Common variants at *VRK2* and *TCF4* conferring risk of schizophrenia

Stacy Steinberg[1], **Simone de Jong**[2], **Irish Schizophrenia Genomics Consortium**[†],
Ole A. Andreassen[3], **Thomas Werge**[4], **Anders D. Børglum**[5,6], **Ole Mors**[5], **Preben B. Mortensen**[7],
Omar Gustafsson[1,3], **Javier Costas**[8], **Olli P. H. Pietiläinen**[10,11], **Ditte Demontis**[6], **Sergi Papiol**[12,13],
Johanna Huttenlocher[1,51], **Manuel Mattheisen**[15], **René Breuer**[16], **Evangelos Vassos**[17,18],
Ina Giegling[19], **Gillian Fraser**[20], **Nicholas Walker**[21], **Annamari Tuulio-Henriksson**[22],
Jaana Suvisaari[22], **Jouko Lönnqvist**[22], **Tiina Paunio**[23], **Ingrid Agartz**[3], **Ingrid Melle**[3],
Srdjan Djurovic[3], **Eric Strengman**[2], **GROUP**[†], **Gesche Jürgens**[24], **Birte Glenthøj**[25],
Lars Terenius[26], **David M. Hougaard**[27], **Torben Ørntoft**[28], **Carsten Wiuf**[29], **Michael Didriksen**[30],
Mads V. Hollegaard[27], **Merete Nordentoft**[31], **Ruud van Winkel**[32,33], **Gunter Kenis**[33],
Lilia Abramova[34], **Vasily Kaleda**[34], **Manuel Arrojo**[9], **Julio Sanjuán**[35], **Celso Arango**[36],
Swetlana Sperling[13], **Moritz Rossner**[12,14], **Michele Ribolsi**[37], **Valentina Magni**[37],
Alberto Siracusano[37], **Claus Christiansen**[38], **Lambertus A. Kiemeney**[39], **Jan Veldink**[40],
Leonard van den Berg[40], **Andres Ingason**[1], **Pierandrea Muglia**[41], **Robin Murray**[17],
Markus M. Nöthen[42], **Engilbert Sigurdsson**[43,44], **Hannes Petursson**[43,44],
Unnur Thorsteinsdottir[1,44], **Augustine Kong**[1], **I. Alex Rubino**[37], **Marc De Hert**[32],
János M. Réthelyi[45], **István Bitter**[45], **Erik G. Jönsson**[26], **Vera Golimbet**[34], **Angel Carracedo**[46],
Hannelore Ehrenreich[12,13], **Nick Craddock**[47], **Michael J. Owen**[47], **Michael C. O'Donovan**[47],
Wellcome Trust Case Control Consortium 2[†], **Mirella Ruggeri**[48], **Sarah Tosato**[48],
Leena Peltonen[10,11,49], **Roel A. Ophoff**[2,52], **David A. Collier**[18], **David St Clair**[20],
Marcella Rietschel[16], **Sven Cichon**[15,42,50], **Hreinn Stefansson**[1], **Dan Rujescu**[19]
and Kari Stefansson[1,44,*]

