Schizophrenia is an idiopathic mental disorder with substantial morbidity, mortality and personal and societal costs\(^1\text{–}^3\). 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\(^4\text{–}^8\). 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\(^9\). Initial exome sequencing studies have not yet identified specific variants of unequivocal genome-wide significance\(^9\text{–}^13\), although larger studies are in progress. Previous GWAS have reported convincing statistical evidence for ~10 genomic regions\(^8\), including the major histocompatibility complex (MHC)\(^14\text{–}^16\) along with \textit{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\(^13\text{,}17\text{,}18\). 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
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 \( \lambda_{\text{GGC}} \) was 1.075, and \( \lambda_{1000} \) (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 \(^1\,^4\,^7\). The association with schizophrenia on chromosome 2 (rs35220450) is new, showed highly consistent effects in the Sw1–Sw6 genotyping batches (between five regions reached genome-wide significance: the MHC locus (chromosome 6: 27,261,324–32,610,445; minimum \( P = 2.18 \times 10^{-10} \)), ASMT-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 \( P = 2.40 \times 10^{-8} \)), RP11-58K62.1 (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 \( \beta \) coefficients with the same sign in the Swedish results, an observation highly inconsistent with the null hypothesis of no association \( (P = 2.2 \times 10^{-10}) \). \( \lambda_{\text{GGC}} \) was 1.186, and \( \lambda_{1000} \) was 1.012, values consistent with a polygenic pattern of association but not with gross inflation due to technical artifacts \(^9\,^10\). 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 \(^1\,^4\,^7\) 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 \( 1 \times 10^{-11} \) to \( 1 \times 10^{-15} \)) (Fig. 2). The increased sample size allowed improved risk profile prediction. The threshold at which the explanatory power of these risk-profile SNPs plateaued has decreased with increasing sample size: \( P = 0.1 \) in this study (Fig. 2) and 0.2 in the PGC report, and there was no plateau in the International Schizophrenia Consortium study \(^1\,^4\,^7\). 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.

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 neureilin 4), a gene previously implicated in mental retardation and autism, and there were multiple possible signals near MECP2 (causal in Rett syndrome; \( P = 9.3 \times 10^{-8} \)).

Most GWAS-associated variants were found outside of protein-coding regions \(^3\,^8\). 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 examined 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

| Table 1 Subject characteristics and sample sizes |
|-----------------------------------------------|
| **Swedish sample characteristics**             |
| Male sex                                       |
| Cases                                         | 0.595 |
| Controls                                      | 0.512 |
| Median age at sampling                        | 54 (45–62) |
| SCZ or SAD                                     | 7 (3–15) |
| Median hospital admissions                     | NA |
| Median total inpatient days                    | 243 (81–696) |
| Median years from first to last HDR admission  | 9.7 (2.9–19.5) |
| **Sample sizes**                              |        |
| Swedish subjects (Sw1–Sw6)                    | 5,001  |
| PGC schizophrenia subjects (excluding Sw1 and Sw2) | 8,832 |
| Replication results for up to 168 genomic regions | 7,413 |
| Total subjects                                | 21,248 |
| Controls                                      | 6,243 |
| NA                                            | 19,762 |
| NA                                            | 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).
replication genotypes were sought were identified using LD clumping, defined by LD (r² > 0.5) and a minimum association P of <1 × 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 × 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 × 10⁻⁸; rs4523957 in SRR, P = 5.69 × 10⁻⁸; and rs6550435 near TRANK1, P = 5.86 × 10⁻⁸, which also had P = 9 × 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 (Ch1orf26), DPYD-MIR137, SDCAG8 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 schizophrenia¹⁷,²⁷,²⁸, we replicated genome-wide significant association for a SNP in CACNA1C (encoding Ca₄,1.2; P = 5.2 × 10⁻¹² at the intronic SNP rs1006737). Intriguingly, we identified a genome-wide significant association with schizophrenia at CACNB2 (P = 1.3 × 10⁻¹⁰ at the intronic SNP rs17691888), which encodes the β₂ subunit of L-type calcium channels (Ca₄β₂). This locus was previously found to be significant when considering five psychiatric disorders as affected³⁰.

