ORIGINAL ARTICLE

Novel loci for major depression identified by genome-wide association study of Sequenced Treatment Alternatives to Relieve Depression and meta-analysis of three studies

SI Shyn[1,12], J Shi[2,12], JB Kraft[1], JB Potash[3], JA Knowles[4], MM Weissman[5], HA Garriock[1], JS Yokoyama[1], PJ McGrath[5], EJ Peters[1], WA Scheftner[6], W Coryell[7], WB Lawson[8], D Jancic[3], PV Gejman[9], AR Sanders[9], P Holmans[10], SL Slager[11], DF Levinson[2] and SP Hamilton[1]

[1]Department of Psychiatry and Institute for Human Genetics, University of California, San Francisco, CA, USA; [2]Department of Psychiatry and Behavioral Sciences, Stanford University, Palo Alto, CA, USA; [3]Department of Psychiatry, Johns Hopkins University, Baltimore, MD, USA; [4]Department of Psychiatry, University of Southern California, Los Angeles, CA, USA; [5]Department of Psychiatry, Columbia University College of Physicians and Surgeons and New York State Psychiatric Institute, New York, NY, USA; [6]Department of Psychiatry, Rush University Hospital, Chicago, IL, USA; [7]Department of Psychiatry, University of Iowa, Iowa City, IA, USA; [8]Department of Psychiatry, Howard University, Washington, DC, USA; [9]NorthShore University HealthCare, Evanston, IL, USA; [10]Department of Psychological Medicine, Cardiff University, Cardiff, UK and [11]Department of Health Sciences Research, Mayo Clinic College of Medicine, Rochester, MN, USA

We report a genome-wide association study (GWAS) of major depressive disorder (MDD) in 1221 cases from the Sequenced Treatment Alternatives to Relieve Depression (STAR*D) study and 1636 screened controls. No genome-wide evidence for association was detected. We also carried out a meta-analysis of three European-ancestry MDD GWAS data sets: STAR*D, Genetics of Recurrent Early-onset Depression and the publicly available Genetic Association Information Network–MDD data set. These data sets, totaling 3957 cases and 3428 controls, were genotyped using four different platforms (Affymetrix 6.0, 5.0 and 500 K, and Perlegen). For each of 2.4 million HapMap II single-nucleotide polymorphisms (SNPs), using genotyped data where available and imputed data otherwise, single-SNP association tests were carried out in each sample with correction for ancestry-informative principal components. The strongest evidence for association in the meta-analysis was observed for intronic SNPs in $ATP6V1B2$ ($P = 6.78 \times 10^{-7}$), $SP4$ ($P = 7.68 \times 10^{-7}$) and $GRM7$ ($P = 1.11 \times 10^{-6}$). Additional exploratory analyses were carried out for a narrower phenotype (recurrent MDD with onset before age 31, $N = 2191$ cases), and separately for males and females. Several of the best findings were supported primarily by evidence from narrow cases or from either males or females. On the basis of previous biological evidence, we consider $GRM7$ a strong MDD candidate gene. Larger samples will be required to determine whether any common SNPs are significantly associated with MDD.
Molecular Psychiatry (2011) 16, 202–215; doi:10.1038/mp.2009.125; published online 29 December 2009
Keywords: major depressive disorder; genetics; GWAS; meta-analysis; neuroscience

Introduction

Major depressive disorder (MDD) is the leading cause of disability for adults under 45 years of age,[1] and has a lifetime incidence of 12–20%.[2] Twin studies suggest a heritability of approximately 40% (perhaps higher in clinical samples), with a two- to threefold increased risk to first-degree relatives of MDD probands.[3] There are no established neurobiological mechanisms or definitive genetic associations. In this study, we report on a new genome-wide association study (GWAS) of MDD in the Sequenced Treatment Alternatives to Relieve Depression (STAR*D) sample, and on a meta-analysis of STAR*D and two other data sets: the Genetics of Recurrent Early-onset Depression (GenRED) GWAS reported in a companion article;[4] and Genetic Association Information Network–MDD (GAIN–MDD), a data set that was analyzed in the first MDD GWAS report[5] and that has been made available to scientists through the database of Genotypes and Phenotypes repository (dbGaP).[6]

The new GWAS sample includes 1221 cases from STAR*D, a multi-center, National Institute of Mental Health (NIMH)-sponsored antidepressant clinical trial.[7,8] The GenRED GWAS[4] included 1020 cases, with 1636 controls from the Molecular Genetics of Schizophrenia (MGS) study[9] (excluding controls who reported any history of MDD). The STAR*D analysis uses the same control data, and our meta-analysis corrects for that overlap. We accessed the GAIN–MDD data set and carried out a new analysis (for methodological consistency) of 1715 cases and 1792 controls, slightly smaller than the published sample[5] but with very similar results.

Genome-wide association study methods evaluate the contribution of common single-nucleotide polymorphisms (SNPs) to common diseases. They have identified robust associations to many non-psychiatric disorders[10] and to bipolar disorder,[11] schizophrenia[12–14] and autism.[15] No genome-wide significant findings were reported for GAIN–MDD[5] or GenRED,[4] or for a GWAS (not included in this meta-analysis) of 1514 recurrent MDD cases and 2052 controls (without lifetime depressive or anxiety disorders) from a German clinical sample and a Swiss population-based sample.[16] This is not surprising, as most GWAS findings have emerged when multiple data sets were combined to achieve large sample sizes (often 10 000–20 000 cases plus controls) with power to detect variants that produce small increases in risk.[10] We have reported separate GenRED and STAR*D analyses, because their distinctive characteristics could prove relevant to interpreting results across studies in the future, but to achieve a larger sample size we also report a meta-analysis of STAR*D, GenRED and GAIN–MDD data.

Materials and methods

Subjects

*STAR*D.* Cases were participants in STAR*D. Individuals (ages 18–75) were enrolled from primary care or psychiatric outpatient clinics if they had a diagnosis of MDD (by clinician rating of Diagnostic and Statistical Manual of Mental Disorders, 4th edition (DSM-IV) criteria) and a current 17-item Hamilton Depression Rating Scale (HAM-D) score of ⩾14 by independent raters[7,8] (although that score did not capture the severity of past depression). Of 1953 participants who donated DNA, we selected the 1500 who self-identified as 'white' as they represented most of the sample and European-ancestry controls were available. After quality control (QC) procedures (described below), 1221 cases were available for analyses. All subjects signed informed consent for genetic studies. Work described here was approved by the institutional review board of the University of California, San Francisco.

Controls were the same as those used in the GenRED GWAS analysis.[4] Details are described elsewhere.[9,13] They were recruited for MGS by a survey research company (Knowledge Networks, Inc., Menlo Park, CA, USA) from a nationally representative internet-based panel that was selected by random digit dialing. Participants had completed an online version of the Composite International Diagnostic Interview-Short Form[17] for lifetime history of common mood, anxiety and substance use disorders. They consented to anonymization and deposition of their DNA and clinical information in the NIMH repository for use in any medical research. The 1636 European-ancestry controls used here had no lifetime history of MDD (or of recurrent depression missing MDD by one criterion) by Composite International Diagnostic Interview-Short Form criteria (which over-diagnose MDD[18]). The MGS collaboration gave permission for us to use genotypes for the part of the control sample that is still under a dbGaP publication embargo. Clinical and demographic characteristics are summarized in Table 1.

Meta-analysis. GenRED cases (N = 1020) were recruited from multiple clinical settings and media and internet announcements and advertisements. Cases were assessed with the diagnostic interview for genetic studies[19] (version 3; http://nimhgenetics.org) and consensus best estimate diagnoses were assigned by review of Diagnostic Interview for Genetic Studies, informant report and available psychiatric records.[4] Probands had recurrence (two or more episodes, or one episode lasting at least 3 years), onset before age 31, and recurrent MDD in a sibling or parent with onset before age 41 (but no suspected bipolar-I disorder in a sibling or parent), features that predict greater familial liability to MDD.[3,20,21] The GenRED GWAS used the same MGS controls as STAR*D (see above). *GAIN–MDD* recruited individuals from a twin registry and two population-based samples in the Netherlands, selecting

Table 1 Demographics of STAR*D participants

	Cases	Controls
N	1221	1636
Age	42.8 ± 13.6	52.5 ± 17.2
Female	58.6% (715)	43.9% (718)
Initial HAM-D	18.4 ± 6.6	
Age of first MDE	27.3 ± 12.9	
Recurrent MDD	73.8% (901)	
Presence of comorbid anxiety disorder (GAD, panic, social phobia, OCD)	463 (37.9%)	

Abbreviations: GAD, generalized anxiety disorder; HAM-D, Hamilton Depression Rating Scale; MDD, major depressive disorder; MDE, major depressive episode; STAR*D, Sequenced Treatment Alternatives to Relieve Depression; OCD, obsessive-compulsive disorder.
Shown are mean ± s.d. for age, first HAM-D score after study entry, and age of first MDE; and numbers for other variables (i.e., first major depressive episode) are percents, with corresponding *N*s in parentheses.

cases who received MDD diagnoses based on a Composite International Diagnostic Interview (CIDI), and controls without MDD and without high neuroticism scores.[5] Each study excluded bipolar disorder, schizophrenia or schizoaffective disorder and more severe substance use disorders, with minor differences in exclusion criteria.

