Molecular Psychiatry (2010) 15, 589–601

www.nature.com/mp

ORIGINAL ARTICLE

Genome-wide association study of recurrent major depressive disorder in two European case–control cohorts

P Muglia[1], F Tozzi[1], NW Galwey[2], C Francks[1], R Upmanyu[2], XQ Kong[3], A Antoniades[1,11], E Domenici[4], J Perry[5], S Rothen[6,7], CL Vandeleur[7], V Mooser[8], G Waeber[9], P Vollenweider[9], M Preisig[6], S Lucae[10], B Müller-Myhsok[10], F Holsboer[10], LT Middleton[5,12] and AD Roses[3,13]

[1]*Genetics Division, Drug Discovery, GlaxoSmithKline R&D, Verona, Italy;* [2]*Genetics Division, Drug Discovery, GlaxoSmithKline R&D, Harlow, Essex, UK;* [3]*Genetics Division, Drug Discovery, GlaxoSmithKline R&D, Research Triangle Park, NC, USA;* [4]*Molecular Psychiatry, Neurosciences Centre of Excellence in Drug Discovery, GlaxoSmithKline R&D, Verona, Italy;* [5]*Genetics Division, Drug Discovery, GlaxoSmithKline R&D, Greenford, Middlesex, UK;* [6]*Department of Psychiatry, University Hospital Center and University of Lausanne, Lausanne, Switzerland;* [7]*Department of Psychiatry, University Hospital of Geneva, Geveva, Switzerland;* [8]*Genetics Division, Drug Discovery, GlaxoSmithKline R&D, Upper Merion, PA, USA;* [9]*Department of Internal Medicine, University Hospital Center and University of Lausanne, Lausanne, Switzerland and* [10]*Max-Planck Institute of Psychiatry, Munich, Germany*

Major depressive disorder (MDD) is a highly prevalent disorder with substantial heritability. Heritability has been shown to be substantial and higher in the variant of MDD characterized by recurrent episodes of depression. Genetic studies have thus far failed to identify clear and consistent evidence of genetic risk factors for MDD. We conducted a genome-wide association study (GWAS) in two independent datasets. The first GWAS was performed on 1022 recurrent MDD patients and 1000 controls genotyped on the Illumina 550 platform. The second was conducted on 492 recurrent MDD patients and 1052 controls selected from a population-based collection, genotyped on the Affymetrix 5.0 platform. Neither GWAS identified any SNP that achieved GWAS significance. We obtained imputed genotypes at the Illumina loci for the individuals genotyped on the Affymetrix platform, and performed a meta-analysis of the two GWASs for this common set of approximately half a million SNPs. The meta-analysis did not yield genome-wide significant results either. The results from our study suggest that SNPs with substantial odds ratio are unlikely to exist for MDD, at least in our datasets and among the relatively common SNPs genotyped or tagged by the half-million-loci arrays. Meta-analysis of larger datasets is warranted to identify SNPs with smaller effects or with rarer allele frequencies that contribute to the risk of MDD.
Molecular Psychiatry (2010) **15,** 589–601; doi:10.1038/mp.2008.131; published online 23 December 2008

Keywords: major depressive disorder; genetics; GWAS; psychiatry; neuroscience

Introduction

Major depressive disorder (MDD) is a common disorder with lifetime prevalence estimated to be in Caucasians in the range 6–30%.[1] A number of efficacious treatment for MDD are available, but there is an ample margin for improving efficacy and side effects profile of existing treatments.[2] The World Health Organization has estimated that in the year 2000 depression was the fourth contributor to the global burden of disease and predicted that MDD will become the second leading contributor by the year 2020.[3] There is therefore a great need and scope for improving treatment strategies for MDD, which will need to be based on a better understanding of its still elusive aetiology.[4] The identification of genetic risk factors for MDD holds the promise of improving our understanding of the neurobiological basis of the disorder, which may ultimately lead to the identification of novel treatment and prevention strategies. Genetics has a recognized role in increasing the susceptibility to MDD, with an estimated λ-sibling value ranging between three and nine[5,6] and a substantial difference in concordance rates between monozygotic and dizygotic twins.[7] MDD with recurrent episodes have been consistently shown to identify a subset of MDD with higher familiality.[8]

A number of not yet conclusive molecular genetics studies, including linkage and association approaches, have been conducted for MDD. Suggestive evidence for linkage has been reported for few

Correspondence: Dr P Muglia, Genetics Division, Drug Discovery, GlaxoSmithKline R&D, Via Fleming 4, Verona 37135, Italy.
E-mail: pierandea.2.muglia@gsk.com
[11]Current address: University of Cyprus, Cyprus.
[12]Current address: Division of Neurosciences and Mental Health, Imperial College, London, UK.
[13]Current address: Duke University, Durham, NC, USA.
Received 2 September 2008; revised 5 November 2008; accepted 11 November 2008; published online 23 December 2008

genomic regions.[9–12] Association studies of MDD, which have greater power[13] to test the common disease–common variant (CDCV) hypothesis,[14] have thus far relied only on candidate gene studies.[15] However, array genotyping, capable to query most of the human genome[16] at affordable costs, now allows testing the CDCV hypothesis at the whole-genome level with no need of *a priori* hypotheses on disease aetiology.

Here, we report the results from two genome-wide association studies (GWAS) conducted in two independent cohorts of recurrent MDD cases and controls, and the results of the meta-analysis of these two GWAS incorporating 1359 cases and 1782 controls, all of declared Caucasian origin.

Materials and methods

Subjects

The present study was performed on two datasets. Participants in these studies that were approved by the respective Ethics Board, received a detailed description of the goal and funding of the study, and signed a written informed consent.

Sample I. A total of 1022 Caucasian individuals diagnosed with recurrent MDD and 1000 Caucasian age- and gender-matched non-affected controls were recruited at the Max-Planck Institute of Psychiatry in Munich, Germany; patients were also recruited at two satellite recruiting hospitals (BKH Augsburg and Klinikum Ingolstadt) in the Munich area. All patients were evaluated using the semi-structured Schedules for Clinical Assessment in Neuropsychiatry (SCAN) instrument.[17,18] The SCAN interview was administered by experienced research assistants who had received proper training at WHO Training and Research Centres. Each participant completed a questionnaire regarding his/her demographics, family and individual history as well as medical history and ethnicity. Patients were included in the study if they received a diagnosis of recurrent MDD (that is, at least two separate episodes of depression) of moderate or severe intensity according to *Diagnostic and Statistical Manual of Mental Disorders*, fourth edition (DSM-IV)[19] or ICD-10,[20] as resulted by analysis of the SCAN2.1 interview using the computerized algorithm CATEGO.[17] Patients were excluded from the study if they had: experienced mood incongruent psychotic symptoms, a lifetime history of intravenous drug use or diagnosis of drug dependency, depression secondary to alcohol or substance abuse or depression as clear consequence of medical illnesses or use of medications. Patients with diagnosis of schizophrenia, schizoaffective disorders and other axis I disorders other than anxiety disorders were excluded from the study. Patients with co-morbid anxiety disorders with the exception of obsessive compulsive and post-traumatic stress disorders, were included considering that anxiety and MDD very often co-occur and are also likely to share some of the genetic factors.[21] Approximately 26% of cases ($n=262$) showed presence of at least one anxiety symptom during their worst episode as assessed by the SCAN interview. The anxiety symptoms considered were the presence of general rating of anxiety, general rating of phobia, free-floating anxiety or anxious foreboding with autonomic symptoms (see Tozzi *et al.*, 2008 for more details on symptom selection).[22] Within patients with these anxiety symptoms 51% had a formal diagnosis of anxiety disorders (including generalized anxiety disorder, panic disorder with or without agoraphobia, social and specific phobia) according to DSM-IV. The SCAN interview did not assess axis II disorders so we cannot exclude the presence of personality disorders among our cases. Controls were selected randomly from a Munich-based community sample and recruited at the Max-Planck Institute of Psychiatry. They were screened for the presence of anxiety and mood disorders using the Composite International Diagnostic Screener.[23] Only individuals without mood and anxiety disorders were collected as controls.

