Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near *ODZ4*

Psychiatric GWAS Consortium Bipolar Disorder Working Group[1]

We conducted a combined genome-wide association study (GWAS) of 7,481 individuals with bipolar disorder (cases) and 9,250 controls as part of the Psychiatric GWAS Consortium. Our replication study tested 34 SNPs in 4,496 independent cases with bipolar disorder and 42,422 independent controls and found that 18 of 34 SNPs had $P < 0.05$, with 31 of 34 SNPs having signals with the same direction of effect ($P = 3.8 \times 10^{-7}$). An analysis of all 11,974 bipolar disorder cases and 51,792 controls confirmed genome-wide significant evidence of association for *CACNA1C* and identified a new intronic variant in *ODZ4*. We identified a pathway comprised of subunits of calcium channels enriched in bipolar disorder association intervals. Finally, a combined GWAS analysis of schizophrenia and bipolar disorder yielded strong association evidence for SNPs in *CACNA1C* and in the region of *NEK4-ITIH1-ITIH3-ITIH4*. Our replication results imply that increasing sample sizes in bipolar disorder will confirm many additional loci.

Bipolar disorder is a severe mood disorder affecting more than 1% of the population[1]. Family, twin and adoption studies consistently have found relative risks to first-degree relatives of affected individuals of ~8 and a concordance of ~40–70% for a monozygotic co-twin of affected individuals[1,2]. Bipolar disorder shares phenotypic similarities with other psychiatric diseases, and relatives of individuals with bipolar disorder are at an increased risk of schizophrenia, major depression and schizoaffective disorder, suggesting a partially shared genetic basis for these disorders[3,4]. Despite robust evidence for heritability, causal mutations have not been identified through linkage or candidate gene association studies[1].

GWAS for bipolar disorder have been performed with multiple partially overlapping samples[5–11]. In a previous small study, researchers reported a genome-wide significant association to *DGKH* (encoding diacylglycerol kinase eta)[5]. Subsequently, researchers in another study[8] identified the region of *ANK3* (encoding ankyrin 3) and those from a third[12] recently reported an association to *NCAN* (encoding neurocan); other studies did not report genome-wide significant loci[5,9,10,13]. The Psychiatric GWAS Consortium (PGC) was established to facilitate the combination of primary genotype data and to allow analyses both within and across psychiatric disorders[14,15]. Here, the PGC Bipolar Disorder Working Group reports results from our association study of bipolar disorder from 16,731 samples and a replication sample of 46,918 individuals.

We received primary genotype and phenotype data (**Table 1**, **Supplementary Note** and **Supplementary Table 1**). Results have been reported singly[6,7,9–11] and in combinations[8,9,12] in seven publications with case and control overlap. We divided the data into 11 case and control groupings and assigned each individual to only one group (**Table 1**). The final dataset comprised 7,481 unique cases and 9,250 unique controls. Cases had the following diagnoses: bipolar disorder type 1 (n = 6,289; 84%), bipolar disorder type 2 (n = 824; 11%) and schizoaffective disorder bipolar (n = 263; 4%), and there were 105 individuals with other bipolar diagnoses (1%). We directly genotyped 46,234 SNPs in all 11 groups and genotyped 1,016,924 SNPs in between 2 and 11 groups. Based on reference haplotypes from the HapMap phase 2 CEU sample, we imputed missing genotypes using BEAGLE[16]. We analyzed imputed SNP dosages from 2,415,422 autosomal SNPs with a minor allele frequency ≥1% and imputation quality score $r^2 > 0.3$. We performed logistic regression of case status on imputed SNP dosage including as covariates five multidimensional scaling components (**Supplementary Fig. 1**) and indicator variables for the sample group using PLINK[17]. The genomic control[18] inflation factor (λ_{GC}) was 1.148. Consistent with previous work suggesting a highly polygenic architecture for schizophrenia and bipolar disorder[19], this estimate likely reflects a mixture of signals arising from a large number of true risk variants of weak effect and some degree of residual confounding. Nonetheless, we designated an association as genome-wide significant only if the genomic-control P value (P_{GC}) was below 5×10^{-8} (the nominal P values are labeled 'P_{raw}'). Results for the primary analyses can be found in **Supplementary Figures 2–4**. **Supplementary Table 2** lists regions containing SNPs with $P_{GC} < 5 \times 10^{-5}$.

Table 2 lists four regions from our primary GWAS analysis that contain SNPs with $P_{raw} < 5 \times 10^{-8}$; two regions reached $P_{GC} \leq 5 \times 10^{-8}$ (**Supplementary Fig. 4**). We detected association in *ANK3* (encoding ankyrin 3) on chromosome 10q21 for the imputed SNP rs10994397 ($P_{GC} = 7.1 \times 10^{-9}$, odds ratio (OR) = 1.35). The second SNP, rs9371601, was located in *SYNE1* (encoding synaptic nuclear envelope protein 1) on chromosome 6q25 ($P_{GC} = 4.3 \times 10^{-8}$, OR = 1.15). The intergenic SNP rs7296288 ($P_{GC} = 8.4 \times 10^{-8}$, OR = 1.15) is in a region of linkage disequilibrium (LD) of ~100 kb on chromosome 12q13 that contains seven genes. rs12576775 (P_{GC} = 2.1 ×

[1]A full list of members and affiliations appears at the end of the paper.

Received 28 March; accepted 23 August; published online 18 September 2011; corrected after print 21 June 2012; doi:10.1038/ng.943

Systematic Treatment Enhancement Program for Bipolar Disorder (STEP1)	European-American	922	645	500K	7,8
Systematic Treatment Enhancement Program for Bipolar Disorder (STEP2)	European-American	659	192	5.0	8
Thematically Organized Psychosis (TOP) Study	Norwegian	203	349	6.0	11
Trinity College Dublin	Irish	150	797	6.0	8
University College London (UCL)	British	457	495	500K	7,8
University of Edinburgh	Scottish	282	275	6.0	8
Wellcome Trust Case-Control Consortium (WTCCC)	British	1,571	2,931	500K	6,8,9
TOTAL		**7,481**	**9,250**		

[a]Cases include BD1, BD2, SAB, BD-NOS (see **Supplementary Table 1**). [b]Most controls were not screened for psychiatric disease. A subset of 33%, however, were screened, see the **Supplementary Note**. [c]Platforms are 6.0, Affymetrix Genome-Wide Human SNP Array 6.0; 5.0, Affymetrix Genome-Wide Human SNP Array 5.0; 500K, Affymetrix GeneChip Human Mapping 500K Array; 550, Illumina HumanHap 550. [d]Primary publication reporting individual sample level genotypes for bipolar disorder are listed. See the **Supplementary Note** for a fuller description of publications and **Supplementary Table 1** for the sample origins in the primary GWAS analyses.

10^{-7}, OR = 1.18) is found at chromosome 11q14 in the first intron of *ODZ4*, a human homolog of the *Drosophila* pair-rule gene *ten-m* (*odz*). We observed generally consistent signals, with no single study driving the overall association results (**Supplementary Fig. 5**). A meta-analysis, under both fixed- and random-effects models, yielded similar results (**Supplementary Tables 3** and **4**).

