UC Riverside
UC Riverside Previously Published Works

Title
Sparse whole-genome sequencing identifies two loci for major depressive disorder.

Permalink
https://escholarship.org/uc/item/1vj5867b

Journal
Nature, 523(7562)

ISSN
0028-0836

Author
CONVERGE consortium

Publication Date
2015-07-01

DOI
10.1038/nature14659

Peer reviewed

Sparse whole genome sequencing identifies two loci for major depressive disorder

CONVERGE Consortium[1]

Abstract

Major depressive disorder (MDD), one of the most frequently encountered forms of mental illness and a leading cause of disability worldwide[1], poses a major challenge to genetic analysis. To date no robustly replicated genetic loci have been identified [2], despite analysis of more than 9,000 cases[3]. Using low coverage genome sequence of 5,303 Chinese women with recurrent MDD selected to reduce phenotypic heterogeneity, and 5,337 controls screened to exclude MDD, we identified and replicated two genome-wide significant loci contributing to risk of MDD on chromosome 10: one near the *SIRT1* gene (P-value = 2.53×10^{-10}) the other in an intron of the *LHPP* gene (P = 6.45×10^{-12}). Analysis of 4,509 cases with a severe subtype of MDD, melancholia, yielded an increased genetic signal at the *SIRT1* locus. We attribute our success to the recruitment of relatively homogeneous cases with severe illness.

The existence and number of subtypes of depression have been debated over the past 100 years. The current consensus is that depression may be a collection of partly distinct diseases, with overlapping causal pathways. This etiologic heterogeneity might therefore substantially reduce the power of genetic association studies and hence explain the failure to find genetic risk loci [3]. For example, there may be cases of MDD of largely environmental origin whose presence reduces power to detect genetic effects. Genetic risk factors for mild depressive syndromes may not be entirely the same as those for more severe cases [4].

For these reasons, we investigated the genetic basis of MDD in subjects for whom known sources of phenotypic and genetic heterogeneity were minimized and known risk factors documented. The CONVERGE (**C**hina, **O**xford a**n**d **V**irginia Commonwealth University **E**xperimental **R**esearch on **G**enetic **E**pidemiology) consortium recruited 11,670 Han Chinese women through a collaboration involving 58 hospitals in China. We studied women only because about 45% of the genetic liability to MDD is not shared between sexes [5,6]. In an attempt to obtain severe cases of MDD, we recruited only recurrent cases (mean number of episodes was 5.6).

We used low coverage sequencing to genotype our sample[7]. Whole genome sequence was acquired to a mean depth of 1.7× (95% CIs 0.7-4.3) per individual from which 32,781,340

SNP sites were identified. After applying stringent quality controls (Supplemental Methods), we obtained 10,640 samples (5,303 cases of MDD, 5,337 controls) and 6,242,619 SNPs for inclusion in GWAS. We compared genotypes from the low coverage sequencing to genotypes called with $10\times$ coverage sequence and to genotypes called from genotyping arrays and a mass spectrometer platform. The mean percentage concordance between genotypes from nine individuals with both low and 10X coverage across all sites was 98.1% (Supplemental Table 1). We compared imputed genotypes to those acquired for 72 individuals using an array and to 21 SNPs genotyped on all individuals with the MassARRAY™ system mass spectrometer (Supplemental Methods). Overall concordance was 98.0% (Supplemental Tables 2 and 3).

Genetic association analysis was carried out with a linear mixed model with a genetic relatedness matrix (GRM) as a random effect and principal components from eigen-decomposition of the GRM as fixed effects (Methods, Supplemental Methods)[8,9]. Figure 1a and Extended Data Figure 1 show the Manhattan and quantile-quantile plots respectively for the analysis. The genomic-control inflation factor (λ, the ratio of the observed median χ^2 to that expected by chance) for association with MDD was 1.070 (for common SNPs > 2 %, λ = 1.074). The adjusted measure for sample size to that of 1,000 cases and 1,000 controls (λ_{1000}) was 1.013.

Two loci exceeded genome-wide significance in association with MDD: one 5′ to the sirtuin1 (*SIRT1*) gene on chromosome 10 (SNP = rs12415800, chromosome 10:69624180, minor allele frequency (MAF) = 45.2%, P = 1.92×10^{-8}, Figure 1b) and the other in an intron of the phosphorlysine phosphohistidine inorganic pyrophosphate phosphatase (*LHPP*) gene (SNP = rs35936514, chromosome 10:126244970, MAF = 26.0%, P = 1.27×10^{-8}, Figure 1c). All SNPs with P-values of association <10^{-6} with MDD are listed in Supplemental Table 4.

We checked the accuracy of the imputed genotypes at 12 SNPs with P < 1×10^{-5}, by re-genotyping the CONVERGE samples using a MassARRAY™ system mass spectrometer, thereby confirming their association with MDD. Extended Data Table 1 shows that the correlation between the two assays was high (mean R^2 = 0.984), and the odds ratios for the two genome-wide significant SNPs assessed by the two methods were almost identical with highly overlapping confidence intervals (rs12415800 odds ratios: 1.167 vs 1.167; rs35936514 odds ratios: 0.845 vs 0.842).

We replicated the associations by genotyping the same 12 SNPs in a separate Han Chinese cohort of 3,231 cases with recurrent MDD, and 3,186 controls (both sexes). Two SNPs at the peaks of association for *SIRT1* and *LHPP* loci (rs12415800 and rs35936514 respectively) for MDD in CONVERGE were significantly associated with MDD (Table 1). Analysis of the combined samples gave P-values for association with MDD at these two SNPs of 2.53×10^{-10} and 6.45×10^{-12}, respectively. Extended Data Table 2 shows the genotype distribution and p-values for tests of violation of Hardy-Weinberg Equilibrium (HWE) in both CONVERGE and the replication cohort at both SNPs.

Comparison with results from the Psychiatric Genomics Consortium (PGC) mega-analysis of European studies[3] failed to provide robust replication for our top SNPs (Extended Data

Figure 2 and Extended Data Table 3). However the proportion of associations in the same direction in the two studies exceeded chance expectations (P<0.001), and polygenic risk scores from the PGC mega-analysis applied to the CONVERGE samples were of significant (P<0.01) but limited predictive value, accounting for 0.1% of MDD risk in CONVERGE (Extended Data Table 4). It is unclear to what extent differences in sample ascertainment, ethnicity, or other factors contribute to the failure to replicate genetic effects in the PGC sample. Notably, variants at our most strongly associated loci are much rarer in European populations, where rs12415800 (*SIRT1*) and rs35936514 (*LHPP*) have frequencies of 3% and 8% respectively, compared to 45% and 26% in CONVERGE.

We considered whether successful mapping of MDD in CONVERGE was attributable to the recruitment of a severe, more genetically determined form of the disease. We tested that hypothesis by looking within CONVERGE at a particularly severe, and more heritable form of MD: melancholia[10]. Prior research has suggested that MDD patients with melancholia have more impairing, recurrent episodes and that risk for MD is increased in the co-twins of probands with the melancholic subtype[11]. This increase is greater in monozygotic than dizygotic pairs [11], as would be expected if the subtype were associated with greater genetic risk.

In CONVERGE, 85% of cases met DSM-IV melancholia criteria[12]. We searched for genetic association in 9,846 samples (4,509 cases and 5,337 controls) and identified the same two loci that exceeded genome-wide significance on chromosome 10. The genomic-control inflation factor λ for melancholia was 1.069, and λ_{1000} was 1.014. Even though the sample for melancholia was smaller than for MDD, at the *SIRT1* locus the significance of association was two orders of magnitude greater than for (top SNP = rs80309727, chromosome 10:69617347, MAF = 45.2%, P = 2.95×10^{-10}). Extended Data Figure 3 shows the Manhattan plot, quantile-quantile plot and detailed views of the *SIRT1* locus associated with melancholia. All SNPs with P-values of association $<10^{-6}$ with melancholia are listed in Supplemental Table 5. To determine whether the increased association might have arisen by chance, we generated an empirical distribution of odds ratios by randomly selecting 4,509 cases from the total set and re-analysing the association with each of the genome-wide significant variants. We found that the observed value lay on the 98.8 percentile at the *SIRT1* locus, but at the 61.6 percentile at the *LHPP* locus (Extended Data Figure 4).

Our results indicate that, as others have suggested[13], obtaining low sequence coverage of a large number of individuals can be an effective way to screen the genome for association signals. We were able to genotype more variants than present on typical genotyping arrays and our set is larger than publicly available sources for imputation[14]. The imputation pipeline we used employed standard tools, and it is likely that imputation accuracy could be improved with further algorithmic research.

