nature genetics

Genome-wide association analysis identifies 13 new risk loci for schizophrenia

Stephan Ripke[1,2,44], Colm O'Dushlaine[2,44], Kimberly Chambert[2], Jennifer L Moran[2], Anna K Kähler[3–5], Susanne Akterin[4], Sarah E Bergen[4], Ann L Collins[3], James J Crowley[3], Menachem Fromer[1,2,6], Yunjung Kim[3], Sang Hong Lee[7], Patrik K E Magnusson[4], Nick Sanchez[2], Eli A Stahl[6], Stephanie Williams[3], Naomi R Wray[7], Kai Xia[8], Francesco Bettella[9], Anders D Borglum[10–12], Brendan K Bulik-Sullivan[1], Paul Cormican[13], Nick Craddock[14], Christiaan de Leeuw[15,16], Naser Durmishi[17], Michael Gill[13,18], Vera Golimbet[19], Marian L Hamshere[14], Peter Holmans[14], David M Hougaard[20], Kenneth S Kendler[21,22], Kuang Lin[23], Derek W Morris[13,18], Ole Mors[12,24], Preben B Mortensen[12,25], Benjamin M Neale[1,2], Francis A O'Neill[26], Michael J Owen[14], Milica Pejovic Milovancevic[27], Danielle Posthuma[15,28], John Powell[23,29], Alexander L Richards[14], Brien P Riley[21,22,30], Douglas Ruderfer[6], Dan Rujescu[31,32], Engilbert Sigurdsson[33], Teimuraz Silagadze[34], August B Smit[35,36], Hreinn Stefansson[9], Stacy Steinberg[9], Jaana Suvisaari[37], Sarah Tosato[38], Matthijs Verhage[15], James T Walters[14], Multicenter Genetic Studies of Schizophrenia Consortium[39], Psychosis Endophenotypes International Consortium[39], Wellcome Trust Case Control Consortium 2[39], Elvira Bramon[23,40,41], Aiden P Corvin[13,18], Michael C O'Donovan[14], Kari Stefansson[9], Edward Scolnick[2], Shaun Purcell[1,2,6,42], Steven A McCarroll[2,43], Pamela Sklar[6], Christina M Hultman[4,45] & Patrick F Sullivan[3,4,8,45]

Schizophrenia is an idiopathic mental disorder with a heritable component and a substantial public health impact. We conducted a multi-stage genome-wide association study (GWAS) for schizophrenia beginning with a Swedish national sample (5,001 cases and 6,243 controls) followed by meta-analysis with previous schizophrenia GWAS (8,832 cases and 12,067 controls) and finally by replication of SNPs in 168 genomic regions in independent samples (7,413 cases, 19,762 controls and 581 parent-offspring trios). We identified 22 loci associated at genome-wide significance; 13 of these are new, and 1 was previously implicated in bipolar disorder. Examination of candidate genes at these loci suggests the involvement of neuronal calcium signaling. We estimate that 8,300 independent, mostly common SNPs (95% credible interval of 6,300–10,200 SNPs) contribute to risk for schizophrenia and that these collectively account for at least 32% of the variance in liability. Common genetic variation has an important role in the etiology of schizophrenia, and larger studies will allow more detailed understanding of this disorder.

Schizophrenia is an idiopathic mental disorder with substantial morbidity, mortality and personal and societal costs[1–3]. The presence of an important genetic component is indicated by a sibling recurrence risk ratio of 8.6, by high heritability estimates (0.64 in a national family study, 0.81 in a meta-analysis of twin studies and 0.23 estimated directly from common SNPs) and by previous genomic findings[4–8].

There are only a handful of robust reported genetic associations for schizophrenia. Genome-wide linkage studies so far have been inconclusive[9], and no compelling mendelian variants have been identified[8]. Eight rare copy number variants of strong effect (genotypic relative risks of 4–20) with consistent replication have been described (for example, at 16p11.2 and 22q11.21); however, these associations are generally not disease specific and can also be associated with autism, mental retardation or epilepsy[8]. Initial exome sequencing studies have not yet identified specific variants of unequivocal genome-wide significance[9–13], although larger studies are in progress. Previous GWAS have reported convincing statistical evidence for ~10 genomic regions[8], including the major histocompatibility complex (MHC)[14–16] along with *MIR137* and targets of miR-137 (ref. 17). These previous GWAS suggested that additional common variant associations were likely to be discovered with larger sample sizes[13,17,18]. We therefore sought to achieve a substantially larger sample size in a multi-stage GWAS.

RESULTS

Results for Swedish samples

We analyzed genome-wide SNP data in 5,001 schizophrenia cases and 6,243 controls from a population-based sampling frame in Sweden (N = 11,244; **Table 1**). Most subjects (57.4%) were not previously reported. After genotyping and imputation with the 1000 Genomes

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

Received 9 December 2012; accepted 1 August 2013; published online 25 August 2013; doi:10.1038/ng.2742

Project Phase 1 reference panel, we obtained allelic dosages for 9,871,789 high-quality polymorphic SNPs. Given that this imputation panel is based on >800 chromosomes from individuals of European ancestry and includes the detail afforded by genome sequencing, we anticipated increased power in finding and describing association signals. Indeed, we observed 10,201 SNPs and 187 genomic regions with association $P < 1 \times 10^{-5}$ using 1000 Genomes Project imputation compared with 1,594 SNPs and 133 genomic regions for HapMap 3 imputation (counts include only one region from the MHC).

The resulting genomic inflation factor (λ_{GC}) was 1.075, and λ_{1000} (refs. 19–21) was 1.013. Quantile-quantile and Manhattan plots are shown in **Supplementary Figures 1** and **2**. For association with schizophrenia, 312 SNPs met a genome-wide significance threshold of $P < 5 \times 10^{-8}$ (ref. 22). These SNPs were in two genomic regions (**Supplementary Fig. 3**): 241 SNPs were in the MHC region (chromosome 6: 28,502,794–32,536,501; minimum $P = 4.07 \times 10^{-11}$ at rs115939516), and 71 SNPs were from chromosome 2 (200,715,388–201,040,981; minimum $P = 3.33 \times 10^{-10}$ at rs35220450). We replicated the MHC association reported in previous studies[14–17]. The association with schizophrenia on chromosome 2 (rs35220450) is new, showed highly consistent effects in the Sw1–Sw6 genotyping batches of the Swedish cohort and encompasses *C2orf69*, *C2orf47* and *TYW5* (also known as *C2orf60*).

Results for Swedish and PGC samples

We reanalyzed schizophrenia data from the Psychiatric Genomics Consortium (PGC) using 1000 Genomes Project imputation (8,832 cases and 12,067 controls, excluding Swedish samples)[17]. SNPs within five regions reached genome-wide significance: the MHC locus (chromosome 6: 27,261,324–32,610,445; minimum $P = 2.18 \times 10^{-10}$), *AS3MT-CNNM2-NT5C2* (chromosome 10: 104,635,103–104,960,464; minimum $P = 4.29 \times 10^{-10}$), *MAD1L1* (chromosome 7: 2,005,747–2,098,238; minimum $P = 2.40 \times 10^{-8}$), *RP11-586K2.1* (chromosome 8: 89,585,639–89,760,620; minimum $P = 2.37 \times 10^{-8}$) and SNPs near *TCF4* (chromosome 18: 53,311,001–53,423,307; minimum $P = 3.00 \times 10^{-8}$).

We then conducted a meta-analysis of the Swedish and independent PGC schizophrenia samples using the same quality control, imputation and analysis pipeline. This GWAS meta-analysis of 13,833 schizophrenia cases and 18,310 controls (**Table 1**) afforded power to detect genotypic relative risks of 1.10–1.14 for reference allele frequencies of 0.15–0.85 (power = 0.8; $\alpha = 5 \times 10^{-8}$, log-additive model). We evaluated the comparability of the Swedish and PGC studies using sign tests: of 608 SNPs selected from the PGC results with association $P < 0.0001$ and in approximate linkage equilibrium, 62.6% had logistic regression β coefficients with the same sign in the Swedish results, an observation highly inconsistent with the null hypothesis of no association ($P = 2.2 \times 10^{-10}$). λ_{GC} was 1.186, and λ_{1000} was 1.012, values consistent with a polygenic pattern of association but not with gross inflation due to technical artifacts[20]. Manhattan and quantile-quantile plots are shown in **Figure 1** and **Supplementary Figure 4**, respectively, and genome-wide significance was exceeded by 3,538 SNPs in 12 genomic regions.

We used risk score profiling[14,17] to evaluate the capacity of 130,000 SNPs derived from PGC to predict case-control status in the Swedish samples. These SNPs were selected for high confidence and approximate linkage equilibrium, without regard to association *P* value. PGC risk scores had a highly significant capacity to predict case-control status in the independent Swedish samples (*P* values from 1×10^{-26} to 1×10^{-114}) (**Fig. 2**). The increased sample size allowed improved risk profile prediction. The threshold at which the explanatory power of these risk-profile SNPs plateaus has decreased with increasing sample size: P_T was 0.1 in this study (**Fig. 2**) and 0.2 in the PGC report, and there was no plateau in the International Schizophrenia Consortium study[14,17]. Although the mean risk profiles showed highly significant differences between cases and controls, the distributions overlapped substantially (**Supplementary Fig. 5**) and are insufficient for diagnostic purposes (area under the receiver operating characteristic curve of 0.65). However, these results strongly support the comparability of the Swedish and PGC samples and the validity of the meta-analysis results.

Table 1 Subject characteristics and sample sizes

	Cases	Controls
Swedish sample characteristics		
Male sex	0.595	0.512
Median age at sampling	54 (45–62)	57 (48–65)
Median hospital admissions for SCZ or SAD	7 (3–15)	NA
Median total inpatient days	243 (81–696)	NA
Median years from first to last HDR admission	9.7 (2.9–19.5)	NA
Sample sizes		
Swedish subjects (Sw1–Sw6)	5,001	6,243
PGC schizophrenia subjects (excluding Sw1 and Sw2)	8,832	12,067
Replication results for up to 168 genomic regions	7,413	19,762
Total subjects	21,246	38,072

Values in parentheses are interquartile ranges. The case group had significantly more males ($P < 0.0001$) and was significantly younger ($P < 0.0001$) than the control group, although these differences were not of large magnitude. The higher median age in controls is in the direction of greater confidence in control classification (controls had greater time at risk for psychiatric hospitalization). Cases tended to have had considerable hospitalizations, inpatient lengths of stay and years of observation. All cases and controls are independent. The Swedish sample had a total of 11,244 subjects, PGC had a total of 20,899 subjects, and the replication had a total of 27,175 subjects. Swedish and PGC meta-analysis results are based on data from 32,143 subjects. The Swedish sample plus the PGC sample plus the replication sample included a total of 59,318 subjects (these counts exclude 511 trios). SCZ, schizophrenia; SAD, schizoaffective disorder; HDR, hospital discharge register; NA, not applicable.

GWAS often omit the X chromosome. This omission is problematic, as the X chromosome is approximately as large as chromosome 8 and is enriched for genes important in brain development. Using a previously described approach, we imputed X-chromosome SNPs using the 1000 Genomes Project reference panel[23]. Joint analysis of all subjects as well as of males and females separately did not identify any association meeting the genome-wide significance threshold. The strongest association (rs12845396; chromosome X: 6,029,533; $P = 3.46 \times 10^{-7}$) was in an intron of *NLGN4X* (encoding neuroligin 4), a gene previously implicated in mental retardation and autism, and there were multiple possible signals near *MECP2* (causal in Rett syndrome; $P = 9.3 \times 10^{-6}$).