We sought to replicate these findings in independent samples. We selected 38 SNPs with $P_{GC} < 5 \times 10^{-5}$ (**Supplementary Table 2**). Of these, four SNPs were not completely independent signals, and we did not use these SNPs for further analyses (Online Methods). We received unpublished data from investigators on a further 4,496 cases and 42,422 controls for the top 34 independent SNPs (**Supplementary Table 5**). Significantly more SNPs replicated than would be expected by chance (**Table 3**). Four of 34 SNPs had replication $P < 0.01$, 18 of 34 SNPs had replication $P < 0.05$, and 31 of 34 SNPs had a signal in the same direction of effect (binomial test $P = 3.8 \times 10^{-7}$). Within the replication samples, two SNPs remained significant following multiple testing correction. The first SNP, rs4765913, is on chromosome 12 in *CACNA1C*, which encodes the α subunit of the L-type voltage-gated calcium channel (replication $P = 1.6 \times 10^{-4}$, OR = 1.13). The second SNP, rs10896135, is in a 17-exon 98-kb open reading frame of *C11orf80* (replication $P = 0.0015$, OR = 0.91), and variants in several other genes are in strong linkage disequilibrium with this SNP. We obtained nominally significant replication *P* values in another gene encoding a calcium channel subunit, *CACNB3* (replication $P = 0.0025$, OR = 0.93). Only two of the four SNPs listed in **Table 2** had replication $P < 0.05$; the genome-wide significant SNPs rs10994397 and rs9371601 did not have $P < 0.05$ (replication $P = 0.12$ and $P = 0.10$, respectively). Finally, we performed a fixed-effects meta-analysis on our effect estimates from the primary and replication data and established genome-wide significant evidence for association with rs4765913 in *CACNA1C* ($P = 1.52 \times 10^{-8}$, OR = 1.14) and rs12576775 in *ODZ4* ($P = 4.40 \times 10^{-8}$, OR = 0.89) (**Fig. 1**). As in the primary analyses, we observed consistent signals, and a meta-analysis of the replication data did not reveal significant heterogeneity between the samples (**Supplementary Tables 6** and **7**).

To interpret why two significant associations found in the primary analysis fail to replicate, we quantified the effect of the 'winner's curse'. Given a polygenic model, power will be low to detect a particular variant at genome-wide significant levels, but there will be many chance opportunities to identify at least one variant. Simulation of the distribution of ORs around several 'true' ORs, conditioning on a genome-wide significant *P* value of 5×10^{-8}, a fixed minor allele frequency of 0.20 and our sample size showed a distinct inflation of the estimated OR, leading to a marked overestimate of the power to replicate an individual result (**Supplementary Table 8**). For example, for a true genotypic relative risk of 1.05, the mean estimated OR is 1.17 conditioning on $P < 5 \times 10^{-8}$. Although the nominal power for replication is 100% for the inflated OR, the true power to replicate at $P < 0.05$ is only 30%. Thus, any single replication failure is by itself less informative. This simulation is consistent with our observations of higher than expected rate of nominal replications.

We assessed enrichment of Gene Ontology (GO) terms for regions containing the top 34 independent SNPs listed in **Table 3** ($P_{GC} < 5 \times 10^{-5}$) using a permutation-based approach that controlled for potential biases caused by SNP density, gene density and gene size and found enrichment in GO:0015270, the category containing voltage gated calcium channel activity. This GO category contains eight genes, three of which (*CACNA1C*, *CACNA1D* and *CACNB3*) are present among the 34 independent association intervals tested ($P = 0.00002$); the probability of observing an empirical *P* value this small, given all the targets tested, is $P = 0.021$. Thus, intervals ranked highly in our GWAS likely do not represent a random set with respect to gene function. Studies based on a larger number of loci, defined by more liberal *P*-value cutoffs, could indicate other promising areas for biological investigation.

Table 2 Primary GWAS association results for four most significant regions

SNP	Chr.	Position[a]	Nearest gene	A1[b]/A2	A1 frequency[c]	OR[d] (95% CI)	P_{raw}	P_{GC}
rs10994397	10	61,949,130	*ANK3*[d]	T/C	0.06	1.35 (1.48–1.23)	5.5×10^{-10}	7.1×10^{-9}
rs9371601	6	152,832,266	*SYNE1*	T/G	0.36	1.15 (1.21–1.10)	4.3×10^{-9}	4.3×10^{-8}
rs7296288	12	47,766,235	Many	C/A	0.48	1.15 (1.20–1.09)	9.4×10^{-9}	8.4×10^{-8}
rs12576775	11	78,754,841	*ODZ4*	G/A	0.18	1.18 (1.25–1.11)	2.7×10^{-8}	2.1×10^{-7}

[a]SNP basepair position on Build 36. [b]A1, allele 1; A2, allele 2. [c]Allele frequency in the total sample. [d]Odds ratio (OR) is predicted toward allele A1. *ANK3* was previously reported[8]. Chr., chromosome.

Table 3 Association results for the primary GWAS, replication and combined samples

