

Genome-wide association study identifies five new schizophrenia loci

The Schizophrenia Psychiatric Genome-Wide Association Study (GWAS) Consortium¹

We examined the role of common genetic variation in schizophrenia in a genome-wide association study of substantial size: a stage 1 discovery sample of 21,856 individuals of European ancestry and a stage 2 replication sample of 29,839 independent subjects. The combined stage 1 and 2 analysis yielded genome-wide significant associations with schizophrenia for seven loci, five of which are new (1p21.3, 2q32.3, 8p23.2, 8q21.3 and 10q24.32-q24.33) and two of which have been previously implicated (6p21.32-p22.1 and 18q21.2). The strongest new finding ($P = 1.6 \times 10^{-11}$) was with rs1625579 within an intron of a putative primary transcript for *MIR137* (microRNA 137), a known regulator of neuronal development. Four other schizophrenia loci achieving genome-wide significance contain predicted targets of *MIR137*, suggesting *MIR137*-mediated dysregulation as a previously unknown etiologic mechanism in schizophrenia. In a joint analysis with a bipolar disorder sample (16,374 affected individuals and 14,044 controls), three loci reached genome-wide significance: *CACNA1C* (rs4765905, $P = 7.0 \times 10^{-9}$), *ANKK3* (rs10994359, $P = 2.5 \times 10^{-8}$) and the *ITIH3-ITIH4* region (rs2239547, $P = 7.8 \times 10^{-9}$).

In stage 1, we conducted a mega-analysis combining genome-wide association study (GWAS) data from 17 separate studies (with a total of 9,394 cases and 12,462 controls; **Table 1** and **Supplementary Tables 1,2**). We imputed allelic dosages for 1,252,901 autosomal SNPs (**Table 1**, **Supplementary Table 3** and **Supplementary Note**) using HapMap3 as the reference panel¹. We tested for association using logistic regression of imputed dosages with sample identifiers and three principal components as covariates to minimize inflation in significance testing caused by population stratification. The quantile-quantile plot (**Supplementary Fig. 1**) deviated from the null distribution with a population stratification inflation factor of $\lambda = 1.23$. However, λ_{1000} , a metric that standardizes the degree of inflation by sample size, was only 1.02, similar to that observed in other GWAS meta-analyses^{2,3}. This deviation persisted despite comprehensive quality control and inclusion of up to 20 principal components (**Supplementary Fig. 1**). Thus, we interpret this deviation as indicative of a large number of weakly associated SNPs consistent with polygenic inheritance⁴. We also examined 298 ancestry-informative markers (AIMs) that reflect European-ancestry population substructure⁵. Unadjusted analyses

showed greater inflation in the test statistics than we saw for all markers (AIMs $\lambda = 2.26$ compared to all markers $\lambda = 1.56$). After inclusion of principal components, the distributions of the test statistics did not differ between AIMs ($\lambda = 1.18$) and all markers ($\lambda = 1.23$), a result inconsistent with population stratification explaining the residual deviation seen in **Supplementary Figure 1**. Moreover, the results of a meta-analysis using summary results generated using study specific principal components (**Supplementary Note**) were highly correlated with those from the mega-analysis (Pearson correlation = 0.94, with a similar $\lambda = 1.20$; **Supplementary Fig. 2**). Of the ten SNPs in **Table 2**, four increased and six decreased in significance, suggesting that the most extreme values did not result from systematic inflation artifacts. Therefore, our primary analysis used unadjusted P values (nevertheless, see **Table 2** for stage 1 P values adjusted for λ (ref. 6)).

In stage 1 (**Table 2**, **Supplementary Table 4** and **Supplementary Figs. 3 and 4**), 136 associations reached genome-wide significance ($P < 5 \times 10^{-8}$)⁷. The majority of these associations ($N = 129$) mapped to 5.5 Mb in the extended major histocompatibility complex (MHC, 6p21.32-p22.1), a region of high linkage disequilibrium (LD) previously implicated in schizophrenia in a subset of the samples used here^{4,8,9}. The other stage 1 regions included new regions (10q24.33 and 8q21.3) and previously reported regions (18q21.2 at *TCF4* (encoding transcription factor 4) and 11q24.2 (ref. 8)). The signal at 11q24.2 is ~0.85 Mb from *NRGN* (encoding neurogranin) and is uncorrelated with the previously associated variant near this gene⁸.

In **Table 2** and **Supplementary Table 4**, we denote regions of association by the most significant marker. Associated SNPs with $r^2 \geq 0.2$ in HapMap3 (CEU+TSI populations) were not considered independent. However, we noticed instances where multiple SNPs within 250 kb of each other yielded evidence for association ($P < 10^{-5}$) despite weak LD ($r^2 < 0.2$) between them. For regions with $P < 10^{-6}$, we performed a conditional analysis using as covariates the dosages of the strongest associated SNP, principal components 1–4 and 6 and study indicator. We observed multiple statistically independent signals at the MHC. Although a number of SNPs within the MHC were potentially independent per HapMap r^2 values, only rs9272105 withstood formal conditional analysis, showing $P = 1.8 \times 10^{-6}$ conditional on association to the best SNP, rs2021722 (stage 1 $P = 4.3 \times 10^{-11}$, inter-SNP distance = 2.4 Mb, $r^2 = 0.01$ in HapMap). Excluding the MHC region, we identified six regions with at least one SNP associated at $P < 10^{-5}$ and a second SNP with a conditionally independent $P < 10^{-3}$

¹A full list of authors and affiliations appears at the end of the paper.

Table 1 Study design and samples

Collection	Country	Platform	Cases included by sex				Controls included by sex			
			Male	Female	Unknown	Total	Male	Female	Unknown	Total
Cardiff UK	UK	Affymetrix 500K	320	152	0	472	1,442	1,492	0	2,934
CATIE	United States	Affymetrix 500K; Perlegen 164K	308	94	0	402	161	46	0	207
ISC-Aberdeen	UK	Affymetrix 5.0	536	184	0	720	447	251	0	698
ISC-Cardiff	Bulgaria	Affymetrix 6.0	270	257	0	527	291	318	0	609
ISC-Dublin	Ireland	Affymetrix 6.0	188	82	0	270	258	602	0	860
ISC-Edinburgh	UK	Affymetrix 6.0	267	101	0	368	146	138	0	284
ISC-London	UK	Affymetrix 5.0; Affymetrix 500K	369	149	0	518	207	284	0	491
ISC-Portugal	Portugal	Affymetrix 5.0	213	133	0	346	80	135	0	215
ISC-SW1	Sweden	Affymetrix 5.0	93	75	0	168	82	85	0	167
ISC-SW2	Sweden	Affymetrix 6.0	231	159	0	390	116	113	0	229
MGS	United States, Australia	Affymetrix 6.0	1,863	816	0	2,679	1,140	1,344	0	2,484
SGENE-Bonn	Germany	Illumina 550K	238	236	0	474	664	640	0	1,304
SGENE-Copenhagen	Denmark	Illumina Human 610-Quad	280	202	0	482	268	189	0	457
SGENE-Munich	Germany	Illumina 300K	279	155	0	434	167	184	0	351
SGENE-TOP3	Norway	Affymetrix 6.0	132	116	0	248	176	175	0	351
SGENE-UCLA	The Netherlands	Illumina 550K	529	175	0	704	310	321	–	631
Zucker Hillside	United States	Affymetrix 500K	128	64	0	192	92	98	0	190
Grand totals for the GWAS			6,244	3,150	0	9,394	6,047	6,415	–	12,462
Multicenter Pedigree	Europe, United States, Australia	Illumina Human 610-Quad	n.a.	n.a.	0	583	0	0	0	0
SGENE-Aarhus	Denmark	Illumina Human 610-Quad	477	399	0	876	477	397	0	874
SGENE-Aarhus	Denmark	Centaurus	114	102	1	217	176	317	0	493
SGENE-Belgium	Belgium	Centaurus; Illumina 370K	326	184	0	510	149	192	0	341
SGENE-Copenhagen	Denmark	Centaurus	264	198	0	462	499	375	0	874
SGENE-Iceland	Iceland	Illumina 300K	346	185	0	531	5,802	5,813	0	11,615
SGENE-England	UK	Illumina 300K	71	22	0	93	48	40	0	88
SGENE-Helsinki	Finland	Illumina 300K	112	70	0	59	122	75	0	147
SGENE-Kuusamo	Finland	Illumina 300K				123				50
SGENE-Hungary	Hungary	Centaurus	105	136	0	241	89	125	0	214
SGENE-Italy	Italy	Illumina 300K	48	36	0	84	50	39	0	89
SGENE-Munich	Germany	Illumina 300K	280	186	0	163	887	912	0	185
SGENE-Munich	Germany	Centaurus				303				1,614
SGENE-Russia	Russia	Centaurus	132	343	0	475	178	290	0	468
SGENE-Sweden	Sweden	Centaurus	158	94	0	252	178	109	0	287
SW3	Sweden	Affymetrix 6.0	327	212	0	539	457	448	0	905
SW4	Sweden	Affymetrix 6.0	656	407	0	1,063	605	568	0	1,173
UQ and ASRB	Australia	SequenomMassArray	347	190	21	558	487	455	15	957
ISGC and WTCCC2	Ireland	Affymetrix 6.0	968	342	0	1,310	245	778	0	1,023
Grand totals for the replication follow up			4,731	3,106	22	8,442	10,449	10,933	15	21,397

Stage 1 describes the 17 samples that provided full GWAS genotyping data, and stage 2 describes the 19 studies that provided results for the top SNPs identified in the combined analysis of stage 1 studies. Stage 2 replication SGENE-Belgium had four cases missing sex information. Stage 2 replication SGENE-Aarhus (focused genotyping sample) had one case missing sex information. Stage 2 replication-University of Queensland had 21 cases and 15 controls missing sex information. Sex information for the two stage 2 replication SGENE-Munich samples are combined. Sex information for the two stage 2 replication SGENE-Finnish (Helsinki and Kuusamo) samples are combined to enable that these two samples are located adjacent to each other in the table (rather than alphabetically). Multicenter Pedigree was a family sample, and so case sex counts are not applicable (n.a). SGENE, Schizophrenia Genetics Consortium; ISC, International Schizophrenia Consortium; TOP3, Thematic Organized Psychoses Research 3; UCLA, University of California at Los Angeles; SW1, Sweden 1; SW2, Sweden 2; WTCCC, Wellcome Trust case Control Consortium; for the Multicenter Pedigree study, the number of cases indicates the number of families; CATIE, Clinical Antipsychotic Trials of Intervention Effectiveness; MGS, Molecular Genetics of Schizophrenia; UQ, University of Queensland; ASRB, Australian Schizophrenia Research Bank; ISGC, Irish Schizophrenia Genomics Consortium.