MDD is most likely highly polygenic[3] and many additional loci remain to be discovered. We attribute the discovery and replication of two SNPs associated with MDD in CONVERGE to the recruitment of severely ill cases in a population where the disease has a relatively low prevalence (e.g., life-time prevalence rate for MDD in China of 3.6% [15] versus 16.2% in the US [16]), at least in part because individuals who report depressive illness

in East Asian cultures are more impaired than those in Western cultures[17]. Consistent with this interpretation, 85% of CONVERGE has melancholia, a severe subtype of MDD; moreover, mapping melancholia led to significant increase in genetic signal at one locus. However, in the absence of replication, this result should be regarded as tentative. Finally we note that one of the replicated risk loci lies close to a gene involved in mitochondrial biogenesis (*SIRT1*)[18], which, together with our finding that MDD is associated with increased amounts of mitochondrial DNA[19], suggests an unexpected origin for at least some of the phenotypic manifestations of MDD.

Methods

Sample collection

CONVERGE collected cases of recurrent major depression from 58 provincial mental health centres and psychiatric departments of general medical hospitals in 45 cities and 23 provinces of China. Controls were recruited from patients undergoing minor surgical procedures at general hospitals (37%) or from local community centres (63%). A sample size of 6,000 cases and 6,000 controls was chosen on the basis of evidence available when the study was designed (in 2007) of the likely existence of genetic loci with odds ratio of 1.2 and above. All subjects were Han Chinese women with four Han Chinese grandparents. Cases were excluded if they had a pre-existing history of bipolar disorder, psychosis or mental retardation. Cases were aged between 30 and 60 and had two or more episodes of MDD meeting DSM-IV criteria[20] with the first episode occurring between 14 and 50 years of age, and had not abused drugs or alcohol before their first depressive episode. All subjects were interviewed using a computerized assessment system. Interviewers were postgraduate medical students, junior psychiatrists or senior nurses, trained by the CONVERGE team for a minimum of one week. The diagnosis of MDD was established with the Composite International Diagnostic Interview (CIDI) (WHO lifetime version 2.1; Chinese version), which utilized DSM-IV criteria. The interview was originally translated into Mandarin by a team of psychiatrists in Shanghai Mental Health Centre, with the translation reviewed and modified by members of the CONVERGE team.

The replication sample was obtained from five hospitals in the north of China. Patients were diagnosed as having MDD by at least two consultant psychiatrists by DSM-IV criteria. Samples were of both sexes, and all four grandparents were Han Chinese. Cases were aged between 30 and 60, and had two or more episodes of MDD meeting DSM-IV criteria. Exclusion criteria were pregnancy, severe medical conditions, abnormal laboratory baseline values, unstable psychiatric features (e.g., suicidal), a history of alcoholism or drug abuse, epilepsy, brain trauma with loss of consciousness, neurological illness, or a concomitant Axis I psychiatric disorder. Control subjects were recruited from local communities and provided information about medical and family histories. Exclusion criteria were a history of major psychiatric or neurological disorders, psychiatric treatment or drug abuse, or family history of severe forms of psychiatric disorders.

The study protocol was approved centrally by the Ethical Review Board of Oxford University (Oxford Tropical Research Ethics Committee) and the ethics committees of all participating hospitals in China. All interviewers were mental health professionals who are

well able to judge decisional capacity. The study posed minimal risk (an interview and saliva sample). All participants provided their written informed consent.

DNA sequencing

DNA was extracted from saliva samples using the Oragene protocol. A barcoded library was constructed for each sample. Sequencing reads obtained from Illumina Hiseq machines were aligned to Genome Reference Consortium Human Build 37 patch release 5 (GRCh37.p5) with Stampy (v1.0.17) [21] using default parameters after filtering out reads containing adaptor sequencing or consisting of more than 50% poor quality (base quality <= 5) bases. Samtools (v0.1.18)[22] was used to index the alignments in BAM format[22] and Picardtools (v1.62) was used to mark PCR duplicates for downstream filtering. The Genome Analysis Toolkit's (GATK, version 2.6) [23] BaseRecalibrator was then run on the BAM files to create base quality score recalibration tables, masking known SNPs and INDELs from dbSNP (version 137, excluding all sites added after version 129). Base quality recalibration (BQSR) was then performed on the BAM files using GATKlite (v2.2.15)[23] while also removing read pairs that did not have the "properly aligned segment" bit set by Stampy (1-5% of reads per sample).

Calling and imputation of genotypes at SNPs included in GWAS

Variant discovery and genotyping at all polymorphic SNPs in 1000G Phase1 East Asian (ASN) reference panel [14] was performed simultaneously using post-BQSR sequencing reads from all samples using the GATK's UnifiedGenotyper (version 2.7-2-g6bda569). Variant quality score recalibration was then performed on these sites using the GATK's VariantRecalibrator (version 2.7-2-g6bda569) and the biallelic SNPs from 1000G Phase1 ASN samples as a true positive set of variants. A sensitivity threshold of 90% to SNPs in the 1000G Phase1 ASN panel was applied for SNP selection for imputation after optimizing for Transition to Transversion (TiTv) ratios in SNPs called. Genotype likelihoods (GLs) were calculated at selected sites using a sample-specific binomial mixture model implemented in SNPtools (version 1.0) [24], and imputation was performed at those sites without a reference panel using BEAGLE (version 3.3.2) [25]. A second round of imputation was performed with BEAGLE on the same GLs, but only at biallelic SNPs polymorphic in the 1000G Phase 1 ASN panel using the 1000G Phase 1 ASN haplotypes as a reference panel (see Supplemental Methods). A final set of allele dosages and genotype probabilities was generated from these two datasets by replacing the results in the former with those in the latter at all sites imputed in the latter. We then applied a conservative set of inclusion threshold for SNPs for genome-wide association study (GWAS): a) p-value for violation HWE $> 10^{-6}$, b) Information score > 0.9, c) MAF in CONVERGE $> 0.5\%$ to arrive at the final set of 6,242,619 SNPs for GWAS.

Sample selection for GWAS

Using both processed sequencing data and imputed dosages at SNPs that passed quality control, we assessed the sequencing and imputation quality of all 11,670 samples whose genomic variants we imputed. We first assessed the number of private variants called from processed sequencing data of all samples, and excluded 117 samples with excess number of private variants in the genic regions of their nuclear genome. Also using processed

sequencing data, we excluded a further 90 samples with excess number of heteroplasmic sites in their mitochondrial genome. Both filters were applied to exclude DNA samples that were likely contaminated. Using imputed data at all sites in the rest of the 11,463 samples, we excluded 29 samples for low imputation quality (maximum genotype probability < 0.9) in more than 10% of imputed sites. Using a subset of 399,211 tagging SNPs (MAF >1%, LD $R^2 < 0.5$, all known in 1000G Phase 1 ASN Panel) across all autosomes, we assessed the pairwise identity by state (IBS) between all remaining samples, and excluded 392 samples for being likely first-degree relatives. Finally, we excluded 590 samples who had incomplete phenotype information from the phenotyping interviews to arrive at the final 10,640 samples for GWAS. These analyses and filters are detailed in Supplemental Methods.

GWAS using linear mixed model and liability score estimates

We performed GWAS one chromosome at a time, with a mixed-linear model including a genetic relationship matrix (GRM) constructed from SNPs in the tagging set other than those on the chromosome being studied. This method is implemented in Factored Spectrally Transformed Linear Mixed Models (FastLMM version 2.06.20130802). Manhattan plots and quantile-quantile plots of the log10 of P-values of the GWAS were generated with custom code in R[26]. Genomic-control inflation factor lambda was calculated using custom code in R[26].

Replication and joint analyses

We genotyped the replication sample on a MassARRAY™ system mass spectrometer. TYPER4.0 was used to assess the reliability of genotype calls generated by SpectroREAD from the mass spectra. Default genotype call inclusion criteria were used. To perform the association analysis with MDD case-control status at these 12 sites in the replication sample, we obtained effect sizes for discovery from logistic regression with PC correction, and then for replication from logistic regression, and then performed fixed-effects meta-analysis.