For *meta-analysis*, we defined two phenotypic models: *Broad* (all 3957 MDD cases from the three samples, vs 3428 controls), and *Narrow* (2191 cases with onset before age 31 and recurrence, including GenRED chronic cases). We did not require positive family history because STAR*D and GAIN–MDD assessed this by proband response to a single question. Exploratory separate analyses of males and females were carried out for each phenotype, because females are at a twofold increased risk, and twin studies suggest partial independence of genetic risk factors for females and males.[22,23] Characteristics of the three samples are summarized in Table 2.

Software

Genotypic data were managed and analyzed using PLINK v1.04–1. 06, except for imputation analyses and analysis of imputed data as described below and in Supplementary Methods.[24] STAR*D results were compiled and visualized using WGAViewer v1.25T-Z[25] and HaploView v4.1.[26]

Genotyping

*STAR*D cases.* Genotyping was conducted for 754 cases by Affymetrix, Inc. (South San Francisco, CA, USA), with the Affymetrix GeneChip Human Mapping 500 K Array Set and genotypes called with the Bayesian Robust Linear Model with Mahalanobis distance classifier.[27] We genotyped the remaining 746 cases with the Affymetrix Genome-Wide Human SNP 5.0 Array and called genotypes with the updated Bayesian Robust Linear Model with Mahalanobis-P algorithm. There were 500 568 SNPs that were assayed by both arrays.

The *GenRED cases* and *MGS controls* were genotyped at the Broad Institute on the Affymetrix Genome-Wide Human SNP 6.0 Array, and genotypes were called with Birdseed version 2.[4,13] The *GAIN–MDD* sample was genotyped with the Perlegen platform.[5]

Quality control analyses

*Single-nucleotide polymorphisms. STAR*D:* DNA samples were genotyped on three related platforms: cases on Affymetrix 500 K and 5.0, and controls on Affymetrix 6.0, resulting in 382 598 SNPs that were assayed on all three platforms and that passed QC for the MGS/GenRED controls. To ensure consistency of results, we then excluded SNPs for all samples in the STAR*D analysis based on cross-platform data as follows:

Table 2 Samples and SNPs included in meta-analysis

	GAIN	GenRED	STAR*D	Total
All subjects (males + females)				
Broad cases	1716	1020	1221	3957
Narrow cases	469	1020	702	2191
Controls	1792		1636[a]	3428
Males				
Broad cases	524	298	506	1328
Narrow cases	113	298	276	687
Controls	681		918[a]	1599
Females				
Broad cases	1192	722	715	2629
Narrow cases	356	722	426	1504
Controls	1111		718[a]	1829
Proportion of cases with the clinical features defining the Narrow phenotype				
Recurrent	39.6%	100%	73.8%	
Onset < 31	59.1%	100%	69.0%	
Recurrent + onset < 31	27.3%	100%	57.9%	
Genotyping platform				
	Perlegen	Affy 6.0	Affy 5.0 Affy 500 K[b]	
Genotyped SNPs (post-QC)				
Autosomal	427 874	646 431	254 857	
X	6438	22 546	5617	
HapMap II SNPs in final meta-analysis				
Autosomal	2 339 408			
X	51 795			
Total	2 391 203			

Abbreviations: GAIN, genetic association information network; GenRED, genetics of recurrent early-onset depression; MDD, major depressive disorder; QC, quality control; SNPs, single-nucleotide polymorphisms; STAR*D, sequenced treatment alternatives to relieve depression.
[a]The same controls were used in the GenRED and STAR*D analyses (although with separate imputation procedures using the SNPs available for cases in each data set), with statistical correction for this correlation in the meta-analyses. See online Supplementary Methods for details.
[b]Affymetrix 5.0 for 606 cases; Affymetrix 500 K for 639 cases.
Shown are the Ns for each sample in each analysis after all QC filtering. Slightly smaller samples were available for X chromosome analyses. See online Supplementary Methods for further details of QC procedures and exclusions. The GAIN–MDD sample sizes are slightly different than those in the published report,[5] because of the independent Quantile-quantile (QC) analyses, but association test results are quite similar. The HapMap II SNPs selected for association analyses had MAF > 1% and imputation $r^2 > 0.3$ in all three datasets.

1. Using data for 806 controls genotyped on *Affy 6.0 and 500 K*,[28] 61 440 SNPs were excluded for which more than 1% of samples had discordant calls (> 8 for autosomal SNPs, > 7 for chromosome X);

2. Using 12 cases genotyped with *Affy 500 K and 5.0*, 4049 SNPs had one or more discordant calls and were excluded;
3. We also examined data for 12 controls genotyped by us with *Affy 5.0 and 6.0*, but found no additional SNPs (not already excluded) with one or more discordancies.

Single-nucleotide polymorphisms were also excluded for deviation from Hardy–Weinberg equilibrium in controls at a $P < 1 \times 10^{-6}$, SNP call rate $< 98\%$ in either cases or controls, a 2% or greater difference in call rate between cases and controls, or minor allele frequency < 0.05. After all QC there were 260 474 SNPs available for analysis that captured an estimated 52.2% of common variation at an r^2 threshold of 0.8 and 66.3% at a threshold of 0.5 (that better reflects the power of a GWAS[29]). Total genotyping rates in the final post-QC data sets were 99.8 and 99.9% for autosomal and X SNPs, respectively.

GenRED and GAIN–MDD: SNP QC for the GenRED sample is described in the companion paper[4] and Supplementary Methods. We carried out new QC analyses of the GAIN–MDD data set (Supplementary Methods), to ensure consistency across the data sets and because final post-QC data were not available from dbGaP. We included 434 312 SNPs (vs 435 291 in the published GWAS report[5]).

Cluster plots of genotype intensity data were visually examined for all top results discussed below for STAR*D or the meta-analysis, including genotyped SNPs or (for the meta-analysis) those critical for the imputation of ungenotyped SNPs that produced strong signals.

Table 2 summarizes the numbers of SNPs available for each data set for meta-analysis.

Individuals

STAR*D. Cases were initially evaluated with PLINK[24] using a subset of approximately 85 000 SNPs. Pairwise estimates of identity-by-descent detected three unexpected duplicates and 21 cryptic relatives (estimated kinship $\geqslant 0.1$); for each pair the sample with the lower call rate was excluded. Four additional cases were removed for unusual degrees of SNP heterozygosity. To evaluate ancestry differences, multidimensional scaling vectors were computed and plotted, and 230 outliers to the main European-ancestry cluster were removed—most self-identified Hispanics were excluded, but 24 had scores within the main European cluster and were retained. We also removed cases with ambiguous gender ($N = 20$), or call rate $< 97\%$ ($N = 1$ for autosomal and 11 for chromosome X analyses), leaving 1221 cases for autosomal analyses and 1211 for chromosome X. QC procedures for the 1636 controls have been described in the companion paper[4] and in Supplementary Methods; briefly, samples were excluded for genotyping call rate $< 97\%$; inconsistency between

Table 3 Genomic control λ values for genotyped and imputed autosomal SNPs in the meta-analysis

	All cases (Broad)			Narrow cases		
	M + F	M	F	M + F	M	F
GAIN	1.047	1.029	1.036	1.025	1.030	1.028
GenRED	1.034	1.007	1.022	1.034	1.007	1.022
STAR*D	1.023	1.025	1.029	1.021	1.009	1.030
Combined	1.046	0.996	1.036	1.029	0.998	1.023

Abbreviations: F, females; GAIN, genetic association information network; GenRED, genetics of recurrent early-onset depression; M, males; SNPs, single-nucleotide polymorphisms; STAR*D, sequenced treatment alternatives to relieve depression.

reported and genotypic gender; outlier values for mean heterozygosity across genotypes; outliers in the distributions of principle component scores for ancestry; outliers in the number of other subjects with which kinship was estimated at $> 10\%$; and cryptic relatives (retaining the sample with the best call rate).

Meta-analysis. Quality control procedures for GenRED and GAIN–MDD (similar to methods described above for controls) are described in Supplementary Methods. For GAIN–MDD, we excluded slightly more ancestry outliers based on principal component scores. Genomic control λ values are shown in Table 3 for each analysis. Quantile-quantile plots are shown in Supplementary Tables S8–11.

Population substructure. To obtain consistent ancestry-informative covariates, we carried out a final principal components analysis[30] of all subjects, using the 82 361 autosomal SNPs common to the three data sets. Subjects who were outliers to the distributions of the two largest components were excluded (no additional STAR*D cases had to be excluded beyond those noted above), and the first 10 principal components (PC) scores were entered into the analyses as covariates to correct for population substructure.

