intriguing new insight into the pathogenesis of schizophrenia.

**URLs**


**METHODS**

Methods and any associated references are available in the online version of the paper at http://www.nature.com/naturegenetics/.

**Supplementary Material**

Refer to Web version on PubMed Central for supplementary material.

**Acknowledgments**

We thank the study participants and the research staff at the many study sites. Over 40 US National Institutes of Health grants and similar numbers of government grants from other countries, along with substantial private and foundation support, enabled this work. We greatly appreciate the sustained efforts of T. Lehner (National Institute of Mental Health) on behalf of the Schizophrenia Psychiatric Genome-Wide Association Study (GWAS) Consortium (PGC). Detailed acknowledgments, including grant support, are listed in the

**References**


Appendix

Stephan Ripke1, Alan R Sanders2,3, Kenneth S Kendler4–6, Douglas F Levinson7, Pamela Sklar1,8, Peter A Holmans9,10, Dan-Yu Lin11, Jubao Duan2,3, Roel A Ophoff12–15, Ole A Andreassen16,17, Edward Scolnick18, Sven Cichon19–21, David St. Clair22, Aiden Corvin23, Hugh Gurling24, Thomas Werge25, Dan Rujescu26, Douglas H R Blackwood27, Carlos N Pato28, Anil K Malhotra5,29,30, Shaun Purcell18, Frank Dudbridge32, Benjamin M Neale18, Lizzy Rossin1, Peter M Visscher33, Danielle Posthumus34,35, Douglas M Ruderfer1, Ayman Fanous36,37, Hreinn Stefansson38, Stacy Steinberg38, Bryan J Mowry39,40, Vera Golimbet41, Marc De Hert42, Erik G Jönsson43, István Bitter44, Olli P H

1Center for Human Genetic Research, Massachusetts General Hospital, Boston, Massachusetts, USA. 2Department of Psychiatry and Behavioral Sciences, NorthShore University HealthSystem, Evanston, Illinois, USA. 3Department of Psychiatry and Behavioral Sciences, University of Chicago, Chicago, Illinois, USA. 4Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University School of Medicine, Richmond, Virginia, USA. 5Department of Human and Molecular Genetics, Virginia Commonwealth University School of Medicine, Richmond, Virginia, USA. 6Department of Psychiatry and Behavioral Sciences, Stanford University, Stanford, California, USA. 7Department of Psychiatry, Mount Sinai School of Medicine, New York, New York, USA. 8Medical Research Council (MRC) Centre for Neuropsychiatric Genetics and Genomics, School of Medicine, Cardiff University, Cardiff, UK. 9Department of Psychological Medicine and Neurology, School of Medicine, Cardiff University, Cardiff, UK. 10Department of Biostatistics, University of North Carolina, Chapel Hill, North Carolina, USA. 11Department of Medical Genetics, University Medical Center Utrecht, Utrecht, The Netherlands. 12Division of Psychiatry, University of Edinburgh, Royal Edinburgh Hospital, Edinburgh, UK. 13Keck School of Medicine, University of Southern California, Los Angeles, California, USA. 14Department of Psychiatry, Division of Research, The Zucker Hillside Hospital Division of the North Shore-Long Island Jewish Health System, Glen Oaks, New York, USA. 15Department of Psychiatry and Behavioral Science, Albert Einstein College of Medicine of Yeshiva University, New York, New York, USA. 16Department of Non-Communicable Disease Epidemiology, London School of Hygiene and Tropical Medicine, London, UK. 17Queensland Statistical Genetics Laboratory, Queensland Institute of Medical Research, Brisbane, Queensland, Australia. 18Vrije Universiteit (VU), Center for Neurogenomics and Cognitive Research (CNCR), Department of Functional Genomics, Amsterdam, The Netherlands. 19VU Medical Centre, Department of Medical Genomics, Amsterdam, The Netherlands. 20Department of Psychiatry, Virginia Commonwealth University School of Medicine, Richmond, Virginia, USA. 21Washington Veteran’s Affairs Medical Center, Washington, DC, USA. 22Department of Psychiatry, Georgetown University School of Medicine, Washington, DC, USA. 23Department of Neurology, University of Cambridge, Cambridge, Cambridge, UK. 24Molecular Psychiatry Laboratory, Research Department of Mental Health Sciences, University College London Medical School, Windeyer Institute of Medical Sciences, London, UK. 25Institute of Biological Psychiatry, Mental Health Center (MHC) St. Hans, Copenhagen University Hospital, Roskilde, Denmark. 26Molecular and Clinical Neurobiology, Department of Psychiatry, Ludwig-Maximilians-University, Munich, Germany. 27Division of Psychiatry, University of Edinburgh, Royal Edinburgh Hospital, Edinburgh, UK. 28Queensland Brain Institute, University of Queensland, Brisbane, Queensland, Australia. 29Queensland Centre for Mental Health Research, University of Queensland, Brisbane, Queensland, Australia. 30Mental Health Research Center, Russian Academy of Medical Sciences, Moscow, Russia. 31University Psychiatric Centre, Catholic University Leuven, Kortenberg, Belgium. 32Department of Clinical Neuroscience, Human Brain Informatics (HUBIN) Project, Karolinska Institutet and Hospital, Stockholm, Sweden.

