Supplemental Methods: GWA for MDD

Authors Sullivan, et al.

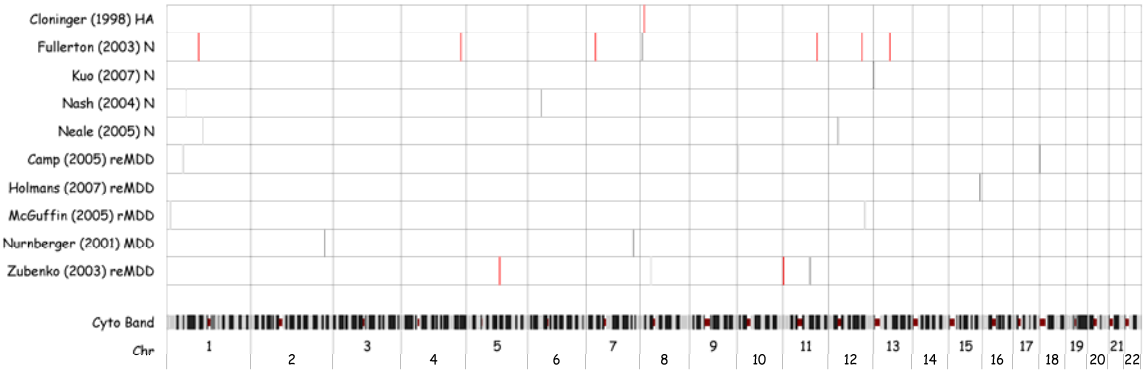
Document: "MDD-etiology-supplement.doc"

Date: 5/21/2010 12:33 PM

Table of Contents

GENOMEWIDE LINKAGE STUDIES OF MDD OR NEUROTICISM

The figure below depicts results of genomewide linkage studies of MDD. This figure was generated using the SLEP database [1]. The genomic position is shown on the x-axis. Each row corresponds to an individual study [2-10] with LOD scores or equivalents plotted (red LOD≥3, gray LOD≥2, & light gray LOD≥1.5).



SUPPLEMENTAL METHODS

Assessment of MDD, depressive symptoms, and other psychopathology indicators

NESDA - The Composite International Diagnostic Interview (CIDI), section E, version 2.1 [11] was used to diagnose MDD. The interview also provided information on age of onset, number of episodes of MDD, and specific symptoms of depression. Information on lifetime comorbid panic disorder with or without agoraphobia, generalized anxiety disorder, social phobia, alcohol dependence was also collected with the CIDI. Depression and anxiety severity indicators include the Inventory of Depressive Symptoms-self-report (IDS-SR) [12] , the Fear questionnaire [13], and the Beck Anxiety Inventory (BAI) [14]. Neuroticism, an endophenotype for MDD, was assessed with the NEO [15]. The family tree inventory was used to examine depression in first degree relatives.

NTR - Phenotyping for depression in the NTR survey studies took place with the Beck depression inventory (BDI) [16] in 1993 and 1997 and with the depression scale from the YASR [17]

in 1991, 1995, 1997 and 2000. Neuroticism was assessed with the ABV [18] in 5 out of 6 surveys (not in 1995) and with the NEO [15] in 2004. Anxiety (STAI) [19]a was assessed at 5 out 7 surveys (not in 1995 and 2004). NTR cases and a subsample of controls underwent the CIDI protocol by phone; either as part of an earlier study [20] on the heritability of major depression and anxiety disorders or as part of the selection procedure for GAIN.

Genotyping.

Perlegen performed all laboratory processes strictly according to standard operating procedures (SOPs). The use of a Laboratory Information Management System (LIMS) is an integral part of each of these steps, ensuring adherence to experimental design, and use of the correct PCR primer plates, arrays, reagents, instruments, and samples.

DNA samples were shipped to Perlegen in 96 well plates and samples were identified only by barcode. DNA samples were quantified using multiple independent measurements of DNA concentration by PicoGreen (Invitrogen Corporation, Carlsbad, CA, USA) and samples were normalized to equal concentrations (CV ± 5%).

DNA from 96-well plates was used to inoculate 4-8 PCR reactions for each sample. These PCR reactions assay a different set of a few hundred “IDQC” SNPs for each of the samples. These 96 samples are then genotyped together on a chip. When the sample is subsequently genotyped for all of the SNPs used in the study, the genotypes of these few hundred IDQC SNPs are compared between the IDQC scan (as the DNA plate as it came in the door) and the experimental scans for that sample. If there was a sample mix-up in the processing of the sample, a pair of samples will match each other’s IDQC SNP genotypes. These samples can then be re-processed to verify the correct sample identity.

Perlegen uses high-density oligonucleotide arrays for custom genotyping [21]. These arrays are generated by light-directed photolithography in conjunction with chemical coupling to direct the synthesis of oligonucleotides of specific DNA sequence in pre-determined positions on a

glass surface. To genotype each SNP, there are 24 oligonucleotide probes with each probe 25 nucleotides in length. The 24 probes per SNP are sub-divided into sets. Four sets total 20 of the 24 probes and are “perfect match” probes. The 20 perfect match probes correspond to the four combinations of forward/reverse strand and reference/alternate SNP allele and five probes per combination. Each of these five probes are slightly different with offsets of -2, -1, 0, +1, and +2 bases between the center of the 25 nucleotide probe and the SNP position. Finally, four of the 24 probes per SNP are “mismatch” probes at offset 0 for the four combinations of forward/reverse strand and reference/alternate SNP allele. SNP selection is described elsewhere [22-24].

Genotype Determination From Probe Intensities.

Individual genotypes were determined by clustering all SNP-scans in the 2-dimensional space defined by $\hat{p}$ values and signals after the processing steps described below.

The term “SNP-scan” refers to the collection of both reference and alternate allele-specific probe intensities for a given SNP from GWA scan from an individual subject. These intensities were used to determine the individual’s genotype. The $\hat{p}$ and signal values were obtained by transformations from the reference and alternate allele trimmed mean intensities (described below), and correspond to the intensity amplitude and angle (which reflects the allele frequency) in the reference and alternate intensity space. Background was computed similarly as signal from the mismatch probe intensities as described below.

Before the transformation of intensities to final signal and $\hat{p}$ values for genotyping, the intensities were preprocessed by normalizing signal/background ratios to a target value and by normalizing average signals of groups of SNP-scans. The SNP-scans were grouped by processing the density of $\hat{p}$ values smoothed by a Gaussian kernel. This grouping results in a number of SNP-scan groups (<10) that reflect the clustering of $\hat{p}$ values. Each group was represented by a median reference and alternate intensity and a transformation was applied to

the underlying intensities to equalize signals of these median intensities. Any consistent bias between intensity values produced by the reference and alternate allele probes is thus removed. This normalization helps to center the SNP-scans such that heterozygous genotypes are distributed around $\hat{p}$ =0.5 and the Reference and Alternate allele genotypes are symmetrically on opposite sides of the heterozygous genotype cluster. The signal/background normalization, that precedes the above normalization of group signals, corrects the underlying intensities and tightens the $\hat{p}$ variance of genotype clusters. It is the property of the intensities that SNP-scans with low signal/background ratio tend to have $\hat{p}$ values closer to 0.5 than SNP-scans with high signal/background ratio. The signal/background ratio normalization transforms all SNP-scans onto the same target signal/background ratio, thus correcting the mapping between the underlying allele frequencies and the $\hat{p}$ values.

After the normalization step the SNP-scans are assigned genotypes in a genotyping cluster search procedure that aimed to maximize the posterior probability of a genotype assignment given the $\hat{p}$ distribution for the genotyped samples. The posterior probability is computed using a Bayes rule from the conditional likelihood of the $\hat{p}$ values given the clustering model and from a set of prior probability distributions for clustering model parameters. The conditional likelihood of $\hat{p}$ values given the clustering model was estimated using a normal mixture model for the distribution of $\hat{p}$ values around the cluster means. The prior probability distributions for the normal mixture model parameters and the genotypes assigned to the clusters based on their mean $\hat{p}$ values were derived from high quality SNPs previously genotyped using a different algorithm.

The maximally likely genotype assignment was selected from a set of possible genotype clusterings that resulted from a genotyping cluster search procedure. For these diploid individuals this procedure explored possible one-cluster two-cluster and three-cluster $\hat{p}$

groupings with all possible genotype assignments and selected the genotype clustering with the highest posterior probability. To limit the search space of the $\hat{p}$ groupings, the $\hat{p}$ values were grouped using their density distribution smoothed with a Gaussian kernel. This procedure resulted in a maximum of 10 groups that were combined to produce all possible one-cluster, two-cluster or three-cluster genotype solutions.

A genotyping quality metric was compiled for each genotype from 28 input metrics that described the quality of the SNP, the particular scan, and the genotype. The genotyping quality metric correlated with a probability of having a discordant call between the Perlegen platform and outside genotyping platforms (i.e., non-Perlegen HapMap project genotypes). A system of 10 bootstrap aggregated regression trees ("bagging") was trained using an independent data set of concordance data between Perlegen genotypes and HapMap project genotypes. The trained predictor was then used to predict the genotyping quality for each of the genotypes in this data set. The performance of the trained predictor was evaluated with an independent data set and produced a receiver operating characteristic (ROC) curve with an area-under-curve of 0.95. Such high area under the ROC curve indicates high sensitivity and specificity of the predictor at optimal cutoffs for the produced quality metric.

