Published in final edited form as:

Nat Genet. 2013 October ; 45(10): 1150–1159. doi:10.1038/ng.2742.

Genome-wide Association Analysis Identifies 14 New Risk Loci for Schizophrenia

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

Abstract

Schizophrenia is a heritable disorder with 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, 6,243 controls) followed by meta-analysis with prior schizophrenia GWAS (8,832 cases, 12,067 controls) and finally by replication of SNPs in 168 genomic regions in independent samples (7,413 cases, 19,762 controls, and 581 trios). In total, 22 regions met genome-wide significance (14 novel and one previously implicated in bipolar disorder). The results strongly implicate calcium signaling in the etiology of schizophrenia, and include genome-wide significant results for *CACNA1C* and *CACNB2* whose protein products interact. We estimate that ~8,300 independent and predominantly common SNPs contribute to risk for schizophrenia and that these collectively account for most of its heritability. Common genetic variation plays an important role in the etiology of schizophrenia, and larger studies will allow more detailed understanding of this devastating disorder.

Keywords

schizophrenia; genetics; genome-wide association; meta-analysis

Schizophrenia is an idiopathic mental disorder with substantial morbidity, mortality, and personal and societal costs. [1-3] An important genetic component is indicated by a sibling recurrence risk ratio of 8.6, 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 prior genomic findings. [4-8]

Correspond with Dr Sullivan: Department of Genetics, CB#7264, 5097 Genomic Medicine, University of North Carolina, Chapel Hill, NC, 27599-7264, USA. Voice: +919-966-3358, FAX: +919-966-3630, pfsulliv@med.unc.edu.

†* denote equal contributions by these pairs of authors.

URLs: Results can be downloaded from the Psychiatric Genomics Consortium website (http://pgc.unc.edu) and visualized using Ricopili (http://www.broadinstitute.org/mpg/ricopili). Genotype data are available upon application from the NIMH Genetics Repository (https://www.nimhgenetics.org).

Author Contributions: SR, COD, EAS, MF, NRW, NS, SB, SHL, ABS, ALR, BKBS, BMN, CdL, DP, DR, FB, JP, KL, MLH, MV, PH, SS, SM, SP, and PFS conducted statistical analyses. ADB, DMH, DR, ES, JS, MPM, ND, OM, PBM, ST, TS, and VG ascertained subjects. ALC, JJC, SW, YK, KX, and PFS performed bioinformatic analyses. KC, JLM, and SA managed the project. BPR, DWM, FAON, HS, JTW, KSK, MG, MJO, NC, PC, MGS, PEC, WTCCC2, APC, EB, KS, and MCOD provided replication samples and genotypes. AKK interfaced with Swedish national registers. The manuscript was written by PKEM, SM, SP, PS, CMH, and PFS. The study was designed by SP, PS, CMH, and PFS. Funding was obtained by ES, PS, CMH, and PFS.

Conflicts of Interest: Dr Sullivan was on the SAB of Expression Analysis (Durham, NC, USA). Dr Sklar is on the Board of Directors of Catalytic, Inc. The other authors report no conflicts.

Although the rationale for genomic searches is strong, there are only a handful of robust empirical findings for schizophrenia. Genome-wide linkage studies to date have been inconclusive, [9] and no compelling Mendelian variants have been identified.[8] Eight rare copy number variants of strong effect (genotypic relative risks 4-20)with consistent replication have been described (e.g., 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. Prior GWAS for common variation have yielded statistical evidence for ∼10 genomic regions[8] including the major histocompatibility complex (MHC) [14-16] along with *MIR137* and targets of miR-137. [17]

The prior studies contained indications that more common variant associations were likely to be discovered with larger sample sizes. [13,17,18] We therefore sought to increase substantially the number of cases using a multistage GWAS.

Results from Sweden

We analyzed genome-wide 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%) have never been previously reported. Following genotyping and imputation with the 1000 Genomes Project Phase 1 reference panel, the genetic data consisted of allelic dosages for 9,871,789 high-quality polymorphic SNPs. Given that this imputation panel is based on >800 chromosomes 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 $P < 1\times10^{-5}$ using 1000 Genomes imputation compared with 1,594 SNPs and 133 regions for HapMap3 imputation (counts include only one region from the MHC).

The resulting λ_{GC} was 1.075 and λ_{1000} (references [19-21]) was 1.013. Quantile-quantile and Manhattan plots are given in Supplemental Figures 5-6. For association with schizophrenia, 312 SNPs met a genome-wide significance threshold of 5×10^{-8} (reference [22]). These SNPs were in two genomic regions (Supplemental Figure 7): 241 SNPs in the MHC region (chr6:28,502,794-32,536,501, minimum $P=4.07\times10^{-11}$ at rs115939516) and 71 SNPs from chr2:200,715,388-201,040,981 (minimum $P=3.33\times10^{-10}$ at rs35220450). We replicated the MHC association reported in prior studies. [14-17] The chr2 association with schizophrenia is novel, shows highly consistent effects in the Sw1-6 genotyping batches and encompasses *C2orf69*, *C2orf47*, *C2orf60*, and *TYW5*.

Sweden + PGC

We re-analyzed the PGC schizophrenia data using 1000 Genomes imputation (8,832 cases and 12,067 controls, excluding Swedish samples). [17] Five regions met genome-wide significance: the MHC locus (chr6:27,261,324-32,610,445, minimum $P=2.18\times10^{-10}$), AS3MT-CNNM2-NT5C2 (chr10:104,635,103-104,960,464, minimum $P=4.29\times10^{-10}$), MAD1L1 (chr7:2,005,747-2,098,238, minimum $P=2.40\times10^{-8}$), RP11-586K2.1,

(chr8:89,585,639-89,760,620, minimum $P=2.37\times10^{-8}$), and SNPs nearTCF4 (chr18:53,311,001-53,423,307, minimum $P=3.00\times10^{-8}$).

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 0.15-0.85 (power=0.8, $\alpha=5\times10^{-8}$, 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 $P < 0.0001$ and in approximate linkage equilibrium, 62.6% had logistic regression beta coefficients with the same sign in the Swedish results, an observation highly inconsistent with the null ($P=2.2\times10^{-10}$). λ_{GC} was 1.186 and λ_{1000} was 1.012, values consistent with a polygenic pattern of association but not gross inflation due to technical artifacts. [20] Quantile-quantile and Manhattan plots are shown in Supplemental Figure 11 and Figure 1, 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 130K SNPs derived from the PGC to predict case-control status in the Swedish samples. These SNPs were selected for high-confidence and approximate linkage equilibrium but without regard to association *P* value. As shown in Figure 2, PGC risk scores had a highly significant capacity to predict case-control status in the independent Swedish samples (*P* values from 10^{-26} – 10^{-114}). The increased sample size allowed improved risk profile prediction as more of the SNPs in the lower bins are replicable signals. The threshold at which the explanatory power of these risk profile SNPs plateaus has decreased with increasing sample size: $P_T=0.1$ in Figure 2, 0.2 in the PGC report, and no plateau in the ISC study). [14,17] Although the mean risk profiles were highly significantly different between cases and controls, the distributions overlap substantially (Supplemental Figure 9) and are insufficient for diagnostic purposes (area under the receiver operating characteristic curve 0.65). However, these results strongly support the comparability of the Swedish and PGC samples and the validity of the meta-analysis.

GWAS often omit the X chromosome (chrX). This omission is problematic as chrX is approximately as large as chromosome 8 and is enriched for genes important in brain development. Using a previously described approach, we imputed genotyped chrX SNPs to the 1000 Genomes reference panel. [23] Joint analysis of all subjects as well as males and females separately revealed no association exceeding genome-wide significance. The strongest association (rs12845396, chrX:6,029,533, $P=3.46\times10^{-7}$) was in an intron of *NLGN4X* (neuroligin 4), a gene previously implicated in mental retardation and autism, and there were multiple possible signals nearMECP2(causal to Rett syndrome, $P=9.3\times10^{-6}$).

GWAS results generally do not lie in protein coding regions. [24] A recent report suggested that most SNPs in the NHGRI GWAS catalog [24] were in or in perfect LD with DNase 1 hypersensitive sites. [25] We thus evaluated whether the Sweden + PGC results had significant overlap with DNase 1 hypersensitive sites generated as part of the ENCODE project. [26] We did not find evidence of enrichment (Supplemental Table 8 and Supplemental Figure 10).

However, this negative result is strongly qualified by the lack of DNase 1 hypersensitivity data directly relevant to psychiatric disorders.