Semmelweis University, Department of Psychiatry and Psychotherapy, Budapest, Hungary.  
Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, Finland.  
Department of Molecular Genetics, University of Helsinki, Helsinki, Finland.  
Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, King’s College, London, UK.  
Section of Psychiatry and Clinical Psychology, University of Verona, Verona, Italy.  
Department of Research, Diakonhjemmet Hospital, Oslo, Norway.  
State Mental Hospital, Haar, Germany.  
Department of Psychiatry, Behavioral Sciences, Emory University, Atlanta, Georgia, USA.  
Department of Psychiatry and Behavioral Sciences, Atlanta Veterans Affairs Medical Center, Atlanta, Georgia, USA.  
Department of Psychiatry, University of Iowa Carver College of Medicine, Iowa City, Iowa, USA.  
Institute of Human Genetics, University of Aarhus, Aarhus, Denmark.  
Centre for Psychiatric Research, Aarhus University Hospital, Risskov, Denmark.  
University of Queensland Diamantina Institute, Princess Alexandra Hospital, University of Queensland, Brisbane, Queensland, Australia.  
University Medical Center Groningen, Department of Psychiatry, University of Groningen, Groningen, The Netherlands.  
School of Nursing, Louisiana State University Health Sciences Center, New Orleans, Louisiana, USA.  
Department of Psychiatry, University of California at San Francisco, San Francisco, California, USA.  
UCIR (Northern California Institute for Research and Education), San Francisco, California, USA.  
Department of Psychiatry, Rudolf Magnus Institute of Neuroscience, University Medical Center Utrecht, Utrecht, The Netherlands.  
University of California at Los Angeles (UCLA) Center for Neurobehavioral Genetics, University of California at Los Angeles, Los Angeles, California, USA.  
School of Psychiatry, University of New South Wales and Schizophrenia Research Institute, Sydney, New South Wales, Australia.  
Department of Psychiatry, University of Queensland, Royal Brisbane Hospital, Brisbane, Australia.  
Department of Psychiatry, Washington University, St. Louis, Missouri, USA.  
Academic Medical Centre, University of Amsterdam, Department of Psychiatry, Amsterdam, The Netherlands.  
Department of Psychiatry, University of Athens Medical School, Athens, Greece.  
Department of Medical Genetics, Oslo University Hospital, Oslo, Norway.  
Wellcome Trust Centre for Human Genetics, Oxford, UK.  
Department of Statistics, University of Oxford, Oxford, UK.  
Mental Health Center Copenhagen, Copenhagen University Hospital, Copenhagen, Denmark.  
Department of Psychiatry, University of Colorado Denver, Aurora, Colorado, USA.  
Center for Clinical Intervention and Neuropsychiatric Schizophrenia Research, Mental Health Center Glostrup, Copenhagen University Hospital, Glostrup, Denmark.  
INSERM, Institut de Myologie, Hôpital de la Pitié-Salpêtrière, Paris, France.  
Illunina, Inc., La Jolla, California, USA.  
School of Electrical Engineering and Computing Science, University of Newcastle, Newcastle, New South Wales, Australia.  
Section of Neonatal Screening and Hormones, Department of Clinical Chemistry and Immunology, The State Serum Institute, Copenhagen, Denmark.  
Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden.  
Centre for Clinical Research in Neuropsychiatry, School of Psychiatry and Clinical Neurosciences, The University of Western Australia, Perth, Western Australia, Australia.  
Department of Child and Adolescent Psychiatry, Pierre and Marie Curie Faculty of Medicine, Paris, France.  
Deceased. Correspondence should be addressed to P.V.G. (pgejman@gmail.com).  
Department of Clinical Pharmacology, Bispebjerg University Hospital, Copenhagen, Denmark.  
Department of Psychology, University of Colorado, Boulder, Colorado, USA.  
Department of Psychiatry and Psychology, School of Mental Health and Neuroscience, European Graduate School of Neuroscience (EURON), South Limburg Mental Health Research and Teaching Network (SEARCH), Maastricht University Medical Centre, Maastricht, The Netherlands.  
Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA.