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

Central Institute of Mental Health, University of Heidelberg, D-68159 Mannheim, Germany, [17]Division of Psychological Medicine and [18]Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, King's College, London SE5 8AF, UK, [19]Division of Molecular and Clinical Neurobiology, Department of Psychiatry, Ludwig-Maximilians University, D-80336 Munich, Germany, [20]Department of Mental Health, University of Aberdeen, Royal Cornhill Hospital, Aberdeen AB25 2ZD, UK, [21]Ravenscraig Hospital, Greenock PA16 9HA, UK, [22]Department of Mental Health and Substance Abuse Services, National Institute for Health and Welfare, FI-00300 Helsinki, Finland, [23]Public Health Genomics Unit, National Institute for Health and Welfare THL, FI-00270 Helsinki, Finland, [24]Department of Clinical Pharmacology, Bispebjerg University Hospital, DK-2400 Copenhagen NV, Denmark, [25]Centre for Clinical Intervention and Neuropsychiatric Schizophrenia Research, Copenhagen University Hospital, Psychiatric Centre, DK-2600 Glostrup, Denmark, [26]Department of Clinical Neuroscience, HUBIN project, Karolinska Institutet and Hospital, SE-171 76 Stockholm, Sweden, [27]Statens Serum Institute, DK-2300 Copenhagen S, Denmark, [28]AROS Applied Biotechnology A/S and Department of Molecular Medicine, Aarhus University Hospital, Skejby, DK-8200 Aarhus N, Denmark, [29]Bioinformatics Research Centre, Aarhus University, DK-8000 Aarhus C, Denmark, [30]Discovery Biology Research, H. Lundbeck A/S, DK-2500 Valby, Denmark, [31]Psychiatric Centre Copenhagen, Copenhagen University Hospital, DK-2400 Copenhagen, Denmark, [32]University Psychiatric Centre, Catholic University Leuven, B-3070 Kortenberg, Belgium, [33]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, [34]Mental Health Research Centre, Russian Academy of Medical Sciences, 117152 Moscow, Russia, [35]Unit of Psychiatry, Faculty of Medicine, University of Valencia, Network Centre of Biomedical Research on Mental Health (CIBERSAM), Valencia, Spain, [36]Hospital General Universitario Gregorio Marañón, Biomedical Network Research Centre on Mental Health (CIBERSAM), Madrid, Spain, [37]Department of Neuroscience, Section of Psychiatry, University of Rome-Tor Vergata, 00137 Rome, Italy, [38]Nordic Bioscience, DK-2730 Herlev, Denmark, [39]Department of Epidemiology and Biostatistics and Department of Urology, Radboud University Nijmegen Medical Centre, 6500 HB Nijmegen, the Netherlands, [40]Rudolf Magnus Institute of Neuroscience and Department of Neurology, University Medical Centre, 3584 CX Utrecht, the Netherlands, [41]Medical Genetics, GlaxoSmithKline R&D, 37135 Verona, Italy, [42]Institute of Human Genetics, University of Bonn, D-53111 Bonn, Germany, [43]Department of Psychiatry, National University Hospital, IS-101 Reykjavik, Iceland, [44]School of Medicine, University of Iceland, IS-101 Reykjavik, Iceland, [45]Department of Psychiatry and Psychotherapy, Semmelweis University, 1083 Budapest, Hungary, [46]Genomic Medicine Group - Galician Foundation of Genomic Medicine-Biomedical Network Research Centre on Rare Diseases (CIBERER), University of Santiago de Compostela, Spain, [47]MRC Centre for Neuropsychiatric Genetics and Genomics, Department of Psychological Medicine and Neurology, School of Medicine, Cardiff University, Cardiff CF14 4XN, UK, [48]Section of Psychiatry and Clinical Psychology, University of Verona, 37134 Verona, Italy, [49]The Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA, [50]Institute of Neurosciences and Medicine (INM-1), D-52425 Juelich, Germany, [51]Department of Medical Genetics, Institute of Human Genetics, University of Tübingen, D-72076 Tübingen, Germany and [52]Centre for Neurobehavioural Genetics, UCLA, Los Angeles, California 90095, USA

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 × 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 × 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 polygenetic 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 $< 1 \times 10^{-5}$ in the combined SGENE-plus-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 $P < 1 \times 10^{-4}$ in the SGENE-plus-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 1 014 cases and 1 144 controls from Germany.

RESULTS

We chose SNPs for follow-up by filtering SNPs having P-values $< 1 \times 10^{-4}$ in the SGENE-plus-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 SGENE-plus-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 ($P = 7.0 \times 10^{-7}$) implying enrichment of the follow-up set of variants for schizophrenia risk alleles.

In the combination of the follow-up and SGENE-plus-ISC-MGS results, eight SNPs, including two that were novel, showed genome-wide significance ($P < 5 \times 10^{-8}$)

(Supplementary Material, Table S2). These SNPs continued to show genome-wide significant association after the addition of data from a further 1 014 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^{-6}; in the same groups, rs9960767 had a P-value of 1.1×10^{-5}, 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 subregions 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, $r^2 \sim 0.2$, $D' \sim 0.6$), 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^{-8}, 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 SGENE-plus-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* (*AS3MT*)—warrant further study.