Sweden + PGC + replication

We then obtained association results for SNPs in 194 genomic regions in six independent samples for a total sample size of over 21,000 cases and 38,000 controls(Table 1). The genomic regions for which replication genotypes were sought were identified using LD clumping defined by LD (r^2> 0.5) and a minimum $P < 1\times10^{-5}$ in the Sweden-PGC meta-analysis. Only one MHC SNP was included. The Sweden-PGC meta-analysis and replication results were highly concordant with 76.3% of the logistic regression beta coefficients having the same direction of effect (sign test $P=1.5\times10^{-17}$). Indeed, of the top 100 SNPs in the Sweden-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.

Table 2 shows the combined results in which 24 regions reached genome-wide significance. As two pairs of these regions overlap (chr1:243Mb and chr5:152Mb), there are associations with schizophrenia in 22 genomic regions. Three additional regions nearly met genome-wide significance (rs4380187 near *ZNF804A* $P=5.66\times10^{-8}$, rs4523957 in *SRR* $P=5.69\times10^{-8}$, and rs6550435 near *TRANK1* $P=5.86\times10^{-8}$ which also had $P=9\times10^{-6}$ in a bipolar disorder GWAS). [27]

Of these 22 regions (Table 3), five regions have been reported previously as meeting genome-wide significance for schizophrenia (*MHC*, *C10orf26*, *DPYD-MIR137*, *SDCCAG8*, and *MMP16*) and two for schizophrenia, bipolar disorder, and a combined phenotype (*CACNA1C* and *ITIH3-ITIH4*). [14-17,27-29] For the remaining 15 regions, we now find genome-wide significance for a locus previously implicated only for bipolar disorder (*NCAN*)[30] along with 14 novel regions.

Themes

We highlight four themes from these results (see also Supplemental Table 9). First, these results implicate calcium signaling in the etiology of schizophrenia. As in prior studies of bipolar disorder and schizophrenia, [17,27,28] we found genome-wide significant support for *CACNA1C* ($Ca_v1.2$, $P=5.2\times10^{-12}$ at the intronic SNP rs1006737). Intriguingly, we identified a novel genome-wide significant association for *CACNB2* ($P=1.3\times10^{-10}$ at the intronic SNP rs17691888) which encodes the β_2 subunit of L-type calcium channels ($Ca_v\,\beta2$). A gene-set test supported the involvement of calcium channel subunits in the etiology of schizophrenia (Supplemental Table 7).

In L-type calcium channels, the α_{1c} subunit forms the transmembrane pore, and directly interacts with the intracellular β_2 subunit. [31] The β_2 subunit also antagonizes an endoplasmic reticulum retention motif on the α_{1c} subunit to facilitate transport to the plasma membrane. [32]Additional genes with genome-wide significant evidence were implicated based on membership in a proteomic network centered on Ca_v2 (reference [33]): the protein products of *ACTR1A* (α-centractin), the divalent metal cation transporter *CNNM2*

(P=3.7×10^{-13}, chr10:103,009,986-105,512,924), and *CACNB2*. A broad genomic region containing the calcium binding protein troponin C (*TNNC1*) also met genome-wide significance (P=1.1×10^{-8}) as well as three calcium homeostasis modulator genes (*CALHM1, CALHM2*,and *CALHM3* in same chr10 region 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 knock-out mice show notable deficits in long term potentiation. [34-37] Calcium "channelopathies" include mutations in *CACNA1C* and *CACNB2* that cause Brugada syndrome types 3 and 4 (OMIM #611875 and #611876). [38]In addition, Timothy syndrome (OMIM #601005), caused by mutations in *CACNA1C*, is a multisystem disorder including cognitive impairment and autism spectrum disorder. [39] Although Mendelian disorders are usually characterized by persistent pathological features, Mendelian calcium channelopathies can have episodic phenomena perhaps reminiscent of the episodic nature of psychotic disorders – for example, intermittent hypoglycemia and hypocalcemia in Timothy syndrome (*CACNA1C*), episodic ataxia (*CACNA1A, CACNB4*), migraine (*CACNA1A*), epilepsy (*CACNA1H, CACNB4*), periodic paralysis (*CACNA1S*), and malignant hyperthermia (*CACNA1S, CACNA2D1*). [31,39]

GWAS findings for schizophrenia have converged on genome-wide significant evidence for a calcium channel functional complex that has also been implicated in bipolar disorder and autism. These genomic results support increased attention to this pathway, and suggest hypotheses for clinical translation. Multiple approved medications act at calcium channels including some antipsychotics (e.g., pimozide) along with adjuvants for treatment non-response for schizophrenia and bipolar disorder (e.g., the calcium channel blockers verapamil and nifedipine). It is possible that drugs that act on the protein products of *CACNA1C* and *CACNB2* for a different therapeutic indication could be "re-purposed" 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 that many approved antipsychotics increase the cardiac QT interval, genetic variation in calcium channel genes might identify individuals at higher risk of sudden cardiac death. [40,41]

Second, as reported previously, [14-17] the strongest association P=9.1×10^{-14})with schizophrenia is in the extended MHC (chr6:25-34Mb), a region of both exceptional importance and complexity. The MHC comprises 0.3% of the genome but contains 1.5% of the genes in OMIM [42] 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 but will also complicate efforts to identify causal variation. Genome-wide significant associations with schizophrenia extend over 7Mb, but Supplemental Figure 12 suggests that larger samples may resolve this association into sub-regions near TRIM26 (tripartite motif containing 26, chr6:30.1Mb) and the HLA-DRB9 unprocessed pseudogene (chr6:32.4Mb, intergenic *HLA-DRA – HLA-DRB5*).

Third, multiple genomic lines of evidence support a role for *MIR137* in the etiology of schizophrenia. We provide increased support for a common variant association located upstream of the *MIR137* transcript ($P=1.7\times10^{-12}$, Supplemental Figure 13). Fourteen genes in the regions in Table 3 have miR-137 target sites predicted by TargetScan (v6.2) [43] (*C6orf47, HLA-DQA1, TNXB, VARS, C10orf26, 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 afar larger set of genes with predicted miR-137 target sites (Supplemental Table 7). In addition, our unpublished work shows enrichment for smaller GWAS *P*-values in genes down-regulated following over-expression of miR-137 in human neural stem cells (Collins, in preparation). Given the role of miR-137 in fundamental neuronal processes, [44-46] these results support investigation of pathways influenced by miR-137in regard to a role in the pathogenesis of schizophrenia.

The SNP with the strongest association to schizophrenia (rs1198588) is 39kb 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* (169kb upstream of rs1198588),and rs1198588 is a significant local expression quantitative trait locus (eQTL) with *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 in Table 3 contain long intergenic non-coding RNAs (lincRNAs). lincRNAs have multiple known or suspected functions including epigenetic regulation and development. [47] Using pathway analysis,[48] there was modest enrichment ($P=0.06$) for smaller association *P* values in a conservative set of lincRNAs derived from sequencing of poly-A RNA from multiple tissues. [47] This observation is consistent with a general role for GWAS findings in the regulation of gene expression rather than alteration of protein sequence. eQTLs [49,50] overlap with SNPs implicated by GWAS over all traits [51-53] as well as for specific traits like height, adiposity, cardiovascular risk factors, chemotherapy-induced cytotoxicity, autism, schizophrenia, and Crohn's disease. [54-61]An estimated 55% of eQTL SNPs lie in DNase I hypersensitivity sites (a marker for open chromatin subject to transcriptional regulation) and 77% of SNPs implicated in GWAS are in or in high LD with SNPs inDNase I hypersensitivity sites. [25,62,63]

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. [64] Traditional genetic epidemiological studies use the phenotypic resemblance of relatives to estimate the proportion of variance in liability using theoretic resemblance assumptions. GCTA uses genome-wide SNP genotypes to calculate the heritability in the population from the 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 (SE 0.01) using HapMap3 imputation and assuming a population risk of 0.01. [7] Using

the same imputation reference and population risk, SNP heritability was substantially higher in the Swedish samples (0.32, SE 0.03) possibly due to the greater phenotypic and genetic homogeneity in the Swedish sample compared to the PGC samples of mixed European ancestry. We obtained a similar estimate of SNP heritability using 1000 Genomes imputed data (0.33, SE 0.03, population risk 0.01). For a population risk of 0.004, [4,65] SNP heritability was 0.26 (SE 0.02) using HapMap3 and 0.27 (SE 0.02) using 1000 Genomes imputation. Partitioning of the SNP-heritability by minor allele frequency is consistent with 80% of the signal reflecting causal variants with MAF > 0.1 (Supplemental Table 5).

To complement the GCTA analyses, we also applied ABPA (approximate Bayesian polygenic analysis) [66] to the Sweden + PGC results. Compared to GCTA, ABPA yielded somewhat larger but generally congruent estimates of variance in liability to schizophrenia using HapMap3 data: 0.43 for population risk of 0.01 (95% credible interval 0.38-0.48) and 0.34 for population risk of 0.004 (95% credible interval 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 + PGC results reflect only partly the underlying genetic architecture of schizophrenia due to inadequate sample size. Using 1000 Genomes results for Sweden + PGC and assuming 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 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 (not shown) indicate that most of the signal was derived from SNPs with allele frequencies > 0.1; low-frequency imputed SNPs were not generally inferred to be associated with schizophrenia. Figure 3 compares ABPA estimates of the genetic architecture of schizophrenia and four biomedical diseases. [66] 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 left-shifted with greater numbers of SNPs with smaller effects.