Allele frequencies were approximated using the intensities collected from the high-density oligonucleotide arrays. A SNPs allele frequency *p* was a ratio of the relative amount of the DNA with reference allele to the total amount of DNA, and thus can have values 0, 0.5 or 1 for diploid DNA:

$$p = \frac{C_{\mathrm{Ref}}}{C_{\mathrm{Ref}} + C_{\mathrm{Alt}}}$$

where C_{REF} and C_{ALT} are the concentrations of reference allele and alternate allele, respectively. As probe intensities were directly related to the concentrations of the SNP alleles, the $\hat{p}$ computed from the intensities of reference and alternate features was a good

approximation of the true allele frequency p. The $\hat{p}$ value was computed from the trimmed mean intensities of perfect match features:

$$\hat{p} = \frac{I^{TM}_{PM,\mathrm{Re}f}}{I^{TM}_{PM,\mathrm{Re}f} + I^{TM}_{PM,Alt}}$$

where:

$$I^{TM}_{PM,\mathrm{Re}f} = \frac{I^{TM}_{PM,\mathrm{Re}f,Fwd} + I^{TM}_{PM,\mathrm{Re}f,\mathrm{Re}v}}{2}$$

$$I^{TM}_{PM,Alt} = \frac{I^{TM}_{PM,Alt,Fwd} + I^{TM}_{PM,Alt,\mathrm{Re}v}}{2}$$

and $I^{TM}_{PM,\mathrm{Re}f,Fwd}$, $I^{TM}_{PM,\mathrm{Re}f,\mathrm{Re}v}$, $I^{TM}_{PM,Alt,Fwd}$ and $I^{TM}_{PM,Alt,\mathrm{Re}v}$ were the trimmed means of perfect match intensities for a given allele and strand denoted by the subscript. The trimmed mean disregarded the highest and the lowest intensity from the five perfect match intensities before computing the arithmetic mean.

The signal and background were computed as follows:

$$signal = \sqrt{(I^{TM}_{PM,\mathrm{Re}f})^2 + (I^{TM}_{PM,Alt})^2}$$

$$background = \sqrt{((I_{MM,\mathrm{Re}f,Fwd} + I_{MM,\mathrm{Re}f,\mathrm{Re}v})/2)^2 + ((I_{MM,Alt,Fwd} + I_{MM,Alt,\mathrm{Re}v})/2)^2}$$

where $I_{MM,\mathrm{Re}f,Fwd}$, $I_{MM,\mathrm{Re}f,\mathrm{Re}v}$, $I_{MM,Alt,Fwd}$ and $I_{MM,Alt,\mathrm{Re}v}$ are the intensities of the four mismatch probes.

Copy Number Variation.

In order to maximize genotyping signal, Perlegen chose the number of PCR cycles such that product amplification was maximized. This decision makes complicates considerably detection of subtle differences in allele intensities consistent with copy number variation (CNV). Algorithm development in this area is in progress. However, Dr Dennis Ballinger of Perlegen conveyed that they were unable to develop a CNV calling algorithm, even with the use of the

HapMap samples as positive and negative controls. This issue had been encountered previously with the Perlegen HapMap 2 [22] genotypes (Drs. Gonçalo Abecasis and Mark Daly, personal communication).

Use of More Stringent Genotype Quality Score Thresholds

The algorithms described above convert raw intensity scores for the multiple probes that interrogate a SNP allele to allele intensities and then to genotypes. As part of this process, the algorithm yields a quality score for each genotype. In discussion with the GAIN Statistical analysis group (chaired by Dr Gonçalo Abecasis of the University of Michigan), we increased the stringency of the passing quality score threshold from ≥ 7 to ≥ 10 meaning that all genotypes with quality scores from [0-9) were set to missing.

Plate Layout

Forty 96-well plates were sent to Perlegen for GWAS genotyping. Genotyping was conducted blind to case-control status. Cases and controls were randomly allocated to plates and positions within plates. Each plate contained DNA samples from 93 Dutch subjects plus 3 QC samples. The three QC samples included: two parents of one control sample on that plate (40 complete trios in total); and half the plates contained the same HapMap CEU sample (used for GC in all GAIN projects) and half contain a randomly selected duplicate case sample. The total number of samples was 3,840 (=40 plates x 96 samples per plate) or 1,860 cases + 1,860 controls + 80 parents + 20 duplicate samples + 20 HapMap samples.

Quality Control – Subjects

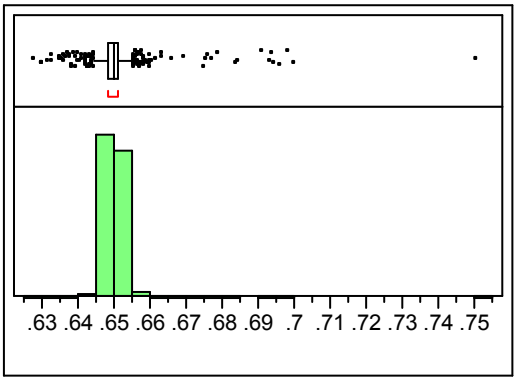
After eliminating the 20 HapMap samples used as internal controls for GAIN, there were 3,820 Dutch samples sent to Perlegen. Genotypes were delivered for 3,761 samples (98.5%). The quality control process for subjects had the following steps:

(a) There were 59 samples excluded (=3,820-3,761) for the following reasons: 39 samples with uncertain linkage between genotype and phenotype records, 7 samples with evidence of contamination, 6 samples that failed genotyping, and 7 miscellaneous failures.

(b) Initial analysis of the GWAS data showed that 8 subjects had missing data for >25% of the SNPs, consistent with failure of one of the four chips that constitute the Perlegen GWAS platform. These individuals were excluded.

(c) Three subjects were genomic outliers. One case had high genomewide homozygosity (~75%, ***Supplemental Figure 1***), one case had chrX homozygosity plus no chrY genotype calls, and one case had a substantial number of chrX heterozygous genotypes plus multiple chrY genotype. These three subjects were excluded.

Supplemental Figure 1: Subject homozygosity

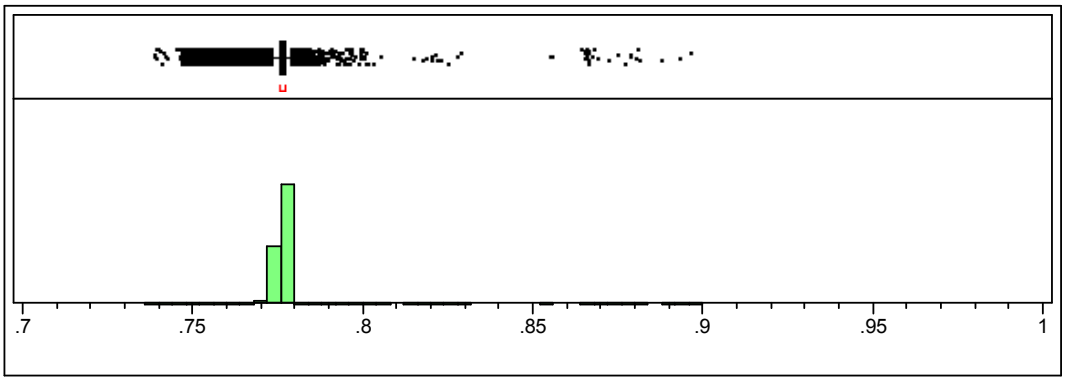


(d) After preliminary SNP exclusions (SNP missingness > 0.05, MAF < 0.01, ≥2 disagreements in duplicated samples, or ≥2 Mendelian inheritance errors), a PLINK [25] dataset containing GWAS data for cases and controls was created and subject characteristics explored in multiple ways.

We used PLINK to select a subset of SNPs (~127K) in linkage equilibrium (LE). The LE subset of the full GWAS data was used to explore relatedness and ancestry. The LE SNP subset was used: (i) to compute genomewide IBS (identity-by-state) estimates for all pairs of subjects in the samples ($_{3,635}C_2$ = 6,604,795 unique pairs of subjects) using PLINK; (ii) to detect

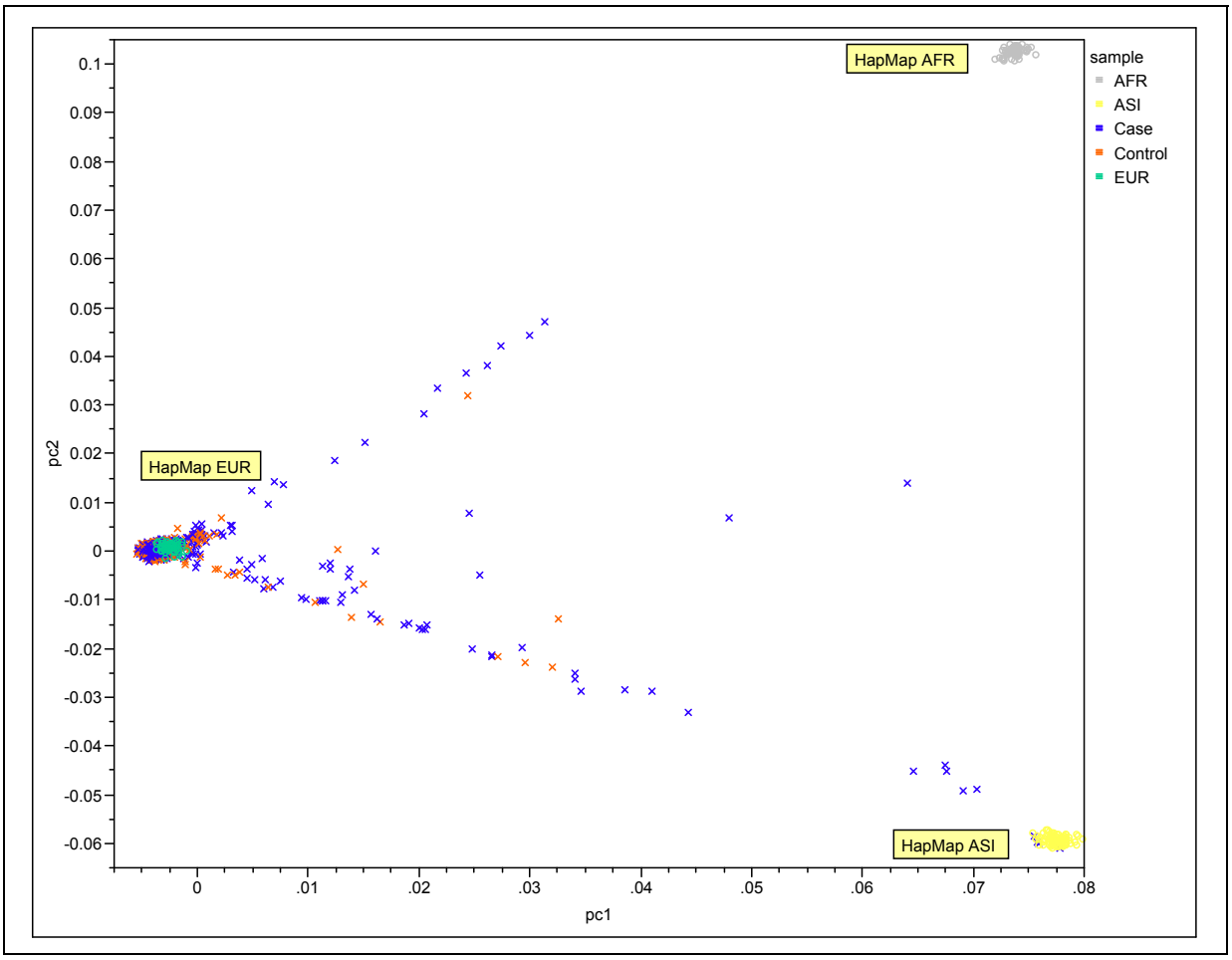
sample outliers using “nearest neighbour” statistical estimates calculated from genomewide IBS estimates; and (iii) to compute principal component estimates using the “smartpca” module of EigenSoft [26] with genotype data (corresponding to the LE SNP subset) from the HapMap [22] AFR, ASI, and EUR panels included.