(Supplementary Table 5). We performed 100 simulations after permuting case-control status randomly within each study. In contrast to the six regions in the real dataset, we never observed more than a single region with co-localized statistically independent signals in any simulated genome-wide scan, indicating our observation is highly unlikely to have occurred by chance.

Noteworthy co-localizing independent signals occurred at three regions (Supplementary Table 5): one region with a genome-wide significant association at 10q24.32-q24.33 (Table 2), a second region that nearly met this threshold at *MAD1L1* (encoding mitotic arrest deficient-like 1; rs10226475, $P = 5.06 \times 10^{-8}$; Supplementary Table 4)

and a third region at *CACNA1C* (encoding calcium channel, voltage-dependent, L type, $\alpha 1C$ subunit), the latter of which has previously been associated with bipolar disorder¹⁰ and other psychiatric phenotypes including schizophrenia¹¹. The conditionally independent signal at *CACNA1C* was more significant than any observation made in 100 permutations of the entire experiment (both conditional $P < 10^{-5}$) and supports *CACNA1C* in schizophrenia after genome-wide correction ($P < 0.01$), even without considering these prior reports.

In stage 2, we evaluated in 29,839 independent subjects (8,442 cases and 21,397 controls) the most significant SNPs ($N = 81$) in each LD region where at least one SNP had surpassed $P < 2 \times 10^{-5}$

Table 2 Top genome-wide association results for schizophrenia

SNP	Chr.	Mb	Alleles	Frequency	<i>P</i> (GC-adjusted <i>P</i>)	OR (95% CI)	Consistency of direction	Gene	Distance (kb)
rs1625579	1p21.3 ^a	98.3	TG	0.80	5.72×10^{-7} (6.52×10^{-6})	1.14 (1.08–1.19)	+++++++	<i>MIR137</i>	Intragenic
					2.65×10^{-6} (n.a.)	1.11 (1.07–1.16)			
					1.59×10^{-11} (6.87×10^{-10})	1.12 (1.09–1.16)			
rs17662626	2q32.3 ^a	193.7	AG	0.91	3.09×10^{-6} (2.60×10^{-5})	1.22 (1.13–1.30)	+ -- +++	<i>PCGEM1</i>	343
					1.70×10^{-3} (n.a.)	1.16 (1.06–1.27)			
					4.65×10^{-8} (1.25×10^{-6})	1.20 (1.13–1.26)			
rs2021722	6p21.3-p22.1	30.3	CT	0.78	3.84×10^{-7} (4.71×10^{-6})	1.14 (1.09–1.19)	+ + + ---	<i>TRIM26</i>	Intragenic
					1.55×10^{-3} (n.a.)	1.10 (1.03–1.17)			
					2.18×10^{-12} (2.88×10^{-10})	1.15 (1.11–1.19)			
rs10503253	8p23.2 ^a	4.2	AC	0.19	3.84×10^{-7} (4.71×10^{-6})	1.14 (1.09–1.19)	+ + + ---	<i>CSMD1</i>	Intragenic
					7.60×10^{-3} (n.a.)	1.08 (1.01–1.14)			
					4.14×10^{-8} (8.98×10^{-7})	1.11 (1.07–1.15)			
rs7004633	8q21.3 ^a	89.8	GA	0.18	1.58×10^{-7} (2.27×10^{-6})	1.11 (1.07–1.15)	+++++++	<i>MMP16</i>	421
					0.011 (n.a.)	1.05 (1.01–1.10)			
					2.75×10^{-8} (7.03×10^{-7})	1.10 (1.07–1.14)			
rs7914558	10q24.32 ^a	104.8	GA	0.59	1.58×10^{-7} (2.27×10^{-6})	1.11 (1.07–1.15)	+ + + +++	<i>CNNM2</i>	Intragenic
					1.07×10^{-3} (n.a.)	1.08 (1.03–1.13)			
					1.82×10^{-9} (3.11×10^{-8})	1.10 (1.07–1.13)			
rs11191580	10q24.33 ^a	104.9	TC	0.91	1.58×10^{-7} (2.27×10^{-6})	1.11 (1.07–1.15)	+++++++	<i>NT5C2</i>	Intragenic
					5.09×10^{-3} (n.a.)	1.09 (1.02–1.16)			
					1.11×10^{-8} (3.72×10^{-7})	1.15 (1.10–1.20)			
rs548181	11q24.2	125.0	GA	0.88	1.00×10^{-6} (1.03×10^{-5})	1.10 (1.06–1.14)	+++++++	<i>STT3A</i>	1
					0.068 (n.a.)	1.04 (0.98–1.11)			
					8.87×10^{-7} (1.74×10^{-5})	1.11 (1.07–1.16)			
rs12966547	18q21.2	50.9	GA	0.58	1.00×10^{-6} (1.03×10^{-5})	1.10 (1.06–1.14)	+++++++	<i>CCDC68</i>	126
					2.29×10^{-5} (n.a.)	1.08 (1.04–1.12)			
					2.60×10^{-10} (5.99×10^{-9})	1.09 (1.06–1.12)			
rs17512836	18q21.2	51.3	CT	0.02	2.35×10^{-8} (4.78×10^{-7})	1.40 (1.28–1.52)	+++++++	<i>TCF4</i>	Intragenic
					0.085 (n.a.)	1.08 (0.96–1.20)			
					1.05×10^{-6} (2.86×10^{-5})	1.23 (1.14–1.31)			

The SNPs listed are those with a stage 1 $P < 5 \times 10^{-8}$ and/or a combined stage 1 and 2 $P < 5 \times 10^{-8}$. These ten independent ($r^2 < 0.2$) SNPs represent eight physically distinct genomic loci, as there are two SNPs listed for two loci (10q24.32-q24.33 and 18q21.2). For the MHC region, only one SNP is listed for clarity. The eight susceptibility loci represent three previously reported and five new loci (Supplementary Table 7). Stage 1 is the discovery GWAS mega-analysis. Stage 2 is the replication sample (single-tailed meta-analysis *P* values are weighted by 1/s.e.), and because the *P* values are single tailed, some 95% confidence intervals contain 1 (if $0.10 < P < 0.05$). Combined values include stages 1 and 2 (two-tailed meta-analysis *P* values are weighted by 1/s.e.). For each SNP, *P* values and odds ratios are listed for stage 1 (top), stage 2 (middle) and combined stage 1 and 2 analysis (bottom) with the genomic control (GC)-adjusted values bracketed (n.a., not applicable for stage 2). Alleles are listed with the stage 1 risk allele first; the frequency (in stage 1 controls) and odds ratio (OR) refer to the stage 1 risk allele. Bolded *P* values indicate $P < 5 \times 10^{-8}$, except for in the stage 2 data, where bolded values indicate $P < 0.05$. The directions of association in eight replication samples are represented by + if the associations are in the same direction, – if they are in opposite directions and a blank space if the data are not available. Mb is the base position based on hg18. Cytogenetic bands are listed for each SNP, though because only one of multiple MHC SNPs are listed, a band range is given in that instance. The nearest gene (or microRNA) is listed, with the distance (kb) from the gene (or if the SNP is intragenic) noted. None of these SNPs showed a significant test for heterogeneity among the samples. Chr., chromosome.

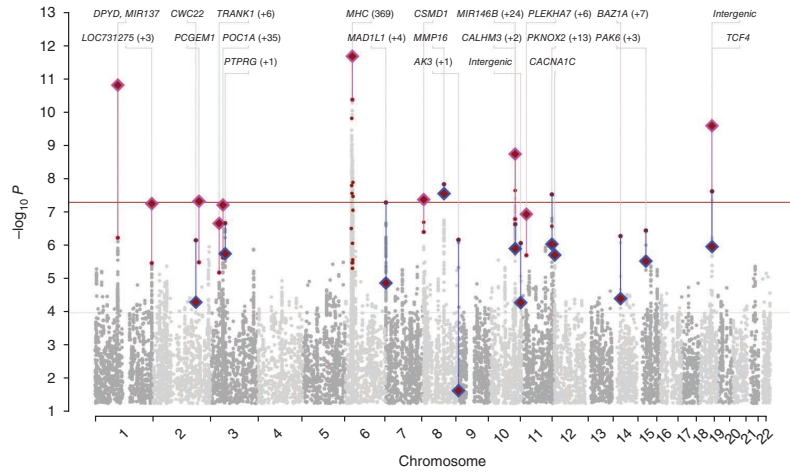
^aNew finding.