We previously estimated the heritability of schizophrenia in Sweden to be 0.64 (95% CI 0.617-0.675) using a national pedigree sample of 9.0M individuals,[5] and a Danish national pedigree study of 2.6M individuals reported a similar estimate (0.67, 95% CI 0.65-0.71). [5,67]Using the 1000 Genomes data with population risk of 0.01, the variance in liability estimate from GCTA accounts for 52% of the heritability (0.33/0.64) and ABPA accounts for 78% of the heritability (0.50/0.64). Imprecision is inherent to these estimates and future work or the use of a twin meta-analytic estimate of the heritability of schizophrenia (0.81, 95% CI 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.

Conclusions

These results provide deeper insight into the genetic architecture of schizophrenia than ever before. We find support for 22 common variant loci (14 novel) 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. [8,18,68,69]

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, that the number of 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 appear 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 models containing a handful of SNPs is unlikely to account for more than a small fraction of the heritability. [70] Ourresults 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 utility for study planning. We used the ABPA estimates of the posterior distribution of genotypic relative risks (Figure 3) to inform power calculations by estimating the numbers of independent loci that could be detected for different sample sizes (Supplemental Table 6 and Supplemental Figure 8). 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 362-1154 SNPs).

Thus, for the first time, we now have a clear path to increased knowledge about the etiology of schizophrenia via 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: identification of the top 2,000 loci (for example) might be sufficient confidently and clearly to reveal 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 (e.g., psychotic prodrome), and appropriate cellular measure for 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.

Online Methods

Overview

We present here the pre-planned principal analyses for this project. In order to advance knowledge of schizophrenia, a minority of samples were included in prior reports. Genotyping was conducted in six batches (denoted Sw1-Sw6) with total sample sizes of 464, 694, 1498, 2388, 4461, and 2345. Genotypes were generated as sufficient numbers of samples accumulated from the field work in Sweden. The 2009 International Schizophrenia Consortium report contained GWAS data from the Sw1-2 subjects (N=1158, 9.8% of the sample before quality control). [14]. The 2011 PGC schizophrenia paper also contained GWAS data from the Sw1-2 subjects plus ∼80 SNPs from Sw3-4 in the replication phase. [17] The 2012 Bergen et al. paper had a particular focus contrasting schizophrenia with bipolar disorder and reported GWAS results from Sw1-4 (N=4044, 42.6% of the full sample). [75] Thus, of the total sample of 11,850 Swedish subjects before quality control (5,351 cases, 6,509 controls), 57.4% have never been reported previously.

Subjects

All procedures were approved by ethical committees at the Karolinska Institutet and University of North Carolina, and all subjects provided written informed consent (or legal guardian consent and subject assent). Sample collection was from 2005-11.

Cases with schizophrenia were identified via the Swedish Hospital Discharge Register [76,77] which captures all public and private inpatient hospitalizations. The register is complete from 1987 and augmented by psychiatric data from 1973-86. The register contains ICD discharge diagnoses [78-80] made by attending physicians for each hospitalization. [81-84] Case inclusion criteria: ≥2 hospitalizations with a discharge diagnosis of schizophrenia, both parents born in Scandinavia, and age ≥18 years. Case exclusion criteria: hospital register diagnosis of any medical or psychiatric disorder mitigating a confident diagnosis of schizophrenia as determined by expert review, and included removal of 3.4% of eligible cases due to the primacy of another psychiatric disorder (0.9%) or a general medical condition (0.3%) or uncertainties in the Hospital Discharge Register (e.g., contiguous admissions with brief total duration, 2.2%).

The validity of this case definition of schizophrenia is described at length in the Supplement, and 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 “super-normal” controls. [85] Control inclusion criteria: never hospitalized for schizophrenia or bipolar disorder (given evidence of genetic overlap with schizophrenia), [5,14,86] both parents born in Scandinavia, and age ≥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), [87,88] and in a recent large Norwegian longitudinal study (58%). [89] For cases, comorbidity with drug/alcohol abuse or dependence did not predict participation nor did any subtype of schizophrenia (e.g., paranoid or disorganized types). The sample was approximately representative of the Swedish populace in regard to county of birth (Supplemental Figure 4).

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 (Supplemental Table 3). Genotype calling, quality control, and imputation were done in four sets corresponding to data from Affymetrix 5.0 (Sw1), Affymetrix 6.0 (Sw2-4), and the OmniExpress batches (Sw5, Sw6). Genotypes were called using Birdsuite (Affymetrix) or BeadStudio (Illumina). The quality control parameters applied were: SNP missingness < 0.05 (before sample removal); subject missingness < 0.02;autosomal heterozygosity deviation; SNP missingness < 0.02 (after sample removal);difference in SNP missingness between cases and controls < 0.02; and deviation from Hardy-Weinberg equilibrium ($P < 10^{-6}$ in controls or $P < 10^{-10}$ in cases).

After basic quality control, 77,986 autosomal SNPs directly genotyped on all four GWAS platforms were extracted and pruned to remove SNPs in LD ($r^2 > 0.05$) or with minor allele frequency < 0.05, leaving 39,239 SNPs suitable for robust relatedness testing and population structure analysis. Relatedness testing was done with PLINK[90] and pairs of subjects with π(x00302) > 0.2 were identified and one member of each relative pair removed at random. Principal component estimation was done 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 [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 pre-phasing/imputation stepwise approach implemented in IMPUTE2 / SHAPEIT (chunk size of 3 Mb and default parameters). [91,92] The imputation reference set consisted of 2,186 phased haplotypes from the full 1000Genomes Project dataset (March 2012, 40,318,245 variants). Evaluation of λ_{GC} led to the removal of SNPs with control allele frequencies < 0.005 or > 0.995, imputation “info” values < 0.2, or that were genotyped only in the smallest sample set (Sw1). Given that male sex is a risk factor for schizophrenia, [93] chromosome X imputation was conducted for subjects passing QC for the autosomal analysis (excluding chrX SNPs with missingness ≥ 0.05 or HWE $P < 10^{-6}$ in females). Imputation was performed separately for males and females, gene dosages 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 Sw1-2). [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 significant SNPs. 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 highest association and other SNPs in high linkage disequilibrium with the index SNP. To summarize GWAS findings, we used the following settings in PLINK:

```
--clump-p1 1e-4 --clump-p2 1e-4 --clump-r2 0.2 --clump-kb 500
```

to retain SNPs with association $P < 0.0001$ and $r^2 < 0.2$ within 500 kb windows.

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. Due to the strong signal and high linkage disequilibrium in the MHC, only one SNP was kept from the extended MHC region. We then determined the number of SNPs whose logistic regression beta 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 done in two ways: selecting SNPs from Sw1-6 results and evaluating the signs in the independent PGC results, and by reversing the procedure (select from PGC, evaluate signs in Sw1-6). Similar results were obtained selecting SNPs for: (a) $P < 1\times10^{-5}$, (b) $P < 1\times10^{-6}$, (c) keeping one SNP every 3 Mb (effectively removing or greatly minimizing the effects of residual linkage disequilibrium).

Risk score profiles (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 a high-quality, relatively independent SNPs with unambiguous directions of effects: from the PGC imputed results file, we made a subset of results containing SNPs with allele frequency 0.02-0.98 and

imputation INFO scores > 0.9. We then removed SNPs in high LD using via clumping (i.e., retain all SNPs with $r^2 < 0.25$ 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, due to the strong signal and high linkage disequilibrium in the MHC, only one SNP was kept from the extended MHC region.