Supplemental Figure 2: Genomewide IBS estimates for all pairs of subjects



Genomewide IBS estimates are shown in ***Supplemental Figure 2***. There appeared to be four clusters of subject pairs – those with low values (~0.74), putative first degree relatives (0.85-0.90), putative second degree relatives (0.81-0.83), and the rest of the distribution who seem to be “unrelated” in the conventional sense of the term. All pairs of subjects with low values included the individual with conspicuously high genomewide homozygosity. For the pairs with high relatedness (≥0.81, the arbitrary but natural cutoff in the data), single subjects were removed in a manner that maximized the number of cases remaining.

Supplemental Figure 3: Plot of PC2 x PC1 based on 127K GWAS SNPs in LE.



A plot of the second by the first principal component is shown in ***Supplemental Figure 3***. The three HapMap panels were included in order to provide context. Several observations are supported by this figure. Inclusion of the HapMap samples allows interpretation of the PCAs, and PC1 separates EUR versus AFR/ASI groups and PC2 distinguishes all three groups. The majority of subjects (97.6%) are in the cluster labeled “HapMap EUR”; however, there is stratification in the sample despite all subjects endorsing “Dutch” ancestry. A sizeable number of subjects have proportions of ASI and AFR ancestry. We hypothesize that this reflects the geopolitical history of the Netherlands and its colonization efforts in Indonesia and Surinam. Finally, we note that more cases than controls have ancestries that diverge from the EUR group.

(e) Final subject exclusions. After excluding subjects from Steps a, b, and c above, we integrated the information regarding relatedness and ancestry (from nearest neighbor and principal components). To evaluate the impact of different exclusion strategies, we used the lambda statistic [27] evaluated blind to SNP identity (i.e., as this was used only for quality control, it does not constitute an analysis for which multiple testing adjustments needs to be made). Lambda is calculated as the median χ^2 from Cochran-Armitage trend association tests for all autosomal SNPs divided by the χ^2 expectation under a null model (~0.455). For an ancestrally homogeneous sample of unrelated individuals with no true associations, the value of lambda should approach 1.0.

Supplemental Table 4: Effect of exclusions on lambda	
Lambda	**Sample set**
1.111	all subjects
1.109	drop 38 subjects with high relatedness
1.050	drop 140 principal component (PC) outliers
1.047	drop 133 nearest neighbour (NN) outliers
1.046	drop 215 PC or NN outliers
1.053	drop 58 PC and NN outliers
1.046	**final drop 95 (58 PC & NN plus 38 related, 1 overlap)**

Supplemental Table 4 shows the effect on lambda for different exclusion possibilities. Again, lambda was computed blind to SNP identities. The overall lambda for the sample shown in Supplemental Figure 3 was elevated ($\lambda = 1.111$). Removing 38 subjects with high relatedness led to a small decrease ($\lambda = 1.109$). Dropping 140 subjects who were outliers in Supplemental Figure 3 dropped lambda considerably ($\lambda = 1.050$) as did removing 133 individuals with extreme nearest neighbour scores ($|Z| > 3$, $\lambda = 1.047$). Interestingly, these two ways of assessing stratification effects identified substantially different groups of subjects – the union of the two approaches had $\lambda = 1.046$ but required dropping 215 subjects whereas the intersection did nearly as well ($\lambda = 1.053$) but required dropping only 58 subjects.

Therefore, the final subject list was obtained by dropping 95 individuals who were either highly related (N=38) or were identified as an outlier on by both principal components and nearest neighbour analyses (N=58, one overlap). The final lambda was acceptably low (λ = 1.046), and similar to that in other GWAS reports in EUR samples [28, 29].

After these exclusions and removing quality control and trio samples, there were 3,540 subjects in the final analysis (95.2% of the 3,720 initial case-control samples) including 1,738 cases (93.4%) and 1,802 controls (96.9%). The final list of subjects is included as a ***Supplemental File*** in linkage format ("mddC.fam").

Quality Control – SNPs

The initial SNP quality control work was conducted by the GAIN Statistical analysis group (chaired by Dr Gonçalo Abecasis of the University of Michigan) prior to the data release and additional work was done subsequently. The overall intention of SNP quality control is to remove SNPs whose genotypes were untrustworthy in order to avoid false positive and misleading results due to low confidence genotyping. We began with the unfiltered GWAS data obtained from dbGaP and (as noted above) applied a quality score passing threshold of ≥10 to all genotype data.

Perlegen delivered 599,164 SNPs and 8 SNPs with duplicate rs numbers were deleted and 73 mitochondrial SNPs were removed for later analysis, leaving 599,083 SNPs. All SNP locations are per NCBI Build 35/UCSC hg17 and the TAMAL database created by PFS [30]. There were no positional duplicates.

Supplemental Table 5 shows the SNP filters applied to the data and the number of SNPs that failed each filter. Any SNP can be excluded for multiple reasons. SNPs failed quality control for any of the following reasons: gross mapping problems (i.e., mapped to multiple genomic locations, to a random or unknown chromosome, or had very different coordinates in hg17 and hg18; ≥2 genotype disagreements in the duplicated samples genotyped by Perlegen; ≥2

Mendelian errors in the trios genotyped by Perlegen; minor allele frequency < 0.01; or > 0.05 missing genotypes in either cases or controls. For chrY SNPs, only the MAF filter was used. A total of 432,889 SNPs passed all quality control filters on chr1-chr22, chrX, chrY, and chrXY (PAR1). The final list of SNPs is included as a ***Supplemental File*** ("mddC.bim").

Supplemental Table 5		
Step	SNPs	Proportion
Total SNPs	599,083	1.000
Mapping error	(1,487)	(0.002)
Duplicates, ≥ 2 disagreements	(1,143)	(0.002)
Trios, ≥ 2 Mendel errors	(536)	(0.001)
MAF < 0.01	(41,495)	(0.069)
Missing > 0.05 in cases or in controls	(154,673)	(0.258)
SNPs passing all QC steps (chrY only MAF)	435,291	0.727

Supplemental Table 6 shows the inter-relationship between the SNP failure flags. Many SNPs failed for more than one reason.

Supplemental Table 6						
SNP_QC	flag_mapping	flag_dupdis	flag_menerr	flag_miss	flag_maf	SNPs
Pass	0	0	0	0	0	435,291
Fail	0	0	0	1	0	120,810
Fail	0	0	0	1	1	33,891
Fail	0	0	0	0	1	5,928
Fail	1	0	0	0	0	1,484
Fail	0	1	0	0	1	1,140
Fail	0	0	1	0	1	536
Fail	1	1	0	0	0	3

After deleting the individuals in the "final drop" list and considering the 435,291 SNPs that passed quality control, we searched for multivariate patterns of SNP features (i.e., missingness, MAF, and minor allele base) associated with higher lambda values. No meaningful patterns were identified.

We chose not to use Hardy-Weinberg disequilibrium exact p-values [31] as a SNP filter for theoretical and practical reasons. There are ample grounds from the literature to suggest that its use is problematical [32-35]. Practically, 53,994 SNPs had HWD p < 1e-5 in controls and 95.6% of these SNPs were excluded for some other reason (e.g., high missingness or low MAF). We thus

included 2,402 SNPs that failed the HWD filter and used these data as part of the process by which to evaluate a SNP-phenotype association. Inclusion of these 2,402 SNPs had a trivial effect on the overall lambda.

Plate Effects

Perlegen returned a code that contained the plate to which each subject was assigned. This variable has 40 levels, equal to the plate assignments described above. We investigated plate effects on a number of outcomes as shown in ***Supplemental Table 7***.