(Supplementary Table 6) in the mega-analysis. Of 22 SNPs from the MHC, 5 surpassed the genome-wide significant threshold in stages 1 and 2 combined (minimum $P = 2.2 \times 10^{-12}$ at rs2021722; Supplementary Table 6). Excluding the MHC region, a sign test for consistency between stages 1 and 2 was highly significant ($P < 10^{-6}$), with the same direction of effect as observed stage 1 also being observed in stage 2 for 49 of 59 SNPs. A Fisher's combined test revealed the distribution of stage 2 *P* values was unlikely to have occurred by chance ($P < 10^{-15}$). We also performed a transmission analysis using the family based Multicenter Pedigree replication sample in conjunction with a GWAS of 622 parent-offspring schizophrenia trios from Bulgaria¹², and the stage 1 associated allele was over-transmitted to cases for 44 of the 59 SNPs (one-sided $P = 1.0 \times 10^{-4}$). Thus, the stage 2 replication results are highly consistent with the stage 1 discovery results.

In the combined dataset (stages 1 and 2), five new (1p21.3, 2q32.3, 8p23.2, 8q21.3 and 10q24.32-q24.33) and two previously reported (6p21.32-p22.1 and 18q21.2) loci met genome-wide significance

(Figs. 1, 2, Table 2, Supplementary Tables 6, 7 and Supplementary Fig. 4). After adjusting for λ (ref. 6), four loci (1p21.3, 6p21.32-p22.1, 10q24.32-q24.33 and 18q21.2) remained significant at $P \leq 5 \times 10^{-8}$. For the primary analyses (unadjusted for λ), the strongest new association was at 1p21.3 (rs1625579; $P = 1.6 \times 10^{-11}$), which is over 100 kb from any RefSeq protein-coding gene but is within intron 3 of AK094607, which contains the primary transcript for *MIR137* (ref. 13). The next best locus, 10q24.32 (Supplementary Table 5 and Supplementary Fig. 5), has independent associations 130 kb apart at rs7914558 ($P = 1.8 \times 10^{-9}$) and rs11191580 ($P = 1.1 \times 10^{-8}$), implicating a 0.5-Mb region containing multiple genes (Supplementary Fig. 5). The third best locus, rs7004633 ($P = 2.8 \times 10^{-8}$) on 8q21.3, is 400 kb from the nearest gene (*MMP16*, encoding matrix metalloproteinase 16). The fourth best locus, rs10503253 ($P = 4.4 \times 10^{-8}$) at 8p23.2, is in an intron of *CSMD1* (encoding CUB and Sushi multiple domains 1). Finally, rs17662626 ($P = 4.7 \times 10^{-8}$) at 2q32.3 is intergenic, mapping 300 kb from a non-coding RNA, *PCGEM1* (prostate-specific transcript 1)¹⁴.

Figure 1 Manhattan plot for stages 1 and 2. Standard $-\log_{10} P$ plot of the study results. For the stage 1 results, 16 regions with one or more SNP achieving $P < 10^{-6}$ are highlighted in color and labeled with the name of the nearest gene. SNPs selected for stage 2 replication are highlighted, with the resulting combined P value after replication (that is, after incorporation of stage 2 results) indicated by the large diamonds. Blue highlighting indicates SNPs that were less significantly associated after replication, and pink highlighting indicates SNPs that were more significantly associated after replication.



MIR137 has been implicated in regulating adult neurogenesis^{15,16} and neuronal maturation¹⁷, mechanisms through which variation at this locus could contribute to brain development abnormalities in schizophrenia. Of relevance, two independent schizophrenia imaging studies found *MIR137* to be one of three microRNAs with targets significantly enriched for association¹⁸. In stage 1, SNPs in or near 301

high-confidence predicted *MIR137* targets (with a TargetScan¹⁹ probability of conserved targeting ≥ 0.9) were enriched for association compared with genes matched for size and marker density: 17 predicted *MIR137* targets (**Supplementary Table 8**) had at least one SNP with

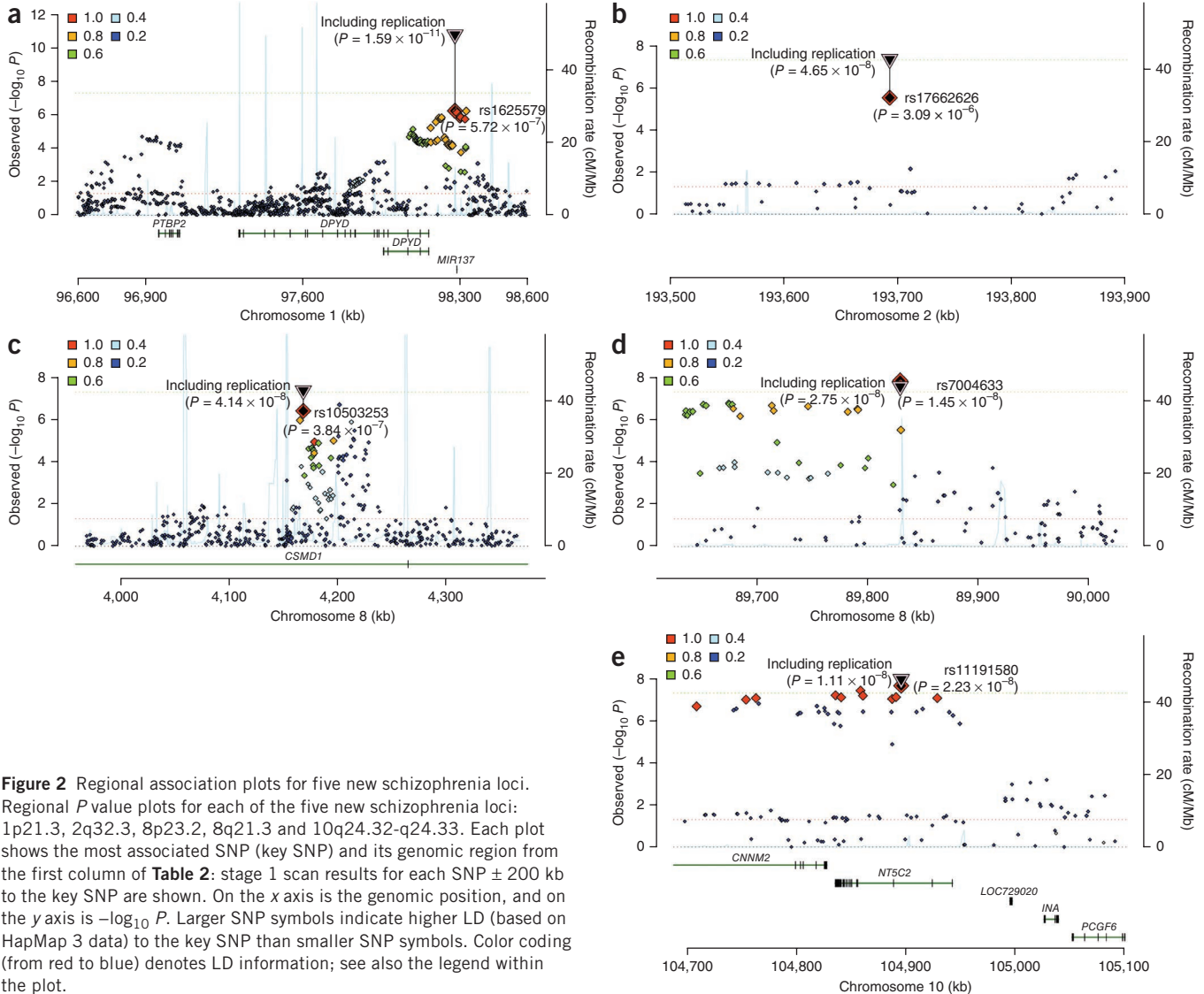


Figure 2 Regional association plots for five new schizophrenia loci. Regional P value plots for each of the five new schizophrenia loci: 1p21.3, 2q32.3, 8p23.2, 8q21.3 and 10q24.32-q24.33. Each plot shows the most associated SNP (key SNP) and its genomic region from the first column of **Table 2**: stage 1 scan results for each SNP ± 200 kb to the key SNP are shown. On the x axis is the genomic position, and on the y axis is $-\log_{10} P$. Larger SNP symbols indicate higher LD (based on HapMap 3 data) to the key SNP than smaller SNP symbols. Color coding (from red to blue) denotes LD information; see also the legend within the plot.



$P < 10^{-4}$, which is more than twice as many as the control gene sets ($P < 0.01$). Excluding the MHC and *MIR137*, of the nine loci with genome-wide significant support either in stage 1 or in the combined set (six loci, 2q32.3, 8p23.2, 8q21.3, 10q24.32-q24.33, 11q24.2 and 18q21.2; **Table 2** and **Supplementary Tables 6,7**) or in a joint analysis with bipolar disorder (three genes, *CACNA1C*, *ANK3* and *ITIH3-ITIH4*, described below), four genes (*TCF4*, *CACNA1C*, *CSMD1* and *C10orf26*) have predicted *MIR137* target sites according to analyses using three different prediction programs (TargetScan¹⁹, PicTar²⁰ and miRanda²¹). *In vitro* overexpression and locked nucleic acid-mediated knockdown of *MIR137* in neuronal cell line N2a leads to changes in expression levels of *TCF4* protein, strongly supporting the prediction that *TCF4* is a target of *MIR137* (L.-H. Tsai, personal communication). Our observations suggest *MIR137*-mediated dysregulation as a new etiologic mechanism in schizophrenia.

