Common variants at VRK2 and TCF4 conferring risk of schizophrenia


1deCODE Genetics, IS-101 Reykjavik, Iceland, 2Rudolf Magnus Institute of Neuroscience and Department of Medical Genetics, University Medical Centre, 3584 CG Utrecht, The Netherlands, 3Institute of Clinical Medicine, Division of Mental Health and Addiction, University of Oslo and Oslo University Hospital, N-0407 Oslo, Norway, 4Research Institute of Biological Psychiatry, Mental Health Centre ScHans Copenhagen University Hospital, DK-4000 Roskilde, Denmark, 5Centre for Psychiatric Research, Aarhus University Hospital, DK-8240 Risskov, Denmark, 6Department of Human Genetics and 7National Centre for Register-based Research, Aarhus University, DK-8000 Aarhus, Denmark, 8Galician Foundation of Genomic Medicine-SERGAS and 9Service of Psychiatry, Complexo Hospitalario Universitario de Santiago (CHUS), Santiago de Compostela, Spain, 10Institute for Molecular Medicine Finland FIMM and Department of Medical Genetics, University of Helsinki, FI-00014 Helsinki, Finland, 11Wellcome Trust Sanger Institute, Hinxton, Cambridge CB10 1SA, UK, 12DFG Research Centre for Molecular Physiology of the Brain (CMPB), D-37075 Göttingen, Germany, 13Division of Clinical Neuroscience and 14Department of Neurgenetics, Max Planck Institute of Experimental Medicine, D-37075 Göttingen, Germany, 15Department of Genomics, Life & Brain Centre, University of Bonn, D-53127 Bonn, Germany, 16Department of Genetic Epidemiology in Psychiatry.

To whom correspondence should be addressed. Tel: +354 5701900; Fax: +354 5701903; Email: kari.stefansson@decode.is

The full list of contributors in these consortia is presented in the Supplementary Material.
Common sequence variants have recently joined rare structural polymorphisms as genetic factors with strong evidence for association with schizophrenia. Here we extend our previous genome-wide association study and meta-analysis (totalling 7,946 cases and 19,036 controls) by examining an expanded set of variants using an enlarged follow-up sample (up to 10,260 cases and 23,500 controls). In addition to previously reported alleles in the major histocompatibility complex region, near neurogranin (NRGN) and in an intron of transcription factor 4 (TCF4), we find two novel variants showing genome-wide significant association: rs2312147[C], upstream of vaccinia-related kinase 2 (VRK2) [odds ratio (OR) = 1.09, \( P = 1.9 \times 10^{-9} \)] and rs4309482[A], between coiled-coiled domain containing 68 (CCDC68) and TCF4, about 400 kb from the previously described risk allele, but not accounted for by its association (OR = 1.09, \( P = 7.8 \times 10^{-9} \)).
INTRODUCTION

Recent results support a contribution of common variants to schizophrenia heritability. Variants showing genome-wide significant association with schizophrenia have been reported in zinc finger binding protein 804A (ZNF804A) at 2q32.1 (1), in the major histocompatibility complex (MHC) region at 6p21.3–22.1 (2–4), near NRGN at 11q24.2 (4) and in TCF4 at 18q21.2 (4). In addition, score alleles—or alleles with some, although not necessarily strong, evidence for association in a discovery data set—have been shown to be significantly enriched in schizophrenia patients of independent data sets (2). Based on comparison of various score allele sets and results from simulations under different genetic models, the best-fitting heritability model has been determined to include a polygenic component composed of a large number of common alleles, each conferring a small risk but together accounting for at least one-third of narrow-sense heritability (2).

We previously carried out the SGene-plus-genome-wide association (GWA) study of schizophrenia (2,663 cases; 13,498 controls) with subsequent combination of results from the top 1500 single-nucleotide polymorphisms (SNPs) with data from the International Schizophrenia Consortium (ISC) (2,602 cases; 2,885 controls) and the Molecular Genetics of Schizophrenia (MGS) group (2,681 cases; 2,653 controls) (4). SNPs having P-values < 1 × 10^-5 in the combined SGeneplus-ISC-MGS data set were followed-up in an additional 4,999 cases and 15,555 controls.