Most GWAS-associated variants were found outside of protein-coding regions[24]. A recent report suggested that most SNPs in the National Human Genome Research Institute (NHGRI) GWAS catalog[24] coincided with or were in perfect linkage disequilibrium (LD) with DNase I–hypersensitive sites[25]. We thus evaluated whether the association signals in the Swedish and PGC meta-analysis results showed significant overlap with DNase I–hypersensitive sites identified by the Encyclopedia of DNA Elements (ENCODE) Project[26]. We did not find evidence of enrichment (**Supplementary Fig. 6** and **Supplementary Table 1**). However, this negative result is strongly qualified by the lack of DNase I hypersensitivity data directly relevant to psychiatric disorders.

Results from Swedish, PGC and replication samples

We then obtained association results for SNPs in 194 genomic regions in 6 independent samples for a total sample size of over 21,000 cases and 38,000 controls (**Table 1**). The genomic regions for which

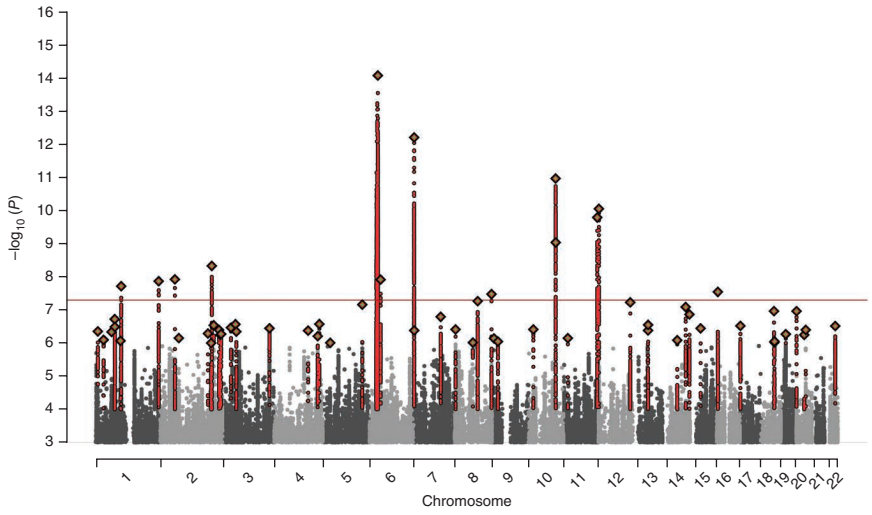

Figure 1 Manhattan plot of the Swedish and PGC schizophrenia meta-analysis results. The *x* axis shows chromosomal position, and the *y* axis shows $-\log_{10}(P)$. The red line is the genome-wide significance level (5×10^{-8}).

replication genotypes were sought were identified using LD clumping, defined by LD ($r^2 > 0.5$) and a minimum association *P* of $<1 \times 10^{-5}$ in the Swedish and PGC meta-analysis. Only one MHC SNP was included. The Swedish and PGC meta-analysis and replication results were highly concordant, with 76.3% of the logistic regression β coefficients having the same direction of effect (sign test $P = 1.5 \times 10^{-17}$). Indeed, of the top 100 SNPs in the Sweden and PGC meta-analysis, 90% had the same sign in the replication results. This result strongly suggests that many more loci will achieve genome-wide significance with further increases in sample size.

The combined results in which SNPs at 24 loci reached genome-wide significance is shown in **Table 2**. As two pairs of these regions overlapped (chromosome 1: 243 Mb and chromosome 5: 152 Mb), there were associations with schizophrenia in 22 genomic regions. Three additional regions nearly met genome-wide significance (rs4380187 near *ZNF804A*, $P = 5.66 \times 10^{-8}$; rs4523957 in *SRR*, $P = 5.69 \times 10^{-8}$; and rs6550435 near *TRANK1*, $P = 5.86 \times 10^{-8}$, which also had $P = 9 \times 10^{-6}$ in a bipolar disorder GWAS)[27].

Of these 22 associated loci (**Table 3**), 5 have been reported previously as meeting genome-wide significance for schizophrenia alone (MHC, *WBP1L* (*C10orf26*), *DPYD-MIR137*, *SDCCAG8* and *MMP16*), and 3 have been reported to be associated with a combined phenotype including schizophrenia and bipolar disorder (*CACNA1C*, *CACNB2* and *ITIH3-ITIH4*)[14–17,27–30]. We now identify 13 newly associated loci, as well as a genome-wide significant association at a locus previously implicated in bipolar disorder (*NCAN*)[31].

Themes

We highlight four themes from these results (see also **Supplementary Table 2**). First, these results implicate calcium signaling in the etiology of schizophrenia. As in previous studies of bipolar disorder and schizophrenia[17,27,28], we replicated genome-wide significant association for a SNP in *CACNA1C* (encoding $Ca_v1.2$; $P = 5.2 \times 10^{-12}$ at the intronic SNP rs1006737). Intriguingly, we identified a genome-wide significant association with schizophrenia in *CACNB2* ($P = 1.3 \times 10^{-10}$ at the intronic SNP rs17691888), which encodes the β_2 subunit of L-type calcium channels ($Ca_v\beta 2$). This locus was previously found to be significant when considering five psychiatric disorders as affected[30]. A gene-set test showed enrichment of smaller *P* values in genes encoding calcium channel subunits (**Supplementary Table 3**).

In L-type calcium channels, the α_{1c} subunit forms the transmembrane pore and directly interacts with the intracellular β_2 subunit[32]. The β_2 subunit also antagonizes an endoplasmic reticulum retention motif on the α_{1c} subunit to facilitate transport to the plasma membrane[33]. Additional genes with genome-wide significant evidence of association were implicated on the basis of membership in a proteomic network centered on Ca_v2 (ref. 34), including the protein products of *ACTR1A* (α-centractin), the divalent metal cation transporter *CNNM2* ($P = 3.7 \times 10^{-13}$; chromosome 10: 103,009,986–105,512,924) and *CACNB2*. Variants within a locus containing *TNNC1* (encoding calcium-binding protein troponin C) also reached genome-wide significance ($P = 1.1 \times 10^{-8}$), as did variants at loci containing three calcium homeostasis modulator genes (*CALHM1*, *CALHM2* and *CALHM3* in the same region of chromosome 10 as *CNNM2*).

The genetics and biology of calcium channels have been the subject of considerable investigation, owing to their importance in fundamental neuronal processes and human diseases. L-type voltage-gated calcium channels are involved in learning, memory and synaptic plasticity, and *Cacna1c* knockout mice show notable deficits in long-term potentiation[35–38]. Calcium 'channelopathies' involve mutations in *CACNA1C* and *CACNB2* that cause Brugada syndrome types 3 and 4 (OMI 611875 and 611876, respectively)[39]. In addition, Timothy syndrome (MIM 601005), caused by mutations in *CACNA1C*, is a multisystem disorder including cognitive impairment and autism spectrum disorder[40]. Although mendelian disorders are usually characterized by persistent pathological features,

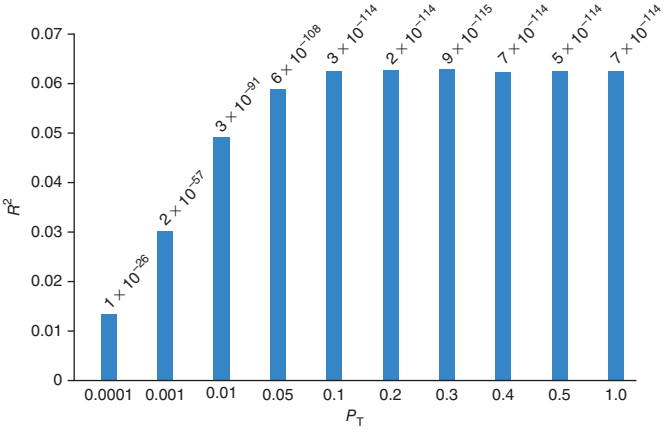

Figure 2 Risk score profiling results using the PGC schizophrenia results as the discovery set and the Swedish data as the testing set. The *x* axis shows ten *P*-value thresholds ($P_T = 1 \times 10^{-4}$, 1×10^{-3}, ..., 1). The *y* axis shows the Nagelkerke pseudo R^2, the proportion of variance in case-control status explained by the risk score profile. The number above each bar is the *P* value for the capacity of the risk score profile to predict case-control status for that P_T.

mendelian calcium channelopathies can have episodic phenomena, perhaps reminiscent of the episodic nature of psychotic disorders—examples of such episodic phenotypes and the underlying genes include intermittent hypoglycemia and hypocalcemia in Timothy syndrome (*CACNA1C*), episodic ataxia (*CACNA1A* and *CACNB4*), migraine (*CACNA1A*), epilepsy (*CACNA1H* and *CACNB4*), periodic paralysis (*CACNA1S*) and malignant hyperthermia (*CACNA1S* and *CACNA2D1*)[32,40].

Our GWAS for schizophrenia suggests candidate genes involved in calcium channels. A calcium channel functional complex has also been suggested as a mechanism in the etiology of bipolar disorder and autism. These results suggest hypotheses for clinical translation. Multiple approved medications act at calcium channels, including some antipsychotics (for example, pimozide) along with adjuvants for treatment non-response for schizophrenia and bipolar disorder (for example, the calcium channel blockers verapamil and nifedipine). It is

Table 2 Association results for Sweden and PGC meta-analysis, replication samples and combined analysis