Page 9
Figure 1.
Manhattan plot for stages 1 and 2. Standard $-\log_{10} P$ plot of the study results. For the stage 1 results, 16 regions with one or more SNP achieving $P < 10^{-6}$ are highlighted in color and labeled with the name of the nearest gene. SNPs selected for stage 2 replication are highlighted, with the resulting combined $P$ value after replication (that is, after incorporation of stage 2 results) indicated by the large diamonds. Blue highlighting indicates SNPs that were less significantly associated after replication, and pink highlighting indicates SNPs that were more significantly associated after replication.
Figure 2.
Regional association plots for five new schizophrenia loci. Regional $P$ value plots for each of the five new schizophrenia loci: 1p21.3, 2q32.3, 8p23.2, 8q21.3 and 10q24.32-q24.33. Each plot shows the most associated SNP (key SNP) and its genomic region from the first column of Table 2: stage 1 scan results for each SNP ± 200 kb to the key SNP are shown. On the x axis is the genomic position, and on the y axis is $-\log_{10} P$. Larger SNP symbols indicate higher LD (based on HapMap 3 data) to the key SNP than smaller SNP symbols. Color coding (from red to blue) denotes LD information; see also the legend within the plot.
### Table 1

Study design and samples

<table>
<thead>
<tr>
<th>Collection</th>
<th>Country</th>
<th>Platform</th>
<th>Cases included by sex</th>
<th>Controls included by sex</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Male</td>
<td>Female</td>
</tr>
<tr>
<td>Cardiff UK</td>
<td>UK</td>
<td>Affymetrix 500K</td>
<td>320</td>
<td>152</td>
</tr>
<tr>
<td>CATIE</td>
<td>United States</td>
<td>Affymetrix 500K; Perlegen 164K</td>
<td>308</td>
<td>94</td>
</tr>
<tr>
<td>ISC-Abedeen</td>
<td>UK</td>
<td>Affymetrix 5.0</td>
<td>536</td>
<td>184</td>
</tr>
<tr>
<td>ISC-Cardiff</td>
<td>Bulgaria</td>
<td>Affymetrix 6.0</td>
<td>270</td>
<td>257</td>
</tr>
<tr>
<td>ISC-Dublin</td>
<td>Ireland</td>
<td>Affymetrix 6.0</td>
<td>188</td>
<td>82</td>
</tr>
<tr>
<td>ISC-Edinburgh</td>
<td>UK</td>
<td>Affymetrix 6.0</td>
<td>267</td>
<td>101</td>
</tr>
<tr>
<td>ISC-London</td>
<td>UK</td>
<td>Affymetrix 5.0; Affymetrix 500K</td>
<td>369</td>
<td>149</td>
</tr>
<tr>
<td>ISC-Portugal</td>
<td>Portugal</td>
<td>Affymetrix 5.0</td>
<td>213</td>
<td>133</td>
</tr>
<tr>
<td>ISC-SW1</td>
<td>Sweden</td>
<td>Affymetrix 5.0</td>
<td>93</td>
<td>75</td>
</tr>
<tr>
<td>ISC-SW2</td>
<td>Sweden</td>
<td>Affymetrix 6.0</td>
<td>231</td>
<td>159</td>
</tr>
<tr>
<td>MGS</td>
<td>United States, Australia</td>
<td>Affymetrix 6.0</td>
<td>1,863</td>
<td>816</td>
</tr>
<tr>
<td>SGENE-Bonn</td>
<td>Germany</td>
<td>Illumina 500K</td>
<td>238</td>
<td>226</td>
</tr>
<tr>
<td>SGENE-Copenhagen</td>
<td>Denmark</td>
<td>Illumina Human 610-Quad</td>
<td>280</td>
<td>202</td>
</tr>
<tr>
<td>SGENE-Munich</td>
<td>Germany</td>
<td>Illumina 300K</td>