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 α₁c subunit forms the transmembrane pore and directly interacts with the intracellular β₂ subunit³². The β₂ subunit also antagonizes an endoplasmic reticulum retention motif on the α₁c subunit to facilitate transport to the plasma membrane³³. Additional genes with genome-wide significant evidence of association were implicated on the basis of membership in a proteomic network centered on Ca₄₂ (ref. 34), including the protein products of ACTRIA (α-centractin), the divalent metal cation transporter CNNM2 (P = 3.7 × 10⁻¹³; chromosome 10: 103,009,986–105,512,924) and CACNB2. Variants within a locus containing TNNCCI (encoding calcium-binding protein troponin C) also reached genome-wide significance (P = 1.1 × 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 potentiation³⁵–³⁸. 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 disorder⁴⁰. 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—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). 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).

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

| Chromosomal region | Length (kb) | SNP | rsID | A1,\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 x 10^-16 | 6.93 x 10^-2 | 9.14 x 10^-14 |
| Chr. 10: 104,487,871–105,245,420 | 757.5 | 362 | rs7085104 | AG | 104,628,873 | 0.645 | 1.07 x 10^-11 | 2.10 x 10^-3 | 3.68 x 10^-13 |
| Chr. 7: 1,827,717–2,346,115 | 518.4 | 566 | rs6461049 | TC | 2,017,445 | 0.571 | 6.17 x 10^-13 | 1.85 x 10^-2 | 5.93 x 10^-13 |
| Chr. 1: 98,141,112–98,664,991 | 524.1 | 50 | rs4380187 | AC | 185,811,940 | 0.529 | 1.92 x 10^-8 | 1.91 x 10^-3 | 1.72 x 10^-11 |
| Chr. 8: 89,188,454–243,627,135 | 243.4 | 252 | rs4523957 | TG | 2,208,899 | 0.616 | 3.17 x 10^-8 | 6.78 x 10^-2 | 5.86 x 10^-8 |

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 x 10^-8.