			Index SNP[a]				*P* value[b]			OR (s.e.)[c]		
Chromosomal region	Length (kb)	SNP	rsID	A_{12}[d]	Position (bp)	Freq.	Sweden + PGC	Replication	Combined	Sweden + PGC	Replication	Combined
Chr. 6: 31,596,138–32,813,768	1,217.6	1,412	rs114002140	AG	32,431,962	0.763	**8.28×10^{-15}**	6.93×10^{-2}	**9.14×10^{-14}**	1.213 (0.025)	1.070 (0.037)	1.167 (0.021)
Chr. 10: 104,487,871–105,245,420	757.5	362	rs7085104	AG	104,628,873	0.645	**1.07×10^{-11}**	2.10×10^{-3}	**3.68×10^{-13}**	1.129 (0.018)	1.076 (0.024)	1.110 (0.014)
Chr. 7: 1,827,717–2,346,115	518.4	566	rs6461049	TC	2,017,445	0.571	**6.17×10^{-13}**	1.85×10^{-2}	**5.93×10^{-13}**	1.132 (0.017)	1.059 (0.024)	1.107 (0.014)
Chr. 1: 98,141,112–98,664,991	523.9	307	rs1198588	AT	98,552,832	0.214	**1.92×10^{-8}**	1.91×10^{-5}	**1.72×10^{-12}**	0.889 (0.021)	0.888 (0.028)	0.889 (0.017)
Chr. 12: 2,285,731–2,440,464	154.7	129	rs1006737	AG	2,345,295	0.332	**8.79×10^{-11}**	3.76×10^{-3}	**5.22×10^{-12}**	1.122 (0.018)	1.070 (0.023)	1.103 (0.014)
Chr. 10: 18,601,928–18,934,390	332.5	147	rs17691888	AG	18,734,528	0.114	3.86×10^{-7}	6.09×10^{-5}	**1.27×10^{-10}**	0.870 (0.028)	0.842 (0.043)	0.862 (0.023)
Chr. 8: 143,297,312–143,410,423	113.1	117	rs4129585	AC	143,312,933	0.439	**3.32×10^{-8}**	1.20×10^{-3}	**2.19×10^{-10}**	1.098 (0.017)	1.077 (0.023)	1.091 (0.014)
Chr. 1: 73,275,828–74,099,273	823.4	1,026	rs10789369	AG	73,824,909	0.383	4.68×10^{-7}	1.99×10^{-4}	**3.64×10^{-10}**	1.091 (0.017)	1.106 (0.027)	1.095 (0.015)
Chr. 11: 130,706,918–130,894,976	188.1	269	rs7940866	AT	130,817,579	0.513	**1.61×10^{-10}**	1.30×10^{-1}	**1.83×10^{-9}**	0.896 (0.017)	0.966 (0.023)	0.921 (0.014)
Chr. 5: 151,888,959–152,835,304	946.3	79	rs17504622	TC	152,654,479	0.050	6.88×10^{-8}	1.02×10^{-2}	**2.65×10^{-9}**	1.250 (0.041)	1.202 (0.072)	1.238 (0.036)
Chr. 19: 19,354,937–19,744,079	389.1	294	rs2905424	TC	19,473,445	0.348	5.38×10^{-7}	1.64×10^{-3}	**3.44×10^{-9}**	1.092 (0.018)	1.093 (0.028)	1.092 (0.015)
Chr. 2: 37,422,072–37,592,628	170.6	10	rs2373000	TC	37,592,628	0.402	9.17×10^{-6}	1.38×10^{-4}	**6.78×10^{-9}**	1.079 (0.017)	1.108 (0.027)	1.087 (0.014)
Chr. 5: 101,581,848–101,870,822	289	367	rs6878284	TC	101,769,726	0.637	1.47×10^{-6}	1.61×10^{-3}	**9.03×10^{-9}**	0.917 (0.018)	0.925 (0.025)	0.920 (0.015)
Chr. 3: 52,215,002–53,175,017	960	533	rs4687552	TC	52,838,402	0.641	9.31×10^{-7}	3.23×10^{-3}	**1.16×10^{-8}**	1.092 (0.018)	1.074 (0.024)	1.086 (0.014)
Chr. 2: 145,139,727–145,214,607	74.9	4	rs12991836	AC	145,141,541	0.652	2.25×10^{-6}	1.30×10^{-3}	**1.19×10^{-8}**	0.918 (0.018)	0.928 (0.023)	0.922 (0.014)
Chr. 2: 200,628,118–201,293,421	665.3	249	rs2949006	TG	200,715,388	0.192	**4.67×10^{-9}**	9.18×10^{-2}	**1.21×10^{-8}**	1.132 (0.021)	1.049 (0.029)	1.102 (0.017)
Chr. 18: 52,722,378–52,827,668	105.3	39	rs4801131	TC	52,752,700	0.418	6.46×10^{-6}	5.27×10^{-4}	**1.22×10^{-8}**	0.926 (0.017)	0.924 (0.023)	0.925 (0.014)
Chr. 2: 233,550,961–233,808,241	257.3	197	rs778371	AG	233,743,109	0.719	5.66×10^{-7}	5.93×10^{-3}	**1.51×10^{-8}**	0.911 (0.019)	0.935 (0.025)	0.920 (0.015)
Chr. 1: 243,593,066–244,025,999	432.9	133	rs14403	TC	243,663,893	0.227	**1.35×10^{-8}**	8.34×10^{-2}	**1.80×10^{-8}**	0.889 (0.021)	0.952 (0.029)	0.910 (0.017)
Chr. 12: 123,447,928–123,913,433	465.5	353	rs11532322	AG	123,731,423	0.318	1.37×10^{-6}	4.77×10^{-3}	**2.28×10^{-8}**	1.099 (0.020)	1.084 (0.029)	1.094 (0.016)
Chr. 1: 243,418,063–243,627,135	209.1	115	rs1538774	CG	243,544,827	0.260	6.11×10^{-7}	8.38×10^{-3}	**2.53×10^{-8}**	0.907 (0.020)	0.934 (0.026)	0.917 (0.016)
Chr. 8: 89,188,454–89,761,163	572.7	402	rs11995572	TG	89,592,083	0.135	5.39×10^{-8}	5.02×10^{-2}	**3.33×10^{-8}**	1.150 (0.026)	1.069 (0.034)	1.120 (0.021)
Chr. 5: 60,484,179–60,843,706	359.5	100	rs171748	AG	60,499,131	0.471	1.62×10^{-6}	5.36×10^{-3}	**3.78×10^{-8}**	1.084 (0.017)	1.068 (0.024)	1.078 (0.014)
Chr. 5: 152,505,453–152,707,306	201.9	8	rs2910032	TC	152,540,354	0.531	8.90×10^{-6}	1.22×10^{-3}	**4.12×10^{-8}**	0.928 (0.017)	0.916 (0.027)	0.925 (0.014)
Chr. 2: 185,533,580–186,057,716	524.1	50	rs4380187	AC	185,811,940	0.529	5.14×10^{-7}	1.98×10^{-2}	5.66×10^{-8}	1.089 (0.017)	1.056 (0.024)	1.078 (0.014)
Chr. 17: 2,015,612–2,256,111	240.5	252	rs4523957	TG	2,208,899	0.616	3.01×10^{-7}	2.66×10^{-2}	5.69×10^{-8}	1.096 (0.018)	1.057 (0.025)	1.083 (0.015)
Chr. 3: 36,834,099–36,964,583	130.5	66	rs6550435	TG	36,864,489	0.656	1.65×10^{-6}	8.24×10^{-3}	5.86×10^{-8}	0.917 (0.018)	0.939 (0.024)	0.925 (0.014)

We used LD clumping to aggregate association findings into genomic regions. All positions are relative to UCSC hg19. Chr., chromosome; freq., frequency. Boldface indicates $P < 5 \times 10^{-8}$.
[a]Details for the index SNP, the SNP with the strongest association in the genomic region. [b]*P* values in the meta-analysis of Sw1–Sw6 with the PGC schizophrenia results, the replication samples alone and the final combined analysis of Sw1–Sw6, PGC and replication samples. [c]Odds ratio (OR) estimates and standard errors. [d]A_{12}, reference and alternate alleles.

possible that drugs that act on the protein products of *CACNA1C* and *CACNB2* for a different therapeutic indication could be repurposed for the treatment of schizophrenia. For example, there has been at least one clinical trial of the efficacy of isradipine in bipolar disorder (an approved antihypertensive acting at the protein product of *CACNA1C*; R. Perlis, personal communication). In addition, given

Table 3 Description of the 22 genome-wide significant loci in the combined analysis

Chromosomal region	*P* value	Previous association[a]	Candidate gene in relation to index SNP[b]	Other genes in genomic region defined by LD[c]	eQTL[d]	Disease associations[e]
Chr. 6: 31,596,138–32,813,768	9.14×10^{-14}	SCZ	*HLA-DRB9*	MHC class II, many other genes, lincRNA	Many	Many
Chr. 10: 104,487,871–105,245,420	3.68×10^{-13}	SCZ	*C10orf32-AS3MT*	*CALHM1*, *CALHM2*, *CALHM3*, *CNNM2*, *CYP17A1*, *INA*, *MIR1307*, *NT5C2*, *PCGF6*, *PDCD11*, *SFXN2*, *ST13P13*, *TAF5*, *USMG5*, *WBP1L*	*ACTR1A*, *ARL3*, *AS3MT*, *C10orf32*, *C10orf78*, *NT5C2*, *TMEM180*, *TRIM8*, *WBP1L*	GWAS: blood pressure, CAD, aneurysm
Chr. 7: 1,827,717–2,346,115	5.93×10^{-13}	No	*MAD1L1*	*FTSJ2*, *NUDT1*, *SNX8*	*C7orf27*, *FTSJ2*, *MAD1L1*, *NUDT1*	
Chr. 1: 98,141,112–98,664,991	1.72×10^{-12}	SCZ	(*MIR137*, 37 kb)	*DPYD*, lincRNA	***DPYD***	*DPYD*: mental retardation
Chr. 12: 2,285,731–2,440,464	5.22×10^{-12}	SCZ, BPD	*CACNA1C*	–	No data	*CACNA1C*: autism, Timothy syndrome, Brugada syndrome 3
Chr. 10: 18,601,928–18,934,390	1.27×10^{-10}	5 disorders	*CACNB2*	*NSUN6*	No data	*CACNB2*: Brugada syndrome 4; GWAS: blood pressure
Chr. 8: 143,297,312–143,410,423	2.19×10^{-10}	No	*TSNARE1*	–	No data	
Chr. 1: 73,275,828–74,099,273	3.64×10^{-10}	No	(x10NST00000415686.1, 4 kb)	lincRNA	No data	
Chr. 11: 130,706,918–130,894,976	1.83×10^{-9}	No	(*SNX19*, 31 kb)	lincRNA	***SNX19***	
Chr. 5: 151,888,959–152,835,304	2.65×10^{-9}	No	ENST00000503048.1	lincRNA (*GRIA1*)	No data	
Chr. 5: 152,505,453–152,707,306	4.12×10^{-8}	No				
Chr. 19: 19,354,937–19,744,079	3.44×10^{-9}	BPD	(*MAU2*, 4 kb)	*CILP2*, *GATAD2A*, *GMIP*, *HAPLN4*, *LPAR2*, *MIR640*, *NCAN*, *NDUFA13*, *PBX4*, *SUGP1*, *TM6SF2*, *TSSK6*, *YJEFN3*	No data	GWAS: lipid levels
Chr. 2: 37,422,072–37,592,628	6.78×10^{-9}	No	*QPCT*	*C2orf56*, *CEBPZ*, *PRKD3*, *SULT6B1* lincRNA	No eQTL	
Chr. 5: 101,581,848–101,870,822	9.03×10^{-9}	No	*SLCO6A1*	lincRNA	No data	
Chr. 3: 52,215,002–53,175,017	1.16×10^{-8}	SCZ, BPD	*ITIH3*	*ALAS1*, *ALDOAP1*, *BAP1*, *C3orf78*, *DNAH1*, *GLT8D1*, *GLYCTK*, *GNL3*, *ITIH1*, *ITIH4*, *MIR135A1*, *MIRLET7G*, *MUSTN1*, *NEK4*, *NISCH*, *NT5DC2*, *PBRM1*, *PHF7*, *PPM1M*, *RFT1*, *SEMA3G*, *SFMBT1*, *SPCS1*, *STAB1*, *TLR9*, *TMEM110*, *TNNC1*, *TWF2*, *WDR82*, lincRNA	No data (*ITIH1-ITIH3-ITIH4*)	*GLYCTK*: D-glyceric aciduria, mental retardation; *RTF1*: mental retardation; GWAS: adiponectin, height, waist-hip ratio
Chr. 2: 145,139,727–145,214,607	1.19×10^{-8}	No	*ZEB2*	–	No eQTL	*ZEB2*: Mowat-Wilson syndrome, mental retardation
Chr. 2: 200,628,118–201,293,421	1.21×10^{-8}	No	*FONG*	*C2orf47*, *C2orf69*, *SPATS2L*, *TYW5*, lincRNA	No data	GWAS: osteoporosis
Chr. 18: 52,722,378–52,827,668	1.22×10^{-8}	No	(ENST00000565991.1, 21 kb)	lincRNA (*TCF4*)	No data	
Chr. 2: 233,550,961–233,808,241	1.51×10^{-8}	No	*C2orf82*	*GIGYF2*, *KCNJ13*, *NGEF*	No data	
Chr. 1: 243,593,066–244,025,999	1.80×10^{-8}	No	*AKT3*	*CEP170*	***AKT3***	
Chr. 1: 243,418,063–243,627,135	2.53×10^{-8}	Yes	*SDCCAG8*		***SDCCAG8***	
Chr. 12: 123,447,928–123,913,433	2.28×10^{-8}	No	*C12orf65*	*ABCB9*, *ARL6IP4*, *CDK2AP1*, *MIR4304*, *MPHOSPH9*, *OGFOD2*, *PITPNM2*, *RILPL2*, *SBNO1*, *SETD8*, lincRNA	*ARL6IP4*, *CDK2AP1*, *OGFOD2*, *SBNO1*	*C12orf65*: mental retardation; GWAS: HDL, height, head size
Chr. 8: 89,188,454–89,761,163	3.33×10^{-8}	SCZ	Intergenic	*MMP16*, lincRNA	*MMP16*	
Chr. 5: 60,484,179–60,843,706	3.78×10^{-8}	No	ENST00000506902.1	*ZSWIM6*, *C5orf43*, lincRNA	*C5orf43*, *ZSWIM6*	

[a]Regions reported to meet genome-wide significance thresholds of association for schizophrenia (SCZ) or bipolar disorder (BPD). [b]The gene within which an index SNP is located is given. For intergenic index SNPs, the nearest gene is given in parentheses. [c]Other named genes in the genomic interval. [d]SNP-transcript associations with $q < 0.05$ in peripheral blood. eQTLs with the SNP with the strongest association are shown in bold. [e]Data from the NHGRI GWAS catalog[24], OMIM[43] and a compilation of genes related to autism[73] and mental retardation[43,74,75]. No data means no Affymetrix U219 probe sets or low expression in peripheral blood. The *CACNB2* association emerged when considering attention deficit/hyperactivity disorder (ADHD), autism, bipolar disorder, major depressive disorder and schizophrenia as affected[30]. CAD, coronary artery disease; HDL, high-density lipoprotein.

that many approved antipsychotics increase the cardiac QT interval, genetic variation in calcium channel genes might be used to identify individuals at higher risk of sudden cardiac death[41,42].