					Primary GWAS		Replication[a]		Combined GWAS and replication		
SNP	Chr.[b]	Position[c]	A1	A2	P_{GC}	OR[d]	$P_{1\text{-sided}}$	OR	P_{GC}	OR	Genes in the LD region
rs4765913	12	2,290,157	A	T	6.50×10^{-6}	1.15	1.60×10^{-4}	1.13	**1.52×10^{-8}**	**1.14**	*CACNA1C*
rs10896135	11	66,307,578	C	G	8.46×10^{-6}	0.88	1.47×10^{-3}	0.91	1.56×10^{-7}	0.89	*ZDHHC24, YIF1A, TMEM151A, SYT12, SPTBN2, SLC29A2, SF3B2, RIN1, RCE1, RBM4B, RBM4, RBM14, RAB1B, PELI3, PC, PACS1, NPAS4, MRPL11, LRFN4, KLC2, GAL3ST3, DPP3, CTSF, CNIH2, CD248, CCS, CCDC87, C11orf86, C11orf80, BRMS1, BBS1, B3GNT1, ACTN3*
rs2070615*	12	47,504,438	A	G	4.00×10^{-5}	0.90	2.52×10^{-3}	0.93	1.02×10^{-6}	0.91	*RND1, DDX23, CACNB3*
rs12576775	11	78,754,841	A	G	2.09×10^{-7}	0.85	7.59×10^{-3}	0.92	**4.40×10^{-8}**	**0.88**	*ODZ4*
rs2175420*	11	78,801,531	C	T	2.90×10^{-5}	0.87	7.80×10^{-3}	0.92	2.35×10^{-6}	0.89	*ODZ4*
rs3845817	2	65,612,029	C	T	1.65×10^{-5}	0.90	8.98×10^{-3}	0.94	1.76×10^{-6}	0.91	
rs2176528	2	194,580,428	C	G	3.98×10^{-5}	1.15	0.0104	1.09	3.71×10^{-6}	1.12	
rs4660531	1	41,612,409	G	T	3.16×10^{-5}	0.89	0.0111	0.93	3.44×10^{-6}	0.91	
rs7578035	2	98,749,324	G	T	1.83×10^{-5}	1.12	0.0129	1.06	2.77×10^{-6}	1.09	*TXNDC9, TSGA10, REV1, MRPL30, MITD1, MGAT4A, LYG1, LYG2, LIPT1, EIF5B, C2orf55, C2orf15*
rs2287921	19	53,920,084	C	T	1.68×10^{-5}	1.12	0.0137	1.06	3.08×10^{-6}	1.10	*SPHK2, SEC1, RPL18, RASIP1, NTN5, MAMSTR, IZUMO1, FUT2, FUT1, FGF21, FAM83E, DBP, CA11*
rs11168751*	12	47,505,405	C	G	1.80×10^{-5}	0.84	0.0143	0.90	2.51×10^{-6}	0.86	*CACNB3*
rs7296288	12	47,766,235	A	C	8.39×10^{-8}	0.87	0.0150	0.94	5.41×10^{-8}	0.90	*TUBA1B, TUBA1A, RHEBL1, PRKAG1, MLL2, LMBR1L, DHH, DDN*
rs7827290	8	142,369,497	G	T	3.54×10^{-5}	1.13	0.0167	1.06	8.75×10^{-6}	1.10	*LOC731779, GPR20*
rs12730292	1	79,027,350	C	G	2.37×10^{-5}	1.12	0.0171	1.06	5.02×10^{-6}	1.10	
rs12912251	15	36,773,660	G	T	9.57×10^{-6}	1.13	0.0204	1.06	3.27×10^{-6}	1.10	*C15orf53*
rs4332037	7	1,917,335	C	T	1.78×10^{-5}	0.87	0.0300	0.93	7.25×10^{-6}	0.90	*MAD1L1*
rs6550435	3	36,839,493	G	T	1.97×10^{-5}	1.12	0.0326	1.05	9.32×10^{-6}	1.09	*LBA1*
rs17395886	4	162,498,835	A	C	2.18×10^{-5}	0.86	0.0351	0.93	1.06×10^{-5}	0.89	*FSTL5*
rs6746896	2	96,774,676	A	G	2.33×10^{-6}	1.14	0.0386	1.05	2.36×10^{-6}	1.10	*LMAN2L, FER1L5, CNNM4*
rs736408	3	52,810,394	C	T	1.22×10^{-6}	1.14	0.0465	1.05	2.19×10^{-6}	1.10	*WDR82, TWF2, TNNC1, TMEM110, TLR9, STAB1, SPCS1, SNORD69, SNORD19, SNORD19B, SFMBT1, SEMA3G, RFT1, PRKCD, PPM1M, PHF7, PBRM1, NT5DC2, NISCH, NEK4, MUSTN1, LOC440957, ITIH1, ITIH3, ITIH4, GNL3, GLYCTK, GLT8D1, DNAH1, BAP1, ALAS1*
rs11162405	1	78,242,248	A	G	2.54×10^{-5}	0.90	0.0476	0.96	1.82×10^{-5}	0.92	*ZZZ3, USP33, NEXN, MGC27382, GIPC2, FUBP1, FAM73A, DNAJB4, AK5*
rs9804190	10	61,509,837	C	T	3.06×10^{-5}	1.17	0.0963	1.04	1.20×10^{-4}	1.10	*ANK3*
rs9371601	6	152,832,266	G	T	4.27×10^{-8}	0.87	0.103	0.97	6.71×10^{-7}	0.91	*SYNE1*
rs3774609	3	53,807,943	G	T	1.14×10^{-5}	0.89	0.107	0.97	3.73×10^{-5}	0.92	*CHDH, CACNA1D*
rs10994397	10	61,949,130	C	T	7.08×10^{-9}	0.74	0.116	0.94	3.08×10^{-7}	0.82	*ANK3*
rs4668059	2	168,874,528	C	T	4.45×10^{-5}	1.18	0.158	1.04	1.32×10^{-4}	1.12	*STK39*
rs16966413	15	36,267,191	A	G	4.74×10^{-5}	0.84	0.160	0.95	9.97×10^{-5}	0.88	*SPRED1*
rs6102917	20	40,652,833	C	G	3.88×10^{-5}	1.44	0.165	1.11	8.46×10^{-5}	1.31	*PTPRT*
rs11085829	19	13,035,312	A	G	4.03×10^{-6}	0.87	0.175	0.97	6.96×10^{-5}	0.92	*NFIX*
rs875326	1	173,556,022	C	T	2.51×10^{-5}	1.15	0.183	1.03	1.11×10^{-4}	1.10	*TNR*
rs13245097*	7	2,307,581	C	T	3.81×10^{-5}	1.13	0.196	1.02	3.24×10^{-4}	1.08	*SNX8, NUDT1, MAD1L1, FTSJ2*
rs780148	10	80,605,089	C	G	4.66×10^{-5}	1.12	0.230	1.03	1.40×10^{-4}	1.09	*ZMIZ1*
rs2281587	10	105,367,339	C	T	1.96×10^{-5}	1.12	0.372	1.01	3.78×10^{-4}	1.07	*SH3PXD2A, NEURL*
rs10776799	1	115,674,570	G	T	4.84×10^{-5}	1.15	0.434	1.01	1.25×10^{-3}	1.08	*NGF*
rs263906	1	101,750,922	C	T	2.42×10^{-5}	1.13	0.440	1.01	4.43×10^{-4}	1.08	
rs10028075	4	87,186,854	C	T	8.96×10^{-6}	0.89	1.00	1.02	2.04×10^{-3}	0.95	*MAPK10*
rs3968	9	4,931,997	C	G	2.09×10^{-5}	1.17	1.00	0.92	0.0174	1.04	
rs8006348	14	50,595,223	A	G	4.91×10^{-5}	0.89	1.00	1.05	0.0637	0.95	*TRIM9*

[a]SNP bp position on Build 36. [b]The odds ratio (OR) is predicted toward allele A1. [c]Replication case and control sample details can be found in the **Supplementary Note**; 4% of the controls were screened for psychiatric disorders. [d]Odds ratio. Bold indicates genome-wide statistical significance. P_{GC}, genomic control *P* value; Chr., chromosome; LD, linkage disequilibrium. *SNPs not representing completely independent signals (see Online Methods).

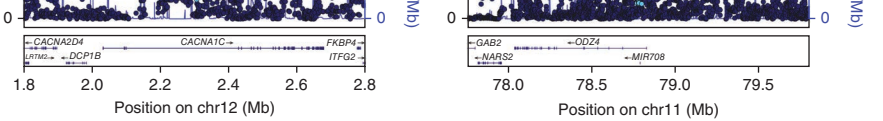

Figure 1 Results are shown as $-\log_{10} P$ for genotyped and imputed SNPs. The most associated SNP in the primary analysis is shown as a small purple triangle. The most associated SNP in the combined analysis is shown as a large purple triangle. The colors of the remaining markers reflects r^2 values with the most associated SNP. The recombination rate from CEU HapMap data (second *y* axis) is shown in light blue.

5′ end of *ANK3*, in an intron of *ANK3* and at the 3′ end of the longest transcript (705 kb in length). In each of these three regions, the association signals remaining after conditioning could arise from multiple causal variants, from a single rare causal variant in incomplete LD with the tested SNPs or could represent false-positive associations.

To provide direct evidence for a polygenic basis for bipolar disorder—as implied by the polygenic component shared between bipolar disorder and schizophrenia previously reported[19]—we repeated the analysis performed by the International Schizophrenia Consortium using bipolar disorder discovery samples. We observed enrichment of putatively associated bipolar disorder 'score alleles' in target sample cases compared to controls for all discovery *P* value thresholds (**Supplementary Table 9**).