Polygenic risk profiling and binomial sign-test

Single SNP association results were obtained from the Psychiatric Genetics Consortium (PGC) study of MDD[3]. Prior to analysis, SNPs were lifted-over to GRCh37/hg19 coordinates and excluded if (a) monomorphic in either European ($N = 379$) or East Asian ($N = 286$) populations from the 1000 Genomes Project Phase 1 reference data [14] or (b) absent from the filtered CONVERGE dataset. To construct the PGC-trained polygenic score, we initially selected autosomal SNPs with statistical imputation information (INFO) greater than 0.9 and MAF greater than 1% in both studies, and performed subsequent LD-based "clumping" to remove markers from highly-correlated SNP pairs (pairwise $R^2 > 0.2$ in East Asians, 500kb window) while preferentially retaining SNPs with smaller PGC P-values. Using the resultant SNP set, we constructed polygene scores based on varying P-value thresholds (10^{-6}, 10^{-5}, 10^{-4}, 10^{-3}, 0.01, 0.1, 0.2, 0.3, 0.4, 0.5, and 1) as previously described [28] We assessed the predictive value of polygenic scores in a genetically unrelated subset of CONVERGE (with pairwise relatedness less than 0.1) by logistic regression, adjusting for ancestry principal components (2) demonstrating significant association with MDD status. The estimated variance in MDD risk accounted for the polygenic score is given

by Nagelkerke's R^2. Using the same P-value thresholds, we tabulated the number of independent SNPs with the same direction of allelic effect in the PGC results as observed in CONVERGE. Filtering criteria for SNPs was INFO score greater than 0.9 in CONVERGE and MAF greater than 1% in both studies, and an analogous LD-clumping procedure was performed (pairwise $R^2 > 0.2$ in Europeans, 500kb window). A one-sided binomial sign test was used to assess whether this observed fraction was significantly greater than expected by chance.
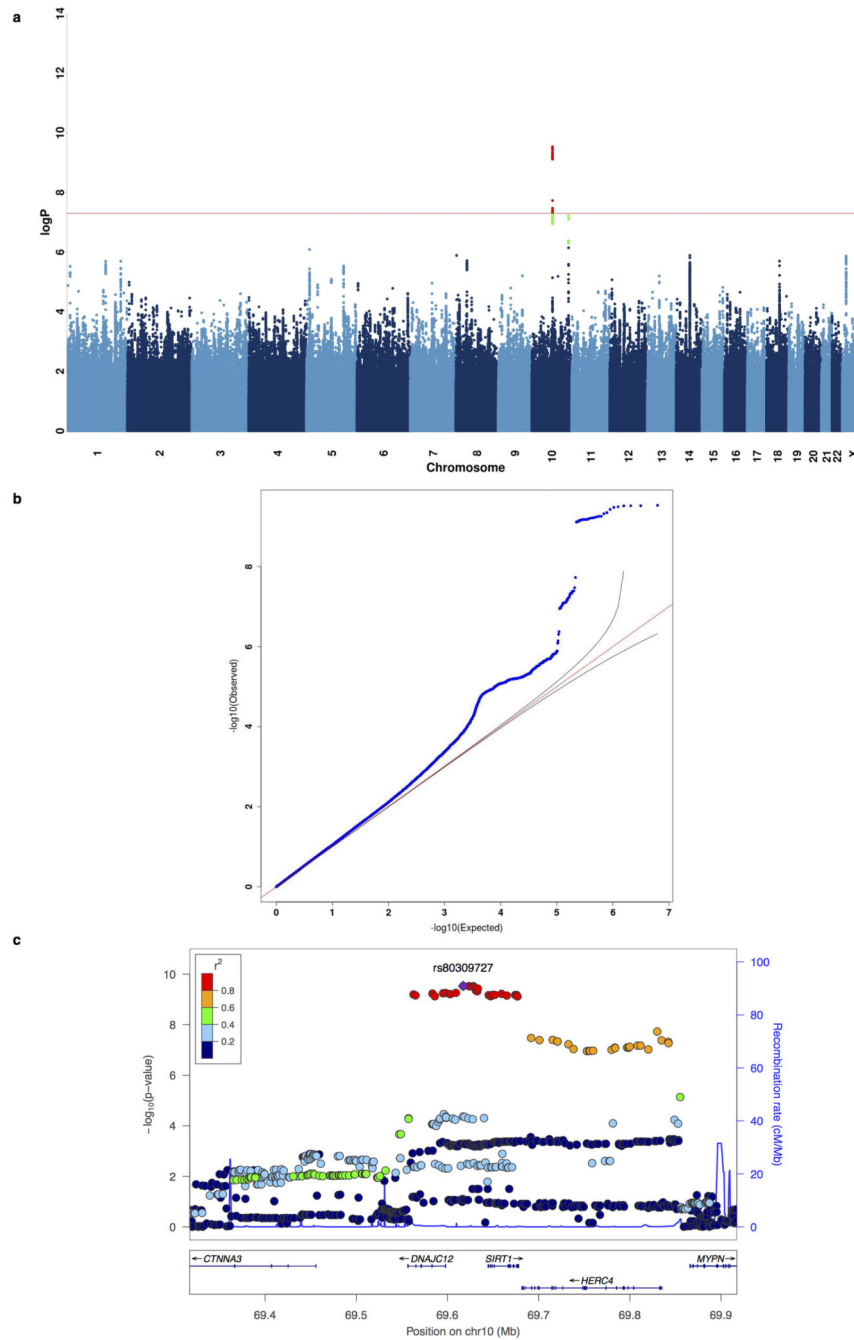
Extended Data
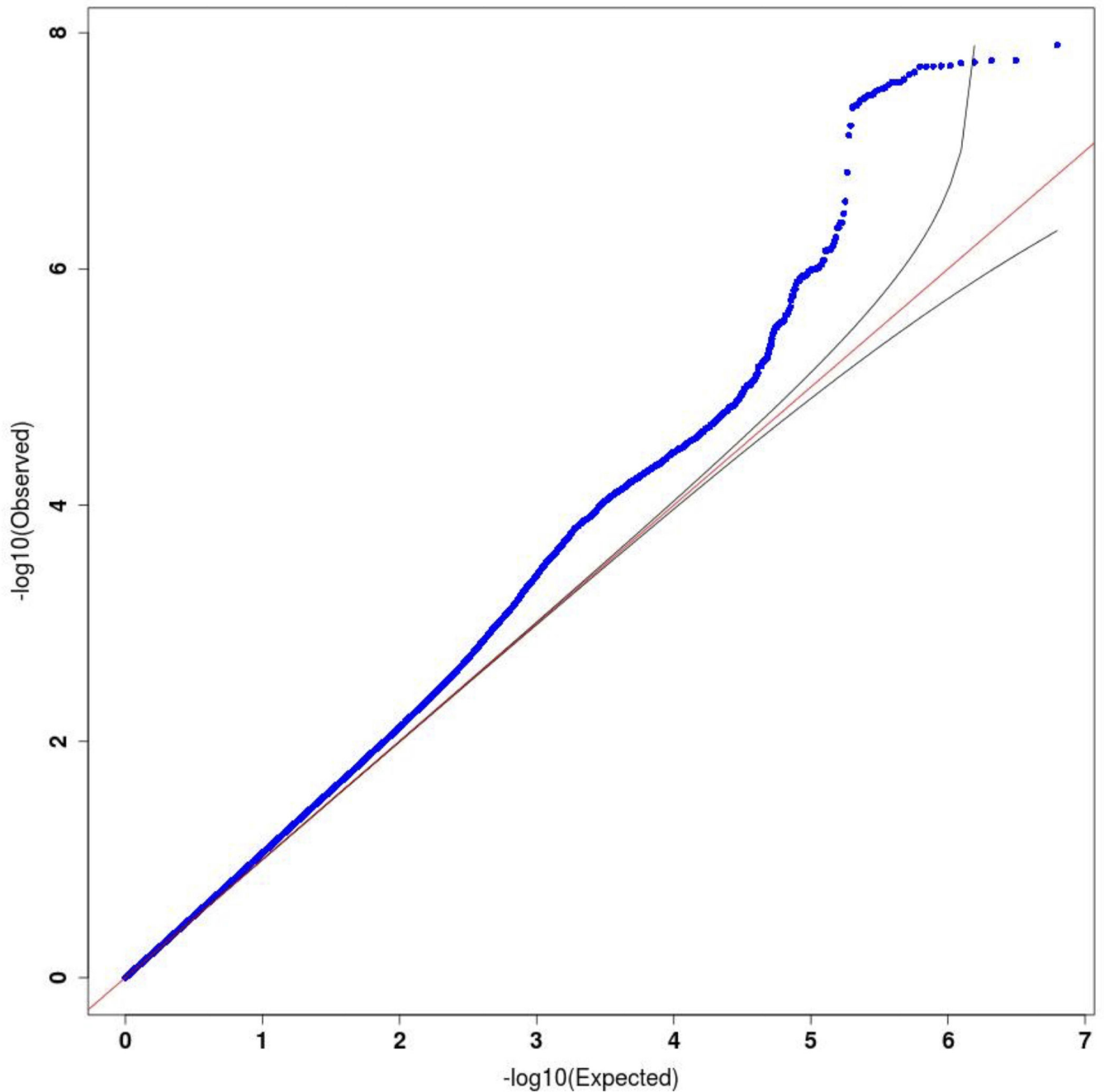
Extended Data Figure 1:

Quantile quantile plots for major depressive disorder.