We used the resulting list from the PGC to calculate schizophrenia risk profile scores in the independent Swedish samples using the

```
--score
```

function in PLINK. We did this 10 times using different subsets of the PGC SNPs selected by increasing P value thresholds. From the set of filtered SNPs from the PGC, we evaluated 10 different association *P* 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., include all SNPs). For each of these 10 sets of SNPs derived from the PGC, the schizophrenia risk profile score (the number of schizophrenia risk alleles weighted by the logistic regression beta) was calculated for each case and control in Sw1-6. 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 risk profile score from the PGC, we used the difference in the Nagelkerke pseudo R^2 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 if 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 based on 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, http://ctglab.nl/software) to conduct gene-set analyses. This method has previously been applied to the International Schizophrenia Consortium data by Lips et al. [94] JAG tests for the association of specified gene-sets with schizophrenia as applied to individual-level genotype data which tends 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 the some of the properties of the SNPs being studied (e.g., 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 is of the null hypothesis is that these genes are not more strongly associated than 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 like 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 (Supplemental Table 4). These subjects are not included in the Swedish samples or in the PGC mega-analysis. [17] The independent samples were from SGENE+, [16], CLOZUK, [29] the Irish Schizophrenia Genomics Consortium, [95] the Psychosis Endophenotype Consortium, [96], and the Multicenter Family Study. [97] After selecting for $P < 1\times10^{-5}$ in the Sweden and PGC meta-analysis and accounting for linkage disequilibrium, we requested association results for 194 genomic regions.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Authors

Stephan Ripke[1,†], Colm O'Dushlaine[1,†], Kimberly Chambert[1], Jennifer L Moran[1], Anna K Kähler[2], Susanne Akterin[2], Sarah Bergen[2], Ann L Collins[3], James J Crowley[3], Menachem Fromer[4], Yunjung Kim[3], Sang Hong Lee[5], Patrik KE Magnusson[2], Nick Sanchez[1], Eli A Stahl[4], Stephanie Williams[3], Naomi R Wray[5], Kai Xia[3], Francesco Bettella[6], Anders D Børglum[7], Brendan K Bulik-Sullivan[1], Paul Cormican[8], Nick Craddock[9], Christiaan de Leeuw[10], Naser Durmishi[11], Michael Gill[8], Vera Golimbet[12], Marian L Hamshere[9], Peter Holmans[9], David M Hougaard[13], Kenneth S Kendler[14], Kuang Lin[15], Derek W Morris[8], Ole Mors[16], Preben B Mortensen[7], Benjamin M Neale[1], Francis A O'Neill[17], Michael J Owen[9], MilicaPejovic Milovancevic[18], Danielle Posthuma[10], John Powell[15], Alexander L Richards[9], Brien P Riley[14], Douglas Ruderfer[4], Dan Rujescu[19], Engilbert Sigurdsson[20], Teimuraz Silagadze[21], August B Smit[10], Hreinn Stefansson[6], Stacy Steinberg[6], Jaana Suvisaari[22], Sarah Tosato[23], Matthijs Verhage[10], James T Walters[9], Multicenter Genetic Studies of Schizophrenia Consortium[24], Psychosis Endophenotypes Consortium, Wellcome Trust Case-Control Consortium[2], Elvira Bramon[25], Aiden P Corvin[8], Michael C O'Donovan[9], Kari Stefansson[6], Edward Scolnick[1], Shaun Purcell[4], Steve McCarroll[1], Pamela Sklar[4], Christina M Hultman[2,*], and Patrick F Sullivan[2,3,*]

Affiliations

[1]Broad Institute of MIT and Harvard, 7 Cambridge Center, Cambridge, MA, USA [2]KarolinskaInstitutet, Stockholm, Sweden [3]University of North Carolina, CB# 7264, Chapel Hill, NC, USA [4]Mt Sinai School of Medicine, L3-43 Icahn, New York, NY, USA [5]University of Queensland, Brisbane, Australia [6]deCODE Genetics, Sturlugata 8, Iceland [7]Aarhus University, Aarhus, Denmark [8]Trinity College Dublin, Dublin, Ireland [9]Cardiff University School of Medicine, Heath Park, Cardiff, UK [10]VU University, Amsterdam, NL [11]University of Skopje, Skopje, Macedonia [12]Russian Academy of Medical Sciences, Moscow, Russia [13]Statens Serum Institut, Copenhagen, Denmark [14]Virginia Commonwealth University, Richmond, VA, USA [15]King's College London, London, UK [16]Aarhus University Hospital, Risskov, Denmark [17]Queen's University Belfast, Belfast, Northern Ireland [18]University of Belgrade, Belgrade, Serbia [19]University of Halle-Wittenberg, Halle, Germany [20]University of Iceland, Reykjavik, Iceland [21]Tbilisi State Medical University, Tbilisi, Georgia [22]National Institute for Health and Welfare THL, Helsinki, Finland [23]University of Verona, Verona, Italy [24]See collaborator list [25]University College London, London, UK

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: Emma Flordal-Thelander, Ann-Britt Holmgren, Marie Hallin, Marie Lundin, Ann-Kristin Sundberg, Christina Pettersson, Radja Satgunanthan-Dawoud, Sonja Hassellund, MalinRådstrom, BirgittaOhlander, Leila Nyrén, and Isabelle Kizling. Funding support was provided by the NIMH R01 MH077139 (Sullivan), NIMH R01 MH095034 (Sklar),Stanley Center for Psychiatric Research, the Sylvan Herman Foundation, the Friedman Brain Institute at Mount Sinai School of Medicine, the Karolinska Institutet, Karolinska University Hospital, the Swedish Research Council, the Swedish County Council, the SöderströmKönigska Foundation, and the Netherlands Scientific Organization (NWO 645-000-003).SGENE was supported by EU 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, 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, analysis, and manuscript preparation.

Appendix

COLLABORATOR LIST - Multicenter Genetic Studies of Schizophrenia Consortium

Prof Douglas F Levinson MD	dflev@stanford.edu	Psychiatry and Behavioral Sciences, Stanford University, Stanford, California, USA
Prof Pablo V Gejman MD	pgejman@gmail.com	Psychiatry and Behavioral Sciences, NorthShore University HealthSystem and University of Chicago, Evanston, Illinois, USA
Dr Claudine Laurent MD PhD	claudinelaurent54@yahoo.fr	Child and Adolescent Psychiatry, Pierre and Marie Curie Faculty of Medicine and Brain and Spinal Cord Institute (ICM), Paris, France
Prof Bryan J Mowry MD FRANZCP	b.mowry@uq.edu.au	Psychiatry, Queensland Brain Institute and Queensland Centre for Mental Health

		Research, University of Queensland; Brisbane, Queensland, Australia
Prof Ann E Pulver PhD	aepulver@jhmi.edu	Psychiatry, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA
Prof Sibylle G Schwab PhD	sibylle.schwab@uk-erlangen.de	Psychiatry, Friedrich-Alexander University, Erlangen-Nuremberg, Erlangen, Germany
Prof Dieter B Wildenauer PhD	dieter.wildenauer@uwa.edu.au	Psychiatry and Clinical Neurosciences, Western Australian Institute for Medical Research & Centre for Medical Research, The University of Western Australia, Nedlands, Australia
Dr Frank Dudbridge PhD	frank.dudbridge@lshtm.ac.uk	Non-Communicable Disease Epidemiology, London School of Hygiene and Tropical Medicine, London, UK
Dr Jianxin Shi PhD	jianxins@mail.nih.gov	Biostatistics, National Cancer Institute, Bethesda, MD, USA
Prof Margot Albus MD	margot.albus@iak-kmo.de	State Mental Hospital, Haar, Germany
Dr Madeline Alexander PhD	mmga1@comcast.net	Psychiatry and Behavioral Sciences, Stanford University, Stanford, California, USA
Prof Dominique Campion PhD	dominique.campion@univ-rouen.fr	INSERM U614, University of Medicine, Rouen, France
Prof David Cohen MD PhD	dcohen55@noos.fr	Child and Adolescent Psychiatry, Pierre and Marie Curie Faculty of Medicine, Institute for Intelligent Systems and Robotics (ISIR), Paris, France
Prof DimitrisDikeos MD	ddikeos@med.uoa.gr	First Department of Psychiatry, University of Athens Medical School, Athens, Greece
Dr JubaoDuan PhD	jduan69@gmail.com	Psychiatry and Behavioral Sciences, NorthShore University HealthSystem and University of Chicago; Evanston, Illinois, USA
Prof Peter Eichhammer MD PhD	Peter.Eichhammer@medbo.de	Psychiatry, University of Regensburg, Regensburg, Germany
Stephanie Godard	stephanie.godard-bauche@upmc.fr	Psychiatry and Genetics, INSERM, Institut de Myologie, Hôpital Pitié Salpêtrière, Paris, France
Dr Mark Hansen PhD	mhansen@illumina.com	Illumina, Inc., La Jolla, California, USA
Prof F Bernard Lerer MD	lerer@cc.huji.ac.il	Psychiatry, Hadassah-Hebrew University Medical Center, Jerusalem, Israel
Prof Kung-Yee Liang PhD	kyliang@jhsph.edu	Biostatistics, Johns Hopkins University, Baltimore, Maryland, USA
Prof Wolfgang Maier MD	wolfgang.maier@ukb.uni-bonn.de	Psychiatry, University of Bonn, Bonn, Germany
Prof Jacques Mallet PhD	mallet@chups.jussieu.fr	Centre National de la Recherche Scientifique, Laboratoire de Génétique Moléculaire de la Neurotransmission et des Processus Neurodégénératifs, Hôpital Pitié Salpêtrière, Paris, France
Deborah A Nertney	deb_nertney@qcmhr.uq.edu.au	Psychiatry, Queensland Brain Institute and Queensland Centre for Mental Health Research, University of Queensland, Brisbane, Queensland, Australia
Prof Gerald Nestadt MD	gnestadt@jhmi.edu	Psychiatry and Behavioral Sciences, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA

Dr Nadine Norton PhD	nortonn@cardiff.ac.uk	Psychological Medicine and Neurology, MRC Centre for Neuropsychiatric Genetics and Genomics, School of Medicine, Cardiff University, Cardiff, Wales, UK
Prof George N Papadimitriou MD	gnpapad@med.uoa.gr	First Department of Psychiatry, University of Athens Medical School, Athens, Greece
Robert Ribble	rcribble@vcu.edu	Psychiatry, VIPBG, VCU, Richmond, Virginia, USA
Dr Alan R Sanders MD	alan.sanders.md@gmail.com	Psychiatry, NorthShore University HealthSystem and University of Chicago, Evanston, Illinois, USA
Prof Jeremy M Silverman PhD	jeremy.silverman@mssm.edu	Psychiatry, Mount Sinai School of Medicine, New York, NY and VAMC,Bronx, New York, USA
Prof Dermot Walsh MD	dwalsh@hrb.ie	The Health Research Board, Dublin, Ireland
Dr Nigel M Williams PhD	williamsnm@cf.ac.uk	Psychological Medicine, MRC Centre for Neuropsychiatric Genetics and Genomics, School of Medicine, Cardiff University, Cardiff, Wales, UK
Brandon Wormley	bwormley@hsc.vcu.edu	Psychiatry, VIPBG, VCU, Richmond, Virginia, USA

References

1. Saha S, Chant D, McGrath J. A systematic review of mortality in schizophrenia: is the differential mortality gap worsening over time? Archives of general psychiatry. 2007; 64:1123–31. [PubMed: 17909124]
2. World Health Organization. The Global Burden of Disease: 2004 Update. WHO Press; Geneva: 2008.
3. Knapp M, Mangalore R, Simon J. The global costs of schizophrenia. Schizophrenia bulletin. 2004; 30:279–93. [PubMed: 15279046]
4. Lichtenstein P, et al. Recurrence risks for schizophrenia in a Swedish national cohort. Psychological Medicine. 2006; 36:1417–26. [PubMed: 16863597]
5. Lichtenstein P, et al. Common genetic influences for schizophrenia and bipolar disorder: A population-based study of 2 million nuclear families. Lancet. 2009; 373:234–9. [PubMed: 19150704]
6. Sullivan PF, Kendler KS, Neale MC. Schizophrenia as a complex trait: evidence from a meta-analysis of twin studies. Archives of General Psychiatry. 2003; 60:1187–92. [PubMed: 14662550]
7. Lee S, et al. Estimating the proportion of variation in susceptibility to schizophrenia captured by common SNPs. Nature Genetics. 2012; 44:247–50. [PubMed: 22344220]
8. Sullivan PF, Daly MJ, O'Donovan M. Genetic architectures of psychiatric disorders: the emerging picture and its implications. Nature Reviews Genetics. 2012; 13:537–51.
9. Ng MY, et al. Meta-analysis of 32 genome-wide linkage studies of schizophrenia. Mol Psychiatry. 2009; 14:774–85. [PubMed: 19349958]
10. Girard SL, et al. Increased exonic de novo mutation rate in individuals with schizophrenia. Nature genetics. 2011; 43:860–3. [PubMed: 21743468]
11. Xu B, et al. De novo gene mutations highlight patterns of genetic and neural complexity in schizophrenia. Nature genetics. 2012
12. Need AC, et al. Exome sequencing followed by large-scale genotyping suggests a limited role for moderately rare risk factors of strong effect in schizophrenia. American Journal of Human Genetics. 2012
13. Kim Y, Zerwas S, Trace SE, Sullivan PF. Schizophrenia genetics: where next? Schizophrenia Bulletin. 2011; 37:456–63. [PubMed: 21505112]

14. International Schizophrenia Consortium. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. Nature. 2009; 460:748–52. [PubMed: 19571811]
15. Shi J, et al. Common variants on chromosome 6p22.1 are associated with schizophrenia. Nature. 2009; 460:753–7. [PubMed: 19571809]
16. Stefansson H, et al. Common variants conferring risk of schizophrenia. Nature. 2009; 460:744–7. [PubMed: 19571808]
17. Schizophrenia Psychiatric Genome-Wide Association Study Consortium. Genome-wide association study identifies five new schizophrenia loci. Nature Genetics. 2011; 43:969–76. [PubMed: 21926974]
18. Wray NR, Visscher PM. Narrowing the boundaries of the genetic architecture of schizophrenia. Schizophr Bull. 2010; 36:14–23. [PubMed: 19996148]
19. Devlin B, Roeder K. Genomic control for association studies. Biometrics. 1999; 55:997–1004. [PubMed: 11315092]
20. Yang J, et al. Genomic inflation factors under polygenic inheritance. European journal of human genetics : EJHG. 2011; 19:807–12. [PubMed: 21407268]
21. de Bakker PI, et al. Practical aspects of imputation-driven meta-analysis of genome-wide association studies. Hum Mol Genet. 2008; 17:R122–8. [PubMed: 18852200]
22. Pe'er I, Yelensky R, Altshuler D, Daly MJ. Estimation of the multiple testing burden for genomewide association studies of nearly all common variants. Genet Epidemiol. 2008; 32:381–385. [PubMed: 18348202]
23. Major Depressive Disorder Working Group of the PGC. A mega-analysis of genome-wide association studies for major depressive disorder. Molecular Psychiatry. 2012
24. Hindorff LA, et al. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. Proc Natl Acad Sci U S A. 2009; 106:9362–7. [PubMed: 19474294]
25. Maurano MT, et al. Systematic localization of common disease-associated variation in regulatory DNA. Science. 2012; 237:1190–1195. [PubMed: 22955828]
26. Dunham I, et al. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012; 489:57–74. [PubMed: 22955616]
27. Psychiatric GWAS Consortium Bipolar Disorder Working Group. Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near ODZ4. Nature genetics. 2011; 43:977–83. [PubMed: 21926972]
28. Ferreira M, et al. Collaborative genome-wide association analysis of 10,596 individuals supports a role for Ankyrin-G (ANK3) and the alpha-1C subunit of the L-type voltage-gated calcium channel (CACNA1C) in bipolar disorder. Nature Genetics. 2008; 40:1056–8. [PubMed: 18711365]
29. Hamshere ML, et al. Genome-wide significant associations in schizophrenia to ITIH3/4, CACNA1C and SDCCAG8, and extensive replication of associations reported by the Schizophrenia PGC. Molecular psychiatry. 2012
30. Cichon S, et al. Genome-wide Association Study Identifies Genetic Variation in Neurocan as a Susceptibility Factor for Bipolar Disorder. Am J Hum Genet. 2011; 88:372–81. [PubMed: 21353194]
31. Bidaud I, Mezghrani A, Swayne LA, Monteil A, Lory P. Voltage-gated calcium channels in genetic diseases. Biochimica et biophysica acta. 2006; 1763:1169–74. [PubMed: 17034879]
32. Bichet D, et al. The I-II loop of the Ca2+ channel alpha1 subunit contains an endoplasmic reticulum retention signal antagonized by the beta subunit. Neuron. 2000; 25:177–90. [PubMed: 10707982]
33. Muller CS, et al. Quantitative proteomics of the Cav2 channel nano-environments in the mammalian brain. Proceedings of the National Academy of Sciences of the United States of America. 2010; 107:14950–7. [PubMed: 20668236]
34. Woodside BL, Borroni AM, Hammonds MD, Teyler TJ. NMDA receptors and voltage-dependent calcium channels mediate different aspects of acquisition and retention of a spatial memory task. Neurobiology of learning and memory. 2004; 81:105–14. [PubMed: 14990230]
35. Moosmang S, et al. Role of hippocampal Cav1.2 Ca2+ channels in NMDA receptor-independent synaptic plasticity and spatial memory. The Journal of neuroscience : the official journal of the Society for Neuroscience. 2005; 25:9883–92. [PubMed: 16251435]