<td>279</td>
<td>155</td>
</tr>
<tr>
<td>SGENE-TOP3</td>
<td>Norway</td>
<td>Affymetrix 6.0</td>
<td>132</td>
<td>116</td>
</tr>
<tr>
<td>SGENE-UCLA</td>
<td>The Netherlands</td>
<td>Illumina 500K</td>
<td>529</td>
<td>175</td>
</tr>
<tr>
<td>Zucker Hillside</td>
<td>United States</td>
<td>Affymetrix 500K</td>
<td>128</td>
<td>64</td>
</tr>
<tr>
<td><strong>Grand totals for the GWAs</strong></td>
<td></td>
<td></td>
<td>6,244</td>
<td>3,150</td>
</tr>
<tr>
<td>Multicenter Pedigree</td>
<td>Europe, United States, Australia</td>
<td>Illumina Human 610-Quad</td>
<td>n.a.</td>
<td>n.a.</td>
</tr>
<tr>
<td>SGENE-Aarhus</td>
<td>Denmark</td>
<td>Illumina Human 610-Quad</td>
<td>477</td>
<td>399</td>
</tr>
<tr>
<td>SGENE-Aarhus</td>
<td>Denmark</td>
<td>Centaurus</td>
<td>114</td>
<td>102</td>
</tr>
<tr>
<td>SGENE-Belgium</td>
<td>Belgium</td>
<td>Centaurus; Illumina 370K</td>
<td>326</td>
<td>184</td>
</tr>
<tr>
<td>SGENE-Copenhagen</td>
<td>Denmark</td>
<td>Centaurus</td>
<td>264</td>
<td>198</td>
</tr>
<tr>
<td>SGENE-Iceland</td>
<td>Iceland</td>
<td>Illumina 300K</td>
<td>346</td>
<td>185</td>
</tr>
<tr>
<td>SGENE-England</td>
<td>UK</td>
<td>Illumina 300K</td>
<td>71</td>
<td>22</td>
</tr>
<tr>
<td>SGENE-Helsinki</td>
<td>Finland</td>
<td>Illumina 300K</td>
<td>112</td>
<td>70</td>
</tr>
<tr>
<td>SGENE-Kuusamo</td>
<td>Finland</td>
<td>Illumina 300K</td>
<td>123</td>
<td>50</td>
</tr>
<tr>
<td>Collection</td>
<td>Country</td>
<td>Platform</td>
<td>Cases included by sex</td>
<td>Controls included by sex</td>
</tr>
<tr>
<td>------------------</td>
<td>-----------</td>
<td>--------------</td>
<td>------------------------</td>
<td>--------------------------</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Male</td>
<td>Female</td>
</tr>
<tr>
<td>SGENE-Hungary</td>
<td>Hungary</td>
<td>Centaurus</td>
<td>105</td>
<td>136</td>
</tr>
<tr>
<td>SGENE-Italy</td>
<td>Italy</td>
<td>Illumina 300K</td>
<td>48</td>
<td>36</td>
</tr>
<tr>
<td>SGENE-Munich</td>
<td>Germany</td>
<td>Illumina 300K</td>
<td>280</td>
<td>186</td>
</tr>
<tr>
<td>SGENE-Munich</td>
<td>Germany</td>
<td>Centaurus</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SGENE-Russia</td>
<td>Russia</td>
<td>Centaurus</td>
<td>132</td>
<td>343</td>
</tr>
<tr>
<td>SGENE-Sweden</td>
<td>Sweden</td>
<td>Centaurus</td>
<td>158</td>
<td>94</td>
</tr>
<tr>
<td>SW3</td>
<td>Sweden</td>
<td>Affymetrix 6.0</td>
<td>327</td>
<td>212</td>
</tr>
<tr>
<td>SW4</td>
<td>Sweden</td>
<td>Affymetrix 6.0</td>
<td>656</td>
<td>407</td>
</tr>
<tr>
<td>UQ and ASRB</td>
<td>Australia</td>
<td>Sequenom MassArray</td>
<td>347</td>
<td>190</td>
</tr>
<tr>
<td>ISGC and WTCCC2</td>
<td>Ireland</td>
<td>Affymetrix 6.0</td>
<td>968</td>
<td>342</td>
</tr>
<tr>
<td><strong>Grand totals for the replication follow up</strong></td>
<td></td>
<td></td>
<td><strong>4,731</strong></td>
<td><strong>3,106</strong></td>
</tr>
</tbody>
</table>