Here we extended our earlier study by examining loci having P < 1 × 10^-4 in the SGeneplus-ISC-MGS data set. All SNPs were tested in 19 European and European-American study groups (9,246 cases and 22,356 controls; Supplementary Material, Table S1), and those attaining genome-wide significance in a joint analysis (SGene-plus-ISC-MGS and follow-up) were investigated in an additional 10,144 cases and 1,144 controls from Germany.

RESULTS

We chose SNPs for follow-up by filtering SNPs having P-values < 1 × 10^-4 in the SGeneplus-MGS-ISC data set so that no SNP in the final set had r^2 > 0.3 with another—except in the MHC region where the two SNPs having the most significant combined P-values in our previous study (4) were retained. Altogether, 39 SNPs were examined, one per genomic region apart from the MHC region and at 18q21.2 where 6 and 2 SNPs, respectively, were investigated. Allelic association analysis was carried out in 19 European and European-American study groups, and inverse-variance-weighted meta-analysis was used to combine the results.

Of the 33 genomic regions tested, 30 had, for the most significant regional SNP in the SGeneplus-ISC-MGS data set, an effect in the same direction as in the discovery data set (Supplementary Material, Table S2). This pattern of excess is unlikely to occur by chance (P = 7.0 × 10^-7) implying enrichment of the follow-up set of variants for schizophrenia risk alleles.

In the combination of the follow-up and SGeneplus-ISC-MGS results, eight SNPs, including two that were novel, showed genome-wide significance (P < 5 × 10^-8) (Supplementary Material, Table S2). These SNPs continued to show genome-wide significant association after the addition of data from a further 10,144 cases and 1,144 controls (Tables 1 and 2).

The newly identified variants showing genome-wide significant association were located at 2p15.1, near VRK2, and at 18q21.2, between CCDC68 and TCF4 (Table 1 and see Supplementary Material, Figures S1 and S2 for illustrations of the regions). In both cases, association did not deviate from the multiplicative model for risk (P = 0.33 for rs2312147; P = 0.68 for rs4309482), and there was no evidence of odds ratio (OR) heterogeneity between the study groups (P = 0.82, I^2 = 0 for rs2312147; P = 0.74, I^2 = 0 for rs4309482; see Supplementary Material, Tables S3 and S4 for association results by study group). Although rs4309482 is located ~400 kb from rs9960767, a SNP previously shown to be associated with schizophrenia by us, the two SNPs are only very weakly correlated (in the study groups with individual genotypes available, D' < 0.8, r^2 < 0.03), and the association of neither SNP could be explained by the other (in the study groups with individual genotypes available, rs4309482 had a P-value of 6.3 × 10^-7 while, conditional on rs9960767, it had a P-value of 4.4 × 10^-4; in the same groups, rs9960767 had a P-value of 1.1 × 10^-7; while, conditional on rs4309482, it had a P-value of 7.8 × 10^-5).

The previously identified SNPs showing genome-wide significance were located in three genomic locations: two sub-regions of the MHC (one extending from 27.2 to 28.4 Mb and the other ~32.3 Mb), near NRGN, and in an intron of TCF4 (Table 2 and see Supplementary Material, Figures S2–S4 for illustrations of the regions). The strongest association in the telomeric MHC sub-region was at rs13211507 (Table 2), in contrast to our previous investigation where the most significant association in that sub-region was at rs6932590. However, as in our earlier work, the most significant SNP in the telomeric sub-region was not strongly correlated with the significant SNP in the centromeric sub-region (in the study groups for which genotypes were available, r^2 ~ 0.2, D' ~ 0.6), and the association of neither SNP could explain that of the other (in groups with individual genotypes available, rs13211507 had a P-value of 1.3 × 10^-3, while conditional on rs3131296, the P-value was 2.1 × 10^-4; in the same groups, rs3131296 had a P-value of 3.9 × 10^-8, while conditional on rs13211507, the P-value was 6.8 × 10^-4). No deviation from the multiplicative model for risk (P > 0.12), or evidence of OR heterogeneity between the study groups (P > 0.13, I^2 < 23, see Supplementary Material, Tables S5–S8 for results by study group), was observed for any of the previously identified SNPs showing genome-wide significance.