A parallel study was performed by PGC schizophrenia investigators. We tested whether a combined analysis of PGC bipolar disorder and PGC schizophrenia data (eliminating known overlapping control samples) would show stronger association for the five most strongly associated SNPs supplemented by the additional genome-wide significant replication region in *CACNA1C*. In the combined bipolar disorder and schizophrenia analysis, two SNPs showed stronger association compared to the bipolar disorder GWAS alone: rs4765913 in *CACNA1C* (combined $P_{raw} = 7.7 \times 10^{-8}$ compared to bipolar disorder $P_{raw} = 1.35 \times 10^{-6}$) and rs736408 in a multigene region containing *NEK4-ITIH1-ITIH3-ITIH4* (combined $P_{raw} = 8.4 \times 10^{-9}$ compared to bipolar disorder $P_{raw} = 2.00 \times 10^{-7}$) (**Supplementary Table 10**).

In summary, we observed primary association signals that reached genome-wide significance in the regions of *ANK3* and *SYNE1* and two signals near genome-wide significance on chromosome 12 and in the region of *ODZ4*. Although in our independent replication sample we did not find additional support for *ANK3* or *SYNE1*, this is consistent with overestimation of the original ORs and should not be taken to disprove association. Data from additional samples are needed to resolve this.

The most notable finding is the abundance of replication signals. The number of nominal associations in the same direction of effect is highly unlikely to be a chance observation and strongly implies that many of the signals will ultimately turn out to be true associations. Such results are expected under a highly polygenic model, where there are few or no variants of large effect. As is typical in studies of complex genetic disorders, our findings explain only a small fraction of bipolar disorder heritability. Our data are consistent with many common susceptibility variants of relatively weak effect[19] potentially power to detect loci with relatively specific phenotypic effects.

A pathway analysis showed significant enrichment of *CACNA1C* and *CACNA1D*, which encode the major L-type α subunits found in the brain, consistent with a prior literature regarding the role of ion channels in bipolar disorder, the mood stabilizing effects of ion channel modulating drugs and the specific treatment literature suggesting direct efficacy of L-type calcium channel blockers in the treatment of bipolar disorder[21]. The *CACNA1C* SNP rs1006737 has been associated with several alterations in structural[22] and functional magnetic resonance imaging[23–25]. Several groups have previously implicated *CACNA1C* in other adult psychiatric disorders, in particular, schizophrenia and major depression[26–29]. L-type calcium channels regulate changes in gene regulation responsible for many aspects of neuronal plasticity and may have direct effects on transcription[29]. Taken together, this should lead to renewed biological investigation of calcium channels in psychiatric disease. *ODZ4*, located on chromosome 11, encodes a member of a family of cell surface proteins, the teneurins, and is related to the *Drosophila* pair-rule gene *ten-m* (*odz*). These genes are likely involved in cell surface signaling and neuronal pathfinding.

Three of our top five regions harbor non-coding RNAs. miR-708, a member of a conserved mammalian microRNA family, is located in the first intron of *ODZ4*. Three small nucleolar RNAs (snoRNAs), SNORD69, SNORD19 and SNORD19B, are located on chromosome 3p21.1 and belong to the C/D family of snoRNAs. Finally, a 121-base non-coding RNA with homology to 5S-ribosomal RNA is within the *SYNE1* association region. The role of microRNAs and non-coding RNAs in neurodevelopmental disorders is increasingly apparent in Rett's syndrome, fragile X syndrome and schizophrenia. Our study represents, to our knowledge, the first connection of these regions to bipolar disorder.

In conclusion, we obtained strong evidence for replication of multiple signals in bipolar disorder. In particular, we support prior findings in *CACNA1C* and now identify an intronic variant in *ODZ4* as being associated with bipolar disorder. These replication results imply that data from additional samples, both from GWAS and sequencing, will identify more of the genetic architecture of bipolar disorder. Finally, our combined analysis with schizophrenia illuminates the growing appreciation of the shared genetic epidemiology of these two disorders[30] and the shared polygenic contribution to risk[19].

URLs. Genetic Cluster Computer, http://www.geneticcluster.org/; NCBI gene2go, ftp.ncbi.nlm.nih.gov/gene/DATA/gene2go.gz; INRICH, http://atgu.mgh.harvard.edu/inrich.

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.

1U54RR025204). Other sources of support include: the Genetic Association Information Network (GAIN), the NIMH Intramural Research Program, the Tzedakah Foundation, the American Philosophical Society, the Stardust foundation, the National Library of Medicine, the Stanley Medical Research Institute, the Merck Genome Research Institute and the Wellcome Trust, the Pritzker Neuropsychiatric Disorders Research Fund L.L.C., GlaxoSmithKline, as well as grants for individual studies (see the **Supplemental Note** for a full list of Acknowledgements). The TOP Study was supported by grants from the Research Council of Norway (167153/V50, 163070/V50 and 175345/V50), the South-East Norway Health Authority (123-2004) and the EU (ENBREC). Additional acknowledgments can be found in the **Supplementary Note**.

Genotype data from this manuscript for the 10,257 samples can be obtained from the Center on Collaborative Genetic Studies of Mental Disorders in accordance with NIMH data release policies (http://zork.wustl.edu/nimh/). Genotype data from the WTCCC sample can be obtained from https://www.wtccc.org.uk/info/access_to_data_samples.shtml. Genotype data from the BOMA-Bipolar Study can be obtained by contacting S. Cichon directly (sven.cichon@uni-bonn.de).

AUTHOR CONTRIBUTIONS

Manuscript preparation: P. Sklar. (chair), O.A.A., S. Cichon , N.C., H.J.E., J.R.K., J.I.N., S.M.P., M.R., S. Ripke, L.J.S. **Analysis group:** S.M.P. (chair), D.B., A.C., N.C., M.F., W.G., J.R.K., M. Mattingsdal, A. McQuillin, P.K., S. Ripke, L.J.S., P. Sklar, T.F.W. **PGC central analysis group**: S. Ripke (chair), M.D., F.D., P.A.H., D.L. **Management group:** J.R.K. (co-chair), P. Sklar. (co-chair), O.A.A., D.B., M. Burmeister, S. Cichon , N.C., T.A.G., M.L.H., S.M.P., P. Muglia, J.I.N., L.J.S., E.N.S., P.P.Z.