Imputation of data for non-genotyped single-nucleotide polymorphisms

For the meta-analysis, to create genotypic data for the same SNPs for all data sets, we imputed data for each sample for HapMap II SNPs that were not genotyped in that sample, using MACH 1.0.[31] (autosomal SNPs) or IMPUTE[32] (X chromosome). For each data set, imputation was based on SNPs that passed QC for both cases and controls. MACH and IMPUTE are two of several available methods with similar accuracy.[33] Using a Hidden Markov Model algorithm with phased Centre d'Etude du Polymorphisme Humain from Utah (CEU) HapMap haplotypes as training data, a non-integer 'allele dosage' is assigned to each individual

for each SNP based on weighted probabilities of possible genotypes. For each SNP, an r^2 value estimates concordance with actual genotypes (and thus the predicted concordance with the association tests they would produce). A low r^2 predicts greater variance in the concordance of genotypes and of test statistics. This uncertainty is taken into account in the meta-analysis procedure. SNPs have been excluded from analysis if minor allele frequency was <1% in any data set or if imputation r^2 was <0.3. This threshold was used in four previous large meta-analyses because it removed most poorly imputed SNPs but few well-imputed SNPs.[34–37] The meta-analysis included 2 391 203 SNPs (2 339 408 autosomal and 51 795 X chromosome SNPs).

Statistical analyses

Analysis of genetic association. For each data set, separate association analyses were carried out for Broad and Narrow phenotypes (all GenRED cases were Narrow) for all subjects and then for males and for females separately. The *a priori* primary analyses (for STAR*D and for the meta-analysis) considered the Broad phenotype for all subjects. For STAR*D, the primary analysis was limited to genotyped SNPs; for the meta-analysis it included genotyped plus imputed SNPs.

For each analysis, single-SNP tests were carried out for each data set by logistic regression for genotyped and imputed SNPs. For discrete genotypes without covariates, logistic regression is asymptotically equivalent to a trend test for additive effects, while permitting covariates. We used custom software to implement the same logistic regression approach for imputed non-integer genotype 'dosages'. Covariates included the first 10 ancestry-informative PCs, plus an indicator for sex for X chromosome SNPs. Combined analysis ('mega-analysis') of genotypes was not straightforward because of the overlapping STAR*D/ GenRED controls, with different numbers of genotyped SNPs for the two case groups. We could have assigned unique subsets of controls to GenRED and STAR*D, but some power is lost when imputation information content is much lower in one sample (see Supplementary Methods). Therefore, we used a meta-analysis procedure as described in Supplementary Methods. Briefly, for each SNP, the procedure weights the *Z*-score for each data set by the case and control sample sizes and imputation r^2 values ($r^2 = 1$ for genotyped SNPs), while correcting for the shared controls between STAR*D and GenRED. Combined odds ratios were obtained with a similar procedure. This method takes into account the direction of association in the data sets (that is, which allele is associated), assuming that the same allele should be associated in samples with closely related ancestries. This increases power compared with the classical procedure, which ignores direction. For the primary analysis, $P < 5 \times 10^{-8}$ was considered the 5% genome-wide significance threshold.[38–40]

We also examined STAR*D and meta-analysis results for SNPs within 50 kb of 41 earlier noted MDD candidate genes. For the meta-analysis, we used a permutation-based procedure to determine whether the distribution of *P*-values observed for these SNPs deviated from chance expectation (see Supplementary Methods for details).

Power analyses. Power analysis methods are described on page S-19 and results shown in Supplementary Tables S3 and S4 and Figure S13. Power was computed for a genome-wide significance threshold of $P < 5 \times 10^{-8}$ and additive inheritance. For the primary STAR*D analysis, there was 80% power to detect an allele with a genotypic relative risk of 1.70, 1.50 and 1.43 for allele frequencies of 0.1, 0.2 and 0.3, respectively; and for the primary meta-analysis, power was approximately 50% for an allele with genotypic relative risks of 1.19 or 1.16 for allele frequencies of 20 or 50%, and was approximately 80% with genotypic relative risk of 1.20 and frequency of 30%.

Data sharing

Genotypic and clinical data are available to qualified scientists through controlled-access repository programs: the NIMH repository program (http://nimh genetics.org) for the GenRED and STAR*D case samples; dbGaP for the MGS control sample and the GAIN–MDD sample.

Results

STAR*D

The distribution of *P*-values is similar to chance expectation (Figure 1), with a genomic control λ value of 1.022. Figure 1 also summarizes association findings by chromosomal location. The top 25 findings are listed in Table 4, and all results with $P < 0.001$ in any analysis are provided online in stard_supplementary_data.txt. There were no genome-wide significant findings. Our top finding (rs12462886, $P = 1.73 \times 10^{-6}$) is located in a gene desert in 19q12. Brain-expressed genes tagged by the top 100 SNPs include: *LPHN2, SRD5A2, DYSF, RPRM, CCDC109B, CTNND2, MSR1, SLC18A1, ANKRD46, CSMD3, SLC5A12, MARK2, RCOR2, KCTD14, SYN3, NLGN4X* and *FGF13*. None of the genes had strong signals in more than one linkage disequilibrium block, but in several instances there were clusters of SNPs with strong signals within an linkage disequilibrium block, which is evidence against genotyping error. For sex-specific analyses, signals (among the top 100 for either sex) in genes of known neurobiological function or expressed in brain include: in males, SNPs in *CTNND2, GRIA1, SLC18A1, PLEKHA7, ERBB2IP, KIFAP3, CLTCL1, THRB* and *SYN3*; and in females, SNPs in *CSMD3, CACNA2D4, SV2B* and *NRXN3*.

Results for SNPs in 41 previous MDD candidate genes are shown in Supplementary Table S7. The best finding was for rs3788477, a SNP intronic to *SYN3*
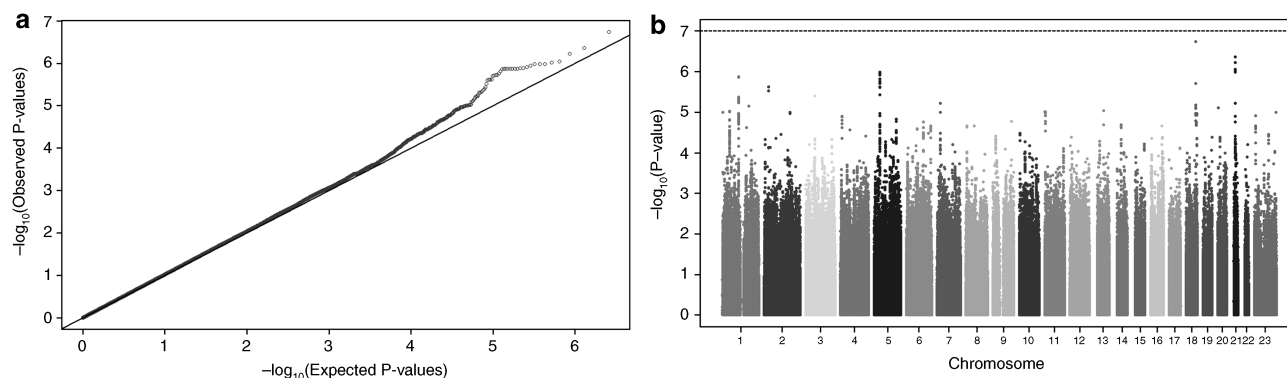
Figure 1 Overview of STAR*D GWAS results for 260 474 single-nucleotide polymorphisms (SNPs). (**a**) Quantile-quantile plot of observed vs expected $-\log$ (P-value). λ, the genomic inflation factor, is estimated at 1.022. (**b**) Manhattan plot of all results by chromosomal location.

Table 4 STAR*D GWAS results

Band	SNP	bp	A1	A2	frq	OR	P	Annotation
19q12	rs12462886	33955530	G	T	0.40	0.76	1.73E-06	
11p14.2	rs10835065	26729194	T	C	0.27	0.76	1.87E-05	SLC5A12 ($-27\,645$)
8q22.2	rs2844043	101557997	C	A	0.46	1.27	1.96E-05	ANKRD46 (33 984)
8q22.3	rs1786330	101761306	T	C	0.32	0.78	2.96E-05	PABPC1 (23 014)
2p25.1	rs7566637	10340390	T	C	0.16	1.35	3.20E-05	HPCAL1 ($-20\,101$)
18q22.1	rs627419	64466738	T	C	0.15	1.35	3.85E-05	TMX3 (25 169)
2p23.1	rs13027103	31745075	A	G	0.11	0.68	3.98E-05	SRD5A2 ($-85\,531$)
1q32.1	rs493474	197627600	T	C	0.34	1.26	4.13E-05	AK125573 (intron)
1q41	rs12125058	219344131	C	T	0.43	0.80	4.65E-05	HLX (219 108)
21q21.3	rs2831649	28504218	A	G	0.26	0.77	5.01E-05	C21orf94 (187 075)
7p21.3	rs11764174	12563141	T	C	0.40	0.80	6.05E-05	SCIN ($-13\,729$)
2q35	rs934036	218577633	A	G	0.26	0.77	6.23E-05	RUFY4 ($-30\,323$)
2q23.3	rs1221754	154081012	C	T	0.27	0.77	6.84E-05	RPRM ($-37\,444$)
11p14.3	rs4550218	22820234	G	C	0.35	1.26	7.10E-05	SVIP ($-12\,276$)
11p15.4	rs7942744	6700466	C	T	0.45	0.80	7.14E-05	GVIN1 (-3309)
3q26.1	rs1517057	167926707	T	C	0.38	1.25	8.10E-05	
11q13.1	rs11231662	63495218	G	T	0.38	1.25	8.43E-05	COX8A (-3437)
Xp22.32	rs5916245	5671536	C	T	0.35	0.76	1.04E-04	NLGN4X (146 549)
5p15.2	rs27520	11271488	C	T	0.47	1.24	1.05E-04	CTNND2 (intron)
8q12.1	rs7013994	58388614	C	T	0.18	1.30	1.10E-04	C8orf71 (28 772)
7p21.2	rs2116624	13371189	A	G	0.31	0.79	1.18E-04	
11q14.1	rs7127866	82487248	T	A	0.25	0.78	1.22E-04	RAB30 ($-26\,716$)
18q21.33	rs8099455	57412153	A	G	0.18	1.31	1.26E-04	CDH20 (38 808)
6p22.3	rs10946320	19569533	C	T	0.36	0.80	1.32E-04	
8q23.3	rs7012271	114386848	C	T	0.29	1.26	1.33E-04	CSMD3 (intron)