The International Schizophrenia Consortium (ISC) reported evidence for a polygenic contribution to schizophrenia⁴. An independent family based study confirmed these results, greatly minimizing the possibility of population stratification artifact¹². We reevaluated the polygenic model, dividing stage 1 samples into independent training and testing sets (**Supplementary Note**). The training set had 15,429 subjects (over twice the size of the ISC training set), and the testing set consisted of 6,428 individuals independent of the ISC report. The proportion of variance (Nagelkerke's r^2) explained in the testing set increased from 3% in the ISC to around 6% here (**Supplementary Table 9** and **Supplementary Fig. 6**). This estimate is much lower than the true total variation in liability that is tagged by all SNPs because SNP effects are estimated with error^{3,4,22-25}. The polygenic model appears to explain a substantial fraction of the heritability of schizophrenia⁴, as has been shown for other complex traits^{3,26-28}. Some of these additional risk loci are likely contained near the most highly significant results of our stage 1 analysis. Supporting this hypothesis, of the top loci that did not reach genome-wide significance in the combined stage 1 and 2 analysis, a sign test ($P < 10^{-4}$) and a Fisher's combined test ($P < 10^{-5}$) both showed an excess of same-direction allelic association (41 of 51 non-MHC SNPs) in the discovery and replication datasets.

Clinical, epidemiological and genetic findings suggest shared risk factors between bipolar disorder and schizophrenia²⁹. In stage 1, three genes with strong support had prior genome-wide significant associations with bipolar disorder: *CACNA1C*, the region containing *ITIH3-ITIH4* (encoding inter- α (globulin) inhibitors H3 and H4) and *ANK3* (encoding ankyrin 3, node of Ranvier (ankyrin G))^{10,11,30} (**Supplementary Table 10**). We performed a joint analysis with the Schizophrenia Psychiatric Genome-Wide Association Study (GWAS) Consortium (PGC) for bipolar disorder applying identical analytical methods. After removing duplicate subjects, we analyzed 16,374 cases with schizophrenia, schizoaffective disorder or bipolar disorder and 14,044 controls. Support for shared susceptibility was strengthened (**Supplementary Table 11**) at *CACNA1C* (rs4765905, $P = 7.0 \times 10^{-9}$), *ANK3* (rs10994359, $P = 2.5 \times 10^{-8}$) and the *ITIH3-ITIH4* region (rs2239547, $P = 7.8 \times 10^{-9}$), each of which reached genome-wide significance. A coding variant in *ITIH4* (p.Pro698Thr; rs4687657) is in perfect LD with the most associated SNP. Although we included all subjects from an earlier report¹⁰, the increased support found with additional independent cases ($N = 11,987$) and controls ($N = 7,835$) provides further evidence for shared risk effects of schizophrenia and bipolar disorder.

The risk variants implicated here confer small risks (odds ratios ~1.10), but the polygenic analysis shows many more susceptibility variants with effects for which our sample is underpowered (**Supplementary Table 12**). At every stage where samples were added, we found an increase in the number of genome-wide significant loci

and enhancement of signals at *CACNA1C*, *ANK3* and *ITIH3-ITIH4* when schizophrenia and bipolar disorder were jointly analyzed. Thus, gains in power offset any penalty for increased heterogeneity.

In summary, we report seven genome-wide significant schizophrenia associations (five of which are new) in a two-stage analysis of 51,695 individuals. We also report loci that confer susceptibility to both bipolar disorder and schizophrenia. The association near *MIR137*, associations in multiple predicted *MIR137* targets and the known role of *MIR137* in neuronal maturation and function together suggest an intriguing new insight into the pathogenesis of schizophrenia.

URLs. PLINK, <http://pngu.mgh.harvard.edu/~purcell/plink/>; Haploview, <http://www.broadinstitute.org/scientific-community/science/programs/medical-and-population-genetics/haploview/haploview>.

METHODS

Methods and any associated references are available in the online version of the paper at <http://www.nature.com/naturegenetics/>.

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS

We thank the study participants and the research staff at the many study sites. Over 40 US National Institutes of Health grants and similar numbers of government grants from other countries, along with substantial private and foundation support, enabled this work. We greatly appreciate the sustained efforts of T. Lehner (National Institute of Mental Health) on behalf of the Schizophrenia Psychiatric Genome-Wide Association Study (GWAS) Consortium (PGC). Detailed acknowledgments, including grant support, are listed in the **Supplementary Note**.

AUTHOR CONTRIBUTIONS

The Schizophrenia Psychiatric Genome-Wide Association Study (GWAS) Consortium (PGC): **overall coordination:** P.V.G. **Coordination of statistical analyses:** M.J.D. **Coordination of phenotypic analyses:** K.S.K. **Statistical analyses:** S.R., M.J.D., P.A.H., D.-Y.L., S.P., F.D., B.M.N., L.R., P.M.V., D.P., D.M.R. **Manuscript preparation:** P.V.G. (primary), M.J.D. (primary), A.R.S. (primary), S.R. (primary), M.C.O. (primary), K.S.K., D.E.L., P.S., P.A.H., P.F.S. (primary), D.-Y.L., J.D., R.A.O., O.A.A., E. Scolnick. **Phenotypic analyses:** K.S.K., A.F., A.C., R.L.A. **Stage 1 GWAS sample 1—Cardiff, UK:** M.C.O., N.C., P.A.H., M. Hamshere, H.J.W., V. Moskvina, S. Dwyer, L.G., S.Z., M.J.O. **Stage 1 GWAS sample 2—Clinical Antipsychotic Trials of Intervention Effectiveness (CATIE):** P.F.S., D.-Y.L., E.v.d.O., Y.K., T.S.S., J.A.L. **Stage 1 GWAS sample 3—International Schizophrenia Consortium (ISC)—Aberdeen:** D.St.C. **Stage 1 GWAS sample 4—ISC—Cardiff:** G.K.K., M.C.O., P.A.H., L.G., I.N., H.J.W., D.T., V. Milanova, M.J.O. **Stage 1 GWAS sample 5—ISC—Dublin:** D.W.M., C.T.O., E.K., E.M.Q., M.G., A.C. **Stage 1 GWAS sample 6—ISC—Edinburgh:** D.H.R.B., K.A.M., B.P., P. Malloy, A.W.M., A. McIntosh. **Stage 1 GWAS sample 7—ISC—London:** A. McQuillin, K.C., S. Datta, J.P., S. Thirumalai, V.P., R.K., J. Lawrence, D.Q., N.B., H.G. **Stage 1 GWAS sample 8—ISC—Portugal:** M.T.P., C.N.P., A.F. **Stage 1 GWAS sample 9—ISC—SW1—Sweden, stage 1 GWAS sample 10—ISC—SW2—Sweden, stage 2 replication follow-up sample 16—SW3—Sweden, stage 2 replication follow-up sample 17—SW4—Sweden:** C.M.H., P.L., S.E.B., S.P., E. Scolnick, P.S., P.F.S. **Stage 1 GWAS sample 11—Molecular Genetics of Schizophrenia (MGS):** J. Shi, D.F.L., J.D., A.R.S., M.C.K., B.J.M., A.O., F.A., C.R.C., J.M.S., N.G.B., W.F.B., D.W.B., K.S.K., R.F., P.V.G. **Stage 1 GWAS sample 12—Schizophrenia Genetics Consortium (SGENE)—Bonn:** S.C., M. Rietschel, M.M.N., W.M., T.G.S., M. Mattheisen. **Stage 1 GWAS sample 13—SGENE—Copenhagen, stage 2 replication follow-up sample 5—SGENE—Copenhagen:** T.H., A.I., K.D.J., L.D., G.J., H.B.R., B.G., J.N., S. Timm, L.O., A.G.W., A.F.-J., J.H.T., T.W. **Stage 1 GWAS sample 14—SGENE—Munich, stage 2 replication follow-up sample 12—SGENE—Munich, stage 2 replication follow-up sample 13—SGENE—Munich:** I.G., A.M.H., H.K., M.F., B.K., P. Muglia, D.R. **Stage 1 GWAS sample 15—SGENE—Thematic Organized Psychoses Research 3 (TOP3):** S. Djurovic, M. Mattingsdal, I.A., I.M., O.A.A. **Stage 1 GWAS sample 16—SGENE—UCLA:** R.A.O., R.M.C., N.B.F., R.S.K., D.H.L., J.v.O., D. Wiersma, R.B., W.C., L.d.H., L.K., I.M.-G., E. Strengman. **Stage 1 GWAS sample 17—Zucker Hillside:** A.K.M., T.L. **Stage 2 replication follow-up sample 1—multicenter pedigree:** P.A.H., B.P.R., A.E.P., M.J.O., D.B.W., P.V.G., B.J.M., C.L., K.S.K., G.N., N.M.W., S.G.S., A.R.S., M. Hansen, D.A.N., J.M., B.W.