Second, as reported previously[14–17], the strongest SNP association (rs114002140; $P = 9.1 \times 10^{-14}$) with schizophrenia is in the extended MHC (chromosome 6: 25–34 Mb), a region of both exceptional importance and complexity. The MHC comprises 0.3% of the genome but contains 1.5% of the genes in Online Mendelian Inheritance in Man (OMIM)[43] and 6.4% of genome-wide significant SNP associations in the NHGRI GWAS catalog[24]. It is the second most gene-dense genomic region and has high LD over its extent. We speculate that these features (high gene density and strong LD) combined with the polygenicity of schizophrenia lead to the strong association observed for this region but will also complicate efforts to identify causal variation. Genome-wide significant associations with schizophrenia extend over 7 Mb, but larger samples may resolve this association into subregions near *TRIM26* (encoding tripartite motif–containing 26; chromosome 6: 30.1 Mb) and the *HLA-DRB9* unprocessed pseudogene (chromosome 6: 32.4 Mb, intergenic to *HLA-DRA*–*HLA-DRB5*) (**Supplementary Fig. 7**).

Third, multiple genomic lines of evidence support a role for *MIR137* in the etiology of schizophrenia. We provide additional support for a common variant association located upstream of the *MIR137* transcript ($P = 1.7 \times 10^{-12}$; **Supplementary Fig. 8**). Fourteen genes in the regions listed in **Table 3** have miR-137 target sites predicted by TargetScan (v6.2)[44] (*C6orf47*, *HLA-DQA1*, *TNXB*, *VARS*, *WBP1L*, *CACNA1C*, *DPYD*, *CACNB2*, *TSSK6*, *NT5DC2*, *PITPNM2*, *SBNO1*, *ZEB2* and *PRKD3*). Using gene-set analysis, we evaluated whether genes with predicted miR-137 target sites were enriched for smaller association *P* values. We confirmed the PGC result[17] and extended the finding by showing more robust enrichment in a far larger set of genes with predicted miR-137 target sites (**Supplementary Table 3**). In addition, our unpublished work shows enrichment for smaller GWAS *P* values in genes downregulated following overexpression of miR-137 in human neural stem cells (A.L.C., Y.K., R. Bloom, D. Rubinow, W. Sun *et al.*, unpublished data). Given the role of miR-137 in fundamental neuronal processes[45–47], these results support the investigation of pathways influenced by miR-137 with regard to their role in the pathogenesis of schizophrenia.

The SNP with the strongest association with schizophrenia (rs1198588) is 39 kb upstream of *MIR137* and might regulate the transcription of *MIR137*. However, this has not been proven experimentally, and there is another candidate gene in the region. rs1198588 is in an LD block that includes *DPYD* (169 kb upstream of rs1198588), and rs1198588 is a significant local expression quantitative trait locus (eQTL) for *DPYD*. We note that *DPYD* also contains a predicted miR-137 target site. An exome sequencing study reported two putative functional *de novo* variants in *DPYD* in cases with schizophrenia[11].

Fourth, 13 of the 22 regions listed in **Table 3** contain long intergenic noncoding RNAs (lincRNAs). LincRNAs have multiple known or suspected functions, including roles in epigenetic regulation and development[48]. Using pathway analysis[49], we found modest enrichment ($P = 0.06$) for smaller association *P* values in a conservative set of lincRNAs derived from sequencing of polyadenylated RNA from multiple tissues[48]. This observation is consistent with a general role for GWAS-identified markers in the regulation of gene expression rather than the alteration of protein sequence. eQTLs[50,51] overlap with SNPs implicated by GWAS over all traits[52–54] as well as for specific traits such as height, adiposity, cardiovascular risk factors, chemotherapy-induced cytotoxicity, autism, schizophrenia and Crohn's disease[55–62]. An estimated 55% of eQTL SNPs lie in DNase I hypersensitivity sites (markers of open chromatin subject to transcriptional regulation), and 77% of SNPs implicated in GWAS coincide with DNase I hypersensitivity sites or are in high LD with SNPs in these sites[25,63,64].

Genetic architecture

There has been considerable debate about the genetic architecture of schizophrenia. We estimated the proportion of variance in liability to schizophrenia explained by SNPs using GCTA (Genome-wide Complex Trait Analysis)[65]. Traditional genetic epidemiological studies use the phenotypic resemblance of relatives to estimate the proportion of variance in liability, using theoretical resemblance assumptions. GCTA uses genome-wide SNP genotypes to calculate heritability in the population from identity-by-state relationships for each pair of individuals. Using the PGC schizophrenia data, we previously estimated the SNP heritability of schizophrenia at 0.23 (s.e. of 0.01), using HapMap 3 imputation and assuming a population risk of 0.01 (ref. 7). Using the same imputation reference and population risk, SNP heritability was substantially higher in the Swedish samples (0.32, s.e. of 0.03), possibly owing to the greater phenotypic and genetic homogeneity in the Swedish sample compared with the PGC

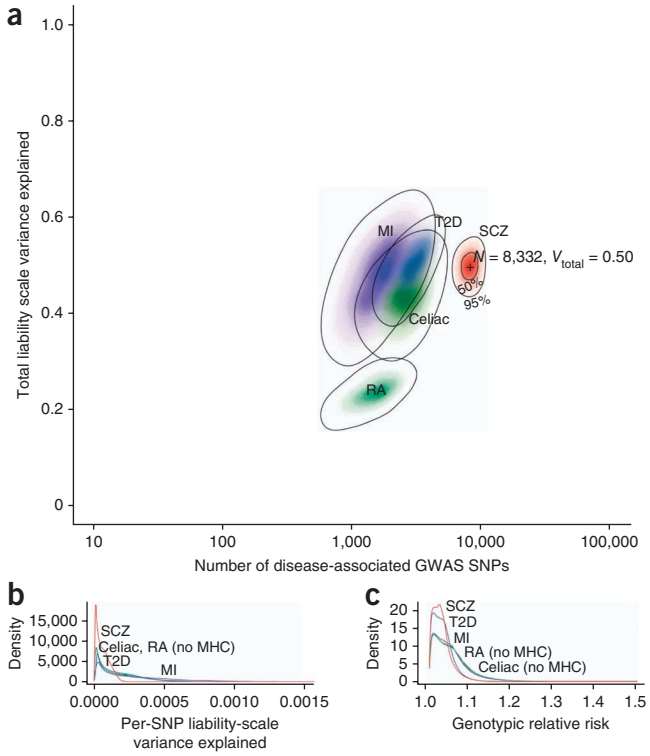

Figure 3 Results of ABPA modeling based on the Swedish and PGC results (population risk of 0.01). (**a**) The *x* axis shows the estimated number of SNPs on a $\log_{10}$ scale, and the *y* axis estimates the total variance in liability explained (V_{total}). The results for five conditions are shown, including schizophrenia from this analysis (SCZ; red) and, for comparison, results from a published analysis of myocardial infarction (MI; purple), type 2 diabetes mellitus (T2D; blue), celiac disease (green) and rheumatoid arthritis (RA; teal)[72]. The schizophrenia results are based on 1000 Genomes Project imputation, and the others are based on HapMap 3 imputation. Color intensity reflects the probability density, with darker colors indicating higher density. Contour lines show 50% and 95% credible regions for schizophrenia, and 95% credible regions for the other diseases. (**b**,**c**) Estimated SNP distributions for the five disorders, including the distribution of SNPs in terms of the variance in liability explained per SNP (**b**) and the estimated distribution of SNP genotypic relative risks (**c**). We again stress that multiple qualifiers are essential in interpreting these estimates.

samples of mixed European ancestry. We obtained a similar estimate of SNP heritability using 1000 Genomes Project data (0.33, s.e. of 0.03; population risk of 0.01). For a population risk of 0.004 (refs. 4,66), SNP heritability was 0.26 (s.e. of 0.02) using HapMap 3 data and 0.27 (s.e. of 0.02) using 1000 Genomes Project data. Partitioning of the SNP heritability by minor allele frequency (MAF) is consistent with 80% of the signal reflecting causal variants with MAF of >0.1 (**Supplementary Table 4**).

To complement the GCTA analyses, we also applied ABPA (approximate Bayesian polygenic analysis)[67] to the Sweden and PGC meta-analysis results. Compared to GCTA, ABPA yielded somewhat larger but generally congruent estimates of variance in liability to schizophrenia using HapMap 3 imputation data, yielding 0.43 for a population risk of 0.01 (95% credible interval of 0.38–0.48) and 0.34 for a population risk of 0.004 (95% credible interval of 0.31–0.37).

The Bayesian framework used by ABPA also allows simultaneous estimation of the number of independent SNP loci that contribute to risk for schizophrenia. Here we assume that the number of genome-wide significant SNP associations and the amount of variance they explain in the Sweden and PGC meta-analysis results only partly reflect the underlying genetic architecture of schizophrenia, owing to inadequate sample size. Using 1000 Genomes Project imputation for Swedish and PGC samples and assuming a population risk of 0.01, we estimated that 8,300 independent SNPs contribute to the genetic basis of schizophrenia and that these SNPs account for 50% of the variance in liability to schizophrenia (95% credible intervals of 6,300–10,200 for the number of SNPs and 0.45–0.54 for total variance explained). We stress that these estimates must be interpreted in the context of the assumptions of ABPA and the strengths and weaknesses of the input data. Additional analyses (data not shown) indicate that most of the signal was derived from SNPs with allele frequencies of >0.1; low-frequency imputed SNPs were not generally inferred to be associated with schizophrenia. ABPA estimates of the genetic architecture of schizophrenia are compared with those for four biomedical diseases (**Fig. 3**)[67]. There are similarities across the estimates for these complex traits, as all are relatively highly polygenic, and common SNPs explain substantial proportions of variation. However, these results suggest that the genetic architecture of schizophrenia is shifted to the left, with greater numbers of SNPs with smaller effects.

We previously estimated the heritability of schizophrenia in Sweden to be 0.64 (95% confidence interval of 0.617–0.675) using a national pedigree sample of 9 million individuals[5], and a Danish national pedigree study of 2.6 million individuals reported a similar estimate (0.67, 95% confidence interval of 0.65–0.71)[5,68]. Using 1000 Genomes Project data with a population risk of 0.01, we found that the variance in liability estimate from GCTA accounts for 52% of the heritability (0.33/0.64), and the variance in liability estimate from ABPA account for 78% of the heritability (0.50/0.64). Imprecision is inherent in these estimates, and future work or the use of a twin meta-analysis estimate of the heritability of schizophrenia (0.81, 95% confidence interval of 0.73–0.90)[6] could revise these estimates downward. However, despite the use of different assumptions and methods, these estimates converge on a crucial qualitative implication: causal variants tagged by common SNPs make substantial contributions to the risk for schizophrenia.

DISCUSSION

These results provide deeper insight into the genetic architecture of schizophrenia than ever before achieved. We find support for 22 common variant loci (13 new) that highlight biological hypotheses for further evaluation. Some findings have immediate translational relevance. Larger studies are highly likely to uncover more common variant associations, as argued elsewhere[8,18,69,70].