Abbreviations: A1, minor allele & tested allele; bp, base pair position; frq, control minor allele frequency; GWAS, genome-wide association study; OR, odds ratio; *P, P*-value; SNP, single nucleotide polymorphism; STAR*D, Sequenced Treatment Alternatives to Relieve Depression. Distance is in base pairs.
Shown are the best 25 results of the STAR*D GWAS, ranked in order of *P*-value. For each region, the SNP with the lowest *P*-value is shown.

($P = 1.64 \times 10^{-4}$). No other SNP in this analysis achieved $P < 10^{-3}$.

Meta-analysis. No genome-wide significant result was observed. Figure 2 illustrates results for all genotyped and imputed SNPs. Table 5 (Broad) and Table 6 (Narrow) summarize results for all regions with at least one SNP with $P < 10^{-5}$. Results for SNPs

with $P < 10^{-3}$ in any analysis are provided in online files meta-analysis_broad_supplementary_data.txt and meta-analysis_narrow_supplementary_data.txt. The Annotation columns of Tables 5 and 6 provide information regarding the closest gene (within 250 kb) or other functional elements annotated in the UCSC browser (full gene names and summaries of known functions are provided in Supplementary Results).
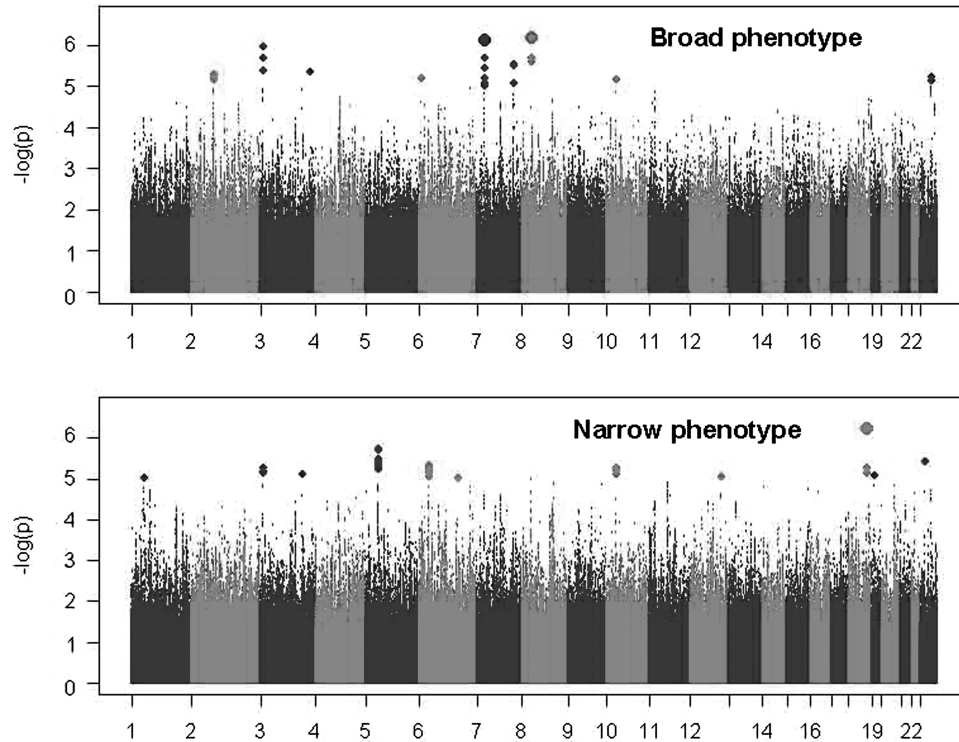
Figure 2 Meta-analysis results. Shown are association test results ($-\log_{10}$(P-values) on the Y axis) for the meta-analyses of the GenRED, STAR*D and GAIN–MDD data sets, for the Broad phenotype (primary analysis) and the Narrow phenotype (recurrent early-onset cases). The X axis shows the start position of each chromosome. Plots for males and females separately are available in online Supplementary Figures S15 and S16.

For all regions with no genes or elements listed, peaks of high homology with known regulatory sequences were detected by the evolutionary and sequence pattern extraction through reduced representations method for estimating regulatory potential.[41]

There are annotated reports of copy number variants in some of these regions, but none were detected in a survey of HapMap data,[42] and Birdsuite[42] (Birdseye module) copy number variant analysis of the GenRED data set showed that no SNP listed in Tables 5 and 6 was spanned by a copy number variant in more than a few subjects.

Figure 3 illustrates annotation information and P-values for all SNPs in the three best-supported gene-containing regions (8p21.2/*ATP6V1B2*, 3p26.1/*GRM7* and 7p15.3/*SP4*).

Results of the analyses of SNPs in or near 41 MDD candidate genes are summarized in Supplementary Table S8 and online file candidate_gene_results.xls. The aggregate analysis did not support the hypothesis of an excess of low P-values among these SNPs.

Discussion

The GWAS of STAR*D for the MDD phenotype (1221 cases and 1636 controls) did not produce genome-wide significant findings. Several regions with modest levels of significance in STAR*D were more

strongly supported in the meta-analysis, including *SLC18A1*, *ATP6V1B2* and *PLEKHA7* for the Broad phenotype and *SYN3* for the Narrow phenotype. As genotypes were assayed on three different platforms, stringent QC measures were required to avoid spurious findings. The very low genomic control inflation factor (λ) suggests that these measures succeeded, but they also reduced the number of SNPs (260 474) available for analysis.

In the meta-analysis of 3957 cases (2191 with a Narrow phenotype) and 3428 controls, genome-wide significant evidence for association to MDD was not observed for 2 391 203 genotyped or imputed HapMap II SNPs, suggesting that if any common SNPs are associated with MDD, their individual genotypic relative risks are likely to be small. Such associations could be detected in future, larger GWAS meta-analyses, a strategy that has succeeded for many other common diseases.[43,10] In samples of one or a few thousand cases, many such loci will produce unimpressive results, but the regions with the strongest evidence for association are statistically most likely to be true associations. We discuss here the three genes in which P-values of approximately $P < 10^{-6}$ were observed in the primary meta-analysis: *ATP6V1B2*, *SP4* and *GRM7*.

ATP6V1B2 encodes a subunit for a vacuolar proton pump ATPase. H^+-ATPases consist of three A, three B and two G domains. In a bipolar disorder GWAS,[28] a

Table 5 Strongest meta-analysis findings for Broad phenotype (all, male or female subjects)