Supplemental Table 7: Assessment of plate effects

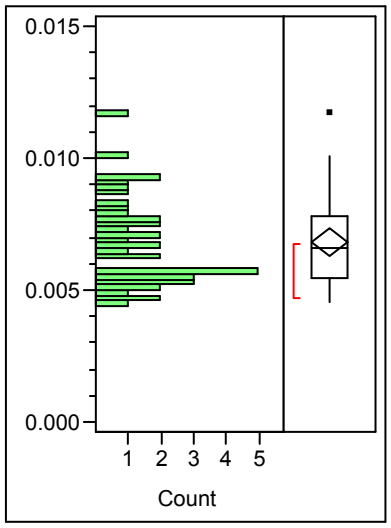
Variable	Test	P-value	R^2
Case/control status	$\chi^2_{39} = 34.4$	0.67	-
Sex	$\chi^2_{39} = 41.3$	0.37	-
Age	$F^{39}_{3500} = 1.05$	0.38	-
Age of onset of MDD	$F^{39}_{1562} = 1.28$	0.12	-
SNP missingness per subject	$F^{39}_{3500} = 12.2$	<0.0001	0.119
PC1	$F^{39}_{3500} = 0.91$	0.64	-
PC2	$F^{39}_{3500} = 1.07$	0.35	-
PC3	$F^{39}_{3500} = 0.85$	0.74	-
PC4	$F^{39}_{3500} = 1.54$	0.02	0.017
PC5	$F^{39}_{3500} = 1.20$	0.19	-
PC6	$F^{39}_{3500} = 0.66$	0.95	-
PC7	$F^{39}_{3500} = 1.54$	0.02	0.017
PC8	$F^{39}_{3500} = 0.96$	0.54	-
PC9	$F^{39}_{3500} = 0.97$	0.53	-
PC10	$F^{39}_{3500} = 0.79$	0.82	-

First, it was evident that the explicit randomization of cases and controls to plates was successful given the absence of significant associations of the plate variable with case-control status, sex, age, and age of onset of MDD. Second, SNP missingness was significantly associated with plate status ($p<0.0001$, $R^2=0.119$). Third, most of the first 10 principal components [26] were not significantly associated with plate assignments with the possible exceptions of PC4 and PC7 which were each associated at the p=0.02 level and may have been due to chance.

The most important association in Supplemental Table 7 was that 11.9% of the variance in SNP missingness (per subject) was explained by plate assignment. Therefore, we explored the plate-SNP missingness per subject association in more detail given its potential importance.

There was no obvious temporal trend in SNP missingness per subject. ***Supplemental Table 8*** shows a histogram of the median of the proportion of missing SNPs per subject for each of the 40 plates. The histogram is reassuring in that the distribution of missingness is approximately continuous over a three-fold range (0.0045 – 0.0117) with no clear evidence of bi-modality.

Supplemental Table 8: Effect of exclusions on lambda

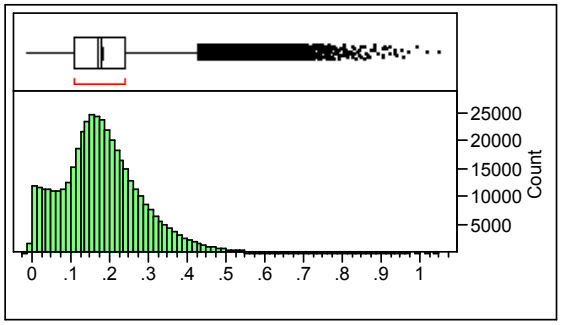
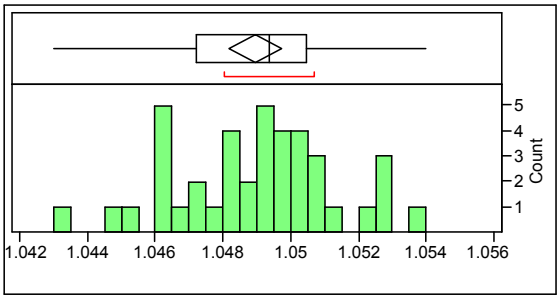


Quantiles		
100.0%	maximum	0.01167
90.0%		0.00931
75.0%	quartile	0.00782
50.0%	median	0.00660
25.0%	quartile	0.00549
10.0%		0.00485
0.0%	minimum	0.00453

To ascertain the practical consequences of plate effects, we reanalyzed the GWAS data using a leave-one-out approach. This required 40 GWAS analyses after removing each plate once (i.e., analysis 1 included plates 2-40, analysis 2 included plates 1, 3-40, etc). PLINK was used to compute estimates of MAF, individual missingness, and case-control association. These data were then analyzed in two ways as shown in ***Supplemental Table 9***. The left panel shows a histogram of SNP-specific plate bias – for each SNP, this equaled the minimum $-\log_{10}(p_{1..40})$ across the 40 p-values from the leave-one-plate-out procedure minus the $-\log_{10}(p_{all})$ for entire sample. Thus, a bias of 2.0 meant that leaving a specific plate out led to a p-value for that SNP was 100 times larger than when that plate was included (i.e., the plate left out was responsible for a marked increase in the association evidence). The median bias was 0.17, the 99.5th percentile was 0.54, and the maximum was 1.05. There were outliers, but such SNPs were rare. These data suggest that plate effects did not have large impacts on individual SNPs

associations. The right panel shows a histogram for the leave-one-plate-out lambdas. The distribution is relatively continuous with no evidence of anti-conservative outliers for higher values of lambda.

Supplemental Table 9: Leave-one-out plate analyses.

SNP-specific bias	Plate-specific lambdas
	

Our major concern in investigating these effects was that one or a few plates might be driving the overall associations. We did not find strong evidence in support of this. However, these plate-wise metrics collapsed across 435,291 SNPs could miss more subtle effects that were operative for a small number of SNPs. Therefore, we saved these data for use as a component in evaluating noteworthy associations.

(We performed analyses analogous to the leave-one-plate-out approach above that included only one plate at a time and obtained similar results, data not shown.)

Evaluation of Noteworthy Associations

For noteworthy associations, there were additional checks to ensure that the association was not due to experimental bias. These checks included: manual inspection of SNP cluster plots to ensure reasonable performance of the genotyping calling algorithm; evaluation of conformation to Hardy-Weinberg equilibrium in controls, cases, and overall (see above for discussion); checks for population stratification effects; evaluation of plate-specific association results to ensure that the overall association was not driven by one plate; and evaluation of the

characteristics of a SNP in high linkage disequilibrium ("proxy association") as a similar association with such a SNP decreases the chance of some forms of method artifacts.

Quantile -Quantile Plots

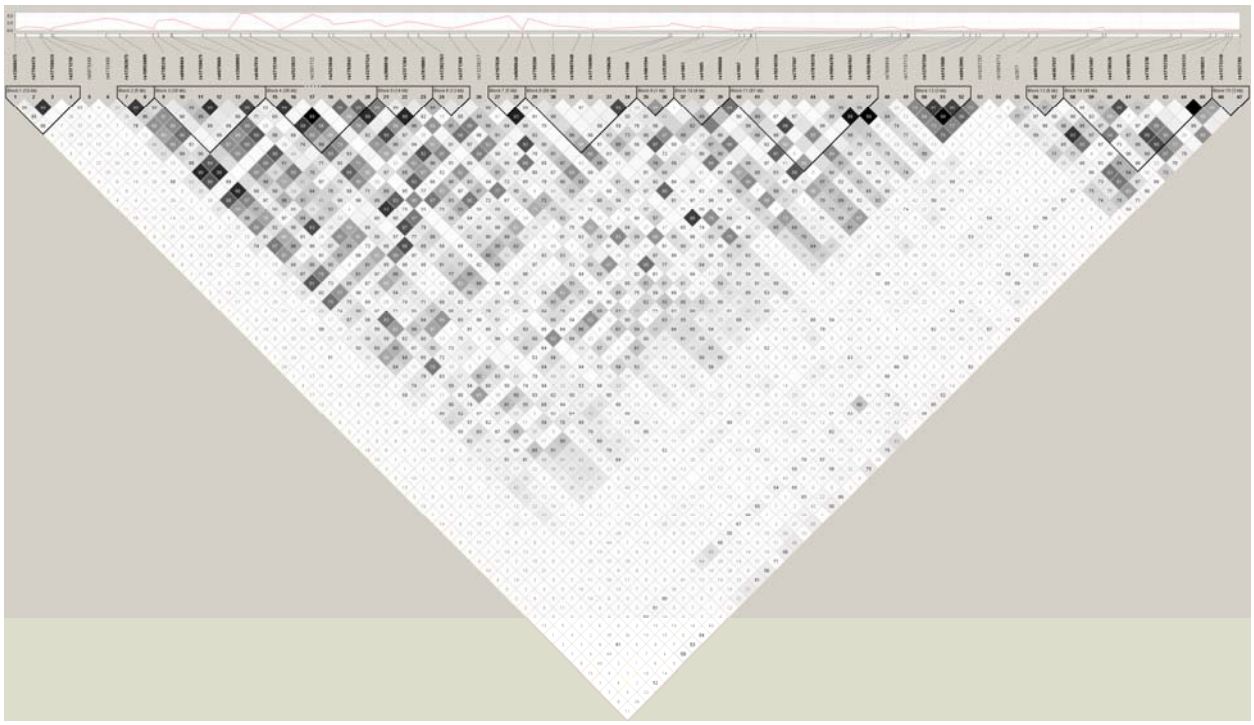
Quantile-quantile plots [36] were created to contrast the observed p-values, $-\log_{10}(p)$, versus those expected by chance, $-\log_{10}\left(\frac{i}{L+1}\right)$, where p is the asymptotic p-value from the Cochran-Armitage trend test, L is the number of SNPs, and i is the rank for each SNP p-value (1=smallest, L=largest). Expected 95% probability intervals for ordered p-values [37] were calculated and included on the plots for reference.