Sample II. Subjects meeting DSM-IV criteria for recurrent MDD ($n=492$) and unaffected controls ($n=1052$) were selected among the Caucasians of a community survey carried out in the city of Lausanne, Switzerland.[24] These individuals, who underwent a comprehensive psychiatric investigation were part of a large survey (CoLaus; $n=6738$) conducted to evaluate risk factors for cardiovascular diseases (CVDs); for a more detailed description of the study see Firmann *et al.*[25] The random drawing procedure was based on a complete list of the Lausanne inhabitants aged 35–75 years ($n=56\,694$ in 2003), provided by the official city population register. All 35- to 66-year old subjects of the somatic survey ($n=5543$) were invited by letters to participate also in the psychiatric evaluation. Those who did not respond to the letter were contacted by telephone. All subjects who were sufficiently fluent in French or English and agreed to participate were included into the psychiatric sub-study (PsyCoLaus) and underwent the psychiatric evaluation between the years 2004 and 2008. In total, 67% of the participants of the CoLaus study in the age range between 35 and 66 years accepted the psychiatric evaluation, which resulted in a sample of 3718 individuals, of whom 92% were Caucasians. In the CoLaus sample comparison between individuals who participated in PsyCoLaus and those who did not revealed similar scores on the General Health Questionnaire,[26] French version,[27] a screen for psychopathology completed during the medical examination, which suggests the absence of selection bias regarding psychopathological characteristics. Psychiatric assessment in the PsyCoLaus sub-study included the semi-structured Diagnostic Interview for Genetic Studies, French version.[28] The family history on all individuals was collected using the Family History-Research Diagnostic Criteria.[29] From a total of 3420 Caucasians

who received full psychiatric assessment and gave consent for genetic testing, 492 subjects with recurrent MDD (at least two distinct major depressive episodes) and 1052 controls devoid of any psychiatric disorders were selected for the purpose of the GWAS analysis reported in this paper. Individuals with comorbid DSM-IV anxiety disorders ($n=184$)—with the exception of obsessive compulsive disorder and post-traumatic stress disorders—were not excluded from the selected cases, to match inclusion criteria of Sample I.

Samples preparation, genotyping and quality control

Sample I. DNA was isolated from whole blood using a standard salting-out procedure. Samples were arrayed at a concentration of 50 ng per μl. Illumina HumanHap550 SNP chip arrays were used to genotype a total of 2068 samples of genomic DNA at Illumina laboratories (San Diego, CA, USA). Of these DNA samples, 1915 gave genotypes that were considered valid (sample success rate = 93.28%). The main reasons for failed genotyping included low concentration DNA samples and low quality DNA. The locus success rate (the percentage of loci assayed that were successfully genotyped) was 99.23% producing 551 101 loci per sample that were available for analysis. The genotype call rate (the percentage of genotypes delivered for successful samples among successful loci) was 99.82%, producing a total of 1 053 473 991 successful genotypings. After this initial quality control (QC) procedure, more detailed QC was conducted as follows.

New genotype calls were generated using a procedure similar to that of Fellay *et al.*[30] with minor modifications dictated by the characteristics of genotyping in our sample. Briefly, two channel signal intensity data, corresponding to the two alleles at each SNP, were brought into the software Beadstudio 3.1 (Illumina Inc., San Diego, CA, USA). The initial genotype calls were generated using the cluster file provided by Illumina with default GenCall cutoff of 0.15: genotypes with a GenCall score less than the threshold value were not accepted for further analysis because they were considered to be too far from their associated cluster to be reliable. We then evaluated the clusters of intensity scores for each SNP using sample call rate, SNP call frequency, cluster separation score, heterozygosity excess (= observed proportion of heterozygotes—proportion expected on the assumption of Hardy–Weinberg equilibrium) and in some cases visual checking of the clustering graph, as follows. Samples with call rate below 95% were removed. We then re-clustered SNPs from autosomal chromosomes with call frequency below 99%, and subsequently updated the success rate statistics on SNPs and samples. Selection on SNPs was then performed by putting the SNPs through filters designed to optimize genotyping call accuracy and call frequency, as follows. All SNPs with call frequency below 95% were removed. For SNPs with call frequency between 95 and 98%, we removed any SNP with cluster separation below 0.3, or heterozygosity excess <-0.1 or $>+0.1$. For SNPs with call frequency above 98%, we removed any SNP with cluster separation below 0.25, or with heterozygosity excess <-0.3 or $>+0.3$. Following these procedures, genotyping accuracy was evaluated using samples genotyped in duplicate. The concordance rate for the 100 duplicate sample pairs was about 99.9%. Of the 535 180 SNPs successfully called, 436 551 (81.6%) show 100% concordance, and 98.85% SNPs showed 98% concordance or better.

After QCs, the final analysis marker set consisted of 522 008 SNPs. SNPs with minor allele frequencies below 0.0015 were excluded from the analysis.

Gender consistency checks were performed using the function available in the software PLINK version 9.99.[31] Presence of cryptic relatedness was established by testing pair-wise identify by descent (IBD) among all subjects via PLINK on autosomal SNPs. Pairs of individuals showing relatedness closer than third degree (estimated IBD >0.125) were identified, and one member from each pair was removed from the analysis. One member of each pair of the 100 duplicated subject pairs was also removed. The number of subjects removed and the reasons for removing them from the overall QC are summarized in Supplementary Table 1. Of the 2068 subjects recruited, 1792 were used for analyses.

Modelling stratification. We tested for population stratification using the method developed by Fellay *et al.*[30] Stratification was modelled using principle components analysis as implemented in the software EIGENSTRAT version 1.01.[32] In preparation for the analysis, we removed autosomal SNPs in high linkage disequilibrium: all SNPs within a window size of 1500 SNPs having $r^2<0.2$ were eliminated. We used a test of population structure derived from Tracy–Widom test[33] to select principal components (PCs) with P-value <0.05, which account for a large proportion of the genetic variation in the sample. We then checked loadings of samples (that is, case–control individuals) on the selected PCs to remove any outlier samples with loadings outside ± 6 standard deviations, after which the test for stratification would have been re-run. However, no such outliers were found. We have identified 15 significant PCs that were used as co-variates in the association analysis to correct for the population stratification that they represented.

Sample II. Nuclear DNA was extracted from whole blood for whole-genome scan analysis and genotyping was performed using the Affymetrix 500K SNP chip, as recommended by the manufacturer. Quality control procedures for genotyping in this dataset were as reported by Sandhu *et al.*[34] In brief, genotypes were

obtained via the BRLMM algorithm. Individual genotypes were removed from the analysis based on gender inconsistency; genotypes calls less than 90%; inconsistent duplicate genotypes. The total final sample from which we selected the individual for our study comprised 5636 participants. SNP quality control checks removed SNPs with genotypes on less than 95% participants; SNPs that were out of Hardy–Weinberg equilibrium ($P<1\times10^{-6}$). After these quality control procedures, 370 697 SNPs remained for analysis for 5367 participants.

Statistical methods

Genome-wide association analysis, Sample I. To evaluate the association of case–control status with the additive genetic effect of each SNP, we performed logistic regression analysis using gender, site and the significant stratification PCs as co-variates, using the software PLINK VERSION 0.9.[31] SNPs located on the mitochondrial DNA, and on X and Y chromosomes were not analysed. As a further test for the presence of population structure, we calculated the genomic control parameter of Devlin and Roeder[35] and obtained $\lambda=1.0104$ that is very close to 1 therefore suggesting the absence of major population structure associated with case–controls status. We have therefore decided not to correct our logistic regression for our genomic control value.