Band	SNP	bp	A1/A2	All Frq	R² GAIN	R² GR	R² SD	GAIN-MDD OR	GAIN-MDD P	GenRED OR	GenRED P	STAR*D OR	STAR*D P	Meta-analysis OR	Meta-analysis P	Annotation
All-Broad																
8p21.3	rs1106634	20110329	A/G	0.12	0.703	0.996	1.000	1.270	3.93E-03	1.270	5.39E-03	1.346	8.18E-05	1.295	**6.78E-07**	ATP6V1B2; SLC18A1 (up); LZTS1 (dwn)
7p15.3	rs17144465	21470952	A/G	0.04	1.000	0.997	0.457	1.440	4.82E-03	1.820	5.97E-06	1.325	1.47E-01	1.561	**7.68E-07**	SP4
3p26.1	rs9870680	7504555	C/T	0.43	0.847	1.000	1.000	1.230	1.23E-04	1.220	4.76E-04	1.104	6.89E-02	1.188	**1.11E-06**	GRM7
7q32.3	rs10265216	130550661	A/T	0.29	1.000	0.993	0.798	1.230	7.64E-05	1.140	3.12E-02	1.161	2.32E-02	1.190	**3.02E-06**	mRNA AK294384 (amygdala)
3q26.32	rs644695	178774391	A/G	0.87	0.739	0.665	0.557	1.610	3.06E-07	1.260	3.20E-02	0.998	9.85E-01	1.354	**4.46E-06**	
2p14	rs724568	67795984	A/C	0.36	0.997	1.000	0.992	1.200	2.58E-04	1.090	1.51E-01	1.191	1.68E-03	1.171	**5.32E-06**	mRNA BC043421 (hypothalamus) (dwn)
Xq21.33	rs5990417	94966526	T/C	0.82	0.940	0.870	0.550	1.200	2.90E-03	1.250	3.70E-03	1.266	5.27E-03	1.230	**6.11E-06**	
6p25.1	rs2326810	6557466	C/G	0.92	0.930	0.863	0.754	1.210	7.29E-02	1.480	1.84E-03	1.751	1.07E-05	1.411	**6.52E-06**	LY86,within
10p11.23	rs1612122	29331904	A/T	0.48	0.959	0.953	0.923	1.150	4.51E-03	1.200	2.38E-03	1.177	4.34E-03	1.169	**6.98E-06**	
Males-Broad																
11p15.1	rs435206	16859754	A/C	0.13	0.994	0.998	0.987	1.460	1.68E-03	1.580	6.70E-04	1.353	6.72E-03	1.444	**1.23E-06**	PLEKHA7,within
6p23	rs12526133	14390002	C/T	0.54	0.954	0.971	0.803	1.330	8.80E-04	1.225	4.07E-02	1.355	7.75E-04	1.311	**1.88E-06**	
10q11.21	rs10900126	44706328	G/T	0.54	0.999	0.977	0.619	1.200	2.75E-02	1.370	1.80E-03	1.522	5.23E-05	1.321	**2.20E-06**	KSP37 (up)
2q24.1	rs10187367	155029880	A/G	0.03	0.990	0.993	0.804	1.260	1.93E-02	2.260	2.92E-04	1.884	3.68E-03	1.635	**5.42E-06**	GALNT13 (dwn)
11q25	rs3329640	133326341	A/G	0.41	0.826	0.847	0.716	1.310	3.38E-03	1.240	4.18E-02	1.380	7.37E-04	1.314	**6.49E-06**	IGSF9B
12q23.3	rs7978310	103272750	A/G	0.05	0.832	0.869	0.776	1.360	8.53E-02	1.870	2.05E-03	2.057	8.66E-05	1.704	**6.84E-06**	TXNRD1 (dwn); EID3 (dwn)
4q22.3	rs7654559	97529818	C/T	0.25	0.969	0.966	0.959	1.260	1.39E-02	1.340	7.27E-03	1.349	8.40E-04	1.309	**7.55E-06**	
9q22.32	rs2147256	97966948	C/T	0.37	0.993	0.792	0.786	1.247	8.53E-03	1.298	1.91E-02	1.367	7.86E-04	1.296	**9.10E-06**	
Females-Broad																
3q26.32	rs644695	178774391	A/G	0.86	0.739	0.665	0.557	1.810	3.23E-07	1.390	1.80E-02	1.374	4.30E-02	1.612	**3.18E-08**	
7p15.3	rs17144465	21470952	A/G	0.03	1.000	0.997	0.457	1.540	7.17E-03	2.110	4.44E-05	1.457	1.69E-01	1.731	**4.54E-06**	SP4
20p13	rs867722	4000298	A/G	0.03	0.827	0.809	0.648	1.970	4.92E-05	1.760	1.14E-02	1.324	3.42E-01	1.821	**4.89E-06**	
7p21.2	rs11772451	15073019	C/G	0.21	0.911	0.972	0.862	1.260	1.93E-03	1.360	9.14E-04	1.259	1.72E-02	1.284	**5.70E-06**	ITPK1 (up); MOAP1 (dwn); C14orf109 (up)
14q32.12	rs7151193	92681288	G/T	0.67	0.960	1.000	1.000	1.200	5.16E-03	1.280	2.46E-03	1.295	9.91E-04	1.240	**5.85E-06**	
10p12.32	rs11011581	20140825	C/T	0.04	0.966	1.000	0.950	1.480	1.86E-03	1.660	1.27E-02	1.518	1.47E-02	1.530	**6.19E-06**	PLXDC2 (dwn)
16q21	rs4620978	60750205	A/C	0.51	0.951	0.891	0.785	1.130	4.61E-02	1.340	2.72E-04	1.474	9.19E-06	1.236	**6.44E-06**	
2q22.1	rs6711718	137123482	C/T	0.39	1.000	0.997	0.749	1.210	1.27E-03	1.200	7.23E-03	1.276	6.05E-03	1.218	**7.80E-06**	

Abbreviations: GAIN, Genetic Association Information Network; GenRED, Genetics of Recurrent Early-onset Depression; MDD, major depressive disorder; mRNA, messenger RNA; SNP, single-nucleotide polymorphism; STAR*D, Sequenced Treatment Alternatives to Relieve Depression.

For each region with at least one $P < 10^{-5}$, the SNP with the lowest P-value is shown. Note that only the best SNP in each region is shown; data for all SNPs with $P < 0.001$ for each meta-analysis (Broad and Narrow; all, males, females) are provided in online files (see text).

For findings listed for males or females, P-values at the same SNP in the other gender were > 0.05. Genes are listed if $P < 10^{-5}$ was observed within the gene, 50-kb upstream (up) or downstream (dwn); or otherwise within 200-kb upstream as noted. All listed genic SNPs are intronic except for PLEKHA7 SNPs at intron–exon boundaries. Non-genic regions all contained peaks of bioinformatically predicted high homology to known regulatory sequences.[41]

A1/A2 = Alleles 1 and 2. Bold font indicates the allele associated with increased risk of MDD. GAIN, GAIN–MDD; GR, GenRED; SD, STAR*D. OR, odds ratio for the risk allele (OR >1) as indicated under A1A/2. FRQ,HapMap CEU frequency of the tested allele (this was always very similar to the frequency in GenRED/STAR*D controls, and similar to GAIN–MDD control frequencies). (Case and control allele frequencies for each sample are shown in online Supplementary Table S12).

R^2 is the value predicted (by MACH 1.0) for the squared correlation between imputed and actual genotypes ($R^2 = 1$ for genotyped SNPs). Lower values predict greater variance between imputed and actual P-values, and thus lower confidence in the P-value.

Table 6 Strongest meta-analysis findings for Narrow phenotype (all, male or female subjects)

Band	SNP	bp	A1/A2	All Frq	R² GAIN	R² GR	R² SD	GAIN-MDD OR	GAIN-MDD P	GenRED OR	GenRED P	STAR*D OR	STAR*D P	Meta-analysis OR	Meta-analysis P	Annotation
									All-narrow							
18q22.1	rs17077540	63436259	A/**G**	0.110	0.331	0.859	0.649	1.250	2.93E-01	1.610	1.83E-07	1.302	3.94E-02	1.481	**6.04E-07**	mRNA BC053410 (LOC643542);DSEL (up)
5p13.2	rs270592	38102343	A/**G**	0.570	0.996	0.989	0.991	1.170	3.62E-02	1.320	2.49E-06	1.150	3.56E-02	1.223	**1.88E-06**	GDNF (220 kb up)
Xp21.1	rs2405829	32364267	A/**G**	0.540	0.960	0.960	0.960	1.280	5.50E-04	1.130	1.87E-02	1.190	2.10E-03	1.180	**3.84E-06**	DMD
6p22.1	rs6930508	27161121	C/**T**	0.560	0.988	0.999	0.999	1.240	5.41E-03	1.180	4.19E-03	1.231	1.55E-03	1.213	**4.73E-06**	histone 2/4 gene cluster; GUSBL1 (up, noncoding)
10p11.23	rs1612122	29331904	A/**T**	0.480	0.959	0.953	0.923	1.280	1.48E-03	1.200	2.38E-03	1.176	1.78E-02	1.218	**5.39E-06**	
3p26.1	rs9870680	7504555	C/**T**	0.433	0.847	1.000	1.000	1.287	1.97E-02	1.220	4.76E-04	1.134	5.13E-02	1.211	**5.56E-06**	GRM7
3q25.1	rs1456139	152123994	A/**G**	0.610	0.999	0.845	0.839	1.180	1.99E-02	1.240	9.11E-04	1.242	3.40E-03	1.219	**7.92E-06**	CLRN1 (dwn)
19p13.11	rs7249956	18576203	A/**C**	0.770	0.893	0.898	0.732	1.270	1.41E-02	1.260	1.69E-03	1.316	2.77E-03	1.277	**8.11E-06**	CRLF1: up: C19orf50,UBA52,C19orf60; dwn:TMEM59L,KLHL26
12q23.3	rs1895943	106764962	A/**C**	0.040	0.989	1.000	0.803	1.350	6.85E-02	1.600	1.67E-04	1.602	4.03E-03	1.519	**8.70E-06**	CpG island at 106.76 Mb; PRDM4 (85 kb up)
6q22.1	rs1855625	114535864	G/**T**	0.684	0.998	1.000	0.996	1.110	2.00E-01	1.283	9.00E-05	1.279	7.64E-04	1.226	**9.34E-06**	HS3ST5, −45130
1p32.2	rs6694643	57121995	A/**T**	0.120	0.990	1.000	0.885	1.510	7.88E-05	1.210	1.16E-02	1.189	8.94E-02	1.294	**9.50E-06**	C8A, within; C8B,45475
									Males-narrow							
3p14.1	rs11710109	66836364	C/**T**	0.380	0.990	0.991	0.843	1.580	2.46E-03	1.500	6.06E-05	1.397	2.70E-03	1.483	**6.54E-08**	
11p15.1	rs389967	16831186	**A**/G	0.100	0.947	1.000	0.975	1.470	8.23E-02	1.710	2.33E-04	1.761	1.64E-04	1.675	**5.75E-07**	PLEKHA7, within
4q23	rs10007831	99911315	C/**T**	0.520	0.977	0.984	0.684	1.600	1.75E-03	1.340	3.39E-03	1.493	1.13E-03	1.441	**9.64E-07**	
10q11.21	rs7079573	44720412	A/**G**	0.540	0.920	0.882	0.624	1.430	2.46E-03	1.380	2.42E-03	1.473	2.58E-03	1.418	**9.53E-06**	KSP37, −6357 check dir
									Females-narrow							
5p13.2	rs270584	38110611	A/**C**	0.560	0.999	1.000	1.000	1.200	3.47E-02	1.390	1.64E-05	1.421	1.53E-04	1.321	**3.71E-07**	GDNF (230 kb up)
2q22.1	rs6711718	137123482	C/**T**	0.390	1.000	0.997	0.749	1.400	1.14E-04	1.200	7.23E-03	1.295	1.33E-02	1.293	**1.71E-06**	
6q25.2	rs644377	153260377	A/**T**	0.420	1.000	1.000	1.000	1.280	5.08E-03	1.330	1.44E-04	1.256	1.13E-02	1.294	**1.91E-06**	
4q28.1	rs2203374	128653784	**C**/G	0.790	0.801	0.814	0.770	1.390	1.36E-02	1.560	5.91E-05	1.400	1.20E-02	1.461	**2.71E-06**	INTU (120 kb up)
12q14.1	rs7977950	60513443	C/**T**	0.120	0.994	1.000	0.991	1.230	9.52E-02	1.400	1.30E-03	1.824	1.01E-06	1.421	**3.40E-06**	FAM19A2, within
11q13.4	rs17133921	74712634	A/**G**	0.060	1.000	1.000	1.000	1.550	3.77E-03	1.550	3.47E-03	1.643	1.96E-03	1.571	**5.29E-06**	ARRB1, within
15q25.3	rs11634319	84224912	C/**T**	0.760	1.000	1.000	0.829	1.450	1.94E-03	1.340	1.98E-03	1.353	1.40E-02	1.385	**5.67E-06**	KLHL25 (86 kb up)
4q34.3	rs346101	179841206	A/**G**	0.510	0.991	0.989	0.886	1.330	1.56E-03	1.160	5.59E-02	1.508	1.70E-05	1.289	**6.01E-06**	