Primary study data was contributed from the following investigators: ***BiGS (Bipolar Genome Study)–University of California San Diego:*** J.R.K., T.A.G., C.M.N., R.M., P.D.S.; ***Scripps Translational Science Institute:*** N.J.S., E.N.S., C.S.B.; ***Indiana University:*** J.I.N., H.J.E., T.F., D.L.K.; ***University of Chicago:*** E.S.G., C.L., J.A.B.; ***Rush University Medical Center:*** W.A.S.; ***Howard University:*** W.B.L., E.A.N., M. Hipolito; ***University of Iowa:*** W.C.; ***Washington University:*** J.R.; ***University of California San Francisco:*** W. Byerley; ***National Institute of Mental Health:*** F.J.M., T.G.S.; ***University of Pennsylvania:*** W. Berrettini, F.W.L.; ***Johns Hopkins Hospital:*** J.B.P., P.P.Z., P.B. Mahon; ***University of Michigan:*** M.G.M., S.Z., P.Z.; ***The Translational Genomics Research Institute:*** D.W.C., S. Szelinger; ***Portland Veterans Affairs Medical Center:*** T.B.B.; ***Central Institute of Mental Health, Mannheim:*** R.B., S.M., M.R., T.G.S., J. Strohmaier, S.H.W.; ***GlaxoSmithKline, Institute of Psychiatry, Centre for Addiction & Mental Health, University of Dundee:*** F.T., P. Muglia, A.F., P. McGuffin, J. Strauss, W.X., J.L.K., J.B.V., K. Matthews, R.D.; ***Massachusetts General Hospital***: M.A.F., C.O.'D., R.P., S.M.P., S. Raychaudhuri, P.H.L., D.R., P. Sklar, J.W.S.; ***NIMH/Pritzker:*** L.J.S., M.F., M. Burmeister, J. Li, W.G., P.K., D.A., R.C.T., F.M., A.F.S., W.E.B., J.D.B., E.G.J., S.J.W., R.M.M., H.A., M. Boehnke; ***Stanley Center Broad Institute:*** K.C., J.M., E. Scolnick; ***TOP Study Group:*** O.A.A., S.D., M. Mattingsdal, I.M., G.M.; ***Trinity College Dublin:*** A.C., M.G., D.M., E.Q.; ***University of Bonn:*** S. Cichon, T.W.M., F.A.D., M. Mattheisen, J. Schumacher, W.M., M. Steffens, T.F.W., P.P., M.M.N.; ***University College London:*** A.A., N.B., H.G., R.K., J. Lawrence, A. McQuillin; ***University of Edinburgh:*** D.B., K. McGhee, A. McIntosh, A.W.M., W.J.M., B.S.P.; ***Wellcome Trust Case Control Consortium Bipolar Disorder Group—Aberdeen:*** G.B., D.St.C.; ***Birmingham:*** S. Caesar, K.G.-S., L.J.; ***Cardiff:*** C.F., E.K.G., D.G., M.L.H., P.A.H., I.R.J., G.K., V.M., I.N., M.C.O., M.J.O., N.C.; ***London:*** D.A.C., A.E., A.F., R.W., P. McGuffin; ***Newcastle:*** A.H.Y., I.N.F.

Replication data contributed by the following investigators: ***Central Institute of Mental Health, Mannheim:*** R.B., S.M., M.R., T.G.S., J. Strohmaier, S.H.W.; ***deCODE genetics:*** H.S., Þ.Þ., S. Steinberg, Ó.G., K.S.; ***FaST:*** S.E.B., K.C., J.M., V.N., S.M.P., P. Sklar, J.W.S.; ***ICCBD:*** C.H., M. Landén, P.L., P. Sullivan, M. Schalling, U.O., L.B., L.F., S.E.B., S.M.P., K.C., J.M., N.L., P. Sklar; ***INSERM U955:*** S.J., M. Leboyer, B.E., F.B.; ***Landspitali University:*** H.P., E. Sigurðsson; ***Max Planck Institute of Psychiatry, Munich:*** B.M.-M., S.L.; ***Psychiatric Center Nordbaden:*** M. Schwarz; ***Neuroscience Research Australia and University of New South Wales:*** P.R.S. and J.M.F.; ***Queensland Institute of Medical Research:*** N.M., G.W.M.; ***Centre National de Génotypage:*** M. Lathrop; ***Therapeia University Hospital, Reykjavík:*** H.Ó.; ***University of Bonn:*** S. Cichon, T.W.M., F.A.D., M. Mattheisen,

1. Craddock, N. & Sklar, P. Genetics of bipolar disorder: successful start to a long journey. *Trends Genet.* **25**, 99–105 (2009).
2. Cardno, A.G. *et al.* Heritability estimates for psychotic disorders: the Maudsley twin psychosis series. *Arch. Gen. Psychiatry* **56**, 162–168 (1999).
3. Craddock, N. & Jones, I. Genetics of bipolar disorder. *J. Med. Genet.* **36**, 585–594 (1999).
4. Craddock, N., O'Donovan, M.C. & Owen, M.J. The genetics of schizophrenia and bipolar disorder: dissecting psychosis. *J. Med. Genet.* **42**, 193–204 (2005).
5. Baum, A.E. *et al.* A genome-wide association study implicates diacylglycerol kinase eta (DGKH) and several other genes in the etiology of bipolar disorder. *Mol. Psychiatry* **13**, 197–207 (2007).
6. Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* **447**, 661–678 (2007).
7. Sklar, P. *et al.* Whole-genome association study of bipolar disorder. *Mol. Psychiatry* **13**, 558–569 (2008).
8. Ferreira, M.A. *et al.* Collaborative genome-wide association analysis supports a role for ANK3 and CACNA1C in bipolar disorder. *Nat. Genet.* **40**, 1056–1058 (2008).
9. 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).
10. Smith, E.N. *et al.* Genome-wide association study of bipolar disorder in European American and African American individuals. *Mol. Psychiatry* **14**, 755–763 (2009).
11. Djurovic, S. *et al.* A genome-wide association study of bipolar disorder in Norwegian individuals, followed by replication in Icelandic sample. *J. Affect. Disord.* **126**, 312–316 (2010).
12. Cichon, S. *et al.* Genome-wide association study identifies genetic variation in neurocan as a susceptibility factor for bipolar disorder. *Am. J. Hum. Genet.* **88**, 372–381 (2011).
13. Schulze, T.G. *et al.* Two variants in Ankyrin 3 (*ANK3*) are independent genetic risk factors for bipolar disorder. *Mol. Psychiatry* **14**, 487–491 (2009).
14. Psychiatric Genome-wide Association Study Consortium. A framework for interpreting genome-wide association studies of psychiatric disorders. *Mol. Psychiatry* **14**, 10–17 (2009).
15. Cichon, S. *et al.* Genomewide association studies: history, rationale, and prospects for psychiatric disorders. *Am. J. Psychiatry* **166**, 540–556 (2009).
16. 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).
17. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
18. Devlin, B. & Roeder, K. Genomic control for association studies. *Biometrics* **55**, 997–1004 (1999).
19. International Schizophrenia Consortium. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* **460**, 748–752 (2009).
20. Zhang, D. *et al.* Singleton deletions throughout the genome increase risk of bipolar disorder. *Mol. Psychiatry* **14**, 376–380 (2009).
21. Levy, N.A. & Janicak, P.G. Calcium channel antagonists for the treatment of bipolar disorder. *Bipolar Disord.* **2**, 108–119 (2000).
22. Kempton, M.J. *et al.* Effects of the *CACNA1C* risk allele for bipolar disorder on cerebral gray matter volume in healthy individuals. *Am. J. Psychiatry* **166**, 1413–1414 (2009).
23. Krug, A. *et al.* Effect of *CACNA1C* rs1006737 on neural correlates of verbal fluency in healthy individuals. *Neuroimage* **49**, 1831–1836 (2010).
24. Bigos, K.L. *et al.* Genetic variation in *CACNA1C* affects brain circuitries related to mental illness. *Arch. Gen. Psychiatry* **67**, 939–945 (2010).
25. Erk, S. *et al.* Brain function in carriers of a genome-wide supported bipolar disorder variant. *Arch. Gen. Psychiatry* **67**, 803–811 (2010).
26. Nyegaard, M. *et al. CACNA1C* (rs1006737) is associated with schizophrenia. *Mol. Psychiatry* **15**, 119–121 (2010).
27. Curtis, D. *et al.* Case-case genome-wide association analysis shows markers differentially associated with schizophrenia and bipolar disorder and implicates calcium channel genes. *Psychiatr. Genet.* **21**, 1–4 (2011).
28. 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).
29. Gomez-Ospina, N., Tsuruta, F., Barreto-Chang, O., Hu, L. & Dolmetsch, R. The C terminus of the L-type voltage-gated calcium channel Ca(V)1.2 encodes a transcription factor. *Cell* **127**, 591–606 (2006).
30. Lichtenstein, P. *et al.* Common genetic determinants of schizophrenia and bipolar disorder in Swedish families: a population-based study. *Lancet* **373**, 234–239 (2009).