36. White JA, et al. Conditional forebrain deletion of the L-type calcium channel Ca V 1.2 disrupts remote spatial memories in mice. Learning & memory. 2008; 15:1–5. [PubMed: 18174367]
37. Mangoni ME, et al. Voltage-dependent calcium channels and cardiac pacemaker activity: from ionic currents to genes. Progress in biophysics and molecular biology. 2006; 90:38–63. [PubMed: 15979127]
38. Perrin MJ, Gollob MH. Genetics of Cardiac Electrical Disease. The Canadian journal of cardiology. 2012
39. Splawski I, et al. Ca(V)1.2 calcium channel dysfunction causes a multisystem disorder including arrhythmia and autism. Cell. 2004; 119:19–31. [PubMed: 15454078]
40. Koponen H, et al. Schizophrenia and sudden cardiac death: a review. Nordic journal of psychiatry. 2008; 62:342–5. [PubMed: 18752109]
41. Stollberger C, Huber JO, Finsterer J. Antipsychotic drugs and QT prolongation. International clinical psychopharmacology. 2005; 20:243–51. [PubMed: 16096514]
42. McKusick VA. Mendelian Inheritance in Man and its online version, OMIM. Am J Hum Genet. 2007; 80:588–604. [PubMed: 17357067]
43. Lewis BP, Burge CB, Bartel DP. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. Cell. 2005; 120:15–20. [PubMed: 15652477]
44. Szulwach KE, et al. Cross talk between microRNA and epigenetic regulation in adult neurogenesis. J Cell Biol. 2010; 189:127–41. [PubMed: 20368621]
45. Smrt RD, et al. MicroRNA miR-137 regulates neuronal maturation by targeting ubiquitin ligase mind bomb-1. Stem Cells. 2010; 28:1060–70. [PubMed: 20506192]
46. Willemsen MH, et al. Chromosome 1p21.3 microdeletions comprising DPYD and MIR137 are associated with intellectual disability. Journal of medical genetics. 2011; 48:810–8. [PubMed: 22003227]
47. Cabili MN, et al. Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. Genes & development. 2011; 25:1915–27. [PubMed: 21890647]
48. Lee P, O'Dushlaine C, Thomas B, Purcell S. InRich: Interval-based enrichment analysis for genome-wide association studies. Bioinformatics. 2012; 28:1797–9. [PubMed: 22513993]
49. Majewski J, Pastinen T. The study of eQTL variations by RNA-seq: from SNPs to phenotypes. Trends in genetics : TIG. 2011; 27:72–9. [PubMed: 21122937]
50. Cookson W, Liang L, Abecasis G, Moffatt M, Lathrop M. Mapping complex disease traits with global gene expression. Nat Rev Genet. 2009; 10:184–94. [PubMed: 19223927]
51. Nicolae DL, et al. Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS. PLoS genetics. 2010; 6:e1000888. [PubMed: 20369019]
52. Stranger BE, et al. Patterns of cis regulatory variation in diverse human populations. PLoS genetics. 2012; 8:e1002639. [PubMed: 22532805]
53. Grundberg E, et al. Mapping cis- and trans-regulatory effects across multiple tissues in twins. Nature genetics. 2012
54. Lango Allen H, et al. Hundreds of variants clustered in genomic loci and biological pathways affect human height. Nature. 2010; 467:832–8. [PubMed: 20881960]
55. Emilsson V, et al. Genetics of gene expression and its effect on disease. Nature. 2008; 452:423–8. [PubMed: 18344981]
56. de Jong S, et al. Expression QTL analysis of top loci from GWAS meta-analysis highlights additional schizophrenia candidate genes. European journal of human genetics : EJHG. 2012; 20:1004–1008. [PubMed: 22433715]
57. 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. Human molecular genetics. 2010; 19:3482–8. [PubMed: 20601676]
58. Luo R, et al. Genome-wide Transcriptome Profiling Reveals the Functional Impact of Rare De Novo and Recurrent CNVs in Autism Spectrum Disorders. American journal of human genetics. 2012; 91:38–55. [PubMed: 22726847]

59. Speliotes EK, et al. Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. Nat Genet. 2010; 42:937–48. [PubMed: 20935630]
60. Zeller T, et al. Genetics and beyond--the transcriptome of human monocytes and disease susceptibility. PloS one. 2010; 5:e10693. [PubMed: 20502693]
61. Gamazon ER, Huang RS, Cox NJ, Dolan ME. Chemotherapeutic drug susceptibility associated SNPs are enriched in expression quantitative trait loci. Proceedings of the National Academy of Sciences of the United States of America. 2010; 107:9287–92. [PubMed: 20442332]
62. Thurman RE, et al. The accessible chromatin landscape of the human genome. Nature. 2012; 489:75–82. [PubMed: 22955617]
63. Degner JF, et al. DNase I sensitivity QTLs are a major determinant of human expression variation. Nature. 2012; 482:390–4. [PubMed: 22307276]
64. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. American journal of human genetics. 2011; 88:76–82. [PubMed: 21167468]
65. Saha S, Chant D, Welham J, McGrath J. A systematic review of the prevalence of schizophrenia. PLoS Medicine. 2005; 2:e141. [PubMed: 15916472]
66. Stahl EA, et al. Bayesian inference analyses of the polygenic architecture of rheumatoid arthritis. Nature Genetics. 2012; 44:483–9. [PubMed: 22446960]
67. Wray NR, Gottesman II. Using summary data from the danish national registers to estimate heritabilities for schizophrenia, bipolar disorder, and major depressive disorder. Frontiers in genetics. 2012; 3:118. [PubMed: 22783273]
68. Lander ES. Initial impact of the sequencing of the human genome. Nature. 2011; 470:187–97. [PubMed: 21307931]
69. Sullivan PF. Don't give up on GWAS. Molecular Psychiatry. 2011
70. Park JH, et al. Estimation of effect size distribution from genome-wide association studies and implications for future discoveries. Nature genetics. 2010; 42:570–5. [PubMed: 20562874]
71. Stahl EA, et al. Genome-wide association study meta-analysis identifies seven new rheumatoid arthritis risk loci. Nat Genet. 2010; 42:508–14. [PubMed: 20453842]
72. Betancur C. Etiological heterogeneity in autism spectrum disorders: more than 100 genetic and genomic disorders and still counting. Brain Res. 2011; 1380:42–77. [PubMed: 21129364]
73. Chiurazzi P, Schwartz CE, Gecz J, Neri G. XLMR genes: update 2007. European journal of human genetics : EJHG. 2008; 16:422–34. [PubMed: 18197188]
74. Inlow JK, Restifo LL. Molecular and comparative genetics of mental retardation. Genetics. 2004; 166:835–81. [PubMed: 15020472]
75. Bergen SE, et al. Genome-wide association study in a Swedish population yields support for greater CNV and MHC involvement in schizophrenia compared to bipolar disorder. Molecular Psychiatry. 2012; 17:880–6. [PubMed: 22688191]
76. Kristjansson E, Allebeck P, Wistedt B. Validity of the diagnosis of schizophrenia in a psychiatric inpatient register. Nordisk Psykiatrik Tidsskrift. 1987; 41:229–34.
77. Dalman C, Broms J, Cullberg J, Allebeck P. Young cases of schizophrenia identified in a national inpatient register--are the diagnoses valid? Social Psychiatry and Psychiatric Epidemiology. 2002; 37:527–31. [PubMed: 12395142]
78. World Health Organization. International Classification of Diseases. World Health Organization; Geneva: 1967.
79. World Health Organization. International Classification of Diseases. World Health Organization; Geneva: 1978.
80. World Health Organization. International Classification of Diseases. World Health Organization; Geneva: 1992.
81. Hultman CM, Sparen P, Takei N, Murray RM, Cnattingius S. Prenatal and perinatal risk factors for schizophrenia, affective psychosis, and reactive psychosis of early onset: case-control study. Bmj. 1999; 318:421–6. [PubMed: 9974454]
82. Zammit S, et al. Investigating the association between cigarette smoking and schizophrenia in a cohort study. Am J Psychiatry. 2003; 160:2216–21. [PubMed: 14638593]

83. Andersson RE, Olaison G, Tysk C, Ekbom A. Appendectomy and protection against ulcerative colitis. N Engl J Med. 2001; 344:808–14. [PubMed: 11248156]
84. Hansson LE, et al. The risk of stomach cancer in patients with gastric or duodenal ulcer disease. N Engl J Med. 1996; 335:242–9. [PubMed: 8657240]
85. Schwartz S, Susser E. Genome-wide association studies: does only size matter? Am J Psychiatry. 2010; 167:741–4. [PubMed: 20595425]
86. Craddock N, Owen MJ. The Kraepelinian dichotomy - going, going… but still not gone. Br J Psychiatry. 2010; 196:92–5. [PubMed: 20118450]
87. Hartge P. Participation in population studies. Epidemiology. 2006; 17:252–4. [PubMed: 16617271]
88. Morton LM, Cahill J, Hartge P. Reporting participation in epidemiologic studies: a survey of practice. Am J Epidemiol. 2006; 163:197–203. [PubMed: 16339049]
89. Bulik CM, 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. 2007; 37:1109–18. [PubMed: 17493296]
90. Purcell S, et al. PLINK: a toolset for whole-genome association and population-based linkage analysis. American Journal of Human Genetics. 2007; 81:559–75. [PubMed: 17701901]
91. Howie B, Marchini J, Stephens M. Genotype imputation with thousands of genomes. G3. 2011; 1:457–70. [PubMed: 22384356]
92. Delaneau O, Marchini J, Zagury JF. A linear complexity phasing method for thousands of genomes. Nature methods. 2012; 9:179–81. [PubMed: 22138821]
93. McGrath JJ. Variations in the incidence of schizophrenia: data versus dogma. Schizophrenia bulletin. 2006; 32:195–7. [PubMed: 16135560]
94. Lips ES, et al. Functional gene group analysis identifies synaptic gene groups as risk factor for schizophrenia. Molecular psychiatry. 2011; 17:996–106. [PubMed: 21931320]
95. 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. Biological Psychiatry. 2012
96. Psychosis Endophenotype Consortium & Wellcome Trust Case-Control Consortium A psychosis cohort for genome-wide analysis of psychiatric disorders. Biological Psychiatry. In press.
97. Levinson DF, et al. Genome-wide Association Study of Multiplex Schizophrenia Pedigrees. American Journal of Psychiatry. 2012