Quantile-quantile plot of GWAS for MDD using the Mixed-Linear-Model with exclusion of chromosome marker is on (MLMe) method implemented in FastLMM on 10,640 samples (5,303 cases, 5,337 controls), genomic inflation factor lambda (λ) = 1.070, rescaled for an equivalent study of 1,000 cases and 1,000 controls (λ_{1000}) = 1.013.

Extended Data Figure 2:

Forest plots of estimated SNP effects in CONVERGE and PGCMDD1 studies.

This figure presents the association odds ratios (OR) at 12 SNPs in CONVERGE and the best available proxy SNPs in PGC MDD (pairwise $R^2 > 0.6$, 500kb window; the proxy SNP is marked by "*"). We present the alternative allele frequency (freq), odds ratio (or) with respect to the alternative allele, standard error of odds ratio (se) and P values of association (pval) for the following analyses (study): primary association analysis with a linear-mixed model using imputed allele dosages in 10,640 samples in CONVERGE (*pri*); validation
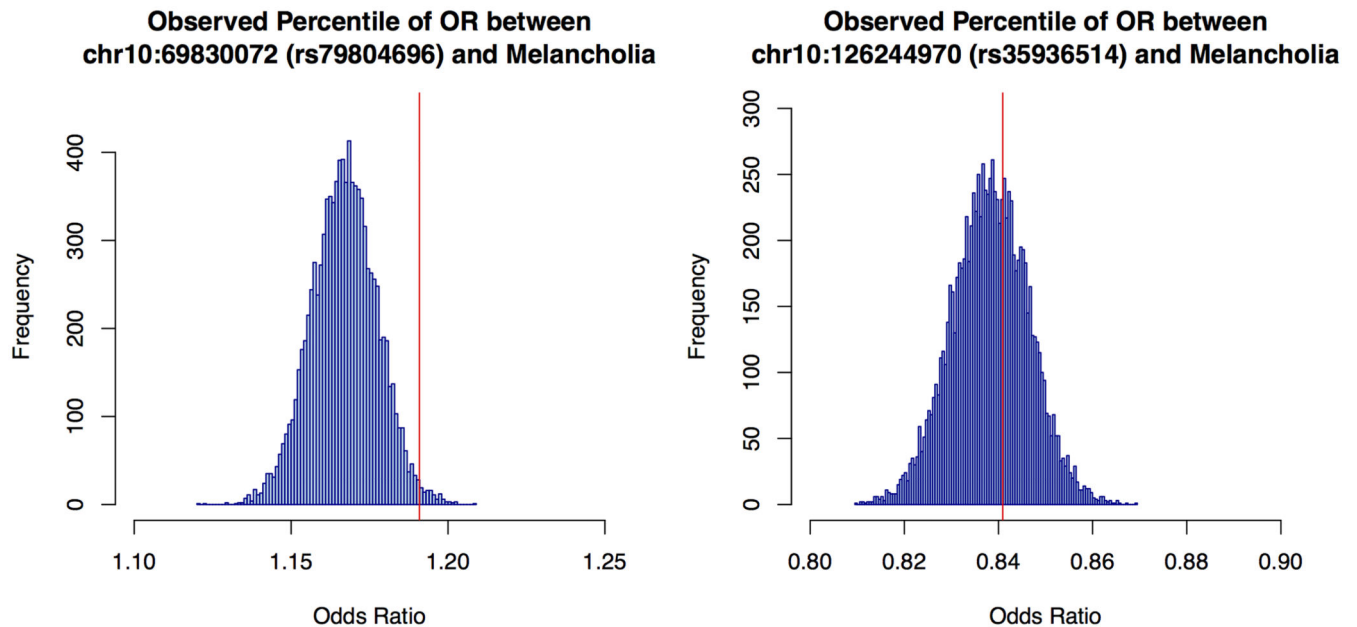
analysis with logistic regression model with PCs as covariates using genotypes from Sequenom on 9,921 samples in CONVERGE (*sqnm*); association with MDD with a logistic regression model in replication cohort of 6,417 samples using genotypes from Sequenom (*repli*); joint association analysis with MDD with a logistic regression model using imputed allele dosages in CONVERGE and genotypes from Sequenom in replication cohort (17,057 samples in total, *joint*).

Extended Data Figure 3:

Manhattan and quantile quantile plots for melancholia.

(a) Manhattan plot of GWAS for melancholia using the Mixed-Linear-Model with exclusion of chromosome marker is on (MLMe) method implemented in FastLMM on 9,846 samples (4,509 cases, 5,337 controls). (b) Quantile-quantile plot of GWAS for melancholia, genomic inflation factor lambda (λ) = 1.069, rescaled for an equivalent study of 1,000 cases and 1,000 controls (λ_{1000}) = 1.014. (c) Regional association plot on GWAS hit on chromosome 10 focusing on top SNP rs80309727 at 5′ of SIRT1 gene, generated with LocusZoom.

Observed Percentile of OR between chr10:69830072 (rs79804696) and Melancholia

Observed Percentile of OR between chr10:126244970 (rs35936514) and Melancholia

Extended Data Figure 4:

Empirical estimation of the odds ratio increases due to the removal of cases not falling under the diagnostic class of melancholia from an association analysis with major depression. The figures show the empirical distributions of the odds ratios for association with each of two SNPs (rs79804696, rs35936514), after removing a random set of 796 samples, equal to the number of cases of MDD not diagnosed as being melancholic. The horizontal axis is the odds ratio for each analysis, and the vertical axis the frequency of occurrence of the odds ratio in 10,000 analyses. The vertical red line is the observed odds ratio after removing cases of MDD not diagnosed as melancholic.

Extended Data Table 1:

Comparison between association results using imputed dosages and directly genotyped markers

		SNP			Imputed Dosages (N=9,921)					Sequenom genotypes		
CHR	POS	RSID	REF	ALT	OR	SE	P	N	R²	OR	SE	P
1	11493832	rs2922240	C	T	1.141	0.029	5.82E-06	9,864	0.991	1.141	0.029	5.72E-06
1	175151950	rs3766688	C	T	0.870	0.029	1.93E-06	9,901	0.995	0.871	0.029	2.32E-06
1	228052027	rs57047840	G	A	1.141	0.032	4.13E-05	9,724	0.974	1.141	0.032	3.91E-05
5	9161674	rs55713588	G	A	1.302	0.052	3.34E-07	9,636	0.925	1.263	0.050	2.87E-06
6	4386107	rs55800092	T	C	0.819	0.040	6.35E-07	9,881	0.992	0.817	0.040	5.52E-07
10	**69624180**	**rs12415800**	**A**	**G**	**1.167**	**0.029**	**8.44E-08**	**9,689**	**0.993**	**1.167**	**0.029**	**9.30E-08**
10	**126244970**	**rs35936514**	**T**	**C**	**0.845**	**0.033**	**2.91E-07**	**9,915**	**0.993**	**0.842**	**0.033**	**1.40E-07**
13	107659212	rs61967003	T	C	1.765	0.116	9.64E-07	9,914	0.997	1.748	0.116	1.53E-06
14	66833851	rs17827252	G	C	0.896	0.029	1.12E-04	8,562	0.999	0.897	0.031	3.94E-04
19	34493757	rs11880240	G	C	1.281	0.056	1.08E-05	9,912	0.996	1.281	0.056	1.14E-05
X	24656658	rs1921918	A	G	0.877	0.032	3.59E-05	9,899	0.994	0.880	0.032	6.39E-05
X	25011374	rs11573525	T	C	1.158	0.033	9.60E-06	9,912	0.969	1.144	0.032	3.21E-05

The table reports results for association between MDD and 12 SNPs. The first five columns give the chromosome (CHR), genomic position (POS), SNP identifier (RSID), reference allele (REF) on Human Genome Reference GRCh37.p5 and alternative allele (ALT) called in CONVERGE. The next three columns show results for imputed allele dosages at 12 SNPs (odds ratio (OR) of association with MDD with respect to the alternative allele and standard error (SE); P-values of association (P). The next two columns present the number of samples successfully genotyped using the Sequenom platform (a high sensitivity and specificity assay), and the Pearson correlation (R^2) between the imputed allele dosages and the genotypes from Sequenom. The final three columns present results of association with MDD using genotypes from the Sequenom genotyping platform. Bold type indicates the genome-wide significant markers: Extended Data Table 1b gives further information on the results for these markers.