FOCUSING ON THE PCLO REGION

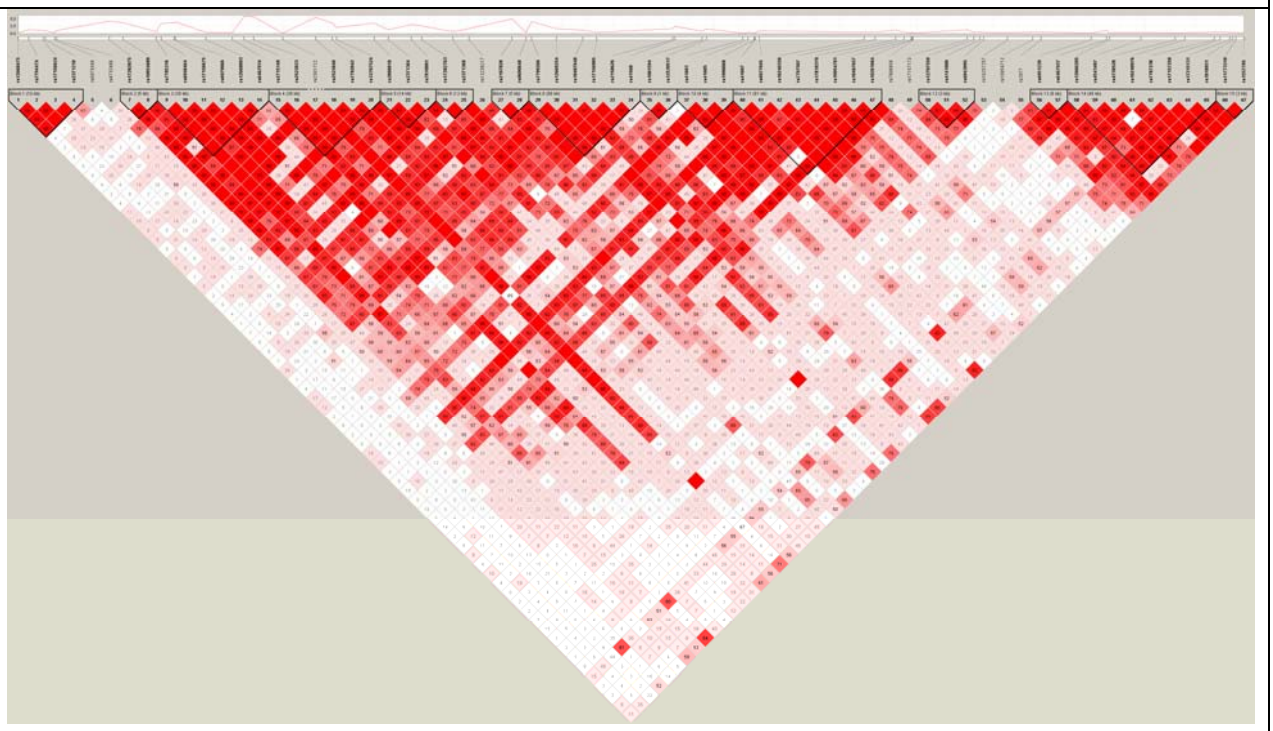
Investigation of Bias.

Supplemental Figure 10 shows the LD patterns for the *PCLO* region (r^2 in top panel and |D'| in lower panel). The region has relatively broad region with little recombination although few of the SNPs are in high LD.

Supplemental Figure 10: Patterns of LD in the *PCLO* region.
r^2



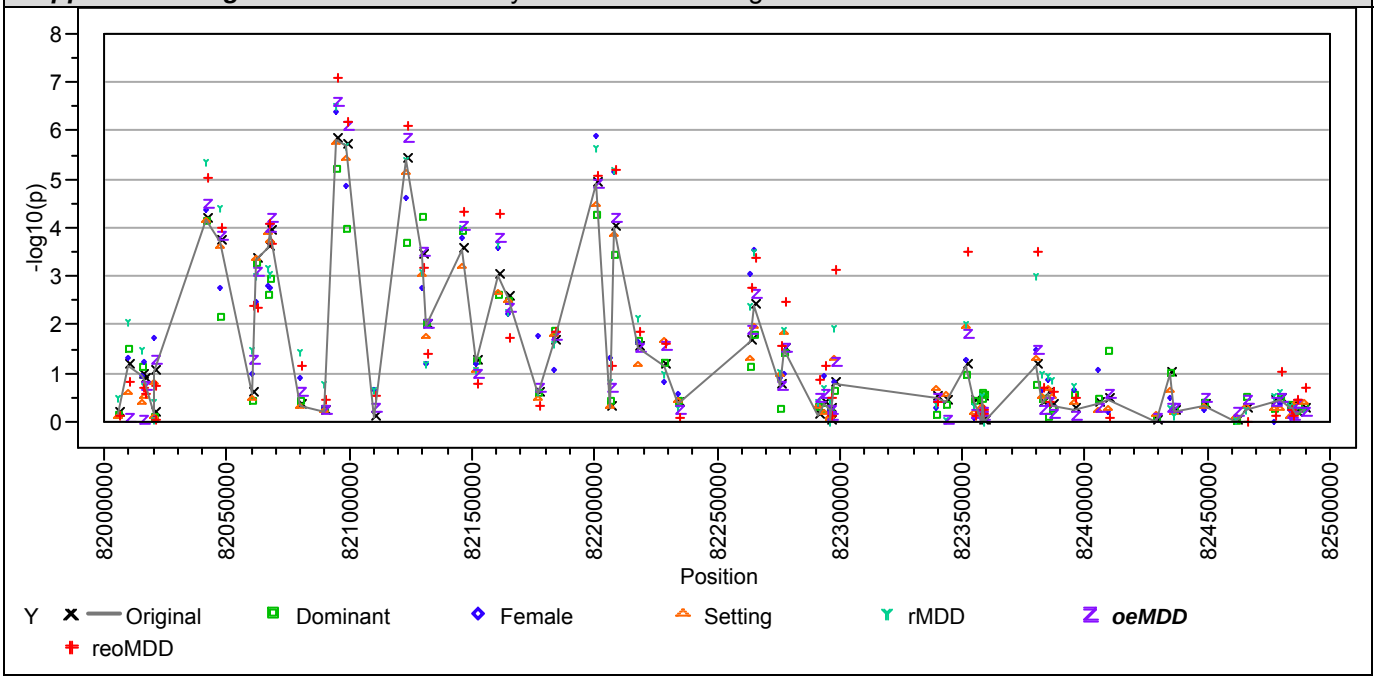
|D'|



Supplemental Figure 11 shows secondary association analyses. These were chosen to expand upon the genetic model used as well as to focus on subtypes of MDD that might be

more likely to have a genetic etiology. The additional genetic model was dominant (green square). The sample was then limited to only females (blue diamonds). We next excluded 423 cases from the original sample who were ascertained from the general population (i.e., limiting cases to those from primary care and mental health organizations, orange triangle). We then excluded 925 cases with single episodes of MDD (i.e., leaving only cases with recurrent MDD, rMDD, green "Y"). Next, we excluded 710 cases with onset of MDD > 30 years of age (i.e., focusing on cases with early-onset MDD, eoMDD, purple "Z"). Finally, we excluded all cases except for the 535 with recurrent, early-onset MDD (reoMDD, red plus). Association results incorporating all cases ("Original", connected black "X") are shown for reference. In all of these subset analyses, we have taken care to ensure that the reference alleles and the directions of association are the same as in the full sample analyses.

Supplemental Figure 11: Additional analyses in the *PCLO* region.



The results of the secondary analyses were generally similar to the primary analysis. For this region, the clinical setting did not make a marked different nor did analysis under a dominant model or limiting the analyses to females. Of interest, the findings for reoMDD –

arguably the most heritable form of MDD [38, 39] – were often more significant than in the primary analyses, particularly for the most significant SNP (rs2715148) where the p-value decreased by 1.2 orders of magnitude to 9.5×10^{-8}.

Replication Studies in the PCLO Region

Replication Strategy & Sample Size Calculation.

The overarching goal in the replication analysis was to have sufficient power to detect a true effect should one exist in the *PCLO* region. This requires greater power than is typical for psychiatric genetics studies (i.e., 95% power for replication rather than the typical 80%) along with recognizing that the observed effect size likely over-estimates the true population effect size (the so-called "winner's curse) [40].

We estimated the necessary replication study sample size using Quanto [41] assuming a log-additive genetic model, MAF=0.45 (similar to rs2522833), MDD population risk of 0.15, a genotypic relative risk of 1.14 ("shrunk" down from the observed GRR of 1.26 for rs2522833 as estimated by the bootstrap method of Sun and Bull) [42], and a two-tailed type 1 error rate of 0.00167 (=0.05/30 replication SNPs). For a total replication set of 6,079 cases and 5,893 controls, statistical power was 97.2% for replication.

Samples.

We were able to identify and to collaborate with seven groups with high quality diagnoses of MDD. These samples all consisted of adult subjects of western European ancestry on whom MDD had been diagnosed according to DSM-IV [43] criteria and the use of a structured clinical interview. Exclusion criteria specified the removal of subjects with MDD-like symptoms that were more parsimoniously due to another psychiatric disorder (bipolar I disorder, schizophrenia, or drug/alcohol dependence) or that were the consequence of a general medical condition (e.g., hypothyroidism). The replication samples were:

- Bonn/Mannheim: Cases for the Bonn/Mannheim MDD study were ascertained from consecutive admissions to the inpatient units of the Department of Psychiatry and Psychotherapy at the University of Bonn and at the Central Institute for Mental Health in Mannheim, Germany. DSM-IV lifetime diagnoses of MDD [43] were made by a consensus best-estimate procedure based on all available information, including a structured interview (SCID-I) [44], medical records, and the family history method. Controls were population based and ascertained from the Bonn and Mannheim environs and subjects with a lifetime history of MDD were excluded.
- Münster: Patients for the Münster MDD study were ascertained from consecutive admissions for inpatient treatment at the Department of Psychiatry, University of Münster, Germany, between 2004 and 2006. Lifetime diagnoses of MDD were established with a structured clinical interview (SCID-I) [44] according to DSM-IV criteria [43]. Patients with schizoaffective disorder or comorbid substance abuse disorders, mental retardation, pregnancy, neurological/neurodegenerative disorders or other clinically unstable medical illnesses impairing psychiatric evaluation were not included. In order to minimize the risk of ethnic stratification, Caucasian descent was ascertained by Caucasian background of both parents. Approval of the ethics committee of the University of Münster, Münster, Germany, and written informed consent from all subjects were obtained. Parts of the present sample have already been described in published studies targeting other gene systems [45-47]. No controls were available.
- Max Planck Institute-Psychiatry Munich (MPIP): Patients with recurrent MDD [48] were recruited from in- and out-patient clinics at the MPIP in Munich and psychiatric hospitals in Augsburg and Ingolstadt, located close to Munich. Patients were diagnosed according to DSM-IV [43] using the Schedule for Clinical Assessment in Neuropsychiatry (SCAN) [49]. Only Caucasian patients over 18 years old with at least two moderate to severe depressive episodes were included. Exclusion criteria were the presence of manic or hypomanic

episodes, mood incongruent psychotic symptoms, the presence of a lifetime diagnosis of intravenous drug abuse and depressive symptoms only secondary to alcohol or substance abuse or dependence, or to a medical illness or medication. Controls were selected randomly from a Munich-based community sample, matched to cases by ancestry, sex, and age, and screened for the presence of anxiety and affective disorders using the Composite International Diagnostic Screener and individuals with MDD were excluded.