In the combination of the follow-up and the SGeneplus-ISC-MGS data sets, four variants had P-values that did not surpass our genome-wide significance threshold but were below 1 × 10^-6 (Supplementary Material, Table S2). These variants—located at 5q21.1, near solute carrier organic anion transporter family member 6A1 (SLCO6A1); at 8q21.3 in matrix metallopeptidase 16 (MMP16); at 11q23.2 between dopamine receptor D2 (DRD2) and transmembrane protease serine 5 (TMPRSS5); and at 10q24.32 in arsenic (+3 oxidation state) methyltransferase (A53MT)—warrant further study.
Table 1. Novel variants showing genome-wide significant association with schizophrenia

<table>
<thead>
<tr>
<th>SNP</th>
<th>Mb</th>
<th>Al</th>
<th>Freq</th>
<th>SGENE-plus + ISC + MGS</th>
<th>Primary follow-up</th>
<th>Secondary follow-up</th>
<th>SGENE-plus + ISC + MGS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(946</td>
<td>cases; 19,036 controls)</td>
<td></td>
<td>(924 cases; 22,356 controls)</td>
<td>(1,014 cases; 1,144 controls)</td>
<td>(18,206 cases; 42,536 controls)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>OR (95% CI)</td>
<td>P-value</td>
<td>OR (95% CI)</td>
<td>P-value</td>
<td>OR (95% CI)</td>
<td>P-value</td>
<td></td>
</tr>
<tr>
<td>2p15.1, VRK2</td>
<td>rs2312147</td>
<td>58.1</td>
<td>C</td>
<td>0.61</td>
<td>1.11 (1.06, 1.17)</td>
<td>3.7 × 10⁻⁶</td>
<td>1.08 (1.03, 1.12)</td>
</tr>
<tr>
<td>18q21.2, TCF4</td>
<td>rs4309482</td>
<td>50.9</td>
<td>A</td>
<td>0.58</td>
<td>1.10 (1.05, 1.15)</td>
<td>7.1 × 10⁻⁵</td>
<td>1.08 (1.04, 1.13)</td>
</tr>
</tbody>
</table>

Mb, megabase based on NCBI Build 36; al, allele; freq, average frequency in SGENE-plus controls.

Table 2. Previously identified variants showing genome-wide significant association with schizophrenia

<table>
<thead>
<tr>
<th>SNP</th>
<th>Mb</th>
<th>Al</th>
<th>Freq</th>
<th>SGENE-plus + ISC + MGS</th>
<th>Primary follow-up</th>
<th>Secondary follow-up</th>
<th>SGENE-plus + ISC + MGS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(946</td>
<td>cases; 19,036 controls)</td>
<td></td>
<td>(924 cases; 22,356 controls)</td>
<td>(1,014 cases; 1,144 controls)</td>
<td>(18,206 cases; 42,536 controls)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>OR (95% CI)</td>
<td>P-value</td>
<td>OR (95% CI)</td>
<td>P-value</td>
<td>OR (95% CI)</td>
<td>P-value</td>
<td></td>
</tr>
<tr>
<td>6p21.3–6p22.1, MHC</td>
<td>rs6913660</td>
<td>27.2</td>
<td>C</td>
<td>0.85</td>
<td>1.19 (1.12, 1.27)</td>
<td>4.6 × 10⁻⁸</td>
<td>1.09 (1.04, 1.15)</td>
</tr>
<tr>
<td></td>
<td>rs6932590</td>
<td>27.4</td>
<td>T</td>
<td>0.78</td>
<td>1.15 (1.09, 1.21)</td>
<td>5.6 × 10⁻⁷</td>
<td>1.10 (1.05, 1.15)</td>
</tr>
<tr>
<td></td>
<td>rs13211507</td>
<td>28.4</td>
<td>T</td>
<td>0.92</td>
<td>1.22 (1.12, 1.33)</td>
<td>5.2 × 10⁻⁶</td>
<td>1.21 (1.13, 1.30)</td>
</tr>
<tr>
<td></td>
<td>rs3131296</td>
<td>32.3</td>
<td>G</td>
<td>0.87</td>
<td>1.18 (1.10, 1.27)</td>
<td>9.8 × 10⁻⁶</td>
<td>1.14 (1.07, 1.21)</td>
</tr>
<tr>
<td>11q24.2, NRGN</td>
<td>rs12807809</td>
<td>124.1</td>
<td>T</td>
<td>0.83</td>
<td>1.16 (1.09, 1.24)</td>
<td>2.3 × 10⁻⁶</td>
<td>1.09 (1.04, 1.15)</td>
</tr>
<tr>
<td>18q21.2, TCF4</td>
<td>rs9860767</td>
<td>51.3</td>
<td>C</td>
<td>0.056</td>
<td>1.27 (1.15, 1.40)</td>
<td>1.8 × 10⁻⁶</td>
<td>1.16 (1.07, 1.25)</td>
</tr>
</tbody>
</table>

Mb, megabase based on NCBI Build 36; al, allele; freq, average frequency in SGENE-plus controls.