Extended Data Table 2:

Genotype distribution and p-values for violation of Hardy-Weinberg Equilibrium in CONVERGE and replication cohorts

MDD Disease State	SNP	CONVERGE		Replication Cohort	
		HomRef/Het/HomAlt	HWE P-value	HomRef/Het/HomAlt	HWE P-value
All	rs12415800	2151/5301/3169	0.445	1212/3037/1920	0.857
	rs35936514	705/4054/5794	0.919	422/2400/3398	0.974
Cases	rs12415800	1178/2626/1490	0.741	654/1538/918	0.829
	rs35936514	318/1919/3027	0.549	190/1136/1783	0.627
Controls	rs12415800	973/2675/1679	0.106	558/1499/1002	0.971
	rs35936514	387/2135/2767	0.389	232/1264/1615	0.503

This table shows the number of samples with the homozygous reference genotype (HomRef), heterozygous genotypes (Het), and homozygous alternative genotype (HomAlt), as well as p-values for violation of Hardy-Weinberg equilibrium (HWE) for both CONVERGE study samples and the replication cohort from North China at the top SNPs rs12415800 in the *SIRT1* locus and rs35936514 in the *LHPP* locus from the GWAS on MDD. The top two rows show these measures for all samples in both the CONVERGE and replication study, the next two rows show that for just cases in CONVERGE and the replication cohort, and the last two shows show that for just the controls. The genotype distributions for CONVERGE are obtained from hard-called genotypes from maximum imputed genotype probabilities at for each sample at each of the two sites. As a genotype will not be called if the maximum genotype probability at a site is lower than 0.9 for any single sample, the total number of CONVERGE samples showing called HomRef/Het/HomAlt genotypes does not equal 10,640 for either SNP. For rs12415800, 19 (9 cases, 10 controls) samples have no genotype calls due to maximum genotype probability being smaller than 0.9, giving a total of 10621 CONVERGE (5,294 cases, 5,327 controls) samples with genotype calls. For rs35936514, 87 (39 cases, 48 controls) samples have no genotype calls due to maximum genotype probability being smaller than 0.9, giving a total of 10,553 (5,264 cases, 5,289 controls) CONVERGE samples with genotype calls.

Extended Data Table 3:

Single-marker association results of top CONVERGE hits in PGC-MDD1

					CONVERGE (10,640 samples)					PGC MDD (18,759 samples)				
CHR	POS	RSID	REF	ALT	FREQ	OR	SE	P	RSID	LD R2	FREQ	INFO	OR	P
1	11493832	rs2922240	C	T	0.3846	1.141	0.028	2.80E-06	rs2922240	1.00	0.495	0.999	0.947	0.011
1	175151950	rs3766688	C	T	0.394	0.875	0.028	1.83E-06	rs3766688	1.00	0.546	0.926	0.970	0.163
1	228052027	rs57047840	G	A	0.2843	1.138	0.031	4.64E-05	rs10916214	0.48	0.133	0.992	0.969	0.300
5	9161674	rs55713588	G	A	0.0956	1.278	0.050	6.04E-07	rs13360003	0.18	0.018	0.774	0.920	0.442
6	4386107	rs55800092	T	C	0.1512	0.824	0.039	1.35E-06	rs17138114	0.89	0.073	0.975	0.972	0.440
10	**69624180**	**rs12415800**	**A**	**G**	**0.4519**	**1.164**	**0.028**	**1.92E-08**	**rs16924945**	**0.32**	**0.005**	**0.355**	**1.629**	**0.496**
10	**126244970**	**rs35936514**	**T**	**C**	**0.2609**	**0.839**	**0.032**	**1.27E-08**	**rs35841851**	**0.98**	**0.023**	**0.92**	**0.970**	**0.553**
13	107659212	rs61967003	T	C	0.0172	1.645	0.109	6.70E-06	rs16969540	0.98	0.055	0.982	0.941	0.211
14	66833851	rs17827252	G	C	0.4624	0.887	0.028	1.44E-05	rs2319184	0.96	0.028	0.941	1.114	0.115
19	34493757	rs11880240	G	C	0.0679	1.291	0.055	8.02E-06	rs7254953	0.65	0.092	0.785	1.054	0.253
X	24656658	rs1921918	A	G	0.7206	0.883	0.031	3.22E-05	rs1921918	1.00	0.581	1.238	0.960	0.035
X	25011374	rs11573525	T	C	0.2602	1.160	0.032	5.86E-06	rs11573525	1.00	0.158	1.081	1.031	0.275

The table compares results from 12 SNPs genotyped in the CONVERGE cohort with either the same SNPs, or best available proxies within a 500kb window reported by the Psychiatric Genetics Consortium (PGC) MDD study. The first five columns give the SNP identifier (RSID), chromosome (CHR), genomic position (POS), reference allele (REF) on Human Genome Reference GRCh37.p5 and alternative allele (ALT) called in CONVERGE. The next four columns show the alternative allele frequency (FREQ) and results of association testing with MDD at the 12 SNPs in CONVERGE: odds ratio (OR) of association with MDD with respect to the alternative allele and standard error (SE) in the odds ratio were obtained from a linear-mixed model with a genetic relatedness matrix containing all samples. The next three columns show the SNP identifier (RSID) of best available proxy of each SNP reported in PGC MDD, the linkage disequilibrium correlation (LD R2) expressed as an R^2 between the SNP in PGC MDD and SNP in CONVERGE, and the alternative allele frequency (FREQ) at the SNP in PGC MDD. The last three columns show the information scores (INFO), odds ratios (OR) and P values of association with MDD in PGC MDD from a logistic regression model. Bold type indicates the genome-wide significant markers.

Extended Data Table 4:
Polygenic risk profiling and binomial sign tests

pT	Polygenic risk profiling		Binomial sign test	
	r^2	P	No. SNPs (%)	P
0.000001	0.000715	**0.0174**	3 (100)	0.125
0.00001	8.40E-05	0.415	12 (66.7)	0.194
0.0001	2.57E-05	0.652	62 (58.1)	0.126
0.001	5.87E-06	0.829	481 (53.6)	0.0605
0.01	8.67E-05	0.407	3632 (51.1)	0.101
0.1	0.00142	0.000797	26106 (50.4)	0.126
0.2	0.00126	0.00156	45166 (50.6)	0.00331
0.3	0.00116	0.00246	61074 (50.5)	0.00627
0.4	0.00125	0.00168	74676 (50.5)	0.00335
0.5	0.0011	0.00317	86429 (50.4)	0.00758
1	0.000924	0.00684	124361 (50.3)	0.0116

The table shows the predictive value of a Psychiatric Genetics Consortium (PGC) MDD-trained polygenic risk score on the CONVERGE results. Predictive values are shown at varying p-value thresholds (*pT*) from P<=1×10^{-6} to 1 (i.e. all results). P is the P-value of the prediction and r^2 the amount of variance explained (thus the table shows that including all SNPs from PGC-MDD explained 0.09% of MDD risk in CONVERGE.). The number of independent SNPs at each threshold is presented (*No. SNPs*); the significance of the observed fraction (%) demonstrating a consistent direction of effect was assessed by a one-sided binomial sign test.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

Author contributions:

Manuscript preparation: Na Cai, Tim Bigdeli, Warren Kretzschmar, Mark Reimers, Todd Webb, Brien Riley, Silviu Bacanu, Roseann E Peterson, Kenneth S Kendler, Jonathan Flint

Replication sample: Qi Xu

CONVERGE Sample collection: Yihan Li, Yiping Chen, Hong Deng, Wenhua Sang, Keqing Li, Jingfang Gao, Baowei Ha, Shugui Gao, Jian Hu, Chunmei Hu, Guoping Huang, Guoqing Jiang, Xiaoyan Zhou, Youhui Li, Kan Li, Qihui Niu, Yi Li, Gongying Li, Lanfen Liu, Zhengrong Liu, Yi Li, Xiang Fang, Runde Pan, Guodong Miao, Qiwen Zhang, Fengyu Yu, Guibing Chen, Min Cai, Donglin Yang, Xiaohong Hong, Yan Song, Chengge Gao,

Jiyang Pan, Yutang Zhang, Tieqiao Liu, Jicheng Dong, Xiaoping Wang, Lina Wang, Qiyi Mei, Zhenming Shen, Xiaojuan Liu, Wenyuan Wu, Danhua Gu, Yunchun Chen, Tiebang Liu, Han Rong, Ying Liu, Luxian Lv, Huaqing Meng, Hong Sang, Jianhua Shen, Tian Tian, Jianguo Shi, Jing Sun, Ming Tao, Xumei Wang, Jing Xia, Qiang He, Gang Wang, Xueyi Wang, Lijun Yang, Kerang Zhang, Ning Sun, Jinbei Zhang, Zhaoyu Gan, Zhen Zhang, Wei Zhang, Hui Zhong, Fuzhong Yang, Enzhao Cong, Shenxun Shi, Guangyi Fu, Jonathan Flint , Kenneth S Kendler