Abbreviations: GAIN, Genetic Association Information Network; GenRED, Genetics of Recurrent Early-onset Depression; MDD, major depressive disorder; mRNA, messenger RNA; SNP, single-nucleotide polymorphism; STAR*D, Sequenced Treatment Alternatives to Relieve Depression.

See Table 5 legend. The following regions also achieved $P<10^{-5}$ in the Broad analysis (Table 5): 3p26.1 (GRM7, All), 10p11.23 (all), 11p15.1 (PLEKHA7, males), 10q11.21 (all), 11q13.4 (PLEKHA7, males) and 2q22.1 (females).
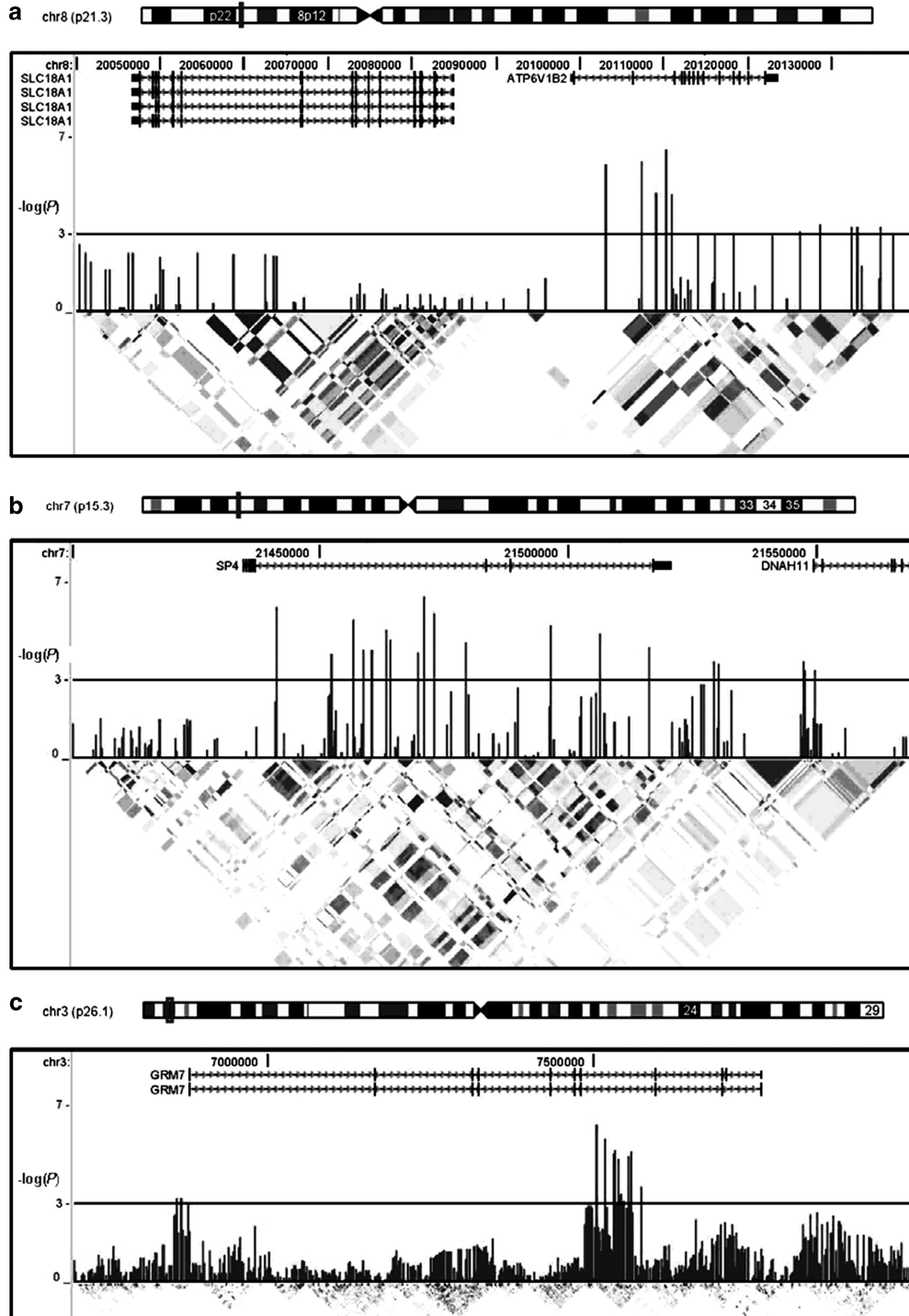
Figure 3 Best-supported regions in the meta-analysis. Shown are plots of association test results (males + females unless noted otherwise) for the three gene-containing regions with the lowest *P*-values in the primary (Broad) meta-analysis (see Table 5): *ATP6V1B2* (Panel a), *SP4* (b), *GRM7* (c). Shown in each panel from top to bottom are: an ideogram of the chromosome with the plotted area marked in red; locations in base pairs; RefSeq genes with arrows representing direction of transcription; association test results as the −log10 of the *P*-value for each genotyped and imputed single-nucleotide polymorphism; and color-coded marker–marker linkage disequilibrium results for phased HapMap II CEU genotypes (UCSC browser). Similar plots for additional top findings are available as online Supplementary Figures.

P-value of 3.32×10^{-5} was observed in *ATP6V1G1*, encoding the G subunit of the same cytosolic V1 domain to which *ATP6V1B2* contributes and which forms a complex with the transmembrane V0 domain for organelle acidification, critical to some forms of receptor-mediated endocytosis and generation of

212

proton gradients across synaptic vesicle membranes. Modest association to bipolar disorder was also reported in an adjacent gene, *SLC18A1* (previously *VMAT1*), which transports monoamines into synaptic vesicles.[44] Our signal lies in a distinct linkage disequilibrium block within *ATP6V1B2*, but *SLC18A1* could conceivably have regulatory sequences in this upstream region.