Common variation is an important (and perhaps predominant) genetic contributor to risk for schizophrenia. We estimated that 6,300–10,200 independent and mostly common SNPs contribute to the etiology of schizophrenia. As one gene or structural element could contain multiple independent associations, the number of genes ultimately determined to harbor causal variation for schizophrenia will be smaller, and we expect that these genes will implicate one or more biological pathways fundamental to disease risk.

Moreover, these thousands of independent loci seem to account for a considerable fraction of the heritability of schizophrenia. It is possible that the commonly used phrase 'missing heritability' lacks precision. Indeed, if thousands of SNPs underlie schizophrenia, a statistical model containing a handful of SNPs is unlikely to account for more than a small fraction of the heritability[71]. Our results imply that the genetic architecture of schizophrenia is not dominated by uncommon variation. However, a balanced plan of attack should include well-powered searches for rare, private or *de novo* genetic variation of strong effect, given that such variants are probably more tractable to current molecular methods.

Power calculations are a fundamental component of the design of genetic studies. However, relatively extensive knowledge of genetic architecture is essential for power calculations to have maximum usefulness in study planning. We used the ABPA estimates of the posterior distribution of genotypic relative risks (**Fig. 3**) to inform power calculations by estimating the numbers of independent loci that could be detected for different sample sizes (**Supplementary Fig. 9** and **Supplementary Table 5**). For example, for 60,000 schizophrenia cases and 60,000 controls, ABPA results project that hundreds of independent SNP loci would reach genome-wide significance (mean of 794 SNPs, 95% credible interval of 362–1,154 SNPs).

Thus, for the first time, there is a clear path to increased knowledge of the etiology of schizophrenia through the application of standard, off-the-shelf genomic technologies for elucidating the effects of common variation. We suggest that a relatively thorough enumeration of the genomic loci conferring risk for schizophrenia (the 'parts list') should be a priority for the field[8]. Identifying all loci would surely be an exercise in diminishing returns. However, we propose a goal for the field: the identification of the top 2,000 loci (for example) might be sufficient to confidently and clearly identify the biological processes that mediate risk and protection for schizophrenia. Achievement of this goal would provide a strong empirical impetus for targeted biological and genetic research into the precise molecular basis of risk for schizophrenia, stratification of at-risk populations (for example, psychotic prodrome) and appropriate cellular measures for the evaluation of novel therapeutics. As indicated by our findings, greater knowledge of the genetic basis of schizophrenia can converge on increasingly specific neurobiological hypotheses that can be prioritized for subsequent investigation.

URLs. Results can be downloaded from the PGC website at http://pgc.unc.edu/ and visualized using Ricopili at http://www.broadinstitute.org/mpg/ricopili. Genotype data are available upon application from the National Institute of Mental Health (NIMH) Genetics Repository at https://www.nimhgenetics.org/. The JAG website is at http://ctglab.nl/software.

METHODS

Methods and any associated references are available in the online version of the paper.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

We are deeply grateful for the participation of all subjects contributing to this research and to the collection team that worked to recruit them: E. Flordal-Thelander, A.-B. Holmgren, M. Hallin, M. Lundin, A.-K. Sundberg, C. Pettersson, R. Satgunanthan-Dawoud, S. Hassellund, M. Rådstrom, B. Ohlander, L. Nyrén and I. Kizling. Funding support was provided by the NIMH (R01 MH077139 to P.F.S. and R01 MH095034 to P.S.), the Stanley Center for Psychiatric Research, the Sylvan Herman Foundation, the Friedman Brain Institute at the Mount Sinai School of Medicine, the Karolinska Institutet, Karolinska University Hospital, the Swedish Research Council, the Swedish County Council, the Söderström Königska Foundation and the Netherlands Scientific Organization (NWO 645-000-003). SGENE was supported by European Union grant HEALTH-F2-2009-223423 (project PsychCNVs). The study of the Aarhus sample was supported by grants from the Danish Strategic Research Council, H. Lundbeck A/S, the Faculty of Health Sciences at Aarhus University, the Lundbeck Foundation and the Stanley Research Foundation. The Wellcome Trust Case Control Consortium 2 project collection was funded by the Wellcome Trust (085475/B/08/Z and 085475/Z/08/Z). The funders had no role in study design, execution or analysis or in manuscript preparation.

AUTHOR CONTRIBUTIONS

S.R., C.O., E.A.S., M.F., N.R.W., N.S., S.E.B., S.H.L., A.B.S., A.L.R., B.K.B.-S., B.M.N., C.d.L., D.P., D. Ruderfer, F.B., J.P., K.L., M.L.H., M.V., P.H., S.S., S.A.M., S.P. and P.F.S. conducted statistical analyses. A.D.B., D.M.H., D. Rujescu, E. Sigurdsson, J.S., M.P.M., N.D., O.M., P.B.M., S.T., T.S. and V.G. ascertained subjects. A.L.C., J.J.C., S.W., Y.K., K.X. and P.F.S. performed bioinformatics analyses. K.C., J.L.M. and S.A. managed the project. B.P.R., D.W.M., F.A.O., H.S., J.T.W., K.S.K., M.G., M.J.O., N.C., P.C., the Multicenter Genetic Studies of Schizophrenia, the Psychosis Endophenotypes International Consortium, the Wellcome Trust Case Control Consortium 2, A.P.C., E.B., K.S. and M.C.O. provided replication samples and genotypes. A.K.K. interfaced with Swedish national registers. The manuscript was written by P.K.E.M., S.A.M., S.P., P.S., C.M.H. and P.F.S. The study was designed by S.P., P.S., C.M.H. and P.F.S. Funding was obtained by E. Scolnick, P.S., C.M.H. and P.F.S.

COMPETING FINANCIAL INTERESTS

The authors declare competing financial interests: details are available in the online version of the paper.

1. Saha, S., Chant, D. & McGrath, J. A systematic review of mortality in schizophrenia: is the differential mortality gap worsening over time? *Arch. Gen. Psychiatry* **64**, 1123–1131 (2007).
2. World Health Organization. *The Global Burden of Disease: 2004 Update* (World Health Organization Press, Geneva, 2008).
3. Knapp, M., Mangalore, R. & Simon, J. The global costs of schizophrenia. *Schizophr. Bull.* **30**, 279–293 (2004).
4. Lichtenstein, P. *et al.* Recurrence risks for schizophrenia in a Swedish national cohort. *Psychol. Med.* **36**, 1417–1425 (2006).
5. Lichtenstein, P. *et al.* Common genetic influences for schizophrenia and bipolar disorder: a population-based study of 2 million nuclear families. *Lancet* **373**, 234–239 (2009).
6. 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).
7. Lee, S.H. *et al.* Estimating the proportion of variation in susceptibility to schizophrenia captured by common SNPs. *Nat. Genet.* **44**, 247–250 (2012).
8. Sullivan, P.F., Daly, M.J. & O'Donovan, M. Genetic architectures of psychiatric disorders: the emerging picture and its implications. *Nat. Rev. Genet.* **13**, 537–551 (2012).
9. Ng, M.Y. *et al.* Meta-analysis of 32 genome-wide linkage studies of schizophrenia. *Mol. Psychiatry* **14**, 774–785 (2009).
10. Girard, S.L. *et al.* Increased exonic *de novo* mutation rate in individuals with schizophrenia. *Nat. Genet.* **43**, 860–863 (2011).
11. Xu, B. *et al. De novo* gene mutations highlight patterns of genetic and neural complexity in schizophrenia. *Nat. Genet.* **44**, 1365–1369 (2012).
12. Need, A.C. *et al.* Exome sequencing followed by large-scale genotyping suggests a limited role for moderately rare risk factors of strong effect in schizophrenia. *Am. J. Hum. Genet.* **91**, 303–312 (2012).
13. Kim, Y., Zerwas, S., Trace, S.E. & Sullivan, P.F. Schizophrenia genetics: where next? *Schizophr. Bull.* **37**, 456–463 (2011).
14. International Schizophrenia Consortium. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* **460**, 748–752 (2009).
15. Shi, J. *et al.* Common variants on chromosome 6p22.1 are associated with schizophrenia. *Nature* **460**, 753–757 (2009).
16. Stefansson, H. *et al.* Common variants conferring risk of schizophrenia. *Nature* **460**, 744–747 (2009).
17. Schizophrenia Psychiatric Genome-Wide Association Study Consortium. Genome-wide association study identifies five new schizophrenia loci. *Nat. Genet.* **43**, 969–976 (2011).
18. Wray, N.R. & Visscher, P.M. Narrowing the boundaries of the genetic architecture of schizophrenia. *Schizophr. Bull.* **36**, 14–23 (2010).
19. Devlin, B. & Roeder, K. Genomic control for association studies. *Biometrics* **55**, 997–1004 (1999).
20. Yang, J. *et al.* Genomic inflation factors under polygenic inheritance. *Eur. J. Hum. Genet.* **19**, 807–812 (2011).
21. de Bakker, P.I. *et al.* Practical aspects of imputation-driven meta-analysis of genome-wide association studies. *Hum. Mol. Genet.* **17**, R122–R128 (2008).
22. Pe'er, I., Yelensky, R., Altshuler, D. & Daly, M.J. Estimation of the multiple testing burden for genomewide association studies of nearly all common variants. *Genet. Epidemiol.* **32**, 381–385 (2008).
23. Major Depressive Disorder Working Group of the PGC. A mega-analysis of genome-wide association studies for major depressive disorder. *Mol. Psychiatry* **18**, 497–511 (2013).
24. Hindorff, L.A. *et al.* Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc. Natl. Acad. Sci. USA* **106**, 9362–9367 (2009).
25. Maurano, M.T. *et al.* Systematic localization of common disease-associated variation in regulatory DNA. *Science* **337**, 1190–1195 (2012).
26. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
27. Psychiatric GWAS Consortium Bipolar Disorder Working Group. Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near *ODZ4*. *Nat. Genet.* **43**, 977–983 (2011).
28. Ferreira, M.A. *et al.* Collaborative genome-wide association analysis of 10,596 individuals supports a role for Ankyrin-G (*ANK3*) and the α-1C subunit of the L-type voltage-gated calcium channel (*CACNA1C*) in bipolar disorder. *Nat. Genet.* **40**, 1056–1058 (2008).
29. Hamshere, M.L. *et al.* Genome-wide significant associations in schizophrenia to *ITIH3/4*, *CACNA1C* and *SDCCAG8*, and extensive replication of associations reported by the Schizophrenia PGC. *Mol. Psychiatry* **18**, 708–712 (2013).
30. Cross-Disorder Group of the Psychiatric Genomics Consortium. Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. *Lancet* **381**, 1371–1379 (2013).
31. Cichon, S. *et al.* Genome-wide association study identifies genetic variation in neurocan as a susceptibility factor for bipolar disorder. *Am. J. Hum. Genet.* **88**, 372–381 (2011).
32. Bidaud, I., Mezghrani, A., Swayne, L.A., Monteil, A. & Lory, P. Voltage-gated calcium channels in genetic diseases. *Biochim. Biophys. Acta* **1763**, 1169–1174 (2006).
33. Bichet, D. *et al.* The I-II loop of the Ca^{2+} channel α1 subunit contains an endoplasmic reticulum retention signal antagonized by the β subunit. *Neuron* **25**, 177–190 (2000).
34. Müller, C.S. *et al.* Quantitative proteomics of the Ca_v2 channel nano-environments in the mammalian brain. *Proc. Natl. Acad. Sci. USA* **107**, 14950–14957 (2010).
35. Woodside, B.L., Borroni, A.M., Hammonds, M.D. & Teyler, T.J. NMDA receptors and voltage-dependent calcium channels mediate different aspects of acquisition and retention of a spatial memory task. *Neurobiol. Learn. Mem.* **81**, 105–114 (2004).
36. Moosmang, S. *et al.* Role of hippocampal $Ca_v1.2$ Ca^{2+} channels in NMDA receptor–independent synaptic plasticity and spatial memory. *J. Neurosci.* **25**, 9883–9892 (2005).
37. White, J.A. *et al.* Conditional forebrain deletion of the L-type calcium channel $Ca_v1.2$ disrupts remote spatial memories in mice. *Learn. Mem.* **15**, 1–5 (2008).
38. Mangoni, M.E. *et al.* Voltage-dependent calcium channels and cardiac pacemaker activity: from ionic currents to genes. *Prog. Biophys. Mol. Biol.* **90**, 38–63 (2006).
39. Perrin, M.J. & Gollob, M.H. Genetics of cardiac electrical disease. *Can. J. Cardiol.* **29**, 89–99 (2013).
40. Splawski, I. *et al.* $Ca_v1.2$ calcium channel dysfunction causes a multisystem disorder including arrhythmia and autism. *Cell* **119**, 19–31 (2004).
41. Koponen, H. *et al.* Schizophrenia and sudden cardiac death: a review. *Nord. J. Psychiatry* **62**, 342–345 (2008).
42. Stöllberger, C., Huber, J.O. & Finsterer, J. Antipsychotic drugs and QT prolongation. *Int. Clin. Psychopharmacol.* **20**, 243–251 (2005).
43. McKusick, V.A. Mendelian Inheritance in Man and its online version, OMIM. *Am. J. Hum. Genet.* **80**, 588–604 (2007).
44. Lewis, B.P., Burge, C.B. & Bartel, D.P. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* **120**, 15–20 (2005).
45. Szulwach, K.E. *et al.* Cross talk between microRNA and epigenetic regulation in adult neurogenesis. *J. Cell Biol.* **189**, 127–141 (2010).
46. Smrt, R.D. *et al.* MicroRNA miR-137 regulates neuronal maturation by targeting ubiquitin ligase mind bomb-1. *Stem Cells* **28**, 1060–1070 (2010).