Genome sequencing and analysis: Jieqin Liang, Jingchu Hu, Qibin Li, Wei Jin, Zhenfei Hu, Guangbiao Wang, Linmao Wang, Puyi Qian, Yuan Liu, Tao Jiang, Yao Lu, Xiuqing Zhang, Ye Yin, Yingrui Li, Huanming Yang, Jian Wang, Xiangchao Gan, Yihan Li, Na Cai, Richard Mott, Jonathan Flint , Jun Wang, Xun Xu

Genotype imputation: Warren Kretzschmar, Jingchu Hu, Li Song, Qibin Li, Na Cai, Jonathan Marchini

Genetic analysis: Na Cai, Tim Bigdeli, Yihan Li, Roseann E Peterson, Silviu Bacanu, Todd Webb, Brien Riley, Kenneth S Kendler, Richard Mott, Jonathan Flint

Author information:

Na Cai 1*, Tim Bigdeli 2*, Warren Kretzschmar 1*, Yihan Li 1*, Jieqin Liang 3, Li Song 3, Jingchu Hu 3, Qibin Li 3, Wei Jin 3, Zhenfei Hu 3, Guangbiao Wang 3, Linmao Wang 3, Puyi Qian 3, Yuan Liu 3, Tao Jiang 3, Yao Lu 3, Xiuqing Zhang 3, Ye Yin 3, Yingrui Li 3, Xun Xu 3, Xiangchao Gan 4, Mark Reimers 2, Todd Webb 2, Brien Riley 2, Silviu Bacanu 2, Roseann E Peterson 2, Yiping Chen 5, Hui Zhong 6, Zhengrong Liu 7, Gang Wang 8, Jing Sun 9, Hong Sang 10, Guoqing Jiang 11, Xiaoyan Zhou 11, Yi Li 12,13, Wei Zhang 14, Xueyi Wang 15, Xiang Fang 16, Runde Pan 17, Guodong Miao 18, Qiwen Zhang 19, Jian Hu 20, Fengyu Yu 21, Bo Du 22, Wenhua Sang 22, Keqing Li 22, Guibing Chen 23, Min Cai 24, Lijun Yang 25, Donglin Yang 26, Baowei Ha 27, Xiaohong Hong 28, Hong Deng 29, Gongying Li 30, Kan Li 31, Yan Song 32, Shugui Gao 33, Jinbei Zhang 34, Zhaoyu Gan 34, Huaqing Meng 35, Jiyang Pan 36, Chengge Gao 37, Kerang Zhang 38, Ning Sun 38, Youhui Li 39, Qihui Niu 39, Yutang Zhang 40, Tieqiao Liu 41, Chunmei Hu 42, Zhen Zhang 43, Luxian Lv 44, Jicheng Dong 45, Xiaoping Wang 46, Ming Tao 47, Xumei Wang 48, Jing Xia 48, Han Rong 49, Qiang He 50, Tiebang Liu 51, Guoping Huang 52, Qiyi Mei 53, Zhenming Shen 54, Ying Liu 55, Jianhua Shen 56, Tian Tian 56, Xiaojuan Liu 57, Wenyuan Wu 58, Danhua Gu 59, Guangyi Fu 1, Yi Li 12,13, Jianguo Shi 60, Yunchun Chen 61, Jingfang Gao 62, Lanfen Liu 63, Lina Wang 63, Fuzhong Yang 64, Enzhao Cong 64, Jonathan Marchini 65,1, Huanming Yang 3, Jian Wang 3, Shenxun Shi 66,64, Richard Mott 1, Qi Xu 67 #, Jun Wang 3,68,69,70 #, Kenneth S Kendler 2 #, Jonathan Flint 71,1 #

* Denotes equal contribution; # Denotes corresponding author

1 Wellcome Trust Centre for Human Genetics, University of Oxford, Roosevelt Drive, Oxford, Oxfordshire, OX3 7BN, United Kingdom

2 Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University, Richmond, Virginia, 23298, USA

3 BGI-Shenzhen, Floor 9 Complex Building, Beishan Industrial Zone, YantianDistrict, Shenzhen, Guangdong, 518083, China

4 Department of Comparative Developmental Genetics, Max Planck Institute for Plant Breeding Research, Carl-von-Linne-Weg 10, Cologne, 50829, Germany

5 CTSU, Richard Doll Building, University of Oxford, Old Road Campus, Oxford, Oxfordshire, OX3 7LF, United Kingdom

6 Anhui Mental Health Center, No.316 Huangshan Road, Hefei, Anhui, 230000, China

7 Anshan Psychiatric Rehabilitation Hospital, No.127 Shuangshan Road,Lishan District, Anshan, Liaoning, 114000, China

8 Beijing Anding Hospital of Capital University of Medical Sciences, No.5 Ankang Hutong, Deshengmen wai, Xicheng District, Beijing, Beijing, 100000, China

9 Brain Hospital of Nanjing Medical University, No.264 Guangzhou Road, Nanjing, Jiangsu, 210000, China

10 Changchun Mental Hospital, No.4596 Beihuan Road, Changchun, Jilin, 130000, China

11 Chongqing Mental Health Center, No.102 Jinzishan,Jiangbei District, Chongqing, 404100, China

12 Dalian No.7 Hospital, No.179 Lingshui Road, Ganjingzi District, Dalian, Liaoning, 116000, China

13 Wuhan Mental Health Center, No.70, Youyi Road, Wuhan, Hubei, 430000, China

14 Daqing No.3 Hospital of Heilongjiang Province, No.54 Xitai Road, Ranghulu district, Daqing, Heilongjiang, 163000, China

15 First Hospital of Hebei Medical University, No.89 Donggang Road,, Shijiazhuang, Hebei, 50000, China

16 Fuzhou Psychiatric Hospital, No.451 South Erhuan Road,Cangshan District, Fuzhou, Fujian, 350000, China

17 Guangxi Longquanshan Hospital, No.1 Jila Road, Yufeng District, Liuzhou, Guangxi Zhuangzu, 545000, China

18 Guangzhou Brain Hospital, Guangzhou Psychiatric Hospital, No.36 Mingxin Road, Fangcun Avenue, Liwan District, Guangzhou, Guangdong, 510000, China

19 Hainan Anning Hospital, No.10 East Nanhai Avenue, Haikou, Hainan, 570100, China

20 Harbin Medical University, No.23 Youzheng street, Nangang District, Haerbin, Heilongjiang, 150000, China

21 Harbin No.1 Special Hospital, No.217 Hongwei Road, Haerbin, Heilongjiang, 150000, China

22 Hebei Mental Health Center, No.572 Dongfeng Road, Baoding, Hebei, 71000, China

23 Huaian No.3 Hospital, No.272 West Huaihai Road, Huaian, Jiangsu, 223001, China

24 Huzhou No.3 Hospital , No.255 Gongyuan Road, Huzhou, Zhejiang, 313000, China

25 Jilin Brain Hospital, No.98 West Zhongyang Road, Siping, Jilin, 136000, China

26 Jining Psychiatric Hospital , North Dai Zhuang,Rencheng District, Jining, Shandong, 272000, China

27 Liaocheng No. 4 Hospital, No.47 North Huayuan Road, Liaocheng, Shandong, 252000, China

28 Mental Health Center of Shantou University, No.243 Daxue Road, Shantou, Guangdong, 515000, China

29 Mental Health Center of West China Hospital of Sichuan University, No.28 South Dianxin Street, Wuhou District, Chengdu, Sichuan, 610000, China

30 Mental Health Institute of Jining Medical College, Dai Zhuang, Bei Jiao, Jining, Shandong, 272000, China

31 Mental Hospital of Jiangxi Province, No.43 Shangfang Road, Nanchang, Jiangxi, 330000, China

32 Mudanjiang Psychiatric Hospital of Heilongjiang Province, Xinglong, Mudanjiang, Heilongjiang, 157000, China

33 Ningbo Kang Ning Hospital, No.1 Zhuangyu Road ,Zhenhai District, Ningbo, Zhejiang, 315000, China

34 No. 3 Hospital of Sun Yat-sen University, No.600 Tianhe Road, Tianhe District, Guangzhou, Guangdong, 510630, China

35 No.1 Hospital of Chongqing Medical University, No.1 Youyi Road,Yuanjiagang,Yuzhong District, Chongqing, Chongqing, 400016, China