SP4 encodes the brain-specific Sp4 zinc-finger transcription factor.[45] In several small samples, modest association to bipolar disorder was observed for SNPs in an Sp4 binding site in the promoter of *ADRBK2* (beta adrenergic receptor kinase 2; earlier G-protein receptor kinase 3)[46] as well as in *SP4* itself.[47] *SP4* mutant mice showed decreased granule cell density in the hippocampal dentate gyrus,[48] deficits in sensorimotor gating and contextual learning,[49] and infertility in surviving male knockout mice despite histologically intact testes and mature sperm, suggesting a possible behavioral deficit.[50] In our data, association is observed primarily in females; it may be noteworthy that Sp4 forms gene-regulating complexes with estrogen receptors.[51] Sp4 may also have a role in glutamate-induced neurotoxicity.[52,53]

GRM7 encodes metabotropic glutamate receptor 7, which may be involved in mood regulation[54,55] Chronic treatment with mood stabilizers (lithium or valproate) decreased a hippocampal micro-RNA, increasing *GRM7* expression.[56] An metabotropic glutamate receptor 7 agonist (AMN082) had antidepressant-like effects in mice that were blocked by knockout of *GRM7*,[57] and chronic antidepressant treatment with citalopram in rodents decreased metabotropic glutamate receptor 7 immunoreactivity in hippocampus and frontal cortex.[58] This is the third GWAS to report evidence of association to mood disorders in this long gene (880 kb). Our lowest *P*-value (7.11×10^{-7}) was at 7.5 Mb (3p26.1), with *P*-values less than 10^{-4} extending to 7.56 Mb. In the German/Swiss recurrent MDD GWAS,[16] the lowest *P*-value (0.0001) was at 7.68 Mb, with *P*-values around 0.01 overlapping our signals. In the Wellcome Trust Case-Control Consortium bipolar disorder GWAS,[59] the best *P*-value in *GRM7* (0.0001 in a genotypic analyses) was at 7.63 Mb. Larger samples will be required to determine the significance of these findings, but the biological evidence suggests that *GRM7* merits further investigation.

The most strongly associated non-genic regions contain multiple peaks of high regulatory potential, but no known regulatory elements. Strong associations in non-genic regions should not be ignored; for example, several cancers are strongly associated with non-genic SNPs on chromosome 8q24,[60] whose functional relevance is now under intensive study. In our secondary analyses, very low *P*-values were observed in non-genic regions (3q26.32 in females, Broad phenotype, $P = 3.85 \times 10^{-8}$; 3p14.1 in males, Narrow phenotype, $P = 3.81 \times 10^{-8}$). These values are not significant after accounting for multiple testing, and

on 3q26.32 there is no support from other SNPs in the region (Figure S16).

For the Narrow (recurrent early-onset) phenotype, the strongest signal was in chromosome 18q22.1. The SNP with the lowest *P*-value had low imputation r^2 values, but two other nearby SNPs had *P*-values less than 10^{-5}. This region has previously been of interest in linkage studies of both bipolar disorder and MDD (see discussion in the companion paper[4]), and given that support for this region varied widely across our three samples, one might wonder whether they differed with respect to bipolar features, but we lacked the relevant data to compare the data sets. GenRED provided the strongest support as well as had the most specific procedures to exclude bipolar disorder in probands and relatives, although the severe, recurrent, early-onset phenotype more closely resembles bipolar disorder. The next strongest signals were in a non-genic region of 5p13.2, 220-kb upstream of GDNF (glial cell-derived neurotrophic factor); and in a cluster of histone genes on 6p22.1, in the same region in which significant association to schizophrenia was recently observed.[12–14] The latter finding was detected in a meta-analysis that included MGS, using a superset of the GenRED/STAR*D controls. However, MGS contributed very little of the statistical support for 6p22.1 association to schizophrenia.

Our meta-analysis findings were generally not more strongly supported by the Narrow analysis, but that sample was also smaller (55% of cases). Narrow cases provided most of the support for such signals in the Broad analysis as *ATP6V1B2*, *GRM7*, *SP4*, *PLEKHA7*, *ITPK1/C14orf109* and regions 10p11.23, 10q11.21, 6p23 and 2q22.1 (Tables 5 and 6 and Supplementary Files). Larger samples of cases with this phenotype might prove useful.

Several candidate genes were supported primarily in one gender such as *SP4* (females) and *PLEKHA7* (males). *PLEKHA7*, which encodes a poorly understood gene (pleckstrin homology domain containing, family A member 7), is associated with systolic blood pressure.[61] Sex differences are likely to exist for genetic effects in MDD.

The strongest signal in the published GAIN–MDD GWAS was in *PCLO* ($P = 7.7 \times 10^{-7}$),[5] encoding Piccolo, a protein involved in cycling of synaptic vesicles including at monoaminergic synapses. The association was supported in only one of five follow-up data sets (that totaled 6079 cases and 5893 controls), and it (like GAIN–MDD) was population-based, suggesting possible phenotypic heterogeneity. *P*-values in PCLO were less significant in our meta-analyses ($\sim 10^{-5}$) than in GAIN–MDD alone. Recurrent early-onset cases provided most of the evidence for association in GAIN–MDD, but the lowest *P*-value in the GenRED sample was 0.017. We have no independent data to test whether association is stronger in population-based samples.

In conclusion, a meta-analysis of three GWAS data sets did not detect genome-wide significant evidence for association to MDD. Of the best-supported genes

and regions, *GRM7* has the greatest previous biological support for involvement in processes such as mediation of response to antidepressant and antimanic drugs. It is likely that much larger samples will be required to clarify the role of common SNPs in genetic susceptibility to MDD. We are participating in the efforts of the Psychiatric GWAS Consortium[10,62] to carry out meta-analyses incorporating additional samples. Given the moderate heritability and clinical heterogeneity of MDD, larger samples with careful phenotypic characterization would be useful.

Conflict of interest

The authors declare no conflict of interest.

Acknowledgments

References

1 World Health Organization. *The Global Burden of Disease: A Comprehensive Assessment of Mortality and Disability from Diseases, Injuries, and Risk Factors in 1990 and Projected to 2020; Summary.* Published by the Harvard School of Public Health on behalf of the World Health Organization and the World Bank; Distributed by Harvard University Press: Cambridge, MA, 1996 p. 43.

2 Belmaker RH, Agam G. Major depressive disorder. *N Engl J Med* 2008; **358**: 55–68.

3 Sullivan PF, Neale MC, Kendler KS. Genetic epidemiology of major depression: review and meta-analysis. *Am J Psychiatry* 2000; **157**: 1552–1562.

4 Shi J, Potash JB, Knowles JA, Weissman MM, Coryell W, Scheftner WA *et al.* Genomewide association study of recurrent early-onset major depressive disorder. *Molecular Psychiatry* (in press).

5 Sullivan PF, de Geus EJ, Willemsen G, James MR, Smit JH, Zandbelt T *et al.* Genome-wide association for major depressive disorder: a possible role for the presynaptic protein piccolo. *Mol Psychiatry* 2009; **14**: 359–375.

6 Mailman MD, Feolo M, Jin Y, Kimura M, Tryka K, Bagoutdinov R *et al.* The NCBI dbGaP database of genotypes and phenotypes. *Nat Genet* 2007; **39**: 1181–1186.

7 Fava M, Rush AJ, Trivedi MH, Nierenberg AA, Thase ME, Sackeim HA *et al.* Background and rationale for the sequenced treatment alternatives to relieve depression (STAR*D) study. *Psychiatr Clin North Am* 2003; **26**: 457–494, x.

8 Rush AJ, Fava M, Wisniewski SR, Lavori PW, Trivedi MH, Sackeim HA *et al.* Sequenced treatment alternatives to relieve depression (STAR*D): rationale and design. *Control Clin Trials* 2004; **25**: 119–142.

9 Sanders AR, Duan J, Levinson DF, Shi J, He D, Hou C *et al.* No significant association of 14 candidate genes with schizophrenia in a large European ancestry sample: implications for psychiatric genetics. *Am J Psychiatry* 2008; **165**: 497–506.

10 Psychiatric GWAS Consortium Coordinating Committee, Cichon S, Craddock N, Daly M, Faraone SV, Gejman PV *et al.* Genomewide association studies: history, rationale, and prospects for psychiatric disorders. *Am J Psychiatry* 2009; **166**: 540–556.

11 Ferreira MAR, O'Donovan MC, Meng YA, Jones IR, Ruderfer DM, Jones L *et al.* Collaborative genome-wide association analysis supports a role for ANK3 and CACNA1C in bipolar disorder. *Nat Genet* 2008; **40**: 1056–1058.

12 Purcell SM, Wray NR, Stone JL, Visscher PM, O'Donovan MC, Sullivan PF *et al.* Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* 2009; **460**: 748–752.

13 Shi J, Levinson DF, Duan J, Sanders AR, Zheng Y, Pe'er I et al. Common variants on chromosome 6p22.1 are associated with schizophrenia. Nature 2009; 460: 753–757.

14 Stefansson H, Ophoff RA, Steinberg S, Andreassen OA, Cichon S, Rujescu D et al. Common variants conferring risk of schizophrenia. Nature 2009; 460: 744–747.

15 Wang K, Zhang H, Ma D, Bucan M, Glessner JT, Abrahams BS et al. Common genetic variants on 5p14.1 associate with autism spectrum disorders. Nature 2009; 459: 528–533.

16 Muglia P, Tozzi F, Galwey NW, Francks C, Upmanyu R, Kong XQ et al. Genome-wide association study of recurrent major depressive disorder in two European case-control cohorts. Mol Psychiatry 2008 (in press).