V.K.L., M.C.O., J.D., M. Albus, M. Alexander, S.G., R.R., K.-Y.L., N.N., W.M., G.P., D. Walsh, M.J., F.A.O., F.B.L., D. Dikeos, J.M.S., D.F.L. **Stage 2 replication follow-up sample 2-SGENE-Aarhus:** A.D.B., D. Demontis, P.B.M., D.M.H., T.F.Ø., O.M. **Stage 2 replication follow-up sample 3-SGENE-Aarhus:** O.M., M.N., A.D.B. **Stage 2 replication follow-up sample 4-SGENE-Belgium:** R.v.W., G.K., M.D.H., J.V. **Stage 2 replication follow-up sample 6-SGENE-Iceland:** H.S., S.S., E. Sigurdsson, H.P., K.S. **Stage 2 replication follow-up sample 7-SGENE-England:** D.A.C. **Stage 2 replication follow-up sample 8-SGENE-Helsinki, stage 2 replication follow-up sample 11-SGENE-Kuusamo:** L.P., O.P.H.P., J. Suvisaari, J. Lönnqvist. **Stage 2 replication follow-up sample 9-SGENE-Hungary:** I.B., J.M.R. **Stage 2 replication follow-up sample 10-SGENE-Italy:** M. Ruggeri, S. Tosato. **Stage 2 replication follow-up sample 14-SGENE-Russia:** V.G. **Stage 2 replication follow-up sample 15-SGENE-Sweden:** E.G.J., I.A., L.T. **Stage 2 replication follow-up sample 18-University of Queensland:** B.J.M., M.A.B., P.A.D., J.J.M., D.E.M. **Stage 2 replication follow-up sample 18-Australian Schizophrenia Research Bank:** B.J.M., V.J.C., R.J.S., S.V.C., F.A.H., A.V.J., C.M.L., P.T.M., C.P., U.S. **Stage 2 replication follow-up sample 19-Irish Schizophrenia Genomics Consortium (ISGC):** A.C., D.W.M., P.C., B.S.M., C.T.O., G.D., F.A.O., M.G., K.S.K., B.P.R., ISGC (see the Acknowledgments in the **Supplementary Note** for additional contributors not listed above). **Stage 2 replication follow-up sample 19-Wellcome Trust Case Control Consortium 2 (WTCCC2):** P.D. (Chair of Management Committee; Data and Analysis Group), C.C.A.S. (Data and Analysis Group; Publications Committee), A.S. (Data and Analysis Group), WTCCC2 (see Acknowledgments in the **Supplementary Note** for additional contributors not listed above). All authors contributed to the current version of the paper.

COMPETING FINANCIAL INTERESTS

The authors declare competing financial interests: details accompany the full-text HTML version of the paper at <http://www.nature.com/naturegenetics/>.

Published online at <http://www.nature.com/naturegenetics/>.

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

- Altshuler, D.M. *et al.* Integrating common and rare genetic variation in diverse human populations. *Nature* **467**, 52–58 (2010).
- Barrett, J.C. *et al.* Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease. *Nat. Genet.* **40**, 955–962 (2008).
- Lango Allen, H. *et al.* Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature* **467**, 832–838 (2010).
- Purcell, S.M. *et al.* Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* **460**, 748–752 (2009).
- Price, A.L. *et al.* Discerning the ancestry of European Americans in genetic association studies. *PLoS Genet.* **4**, e236 (2008).
- Devlin, B. & Roeder, K. Genomic control for association studies. *Biometrics* **55**, 997–1004 (1999).
- Dudbridge, F. & Gusnanto, A. Estimation of significance thresholds for genomewide association scans. *Genet. Epidemiol.* **32**, 227–234 (2008).
- Stefansson, H. *et al.* Common variants conferring risk of schizophrenia. *Nature* **460**, 744–747 (2009).
- Shi, J. *et al.* Common variants on chromosome 6p22.1 are associated with schizophrenia. *Nature* **460**, 753–757 (2009).
- Ferreira, M.A. *et al.* Collaborative genome-wide association analysis supports a role for *ANKK3* and *CACNA1C* in bipolar disorder. *Nat. Genet.* **40**, 1056–1058 (2008).
- Green, E.K. *et al.* The bipolar disorder risk allele at *CACNA1C* also confers risk of recurrent major depression and of schizophrenia. *Mol. Psychiatry* **15**, 1016–1022 (2010).
- Ruderfer, D.M. *et al.* A family-based study of common polygenic variation and risk of schizophrenia. *Mol. Psychiatry* published online, doi:10.1038/mp.2011.34 (12 April 2011).
- Bemis, L.T. *et al.* MicroRNA-137 targets microphthalmia-associated transcription factor in melanoma cell lines. *Cancer Res.* **68**, 1362–1368 (2008).
- Srikantan, V. *et al.* *PCGEM1*, a prostate-specific gene, is overexpressed in prostate cancer. *Proc. Natl. Acad. Sci. USA* **97**, 12216–12221 (2000).
- Szulwach, K.E. *et al.* Cross talk between microRNA and epigenetic regulation in adult neurogenesis. *J. Cell Biol.* **189**, 127–141 (2010).
- Silber, J. *et al.* miR-124 and miR-137 inhibit proliferation of glioblastoma multiforme cells and induce differentiation of brain tumor stem cells. *BMC Med.* **6**, 14 (2008).
- Smrt, R.D. *et al.* MicroRNA miR-137 regulates neuronal maturation by targeting ubiquitin ligase mind bomb-1. *Stem Cells* **28**, 1060–1070 (2010).
- Potkin, S.G. *et al.* Identifying gene regulatory networks in schizophrenia. *Neuroimage* **53**, 839–847 (2010).
- Lewis, B.P., Burge, C.B. & Bartel, D.P. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* **120**, 15–20 (2005).
- Krek, A. *et al.* Combinatorial microRNA target predictions. *Nat. Genet.* **37**, 495–500 (2005).
- John, B. *et al.* Human MicroRNA targets. *PLoS Biol.* **2**, e363 (2004).
- Yang, J. *et al.* Common SNPs explain a large proportion of the heritability for human height. *Nat. Genet.* **42**, 565–569 (2010).
- Yang, J. *et al.* Genomic inflation factors under polygenic inheritance. *Eur. J. Hum. Genet.* **19**, 807–812 (2011).
- Visscher, P.M., Yang, J. & Goddard, M.E. A commentary on 'common SNPs explain a large proportion of the heritability for human height' by Yang *et al.* (2010). *Twin Res. Hum. Genet.* **13**, 517–524 (2010).
- Lee, S.H., Wray, N.R., Goddard, M.E. & Visscher, P.M. Estimating missing heritability for disease from genome-wide association studies. *Am. J. Hum. Genet.* **88**, 294–305 (2011).
- Speliotes, E.K. *et al.* Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nat. Genet.* **42**, 937–948 (2010).
- Wei, Z. *et al.* From disease association to risk assessment: an optimistic view from genome-wide association studies on type 1 diabetes. *PLoS Genet.* **5**, e1000678 (2009).
- Bush, W.S. *et al.* Evidence for polygenic susceptibility to multiple sclerosis—the shape of things to come. *Am. J. Hum. Genet.* **86**, 621–625 (2010).
- Lichtenstein, P. *et al.* Common genetic determinants of schizophrenia and bipolar disorder in Swedish families: a population-based study. *Lancet* **373**, 234–239 (2009).
- Scott, L.J. *et al.* Genome-wide association and meta-analysis of bipolar disorder in individuals of European ancestry. *Proc. Natl. Acad. Sci. USA* **106**, 7501–7506 (2009).

Stephan Ripke¹, Alan R Sanders^{2,3}, Kenneth S Kendler^{4–6}, Douglas F Levinson⁷, Pamela Sklar^{1,8}, Peter A Holmans^{9,10}, Dan-Yu Lin¹¹, Jubao Duan^{2,3}, Roel A Ophoff^{12–15}, Ole A Andreassen^{16,17}, Edward Scolnick¹⁸, Sven Cichon^{19–21}, David St. Clair²², Aiden Corvin²³, Hugh Gurling²⁴, Thomas Werge²⁵, Dan Rujescu²⁶, Douglas H R Blackwood²⁷, Carlos N Pato²⁸, Anil K Malhotra^{29–31}, Shaun Purcell¹⁸, Frank Dudbridge³², Benjamin M Neale¹⁸, Lizzy Rossin¹, Peter M Visscher³³, Danielle Posthuma^{34,35}, Douglas M Ruderfer¹, Ayman Fanous^{5,36,37}, Hreinn Stefansson³⁸, Stacy Steinberg³⁸, Bryan J Mowry^{39,40}, Vera Golimbet⁴¹, Marc De Hert⁴², Erik G Jönsson⁴³, István Bitter⁴⁴, Olli P H Pietiläinen^{45,46}, David A Collier⁴⁷, Sarah Tosato⁴⁸, Ingrid Agartz^{16,49}, Margot Albus⁵⁰, Madeline Alexander⁷, Richard L Amdur^{36,37}, Farooq Amin^{51,52}, Nicholas Bass²⁴, Sarah E Bergen¹, Donald W Black⁵³, Anders D Børglum^{54,55}, Matthew A Brown⁵⁶, Richard Bruggeman⁵⁷, Nancy G Buccola⁵⁸, William F Byerley^{59,60}, Wiepke Cahn⁶¹, Rita M Cantor^{14,15}, Vaughan J Carr⁶², Stanley V Catts⁶³, Khalid Choudhury²⁴, C Robert Cloninger⁶⁴, Paul Cormican²³, Nicholas Craddock^{9,10}, Patrick A Danoy⁵⁶, Susmita Datta²⁴, Lieuwe de Haan⁶⁵, Ditte Demontis⁵⁴, Dimitris Dikeos⁶⁶, Srdjan Djurovic^{16,67}, Peter Donnelly^{68,69}, Gary Donohoe²³, Linh Duong²⁵, Sarah Dwyer^{9,10}, Anders Fink-Jensen⁷⁰, Robert Freedman⁷¹, Nelson B Freimer¹⁴, Marion Friedl²⁶, Lyudmila Georgieva^{9,10}, Ina Giegling²⁶, Michael Gill²³, Birte Glenthøj⁷², Stephanie Godard⁷³, Marian Hamshere^{9,10}, Mark Hansen⁷⁴, Thomas Hansen²⁵, Annette M Hartmann²⁶, Frans A Henskens⁷⁵, David M Hougaard⁷⁶, Christina M Hultman⁷⁷, Andrés Ingason²⁵, Assen V Jablensky⁷⁸, Klaus D Jakobsen²⁵, Maurice Jay^{79,132}, Gesche Jürgens⁸⁰, René S Kahn⁶¹, Matthew C Keller⁸¹, Gunter Kenis⁸², Elaine Kenny²³, Yunjung Kim⁸³,