47. 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).
48. Cabili, M.N. *et al.* Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev.* **25**, 1915–1927 (2011).
49. Lee, P.H., O'Dushlaine, C., Thomas, B. & Purcell, S. InRich: interval-based enrichment analysis for genome-wide association studies. *Bioinformatics* **28**, 1797–1799 (2012).
50. Majewski, J. & Pastinen, T. The study of eQTL variations by RNA-seq: from SNPs to phenotypes. *Trends Genet.* **27**, 72–79 (2011).
51. Cookson, W., Liang, L., Abecasis, G., Moffatt, M. & Lathrop, M. Mapping complex disease traits with global gene expression. *Nat. Rev. Genet.* **10**, 184–194 (2009).
52. Nicolae, D.L. *et al.* Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS. *PLoS Genet.* **6**, e1000888 (2010).
53. Stranger, B.E. *et al.* Patterns of *cis* regulatory variation in diverse human populations. *PLoS Genet.* **8**, e1002639 (2012).
54. Grundberg, E. *et al.* Mapping *cis*- and *trans*-regulatory effects across multiple tissues in twins. *Nat. Genet.* **44**, 1084–1089 (2012).
55. Lango Allen, H. *et al.* Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature* **467**, 832–838 (2010).
56. Emilsson, V. *et al.* Genetics of gene expression and its effect on disease. *Nature* **452**, 423–428 (2008).
57. de Jong, S. *et al.* Expression QTL analysis of top loci from GWAS meta-analysis highlights additional schizophrenia candidate genes. *Eur. J. Hum. Genet.* **20**, 1004–1008 (2012).
58. Fransen, K. *et al.* Analysis of SNPs with an effect on gene expression identifies *UBE2L3* and *BCL3* as potential new risk genes for Crohn's disease. *Hum. Mol. Genet.* **19**, 3482–3488 (2010).
59. Luo, R. *et al.* Genome-wide transcriptome profiling reveals the functional impact of rare *de novo* and recurrent CNVs in autism spectrum disorders. *Am. J. Hum. Genet.* **91**, 38–55 (2012).
60. Speliotes, E.K. *et al.* Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nat. Genet.* **42**, 937–948 (2010).
61. Zeller, T. *et al.* Genetics and beyond—the transcriptome of human monocytes and disease susceptibility. *PLoS ONE* **5**, e10693 (2010).
62. Gamazon, E.R., Huang, R.S., Cox, N.J. & Dolan, M.E. Chemotherapeutic drug susceptibility associated SNPs are enriched in expression quantitative trait loci. *Proc. Natl. Acad. Sci. USA* **107**, 9287–9292 (2010).
63. Thurman, R.E. *et al.* The accessible chromatin landscape of the human genome. *Nature* **489**, 75–82 (2012).
64. Degner, J.F. *et al.* DNase I sensitivity QTLs are a major determinant of human expression variation. *Nature* **482**, 390–394 (2012).
65. Yang, J., Lee, S.H., Goddard, M.E. & Visscher, P.M. GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* **88**, 76–82 (2011).
66. Saha, S., Chant, D., Welham, J. & McGrath, J. A systematic review of the prevalence of schizophrenia. *PLoS Med.* **2**, e141 (2005).
67. Stahl, E.A. *et al.* Bayesian inference analyses of the polygenic architecture of rheumatoid arthritis. *Nat. Genet.* **44**, 483–489 (2012).
68. Wray, N.R. & Gottesman, I.I. Using summary data from the Danish national registers to estimate heritabilities for schizophrenia, bipolar disorder, and major depressive disorder. *Front. Genet.* **3**, 118 (2012).
69. Lander, E.S. Initial impact of the sequencing of the human genome. *Nature* **470**, 187–197 (2011).
70. Sullivan, P. Don't give up on GWAS. *Mol. Psychiatry* **17**, 2–3 (2012).
71. Park, J.H. *et al.* Estimation of effect size distribution from genome-wide association studies and implications for future discoveries. *Nat. Genet.* **42**, 570–575 (2010).
72. Stahl, E.A. *et al.* Genome-wide association study meta-analysis identifies seven new rheumatoid arthritis risk loci. *Nat. Genet.* **42**, 508–514 (2010).
73. Betancur, C. Etiological heterogeneity in autism spectrum disorders: more than 100 genetic and genomic disorders and still counting. *Brain Res.* **1380**, 42–77 (2011).
74. Chiurazzi, P., Schwartz, C.E., Gecz, J. & Neri, G. *XLMR* genes: update 2007. *Eur. J. Hum. Genet.* **16**, 422–434 (2008).
75. Inlow, J.K. & Restifo, L.L. Molecular and comparative genetics of mental retardation. *Genetics* **166**, 835–881 (2004).

[1]Analytical and Translational Genetics Unit, Massachusetts General Hospital, Boston, Massachusetts, USA. [2]Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA. [3]Department of Genetics, University of North Carolina, Chapel Hill, North Carolina, USA. [4]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden. [5]Division of Mental Health and Addiction, Oslo University Hospital, Oslo, Norway. [6]Department of Psychiatry, Division of Psychiatric Genomics, Icahn School of Medicine at Mount Sinai, New York, New York, USA. [7]Queensland Brain Institute, The University of Queensland, Brisbane, Queensland, Australia. [8]Department of Psychiatry, University of North Carolina, Chapel Hill, North Carolina, USA. [9]deCODE Genetics, Reykjavik, Iceland. [10]Aarhus University Hospital, Risskov, Denmark. [11]Centre for Integrative Sequencing (iSEQ), Aarhus University, Aarhus, Denmark. [12]Lundbeck Foundation Initiative for Integrative Psychiatric Research (iPSYCH), Aarhus and Copenhagen, Denmark. [13]Department of Psychiatry, Trinity College Dublin, Dublin, Ireland. [14]Medical Research Council (MRC) Centre for Psychiatric Genetics and Genomics, Cardiff University School of Medicine, Cardiff, UK. [15]Department of Functional Genomics, Clinical Genetics, Center for Neurogenomics and Cognitive Research, VU University Amsterdam and VU Medical Center, Amsterdam, The Netherlands. [16]Institute for Computing and Information Sciences, Radboud University, Nijmegen, The Netherlands. [17]University Clinic of Psychiatry, Department of Child and Adolescent Psychiatry, Skopje, Republic of Macedonia. [18]Neuropsychiatric Genetics Research Group, Trinity College Dublin, Dublin, Ireland. [19]Mental Health Research Center, Russian Academy of Medical Sciences, Moscow, Russia. [20]Statens Serum Institut, Copenhagen, Denmark. [21]Department of Psychiatry, Virginia Commonwealth University, Richmond, Virginia, USA. [22]Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University, Richmond, Virginia, USA. [23]Institute of Psychiatry at King's College London, London, UK. [24]Centre for Psychiatric Research, Aarhus University Hospital, Risskov, Denmark. [25]National Centre for Register-Based Research, Aarhus University, Aarhus, Denmark. [26]Centre for Public Health, Queen's University, Belfast, UK. [27]University of Belgrade, Faculty of Medicine, Belgrade, Serbia. [28]Department of Child and Adolescent Psychiatry, Erasmus University Medical Centre, Rotterdam, The Netherlands. [29]Department of Neuroscience, King's College London, London, UK. [30]Department of Human and Molecular Genetics, Virginia Commonwealth University, Richmond, Virginia, USA. [31]Department of Psychiatry, University of Halle, Halle, Germany. [32]Department of Psychiatry, University of Munich, Munich, Germany. [33]Department of Psychiatry, University of Iceland, Reykjavik, Iceland. [34]Department of Psychiatry, Tbilisi State Medical University, Tbilisi, Georgia. [35]Center for Neurogenomics and Cognitive Research, VU University Amsterdam, Amsterdam, The Netherlands. [36]Department of Molecular and Cellular Neuroscience, VU University Amsterdam, Amsterdam, The Netherlands. [37]Mental Health and Substance Abuse Services, National Institute for Health and Welfare, Helsinki, Finland. [38]Section of Psychiatry, University of Verona, Verona, Italy. [39]Full lists of members and affiliations appear at the end of the paper. [40]Institute of Cognitive Neuroscience, University College London, London, UK. [41]Mental Health Sciences Unit, University College London, London, UK. [42]Psychiatric and Neurodevelopmental Genetics Unit, Massachusetts General Hospital, Boston, Massachusetts, USA. [43]Department of Genetics, Harvard Medical School, Boston, Massachusetts, USA. [44]These authors contributed equally to this work. [45]These authors jointly directed this work. Correspondence should be addressed to P.F.S. (pfsulliv@med.unc.edu).