**DISCUSSION**

In this work, we found two new variants, one at 2p15.1 and the other at 18q21.2, showing genome-wide significant association with schizophrenia. Six other variants—in the MHC region, at 11q24.2 and at 18q21.2—continued to show genome-wide significant association.

The novel variant at 2p15.1, rs2312147[C], is located ~50 kb upstream of VRK2, a gene coding for a serine/threonine kinase belonging to the casein kinase I group (5). VRK2 is widely expressed, with elevated expression in highly proliferative cells (6). Alternatively spliced transcripts result in multiple isoforms (7). VRK2 has been proposed to play a role in the maintenance of appropriate nuclear architecture (8), and to be involved in preventing apoptosis (9). The serine/threonine kinase binds JIP1 [Jun NH₂-terminal Kinase (JNK) Interacting Protein 1] (10) which in neuronal cells, serves as an anti-apoptotic factor in response to stress (11) and plays a role in axonal development (12).

A second gene, Fanconi anemia complementation group L (FANCL), is located in the same linkage disequilibrium block as VRK2 (Supplementary Material, Figure S1). FANCL ubiquitylates Fanconi anemia group D2 (FANCD2) and Fanconi anemia complementation group I (FANCI), promoting DNA-interstrand crosslink repair (13).

The newly identified risk factor at 18q21.2, rs4309482[A], lies approximately equidistant from CCDC68 and TCF4, and could act through either gene or both. The association with schizophrenia of rs9960767[C]—located in intron 3 of TCF4—makes the simplest model one in which TCF4 is affected by both variants.

TCF4 encodes a basic helix-loop-helix transcription factor with many alternatively spliced transcripts resulting in multiple isoforms (14). The gene is expressed in the embryonic central nervous system (CNS) and many adult tissues (15), including the mammalian brain (16). Multiple lines of evidence support a role for TCF4 in the CNS, especially in its development. TCF4 deletions and some coding mutations in the gene cause Pitt–Hopkins syndrome, an encephalopathy characterized by severe intellectual disability, a distinctive facial appearance and additional features including breathing problems, epilepsy and postnatal microcephaly (15,17,18). TCF4 interacts with transcription factors playing key roles in neurodevelopment including mouse atonal (ato) homolog 1 (Math1) (19), humanachaete–scute homolog 1 (HASH1) (20) and neurogenic differentiation factor 2 (neuroD2) (16,21). In Tcf4⁻/⁻ mice, pontine nuclei development is disrupted (19), whereas mice overexpressing Tcf4 postnatally in the forebrain display cognitive impairments and deficits in pre-pulse inhibition (16).

Differential expression of TCF4 related to psychosis has been reported. Decreased expression in blood has been associated with the psychotic state (22), whereas up-regulation has been found in both the cerebellar cortex of schizophrenia.
patients (23) and the brains of mice given phencyclidine (PCP)—a treatment that mimics the behavioural syndrome of schizophrenia (22).

A common SNP in intron 3 of TCF4 is strongly associated with Fuchs’s corneal dystrophy (FCD) (24). This SNP, rs613872, is weakly correlated with rs4309482 and rs9960767 (HapMap CEU $r^2 < 0.1$). We find no reports examining co-occurrence of schizophrenia and FCD.

Two findings of this study, the novel genome-wide significant associations and the excess of ORs in the follow-up set in the same direction as in the discovery data set, support a role for common sequence variants in schizophrenia heritability. The latter finding, like the previous score allele analysis (2), buttresses the notion that larger sample sizes will allow the identification of additional common risk alleles. Previous studies have demonstrated that rare structural variants contribute to schizophrenia susceptibility (25). Next-generation sequencing technology should facilitate the recovery of rare sequence variants conferring risk of schizophrenia. Eventually, a collection of variants—rare and common, structural and single-nucleotide—may account for a substantial proportion of schizophrenia heritability, as has been shown for other common diseases such as type 2 diabetes (26).