Genome-wide association analysis Sample II. Association of case–control status with the additive genetic effect of each Affymetrix 5.0 SNP that passed QC was evaluated by same methods as described in Sample I, except that the study was performed at a single study site and study site was therefore not used as a co-variate. The genotypes at the Illumina HumanHap550 SNP loci in the individuals in Sample II were determined, to the extent possible, by imputation from the Affymetrix 5.0 genotypes using the program IMPUTE.[36] For the association analysis, we selected imputed genotypes that had maximum probability value (that is, >0.9 as in the default option in program IMPUTE.[36] Association of case–control status with the additive genetic effect of the imputed genotypes at each SNP locus was evaluated as described above.

Meta-analysis Sample I and Sample II GWAS. Results from the two GWAS included a total of 1418 recurrent MDD cases and 1918 controls. Imputing genotypings in Sample II and selection of Illumina loci, we obtained shared genotyped or imputed genotypings at 494 678 SNP loci genome-wide for a total of 1359 cases and 1782 controls that were available for meta-analysis. Because the results were derived from two different studies, using different genotyping platforms, it was considered that a single statistical analysis of the combined data was not appropriate. The results were therefore combined using a meta-analytical approach. We have used the meta-analysis approach as implemented in the statistical software METAL (http://www.sph.umich.edu/csg/abecasis/Metal/).

Gene-based P*-values*

To evaluate the presence of association at a gene level, we have performed a gene-based analysis. For this purpose, we have first mapped SNPs available from the two GWAS meta-analysis to genes according to dbSNP (build 125). We included SNPs within a 100 kb region up- and downstream of each gene to capture variants that are not in the coding region but may disrupt nearby regulatory regions. The evidence for association of each gene with the MDD phenotype is conveniently summarized by the lowest *P*-values observed at any SNP locus within the gene. However, reliable interpretation of such minimum *P*-values must take into account the different number of SNPs genotyped in different genes, and the pattern of LD among them. For this purpose, we used a method proposed by Galwey (in preparation), closely related to those of Nyholt[37] and Li and Ji[38] that uses the Šidák correction for multiple testing,[39] but adjusts the actual number of SNPs genotyped in each gene to obtain an effective number of SNPs, on the basis of the eigenvalues of the LD matrix among the SNPs.

Functional annotation

We carried out functional annotation analysis on the genes identified by the most significant SNPs, namely those with association $P<0.0005$ from our SNP-based meta-analysis results. The gene annotation was performed according to National Center for Biotechnology Information dbSNP through the use of WGA-Viewer software.[40] The gene list was submitted to the Ingenuity Pathway Analysis tool (Ingenuity Systems, Mountain View, CA, USA; https://analysis.ingenuity.com/pa/login/login.jsp), a system with a web-based interface providing computational algorithms to functionally analyse large datasets and identify or generate gene/protein networks that are formed by the genes of interest. The analysis is supported by the company's *Knowledge Base*, which contains interactions collated from extensive annotation of literature findings, canonical pathways and functional categories including gene ontologies. The significance value of a given canonical pathway or functional analysis category is a measure of the probability that the observed association of the phenotype with the pathway or function would occur by chance. The value of *P* is calculated using the right-tailed Fisher's exact test, and nominal values of $P<0.05$ were *a priori* assumed to be statistically significant.

We have also searched whether published genetic association studies on mood disorders and schizophrenia reported results for the list of our 104 most significant genes from our meta-analysis results. To do this, we run a batch search on the genetic association database (http://geneticassociationdb.nih.gov/), a web-based National Institutes of Health supported public repository of information from

Table 1 Demographic and main clinical features of samples I and II totaling 1418 cases and 1918 controls

	Mean age at interview (s.d.; range)	*Gender female: male ratio*	*Age of onset[a], M (s.d.; range)*	*Average no. of depression episodes per year, M (s.d; range)*	*Depression familiality[b](%)*
Cases					
Sample I ($n=926$)	50.70 (13.71; 18–87)	2:1	37.16 (13.37; 7–84)	0.73 (1.46)	51.7
Sample II ($n=492$)	50.94 (8.56; 36–67)	2.7:1	27.19 (10.44; 7–56)	0.20 (0.25)	54.4
Controls					
Sample I ($n=866$)	52.37 (13.23;19–91)	2.1:1	NA	NA	11.5
Sample II ($n=1052$)	51.85 (8.74; 36–69)	1:1.5	NA	NA	22.7

Abbreviations: NA, not applicable; M, mean.
[a]Age of onset was defined as follows: Sample I: Assessed retrospectively and defined as the age at which the first manifestation of MDD that required formal advice occurred, as reported by the participant; Sample II: age at which occurred the first full DSM episode.
[b]Family history of depression was defined as follows: Sample I: family history in first-degree relatives as reported by the patients during the SCAN interview, family history in controls was assessed using a self report questionnaire that asked whether their first-degree relative had depression that needed treatment. Sample II: Both patients and controls were assessed using the family history questionnaire (FH-RDC) that asked about the presence of one or more first-degree relatives with a DSM-IV diagnosis of depression.

published disease-based genetic association studies compiled from PubMed by the National Office of Public Health Genomics.[41]

Results

Our two GWAS analysed 926 cases and 866 controls in Sample I and 494 cases and 1052 controls in Sample II. Demographic and main clinical characteristics are shown in Table 1. Cases and controls were matched for gender and age in Sample I as per recruitment criteria. Cases and controls in Sample II that were selected from our population-based cohort, showed similar age means but differences in gender distribution reflecting the higher prevalence of MDD in women in the general population. Sample I had higher average number of major depressive episodes per year when compared with Sample II whereas Sample II showed a younger age at onset. The difference in the number of episodes per year between patients in Samples I and II is likely to reflect the different ascertainment and selection of patients for the two studies: Sample I a clinical-based study recruited more severe cases than the patients selected for Sample II that are derived from a population-based study. Age at onset and family history differences between the two samples could reflect real difference existing in the two samples for these two variables, however, it has to be considered that these data were defined and collected differently in the two studies. (see Table 1 for more details).

A total of 494 678 SNPs that passed QC procedures from SNPs genotyped in Sample I and genotyped or imputed in Sample II were considered for the meta-analysis that considered a total of 1359 recurrent MDD cases and 1782 controls. The genomic control λ calculated according to Devlin and Roder[35] showed the absence of major population structure (values close to 1) in both Samples I and II. Association analysis results in Sample I, with no use of the 15 PCs as co-variates in the logistic regression analysis produced $\lambda=1.002$; and $\lambda=1.0104$ with the use of PCs in the analysis. In Sample II, without correction for population stratification: $\lambda=0.99$; analysis using PCs as co-variates: $\lambda=0.98$. Results from the meta-analysis gave $\lambda=0.99$, suggesting the absence of major population structure.

None of the association *P*-values derived from the GWAS analysis of recurrent MDD in each sample individually, or from the meta-analysis, passed the stringent threshold for genome-wide significance at $P<5\times10^{-8}$.[42] This can be observed in the quantile–quantile plots, in Figure 1, indicating that in all three analyses the observed distribution of *P*-values lies close to that expected by chance on the null hypothesis of no association.