Stage 1 describes the 17 samples that provided full GWAS genotyping data, and stage 2 describes the 19 studies that provided results for the top SNPs identified in the combined analysis of stage 1 studies. Stage 2 replication SGENE-Belgium had four cases missing sex information. Stage 2 replication SGENE-Aarhus (focused genotyping sample) had one case missing sex information. Stage 2 replication University of Queensland had 21 cases and 15 controls missing sex information. Sex information for the two stage 2 replication SGENE-Munich samples are combined. Sex information for the two stage 2 replication SGENE-Finnish (Helsinki and Kuusamo) samples are combined to enable that these two samples are located adjacent to each other in the table (rather than alphabetically). Multicenter Pedigree was a family sample, and so case sex counts are not applicable (n.a.). SGENE, Schizophrenia Genetics Consortium; ISC, International Schizophrenia Consortium; TOP3, Thematic Organized Psychoses Research 3; UCLA, University of California at Los Angeles; SW1, Sweden 1; SW2, Sweden 2; WTCCC, Wellcome Trust case Control Consortium; for the Multicenter Pedigree study, the number of cases indicates the number of families; CATIE, Clinical Antipsychotic Trials of Intervention Effectiveness; MGS, Molecular Genetics of Schizophrenia; UQ, University of Queensland; ASRB, Australian Schizophrenia Research Bank; ISGC, Irish Schizophrenia Genomics Consortium.
Table 2