Table 1. Novel variants showing genome-wide significant association with schizophrenia

SNP	Mb	Al	Freq	SGENE-plus + ISC + MGS (7 946 cases; 19 036 controls)		Primary follow-up (9 246 cases; 22 356 controls)		Secondary follow-up (1 014 cases; 1 144 controls)		SGENE-plus + ISC + MGS and all follow-up (18 206 cases; 42 536 controls)	
				OR (95% CI)	*P*-value	OR (95% CI)	*P*-value	OR (95% CI)	*P*-value	OR (95% CI)	*P*-value
2p15.1, *VRK2*											
rs2312147	58.1	C	0.61	1.11 (1.06, 1.17)	3.7×10^{-6}	1.08 (1.03, 1.12)	0.00024	1.10 (0.97, 1.24)	0.13	1.09 (1.06, 1.12)	1.9×10^{-9}
18q21.2, *TCF4*											
rs4309482	50.9	A	0.58	1.10 (1.05, 1.15)	7.1×10^{-5}	1.08 (1.04, 1.13)	3.6×10^{-5}	1.06 (0.94, 1.20)	0.34	1.09 (1.06, 1.12)	7.8×10^{-9}

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

SNP	Mb	Al	Freq	SGENE-plus + ISC + MGS (7 946 cases; 19 036 controls)		Primary follow-up (9 246 cases; 22 356 controls)		Secondary follow-up (1 014 cases; 1 144 controls)		SGENE-plus + ISC + MGS and all follow-up (18 206 cases; 42 536 controls)	
				OR (95% CI)	*P* value	OR (95% CI)	*P* value	OR (95% CI)	*P* value	OR (95% CI)	*P* value
6p21.3–6p22.1, MHC											
rs6913660	27.2	C	0.85	1.19 (1.12, 1.27)	4.6×10^{-8}	1.09 (1.04, 1.15)	0.00067	1.34 (1.13, 1.57)	0.00050	1.14 (1.10, 1.19)	1.7×10^{-11}
rs6932590	27.4	T	0.78	1.15 (1.09, 1.21)	5.6×10^{-7}	1.10 (1.05, 1.15)	6.9×10^{-5}	1.11 (0.96, 1.28)	0.15	1.12 (1.08, 1.16)	1.6×10^{-10}
rs13211507	28.4	T	0.92	1.22 (1.12, 1.33)	5.2×10^{-6}	1.21 (1.13, 1.30)	1.1×10^{-7}	1.34 (1.08, 1.67)	0.0089	1.22 (1.16, 1.29)	1.4×10^{-13}
rs3131296	32.3	G	0.87	1.18 (1.10, 1.27)	9.8×10^{-6}	1.14 (1.07, 1.21)	1.4×10^{-5}	1.28 (1.07, 1.54)	0.0080	1.16 (1.11, 1.21)	5.1×10^{-11}
11q24.2, *NRGN*											
rs12807809	124.1	T	0.83	1.16 (1.09, 1.24)	2.3×10^{-6}	1.09 (1.04, 1.15)	0.00057	1.16 (0.99, 1.37)	0.063	1.12 (1.08, 1.17)	2.8×10^{-9}
18q21.2, *TCF4*											
rs9960767	51.3	C	0.056	1.27 (1.15, 1.40)	1.8×10^{-6}	1.16 (1.07, 1.25)	0.00025	1.09 (0.85, 1.41)	0.50	1.20 (1.13, 1.27)	4.2×10^{-9}

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 NH$_2$-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), human achaete–scute homologue 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 discovery 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 material 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 factions <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.

REFERENCES

1. Williams, H.J., Norton, N., Dwyer, S., Moskvina, V., Nikolov, I., Carroll, L., Georgieva, L., Williams, N.M., Morris, D.W., Quinn, E.M. *et al.* (2011) Fine mapping of ZNF804A and genome-wide significant evidence for its involvement in schizophrenia and bipolar disorder. *Mol. Psychiatry*, **16**, 429–441.