Furthermore, the two GWAS failed to show agreement of the most significant SNP associations identified. 23 818 SNPs showed association of *P*-values <0.05 in Sample I and 22 587 such SNPs in Sample II; among these SNPs we have observed 1180 SNPs that were shared between the two samples, which is very close to the number we would expect by chance (1201). Similar for SNP that produced association *P*-values <0.0005, we failed to detect much agreement between the two GWAS. In fact,

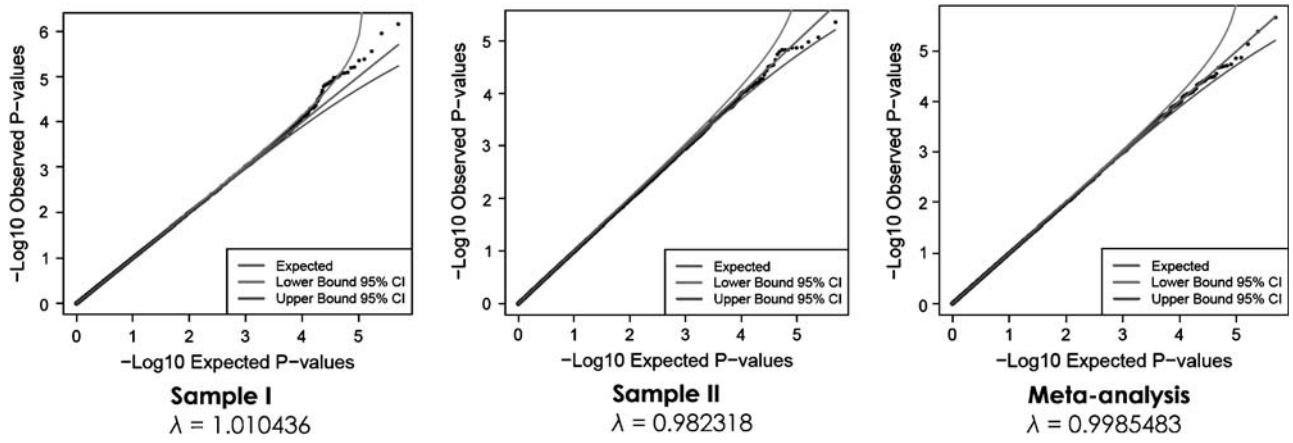

Figure 1 Quantile–quantile plots of genome-wide association study (GWAS) and GWAS meta-analysis of individual loci with major depressive disorder (MDD) case–control status. *P*-values are obtained from logistic regression using the additive genetic model. λ = Genomic control parameter of Devlin *et al*.

there were 250 SNPs with $P < 0.0005$ in Sample I and 220 in Sample II and none of these SNPs was shared between the two GWAS, which is also very close to the number of SNPs we would expect to be shared by chance (0.1).

The most significant *P*-value (0.58×10^{-5}) from the meta-analysis of the two studies was observed for an SNP (rs4238010) on 12p13 that is located over 260 kb away from the closest known gene, named cyclin D2 (*CCND2*). Forty-two SNPs showed meta-analysis association *P*-values $< 10 \times 10^{-5}$. These SNPs mapped to 14 distinct chromosomes and to 27 independent loci. A summary of the most significant SNPs for each of these independent loci is presented in Table 2.

The 42 most significant SNPs are located within 15 known genes. All SNPs in these genes are located in non-coding regions (either in intergenic or intronic regions) except for a non-synonymous SNP (rs590557) that leads to an amino-acid substitution (Leu to Ser) in the EH domain-containing 3 gene that is thought to regulate endocytic trafficking.[43]

To facilitate comparison with future studies, we have provided in the Supplementary material the 200 most significant *P*-values from the GWAS in Samples I and II and from the meta-analysis results in the supplement (Supplementary Tables 2, 3 and 4).

We have tested the presence of associations from published literature for mood disorders and schizophrenia for the 104 most significant genes that were identified in our meta-analysis by SNPs with a $P < 10^{-4}$. For none of these genes previous association has been reported with MDD. Previous association studies were reported for adenylate cyclase 9 (*ADCY9*), adenomatous polyposis coli (*APC*), complexin 2 gene (*CPLX2*), glutamate receptor, ionotropic, δ-1 gene (*GRID1*) and *GRM7* with schizophrenia and bipolar disorder (BD). *ADCY9* showed modest associations with BD,[44] but lack of association in a small Asian sample.[45] *APC*, a tumor suppressor, gene has been reported to be associated with schizophrenia in a small Chinese family-based sample.[46] The *CPLX2*, which is involved in synaptogenesis and in the modulation of neurotransmitter release, showed contrasting results in two Asian schizophrenia datasets.[47,48] The *GRID1* showed association in schizophrenia and BD in Ashkenazi Jewish cohorts[49] as well as in a sample of schizophrenia cases and controls from China.[50]

Our gene-based analysis produced a global *P*-value from each of the genes as calculated by the SNP-based meta-analysis results. As described in more details in our statistical methods the gene-based *P*-values is obtained with a correction for the number of SNPs genotyped and their LD matrix. A total of 17 643 genes were available for this analysis, and the results for the most significant genes are illustrated in Table 3.

The most significant results of the gene-based analysis are similar to those obtained from single SNP analysis. For example, the gene *SMG7*, which showed the strongest adjusted gene-based association ($P = 0.009$), was also among the top loci for a cluster of SNPs on ch1. Similarly, gene *NKKB1* gave the second lowest *P*-value both in the gene-based analysis and in the single SNP meta-analysis.

We have queried our meta-analysis results for a panel of candidate genes for which there are published reports of association with MDD as reviewed by Levinson.[15] We also queried our results for genes that showed most significant associations in the published GWAS of BD.[15,51–54] None of these candidate genes showed association in our results that would survive correction for multiple testing at the whole-genome level. It is however of interest for future studies to highlight some of our results for these genes, and these are summarized in Table 4.

The most significant association was detected for gene *GRM7* (metabotropic glutamate receptor 7 gene, mGluR7). A SNP (rs1485171) in this gene has been reported among the most significant and biologically plausible associations in the bipolar Wellcome Trust Case Control Consortium (WTCCC) results.[53] Our genotypings did not include this SNP, and the SNP

Table 2 Most significant SNPs from meta-analysis, and significance levels of the constituent single-sample results

SNP	*Chr.*	*Location (bp)*	*Distance to closest gene (bp)*[a]	*Closest gene*[a]	*SNP type*	*Meta-analysis P-value × 10*^{5}	*Allele 1/2*[b]	*Meta-analysis, OR*	*P-value*		*Allele frequency*[c]				*Rank*[d]
									Sample I	*Sample II*	*Sample I*		*Sample II*		
											Case	*Control*	*Case*	*Control*	
rs2702199	1	180 133 537	−39 754	SMG7	Intergenic	1.49	t/**c**	0.72	0.0015	0.0031	0.29	0.34	0.29	0.35	3
rs1653765	2	23 640 884	0	KLHL29	Intronic	2.38	a/**g**	0.69	0.000069	0.14	0.51	0.55	0.77	0.81	8
rs590557	2	31 373 046	0	EHD3	Non-syn	3.71	a/**g**	0.65	0.0056	0.0019	0.11	0.14	0.11	0.15	16
rs7596956	2	219 977 864	0	DNAJB2	Intronic	7.99	t/**c**	0.72	0.00087	0.029	0.81	0.85	0.78	0.82	34
rs9311395	3	4 716 821	0	ITPR1	Intronic	9.56	a/**g**	0.78	0.0015	0.022	0.45	0.46	0.46	0.51	40
rs6781822	3	11 650 632	0	VGLL4	Intronic	1.89	**t**/c	1.35	0.0018	0.0032	0.39	0.35	0.40	0.36	6
rs709496	3	109 180 117	−64 514	CD47	Intergenic	9.62	a/**g**	0.78	0.037	0.00039	0.52	0.55	0.49	0.57	41
rs6832167	4	70 655 070	0	UGT2A1	Intronic	2.67	a/**g**	0.68	0.042	0.000013	0.75	0.77	0.86	0.93	10
rs1313925	4	103 734 756	−44 917	NFKB1	Intergenic	1.12	t/**c**	0.74	0.017	0.000079	0.37	0.42	0.34	0.40	2
rs7676614	4	144 973 971	0	FREM3	Intronic	9.52	**a**/g	1.32	0.021	0.0010	0.68	0.65	0.70	0.65	39
rs6813541	4	175 147 298	−385 241	FBXO8	Intergenic	5.49	**t**/g	1.29	0.00010	0.092	0.49	0.43	0.53	0.50	27
rs12651937	5	77 958 295	0	LHFPL2	Intronic	9.78	t/**c**	0.76	0.00024	0.089	0.32	0.34	0.30	0.33	42
rs929122	6	117 712 442	−3781	ROS1	Downstream	4.92	t/**c**	0.68	0.00014	0.14	0.75	0.78	0.86	0.90	25
rs6945001	7	25 901 622	−63 478	NFE2L3	Intergenic	6.63	a/**g**	0.63	0.010	0.0017	0.90	0.92	0.89	0.93	28
rs2800092	10	36 653 560	−19 8190	Q5SYT8_HUMAN	Intergenic	7.97	a/**g**	0.78	0.00058	0.041	0.43	0.47	0.42	0.46	33
rs10833965	11	23 171 456	—	—	Intergenic	3.58	**a**/g	1.32	0.019	0.00032	0.46	0.41	0.47	0.41	15
rs12278579	11	42 694 572	—	—	Intergenic	4.86	t/**c**	0.62	0.00010	0.086	0.94	0.96	0.92	0.94	24
rs4238010	12	3 988 578	−26 4621	CCND2	Intergenic	0.58	a/**g**	0.58	0.0015	0.00098	0.87	0.89	0.96	0.98	1
rs4760933	12	69 430 025	0	PTPRR	Intronic	2.90	a/**g**	0.60	0.000014	0.53	0.79	0.84	0.94	0.94	11
rs1344677	12	103 553 434	0	CHST11	Intronic	2.58	**t**/c	1.32	0.015	0.00029	0.54	0.52	0.51	0.44	9
rs17767562	13	96 073 933	0	HS6ST3	Intronic	4.23	t/**c**	0.76	0.0094	0.0011	0.58	0.60	0.54	0.62	19
rs17105696	14	36 457 578	0	SLC25A21	Intronic	3.22	a/**g**	0.51	0.0000079	0.23	0.96	0.98	0.96	0.97	13
rs8003655	14	85 267 100	—	—	Intergenic	4.99	t/**c**	0.49	0.000044	0.14	0.96	0.98	0.97	0.98	26
rs2239307	16	3 982 450	0	ADCY9	Intronic	3.07	t/**c**	0.65	0.00034	0.032	0.80	0.82	0.98	1.00	12
rs17299191	16	54 029 584	−41 005	MMP2	Intergenic	9.44	t/**c**	0.72	0.0045	0.0073	0.24	0.26	0.19	0.23	38
rs12923267	16	64 095 518	—	—	Intergenic	4.44	t/**c**	0.74	0.0030	0.0048	0.33	0.38	0.31	0.37	21
rs8122984	20	57 786 316	0	PHACTR3	Intronic	7.45	a/**g**	0.76	0.012	0.0016	0.73	0.76	0.68	0.74	30