- Ancestry was recorded using a self-report sheet for perceived nationality, first language and ethnicity of the subject himself, parents and all four grandparents. All included patients were Caucasian.
- University of Edinburgh (UEDIN): cases with recurrent, severe MDD were recruited from inpatient and outpatient services of Royal Edinburgh Hospital and other psychiatric hospitals in southeast Scotland. All subjects are of European ancestry. The diagnosis of MDD was per DSM-IV criteria [43] (consensus between two psychiatrists) using data from direct interview by a psychiatrist (SADS-L) [50] supplemented by clinical interviews and medical record review. Phenotype data include demography, age at onset, and family history. Controls were of European ancestry and recruited from the Scottish Blood Transfusion Service. Controls were not systematically screened for personal history of psychiatric illness although subjects taking psychotropic medication are not accepted routinely as blood donors.
- Depression Case-Control Study (DeCC): All subjects were interviewed using the Schedules for Clinical Assessment in Neuropsychiatry (SCAN) [49]. Items of psychopathology in the SCAN interview were rated for presence and severity according to the worst and second worst episodes of depression identified by the subjects. For the purposes of rating severity, subjects were asked to identify within each of these episodes of depression the period when their symptoms were at their worst (peak intensity). The computerized version of the SCAN2.1 is built on top of the IShell system, which is a

computer aided personal interviewing tool [51] and which provides diagnoses according to DSM-IV [43]. Further details are provided elsewhere [52]. Controls were ascertained from UK and screened for absence of MDD using the Past History Schedule.

- Sequenced Treatment Alternatives to Relieve Depression (STAR*D): DSM-IV MDD cases [43] were recruited from participants in the STAR*D trial and whom consented separately to give DNA for genetic studies [53]. The STAR*D trial was a large NIMH-sponsored treatment trial involving 4,041 subjects that was designed to assess effectiveness of antidepressant treatments in generalizable samples, and to determine outcomes for outpatients with non-psychotic MDD treated with citalopram[54]. Diagnoses were via the Psychiatric Diagnostic Screening Questionnaire, and depressive symptoms were assessed with the 16-item Quick Inventory of Depressive Symptomatology. Patients with bipolar, psychotic, or obsessive-compulsive disorders were excluded, as were those with primary eating disorders, general medical conditions that contraindicated study medications, substance dependence requiring inpatient detoxification, and clear non-response or intolerance to any protocol antidepressant during current episode or those who were pregnant or breast-feeding. Controls were ascertained from a US population sample as described elsewhere [55]. All subjects included in these analyses were of European ancestry.
- Queensland Institute for Medical Research (QIMR): MDD cases and controls were unrelated individuals from the population-based Australian Twin Registry [56]. Both cases and controls completed either the SSAGA-OZ interview (a version of the Semi-Structured Assessment for the Genetics of Alcoholism instrument modified for Australia) [57] or the shortened Composite International Diagnostic Interview (CIDI) [11], administered by trained interviewers. These comprehensive psychiatric interviews assess the physical, psychological, and social manifestations psychiatric disorders in adults according to DSM-IV criteria. The inclusion criterion was DSM-IV MDD (including impairment) [43] and exclusion criteria were any history of bipolar spectrum disorder, a primary diagnosis of

drug/alcohol dependence, or depressive symptoms secondary to a general medical condition or bereavement. Controls were selected to have below average standardized neuroticism scores (assessed by the revised 12-item Eysenck Personality Questionnaire [58] and standardized from a population cohort of over 18,000) and to be from families in which no family members who had completed the study questionnaires had received a diagnosis of either major depression or of an anxiety disorder.

Finally, in the time since the primary samples were genotyped, 160 additional cases were ascertained in NESDA due to its longitudinal design. The original GWAS cases and controls plus these additional subjects were also genotyped for the same set of 30 SNPs.

Supplemental Table 12 summarizes these studies. The Bonn-Mannheim and Münster samples were combined for analysis given their similar ascertainment, geographical location in central/west Germany, and as combining these samples balanced the numbers of cases and controls.

Supplemental Table 12: Samples genotyped in the follow-up phase.

Feature	NESDA-NTR	Bonn-Mannheim	Münster	MPIP	DeCC	STAR*D	QIMR	UEDIN
Type	Original	Replicate	Replicate	Replicate	Replicate	Replicate	Replicate	Replicate
Location	NL	Germany	Germany	Germany	UK	US	Australia	UK
Design	Cohort	Cross-sectional	Cross-sectional	Cross-sectional	Cross-sectional	Clinical trial	Cohort	Cross-sectional
Genotyping	QIMR	QIMR	QIMR	MPIP	IOP	MGH	QIMR	QIMR
SNPs	30	30	30	30	3	30	30	30
Cases								
Ascertainment	Pop-based & clinical	Clinical	Clinical	Clinical	Clinical	Clinical	Pop-based	Clinical
Ancestry	EUR	EUR	EUR	EUR	EUR	EUR	EUR	EUR
Dx criteria	DSM-IV	DSM-IV	DSM-IV	DSM-IV	DSM-IV	DSM-IV	DSM-IV	DSM-IV
Dx definition	MDD	rMDD	MDD	MDD	rMDD	MDD	MDD	rMDD
Exclusion 1	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Exclusion 2	Multiple	-	-	Incongr psychosis	Incongr psychosis	Multiple	Bereave-ment	IQ < 70
Controls								
Ascertainment	Pop-based	Pop-based	-	Pop-based	Pop-based	Pop-based	Pop-based	Pop-based

Ancestry	EUR	EUR	-	EUR	EUR	EUR	EUR	EUR
MDD removed?	Yes	Yes	-	Yes	Yes	Yes	Yes	Yes
Comparisons								
N Case	1866	777	452	940	1403	1187	966	354
N Controls	1792	1260	0	967	1291	864	1039	472
Proportion female (case, control)	0.67, 0.62	0.65, 0.53	0.59, n/a	0.67, 0.67	0.69, 0.57	0.57, 0.47	0.69, 0.64	0.59, 0.24
Mean age (case, control)	42, 45	48, 48	50, n/a	50, 51	47, 42	43, 52	44, 45	33, 49
Proportion of cases reoMDD	0.31	0.28	0.24	0.36	0.57	0.58	0.22	0.67

Ethical Issues.

All samples were collected under IRB approval from the appropriate local committee. All subjects provided written informed consent that allowed de-identified phenotype and DNA samples to be used in this replication study. All information used in the *PCLO* replication analyses were de-identified and effectively anonymized.

SNP Selection.

30 SNPs were genotyped in the original sample and in six of the seven replication samples (***Supplemental Table 13***). Fifteen of these SNPs were present on the Perlegen array and were analyzed in the primary GWAS and the other 15 SNPs were selected using a multi-marker tagging algorithm operationalized in TAGGER [59]. In one sample (DeCC), three SNPs were genotyped (boldfaced) due to DNA constraints.

Supplemental Table 12: Samples genotyped in the follow-up phase.

SNP ID	Chr.	hg17 position	Genotyped in GWAS?	Minor allele	Major allele
rs7780196	7	82021603	no	G	T
rs17282875	7	82041576	yes	A	G
rs10954689	7	82047024	yes	T	C
rs12672552	7	82058350	no	C	T
rs6948464	7	82061983	yes	T	C
rs13227462	7	82065698	no	C	G
rs17156675	7	82067056	yes	T	C
rs6979066	7	82067588	yes	A	G
rs6965452	7	82073522	no	A	G
rs11771757	7	82074382	no	T	C

rs12668093	7	82079684	yes	T	C
rs6954078	7	82080529	no	G	T
rs2715148	**7**	**82094686**	**yes**	A	C
rs2522833	**7**	**82098359**	**yes**	C	A
rs2522840	7	82123066	yes	G	T
rs13233504	7	82142482	no	T	C
rs2888018	7	82145941	yes	G	T
rs2371364	7	82151525	yes	G	A
rs2371367	7	82163042	no	C	A
rs2189972	7	82169314	no	C	T
rs17235252	7	82180688	no	T	C
rs17809157	7	82192478	no	A	T
rs2107828	**7**	**82200320**	**yes**	A	T
rs10954694	7	82201812	no	C	T
rs10487645	7	82203942	no	C	T
rs9690648	7	82205975	yes	C	T
rs17235831	7	82206612	no	A	G
rs6959723	7	82206991	no	G	A
rs7799260	7	82208167	yes	C	G
rs12669254	7	82217749	yes	C	T

Genotyping & Quality Control.