2. Purcell, S.M., Wray, N.R., Stone, J.L., Visscher, P.M., O'Donovan, M.C., Sullivan, P.F. and Sklar, P. (2009) Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature*, **460**, 748–752.

3. Shi, J., Levinson, D.F., Duan, J., Sanders, A.R., Zheng, Y., Pe'er, I., Dudbridge, F., Holmans, P.A., Whittemore, A.S., Mowry, B.J. *et al.* (2009) Common variants on chromosome 6p22.1 are associated with schizophrenia. *Nature*, **460**, 753–757.

4. Stefansson, H., Ophoff, R.A., Steinberg, S., Andreassen, O.A., Cichon, S., Rujescu, D., Werge, T., Pietilainen, O.P., Mors, O., Mortensen, P.B. *et al.* (2009) Common variants conferring risk of schizophrenia. *Nature*, **460**, 744–747.

5. Manning, G., Whyte, D.B., Martinez, R., Hunter, T. and Sudarsanam, S. (2002) The protein kinase complement of the human genome. *Science*, **298**, 1912–1934.

6. Nezu, J., Oku, A., Jones, M.H. and Shimane, M. (1997) Identification of two novel human putative serine/threonine kinases, VRK1 and VRK2, with structural similarity to vaccinia virus B1R kinase. *Genomics*, **45**, 327–331.

7. Blanco, S., Klimcakova, L., Vega, F.M. and Lazo, P.A. (2006) The subcellular localization of vaccinia-related kinase-2 (VRK2) isoforms determines their different effect on p53 stability in tumour cell lines. *FEBS J.*, **273**, 2487–2504.

8. Nichols, R.J., Wiebe, M.S. and Traktman, P. (2006) The vaccinia-related kinases phosphorylate the N' terminus of BAF, regulating its interaction with DNA and its retention in the nucleus. *Mol. Biol. Cell*, **17**, 2451–2464.

9. Li, L.Y., Liu, M.Y., Shih, H.M., Tsai, C.H. and Chen, J.Y. (2006) Human cellular protein VRK2 interacts specifically with Epstein-Barr virus BHRF1, a homologue of Bcl-2, and enhances cell survival. *J. Gen. Virol.*, **87**, 2869–2878.

10. Blanco, S., Sanz-Garcia, M., Santos, C.R. and Lazo, P.A. (2008) Modulation of interleukin-1 transcriptional response by the interaction between VRK2 and the JIP1 scaffold protein. *PLoS ONE*, **3**, e1660.

11. Dong, Z., Zhou, L., Del Villar, K., Ghanevati, M., Tashjian, V. and Miller, C.A. (2005) JIP1 regulates neuronal apoptosis in response to stress. *Brain Res. Mol. Brain Res.*, **134**, 282–293.

12. Dajas-Bailador, F., Jones, E.V. and Whitmarsh, A.J. (2008) The JIP1 scaffold protein regulates axonal development in cortical neurons. *Curr. Biol.*, **18**, 221–226.

13. Garner, E. and Smogorzewska, A. (2011) Ubiquitylation and the Fanconi anemia pathway. *FEBS Lett.* [Epub ahead of print].

14. Liu, Y., Ray, S.K., Yang, X.Q., Luntz-Leybman, V. and Chiu, I.M. (1998) A splice variant of E2–2 basic helix-loop-helix protein represses the brain-specific fibroblast growth factor 1 promoter through the binding to an imperfect E-box. *J. Biol. Chem.*, **273**, 19269–19276.

15. de Pontual, L., Mathieu, Y., Golzio, C., Rio, M., Malan, V., Boddaert, N., Soufflet, C., Picard, C., Durandy, A., Dobbie, A. *et al.* (2009) Mutational, functional, and expression studies of the TCF4 gene in Pitt-Hopkins syndrome. *Hum. Mutat.*, **30**, 669–676.