Abbreviation: bp, base pair; Chr, chromosome; non-syn, nonsynonymous; OR, odd ratio; SNP, single nucleotide polymorphism.
[a]A missing value indicates no known coding region within 500 kb from the SNP.
[b]The allele shown in bold confers greater odds that the carrier is a case, that is, the 'risk allele'. The use of this term does not imply that the observed association is a true-positive finding. The direction of association is consistent between samples for all SNPs presented here.
[c]Frequency of allele 1.
[d]Rank of SNP on the basis of meta-analysis *P*-value.

Table 3 Gene-based *P*-values indicating strongest associations with MDD case–control status from the two study meta-analysis results

Chromosome	*Gene*	*No. of SNPs genotyped*	*Effective no. of SNPs*[a]	*Min. (P)*	*Adjusted, corrected min. (P)*[b]
1	*SMG7*	34	12.84	0.00001	0.00019
4	*NFKB1*	45	20.10	0.00001	0.00022
17	*LOC654346*	1	1.00	0.00035	0.00035
1	*NMNAT2*	60	24.04	0.00001	0.00036
1	*LAMC2*	45	21.13	0.00002	0.00045
4	*UGT2A2*	47	17.52	0.00003	0.00047
4	*UGT2A1*	51	18.78	0.00003	0.00050
3	*ATG7*	79	32.13	0.00002	0.00060
11	*CUGBP1*	19	5.31	0.00014	0.00074
7	*NFE2L3*	24	12.07	0.00007	0.00080
3	*VGLL4*	95	47.52	0.00002	0.00090
2	*ABCB6*	28	12.03	0.00008	0.00100
2	*ZFAND2B*	28	12.03	0.00008	0.00100
2	*DNAJB2*	27	12.15	0.00008	0.00100
2	*ATG9A*	28	12.22	0.00008	0.00100
11	*RAPSN*	19	7.07	0.00014	0.00100
2	*STK16*	28	12.42	0.00008	0.00100
2	*ANKZF1*	30	12.47	0.00008	0.00100
2	*GLB1L*	30	12.47	0.00008	0.00100
2	*TUBA1*	29	13.40	0.00008	0.00110
2	*DNPEP*	28	13.74	0.00008	0.00110
2	*PTPRN*	30	13.74	0.00008	0.00110
11	*PSMC3*	21	7.90	0.00014	0.00110
6	*VGLL2*	45	23.02	0.00005	0.00113
11	*SLC39A13*	22	8.21	0.00014	0.00114
2	*XDH*	67	33.00	0.00004	0.00122
11	*MYBPC3*	28	9.68	0.00014	0.00135
11	*SPI1*	28	10.12	0.00014	0.00141
4	*GYPE*	7	2.40	0.00059	0.00142
16	*ADCY9*	98	49.84	0.00003	0.00153
2	*EHD3*	94	42.92	0.00004	0.00159

Abbreviation: min., minimum.
Genes giving adjusted, corrected min. (P) < 0.0016 are presented, ranked in ascending order of this variable.
[a]Obtained from the observed no. of SNPs genotyped and the LD matrix among them: for details, see text.
[b]Sidak-corrected minimum *P*-value, the correction adjusted by using the effective no. of SNPs in place of the actual no. of SNPs genotyped.

that showed the most significant association in our study (rs162209) shows no LD with it ($r^2 = 0.03$; D' = 0.2 HapMap CEU data). Some degree of association has been observed also for gene *CACNA1C*, where the most significantly associated SNP for BD[52]reported was not genotyped in our datasets and was not in LD with our associated SNP. The most significant association *P*-value for *ANK3* was observed for rs10509123 that showed a meta-analysis $P = 0.06$ that is not in LD with the two most significant SNPs from Ferreira *et al*.[54] study rs10994336 and rs1938526 that were not genotyped in our samples. Finally of some interest is the association with gene *5HT2A* where our best result was observed for SNP rs17289304. The SNP rs6313 (T102C) in this gene, for which previous mixed evidence of associations with MDD has been reported, did not show association in our datasets. The SNP rs17289304, also in this gene and reported to be associated with response with antidepressants[55] was not genotyped in our dataset and is not reported to be in LD with our best signal according to HapMap. For genes *GRM7*, *CACNA1C* and *5HT2A* regional plots with results for all SNPs tested in our study can be seen in the Supplementary material (Supplementary Figure 1).

The pathway analysis of the 104 most significant genes in the meta-analysis (identified by SNPs that displayed a meta-analysis, $P < 0.0005$) revealed that this subset was significantly enriched ($P < 0.05$) in members of four canonical pathways. The pathways identified and genes identified in each pathway were: (1) synaptic long-term depression pathway: 4 genes (*GRM7*, *ADCY9*, *GRID1*, *ITPR1*) out of 162 genes categorized in this pathway were present in our 104 most significant genes. As our top 104 genes represent 0.589% of the total number of genes tested in our GWAS, we should expect to identify about 1 gene belonging to this pathway ($162 \times 0.589/100 = 0.95$); (2)

Table 4 Significance of association with MDD case–control status in selected candidate genes