\dDetails for the index SNP, the SNP with the strongest association in the genomic region. \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. Odds ratio (OR) estimates and standard errors. *P* values reflect differences for the index SNP, the SNP with the strongest association in the genomic region. 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.
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 associationa | Candidate gene in relation to index SNPb | Other genes in genomic region defined by LDc | eQTLd | Disease associationse |
|---------------------|---------|-----------------------|------------------------------------------|-------------------------------------------|-------|----------------------|
| 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, CNM2, CYP17A1, INA, MIR1307, NT5SC2, PCGF6, POCD11, SFXN2, ST13P13, TAF5, USMGS5, WBP1L | ACTR1A, ARL3, AS3MT, C10orf32, C10orf78, NT5SC2, TME1M180, 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 | DPYD, DPYD: mental retardation |
| Chr. 1: 98,141,12–98,664,991 | 1.72 × 10−12 | SCZ | (MIR137, 37 kb) | DPYD, lincRNA | No data | No data |
| Chr. 12: 2,285,731–2,440,464 | 5.22 × 10−12 | SCZ, BPD | CACNA1C | – | No data | No data |
| Chr. 10: 18,601,928–18,934,390 | 1.27 × 10−10 | 5 disorders | CACNB2 | NSUN6 | No data | CACNA1C: autism, Timothy syndrome, Brugada syndrome 3 |
| Chr. 8: 143,297,312–143,410,423 | 2.19 × 10−10 | No | TSNARE1 | – | No data | No data |
| Chr. 11: 130,706,918–130,894,976 | 3.64 × 10−10 | No | (x1NST000004156861.1, 4 kb) | lincRNA | No data | No data |
| Chr. 5: 151,888,959–152,835,304 | 1.83 × 10−9 | No | (SNX19, 31 kb) | lincRNA | No data | SNX19 |
| Chr. 5: 152,505,453–152,707,306 | 2.65 × 10−9 | No | ENST00000503048.1 | lincRNA (GRIA1) | No data | No data |
| Chr. 19: 19,354,937–19,744,079 | 4.12 × 10−8 | No | No | No data | No data | GWAS: lipid levels |
| Chr. 2: 37,422,072–37,592,628 | 6.78 × 10−9 | No | QPCT | C2orf56, CEBPZ, PRKD3, SULT6B1 | No eQTL | No data |
| Chr. 5: 101,581,848–101,870,822 | 9.03 × 10−9 | No | SLC6A1 | lincRNA | No data | GRIA1: o-glyceric aciduria, mental retardation; RTF1: mental retardation; GWAS: adiponectin, height, waist-hip ratio |
| Chr. 3: 52,215,002–53,175,017 | 1.16 × 10−8 | SCZ, BPD | ITIH3 | ALAS1, ALDOB, BAP1, C3orf78, DHAH1, GLYCTK, GNL3, ITIH1, ITIH4, MIRR35, MIRR27, MUSTN1, NEK4, NISCH, NT5DC2, PBMR1, PHF7, PPM1M, RFT1, SEMASG, SFMBT1, SPCs1, STAB1, TLR9, TME1M110, TNDC1, TWF2, WDR82, lincRNA | C2orf75, C2orf69, CEBPZ, PRKD3, SULT6B1 | No data |
| 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 | No data |
| Chr. 2: 233,550,961–233,808,241 | 1.51 × 10−8 | No | C2orf82 | GIGYF2, KCNJ13, NGEF | No data | No data |
| Chr. 1: 243,593,066–244,025,999 | 1.80 × 10−8 | No | AKT3 | CEP170 | AKT3 | No data |
| Chr. 1: 243,418,063–243,627,135 | 2.53 × 10−8 | Yes | SOCCAG8 | – | No data | No data |
| Chr. 12: 123,447,928–123,913,433 | 2.28 × 10−8 | No | C12orf65 | ABCB9, ARL6IP4, CDK2AP1, MIR4304, MIPH506H9, DGFD2, PTPN42, RILPL2, SBN01, SETDB, lincRNA | ARL6IP4, CDK2AP1, DGFD2, SBN01 | 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 | No data |
| Chr. 5: 60,484,179–60,843,706 | 3.78 × 10−8 | No | ENST00000506902.1 | ZSWIM6, C5orf43, lincRNA | C5orf43, ZSWIM6 | No data |
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.\\n
Second, as reported previously, the strongest SNP association (rs114022140; \( P = 9.1 \times 10^{-14} \)) 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) and 6.4% of genome-wide significant SNP associations in the NHGRI GWAS catalog. 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-DRB3) (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) (C6orf47, HLA-DQA1, TNXB, VARS, WBP1L, CACNA1C, DPYD, CACNB2, TSSK6, NT3DC2, PITPNM2, SNON1, ZEB2 and PRKRD3). 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 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, 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.\\n
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. Using pathway analysis, we found modest enrichment (\( P = 0.06 \)) for smaller association \( P \) values in a conservatory set of lincRNAs derived from sequencing of polyadenylated RNA from multiple tissues. 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 overlap with SNPs implicated by GWAS over all traits as well as for specific traits such as height, adiposity, cardiovascular risk factors, chemotherapy-induced cytotoxicity, autism, schizophrenia and Crohn’s disease. 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.

**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). 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).](image-url)
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 Swedish 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 Swedish 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 individuals5, 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 accounts 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 elsewhere6,18,69,70.