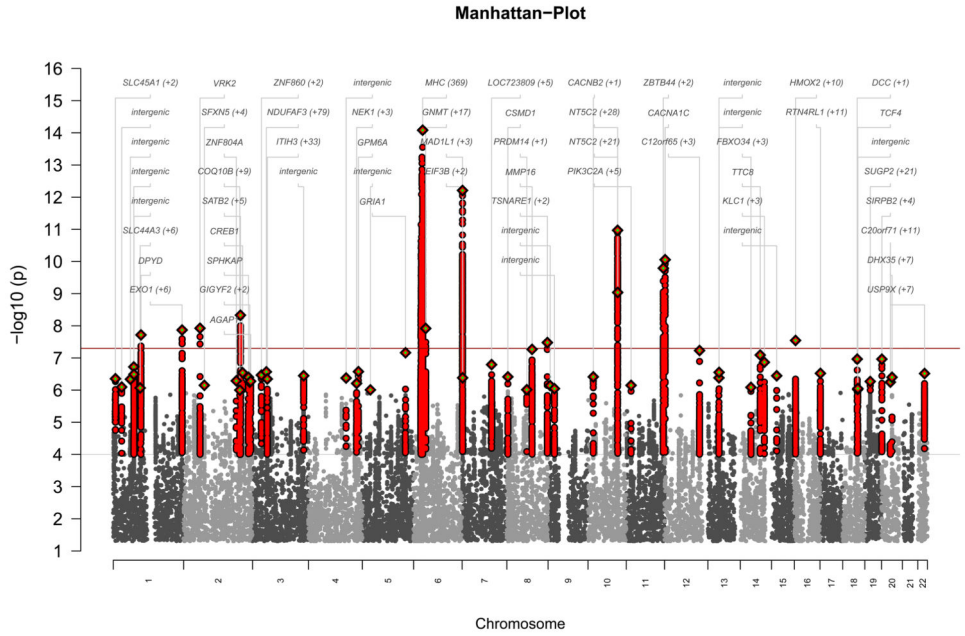

Figure 1.

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

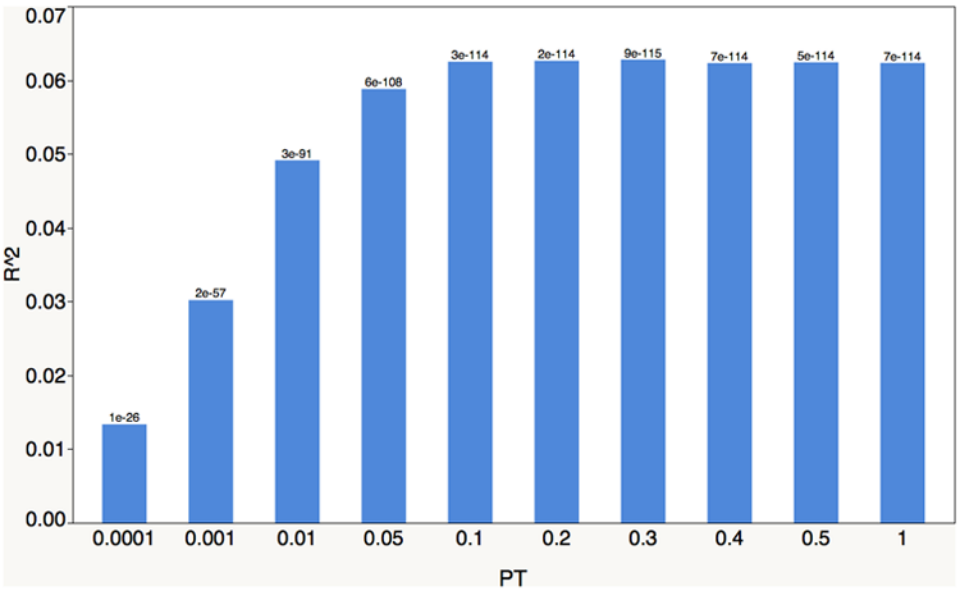

Figure 2.

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

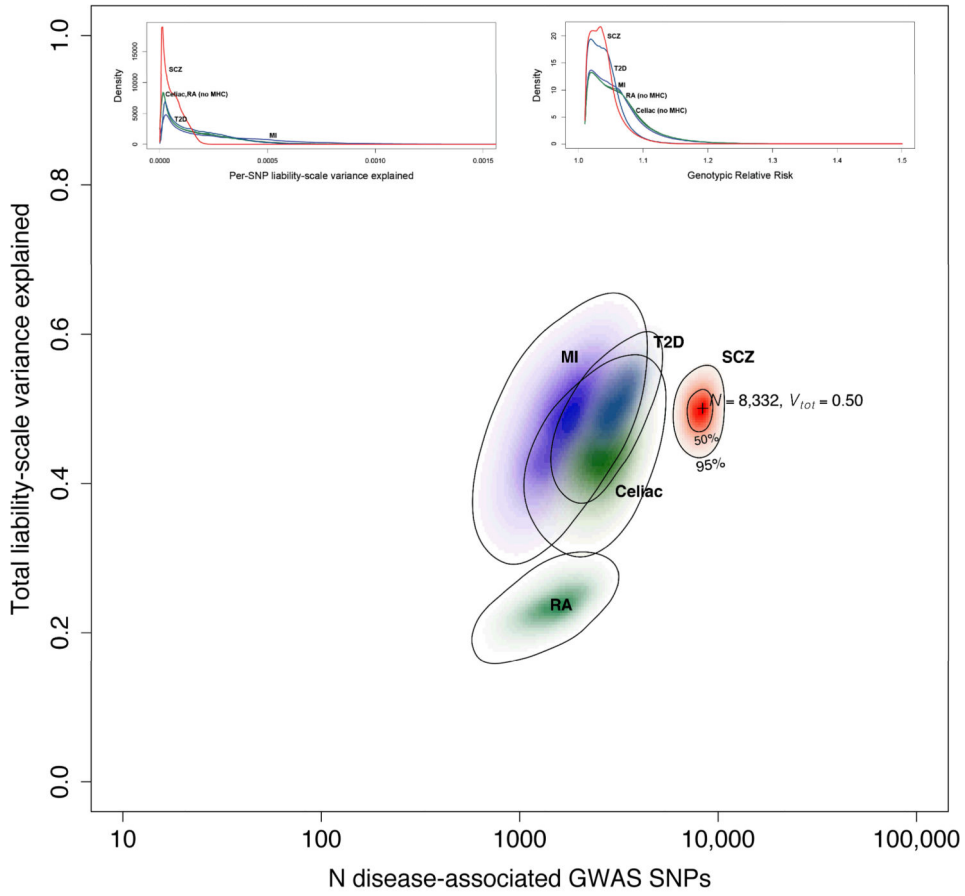

Figure 3.

The main figure shows the results of ABPA modeling based on the Sweden + PGC results (population risk 0.01). The x-axis is the estimated number of SNPs on a $\log_{10}$ scale, and the y-axis estimates the total variance in liability explained. The results for five conditions are shown: schizophrenia (this analysis, 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). [71] The schizophrenia results are based on 1000 Genomes imputation, and the others on HapMap3 imputation. Color intensity reflects the probability density with darker colors indicating higher density. Contour lines show 50% and 95% credible regions for SCZ, and 95% credible regions for the other diseases. The insets depict estimated SNP distributions for the five disorders: (a) distribution of SNPs in terms of the variance in liability explained per SNP and (b) the estimated distribution of SNP genotypic relative risks (GRR). We again stress that multiple qualifiers are essential in interpreting these estimates.

Table 1

Subject characteristics and sample sizes.