Gene name Suggested candidate SNP	*Chr*	*No of SNPs genotyped*[a]	*Most significant SNP ID*	*Min.* P *Sample I*	*Min.* P *Sample II*	*Min.* P *Meta-analysis*	*Direction*[b]	*Adjusted, corrected min. gene-based* (P)[c]
ANK3	10	149	rs10509123	0.11	0.37	0.06	++	0.97
Candidate SNP			rs10994336	NA	NA			
CACNA1C	12	214	rs16929470	0.001	0.30	0.001	++	0.12
Candidate SNP			rs1006737	NA	NA			
BDNF	11	35	rs908867	0.15	0.21	0.05	−−	0.59
Candidate SNP (val66met)			rs6265	0.16	0.50	0.55	+−	
COMT	22	77	rs4646316	0.30	0.014	0.025	++	0.55
Candidate SNP (val158met)			rs4680	0.38	0.24	0.15	−−	
DGKH	13		rs9525590	0.01	0.35	0.01	++	0.46
Candidate SNP			rs1012053	0.06	0.29	0.32		
DRD3	3	51	rs9834217	0.72	0.01	0.06	−−	0.76
Candidate SNP (gly9ser)			rs6280	0.78	0.77	0.97	−+	
GRM7	3	344	rs162209	0.0002	0.10	0.0001	++	0.01
Candidate SNP			rs1485171	NA	NA			
KCNC2	12		rs7304239	0.19	0.27	0.09	−−	0.90
Candidate SNP			rs1526805	NA	NA			
MYO5B	18	165	rs2292382	0.02	0.55	0.03	−−	0.89
Candidate SNP			rs4399921	NA	NA			
GABBR1	4	79	rs3129073	0.34	0.001	0.005	++	0.12
Candidate SNP			rs7680321	NA	NA			
SLC6A4 (5HTT)	17	24	rs11650871	0.37	0.18	0.12	++	0.80
5HT2A	13	85	rs17289304	0.11	0.004	0.002	−−	0.06
Candidate SNP			rs6313	0.93	0.22	0.38	−−	
SYN3	22	284	rs16990428	0.006	0.09	0.001	++	0.16
Candidate SNP			rs11089599	NA	NA			
TSPAN8	12	41	rs10444584	0.008	0.71	0.023	++	0.26
Candidate SNP			rs1705236	NA	NA			
TPH2	12	51	rs328765	0.22	0.059	0.031	−−	0.43
Candidate SNP			rs1386494	NA	NA	NA		

Abbreviations: BDNF, brain-derived neurotrophic factor; COMT, catechol-*O*-methyl transferase; DGKH, diacylglycerol kinase-η; ID, identification; Min., minimum; NA, not applicable; SNP, single nucleotide polymorphism; TPH2, tryptophan hydroxylase 2.

[a]Includes SNPs 100 kb upstream and downstream of coding region available for the meta-analysis. NA: means that the SNP was not genotyped by the illumina array or failed QC.

[b]++ and −− indicate that the direction of association is consistent in the two samples. +− and −+ indicate that it is inconsistent.

[c]For definition of this variable, see footnotes of Table 3.

cAMP-mediated signalling pathway, 4 genes (*GRM7*, *ADCY9*, *RGS7*, *AKAP7*) out of 161 in the pathway were present, compared to 1 expected; (3) G-protein-coupled receptor signalling pathway, 4 genes (*GRM7*, *ADCY9*, *RGS7*, *NFKB1*) out of 203 in the pathway were present, compared to 1 expected; (4) glutamate receptor signalling, 2 genes (*GRM7*, *GRID1*) out of 67 in the pathway were present, whereas none was expected.

Discussion

We report here the results from two GWAS of recurrent MDD and their meta-analysis. Our individual GWAS as the meta-analysis results showed low genomic control values, suggesting that our cases and controls are relatively homogenous, and that we can, with reasonable confidence, exclude a large distortion in our results caused by residual genetic substructure. The total sample size of 1359 recurrent MDD cases and 1782 controls analysed in our meta-analysis represents one of the largest samples used in genetic association studies of MDD. Furthermore, our cases from both GWAS were carefully selected to have at least two distinct major depression episodes (that is, recurrent MDD) that characterize a subtype of MDD suggested to have higher familiality.[8] Nevertheless, the main finding from our meta-analysis is that none of the over half million SNPs analysed reached association at the genome-wide level $P<5\times10^{-8}$,[42] conceivably for the lack of adequate statistical power provided by our sample size.

Recent GWAS of various common medical conditions have identified and confirmed over 70 common SNPs and showed that each of these SNPs has a small contribution to the increased disease risk: mean odds ratios = 1.36.[56] These GWAS clearly indicated that sample size in the order of several thousands are required to provide sufficient statistical power to detect SNPs with such small effect.[57] This is likely to be the case also for psychiatric disorders, as suggested by initial GWAS efforts for schizophrenia and BD[51–54,58–61] and by our work, which represents the first GWAS to be reported for recurrent MDD. The sample size used in our study was clearly not powered enough to detect SNP with small effects. The lack of adequate statistical power is the main limitation of our study.

Previous candidate gene studies of major depression have provided inconclusive evidence of association for genes with strong biological rationale for MDD.[62] We queried our GWAS and meta-analysis results for such candidates and for candidate risk genes identified from BD GWAS as BD relatives are at increased risk for MDD[63] and some degree of genetic overlap between MDD and BD is likely to exist.[51–53] As expected, none of the SNPs in the candidate genes showed *P*-values that would remain significant after correction for multiple testing at the whole-genome level. Although, it has been suggested that given the strong *a priori* biological hypothesis for candidate genes their results should be interpreted with a more liberal statistical criterion, the inconsistent results between previous studies and ours does not support any of the previously reported allelic associations for MDD. More powered studies will indicate whether these genes that continue to be strong biological candidates have indeed an effect in increasing risk for depression. It has to be considered, for future studies that if substantial allelic heterogeneity exists for any of these genes it would become a real challenge to show convincing evidence for association even with substantially more powered studies. For example, our and previously published results suggest that some degree of allelic heterogeneity may exist for *GRM7*. The strongest signal for *GRM7* in our meta-analysis was observed for rs162209 ($P=0.0001$); however, the association was almost exclusively derived from our Study I. Study II showed moderate association ($P=0.005$) for another distant and not correlated SNP (rs983534). Also, the most strongly associated *GRM7* SNP from BD WTCCC GWAS[53] was not correlated to our strongest *GRM7* SNPs. Overall, *GRM7* showed multiple independent SNPs to be associated to some degree with MDD, and this is captured by our gene-based association where the Šidák correction for multiple SNPs tested, adjusted on the basis of the LD matrix among the SNPs, gave the value $P=0.01$. *GRM7* is a strong biological candidate as codes for an mGluR7 that is emerging as a potential target for the treatment of mood disorders.[64] The gene has a suggested role in depression from pre-clinical evidence,[65–67] and future genetic studies are critical for clarifying its role in mood disorders. Of interest to report that a recent genome-wide linkage scan, we have conducted in recurrent MDD sibling pairs that showed the best evidence for linkage, when accounting for disease severity, on chromosome 3 where *GRM7* is located.[12]

Among the other candidate genes investigated *CACNA1C* and *5HT2A* showed modest associations in our study. Again, the lack of consistent SNP showing association and the lack of overlap in SNP genotyped in our study with previously associated SNP for *CACNA1C*, make also difficult to draw any conclusions on *CACNA1C* and *5HT2A*.

We were conscious of the fact that all our most significant associations could all be spurious results. Nevertheless, as it is more likely for real associations to be present among the most significant SNPs, we decided to explore whether within the genes identified by the most significant SNPs there was an enrichment of biological pathways that may be relevant to MDD. We found that genes encoding for proteins involved in synaptic long-term depression, cAMP-mediated signalling, G-protein coupled signalling and glutamate receptor signalling functions were moderately overrepresented among our most significant genes. However, considering that the pathways identified showed statistical significance that does not survive correction for multiple testing results from larger GWAS should be sought before making any conclusion.

The new collaborative attitude in psychiatric genetics has led to the formation of large multicentre efforts to analyse jointly a number of GWAS. The recently formed Psychiatric Genome Wide Association Study Consortium (PGC) is aiming to analyse multiple GWAS on approximately 13 000 cases and 10 000 controls that will include samples used in the present study. PGC would allow identifying SNPs of small effect and also testing, with reasonable power, the presence of genetic risk factors that predispose to MDD subtypes and to relevant symptom dimensions. Our study ascertained MDD cases with recurrent episodes reducing to some degree the substantial heterogeneity that exist for MDD; additional evidences however suggest that other clinical variables as severity[12] and age at onset[15,22] that appear to define more heritable phenotypes, should be considered for future analysis.