George K Kirov^{9,10}, Heike Konnerth²⁶, Bettina Konte²⁶, Lydia Krabbendam⁸⁴, Robert Krasucki²⁴, Virginia K Lasseter^{85,132}, Claudine Laurent⁷⁹, Jacob Lawrence²⁴, Todd Lencz²⁹⁻³¹, F Bernard Lerer⁸⁶, Kung-Yee Liang⁸⁷, Paul Lichtenstein⁷⁷, Jeffrey A Lieberman⁸⁸, Don H Linszen⁶⁵, Jouko Lönnqvist⁸⁹, Carmel M Loughland⁹⁰, Alan W Maclean²⁷, Brion S Maher⁴⁻⁶, Wolfgang Maier⁹¹, Jacques Mallet⁹², Pat Malloy²⁷, Manuel Mattheisen^{19,21,93}, Morten Mattingsdal^{16,94}, Kevin A McGhee²⁷, John J McGrath^{39,40}, Andrew McIntosh²⁷, Duncan E McLean⁹⁵, Andrew McQuillin²⁴, Ingrid Melle^{16,17}, Patricia T Michie^{96,97}, Vihra Milanova⁹⁸, Derek W Morris²³, Ole Mors⁵⁵, Preben B Mortensen⁹⁹, Valentina Moskvina^{9,10}, Pierandrea Muglia^{100,101}, Inez Myin-Germeys⁸⁴, Deborah A Nertney^{39,40}, Gerald Nestadt⁸⁵, Jimmi Nielsen¹⁰², Ivan Nikolov^{9,10}, Merete Nordentoft¹⁰³, Nadine Norton^{9,10}, Markus M Nöthen^{19,21}, Colm T O'Dushlaine²³, Ann Olincy⁷¹, Line Olsen²⁵, F Anthony O'Neill¹⁰⁴, Torben F Ørntoft^{105,106}, Michael J Owen^{9,10}, Christos Pantelis¹⁰⁷, George Papadimitriou⁶⁶, Michele T Pato²⁸, Leena Peltonen^{45,46,108,132}, Hannes Petursson¹⁰⁹, Ben Pickard¹¹⁰, Jonathan Pimm²⁴, Ann E Pulver⁸⁵, Vinay Puri²⁴, Digby Quedsted¹¹¹, Emma M Quinn²³, Henrik B Rasmussen²⁵, János M Réthelyi⁴⁴, Robert Ribble⁴⁻⁶, Marcella Rietschel^{91,112}, Brien P Riley⁴⁻⁶, Mirella Ruggeri⁴⁸, Ulrich Schall^{97,113}, Thomas G Schulze^{112,114}, Sibylle G Schwab¹¹⁵⁻¹¹⁷, Rodney J Scott¹¹⁸, Jianxin Shi¹¹⁹, Engilbert Sigurdsson^{109,120}, Jeremy M Silverman^{8,121}, Chris C A Spencer⁶⁸, Kari Stefansson³⁸, Amy Strange⁶⁸, Eric Strengman^{12,13}, T Scott Stroup⁸⁸, Jaana Suvisaari⁸⁹, Lars Terenius⁴³, Srinivasa Thirumalai¹²², Johan H Thygesen²⁵, Sally Timm¹²³, Draga Toncheva¹²⁴, Edwin van den Oord¹²⁵, Jim van Os⁸⁴, Ruud van Winkel^{42,82}, Jan Veldink¹²⁶, Dermot Walsh¹²⁷, August G Wang¹²⁸, Durk Wiersma⁵⁷, Dieter B Wildenauer^{115,129}, Hywel J Williams^{9,10}, Nigel M Williams^{9,10}, Brandon Wormley⁴⁻⁶, Stan Zammit^{9,10}, Patrick F Sullivan^{77,83,130,131}, Michael C O'Donovan^{9,10}, Mark J Daly¹ & Pablo V Gejman^{2,3}

¹Center for Human Genetic Research, Massachusetts General Hospital, Boston, Massachusetts, USA. ²Department of Psychiatry and Behavioral Sciences, NorthShore University HealthSystem, Evanston, Illinois, USA. ³Department of Psychiatry and Behavioral Sciences, University of Chicago, Chicago, Illinois, USA. ⁴Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University School of Medicine, Richmond, Virginia, USA. ⁵Department of Psychiatry, Virginia Commonwealth University School of Medicine, Richmond, Virginia, USA. ⁶Department of Human and Molecular Genetics, Virginia Commonwealth University School of Medicine, Richmond, Virginia, USA. ⁷Department of Psychiatry and Behavioral Sciences, Stanford University, Stanford, California, USA. ⁸Department of Psychiatry, Mount Sinai School of Medicine, New York, New York, USA. ⁹Medical Research Council (MRC) Centre for Neuropsychiatric Genetics and Genomics, School of Medicine, Cardiff University, Cardiff, UK. ¹⁰Department of Psychological Medicine and Neurology, School of Medicine, Cardiff University, Cardiff, UK. ¹¹Department of Biostatistics, University of North Carolina, Chapel Hill, North Carolina, USA. ¹²Department of Medical Genetics, University Medical Center Utrecht, Utrecht, The Netherlands. ¹³Rudolf Magnus Institute of Neuroscience, University Medical Center Utrecht, Utrecht, The Netherlands. ¹⁴University of California at Los Angeles (UCLA) Center for Neurobehavioral Genetics, University of California at Los Angeles, Los Angeles, California, USA. ¹⁵Department of Human Genetics, University of California at Los Angeles, Los Angeles, California, USA. ¹⁶Psychiatry Section, Institute of Clinical Medicine, University of Oslo, Oslo, Norway. ¹⁷Department of Psychiatry, Oslo University Hospital, Oslo, Norway. ¹⁸Broad Institute, Cambridge, Massachusetts, USA. ¹⁹Department of Genomics, Life and Brain Center, University of Bonn, Bonn, Germany. ²⁰Institute of Neuroscience and Medicine (INM-1), Research Center Juelich, Juelich, Germany. ²¹Institute of Human Genetics, University of Bonn, Bonn, Germany. ²²Institute of Medical Sciences, University of Aberdeen, Foresterhill, Aberdeen, UK. ²³Neuropsychiatric Genetics Research Group, Trinity College Dublin, Dublin, Ireland. ²⁴Molecular Psychiatry Laboratory, Research Department of Mental Health Sciences, University College London Medical School, Windyer Institute of Medical Sciences, London, UK. ²⁵Institute of Biological Psychiatry, Mental Health Center (MHC) Sct. Hans, Copenhagen University Hospital, Roskilde, Denmark. ²⁶Molecular and Clinical Neurobiology, Department of Psychiatry, Ludwig-Maximilians-University, Munich, Germany. ²⁷Division of Psychiatry, University of Edinburgh, Royal Edinburgh Hospital, Edinburgh, UK. ²⁸Keck School of Medicine, University of Southern California, Los Angeles, California, USA. ²⁹Department of Psychiatry, Division of Research, The Zucker Hillside Hospital Division of the North Shore-Long Island Jewish Health System, Glen Oaks, New York, USA. ³⁰Center for Psychiatric Neuroscience, The Feinstein Institute for Medical Research, Manhasset, New York, USA. ³¹Department of Psychiatry and Behavioral Science, Albert Einstein College of Medicine of Yeshiva University, New York, New York, USA. ³²Department of Non-Communicable Disease Epidemiology, London School of Hygiene and Tropical Medicine, London, UK. ³³Queensland Statistical Genetics Laboratory, Queensland Institute of Medical Research, Brisbane, Queensland, Australia. ³⁴Vrije Universiteit (VU), Center for Neurogenomics and Cognitive Research (CNCR), Department of Functional Genomics, Amsterdam, The Netherlands. ³⁵VU Medical Centre, Department of Medical Genomics, Amsterdam, The Netherlands. ³⁶Washington Veteran's Affairs Medical Center, Washington, DC, USA. ³⁷Department of Psychiatry, Georgetown University School of Medicine, Washington, DC, USA. ³⁸deCODE Genetics, Reykjavik, Iceland. ³⁹Queensland Brain Institute, University of Queensland, Brisbane, Queensland, Australia. ⁴⁰Queensland Centre for Mental Health Research, University of Queensland, Brisbane, Queensland, Australia. ⁴¹Mental Health Research Center, Russian Academy of Medical Sciences, Moscow, Russia. ⁴²University Psychiatric Centre, Catholic University Leuven, Kortenberg, Belgium. ⁴³Department of Clinical Neuroscience, Human Brain Informatics (HUBIN) Project, Karolinska Institutet and Hospital, Stockholm, Sweden. ⁴⁴Semmelweis University, Department of Psychiatry and Psychotherapy, Budapest, Hungary. ⁴⁵Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, Finland. ⁴⁶Department of Medical Genetics, University of Helsinki, Helsinki, Finland. ⁴⁷Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, King's College, London, UK. ⁴⁸Section of Psychiatry and Clinical Psychology, University of Verona, Verona, Italy. ⁴⁹Department of Research, Diakonhjemmet Hospital, Oslo, Norway. ⁵⁰State Mental Hospital, Haar, Germany. ⁵¹Department of Psychiatry and Behavioral Sciences, Emory University, Atlanta, Georgia, USA. ⁵²Department of Psychiatry and Behavioral Sciences, Atlanta Veterans Affairs Medical Center, Atlanta, Georgia, USA. ⁵³Department of Psychiatry, University of Iowa Carver College of Medicine, Iowa City, Iowa, USA. ⁵⁴Institute of Human Genetics, University of Aarhus, Aarhus, Denmark. ⁵⁵Centre for Psychiatric Research, Aarhus University Hospital, Risskov, Denmark. ⁵⁶University of Queensland Diamantina Institute, Princess Alexandra Hospital, University of Queensland, Brisbane, Queensland, Australia. ⁵⁷University Medical Center Groningen, Department of Psychiatry, University of Groningen, Groningen, The Netherlands. ⁵⁸School of Nursing, Louisiana State University Health Sciences Center, New Orleans, Louisiana, USA. ⁵⁹Department of Psychiatry, University of California at San Francisco, San Francisco, California, USA. ⁶⁰NCIRE (Northern California Institute for Research and Education), San Francisco, California, USA. ⁶¹Department of Psychiatry, Rudolf Magnus Institute of Neuroscience, University Medical Center Utrecht, Utrecht, The Netherlands. ⁶²School of Psychiatry, University of New South Wales and Schizophrenia Research Institute, Sydney, New South Wales, Australia. ⁶³Department of Psychiatry, University of Queensland, Royal Brisbane Hospital, Brisbane, Australia. ⁶⁴Department of Psychiatry, Washington University, St. Louis, Missouri, USA. ⁶⁵Academic Medical Centre, University of Amsterdam, Department of Psychiatry, Amsterdam, The Netherlands. ⁶⁶Department of Psychiatry, University of Athens Medical School, Athens, Greece. ⁶⁷Department of Medical Genetics, Oslo University Hospital, Oslo, Norway. ⁶⁸Wellcome Trust Centre for Human Genetics, Oxford, UK. ⁶⁹Department of Statistics, University of Oxford, Oxford, UK. ⁷⁰Mental Health Center Copenhagen, Copenhagen University Hospital, Copenhagen, Denmark. ⁷¹Department of Psychiatry, University of Colorado Denver, Aurora, Colorado, USA. ⁷²Center for Clinical Intervention and Neuropsychiatric Schizophrenia Research, Mental Health Center Glostrup, Copenhagen University Hospital, Glostrup, Denmark. ⁷³INSERM, Institut de Myologie, Hôpital de la Pitié-Salpêtrière, Paris, France. ⁷⁴illumina, Inc., La Jolla, California, USA. ⁷⁵School of Electrical Engineering