For 7 of the 8 samples in Supplemental Table 12, a panel of 30 SNPs was selected for genotyping with Sequenom iPlex. This effort was led by Dr. Michael James of QIMR but all investigators participated. Given the broad region of association, half the SNPs selected were genotyped in the original GWAS and the remainder to tag common variation in the region. Primers were designed using SpectroDESIGNER software (Sequenom, San Diego, CA, USA). Polymerase chain reaction was performed followed by homogeneous MassEXTEND reaction. Samples were analyzed in automated mode by a MassARRAY RT mass spectrometer (Sequenom) [60]. The resulting spectra were analyzed by SPECTROTYPER software (Sequenom) after baseline correction and peak identification. Due to sample transfer restrictions, iPlex genotyping was conducted at three sites (QIMR, Munich, and Harvard/MGH) but for the same set of 30 SNPs and on the same platform. At the QIMR site, there were 148 duplicated samples across 15 96-well plates, and there were only 3 genotyping conflicts in 3 different SNPs (genotyping error rate of 0.068%). For one site (Institute of Psychiatry), three

SNPs were genotyped using TaqMan (rs2715148, rs2522833, and rs2107828) using established protocols [61].

All genotype data were imported into PLINK [25] and quality control filters applied (individual missingness < 0.1 and SNP genotype missingness < 0.1). Due to our selection procedure, we did not apply a MAF filter. Exact tests of conformation to Hardy-Weinberg Equilibrium expectations [31] in controls showed no apparent divergence at p<0.001.

Supplemental Table 14 depicts SNP missingness separately per SNP and sample. SNP missingness is quite low (generally < 0.01) for seven samples genotyped using Sequenom iPlex – all SNPs passed for NESDA-NTR, STAR*D, QIMR, and UEDIN and 1 SNP failed for the combined Bonn-Mannhein/Münster sample and 2 failed SNPs failed for MPIP. For the DeCC sample, 3 SNPs were genotyped using TaqMan and 1 failed.

Supplemental Table 14: SNP missingness by SNP and sample.

SNP ID	hg17 position	NESDA-NTR	Bonn-Mannheim + Münster	MPIP	DeCC	STAR*D	QIMR	UEDIN
rs7780196	82,021,603	0	0.003	0		0	0	0
rs17282875	82,041,576	0	0.004	1		0.06	0.001	0
rs10954689	82,047,024	0	0	0.004		0	0	0
rs12672552	82,058,350	0	0	0		0	0	0
rs6948464	82,061,983	0	0.004	0		0	0	0
rs13227462	82,065,698	0	0.002	0.001		0	0	0
rs17156675	82,067,056	0	0	0		0	0	0
rs6979066	82,067,588	0.001	0.006	0.001		0.002	0.011	0
rs6965452	82,073,522	0	0.001	0.002		0	0.002	0
rs11771757	82,074,382	0.004	0.82	0.001		0	0.006	0.001
rs12668093	82,079,684	0	0	0.01		0	0	0
rs6954078	82,080,529	0.007	0.008	0.009		0.007	0.007	0.01
rs2715148	82,094,686	0.002	0.004	0.001	0.039	0.001	0.001	0
rs2522833	82,098,359	0	0.002	0.008	0.005	0	0.001	0
rs2522840	82,123,066	0.001	0.002	0.002		0	0	0
rs13233504	82,142,482	0.001	0.005	0		0	0.001	0
rs2888018	82,145,941	0	0.005	0.002		0.001	0	0
rs2371364	82,151,525	0.001	0.001	0.002		0	0	0
rs2371367	82,163,042	0.001	0.002	0.007		0	0	0
rs2189972	82,169,314	0	0.001	0.001		0	0	0
rs17235252	82,180,688	0	0.001	1		0.002	0	0
rs17809157	82,192,478	0.002	0.003	0.001		0	0.001	0
rs2107828	82,200,320	0	0.001	0.003	0.122	0	0	0
rs10954694	82,201,812	0.001	0.001	0		0	0	0
rs10487645	82,203,942	0	0	0.003		0	0	0

rs9690648	82,205,975	0	0	0		0	0	0
rs17235831	82,206,612	0.002	0	0.002		0	0.001	0.001
rs6959723	82,206,991	0.001	0.003	0.001		0	0.001	0
rs7799260	82,208,167	0	0.001	0.001		0	0	0
rs12669254	82,217,749	0.001	0	0.002		0	0	0

Supplemental Table 15 shows MAF in controls separately for each SNP and sample. All minor alleles are identical. Frequencies in HapMap CEU (EUR) founders are shown for comparison for the25 SNPs that passed the same QC filters as for these samples.

Supplemental Table 15: Minor allele frequencies in controls by SNP and sample.

SNP ID	hg17position	HapMap EUR	NESDA-NTR	Bonn-Mannheim + Münster	MPIP	DeCC	STAR*D	QIMR	UEDIN
rs7780196	82,021,603	*fail*	0.420	0.429	0.399		0.439	0.410	0.424
rs17282875	82,041,576	*fail*	0.471	0.473	*fail*		0.450	0.492	0.471
rs10954689	82,047,024	0.500	0.467	0.458	0.449		0.468	0.442	0.470
rs12672552	82,058,350	0.125	0.119	0.103	0.109		0.111	0.106	0.115
rs6948464	82,061,983	0.367	0.364	0.363	0.357		0.332	0.363	0.339
rs13227462	82,065,698	0.051	0.053	0.057	0.083		0.061	0.053	0.078
rs17156675	82,067,056	*fail*	0.288	0.290	0.278		0.294	0.282	0.304
rs6979066	82,067,588	0.350	0.332	0.342	0.325		0.355	0.328	0.346
rs6965452	82,073,522	0.112	0.145	0.149	0.173		0.159	0.161	0.163
rs11771757	82,074,382	0.050	0.043	*fail*	0.047		0.061	0.045	0.042
rs12668093	82,079,684	0.167	0.160	0.147	0.146		0.157	0.150	0.151
rs6954078	82,080,529	0.483	0.512	0.511	0.528		0.495	0.524	0.504
rs2715148	82,094,686	0.525	0.511	0.510	0.534	0.506	0.477	0.518	0.507
rs2522833	82,098,359	0.425	0.425	0.424	0.408	0.430	0.441	0.414	0.433
rs2522840	82,123,066	0.425	0.430	0.426	0.409		0.445	0.418	0.438
rs13233504	82,142,482	0.183	0.208	0.210	0.193		0.222	0.197	0.202
rs2888018	82,145,941	0.300	0.343	0.341	0.314		0.357	0.324	0.327
rs2371364	82,151,525	0.133	0.111	0.114	0.119		0.122	0.121	0.118
rs2371367	82,163,042	0.283	0.302	0.284	0.291		0.322	0.304	0.289
rs2189972	82,169,314	0.200	0.203	0.191	0.201		0.223	0.213	0.200
rs17235252	82,180,688	0.085	0.089	0.093	*fail*		0.089	0.105	0.085
rs17809157	82,192,478	*fail*	0.118	0.122	0.122		0.097	0.125	0.107
rs2107828	82,200,320	0.500	0.480	0.494	0.506	*fail*	0.450	0.487	0.481
rs10954694	82,201,812	0.358	0.355	0.352	0.347		0.374	0.353	0.371
rs10487645	82,203,942	0.050	0.065	0.065	0.061		0.078	0.073	0.058
rs9690648	82,205,975	0.050	0.066	0.066	0.061		0.082	0.075	0.059
rs17235831	82,206,612	0.108	0.101	0.106	0.106		0.101	0.115	0.099
rs6959723	82,206,991	0.200	0.175	0.184	0.207		0.189	0.192	0.211
rs7799260	82,208,167	0.483	0.477	0.493	0.503		0.449	0.481	0.480
rs12669254	82,217,749	*fail*	0.102	0.092	0.091		0.095	0.093	0.091

Replication Sample Heterogeneity.

We next investigated heterogeneity by studying the case and control samples separately. These analyses included the original sample and the replication cohorts. In order to maximize the number of samples included in PCA (which employs listwise deletion), the DeCC sample was excluded as there were only two analyzable SNPs and the four SNPs with the greatest missingness were also excluded.

Controls. There is substantial evidence that controls are relatively homogeneous for *PCLO* genotypes. First, the median difference between the maximum and minimum MAF across these groups was relatively small 0.033 (IQR 0.025 – 0.043, range 0.011 – 0.061). Second, statistical tests of the number of copies of minor alleles in controls across groups was not significant for 29 of 30 SNPs. The exception was one of the SNPs with the lowest MAF, rs13227462, which had significantly more minor alleles in the MPIP and UEDIN groups. Third, we conducted principal components analysis [26] and neither of the first two principal component was significantly different between groups (ANOVA p=0.44 and 0.06, R^2=8e-4 and 2e-3) and the plot of PC2 x PC1 showed no clustering by sample.

Cases. In contrast, PCA of *PCLO* case genotypes did show evidence for heterogeneity. There were significant differences in the number of copies of minor alleles for most of the SNPs. Moreover, there were significant differences across samples in both the first and second principal components (ANOVA p=0.008 and 4e-7, R2=0.2% and 0.6%). Post-hoc comparisons indicated that NESDA/NTR cases were most similar to cases from QIMR (on both principal components) and to STAR*D cases (first principal component).