36 No.1 Hospital of Jinan University, No.613 West Huangpu Avenue, Guangzhou, Guangdong, 510000, China

37 No.1 Hospital of Medical College of Xian Jiaotong University, No. 277 West Yan Ta Road, Xian, Shaan Xi , 710061, China

38 No.1 Hospital of Shanxi Medical University, No.85 South Jiefang Road, Taiyuan, Shanxi, 30000, China

39 No.1 Hospital of Zhengzhou University, No.1 East Jianshe Road, Zhengzhou, Henan, 450000, China

40 No.2 Hospital of Lanzhou University, No.82, Cuiyingmen, Lanzhou, Gansu, 730000, China

41 No.2 Xiangya Hospital of Zhongnan University, No.139 Middle Renmin Road,Furong District, Changsha, Hunan, 410000, China

42 No.3 Hospital of Heilongjiang Province, No.135 Jiaotong Road, Beian, Heilongjiang, 164000, China

43 No.4 Hospital of Jiangsu University, No.246 Nanmen Street, Zhenjiang, Jiangsu, 212000, China

44 Psychiatric Hospital of Henan Province, No.388 Middle Jianshe Road, Xinxiang, Henan, 453000, China

45 Qingdao Mental Health Center, No.299 Nanjing Road, Shibei District, Qingdao, Shandong, 266000, China

46 Renmin Hospital of Wuhan University, No.238 Jiefang Road, Wuchang District, Wuhan, Hubei, 430000, China

47 Second Affiliated Hospital of Zhejiang Chinese Medical University, No.318 Chaowang Road, Hangzhou, Zhejiang, 310000, China

48 ShengJing Hospital of China Medical University, No.36 Sanhao Street, Heping District, Shenyang, Liaoning, 110001, China

49 Shenzhen Key Lab for Psychological Healthcare, Kangning Hospital, No.1080, Cuizhu Street, Luohu District, Shenzhen, Guangdong, 518000, China

50 Department of General Internal Medicine, Kanazawa Medical University, Kahoku, Ishikawa, 920-0293, Japan

51 Shenzhen Key Lab for Psychological Healthcare;Shenzhen Kangning Hospital, No.1080, Cuizhu Street, Luohu District , Shenzhen, Guangdong, 518000, China

52 Sichuan Mental Health Center, No.190, East Jiannan Road, Mianyang, Sichuan, 621000, China

53 Suzhou Guangji Hospital, No.286, Guangji Road, Suzhou, Jiangsu, 215000, China

54 Tangshan No.5 Hospital, No.57 West Nanxin Road, Lunan District, Tangshan, Hebei, 63000, China

55 The First Hospital of China Medical University, No.155 North Nanjing Street, Heping District, Shenyang, Liaoning, 110001, China

56 Tianjin Anding Hospital, No.13 Liulin Road, Hexi District, Tianjin, 300000, China

57 Tianjin First Center Hospital, No.55 Xuetang Street, Xinkai Road, Hedong District, Tianjin, 300000, China

58 Tongji University Hospital, No.389 Xinchun Road, Shanghai, 200000, China

59 Weihai Mental Health Center, Qilu Avenue, ETDZ, Weihai, Shandong, 264200, China

60 Xian Mental Health Center, No.15 Yanyin Road, New Qujiang District, Xian, Shaanxi, 710000, China

61 Xijing Hospital of No.4 Military Medical University, No.17 West Changle Road, Xian, Shaanxi, 710000, China

62 Zhejiang Traditional Chinese Medical Hospital, No.54 Youdian Road, Hangzhou, Zhejiang, 310000, China

63 Shandong Mental Health Center, No.49 East Wenhua Road, Jinan, Shandong, 250000, China

64 Shanghai Jiao Tong University School of Medicine, Shanghai Mental Health Centre, No. 600 Wan Ping Nan Road, Shanghai, 200030, China

65 Department of Statistics, University of Oxford, Oxford, Oxfordshire, OX1 3TG, United Kingdom

66 Fudan University affiliated Huashan Hospital, No. 12 Wulumuqi Zhong Road, Shanghai, 200040, China

67 National Laboratory of Medical Molecular Biology, Institute of Basic Medical Sciences & Neuroscience Center, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, 10005, China

68 Department of Biology, University of Copenhagen, Ole Maal Oes Vej 5, Copenhagen, 2200, Denmark

69 Macau University of Science and Technology, Avenida Wai long, Taipa, Macau 999078, China, Taipa, Macau, 999078, China

70 Princess Al Jawhara Center of Excellence in the Research of Hereditary Disorders, King Abdulaziz University, Jeddah, 21589, Saudi Arabia

71 East China Normal University, 3663 North Zhongshan Road, Shanghai, 200062, China

References

1. Kessler RC, Bromet EJ. The epidemiology of depression across cultures. Annual review of public health. 2013; 34:119–138.

2. Flint J, Kendler KS. The genetics of major depression. Neuron. 2014; 81:484–503. [PubMed: 24507187]

3. Major Depressive Disorder Working Group of the Psychiatric, G. C. et al. A mega-analysis of genome-wide association studies for major depressive disorder. Mol Psychiatry. 2013; 18:497–511. [PubMed: 22472876]

4. Foley DL, et al. Genetic and environmental risk factors for depression assessed by subject-rated symptom check list versus structured clinical interview. Psychol Med. 2001; 31:1413–1423. [PubMed: 11722156]

5. Kendler KS, et al. Clinical indices of familial depression in the Swedish Twin Registry. Acta Psychiatrica Scandinavica. 2007; 115:214–220. [PubMed: 17302621]

6. Sullivan PF, et al. Genetic epidemiology of major depression: review and meta-analysis. Am J Psychiatry. 2000; 157:1552–1562. [PubMed: 11007705]

7. Li Y, et al. Low-coverage sequencing: implications for design of complex trait association studies. Genome Res. 2011; 21:940–951. [PubMed: 21460063]

8. Lippert C, et al. FaST linear mixed models for genome-wide association studies. Nat Methods. 2011; 8:833–835. [PubMed: 21892150]

9. Widmer C, et al. Further improvements to linear mixed models for genome-wide association studies. Scientific reports. 2014; 4:6874. [PubMed: 25387525]

10. Angst J, et al. Melancholia and atypical depression in the Zurich study: epidemiology, clinical characteristics, course, comorbidity and personality. Acta Psychiatr Scand Suppl. 2007:72–84. [PubMed: 17280573]

11. Kendler KS. The diagnostic validity of melancholic major depression in a population-based sample of female twins. Arch Gen Psychiatry. 1997; 54:299–304. [PubMed: 9107145]

12. Sun N, et al. A comparison of melancholic and nonmelancholic recurrent major depression in Han chinese women. Depress Anxiety. 2011; 29:4–9. [PubMed: 22065498]

13. Pasaniuc B, et al. Extremely low-coverage sequencing and imputation increases power for genome-wide association studies. Nat Genet. 2012; 44:631–635. [PubMed: 22610117]

14. Abecasis GR, et al. An integrated map of genetic variation from 1,092 human genomes. Nature. 2012; 491:56–65. [PubMed: 23128226]

15. Lee S, et al. The epidemiology of depression in metropolitan China. Psychol Med. 2009; 39:735–747. [PubMed: 18713484]

16. Kessler RC, et al. The epidemiology of major depressive disorder: results from the National Comorbidity Survey Replication (NCS-R). Jama. 2003; 289:3095–3105. [PubMed: 12813115]

17. Liao SC, et al. Low prevalence of major depressive disorder in Taiwanese adults: possible explanations and implications. Psychol Med. 2012; 42:1227–1237. [PubMed: 22051196]

18. Gerhart-Hines Z, et al. Metabolic control of muscle mitochondrial function and fatty acid oxidation through SIRT1/PGC-1alpha. EMBO J. 2007; 26:1913–1923. [PubMed: 17347648]

19. Cai N, et al. Molecular signatures of major depression. Current Biology. 2015 in press.

20. Association, A. P.. Diagnostic and statistical manual of mental disorders. 4th edn. American Psychiatric Association; 1994.

21. Lunter G, Goodson M. Stampy: a statistical algorithm for sensitive and fast mapping of Illumina sequence reads. Genome Res. 2011; 21:936–939. [PubMed: 20980556]

22. Li H, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009; 25:2078–2079. [PubMed: 19505943]

23. McKenna A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010; 20:1297–1303. [PubMed: 20644199]

24. Wang Y, et al. An integrative variant analysis pipeline for accurate genotype/haplotype inference in population NGS data. Genome Res. 2013; 23:833–842. [PubMed: 23296920]

25. Browning SR, Browning BL. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. Am J Hum Genet. 2007; 81:1084–1097. [PubMed: 17924348]

26. R-Development-Core-Team. A language and environment for statistical computing. R Foundation for Statistical Computing; 2004.