and Computing Science, University of Newcastle, Newcastle, New South Wales, Australia. ⁷⁶Section of Neonatal Screening and Hormones, Department of Clinical Chemistry and Immunology, The State Serum Institute, Copenhagen, Denmark. ⁷⁷Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden. ⁷⁸Centre for Clinical Research in Neuropsychiatry, School of Psychiatry and Clinical Neurosciences, The University of Western Australia, Perth, Western Australia, Australia. ⁷⁹Department of Child and Adolescent Psychiatry, Pierre and Marie Curie Faculty of Medicine, Paris, France. ⁸⁰Department of Clinical Pharmacology, Bispebjerg University Hospital, Copenhagen, Denmark. ⁸¹Department of Psychology, University of Colorado, Boulder, Boulder, Colorado, USA. ⁸²Department of Psychiatry and Psychology, School of Mental Health and Neuroscience, European Graduate School of Neuroscience (EURON), South Limburg Mental Health Research and Teaching Network (SEARCH), Maastricht University Medical Centre, Maastricht, The Netherlands. ⁸³Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA. ⁸⁴Maastricht University Medical Centre, South Limburg Mental Health Research and Teaching Network, Maastricht, The Netherlands. ⁸⁵Department of Psychiatry and Behavioral Sciences, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA. ⁸⁶Department of Psychiatry, Hadassah-Hebrew University Medical Center, Jerusalem, Israel. ⁸⁷Department of Biostatistics, Johns Hopkins University Bloomberg School of Public Health, Baltimore, Maryland, USA. ⁸⁸Department of Psychiatry, Columbia University, New York, New York, USA. ⁸⁹Department of Mental Health and Substance Abuse Services, National Institute for Health and Welfare, Helsinki, Finland. ⁹⁰Schizophrenia Research Institute, Sydney and Centre for Brain and Mental Health Research, University of Newcastle, Newcastle, New South Wales, Australia. ⁹¹Department of Psychiatry, University of Bonn, Bonn, Germany. ⁹²Laboratoire de Génétique Moléculaire de la Neurotransmission et des Processus Neurodégénératifs, Centre National de la Recherche Scientifique, Hôpital de la Pitié Salpêtrière, Paris, France. ⁹³Institute of Medical Biometry, Informatics and Epidemiology (IMBIE), University of Bonn, Bonn, Germany. ⁹⁴Department of Research, Sørlandet Hospital, Kristiansand, Norway. ⁹⁵Queensland Centre for Mental Health Research, The Park Centre for Mental Health, Wacol, Queensland, Australia. ⁹⁶Functional Neuroimaging Laboratory, School of Psychology, University of Newcastle, Sydney, New South Wales, Australia. ⁹⁷Schizophrenia Research Institute, Sydney, New South Wales, Australia. ⁹⁸Department of Psychiatry, First Psychiatric Clinic, Alexander University Hospital, Sofia, Bulgaria. ⁹⁹National Centre for Register-Based Research, University of Aarhus, Aarhus, Denmark. ¹⁰⁰Department of Psychiatry, University of Toronto, Toronto, Canada. ¹⁰¹NeuroSearch A/S, Ballerup, Denmark. ¹⁰²Unit for Psychiatric Research, Aalborg Psychiatric Hospital, Aalborg, Denmark. ¹⁰³Psychiatric Centre Copenhagen, Copenhagen University Hospital, Copenhagen, Denmark. ¹⁰⁴Department of Psychiatry, Queens University, Belfast, Ireland. ¹⁰⁵ARoS Applied Biotechnology A/S, Skejby, Denmark. ¹⁰⁶Department of Molecular Medicine, Aarhus University Hospital, Skejby, Denmark. ¹⁰⁷Melbourne Neuropsychiatry Centre, Department of Psychiatry, The University of Melbourne and Melbourne Health, Melbourne, Victoria, Australia. ¹⁰⁸Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, UK. ¹⁰⁹Department of Psychiatry, National University Hospital, Reykjavik, Iceland. ¹¹⁰Strathclyde Institute of Pharmacy and Biomedical Sciences, The John Arbuthnott Building, University of Strathclyde, Glasgow, UK. ¹¹¹Department of Psychiatry, University of Oxford, Warneford Hospital, Headington, Oxford, UK. ¹¹²Department of Genetic Epidemiology in Psychiatry, Central Institute of Mental Health, University of Heidelberg, Mannheim, Germany. ¹¹³Priority Centre for Brain and Mental Health Research, University of Newcastle, Sydney, New South Wales, Australia. ¹¹⁴Department of Psychiatry and Psychotherapy, Georg-August-University, Göttingen, Germany. ¹¹⁵School of Psychiatry and Clinical Neurosciences, University of Western Australia, Perth, Western Australia, Australia. ¹¹⁶Department of Psychiatry, University of Erlangen-Nuremberg, Erlangen, Germany. ¹¹⁷Centre for Medical Research, Western Australian Institute for Medical Research, University of Western Australia, Perth, Western Australia, Australia. ¹¹⁸Centre for Information Based Medicine, University of Newcastle, Hunter Medical Research Institute, Newcastle and Schizophrenia Research Institute, Sydney, New South Wales, Australia. ¹¹⁹Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, Maryland, USA. ¹²⁰Department of Psychiatry, University of Iceland, Reykjavik, Iceland. ¹²¹Department of Psychiatry, Veterans Affairs Medical Center, New York, New York, USA. ¹²²West Berkshire National Health Service (NHS) Trust, Reading, UK. ¹²³Mental Health Center Frederiksberg, Copenhagen University Hospital, Copenhagen, Denmark. ¹²⁴Department of Medical Genetics, University Hospital Maichin Dom, Sofia, Bulgaria. ¹²⁵Department of Pharmacy, Virginia Commonwealth University, Richmond, Virginia, USA. ¹²⁶Rudolf Magnus Institute of Neuroscience, Department of Neurology, Universitair Medisch Centrum (UMC) Utrecht, Utrecht, The Netherlands. ¹²⁷The Health Research Board, Dublin, Ireland. ¹²⁸Mental Health Center Amager, Copenhagen University Hospital, Copenhagen, Denmark. ¹²⁹Centre for Clinical Research in Neuropsychiatry, Graylands Hospital, Mt Claremont, Western Australia, Australia. ¹³⁰Department of Psychiatry, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA. ¹³¹Department of Epidemiology, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA. ¹³²Deceased. Correspondence should be addressed to P.V.G. (pgejman@gmail.com).