Multicenter Genetic Studies of Schizophrenia Consortium:

Douglas F Levinson[46], Pablo V Gejman[47], Kenneth S Kendler[21,22], Claudine Laurent[48], Bryan J Mowry[7], Michael C O'Donovan[14], Michael J Owen[14], Ann E Pulver[49], Brien P Riley[21,22,30], Sibylle G Schwab[50], Dieter B Wildenauer[51], Frank Dudbridge[52], Peter Holmans[14], Jianxin Shi[53], Margot Albus[54], Madeline Alexander[46], Dominique Campion[55], David Cohen[56], Dimitris Dikeos[57], Jubao Duan[47], Peter Eichhammer[58], Stephanie Godard[59], Mark Hansen[60], F Bernard Lerer[61], Kung-Yee Liang[62], Wolfgang Maier[63], Jacques Mallet[64], Deborah A Nertney[7], Gerald Nestadt[49], Nadine Norton[14], Francis A O'Neill[26], George N Papadimitriou[57], Robert Ribble[22], Alan R Sanders[47], Jeremy M Silverman[6], Dermot Walsh[65], Nigel M Williams[14] & Brandon Wormley[22]

Psychosis Endophenotypes International Consortium:

Maria J Arranz[23,66], Steven Bakker[67], Stephan Bender[68,69], Elvira Bramon[23,40,41], David Collier[23,70], Benedicto Crespo-Facorro[71,72], Jeremy Hall[73], Conrad Iyegbe[23], Assen Jablensky[74], Rene S Kahn[67], Luba Kalaydjieva[75,76], Stephen Lawrie[73], Cathryn M Lewis[23], Kuang Lin[23], Don H Linszen[77], Ignacio Mata[71,72], Andrew McIntosh[73], Robin M Murray[23], Roel A Ophoff[78], John Powell[23,29], Dan Rujescu[31,32], Jim Van Os[23,79], Muriel Walshe[23], Matthias Weisbrod[69] & Durk Wiersma[80]

Wellcome Trust Case Control Consortium 2:

Management Committee: Peter Donnelly (Chair)[81,82], Ines Barroso (Deputy Chair)[83], Jenefer M Blackwell[84,85], Elvira Bramon[23,40,41], Matthew A Brown[86], Juan P Casas[87,88], Aiden P Corvin[13,18], Panos Deloukas[83], Audrey Duncanson[89], Janusz Jankowski[90], Hugh S Markus[91], Christopher G Mathew[92], Colin N A Palmer[93], Robert Plomin[94], Anna Rautanen[82], Stephen J Sawcer[95], Richard C Trembath[92], Ananth C Viswanathan[96,97] & Nicholas W Wood[98]

Data and Analysis Group: Chris C A Spencer[82], Gavin Band[82], Céline Bellenguez[82], Colin Freeman[82], Garrett Hellenthal[82], Eleni Giannoulatou[82], Matti Pirinen[82], Richard D Pearson[82], Amy Strange[82], Zhan Su[82], Damjan Vukcevic[82] & Peter Donnelly[81,82]

DNA, Genotyping, Data QC and Informatics Group: Cordelia Langford[83], Sarah E Hunt[83], Sarah Edkins[83], Rhian Gwilliam[83], Hannah Blackburn[83], Suzannah J Bumpstead[83], Serge Dronov[83], Matthew Gillman[83], Emma Gray[83], Naomi Hammond[83], Alagurevathi Jayakumar[83], Owen T McCann[83], Jennifer Liddle[83], Simon C Potter[83], Radhi Ravindrarajah[83], Michelle Ricketts[83], Avazeh Tashakkori-Ghanbaria[83], Matthew J Waller[83], Paul Weston[83], Sara Widaa[83], Pamela Whittaker[83], Ines Barroso[83] & Panos Deloukas[83]

Publications Committee: Christopher G Mathew (Chair)[92], Jenefer M Blackwell[84,85], Matthew A Brown[86], Aiden P Corvin[13,18], Mark I McCarthy[99] & Chris C A Spencer[82]

[46]Department of Psychiatry and Behavioral Sciences, Stanford University, Stanford, California, USA. [47]Department of Psychiatry and Behavioral Sciences, NorthShore University HealthSystem and University of Chicago, Evanston, Illinois, USA. [48]Child and Adolescent Psychiatry, Pierre and Marie Curie Faculty of Medicine, Brain and Spinal Cord Institute, Paris, France. [49]Department of Psychiatry and Behavioral Sciences, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA. [50]Psychiatry and Psychotherapy Clinic, Friedrich-Alexander University, Erlangen-Nuremberg, Erlangen, Germany. [51]Department of Psychiatry and Clinical Neurosciences, University of Western Australia, Perth, Western Australia, Australia. [52]Department of Non-Communicable Disease Epidemiology, London School of Hygiene and Tropical Medicine, London, UK. [53]Division of Cancer Epidemiology & Genetics, National Cancer Institute, Bethesda, Maryland, USA. [54]State Mental Hospital, Haar, Germany. [55]INSERM U1079, Faculté de Médecine, Rouen, France. [56]Pierre and Marie Curie Faculty of Medicine, Institute for Intelligent Systems and Robotics, Paris, France. [57]First Department of Psychiatry, University of Athens Medical School, Athens, Greece. [58]Department of Psychiatry, University of Regensburg, Regensburg, Germany. [59]INSERM, Institut de Myologie, Hôpital Pitié Salpêtrière, Paris, France. [60]Illumina, Inc., La Jolla, California, USA. [61]Department of Psychiatry, Hadassah–Hebrew University Medical Center, Jerusalem, Israel. [62]Department of Biostatistics, Johns Hopkins University, Baltimore, Maryland, USA. [63]Department of Psychiatry, University of Bonn, Bonn, Germany. [64]CNRS, Laboratoire de Génétique Moléculaire de la Neurotransmission et des Processus Neurodégénératifs, Hôpital Pitié Salpêtrière, Paris, France. [65]The Health Research Board, Dublin, Ireland. [66]Fundacio de Docencia i Recerca Mutua de Terrassa, Universitat de Barcelona, Barcelona, Spain. [67]Department of Psychiatry, Rudolf Magnus Institute of Neuroscience, University Medical Center Utrecht, Utrecht, The Netherlands. [68]Child and Adolescent Psychiatry, University of Technology Dresden, Dresden, Germany. [69]Section for Experimental Psychopathology, General Psychiatry, Heidelberg, Germany. [70]Discovery Neuroscience Research, Eli Lilly and Company, London, UK. [71]CIBERSAM (Centro Investigación Biomédica en Red Salud Mental), Madrid, Spain. [72]University Hospital Marques de Valdecilla, Instituto de Formacion e Investigacion Marques de Valdecilla, University of Cantabria, Santander, Spain. [73]Division of Psychiatry, University of Edinburgh, Edinburgh, UK. [74]Centre for Clinical Research in Neuropsychiatry, The University of Western Australia, Perth, Western Australia, Australia. [75]Centre for Medical Research, The University of Western Australia, Perth, Western Australia, Australia. [76]Western Australian Institute for Medical Research, The University of Western Australia, Perth, Western Australia, Australia. [77]Department of Psychiatry, Academic Medical Center, University of Amsterdam, Amsterdam, The Netherlands. [78]Center for Neurobehavioral Genetics, University of California, Los Angeles, Los Angeles, California, USA. [79]Maastricht University Medical Centre, South Limburg Mental Health Research and Teaching Network, EURON, Maastricht, The Netherlands. [80]Department of Psychiatry, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands. [81]Department of Statistics, University of Oxford, Oxford, UK. [82]Wellcome Trust Centre for Human Genetics, Oxford, UK. [83]Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, UK. [84]Cambridge Institute for Medical Research, University of Cambridge School of Clinical Medicine, Cambridge, UK. [85]Telethon Institute for Child Health Research, Centre for Child Health Research, University of Western Australia, Subiaco, Western Australia, Australia. [86]Diamantina Institute of Cancer, Immunology and Metabolic Medicine, Princess Alexandra Hospital, University of Queensland, Brisbane, Queensland, Australia. [87]Department of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, London, UK. [88]Department of Epidemiology and Public Health, University College London, London, UK. [89]Molecular and Physiological Sciences, The Wellcome Trust, London, UK. [90]Peninsula School of Medicine and Dentistry, Plymouth University, Plymouth, UK. [91]Clinical Neurosciences, St. George's University of London, London, UK. [92]Department of Medical and Molecular Genetics, School of Medicine, King's College London, Guy's Hospital, London, UK. [93]Biomedical Research Centre, Ninewells Hospital and Medical School, Dundee, UK. [94]Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, King's College London, London, UK. [95]Department of Clinical Neurosciences, University of Cambridge, Addenbrooke's Hospital, Cambridge, UK. [96]Institute of Ophthalmology, University College London, London, UK. [97]National Institute for Health Research (NIHR) Biomedical Research Centre at Moorfields Eye Hospital National Health Service Foundation Trust, London, UK. [98]Department of Molecular Neuroscience, Institute of Neurology, London, UK. [99]Oxford Centre for Diabetes, Endocrinology and Metabolism, Churchill Hospital, Oxford, UK.

ONLINE METHODS

Overview. We present here the preplanned principal analyses for this project. To advance knowledge of schizophrenia, a minority of samples were included in previous reports. Genotyping was conducted in 6 batches (denoted Sw1–Sw6) with total sample sizes of 464, 694, 1,498, 2,388, 4,461 and 2,345. Genotypes were generated as sufficient numbers of samples accumulated from field work in Sweden. The 2009 International Schizophrenia Consortium report contained GWAS data from the Sw1 and Sw2 subjects (N = 1,158; 9.8% of the sample before quality control)[14]. The 2011 PGC schizophrenia paper also contained GWAS data from the Sw1 and Sw2 subjects plus ~80 SNPs from Sw3 and Sw4 in the replication phase[17]. The 2012 Bergen *et al.* paper focused on contrasting schizophrenia with bipolar disorder and reported GWAS results from Sw1–Sw4 (N = 4,044; 42.6% of the full sample)[76]. Thus, of the total sample of 11,850 Swedish subjects before quality control (5,351 cases and 6,509 controls), 57.4% have not been reported previously.

Subjects. All procedures were approved by ethical committees at the Karolinska Institutet and the University of North Carolina, and all subjects provided written informed consent (or legal guardian consent and subject assent). Sample collection occurred from 2005–2011.

Cases with schizophrenia were identified via the Swedish Hospital Discharge Register[77,78], which captures all public and private inpatient hospitalizations. The register is complete from 1987 and is augmented by psychiatric data from 1973–1986. The register contains International Classification of Disease discharge diagnoses[79–81] made by attending physicians for each hospitalization[82–85]. Case inclusion criteria included ≥2 hospitalizations with a discharge diagnosis of schizophrenia, both parents born in Scandinavia and age of ≥18 years. Case exclusion criteria included hospital register diagnosis of any medical or psychiatric disorder mitigating a confident diagnosis of schizophrenia as determined by expert review and involved the removal of 3.4% of the eligible cases owing to the primacy of another psychiatric disorder (0.9%) or a general medical condition (0.3%) or uncertainties in the hospital discharge register (for example, contiguous admissions with brief total duration; 2.2%).

The validity of this case definition of schizophrenia is described in the **Supplementary Note** (**Supplementary Figs. 10** and **11** and **Supplementary Tables 6** and **7**), and this validity is strongly supported by clinical, epidemiological, genetic epidemiological and genetic evidence.

Controls were selected at random from Swedish population registers, with the goal of obtaining an appropriate control group and avoiding 'supernormal' controls[86]. Control inclusion criteria included never being hospitalized for schizophrenia or bipolar disorder (given evidence of genetic overlap with schizophrenia)[5,14,87], both parents born in Scandinavia and age of ≥18 years.

Of the potential cases and controls who were alive and contactable, refusal rates were higher for cases than for controls (46.7% versus 41.7%). However, these proportions compare favorably with modern refusal rates in epidemiology (59% for cross-sectional and 44% for case-control studies)[88,89] and with the refusal rate in a recent large Norwegian longitudinal study (58%)[90]. For cases, comorbidity with drug and/or alcohol abuse or dependence did not predict participation nor did any subtype of schizophrenia (for example, paranoid or disorganized type). The sample was approximately representative of the Swedish populace with regard to county of birth (**Supplementary Fig. 12**).

Genotyping, quality control and imputation. DNA was extracted from peripheral blood samples at the Karolinska Institutet Biobank. Samples were genotyped in six batches at the Broad Institute using Affymetrix 5.0 (3.9%), Affymetrix 6.0 (38.6%) and Illumina OmniExpress (57.4%) chips according to the manufacturers' protocols (**Supplementary Table 8**). Genotype calling, quality control and imputation were carried out in four sets corresponding to data from Affymetrix 5.0 (Sw1), Affymetrix 6.0 (Sw2–Sw4) and OmniExpress (Sw5 and Sw6) batches. Genotypes were called using Birdsuite (Affymetrix) or BeadStudio (Illumina). The quality control parameters applied included SNP missingness of <0.05 (before sample removal); subject missingness of <0.02; autosomal heterozygosity deviation; SNP missingness of <0.02 (after sample removal); difference in SNP missingness between cases and controls of <0.02; and deviation from Hardy-Weinberg equilibrium ($P < 1 \times 10^{-6}$ in controls or $P < 1 \times 10^{-10}$ in cases).