Tatiana Foroud[12], Daniel L Koller[12], Elliot S Gershon[30], Chunyu Liu[30], Judith A Badner[30], William A Scheftner[31], William B Lawson[32], Evaristus A Nwulia[32], Maria Hipolito[32], William Coryell[33], John Rice[34], William Byerley[35], Francis J McMahon[36], Thomas G Schulze[14,36,37], Wade Berrettini[38], Falk W Lohoff[38], James B Potash[28], Pamela B Mahon[28], Melvin G McInnis[39], Sebastian Zöllner[4,39], Peng Zhang[4,39], David W Craig[40], Szabocls Szelinger[40], Thomas B Barrett[41], René Breuer[14], Sandra Meier[14], Jana Strohmaier[14], Stephanie H Witt[14], Federica Tozzi[26], Anne Farmer[42], Peter McGuffin[42], John Strauss[43], Wei Xu[44], James L Kennedy[45], John B Vincent[43], Keith Matthews[46], Richard Day[46], Manuel A Ferreira[2,3,47], Colm O'Dushlaine[2,3], Roy Perlis[2,3], Soumya Raychaudhuri[2,3], Douglas Ruderfer[2,3], Phil H Lee[2], Jordan W Smoller[2,3], Jun Li[48], Devin Absher[49], William E Bunney[50], Jack D Barchas[51], Alan F Schatzberg[52], Edward G Jones[53], Fan Meng[54], Robert C Thompson[54], Stanley J Watson[54], Richard M Myers[49], Huda Akil[54], Michael Boehnke[4], Kim Chambert[3], Jennifer Moran[3], Ed Scolnick[3], Srdjan Djurovic[5,55], Ingrid Melle[5,6], Gunnar Morken[56,57], Michael Gill[17], Derek Morris[17], Emma Quinn[17], Thomas W Mühleisen[7,8], Franziska A Degenhardt[7,8], Manuel Mattheisen[7,20], Johannes Schumacher[7], Wolfgang Maier[58], Michael Steffens[20], Peter Propping[7], Markus M Nöthen[7,8], Adebayo Anjorin[19], Nick Bass[19], Hugh Gurling[19], Radhika Kandaswamy[19], Jacob Lawrence[19], Kevin McGhee[15,16], Andrew McIntosh[15,16], Alan W McLean[15,16], Walter J Muir[15,16], Benjamin S Pickard[15,16], Gerome Breen[42,59], David St. Clair[59], Sian Caesar[60], Katherine Gordon-Smith[10,60], Lisa Jones[60], Christine Fraser[10], Elaine K Green[10], Detelina Grozeva[10], Ian R Jones[10], George Kirov[10], Valentina Moskvina[10,22], Ivan Nikolov[10], Michael C O'Donovan[10], Michael J Owen[10], David A Collier[42], Amanda Elkin[42], Richard Williamson[42], Allan H Young[46,61], I Nicol Ferrier[46], Kari Stefansson[62], Hreinn Stefansson[62], Þorgeir Þorgeirsson[62], Stacy Steinberg[62], Ómar Gustafsson[62], Sarah E Bergen[2,3], Vishwajit Nimgaonkar[63], Christina Hultman[64], Mikael Landén[64,65], Paul Lichtenstein[64], Patrick Sullivan[66], Martin Schalling[67], Urban Osby[67], Lena Backlund[68], Louise Frisén[67], Niklas Langstrom[65], Stéphane Jamain[69–72], Marion Leboyer[69–72], Bruno Etain[69–72], Frank Bellivier[69–72], Hannes Petursson[73], Engilbert Sigurðsson[73], Bertram Müller-Mysok[74], Susanne Lucae[74], Markus Schwarz[75], Janice M Fullerton[76,77], Peter R Schofield[76,77], Nick Martin[47], Grant W Montgomery[47], Mark Lathrop[78], Högni Óskarsson[79], Michael Bauer[80], Adam Wright[81], Philip B Mitchell[81], Martin Hautzinger[82], Andreas Reif[83], John R Kelsoe[25,84] & Shaun M Purcell[1–3]

[1]Division of Psychiatric Genomics, Department of Psychiatry, Mount Sinai School of Medicine, New York, New York, USA. [2]Center for Human Genetic Research, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts, USA. [3]Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA. [4]Department of Biostatistics and Center for Statistical Genetics, School of Public Health, University of Michigan, Ann Arbor, Michigan, USA. [5]Institute of Clinical Medicine, European Network of Bipolar Research Expert Centers (ENBREC) Group, University of Oslo, Oslo, Norway. [6]Division of Mental Health and Addiction, Oslo University Hospital, Oslo, Norway. [7]Institute of Human Genetics, University of Bonn, Bonn, Germany. [8]Department of Genomics, Life and Brain Center, University of Bonn, Bonn, Germany. [9]Institute of Neuroscience and Medicine (INM-1), Research Center Juelich, Juelich, Germany. [10]Medical Research Council (MRC) Centre for Neuropsychiatric Genetics and Genomics, Henry Wellcome Building, School of Medicine, Cardiff University, Heath Park, Cardiff, UK. [11]Department of Biochemistry and Molecular Biology, Department of Medical and Molecular Genetics, Indiana University School of Medicine, Indianapolis, Indiana, USA. [12]Department of Psychiatry, Indiana University School of Medicine, Indianapolis, Indiana, USA. [13]Department of Medical and Molecular Genetics, Indiana University School of Medicine, Indianapolis, Indiana, USA. [14]Department of Genetic Epidemiology in Psychiatry, Central Institute of Mental Health Mannheim, University of Heidelberg, Mannheim, Germany. [15]Division of Psychiatry, University of Edinburgh, Royal Edinburgh Hospital, Edinburgh, UK. [16]Medical Genetics Section, University of Edinburgh Molecular Medicine Centre, Western General Hospital, Edinburgh, UK. [17]Neuropsychiatric Genetics Research Group, Department of Psychiatry and Institute of Molecular Medicine, Trinity College Dublin, Dublin, Ireland. [18]Sorlandet Hospital HF, Kristiansand, Norway. [19]Molecular Psychiatry Laboratory, Research Department of Mental Health Sciences, University College London, Rockefeller Building, London, UK. [20]Institute for Medical Biometry, Informatics and Epidemiology, University of Bonn, Bonn, Germany. [21]London School of Hygiene and Tropical Medicine, University of London, London, UK. [22]Biostatistics and Bioinformatics Unit, Cardiff University School of Medicine, Cardiff, UK. [23]Department of Biostatistics, University of North Carolina, Chapel Hill, North Carolina, USA. [24]Department of Human Genetics, Department of Psychiatry, Molecular and Behavioral Neuroscience Institute, University of Michigan, Ann Arbor, Michigan, USA. [25]Department of Psychiatry, University of California San Diego, La Jolla, California, USA. [26]Neurosciences Centre of Excellence in Drug Discovery, GlaxoSmithKline Research and Development, Verona, Italy. [27]The Scripps Translational Science Institute and Scripps Health, La Jolla, California, USA. [28]Department of Mental Health, Johns Hopkins University and Hospital, Baltimore, Maryland, USA. [29]The Scripps Translational Science Institute and The Scripps Research Institute, La Jolla, California, USA. [30]Department of Psychiatry, University of Chicago, Chicago, Illinois, USA. [31]Rush University Medical Center, Chicago, Illinois, USA. [32]Department of Psychiatry and Behavioral Sciences, Howard University College of Medicine, Washington, DC, USA. [33]Department of Psychiatry, University of Iowa, Iowa City, Iowa, USA. [34]Washington University School of Medicine, St. Louis, Missouri, USA. [35]Department of Psychiatry, University of California San Francisco School of Medicine, San Francisco, California, USA. [36]National Institute of Mental Health, US National Institutes of Health, Bethesda, Maryland, USA. [37]Department of Psychiatry, University Göttingen, Göttingen, Germany. [38]Department of Psychiatry, University of Pennsylvania, Philadelphia, Pennsylvania, USA. [39]Department of Psychiatry, University of Michigan, Ann Arbor, Michigan, USA. [40]The Translational Genomics Research Institute, Phoenix, Arizona, USA. [41]Portland Veterans Affairs Medical Center, Portland, Oregon, USA. [42]Social, Genetic and Developmental Psychiatry (SGDP) Centre, The Institute of Psychiatry, King's College London, De Crespigny Park Denmark Hill, London, UK. [43]Molecular Neuropsychiatry and Development Laboratory, Centre for Addiction and Mental Health, Toronto,