16. Brzozka, M.M., Radyushkin, K., Wichert, S.P., Ehrenreich, H. and Rossner, M.J. (2010) Cognitive and sensorimotor gating impairments in transgenic mice overexpressing the schizophrenia susceptibility gene Tcf4 in the brain. *Biol. Psychiatry*, **68**, 33–40.

17. Marangi, G., Ricciardi, S., Orteschi, D., Lattante, S., Murdolo, M., Dallapiccola, B., Biscione, C., Lecce, R., Chiurazzi, P., Romano, C. *et al.* (2011) The Pitt-Hopkins syndrome: Report of 16 new patients and clinical diagnostic criteria. *Am. J. Med. Genet. A*, **155**, 1536–1545.

18. Pitt, D. and Hopkins, I. (1978) A syndrome of mental retardation, wide mouth and intermittent overbreathing. *Aust. Paediatr. J.*, **14**, 182–184.

19. Flora, A., Garcia, J.J., Thaller, C. and Zoghbi, H.Y. (2007) The E-protein Tcf4 interacts with Math1 to regulate differentiation of a specific subset of neuronal progenitors. *Proc. Natl Acad. Sci. USA*, **104**, 15382–15387.

20. Persson, P., Jogi, A., Grynfeld, A., Pahlman, S. and Axelson, H. (2000) HASH-1 and E2-2 are expressed in human neuroblastoma cells and form a functional complex. *Biochem. Biophys. Res. Commun.*, **274**, 22–31.

21. Ravanpay, A.C. and Olson, J.M. (2008) E protein dosage influences brain development more than family member identity. *J. Neurosci. Res.*, **86**, 1472–1481.

22. Kurian, S.M., Le-Niculescu, H., Patel, S.D., Bertram, D., Davis, J., Dike, C., Yehyawi, N., Lysaker, P., Dustin, J., Caligiuri, M. *et al.* (2009) Identification of blood biomarkers for psychosis using convergent functional genomics. *Mol. Psychiatry*, **16**, 37–58.

23. Mudge, J., Miller, N.A., Khrebtukova, I., Lindquist, I.E., May, G.D., Huntley, J.J., Luo, S., Zhang, L., van Velkinburgh, J.C., Farmer, A.D. *et al.* (2008) Genomic convergence analysis of schizophrenia: mRNA sequencing reveals altered synaptic vesicular transport in post-mortem cerebellum. *PLoS ONE*, **3**, e3625.

24. Baratz, K.H., Tosakulwong, N., Ryu, E., Brown, W.L., Branham, K., Chen, W., Tran, K.D., Schmid-Kubista, K.E., Heckenlively, J.R., Swaroop, A. *et al.* (2010) E2−2 protein and Fuchs's corneal dystrophy. *N. Engl. J. Med.*, **363**, 1016–1024.

25. Sebat, J., Levy, D.L. and McCarthy, S.E. (2009) Rare structural variants in schizophrenia: one disorder, multiple mutations; one mutation, multiple disorders. *Trends Genet.*, **25**, 528–535.

26. Lander, E.S. (2011) Initial impact of the sequencing of the human genome. *Nature*, **470**, 187–197.

27. Ribbe, K., Friedrichs, H., Begemann, M., Grube, S., Papiol, S., Kastner, A., Gerchen, M.F., Ackermann, V., Tarami, A., Treitz, A. *et al.* (2010) The cross-sectional GRAS sample: a comprehensive phenotypical data collection of schizophrenic patients. *BMC Psychiatry*, **10**, 91.

28. Gretarsdottir, S., Thorleifsson, G., Reynisdottir, S.T., Manolescu, A., Jonsdottir, S., Jonsdottir, T., Gudmundsdottir, T., Bjarnadottir, S.M., Einarsson, O.B., Gudjonsdottir, H.M. *et al.* (2003) The gene encoding phosphodiesterase 4D confers risk of ischemic stroke. *Nat. Genet.*, **35**, 131–138.

29. Devlin, B. and Roeder, K. (1999) Genomic control for association studies. *Biometrics*, **55**, 997–1004.