On a broader perspective on the genetics of MDD, we should consider that the initial analytical effort on the high definition whole-genome scans genotypings has concentrated in evaluating the CDCV hypothesis. However, several lines of recent evidence from the analysis of copy number variation (CNV) in schizophrenia[68–70] and autism[71] are suggesting that rare and penetrant gene variants are likely to account to some of the genetic risk contribution to psychiatric disorders. Substantially, larger samples than the one used in our study would also allow testing the common disease-rare variant hypothesis (CDRV) through direct detection of CNVs as well as through other methods such as the population-based linkage analysis of SNP genotypes.[72] In addition, to fully capture the genetic architecture of MDD the analysis of genes would have to include the interaction with known environmental risk factors for MDD (for example, childhood abuse and stressful life events[71,73,74]).

In summary, considerable hope has relied, and continues to rely on human genetics for improving disease prevention and treatment strategies.[58] This is particularly true for depression, where the aetiology is mostly unknown. In the present study, we have been able to exclude the presence of common SNPs, captured by half million genotyping arrays that have a substantial effect in increasing the risk for MDD. Our sample size did not provide adequate statistical power to detect SNPs with small effect. The numerous MDD GWAS in progress and the joint analysis of these studies provide ground for identifying small effect genetic risk factors for MDD.

Acknowledgments

We acknowledge all the participants in the studies. We thank numerous people at GSK and Max-Planck Institute, BKH Augsburg and Klinikum Ingolstadt in Germany and the Staff and the Department of Psychiatry and Internal Medicine at the University of Lausanne who have contributed to this project. In particular, GSK thanks Paul M Matthews for his support to the project over the past year and to Shyama Brewster for initiating and coordinating the first phase of recruitment at the Germany Sites.

References

1 Waraich P, Goldner EM, Somers JM, Hsu L. Prevalence and incidence studies of mood disorders: a systematic review of the literature. *Can J Psychiatry* 2004; **49**: 124–138.
2 Mann JJ. The medical management of depression. *N Engl J Med* 2005; **353**: 1819–1834.
3 World Health Organization. *World Health Report 2001. Mental health: New Understanding, New Hope*, Geneva WHO2 editor 2008.
4 Belmaker RH, Agam G. Major depressive disorder. *N Engl J Med* 2008; **358**: 55–68.
5 Farmer A, Harris T, Redman K, Sadler S, Mahmood A, McGuffin P. Cardiff depression study. A sib-pair study of life events and familiality in major depression. *Br J Psychiatry* 2000; **176**: 150–155.
6 Jones I, Kent L, Craddock N. Genetics of affective disorders. In: McGuffin P, Owen MJ, Gottesman II (eds). *Psychiatric Genetics and Genomics*. Oxford University Press: Oxford, 2002, pp 211–246.
7 Shih RA, Belmonte PL, Zandi PP. A review of the evidence from family, twin and adoption studies for a genetic contribution to adult psychiatric disorders. *Int Rev Psychiatry* 2004; **16**: 260–283.
8 Sullivan PF, Neale MC, Kendler KS. Genetic epidemiology of major depression: review and meta-analysis. *Am J Psychiatry* 2000; **157**: 1552–1562.
9 Abkevich V, Camp NJ, Hensel CH, Neff CD, Russell DL, Hughes DC *et al*. Predisposition locus for major depression at chromosome 12q22-12q23.2. *Am J Hum Genet* 2003; **73**: 1271–1281.
10 Holmans P, Zubenko GS, Crowe RR, DePaulo Jr JR, Scheftner WA, Weissman MM *et al*. Genomewide significant linkage to recurrent, early-onset major depressive disorder on chromosome 15q. *Am J Hum Genet* 2004; **74**: 1154–1167.
11 McGuffin P, Knight J, Breen G, Brewster S, Boyd PR, Craddock N *et al*. Whole genome linkage scan of recurrent depressive disorder from the depression network study. *Hum Mol Genet* 2005; **14**: 3337–3345.
12 Webb BT, van den Oord EJCG, McGuffin P, Knight J, Breen G, Brewster S *et al*. Genome-wide linkage analysis using additional families from the depression network study. *Am J Med Genet* 2006; **141B**: 683–824.
13 Risch NJ. Searching for genetic determinants in the new millennium. *Nature* 2000; **405**: 847–856.
14 Reich DE, Lander ES. On the allelic spectrum of human disease. *Trends Genet* 2001; **17**: 502–510.
15 Levinson DF. The genetics of depression: a review. *Biol Psychiatry* 2006; **60**: 84–92.
16 Barrett JC, Cardon LR. Evaluating coverage of genome-wide association studies. *Nat Genet* 2006; **38**: 659–662.
17 Celik C. *Computer Assisted Personal Interviewing Application for the Schedules for Clinical Assessment in Neuropsychiatry Version 2.1 and Diagnostic Algorithms for WHO ICD 10 chapter V DCR and for Statistical Manual IV*, Release I. Ed. 1.0.3.5. ed. World Health Organization: Geneva, 2003.
18 World Health Organization. *Diagnosis and Clinical Measurement in Psychiatry. A Reference Manual for SCAN*. World Health Organization: Geneva, 1998.
19 American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders (DSM-IV)*, 4th edn American Psychiatric Press: Washington DC, 1994.
20 World Health Organization. *The ICD-10 Classification of Mental and Behavioural Disorders. Diagnostic Criteria for Research*. World Health Organization: Geneva, 1993.
21 Hettema JM. What is the genetic relationship between anxiety and depression? *Am J Med Genet C Semin Med Genet* 2008; **148**: 140–146.
22 Tozzi F, Prokopenko I, Perry JD, Kennedy JL, McCarthy AD, Holsboer F *et al*. Family history of depression is associated with