Feature	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)	n/a
Median total inpatient days	243 (81-696)	n/a
Median years from first to last HDR admission	9.7 (2.9-19.5)	n/a
Sample sizes		
Swedish subjects (Sw1-6)	5,001	6,243
PGC schizophrenia subjects (excluding Sw1-2)	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 inter-quartile ranges. The case group had significantly more males ($p < 0.0001$) and was significantly younger ($p < 0.0001$) than controls although these differences were not of large magnitude. The higher median age in controls is in the direction of greater confidence in control classification (i.e., controls had greater time at risk for psychiatric hospitalization). Cases tended to have had considerable hospitalizations, inpatient lengths of stay, and years of observation. IQQ=interquartile range, SCZ=schizophrenia, SAD=schizoaffective disorder, HDR=Hospital Discharge Register.

All cases and controls are independent. The Swedish sample totals N=11,244, the PGC N=20,899, and the replication samples N=27,175. The Sweden plus PGC meta-analysis is based on N=32, 143. The Swedish sample plus PGC plus replication samples total 59,318. (these counts exclude 511 trios).

Table 2

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

Chromosomal region	kb	SNPs	Index SNP				P-value			OR (SE)		
			rs ID	a12	bp	Freq	Sw+PGC	Replication	Combined	Sw+PGC	Replication	Combined
chr6:31,596,138-32,813,768	1217.6	1412	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)
chr10:104,487,871-105,245,420	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)
chr7:1,827,717-2,346,115	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)
chr1:98,141,112-98,664,991	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)
chr12:2,285,731-2,440,464	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)
chr10:18,601,928-18,934,390	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)
chr8:143,297,312-143,410,423	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)
chr1:73,275,828-74,099,273	823.4	1026	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)
chr11:130,706,918-130,894,976	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)
chr5:151,888,959-152,835,304	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)
chr19:19,354,937-19,744,079	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)
chr2:37,422,072-37,592,628	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)
chr5:101,581,848-101,870,822	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)
chr3:52,215,002-53,175,017	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)
chr2:145,139,727-145,214,607	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)
chr2:200,628,118-201,293,421	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)
chr18:52,722,378-52,827,668	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)
chr2:233,550,961-233,808,241	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)
chr1:243,593,066-244,025,999	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)
chr12:123,447,928-123,913,433	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)
chr1:243,418,063-243,627,135	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)
chr8:89,188,454-89,761,163	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)
chr5:60,484,179-60,843,706	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)
chr5:152,505,453-152,707,306	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)

Chromosomal region	kb	SNPs	Index SNP				P-value			OR (SE)		
			rs ID	a12	bp	Freq	Sw+PGC	Replication	Combined	Sw+PGC	Replication	Combined
chr2:185,533,580-186,057,716	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)
chr17:2,015,612-2,256,111	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)
chr3:36,834,099-36,964,583	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. The first three columns describe the genomic regions and the next three columns the index SNP, the SNP with strongest association in the genomic region. The next three columns show the P-values in the meta-analysis of Sw1-6 with the PGC schizophrenia results, the replication samples alone, and the final combined analysis of Sw1-6, PGC, and replication samples. The final three columns show the odd ratios (OR) and standard errors (SE). All locations UCSC hg19.

Table 3

Description of 22 genome-wide significant regions in the combined analysis

Chromosomal region	P-value	Prior GWSIG?	Gene in relation to index SNP	Other genes in genomic region defined by LD	eQTL	Disease associations
chr6:31,596,138-32,813,768	9.14×10^{-14}	SCZ	*HLA-DRB9*	MHC class II, many other genes, lincRNA	Many	Many
chr10:104,487,871-105,245,420	3.68×10^{-13}	SCZ	*C10orf32-AS3MT*	*C10orf26 CALHM1 CALHM2 CALHM3 CNNM2 CYP17A1 INA MIR1307 NT5C2 PCGF6 PDCD11 SFXN2 ST13P13 TAF5 USMG5*	*ACTR1A ARL3 AS3MT C10orf26 C10orf32 C10orf78 NT5C2 TMEM180 TRIM8*	GWAS-blood pressure, CAD, aneurysm
chr7:1,827,717-2,346,115	5.93×10^{-13}	No	*MAD1L1*	*FTSJ2 NUDT1 SNX8*	*C7orf27 FTSJ2 MAD1L1 NUDT1*	
chr1:98,141,112-98,664,991	1.72×10^{-12}	SCZ	(*MIR137*, 37kb)	*DPYD* & lincRNA	*DPYD*	*DPYD*-MR
chr12:2,285,731-2,440,464	5.22×10^{-12}	SCZ, BIP	*CACNA1C*	-	No data	*CACNA1C*-AUT-Timothy syn Brugada syn 3
chr10:18,601,928-18,934,390	1.27×10^{-10}	No	*CACNB2*	*NSUN6*	No data	*CACNB2*-Brugada syn 4, GWAS-blood pressure
chr8:143,297,312-143,410,423	2.19×10^{-10}	No	*TSNARE1*	-	No data	
chr1:73,275,828-74,099,273	3.64×10^{-10}	No	(x10NST00000415686.1, 4kb)	lincRNA	No data	
chr11:130,706,918-130,894,976	1.83×10^{-9}	No	(*SNX19*, 31kb)	lincRNA	*SNX19*	
chr5:151,888,959-152,835,304 chr5:152,505,453-152,707,306	2.65×10^{-9} 4.12×10^{-8}	No No	ENST00000503048.1	lincRNA (*GRIA1*)	No data	
chr19:19,354,937-19,744,079	3.44×10^{-9}	BIP	(*MAU2*, 4kb)	*CILP2 GATAD2A GMIP HAPLN4 LPAR2 MIR640 NCAN NDUFA13 PBX4 SUGP1 TM6SF2 TSSK6 YJEFN3*	No data	GWAS-lipid levels
chr2:37,422,072-37,592,628	6.78×10^{-9}	No	*QPCT*	*C2orf56 CEBPZ PRKD3 SULT6B1* lincRNA	No eQTL	
chr5:101,581,848-101,870,822	9.03×10^{-9}	No	*SLCO6A1*	lincRNA	No data	
chr3:52,215,002-53,175,017	1.16×10^{-8}	SCZ, BIP	*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-glycericaciduria-MR; RTF1-MR; GWAS-adiponectin, height, waist-hip ratio
chr2:145,139,727-145,214,607	1.19×10^{-8}	No	*ZEB2*	-	No eQTL	ZEB2-Mowat-Wilson syn-MR
chr2:200,628,118-201,293,421	1.21×10^{-8}	No	*FONG*	*C2orf47 C2orf60 C2orf69 SPATS2L TYW5* lincRNA	No data	GWAS-osteoporosis
chr18:52,722,378-52,827,668	1.22×10^{-8}	No	(ENST00000565991.1, 21kb)	lincRNA (*TCF4*)	No data	
chr2:233,550,961-233,808,241	1.51×10^{-8}	No	*C2orf82*	*GIGYF2 KCNJ13 NGEF*	No data	
chr1:243,593,066-244,025,999 chr1:243,418,063-243,627,135	1.80×10^{-8} 2.53×10^{-8}	No Yes	*AKT3* *SDCCAG8*	*CEP170*	*AKT3* *SDCCAG8*	
chr12:123,447,928-123,913,433	2.28×10^{-8}	No	*C12orf65*	*ABCB9 ARL6IP4 CDK2AP1 MIR4304 MPHOSPH9 OGFOD2 PITPNM2 RILPL2 SBNO1 SETD8* lincRNA	*ARL6IP4 CDK2AP1 OGFOD2 SBNO1*	*C12orf65-MR*; GWAS-HDL, height, head size
chr8:89,188,454-89,761,163	3.33×10^{-8}	SCZ	Intergenic	*MMP16* lincRNA	*MMP16*	
chr5:60,484,179-60,843,706	3.78×10^{-8}	No	ENST00000506902.1	*ZSWIM6C5orf43* lincRNA	*C5orf43 ZSWIM6*	

The prior GWSIG column indicates regions reported to meet genome-wide significance for schizophrenia or bipolar disorder. The first gene column shows the gene with respect to the SNP with the strongest association in the interval. Parentheses indicate that a SNP is not within a gene and show the distance to the nearest gene. The second gene column shows the other named genes in the genomic interval. The eQTL column shows SNP-transcript associations with $q < 0.05$ in peripheral blood. Underlining indicates eQTLs with the SNP with the

strongest association. Disease associations contains data from the NHGRI GWAS catalog, [24] OMIM, [42] and compilation of genes related to autism[72] and mental retardation. [42,73,74] No data means no Affymetrix U219 probesets or low expression in peripheral blood.
Abbreviations: GWSIG=genome-wide significant, SCZ=schizophrenia, BIP=bipolar disorder, AUT=autism, MR=mental retardation.