[60]Department of Psychiatry, School of Clinical and Experimental Medicine, Birmingham University, Birmingham, UK. [61]University of British Columbia (UBC) Institute of Mental Health, Vancouver, British Columbia, Canada. [62]deCODE genetics, Reykjavík, Iceland. [63]Department of Human Genetics, University of Pittsburgh, Pittsburgh, Pennsylvania, USA. [64]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden. [65]Institute of Neuroscience and Physiology, University of Gothenburg, Gothenburg, Sweden. [66]Department of Genetics, University of North Carolina at Chapel Hill School of Medicine, Chapel Hill, North Carolina, USA. [67]Department of Molecular Medicine, Karolinska Institutet, Stockholm, Sweden. [68]Department of Clinical Neuroscience, Karolinska Institutet, Stockholm, Sweden. [69]INSERM, U955, Psychiatrie Génétique, Créteil, France. [70]Université Paris Est, Faculté de Médecine, Créteil, France. [71]Assistance Publique–Hôpitaux de Paris (AP-HP), Hôpital H. Mondor–A. Chenevier, Département de Psychiatrie, Créteil, France. [72]ENBREC group, Fondation Fondamental, Créteil, France. [73]Division of Psychiatry, Landspitali University Hospital, University of Iceland, Reykjavik, Iceland. [74]Max Planck Institute of Psychiatry, Munich, Germany. [75]Psychiatric Center Nordbaden, Wiesloch, Germany. [76]Neuroscience Research Australia, Sydney, Australia. [77]University of New South Wales, Sydney, Australia. [78]Centre National de Génotypage, Evry, France. [79]Therapeia, Reykjavik, Iceland. [80]Department of Psychiatry and Psychotherapy, ENBREC Group, University Hospital Carl Gustav Carus, Dresden, Germany. [81]School of Psychiatry, University of New South Wales and Black Dog Institute, Sydney, New South Wales, Australia. [82]Department of Clinical and Developmental Psychology, Institute of Psychology, University of Tubingen, Tubingen, Germany. [83]Department of Psychiatry, University of Würzburg, Würzburg, Germany. [84]Department of Psychiatry, Special Treatment and Evaluation Program (STEP), Veterans Affairs San Diego Healthcare System, San Diego, California, USA. Correspondence should be addressed to P.S. (pamela.sklar@mssm.edu).

the criteria for bipolar disorder within the primary study classification system. Controls were selected from the same geographical and ethnic populations as the cases and had a low probability of having bipolar disorder. Some control selection criteria excluded individuals with a personal history of mood disorder, and other controls were unscreened. The distribution of diagnoses is shown in **Supplementary Table 1**. Protocols and assessment procedures were approved by the institutional review boards of the authors' institutions. All participants provided written informed consent, and consent allowed the samples to be used within the current analyses.

PGC central data quality control pipeline. Primary study genotype data were deposited on the Genetic Cluster Computer (see URLs) hosted by the Dutch National Computing and Networking Services. Data were generated using four different genotyping platforms (Affymetrix 500K, 5.0, 6.0 and Illumina HumanHap 500). Data were processed by the PGC central analysis committee pipeline, which performed semi-automated formatting, quality testing, inter- and intra-study relatedness checks and imputation. First, SNP names, positions and strand were harmonized. For SNPs with <5% missing data, individuals were retained if the missing genotype rate per individual was <0.02. Subsequently, SNPs were retained if the missing genotype rate per SNP was <0.02, the missing genotype rate between cases and controls per SNP was <0.02 (absolute difference), Hardy-Weinberg equilibrium (controls) $P > 1 \times 10^{-6}$ and the frequency difference to the HapMap reference was <0.15. This removed 380,959 SNPs and 177 individuals from the 11 bipolar disorder studies. After these steps, there were 10,926 controls and 8,338 cases for analysis.

Data were imputed using BEAGLE 3.0 (ref. 16), with phased HapMap phase 2 data as a reference. Each dataset was imputed separately, splitting into imputation batches of 300 individuals randomly, keeping the case-control ratio balanced.

Duplicate sample elimination. Using PLINK[17], we found that 3,714 individuals in 2,316 pairs were duplicated, which we defined as a pair of samples with an estimated probability of genome-wide identity-by-descent of sharing two chromosomes above 90%. To remove duplicates, in order to preserve case:control ratios as close to 50:50 as possible and to favor data generated using more recent platforms, we preferentially kept samples from duplicate pairs in the order as follows: BOMA-bipolar study, TOP, STEP2, NIMH/PRITZKER, GAIN/BiGS, STEP1, TRINITY COLLEGE, UEDINBURGH, GSK, UCL and then WTCCC. The final dataset contained only unique individuals, with each individual belonging to exactly one sample. We further detected instances of previously unknown close relatedness. After removing a small number of parent-offspring, full-sibling and half-sibling pairs, we were left with $N = 16{,}731$ individuals in 16,254 families (including 477 known sibling pairs from NIMH/PRITZKER).

Ancestry evaluation. We used the WTCCC control sample to select SNPs in approximate linkage equilibrium to calculate multi-dimensional scaling (MDS) components to assess and correct for population stratification. This yielded $N = 21{,}134$ autosomal SNPs, genotyped on all platforms, which is sufficient for this MDS analysis. We calculated the top 20 MDS components. Based on inspection of between- and within-sample correlation with the phenotype, we retained the top five components, which were used as covariates along with ten binary dummy variables to control for differences between the 11 samples (**Supplementary Fig. 1**).

Association analyses. *Primary analyses.* Following initial quality control and elimination of duplicates, there were 16,731 individuals and 2,541,952 SNPs. Analyses are based on the 2,415,422 SNPs with minor allele frequency >1% and imputation $r^2 > 0.3$ with a HapMap SNP. The primary analysis was a logistic regression of disease state on single-SNP allele dosage including covariates to **Table 2** were based on the clumping approach (PLINK). Specifically, we took all SNPs significant at $P < 5 \times 10^{-5}$ that had not already been clumped (denoting these as index SNPs) and formed clumps of all other SNPs that are within 1 Mb of the index SNP, in LD with the index SNP ($r^2 > 0.2$) or nominally associated with disease ($P < 0.05$). This approach grouped SNPs in LD space rather than physical distance. This clumping approach resulted in 38 SNPs with $P < 5 \times 10^{-5}$.