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 heritability71. 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 field8. 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.
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. Satgunanathan-Dawoud, S. Hasselund, M. Rådstrom, B. Ohlander, L. Nyérn and L. Jakobsson. Funding support was provided by the NMHH (R01 MH077192 to P.S. and R01 MH09034 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 Königsa Foundation and the Netherlands Scientific Organization (NWO 645-000-003). SGENE was supported by European Union grant HEALTH-F2-2009-23243 (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 085472/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., J.P., A.B.S., A.L.R., B.K.R.-S., B.M.N., C.d.L., D.P., D. Ruderfer, F.B., J.P., K.L., M.H., M.V., P.S., S.A.M., S.P. and P.F.S. conducted statistical analyses. A.D.B., D.M.H., D. Ruescu, 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.I.C., S.W., Y.K., K.X. and P.F.S. performed bioinformatics analyses. K.C., J.I.M. and S.A. managed the project. B.P.R., D.W.M., E.A.O., H.S., J.T.W., K.S.K., M.G., M.L.O., N.C., PC., the Multicenter Genetic Studies of Schizophrenia, the Psychosis Endophenotypes International Consortium, the Wellcome Trust Case Control Consortium 2, A.P.C., 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., S.P.S., C.M.H. and P.F.S. Funding was obtained by E. O’Halloran, 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.

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

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.
Multicenter Genetic Studies of Schizophrenia Consortium:

Douglas F Levinson46, Pablo V Gejman47, Kenneth S Kendler21,22, Claudine Laurent48, Bryan J Mowry7, Michael C O'Donovan14, Michael J Owen14, Ann E Pulver49, Brien P Riley21,22,30, Sibylle G Schwaab50, Dieter B Wildenauer51, Frank Dudbridge52, Peter Holmans4, Jianxin Shi53, Margot Albus54, Dermot Walsh65, Nigel M Williams14 & Brandon Wormley22
Psychosis Endophenotypes International Consortium:

Maria J Arranz23,66, Steven Bakker67, Stephan Bender68,69, Elvira Bramon23,40,41, David Collier23,70, Benediço Crespo-Facorro71,72, Jeremy Hall73, Conrad Iyegbe23, Assem Jablensky74, Rene S Kahn67, Luba Kallaydjiieva75,76, Stephen Lawrie73, Cathryn M Lewis23, Kuang Lin21, Don H Linszen77, Ignacio Mata71,72, Andrew McIntosh73, Robin M Murray23, Roel A Ophoff78, John Powell23,29, Dan Rujescu31,32, Jim Van Os23,79, Muriel Walshe23, Matthias Weisbrod69 & Durk Wiersma80

Wellcome Trust Case Control Consortium 2:

Management Committee: Peter Donnelly (Chair)81,82, Ines Barroso (Deputy Chair)83, Jenefer M Blackwell84,85, Elvira Bramon23,40,41, Matthew A Brown86, Juan P Casas87,88, Aiden P Corvin13,18, Panos Deloukas83, Audrey Duncanson89, Janusz Jankowski90, Hugh S Markus91, Christopher G Mathew92, Colin N A Palmer93, Robert Plomin94, Anna Rautanen82, Stephen J Sawcer83, Richard C Trembath83, Ananth C Viswanathan96,97 & Nicholas W Wood98

Data and Analysis Group: Chris C A Spencer82, Gavin Ban82, Céline Bellenguez82, Colin Freeman82, Garrett Hellenthal82, Eleni Giannoulatos82, Matti Pirinen82, Richard D Pearson82, Amy Strange82, Zhan Su82, Damjan Vukcevic82 & Peter Donnelly81,82

DNA, Genotyping, Data QC and Informatics Group: Cordelia Langford83, Sarah E Hunt83, Sarah Edkins83, Rhian Gwilliam83, Hannah Blackburn83, Suzannah J Bumpstead83, Serge Dronov83, Matthew Gillman83, Emma Gray83, Naommi Hammond83, Alagurevathi Jayakumar83, Owen T McCann83, Jennifer Liddle83, Simon C Potter83, Radhi Ravindrarajah83, Michelle Ricketts83, Avazeh Tashakkori-Ghanbaria83, Matthew J Waller83, Paul Weston83, Sara Widdas83, Pamela Whittaker83, Ines Barroso83 & Panos Deloukas83