ONLINE METHODS

Overview. Seventeen samples representing 11 countries (Table 1 and Supplementary Table 1) comprising all known European-ancestry GWAS samples were invited to participate in stage 1. The quality of the clinical data was verified by a systematic review of data collection methods and procedures at each site (Table 1, Supplementary Table 13 and Supplementary Note).

Quality control. Technical quality control was performed on genotypes generated for stage 1 by various GWAS platforms (Table 1), with quality control conducted separately using a common approach; all datasets were separately analyzed (Supplementary Table 3). The SNPs successfully genotyped in each study and common to all platforms ($N = 11,310$) were pruned to remove high LD and lower frequency SNPs and were then used for relatedness testing in each sample and in the combined total sample. Common quality control parameters were applied: (i) missing rate per SNP < 0.05 (before sample removal below), (ii) missing rate per individual < 0.02 , (iii) missing rate per SNP < 0.02 (after sample removal above), (iv) missing rate per SNP difference in cases and controls < 0.02 , (v) SNP frequency difference to HapMap < 0.15 and (vi) Hardy-Weinberg equilibrium (controls) $P < 10^{-6}$. Stage 1 study sample sizes varied between 400 and 5,000 individuals (Table 1 and Supplementary Table 3). The number of SNPs per study after quality control varied between 250,000 and 680,000. On average, the quality control processes excluded 15 individuals per study (with a range of 0–100 individuals) and 38,000 SNPs per study (with a range of 5,000–160,000 SNPs). After quality control, the stage 1 GWAS datasets together comprised 21,856 individuals and, for the next steps of the ‘genetic quality control’ analysis, a set of 17,074 SNPs common to all platforms and successfully genotyped in each stage 1 sample was extracted. These SNPs were then further pruned to remove LD (leaving no pairs with $r^2 > 0.05$) and lower frequency SNPs (minor allele frequency < 0.05), leaving 11,310 SNPs suitable for robust relatedness testing and population structure analysis.

Genetic quality control. Genetic quality control included relatedness testing and principal components analyses. Relatedness testing was done with PLINK (see URLs), reporting pairs with genome identity (π -hat) > 0.9 as ‘identical samples’ and with π -hat > 0.2 as being closely related. After random shuffling, one SNP from each pair was excluded from downstream analysis. From groups with multiple related pairs (for example, a family), only one individual was kept. After quality control, the PGC GWAS European-ancestry sample (stage 1) consisted of 9,394 cases with schizophrenia or schizoaffective disorder and 12,462 controls (52% of whom were psychiatrically screened). Principal component estimation was done with the same collection of SNPs on the non-related subset of individuals. We estimated the first 20 principal components and tested each of them for phenotype association (using logistic regression with study indicator variables included as covariates) and evaluated their impact on the genome-wide test statistics using λ (the genomic control inflation factor based on the median χ^2) after genome-wide association of the specified principal component. Based on this we decided to include principal components 1, 2, 3, 4 and 6 for downstream analysis as associated covariates. The deviation noted in Supplementary Figure 1 was observed more strongly among SNPs with large numbers of neighboring SNPs in high LD (Supplementary Fig. 7); SNPs that tag more variants are more likely to capture true genetic effects.

Imputation. Imputation of untyped SNPs was performed within each study in batches of 300 individuals. These batches were randomly drawn in order to keep the same case-control ratio as in the total sample from that study. We used Beagle 3.0.4 (ref. 31). Imputation was performed with CEU+TSI HapMap phase 3 data (UCSC hg18/NCBI 36)¹ with 410 phased haplotypes encompassing 1,252,901 SNPs. λ was carefully monitored before and after imputation. Each of the 17,074 SNPs present on every platform was dropped and reimputed with high correlation (Pearson $r^2 = 0.99973$).

Stage 1 association analyses. Stage 1 association analyses of the phenotype were performed with the imputed dosages using standard logistic regression, which intrinsically properly addresses uncertainty in the imputation results. As covariates, we used the study indicator as a categorical variable and five principal components derived from the population.

SNPs from the stage 1 association results were grouped into LD regions using a simple iterative process after all SNPs in the genome were ranked from most to least significant: if a SNP had $r^2 > 0.2$ (defined by Haploview (see URLs) analysis of HapMap3 CEU+TSI data) to a more significantly associated SNP within 1 Mb on either side of it, the SNP was assigned to the region defined by the more associated SNP. Distinct regions defined by associated variants physically within ± 200 kb were tested using a conditional analysis to examine the independence of the associations in the PGC schizophrenia data. LD regions with $P < 2 \times 10^{-5}$ are reported in Supplementary Table 6 and were assessed in independent stage 2 (replication) samples.

Power analysis. The power analysis was performed using the Genetic Power Calculator (GPC)³² to calculate the power of the discovery sample (all 17 stage 1 samples together). Supplementary Table 12 shows the power of this sample to detect association at three significance thresholds: a canonical genome-wide significance threshold ($P < 5 \times 10^{-8}$), the threshold we used for follow up ($P < 2 \times 10^{-5}$) and a more modest threshold of $P < 0.001$. The middle of these is most relevant to the power of this study because this threshold represents a level that would have been used to advance to follow up in the larger replication study (stage 2) and reported in this manuscript. Presented are power figures at representative and relevant odds ratios (ORs) (OR = 1.10, OR = 1.15 and OR = 1.20) and allele frequencies (0.05, 0.10, 0.20 and 0.40). Notably, although the power of the sample sizes used in this study to identify loci at the significance threshold used for follow up at ORs greater than 1.15 was high, the power noticeably dropped at an OR of 1.10, where the majority of our confirmed findings were made.

Stage 2 association analyses. Stage 2 association analyses were performed using all accessible replication European-ancestry samples (stage 2), which consisted of 8,442 cases and 21,397 controls (37% of which were screened) in 19 samples representing 14 countries (Table 1 and Supplementary Table 2). We tested the top SNP in each LD region in the replication samples either by looking up the data in GWAS datasets not involved in the PGC analysis or by focused genotyping in unscanned samples (Supplementary Note). Where new genotyping was required, data were provided for the top 15–30 SNPs (depending on available resources) by collaborators. After receiving P values, ORs, standard errors and allele frequencies for the replication datasets, the results were aligned, and a standard error-weighted meta-analysis was performed of (i) the replication experiment (stage 2) alone and (ii) the combination of the PGC discovery and replication data (stages 1 and 2). An analysis of the combination of stages 1 and 2 was also performed with the genomic control-adjusted version of the stage 1 results by converting the genomic control-adjusted P value for stage 1 back into an adjusted z score, which was combined with the stage 2 meta-analysis result (Table 2).

Conditional analysis. A conditional analysis was performed for all regions with $P < 10^{-6}$ in a logistic regression framework using as covariates the dosages of the target SNP of the region, multi-dimensional scaling scores 1–4 and 6 and the study indicator. Any $P < 10^{-4}$ was then assumed to be independent from the main signal in this region. A plot of these regions after conditioning is presented in Supplementary Figure 5. Within a narrow range of ± 200 kb (which is similar to distances used in the analysis of other phenotypes manifesting complex genetics, for example, height³) of the original target SNP, we found three regions with a clear sign (permutation-corrected $P < 0.05$) for one or more independent signals (Supplementary Table 5).

Combined analysis of top-associated SNPs with PGC bipolar disorder. This combined analysis was performed because after the main analysis of the PGC schizophrenia and the PGC bipolar disorder datasets was finalized, we noticed several regions associated at high significance levels in both disorder groups. We therefore applied the same methods as already described for the stage 1 PGC schizophrenia analyses. This included the removal of one individual of each overlapping subject pair (see the description of ‘genetic quality control’). After creating principal components from the SNPs common to each dataset (see ‘genetic quality control’), we tested for association 16,374 joint cases (cases with schizophrenia, schizoaffective disorder or bipolar disorder) and 14,044 joint controls, including principal components 1–7 and

20 and the study indicator as covariates (see 'stage 1 association analyses'). We thereafter focused on regions reported by bipolar disorder GWAS. After LD pruning, we report in **Supplementary Table 11** the three genome-wide significant regions.

Additional analyses. Additional analyses performed included an interaction analysis (see below and **Supplementary Table 14**), stage 1 association analyses with stricter control for population stratification (**Supplementary Figs. 2, 3,8,9** and **Supplementary Note**) and a score analysis to test for a polygenic model of inheritance (**Supplementary Table 9, Supplementary Fig. 6** and **Supplementary Note**). An interaction analysis of associated SNPs was performed between each pair of loci in **Table 2** in a logistic regression model that included the main effect of each SNP. No interaction terms were significant when correcting for the number of pairs tested (**Supplementary**

Table 14). Besides the overall stage 1 quantile-quantile and Manhattan plots (**Supplementary Figs. 1,4**), we present individual stage 1 sample quantile-quantile and Manhattan plots (**Supplementary Figs. 10,11**). To complement the principal component analysis plots for individual stage 1 samples (**Supplementary Fig. 8**), we also present multi-dimensional scaling for all stage 1 samples compared to HapMap3 anchors (**Supplementary Fig. 12**).

31. Browning, S.R. & Browning, B.L. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am. J. Hum. Genet.* **81**, 1084–1097 (2007).
32. Purcell, S., Cherny, S.S. & Sham, P.C. Genetic Power Calculator: design of linkage and association genetic mapping studies of complex traits. *Bioinformatics* **19**, 149–150 (2003).