Replication analyses. From each replication sample, we obtained information on *P* values, ORs, standard errors (SE), minor allele frequencies and the associated risk allele for SNPs listed in **Supplementary Table 2**. If the target SNP listed in **Supplementary Table 2** was not present in the replication dataset, we obtained a proxy SNP in strong LD and weighted the SE to account for the lack of information: SEW = SE/sqrt(r^2). The estimate (ES), β, is the natural logarithm of the odds ratio. We performed a standard meta-analysis to estimate a common odds ratio weighted by individual study's SE (**Supplementary Table 8**).

$$\mathrm{ES} = \frac{\sum ES_i / SE_i^2}{\sum 1 / SE_i^2}$$

$$\mathrm{SE} = \sqrt{\frac{1}{\sum \frac{1}{SE_i^2}}}$$

We combined the odds ratios and standard errors from the discovery and replication samples using a fixed effect meta-analysis. The final *P* values are genomic-control adjusted based on a λ of 1.176, estimated from all available GWAS data (from both wave 1 and the replication samples).

Gene Ontology (GO) enrichment analyses. We looked for terms enriched for genes in the most associated intervals. We started with the 38 intervals described above. Only 34 regions were analyzed by collapsing any regions that (i) physically overlapped, (ii) spanned the same gene or (iii) did not show conditionally independent association signals. Three regions contain SNPs that had low pairwise r^2 but did not show independent association when covarying for the neighboring SNP, reflecting high LD measured in terms of D' and not indicative of truly independent signals. The enrichment analysis depended on the assumption of independence between intervals so as not to double count genes. The final list contained 34 independent regions.

We used NCBI gene2go (see URLs) and mapped Entrez GeneIDs to gene symbols and hg18 coordinates using the UCSC Genome Browser. Of the 9,834 total GO terms, we restricted the analysis to terms with at least 2 and not more than 200 human genes, leaving 6,482 GO terms ('targets'). For each target, we counted the number of association intervals that contained at least one target gene; we required that at least two intervals contained at least one gene from each target. We evaluated the probability of observing the number of intersecting intervals by chance alone using a permutation procedure (implemented in INRICH software (see URLs)). Specifically, we randomly placed each independent interval in an alternate position on the genome matching for the total number of SNPs and implied new size of the interval (the distance in bp was within a factor of 0.8–1.2 of the original) and the total number of genes. In this manner, we controlled for potential biases caused by SNP and gene density and gene size. We repeated the permutation 100,000 times and corrected for multiple testing by evaluating the distribution of minimum empirical *P* values under the null hypothesis given 6,482 tested targets. The corrected empirical *P* values implicitly account for the non-independence of the GO terms.

Conditional analyses. To identify additional signals after accounting for the effects of the initial GWAS signals, we performed a conditional analysis including the most strongly associated SNPs in the analysis of each SNP. In regions in which we detected a potential secondary signal(s) ($P_{GC} < 10^{-4}$), we performed separate conditional analyses using the initial GWAS-identified

librium, yielding ~20,000 SNPs. We estimated the odds ratios from the ten sites, excluding the German sample, by a fixed-effect meta-analysis and took the log of these odds ratios as weights to calculate the scores in the target sample. Following the ISC methods, we selected discovery sample *P*-value thresholds of $P < 0.01$, $P < 0.05$, $P < 0.1$, $P < 0.2$, $P < 0.3$, $P < 0.4$ and $P < 0.5$. For each threshold, we performed a logistic regression of disease state in the target sample on the polygenic score from the remaining independent samples covarying for the rate of genotyping failure and MDS components to adjust for potential confounders. We observed significant enrichment of putatively associated 'score alleles' in the target sample cases compared to the controls (*P* values and pseudo r^2 values presented; all effects were in the expected direction, with a higher score in the cases compared to the controls).

PLINK pi-hat > 0.9 was used to identify identical controls. Only one individual was retained for analysis and was randomly assigned to either the bipolar disorder or the schizophrenia set. The primary analysis was a logistic regression of disease state on single SNP allele dosage similar to those described above for our primary GWAS sample association. We included covariates to account for site as well as the quantitative indices (the first five plus three additional indices that showed some correlation with phenotype) of ancestry based on multi-dimensional scaling.

31. The Schizophrenia Psychiatric Genome-Wide Association Study (GWAS) Consortium. Genome-wide association study identifies five new schizophrenia loci. *Nat. Genet.* advance online publication, doi:10.1038/ng.940 (18 September 2011).

In the version of this article initially published, there were errors in the consortium membership list and the associated affiliations and in the acknowledgements and contributions sections. These errors and their corrections are detailed below by section.

Consortium members:
Janice M. Fullerton was omitted from the membership list and has now been added with affiliations 76 and 77. Phil H. Lee was listed incorrectly as Phil L. Hyoun. Fan Meng was listed incorrectly as Fan Guo Meng, and the associated affiliation has been changed from 51 to 54. Robert Thompson was assigned affiliation 50; the correct affiliation is 54. Marian Hamshere and Valentina Moskvina were assigned affiliation 26; the correct affiliation for both is 22. Richard Day was assigned affiliation 47; the correct affiliation is 46. Jun Li was assigned affiliation 24; the correct affiliation is 48. In addition to the affiliation originally listed for Sebastian Zöllner and Peng Zhang, both have now also been assigned affiliation 4. Howard Endenberg has now also been assigned affiliation 12, and Shaun Purcell has now also been assigned affiliation 1.

Consortium affiliations:
Affiliation 46 was originally given as the University of Dundee School of Medicine, Nethergate, Dundee, UK. The correct affiliation is the Division of Neuroscience, University of Dundee, Ninewells Hospital & Medical School, Dundee, UK. Affiliation 47 was originally given as the School of Neurology, Neurobiology and Psychiatry, Royal Victoria Infirmary, Newcastle upon Tyne, UK. The correct affiliation is the Department of Human Genetics, University of Michigan, Ann Arbor, Michigan, USA. This affiliation was also listed out of order and has now been changed to affiliation 48. Affiliation 76 was originally given as the Prince of Wales Medical Institute, Sydney, Australia. The correct affiliation is Neuroscience Research Australia, Sydney, Australia.

Acknowledgments:
The Stanley Foundation for Medical Research was listed as a source of funding. The correct name is the Stanley Medical Research Institute, and the Merck Genome Research Institute has also been added as a source of support.

Contributions:
In four instances, the contribution of Sven Cichon was indicated with the incorrect spelling S. Chichon instead of S. Cichon. Manuel A. Ferreira (M.A.F.) was incorrectly listed as a contributor to manuscript preparation. Manuel A. Ferreira (M.A.F.) was listed as a contributor to primary study data at the NIMH/Pritzke; the correct contributor was Matthew Flickinger (M.F.). In the section listing contributors to replication data, Neuroscience Research Australia was named incorrectly as the Prince of Wales Medical Institute, and Janice M. Fullerton (J.M.F.) has been added as a contributor at this site and at the University of New South Wales.

The errors detailed above have been corrected in the HTML and PDF versions of the article. In addition, the subsections of the contributions section detailing the individuals contributing to primary study data and replication data were omitted from the original HTML version of the article, and this error has now been corrected.