Publications Committee: Christopher G Mathew (Chair)92, Jenefer M Blackwell84,85, Matthew A Brown86, Aiden P Corvin13,18, Mark I McCarthy99 & Chris C A Spencer82

---

46Department of Psychiatry and Behavioral Sciences, NorthShore University HealthSystem and University of Chicago, Evanston, Illinois, USA. 47Department of Psychiatry and Behavioral Sciences, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA. 48Psychiatry and Psychotherapy Clinic, Friedrich-Alexander University, Erlangen-Nuremberg, Erlangen, Germany. 49Department of Psychiatry and Clinical Neurosciences, University of Western Australia, Perth, Western Australia, Australia. 50Department of Non-Communicable Disease Epidemiology, London School of Hygiene and Tropical Medicine, London, UK. 51Division of Cancer Epidemiology & Genetics, National Cancer Institute, Bethesda, Maryland, USA. 52State Mental Hospital, Haar, Germany. 53INSERM U1079, Faculté de Médecine, Rouen, France. 54Department of Biostatistics, Johns Hopkins University, Baltimore, Maryland, USA. 55Department of Psychiatry, University of Bonn, Bonn, Germany. 56Department of Medicine and Experimental Therapeutics, University of Cambridge, Addenbrooke’s Hospital, Cambridge, UK. 57First Department of Psychiatry, University of Athens Medical School, Athens, Greece. 58Department of Psychiatry, University of Regensburg, Regensburg, Germany. 59INSERM, Institut de Myologie, Hôpital Pitié Salpêtrière, Paris, France. 60Illumiia, Inc., La Jolla, California, USA. 61Department of Psychiatry, Hadassah Hebrew University Medical Center, Jerusalem, Israel. 62Department of Biostatistics, Johns Hopkins University, Baltimore, Maryland, USA. 63Department of Psychiatry, University of Bonn, Bonn, Germany. 64CNRS, Laboratoire de Genétique Moléculaire de la Neurotransmission et des Processus Neurodégénératifs, Hôpital Pitié Salpêtrière, Paris, France. 65The Health Research Board, Dublin, Ireland. 66Fundacion de Docencia e Recerca Mutua de Terrassa, Universitat de Barcelona, Barcelona, Spain. 67Department of Psychiatry, Rudolf Magnus Institute for Neuroscience, University Medical Center Utrecht, Utrecht, The Netherlands. 68Child and Adolescent Psychiatry, University of Technology Dresden, Dresden, Germany. 69Section for Experimental Psychopathology, General Psychiatry, Heidelberg, Germany. 70Discovery Neuroscience Research, Eli Lilly and Company, London, UK. 71CIBERSAM (Centro Investigación Biomédica en Red Salud Mental), Madrid, Spain. 72University Hospital Marques de Valdecilla, Instituto de Formacion e Investigacion Marques de Valdecilla, University of Cantabria, Santander, Spain. 73Division of Neuroscience, University of Edinburgh, Edinburgh, UK. 74Centre for Clinical Research in Neuropsychiatry, The University of Western Australia, Perth, Western Australia, Australia. 75Centre for Medical Research, The University of Western Australia, Perth, Western Australia, Australia. 76Western Australian Institute for Medical Research, The University of Western Australia, Perth, Western Australia, Australia. 77Department of Psychiatry, Academic Medical Center, University of Amsterdam, Amsterdam, The Netherlands. 78Center for Neurobehavioral Genetics, University of California, Los Angeles, Los Angeles, California, USA. 79Maastricht University Medical Centre, South Limburg Mental Health Research and Teaching Network, EURON, Maastricht, The Netherlands. 80Department of Psychiatry, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands. 81Department of Statistics, University of Oxford, Oxford, UK. 82Wellcome Trust Centre for Human Genetics, Oxford, UK. 83Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, UK. 84Cambridge Institute for Medical Research, University of Cambridge School of Clinical Medicine, Cambridge, UK. 85Telethon Institute for Child Health Research, Centre for Child Health Research, University of Western Australia, Subiaco, Western Australia, Australia. 86Diamantina Institute of Cancer, Immunology and Metabolic Medicine, Princess Alexandra Hospital, University of Queensland, Brisbane, Queensland, Australia. 87Department of Epidemiology and Public Health, University College London, London, UK. 88Department of Epidemiology, University College London, London, UK. 89Department of Medical and Molecular Genetics, School of Medicine, King’s College London, Guy’s Hospital, London, UK. 90Biomedical Research Centre, Ninewells Hospital and Medical School, Dundee, UK. 91Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, King’s College London, London, UK. 92Netherlands Institute for Health Research (NHR) Biomedical Research Centre at Moorfields Eye Hospital National Health Service Foundation Trust, London, UK. 93Department of Molecular Neuroscience, Institute of Neurology, London, UK. 94Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, King’s College London, London, UK. 95Department of Clinical Neurosciences, University of Cambridge, Addenbrooke’s Hospital, Cambridge, UK. 96Institute of Ophthalmology, University College London, London, UK. 97National Institute for Health Research (NIHR) Biomedical Research Centre at Moorfields Eye Hospital National Health Service Foundation Trust, London, UK. 98Department of Molecular Neuroscience, Institute of Neurology, London, UK. 99Oxford 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 phase17. 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 Register77,78, 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 diagnoses79–81 made by attending physicians for each hospitalization from 1973–18 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’ controls85. Control inclusion criteria included never being hospitalized for schizophrenia or bipolar disorder (given evidence of genetic overlap with schizophrenia)85,87, 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’ controls85. Control inclusion criteria included never being hospitalized for schizophrenia or bipolar disorder (given evidence of genetic overlap with schizophrenia)85,87, both parents born in Scandinavia and age of ≥18 years.