Top genome-wide association results for schizophrenia

<table>
<thead>
<tr>
<th>SNP</th>
<th>Chr.</th>
<th>Mb</th>
<th>Alleles</th>
<th>Frequency</th>
<th>P (GC-adjusted P)</th>
<th>OR (95% CI)</th>
<th>Consistency of direction</th>
<th>Gene</th>
<th>Distance (kb)</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs1625579</td>
<td>1p21.3</td>
<td>98.3</td>
<td>TG</td>
<td>0.80</td>
<td>$5.72 	imes 10^{-7}$ (6.52 x $10^{-6}$)</td>
<td>1.14 (1.08–1.19)</td>
<td>++++++</td>
<td>MIR137</td>
<td>Intragenic</td>
</tr>
<tr>
<td>rs17662626</td>
<td>2q32.3</td>
<td>193.7</td>
<td>AG</td>
<td>0.91</td>
<td>$2.65 	imes 10^{-6}$ (n.a.)</td>
<td>1.11 (1.07–1.16)</td>
<td>+ +++++</td>
<td>PCGEM1</td>
<td>343</td>
</tr>
<tr>
<td>rs202172</td>
<td>6p21.3-p22.1</td>
<td>30.3</td>
<td>CT</td>
<td>0.78</td>
<td>$1.55 	imes 10^{-3}$ (n.a.)</td>
<td>1.10 (1.03–1.17)</td>
<td>+ + + ++</td>
<td>TRM26</td>
<td>Intragenic</td>
</tr>
<tr>
<td>rs10503253</td>
<td>8p23.2</td>
<td>4.2</td>
<td>AC</td>
<td>0.19</td>
<td>$7.60 	imes 10^{-3}$ (n.a.)</td>
<td>1.08 (1.01–1.14)</td>
<td>+ + +++</td>
<td>CSMD1</td>
<td>Intragenic</td>
</tr>
<tr>
<td>rs7004633</td>
<td>8q21.3</td>
<td>89.8</td>
<td>GA</td>
<td>0.18</td>
<td>$0.011$ (n.a.)</td>
<td>1.05 (1.01–1.10)</td>
<td>++++++</td>
<td>MMP16</td>
<td>421</td>
</tr>
<tr>
<td>rs7914558</td>
<td>10q24.32</td>
<td>104.8</td>
<td>GA</td>
<td>0.59</td>
<td>$1.07 	imes 10^{-3}$ (n.a.)</td>
<td>1.08 (1.03–1.13)</td>
<td>+++++</td>
<td>CNNM2</td>
<td>Intragenic</td>
</tr>
<tr>
<td>rs11191580</td>
<td>10q24.33</td>
<td>104.9</td>
<td>TC</td>
<td>0.91</td>
<td>$5.09 	imes 10^{-3}$ (n.a.)</td>
<td>1.09 (1.02–1.16)</td>
<td>++++++</td>
<td>NT5C2</td>
<td>Intragenic</td>
</tr>
<tr>
<td>rs548181</td>
<td>11q24.2</td>
<td>125.0</td>
<td>GA</td>
<td>0.88</td>
<td>$5.29 	imes 10^{-5}$ (n.a.)</td>
<td>1.08 (1.04–1.12)</td>
<td>++++++</td>
<td>STT3A</td>
<td>1</td>
</tr>
<tr>
<td>rs12966547</td>
<td>18q21.2</td>
<td>50.9</td>
<td>GA</td>
<td>0.58</td>
<td>$2.60 	imes 10^{-18}$ (5.99 x $10^{-18}$)</td>
<td>1.09 (1.06–1.12)</td>
<td>+++++++</td>
<td>CDC68</td>
<td>126</td>
</tr>
<tr>
<td>SNP</td>
<td>Chr.</td>
<td>Mb</td>
<td>Alleles</td>
<td>Frequency</td>
<td>$P$ (GC-adjusted $P$)</td>
<td>OR (95% CI)</td>
<td>Consistency of direction</td>
<td>Gene</td>
<td>Distance (kb)</td>
</tr>
<tr>
<td>---------</td>
<td>------</td>
<td>-----</td>
<td>---------</td>
<td>-----------</td>
<td>----------------------</td>
<td>-------------</td>
<td>--------------------------</td>
<td>------</td>
<td>---------------</td>
</tr>
<tr>
<td>rs17512836</td>
<td>18q21.2</td>
<td>51.3</td>
<td>CT</td>
<td>0.02</td>
<td>$2.35 \times 10^{-8}$ (4.78 x $10^{-7}$)</td>
<td>1.40 (1.28–1.52)</td>
<td>++</td>
<td>TCF4</td>
<td>Intragenic</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$1.05 \times 10^{-6}$ (2.86 x $10^{-5}$)</td>
<td>1.23 (1.14–1.31)</td>
<td>+</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The SNPs listed are those with a stage 1 $P < 5 \times 10^{-8}$ and/or a combined stage 1 and 2 $P < 5 \times 10^{-8}$. These ten independent ($r^2 < 0.2$) SNPs represent eight physically distinct genomic loci, as there are two SNPs listed for two loci (10q24.32-q24.33 and 18q21.2). For the MHC region, only one SNP is listed for clarity. The eight susceptibility loci represent three previously reported and five new loci (Supplementary Table 7). Stage 1 is the discovery GWAS mega-analysis. Stage 2 is the replication sample (single-tailed meta-analysis $P$ values are weighted by 1/s.e.), and because the $P$ values are single tailed, some 95% confidence intervals contain 1 (if 0.10 < $P$ < 0.05). Combined values include stages 1 and 2 (two-tailed meta-analysis $P$ values are weighted by 1/s.e.). For each SNP, $P$ values and odds ratios are listed for stage 1 (top), stage 2 (middle) and combined stage 1 and 2 analysis (bottom) with the genomic control (GC)-adjusted values bracketed (n.a., not applicable for stage 2). Alleles are listed with the stage 1 risk allele first; the frequency (in stage 1 controls) and odds ratio (OR) refer to the stage 1 risk allele. Bolded $P$ values indicate $P < 5 \times 10^{-8}$, except for in the stage 2 data, where bolded values indicate $P < 0.05$. The directions of association in eight replication samples are represented by + if the associations are in the same direction, − if they are in opposite directions and a blank space if the data are not available. Mb is the base position based on hg18. Cytogenetic bands are listed for each SNP, though because only one of multiple MHC SNPs are listed, a band range is given in that instance. The nearest gene (or microRNA) is listed, with the distance (kb) from the gene (or if the SNP is intragenic) noted. None of these SNPs showed a significant test for heterogeneity among the samples. Chr., chromosome.

*New finding.*