27. Purcell S, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007; 81:559–575. [PubMed: 17701901]

28. Dudbridge F. Power and predictive accuracy of polygenic risk scores. PLoS Genet. 2013; 9:e1003348. [PubMed: 23555274]

29. Pruim RJ, et al. LocusZoom: regional visualization of genome-wide association scan results. Bioinformatics. 2010; 26:2336–2337. [PubMed: 20634204]
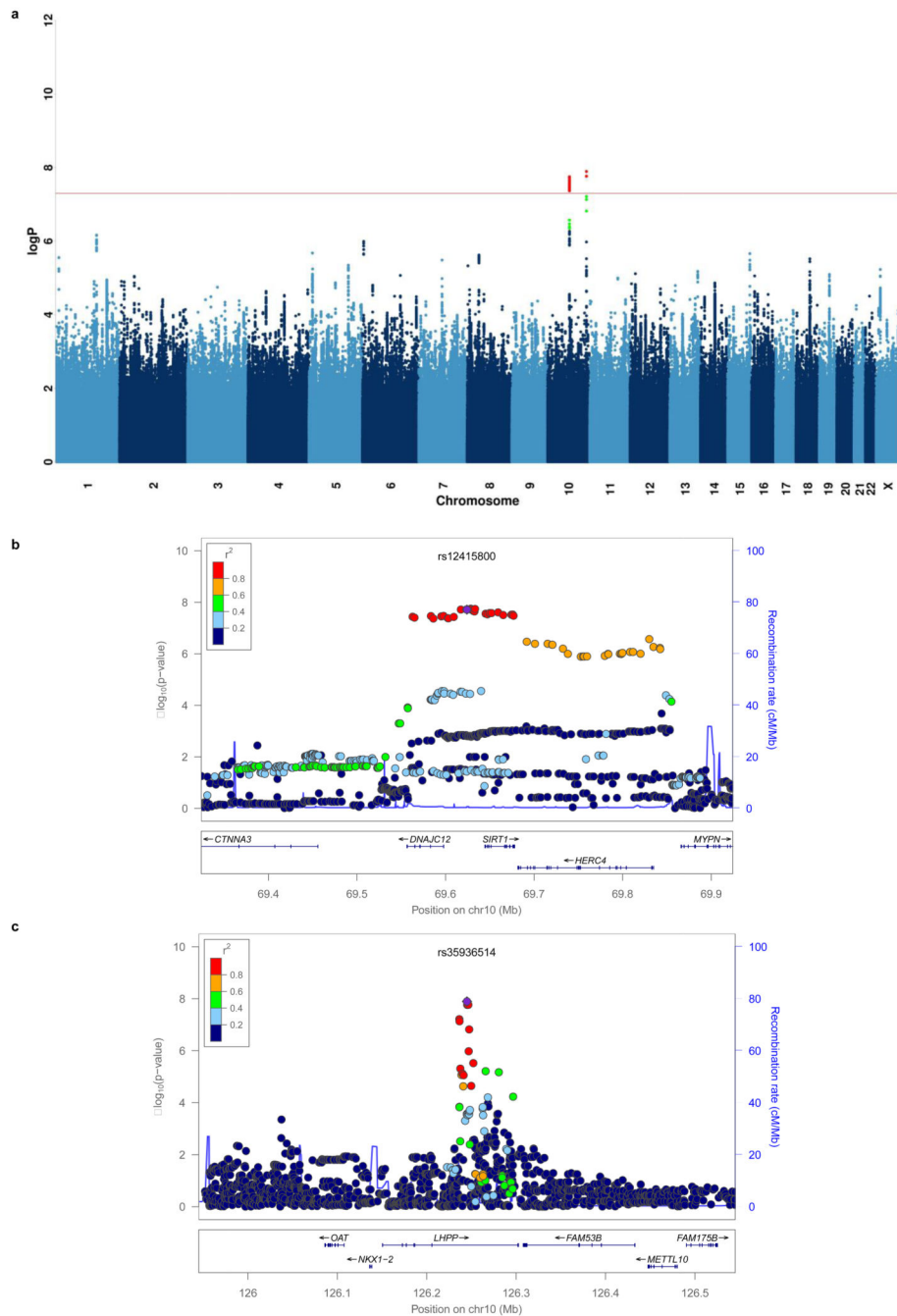
Figure 1.

Two loci associated with MDD in the CONVERGE sample

(a) Manhattan plot of genome wide association for MDD, (b) association at the *SIRT1* region on chromosome 10 at 69.6 Mb and (c) the *LHPP* gene on chromosome 10 at position 126.2 Mb. The −log10 P-values of imputed SNPs associated with MDD are shown on the left y axis. The recombination rates (NCBI Build GRCh37), (light-blue lines), are shown on the right y axis. Position in megabases (Mb) is on the x axis. Linkage disequilibrium of each

SNP with top SNP, shown in large purple diamond, is indicated by its colour. The plots were drawn using LocusZoom[29].

Table 1

Genetic association between MDD and 12 variants in CONVERGE and a replication sample

	SNP				CONVERGE (N=10,640)					Replication (N=6,417)			Joint (N=17,057)		
CHR	POS	RSID	REF	ALT	FREQ	INFO	OR	SE	P	OR	SE	P	OR	SE	P
1	11493832	rs2922240	T	C	0.385	1.018	1.141	0.028	2.80E-06	0.949	0.037	1.54E-01	1.070	0.022	2.46E-03
1	175151950	rs3766688	T	C	0.394	1.003	0.875	0.028	1.83E-06	0.991	0.037	8.15E-01	0.918	0.022	1.34E-04
1	228052027	rs57047840	A	G	0.284	0.970	1.138	0.031	4.64E-05	1.001	0.041	9.90E-01	1.088	0.025	5.57E-04
5	9161674	rs55713588	A	G	0.096	0.893	1.278	0.050	6.04E-07	1.054	0.062	3.93E-01	1.042	0.035	2.08E-01
6	4386107	rs55800092	C	T	0.151	1.001	0.824	0.039	1.35E-06	0.962	0.052	4.49E-01	0.876	0.031	1.82E-05
10	**69624180**	**rs12415800**	**G**	**A**	**0.452**	**0.992**	**1.164**	**0.028**	**1.92E-08**	**1.130**	**0.036**	**7.71E-04**	**1.150**	**0.022**	**2.37E-10**
10	**126244970**	**rs35936514**	**C**	**T**	**0.260**	**0.993**	**0.839**	**0.032**	**1.27E-08**	**0.838**	**0.041**	**1.68E-05**	**0.842**	**0.025**	**6.43E-12**
13	107659212	rs61967003	C	T	0.017	0.999	1.645	0.109	6.70E-06	0.788	0.150	1.11E-01	1.277	0.087	4.81E-03
14	66833851	rs17827252	C	G	0.463	1.011	0.887	0.028	1.44E-05	0.962	0.041	3.41E-01	0.907	0.023	2.20E-05
19	34493757	rs11880240	C	G	0.068	1.019	1.291	0.055	8.02E-06	1.048	0.072	5.12E-01	1.184	0.043	9.15E-05
X	24656658	rs1921918	A	G	0.721	0.995	0.883	0.031	3.22E-05	0.994	0.047	9.01E-01	0.917	0.026	1.09E-03
X	25011374	rs11573525	C	T	0.260	0.971	1.160	0.032	5.86E-06	1.011	0.047	8.19E-01	1.100	0.027	2.18E-04

The table reports results for 12 SNPs in the CONVERGE and replication samples. The first five columns give the chromosome (CHR), genomic position (POS), SNP identifier (RSID), reference allele (REF) on Human Genome Reference GRCh37.p5 and alternative allele (ALT) called in CONVERGE. The next five columns show the alternative allele frequency (FREQ) and results of association testing with MDD using imputed allele dosages in 10,640 CONVERGE samples (5,303 cases, 5,337 controls); information scores (INFO), odds ratio (OR) of association with MDD with respect to the alternative allele and standard error (SE) in the odds ratio were obtained from a logistic regression model; P values of association (P) obtained from a linear-mixed model with a genetic relatedness matrix containing all samples. The next three columns present results of association with MDD in the replication cohort of 6,417 samples (3,231 cases, 3,186 controls) from a logistic regression model. The final three columns present results of association with MDD in a joint analysis with both CONVERGE and replication cohort from a logistic regression model. Bold type indicates the genome wide significant markers.