After basic quality control, 77,986 autosomal SNPs directly genotyped on all 4 GWAS platforms were extracted and pruned to remove SNPs in LD ($r^2 > 0.05$) or with minor allele frequency of <0.05, leaving 39,239 SNPs suitable for robust relatedness testing and population structure analysis (**Supplementary Fig. 13**). Relatedness testing was carried out with PLINK[91] and pairs of subjects with $\hat{\pi}$ of >0.2 were identified, and one member of each relative pair was removed at random. Principal-component estimation was carried out with the same collection of SNPs. We tested 20 principal components for phenotype association (using logistic regression with batch indicator variables included as covariates) and evaluated their impact on the genome-wide test statistics using λ (ref. 19) after genome-wide association of the specified principal component, and 11 principal components were included in all association analyses.

Genotype imputation was performed using the prephasing/imputation stepwise approach implemented in IMPUTE2 and SHAPEIT (chunk size of 3 Mb and default parameters)[92,93]. The imputation reference set consisted of 2,186 phased haplotypes from the full 1000 Genomes Project data set (March 2012; 40,318,245 variants). Evaluation of λ_{GC} led to the removal of SNPs with control allele frequencies of <0.005 or >0.995, imputation 'info' values of <0.2 or that were genotyped only in the smallest sample set (Sw1). Given that male sex is a risk factor for schizophrenia[94], we conducted chromosome X imputation for subjects passing quality control for the autosomal analysis (excluding X-chromosome SNPs with missingness of ≥0.05 or Hardy-Weinberg equilibrium P of $<1 \times 10^{-6}$ in females). Imputation was performed separately for males and females, and gene dosages were tested for association under an additive logistic regression model using the same covariates as for the autosomal analysis. All genomic locations are given in NCBI Build 37/UCSC hg19 coordinates.

Statistical analysis. We first analyzed Swedish cases and controls (N = 11,244) and then conducted a meta-analysis with the PGC results for schizophrenia to evaluate our results with respect to the world's literature (N = 20,899 after removing 954 subjects from Sw1 and Sw2)[17]. To maximize comparability, the Swedish samples were run through the same analytical pipeline used for the PGC samples. Association testing was carried out in PLINK using imputed SNP dosages and the principal components described above as covariates[22]. Meta-analysis was conducted using an inverse-weighted fixed-effects model[21]. To evaluate the comparability of the Swedish results with those from the PGC schizophrenia study, we used sign tests and risk score profiling based on sets of carefully selected SNPs[17].

Summarizing regional data using 'clumping'. Many GWAS findings implicate an extended region containing multiple SNPs with significant association. These are not independent associations but result because of high LD between associated SNPs. It is useful to summarize these associations in terms of the index SNP with the strongest association and other SNPs in high LD with the index SNP. To summarize GWAS findings, we used the following settings in PLINK to retain SNPs with association $P < 0.0001$ and $r^2 < 0.2$ within 500-kb windows:–clump-p1 1e-4–clump-p2 1e-4–clump-r2 0.2–clump-kb 500.

Sign tests. We used sign tests to compare the overall patterns of results between the Swedish and PGC schizophrenia samples. We used the clumping settings above to derive a filtered set of SNPs. Because of the strong signal and high LD in the MHC, we kept only one SNP from the extended MHC region. We then determined the number of SNPs whose logistic regression β coefficient signs were the same between two independent samples. Under the null, the expectation is that 50% of the signs of these SNPs will be the same between two independent sets of results. The significance of the observed proportion was evaluated using the binomial distribution.

The significance test was carried out in two ways: (i) by selecting SNPs from Sw1–Sw6 results and evaluating the signs in the independent PGC results and (ii) by reversing the procedure (selecting from PGC and evaluating signs in Sw1–Sw6). We obtained similar results selecting SNPs for (i) association $P < 1 \times 10^{-5}$, (ii) association $P < 1 \times 10^{-6}$ and (iii) keeping one SNP every 3 Mb (effectively removing or greatly minimizing the effects of residual LD).

Risk profile score (RPS). We used RPS[14] as an alternative and complementary way to compare the overall patterns of results from the PGC schizophrenia

analysis (discovery sample) with the independent Swedish results (target sample). We began by selecting high-quality, relatively independent SNPs with unambiguous directions of effect: from the PGC imputed results file, we generated a subset of results containing SNPs with allele frequency of 0.02–0.98 and imputation info scores of >0.9. We then removed SNPs in high LD using clumping (i.e., retaining all SNPs with r^2 of <0.25 within 500-kb windows): –clump-p1 1–clump-p2 1–clump-r2 0.25–clump-kb 500. For RPS, we wished to evaluate SNP effects across the *P*-value spectrum. Again, owing to the strong signal and high LD in the MHC, only one SNP was kept from the extended MHC region.

We used the resulting list from the PGC to calculate schizophrenia RPSs in the independent Swedish samples using the –score function in PLINK. We did this ten times using different subsets of the PGC SNPs selected by increasing *P*-value threshold. From the set of filtered SNPs from the PGC, we evaluated 10 different association *P*-value thresholds (P_T): 0.0001, 0.001, 0.01, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5 and 1.0 (i.e., including all SNPs). For each of these ten sets of SNPs derived from PGC, the schizophrenia risk profile score (the number of schizophrenia risk alleles weighted by the logistic regression β) was calculated for each case and control in Sw1–Sw6. Logistic regression was then used to test whether Swedish cases had significantly different burden of schizophrenia risk alleles in comparison to controls (including ancestry principal components as covariates). To estimate the proportion of variance of case-control status in the Swedish samples accounted for by the RPS from PGC, we used the difference in the Nagelkerke pseudo R^2 value contrasting a logistic regression model, containing the risk profile score plus ancestry covariates with a logistic regression model containing the covariates alone.

Gene-set analysis. One way to understand polygenic associations for a complex trait is to determine whether the implicated genetic variants are in genes that comprise a biological pathway. Gene-set analysis includes evaluation of genetic variants in genes that are grouped on the basis of their interacting role in biological pathways (biological pathway analysis) and genes that share similar cellular functions (functional gene-set analysis).

We used JAG (Joint Association of Genetic variants; see URLs) to conduct gene-set analyses. This method has previously been applied to the International Schizophrenia Consortium data by Lips *et al.*[95]. JAG tests for the association of specified gene sets with schizophrenia as applied to individual-level genotype data, which tend to be more powerful than using summary statistics. JAG constructs a test statistic for each gene set. JAG includes both self-contained and competitive tests. These two approaches evaluate different null hypotheses. Statistical significance (P_{self} and P_{comp}) are determined using permutation. First, the self-contained test evaluates the null hypothesis that a defined set of genes is not associated with schizophrenia, while accounting for some of the properties of the SNPs being studied (for example, LD structure). Second, the competitive test evaluates whether a specific set of genes has evidence for stronger associations with schizophrenia than randomly selected sets of control genes (with the latter matched to the former using the same effective number of SNPs per gene set). Thus, a competitive test of the null hypothesis is that these genes are not more strongly associated than those in a similar but randomly selected set of genes. That is, the comparison is more one to the average degree of association across genes. The principal comparison is the competitive test, and we present self-contained tests for completeness. Competitive gene-set tests are more appropriate for a polygenic disease such as schizophrenia because they explicitly prioritize gene sets that show a greater average degree of association, over and above the polygenic background, rather than prioritizing larger but more weakly enriched gene sets (as self-contained tests would tend to do).

Replication. We obtained replication association results from six independent samples totaling 7,452 cases, 20,404 controls and 581 trios (**Supplementary Table 9**). These subjects are not included in the Swedish samples or in the PGC mega-analysis[17]. The independent samples were from SGENE+ (ref. 16), CLOZUK[29], the Irish Schizophrenia Genomics Consortium[96], the Psychosis Endophenotype Consortium[97] and the Multicenter Family Study[98]. After selecting for association *P* of $<1 \times 10^{-5}$ in the Sweden and PGC meta-analysis and accounting for LD, we requested association results for 194 genomic regions.

76. Bergen, S.E. *et al.* Genome-wide association study in a Swedish population yields support for greater CNV and MHC involvement in schizophrenia compared to bipolar disorder. *Mol. Psychiatry* **17**, 880–886 (2012).
77. Kristjansson, E., Allebeck, P. & Wistedt, B. Validity of the diagnosis of schizophrenia in a psychiatric inpatient register. *Nord. Psykiatr. Tidsskr.* **41**, 229–234 (1987).
78. Dalman, C., Broms, J., Cullberg, J. & Allebeck, P. Young cases of schizophrenia identified in a national inpatient register—are the diagnoses valid? *Soc. Psychiatry Psychiatr. Epidemiol.* **37**, 527–531 (2002).
79. World Health Organization. *International Classification of Diseases* (World Health Organization, Geneva, 1967).
80. World Health Organization. *International Classification of Diseases* (World Health Organization, Geneva, 1978).
81. World Health Organization. *International Classification of Diseases* (World Health Organization, Geneva, 1992).
82. Hultman, C.M., Sparen, P., Takei, N., Murray, R.M. & Cnattingius, S. Prenatal and perinatal risk factors for schizophrenia, affective psychosis, and reactive psychosis of early onset: case-control study. *Br. Med. J.* **318**, 421–426 (1999).
83. Zammit, S. *et al.* Investigating the association between cigarette smoking and schizophrenia in a cohort study. *Am. J. Psychiatry* **160**, 2216–2221 (2003).
84. Andersson, R.E., Olaison, G., Tysk, C. & Ekbom, A. Appendectomy and protection against ulcerative colitis. *N. Engl. J. Med.* **344**, 808–814 (2001).
85. Hansson, L.E. *et al.* The risk of stomach cancer in patients with gastric or duodenal ulcer disease. *N. Engl. J. Med.* **335**, 242–249 (1996).
86. Schwartz, S. & Susser, E. Genome-wide association studies: does only size matter? *Am. J. Psychiatry* **167**, 741–744 (2010).
87. Craddock, N. & Owen, M.J. The Kraepelinian dichotomy—going, going but still not gone. *Br. J. Psychiatry* **196**, 92–95 (2010).
88. Hartge, P. Participation in population studies. *Epidemiology* **17**, 252–254 (2006).
89. Morton, L.M., Cahill, J. & Hartge, P. Reporting participation in epidemiologic studies: a survey of practice. *Am. J. Epidemiol.* **163**, 197–203 (2006).
90. Bulik, C.M. *et al.* Patterns of remission, continuation and incidence of broadly defined eating disorders during early pregnancy in the Norwegian Mother and Child Cohort Study (MoBa). *Psychol. Med.* **37**, 1109–1118 (2007).
91. Purcell, S. *et al.* PLINK: a toolset for whole-genome association and population-based linkage analysis. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
92. Howie, B., Marchini, J. & Stephens, M. Genotype imputation with thousands of genomes. *G3* **1**, 457–470 (2011).
93. Delaneau, O., Marchini, J. & Zagury, J.F. A linear complexity phasing method for thousands of genomes. *Nat. Methods* **9**, 179–181 (2012).
94. McGrath, J.J. Variations in the incidence of schizophrenia: data versus dogma. *Schizophr. Bull.* **32**, 195–197 (2006).
95. Lips, E.S. *et al.* Functional gene group analysis identifies synaptic gene groups as risk factor for schizophrenia. *Mol. Psychiatry* **17**, 996–1006 (2012).
96. Irish Schizophrenia Genomics Consortium & Wellcome Trust Case Control Consortium. Genome-wide association study implicates *HLA-C*01:02* as a risk factor at the MHC locus in schizophrenia. *Biol. Psychiatry* **72**, 620–628 (2012).
97. Psychosis Endophenotypes International Consortium & Wellcome Trust Case-Control Consortium 2. A genome-wide association analysis of a broad psychosis phenotype identifies three loci for further investigation. *Biol. Psychiatry* doi:10.1016/j.biopsych.2013.03.033 (17 July 2013).
98. Levinson, D.F. *et al.* Genome-wide association study of multiplex schizophrenia pedigrees. *Am. J. Psychiatry* **169**, 963–973 (2012).