The primary goal of uncovering schizophrenia risk alleles is to find novel genes and, through them, pathways involved in the disease. The results revealed here implicate genes involved in signal transduction (VRK2) and gene expression (TCF4). A great deal of additional work is required to understand how the SNPs affect these genes (or possibly other, neighbouring genes) and how these changes, in turn, lead to schizophrenia. Nevertheless, the results described here provide a valuable starting point for investigation of the biological mechanisms leading to schizophrenia.

MATERIALS AND METHODS

Samples

The genome-wide typed (‘SGENE-plus’; 2,663 cases and 13,498 controls) and meta-analysis samples (5,283 cases and 5,538 controls) used here were identical to those of our original GWA paper. Those samples are described in that work (4) and in the companion papers (2,3). The primary follow-up samples employed here consist of the follow-up sample from our original GWA study (excluding 460 cases and 677 controls for which DNA was no longer available) and an additional 4,707 cases and 7,478 controls from Belgium, Denmark (Aarhus), Denmark (Copenhagen), Ireland (the Wellcome Trust Case Control Consortium 2), Italy, the Netherlands, Norway, Russia, the UK (Cardiff University) and the USA [the European-American portion of the Clinical Antipsychotic Trials of Intervention Effectiveness (CATIE) study] (Supplementary Material, Table S1). The secondary follow-up samples consisted of 1,014 cases and 1,144 controls from the Göttingen Research Association for Schizophrenia (GRAS) (27) study. All cases were diagnosed with schizophrenia, schizoaffective disorder (~7%) or persistent delusional disorder (< 1%) (see the Supplementary Material for further information on the various study groups).

Genotyping and association analysis

Genotyping was carried out using Illumina and Affymetrix genome-wide arrays, Centaurus assays (Nanogen), the Sequenom MassArray iPLEX genotyping system and the Roche LightCycler480 system (Supplementary Material, Table S1). The Supplementary Material further describes genotyping as well as sample and marker quality control in each group. Association analysis was carried out using a likelihood model as previously described (28). Imputation was performed for the Ireland (WTCCC2) and UK groups (both typed on Affymetrix chips) as described in the Supplementary Material. For the Finnish and CATIE samples, association results were adjusted for the first 10 or 20, respectively, principal components using logistic regression. The remaining genome-wide-typed groups—all of which had genomic control inflation factors $<1.1$—were corrected for potential population stratification using genomic control (29).

Summary statistics were used to combine association results from the various study groups as described earlier (4).

SUPPLEMENTARY MATERIAL

Supplementary Material is available at HMG online.

ACKNOWLEDGEMENTS

We would like to thank the subjects, their families and the recruitment centre staff. We would also like to acknowledge the help of Maria Dolores Molto (Genetics Department, Valencia University, CIBERSAM); Eduardo Paz and Ramón Ramos-Rios (Complejo Hospitalario de Santiago); and the contribution of the Spanish National Genotyping Centre (CeGen-USC node). The study makes use of data generated by the Clinical Antipsychotic Trials of Intervention Effectiveness (CATIE) project whose principal investigators were Jeffrey A. Lieberman, M.D., T. Scott Stroup, M.D., M.P.H. and Joseph P. McEvoy, M.D.

Conflict of Interest statement. None declared.

FUNDING

This work was supported by the European Union [grant numbers LSHM-CT-2006-037761 (Project SGENE), PIAP-GA-2008-218251 (Project PsychGene), HEALTH-F2-2009-223423 (Project PsychCNVs)]; the National Genome Research Network of the German Federal Ministry of Education and Research (BMBF) [grant numbers 01GS08144 (MooDS-Net), 01GS08147 (NGFNplus)]; the National Institute of Mental Health [R01 MH078075, and N01 MH900001, MH074027 to the Clinical Antipsychotic Trials of Intervention Effectiveness (CATIE) project]; the Centre of Excellence for Complex Disease Genetics of the Academy of Finland (grant numbers 213506, 129680); the Biocentrum Helsinki Foundation and Research Program for Molecular Medicine, Faculty of Medicine, University of Helsinki; the Stanley Medical Research Institute; the Danish Council for Strategic Research (grant number 2101-07-0059); H. Lundbeck A/S; the Research Council of Norway (grant number 163070/V50); the South-East
REFERENCES