younger age of onset in patients with recurrent depression. *Psychol Med* 2008; **38**: 641–649.
23 Wittchen HU, Hoefler M, Gander F, Pfister H, Storz S, Ustun TB *et al.* Screening for mental disorders: performance of the Composite International Diagnostic-Screener (CID-S). *Int J Methods Psychiatr Res* 1999; **8**: 59–70.
24 Preisig M, Vollenweider P, Bovet P, Rothen S, Vandeleur C, Mooser V *et al.* PsyCoLaus: a cross-sectional population-based study on psychiatric disorders and their association with cardiovascular risk factors. 2008 submitted.
25 Firmann M, Mayor V, Vidal PM, Bochud M, Pecoud A, Hayoz D *et al.* The CoLaus study: a population-based study to investigate the epidemiology and genetic determinants of cardiovascular risk factors and metabolic syndrome. *BMC Cardiovasc Disord* 2008; **8**: 6.
26 Goldberg DP. *The Detection of Psychiatric Illness by Questionnaire. Maudsley Monograph 21 ed*. Oxford University Press: Oxford, 1972.
27 Bettschart W, Bolognini M. Questionnaire de santé GHQ-12. In: Guelfi JD (ed). *L'évaluation Clinique Standardisée en Psychiatrie*. Tome I: Boulogne, 1996, p 157.
28 Leboyer M, Barbe B, Gorwood P, Teherani M, Allilaire JF, Preisig M *et al. Interview Diagnostique pour les Etudes Génétiques*. INSERM: Paris, 1995.
29 Andreasen NC, Endicott J, Spitzer RL, Winokur G. The family history method using diagnostic criteria. Reliability and validity. *Arch Gen Psychiatry* 1977; **34**: 1229–1235.
30 Fellay J, Shianna KV, Ge D, Colombo S, Ledergerber B, Weale M *et al.* A whole-genome association study of major determinants for host control of HIV-1. *Science* 2007; **317**: 944–947.
31 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.
32 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.
33 Patterson N, Price AL, Reich D. Population structure and eigen analysis. *PLoS Genet* 2006; **2**: e190.
34 Sandhu MS, Waterworth DM, Debenham SL, Wheeler E, Papadakis K, Zhao JH *et al.* LDL-cholesterol concentrations: a genome-wide association study. *Lancet* 2008; **371**: 483–491.
35 Devlin B, Roeder K. Genomic control for association studies. *Biometrics* 1999; **55**: 997–1004.
36 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.
37 Nyholt DR. A simple correction for multiple testing for single-nucleotide polymorphisms in linkage disequilibrium with each other. *Am J Hum Genet* 2004; **74**: 765–769.
38 Li J, Ji L. Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix. *Heredity* 2005; **95**: 221–227.
39 Šidák Z. Rectangular confidence regions for the means of multivariate normal distributions. *J Am Stat Assoc* 1967; **62**: 626–633.
40 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.
41 Becker KG, Barnes KC, Bright TJ, Wang SA. The genetic association database. *Nat Genet* 2004; **36**: 431–432.
42 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.
43 George M, Ying G, Rainey MA, Solomon A, Parikh PT, Gao Q *et al.* Shared as well as distinct roles of EHD proteins revealed by biochemical and functional comparisons in mammalian cells and *C elegans*. *BMC Cell Biol* 2007; **8**: 3.
44 Toyota T, Yamada K, Saito K, Detera-Wadleigh SD, Yoshikawa T. Association analysis of adenylate cyclase type 9 gene using pedigree disequilibrium test in bipolar disorder. *Mol Psychiatry* 2002; **7**: 450–452.
45 Toyota T, Hattori E, Meerabux J, Yamada K, Saito K, Shibuya H *et al.* Molecular analysis, mutation screening, and association study of adenylate cyclase type 9 gene (ADCY9) in mood disorders. *Am J Med Genet* 2002; **114**: 84–92.
46 Cui DH, Jiang KD, Jiang SD, Xu YF, Yao H. The tumor suppressor adenomatous polyposis coli gene is associated with susceptibility to schizophrenia. *Mol Psychiatry* 2005; **10**: 669–677.
47 Kishi T, Ikeda M, Suzuki T, Kitajima T, Yamanouchi Y, Kinoshita Y *et al.* No association of complexin1 and complexin2 genes with schizophrenia in a Japanese population. *Schizophr Res* 2006; **82**: 185–189.
48 Lee HJ, Song JY, Kim JW, Jin SY, Hong MS, Park JK *et al.* Association study of polymorphisms in synaptic vesicle-associated genes, SYN2 and CPLX2, with schizophrenia. *Behav Brain Funct* 2005; **1**: 15.
49 Fallin MD, Lasseter VK, Avramopoulos D, Nicodemus KK, Wolyniec PS, McGrath JA *et al.* Bipolar I disorder and schizophrenia: a 440-single-nucleotide polymorphism screen of 64 candidate genes among Ashkenazi Jewish case-parent trios. *Am J Hum Genet* 2005; **77**: 918–936.
50 Guo SZ, Huang K, Shi YY, Tang W, Zhou J, Feng GY *et al.* A case-control association study between the GRID1 gene and schizophrenia in the Chinese Northern Han population. *Schizophr Res* 2007; **93**: 385–390.
51 Baum AE, Akula N, Cabanero M, Cardona I, Corona W, Klemens B *et al.* A genome-wide association study implicates diacylglycerol kinase eta (DGKH) and several other genes in the etiology of bipolar disorder. *Mol Psychiatry* 2008; **13**: 197–207.
52 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.
53 WTCCC. Genome-wide association study of 14 000 cases of seven common diseases and 3000 shared controls. *Nature* 2007; **447**: 661–678.
54 Ferreira MA, 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.
55 Paddock S, Laje G, Charney D, Rush AJ, Wilson AF, Sorant AJ *et al.* Association of GRIK4 with outcome of antidepressant treatment in the STAR*D cohort. *Am J Psychiatry* 2007; **164**: 1181–1188.
56 Bodmer W, Bonilla C. Common and rare variants in multifactorial susceptibility to common diseases. *Nat Genet* 2008; **40**: 695–701.
57 Altshuler D, Daly M. Guilt beyond a reasonable doubt. *Nat Genet* 2007; **39**: 813–815.
58 Kingsmore SF, Lindquist IE, Mudge J, Gessler DD, Beavis WD. Genome-wide association studies: progress and potential for drug discovery and development. *Nat Rev Drug Discov* 2008; **7**: 221–230.
59 Lencz T, Morgan TV, Athanasiou M, Dain B, Reed CR, Kane JM *et al.* Converging evidence for a pseudoautosomal cytokine receptor gene locus in schizophrenia. *Mol Psychiatry* 2007; **12**: 572–580.
60 Shifman S, Johannesson M, Bronstein M, Chen SX, Collier DA, Craddock NJ *et al.* Genome-wide association identifies a common variant in the reelin gene that increases the risk of schizophrenia only in women. *PLoS Genet* 2008; **4**: e28.
61 Sullivan PF, Lin D, Tzeng JY, van den OE, Perkins D, Stroup TS *et al.* Genomewide association for schizophrenia in the CATIE study: results of stage 1. *Mol Psychiatry* 2008; **13**: 570–584.
62 Levinson DF. The genetics of depression: a review. *Biol Psychiatry* 2006; **60**: 84–92.
63 Levinson DF. The genetics of depression: a review. *Biol Psychiatry* 2006; **60**: 84–92.
64 Witkin JM, Marek GJ, Johnson BG, Schoepp DD. Metabotropic glutamate receptors in the control of mood disorders. *CNS Neurol Disord Drug Targets* 2007; **6**: 87–100.
65 Callaerts-Vegh Z, Beckers T, Ball SM, Baeyens F, Callaerts PF, Cryan JF *et al.* Concomitant deficits in working memory and fear extinction are functionally dissociated from reduced anxiety in metabotropic glutamate receptor 7-deficient mice. *J Neurosci* 2006; **26**: 6573–6582.
66 Cryan JF, Kelly PH, Neijt HC, Sansig G, Flor PJ, van der PH. Antidepressant and anxiolytic-like effects in mice lacking the group III metabotropic glutamate receptor mGluR7. *Eur J Neurosci* 2003; **17**: 2409–2417.
67 Mitsukawa K, Mombereau C, Lotscher E, Uzunov DP, van der PH, Flor PJ *et al.* Metabotropic glutamate receptor subtype 7 ablation

causes dysregulation of the HPA axis and increases hippocampal BDNF protein levels: implications for stress-related psychiatric disorders. *Neuropsychopharmacology* 2006; **31**: 1112–1122.
68 Walsh T, McClellan JM, McCarthy SE, Addington AM, Pierce SB, Cooper GM *et al.* Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia. *Science* 2008; **320**: 539–543.
69 Stefansson H, Rujescu D, Cichon S, Pietilainen OP, Ingason A, Steinberg S *et al.* Large recurrent microdeletions associated with schizophrenia. *Nature* 2008; **455**: 232–236.
70 Stone JL, O'Donovan MC, Gurling H, Kirov GK, Blackwood DH, Corvin A *et al.* Rare chromosomal deletions and duplications increase risk of schizophrenia. *Nature* 2008; **455**: 237–241.
71 Sebat J, Lakshmi B, Malhotra D, Troge J, Lese-Martin C, Walsh T *et al.* Strong association of *de novo* copy number mutations with autism. *Science* 2007; **316**: 445–449.
72 Francks C, Tozzi F, Farmer A, Vincent JB, Rujescu D, St Clair D *et al.* Population-based linkage analysis of schizophrenia and bipolar case-control cohorts identifies a potential susceptibility locus on 19q13. *Mol Psychiatry*; advance online publication 16 September 2008; doi:10.1038/mp.2008.100.
73 Kendler KS, Gardner CO, Prescott CA. Toward a comprehensive developmental model for major depression in women. *Am J Psychiatry* 2002; **159**: 1133–1145.
74 Kendler KS, Kuhn JW, Prescott CA. Childhood sexual abuse, stressful life events and risk for major depression in women. *Psychol Med* 2004; **34**: 1475–1482.

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