17 Kessler RC, Andrews G, Mroczek D, Ustun TB, Wittchen H-U. The World Health Organization Composite International Diagnostic Interview Short Form (CIDI-SF). Int J Methods Psychiatr Res 1998; 7: 171–185.

18 Aalto-Setala T, Haarasilta L, Marttunen M, Tuulio-Henriksson A, Poikolainen K, Aro H et al. Major depressive episode among young adults: CIDI-SF versus SCAN consensus diagnoses. Psychol Med 2002; 32: 1309–1314.

19 Nurnberger Jr JI, Blehar MC, Kaufmann CA, York-Cooler C, Simpson SG, Harkavy-Friedman J et al. Diagnostic interview for genetic studies. Rationale, unique features, and training. NIMH Genetics Initiative. Arch Gen Psychiatry 1994; 51: 849–859.

20 Kendler KS, Gatz M, Gardner CO, Pedersen NL. Clinical indices of familial depression in the Swedish Twin Registry. Acta Psychiatr Scand 2007; 115: 214–220.

21 Levinson DF, Zubenko GS, Crowe RR, DePaulo RJ, Scheftner WS, Weissman MM et al. Genetics of recurrent early-onset depression (GenRED). Am J Med Genet B NeuropsychiatrGenet 2003; 119: 118–130.

22 Kendler KS, Gardner CO, Neale MC, Prescott CA. Genetic risk factors for major depression in men and women: similar or different heritabilities and same or partly distinct genes? PsycholMed 2001; 31: 605–616.

23 Kendler KS, Gatz M, Gardner CO, Pedersen NL. A Swedish National Twin Study of lifetime major depression. Am J Psychiatry 2006; 163: 109–114.

24 Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 2007; 81: 559–575.

25 Ge D, Zhang K, Need AC, Martin O, Fellay J, Urban TJ et al. WGAViewer: software for genomic annotation of whole genome association studies. Genome Res 2008; 18: 640–643.

26 Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 2005; 21: 263–265.

27 Rabbee N, Speed TP. A genotype calling algorithm for affymetrix SNP arrays. Bioinformatics 2006; 22: 7–12.

28 Sklar P, Smoller JW, Fan J, Ferreira MA, Perlis RH, Chambert K et al. Whole-genome association study of bipolar disorder. Mol Psychiatry 2008; 13: 558–569.

29 Jorgenson E, Witte JS. A gene-centric approach to genome-wide association studies. Nat Rev Genet 2006; 7: 885–891.

30 Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet 2006; 38: 904–909.

31 Huang L, Li Y, Singleton AB, Hardy JA, Abecasis G, Rosenberg NA et al. Genotype-imputation accuracy across worldwide human populations. Am J Hum Genet 2009; 84: 235–250.

32 Marchini J, Howie B, Myers S, McVean G, Donnelly P. A new multipoint method for genome-wide association studies by imputation of genotypes. Nat Genet 2007; 39: 906–913.

33 Nothnagel M, Ellinghaus D, Schreiber S, Krawczak M, Franke A. A comprehensive evaluation of SNP genotype imputation. Hum Genet 2009; 125: 163–171.

34 Kathiresan S, Willer CJ, Peloso GM, Demissie S, Musunuru K, Schadt EE et al. Common variants at 30 loci contribute to polygenic dyslipidemia. Nature genetics 2009; 41: 56–65.

35 Scott LJ, Mohlke KL, Bonnycastle LL, Willer CJ, Li Y, Duren WL et al. A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. Science (New York, NY) 2007; 316: 1341–1345.

36 Willer CJ, Sanna S, Jackson AU, Scuteri A, Bonnycastle LL, Clarke R et al. Newly identified loci that influence lipid concentrations and risk of coronary artery disease. Nat Genet 2008; 40: 161–169.

37 Willer CJ, Speliotes EK, Loos RJ, Li S, Lindgren CM, Heid IM et al. Six new loci associated with body mass index highlight a neuronal influence on body weight regulation. Nat Genet 2009; 41: 25–34.

38 Dudbridge F, Gusnanto A. Estimation of significance thresholds for genome-wide association scans. Genet Epidemiol 2008; 32: 227–234.

39 Hoggart CJ, Clark TG, De Iorio M, Whittaker JC, Balding DJ. Genome-wide significance for dense SNP and resequencing data. Genet Epidemiol 2008; 32: 179–185.

40 Pe'er I, Yelensky R, Altshuler D, Daly MJ. Estimation of the multiple testing burden for genome-wide association studies of nearly all common variants. Genet Epidemiol 2008; 32: 381–385.

41 Taylor J, Tyekucheva S, King DC, Hardison RC, Miller W, Chiaromonte F. ESPERR: learning strong and weak signals in genomic sequence alignments to identify functional elements. Genome Res 2006; 16: 1596–1604.

42 McCarroll SA, Kuruvilla FG, Korn JM, Cawley S, Nemesh J, Wysoker A et al. Integrated detection and population-genetic analysis of SNPs and copy number variation. Nat Genet 2008; 40: 1166–1174.

43 Manolio TA, Brooks LD, Collins FS. A HapMap harvest of insights into the genetics of common disease. J Clin Invest 2008; 118: 1590–1605.

44 Lohoff FW, Dahl JP, Ferraro TN, Arnold SE, Gallinat J, Sander T et al. Variations in the vesicular monoamine transporter 1 gene (VMAT1/SLC18A1) are associated with bipolar i disorder. Neuropsychopharmacology 2006; 31: 2739–2747.

45 Suske G. The Sp-family of transcription factors. Gene 1999; 238: 291–300.

46 Zhou X, Barrett TB, Kelsoe JR. Promoter variant in the GRK3 gene associated with bipolar disorder alters gene expression. Biol Psychiatry 2008; 64: 104–110.

47 Zhou X, Tang W, Greenwood TA, Guo S, He L, Geyer MA et al. Transcription factor SP4 is a susceptibility gene for bipolar disorder. PLoS ONE 2009; 4: e5196.

48 Zhou X, Qyang Y, Kelsoe JR, Masliah E, Geyer MA. Impaired postnatal development of hippocampal dentate gyrus in Sp4 null mutant mice. Genes Brain Behav 2007; 6: 269–276.

49 Zhou X, Long JM, Geyer MA, Masliah E, Kelsoe JR, Wynshaw-Boris A et al. Reduced expression of the Sp4 gene in mice causes deficits in sensorimotor gating and memory associated with hippocampal vacuolization. Mol Psychiatry 2004; 10: 393–406.

50 Supp DM, Witte DP, Branford WW, Smith EP, Potter SS. Sp4, a member of the sp1-family of zinc finger transcription factors, is required for normal murine growth, viability, and male fertility. Develop Biol 1996; 176: 284–299.

51 Safe S, Kim K. Non-classical genomic estrogen receptor (ER)/specificity protein and ER/activating protein-1 signaling pathways. J Mol Endocrinol 2008; 41: 263–275.

52 Mao X, Moerman-Herzog AM, Wang W, Barger SW. Differential transcriptional control of the superoxide dismutase-2 kappaB element in neurons and astrocytes. J Biol Chem 2006; 281: 35863–35872.

53 Mao X, Yang SH, Simpkins JW, Barger SW. Glutamate receptor activation evokes calpain-mediated degradation of Sp3 and Sp4, the prominent Sp-family transcription factors in neurons. J Neurochem 2007; 100: 1300–1314.

54 Pilc A, Chaki S, Nowak G, Witkin JM. Mood disorders: regulation by metabotropic glutamate receptors. Biochem Pharmacol 2008; 75: 997–1006.

55 Witkin JM, Marek GJ, Johnson BG, Schoepp DD. Metabotropic glutamate receptors in the control of mood disorders. CNSNeurol DisordDrug Targets 2007; 6: 87–100.

56 Zhou R, Yuan P, Wang Y, Hunsberger JG, Elkahloun A, Wei Y et al. Evidence for selective microRNAs and their effectors as common long-term targets for the actions of mood stabilizers. Neuropsychopharmacology 2008 08/13/online.

57 Palucha A, Klak K, Branski P, van der Putten H, Flor P, Pilc A. Activation of the mGlu7 receptor elicits antidepressant-like effects in mice. *Psychopharmacology* 2007; **194**: 555–562.

58 Wieronska JM, Klak K, Palucha A, Branski P, Pilc A. Citalopram influences mGlu7, but not mGlu4 receptors' expression in the rat brain hippocampus and cortex. *Brain Res* 2007; **1184**: 88–95.

59 Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 2007; **447**: 661–678.

60 Ghoussaini M, Song H, Koessler T, Al Olama AA, Kote-Jarai Z, Driver KE *et al*. Multiple loci with different cancer specificities within the 8q24 gene desert. *J Natl Cancer Inst* 2008; **100**: 962–966.

61 Levy D, Ehret GB, Rice K, Verwoert GC, Launer LJ, Dehghan A *et al*. Genome-wide association study of blood pressure and hypertension. *Nat Genet* 2009.

62 Psychiatric GWAS Consortium. A framework for interpreting genome-wide association studies of psychiatric disorders. *Mol Psychiatry* 2009; **14**: 10–17.

Supplementary Information accompanies the paper on the Molecular Psychiatry website (http://www.nature.com/mp)