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 generated as described above and with full sample size.

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 epidemiological and genetic evidence.

The sample before quality control was approximately representative of the Swedish populace with regard to county of birth (Supplementary Fig. 12).

After basic quality control, 77,986 autosomal SNPs directly genotyped on all 4 GWAS platforms were extracted and pruned to remove SNPs in LD (r2 > 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 PLINK93 and pairs of subjects with r ≥ 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 λPC 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 schizophrenia94, we conducted chromosome X imputation for subjects passing quality control for the autosomal analysis (excluding X-chromosome SNPs with missingness of 20.05 or Hardy–Weinberg equilibrium P ≤ 1 × 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 covariates22. Meta-analysis was conducted using an inverse-weighted fixed-effects model21. 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 SNPs17.

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 r2 < 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 the 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 × 10−3, (ii) association P < 1 × 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 RPS14 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):

\[
\text{clump-p1 1–clump-p2 1–clump-r2 0.25–clump-kr 500 .}
\]

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 (\( P_i \)):

\[
0.0001, 0.001, 0.01, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5 \text{ and } 1.0 (\text{i.e., including all SNPs}).
\]

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 \( \beta \)) 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 R² 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.\(^9\). 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 (\( P_{\text{self}} \) and \( P_{\text{comp}} \)) 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\(^39\), the Irish Schizophrenia Genomics Consortium\(^86\), the Psychosis Endophenotype Consortium\(^72\) and the Multicenter Family Study\(^88\). After selecting for association \( P < 1 \times 10^{-5} \) in the Swedish 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. Badiani, 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. & Ekberg, A. Appendectomy and protection against ulcerative colitis. N. Engl. J. Med. 344, 808–814 (2001).

85. Harrison, 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. Harte, P. Participation in population studies. Epidemiology 17, 252–254 (2006).

89. Morton, L.M., Cahill, J. & Harte, 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. Am. J. Hum. Genet. 81, 150–159 (2007).

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. Psychology 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).