SECONDARY ANALYSES

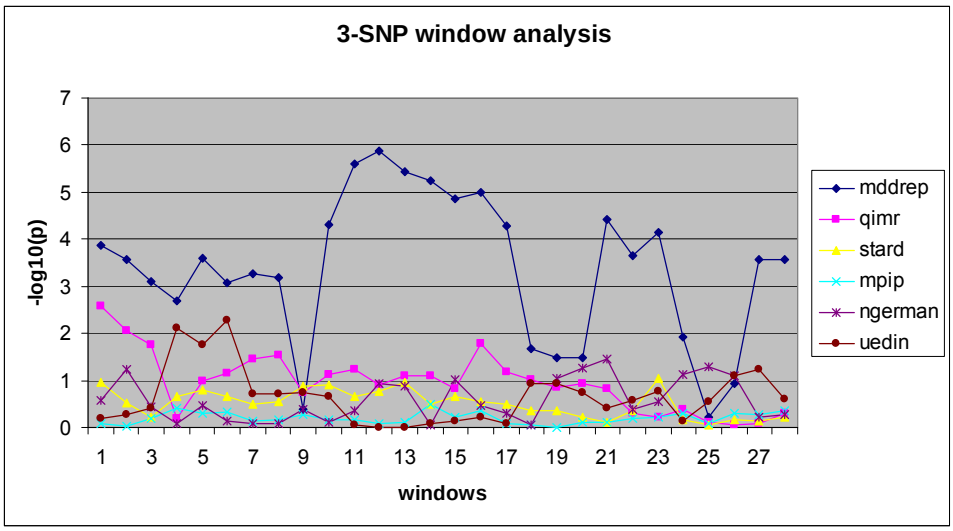
Imputation

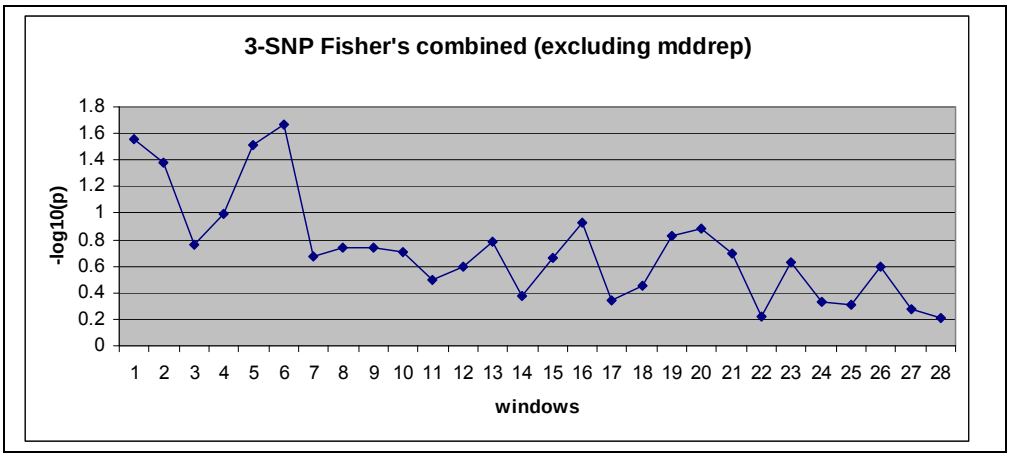
To create the crossover and error model for MACH, we randomly selected 150 cases and 150 controls. Using the HapMap EUR panel as reference, we employed Abecasis' MACH (v1) to impute genotypes for 2,037,829 autosomal SNPs. We used a stringent imputation quality threshold to remove 101,419 SNPs with $R^2 < 0.5$ (a cutoff that removes ~90% of SNPs with unreliable imputation results while sacrificing 2-3% of reliably imputed SNPs). We have demonstrated empirically the similarity of LD patterns in the HapMap EUR panel and in the trio sample from this study (Pardo, in preparation).

PCLO Haplotype Analyses.

Haplotype analyses of *PCLO* replication genotype data using HAPSTAT and three-SNP sliding windows are depicted in ***Supplemental Table 16***. The top panel shows results from the original NESDA/NTR sample and for each of the replication samples. The bottom panel shows the combined analysis of the replication samples.

Supplemental Table 16: Haplotype analyses of PCLO replication genotype data.





MDD Candidate Genes.

This list was culled from a comprehensive review [62] supplemented by PubMed searches.

(a) These genes had zero SNPs in their transcripts: *ADRA2A, ADRB1, APOE, AVPR1B, CCL2, CNTF, CYP2D6, DRD1, DRD4, DRD5, GMIP, GNB3, GPBAR1, GPR50, HP, HTR1A, HTR1B, PENK,* and*. TH*.

(b) These genes had inadequate SNP coverage: *AR*, *FZD3, GABRA3, HTR2C, MAOA, MAOB, PAM, PDE5A,* and *TFCP2*.

(c) MDD candidate genes with minimum p < 0.05.

Supplemental Table 17: Analysis of MDD candidate genes.

Gene	Length	SNPs	Min P-value	Gene Product	Gene size percentile
NOS1	148604	29	0.0005606	nitric oxide synthase 1 (neuronal)	92
DISC1	239711	106	0.001964	disrupted in schizophrenia 1	96
ADCY9	149668	39	0.003151	adenylate cyclase 9	92
FKBP5	66990	9	0.006093	FK506 binding protein 5	81
DRD2	13925	11	0.006189	dopamine receptor D2	47
CHRM2	1401	42	0.009241	cholinergic receptor, muscarinic 2	14
SLC6A2	45652	22	0.00954	solute carrier family 6 member 2	74
GNAS	19103	13	0.009909	GNAS complex locus	54
HTR6	13615	1	0.0121	5-hydroxytryptamine (serotonin) receptor 6	47
CNR1	1419	2	0.01878	central cannabinoid receptor	15

PDE10A	243423	110	0.0198	phosphodiesterase 10A	96
TACR1	149502	29	0.02015	tachykinin receptor 1	92
MTHFR	12437	6	0.0206	5,10-methylenetetrahydrofolate reductase	45
ESR1	291054	50	0.02143	estrogen receptor 1	97
LHPP	151498	49	0.02277	phospholysine phosphohistidine inorganic	93
NPY	6447	2	0.0272	neuropeptide Y	32
OPRM1	79864	56	0.02746	opioid receptor, mu 1 i	84
PLA2G4A	134154	27	0.03478	cytosolic phospholipase A2, group IVA	91
GNAL	191572	26	0.03489	guanine nucleotide binding protein (G protein),	94
ESR2	55453	46	0.03898	estrogen receptor beta	78
NR3C2	355518	70	0.04391	nuclear receptor subfamily 3, group C, member 2	98
IL1B	5869	2	0.04676	interleukin 1, beta proprotein	31
PDE11A	443030	60	0.04789	phosphodiesterase 11A	98
TPH1	19772	4	0.04865	tryptophan hydroxylase 1	55

REFERENCES

1. Konneker T, Barnes T, Furberg H, Losh M, Bulik CM, Sullivan PF. A searchable database of genetic evidence for psychiatric disorders. *American Journal of Medical Genetics (Neuropsychiatric Genetics)* In press.
2. Cloninger CR, Van Eerdewegh P, Goate A, Edenberg HJ, Blangero J, Hesselbrock V, Reich T, Nurnberger J, Jr., Schuckit M, Porjesz B, Crowe R, Rice JP, Foroud T, Przybeck TR, Almasy L, Bucholz K, Wu W, Shears S, Carr K, Crose C, Willig C, Zhao J, Tischfield JA, Li TK, Conneally PM, et al. Anxiety proneness linked to epistatic loci in genome scan of human personality traits. *Am J Med Genet* 1998;81:313-7.
3. Fullerton J, Cubin M, Tiwari H, Wang C, Bomhra A, Davidson S, Miller S, Fairburn CG, Goodwin G, Neale MC, Fiddy S, MMott R, Allison DB, Flint J. Linkage analysis of extremely discordant and concordant sibilng pairs identifies quantiitative-trait loci that influence variation in the human personality trait neuroticism. *American Journal of Human Genetics* 2003;72:879-90.
4. Kuo PH, Neale MC, Riley BP, Patterson DG, Walsh D, Prescott CA, Kendler KS. A genome-wide linkage analysis for the personality trait neuroticism in the Irish affected sib-pair study of alcohol dependence. *Am J Med Genet B Neuropsychiatr Genet* 2007;144B:463-8.
5. Nash MW, Huezo-Diaz P, Williamson RJ, Sterne A, Purcell S, Hoda F, Cherny SS, Abecasis GR, Prince M, Gray JA, Ball D, Asherson P, Mann A, Goldberg D, McGuffin P,

Farmer A, Plomin R, Craig IW, Sham PC. Genome-wide linkage analysis of a composite index of neuroticism and mood-related scales in extreme selected sibships. *Hum Mol Genet* 2004;13:2173-82.

6. Neale BM, Sullivan PF, Kendler KS. A genome scan of neuroticism in nicotine dependent smokers. *Am J Med Genet B Neuropsychiatr Genet* 2004;132B:65-69.
7. Camp NJ, Lowry MR, Richards RL, Plenk AM, Carter C, Hensel CH, Abkevich V, Skolnick MH, Shattuck D, Rowe KG, Hughes DC, Cannon-Albright LA. Genome-wide linkage analyses of extended Utah pedigrees identifies loci that influence recurrent, early-onset major depression and anxiety disorders. *Am J Med Genet B Neuropsychiatr Genet* 2005;135:85-93.
8. Holmans P, Zubenko GS, Crowe RR, DePaulo JR, Jr., Scheftner WA, Weissman MM, Zubenko WN, Boutelle S, Murphy-Eberenz K, MacKinnon D, McInnis MG, Marta DH, Adams P, Knowles JA, Gladis M, Thomas J, Chellis J, Miller E, Levinson DF. Genomewide significant linkage to recurrent, early-onset major depressive disorder on chromosome 15q. *Am J Hum Genet* 2004;74:1154-67.
9. McGuffin P, Knight J, Breen G, Brewster S, Boyd PR, Craddock N, Gill M, Korszun A, Maier W, Middleton L, Mors O, Owen MJ, Perry J, Preisig M, Reich T, Rice J, Rietschel M, Jones L, Sham P, Farmer AE. Whole genome linkage scan of recurrent depressive disorder from the depression network study. *Hum Mol Genet* 2005;14:3337-45.
10. Nurnberger JI, Jr., Foroud T, Flury L, Su J, Meyer ET, Hu K, Crowe R, Edenberg H, Goate A, Bierut L, Reich T, Schuckit M, Reich W. Evidence for a locus on chromosome 1 that influences vulnerability to alcoholism and affective disorder. *Am J Psychiatry* 2001;158:718-24.
11. World Health Organization. Composite International Diagnostic Interview (CIDI), Version 2.1. Geneva, Switzerland: World Health Organization; 1997.
12. Rush AJ, Gullion CM, Basco MR, Jarrett RB, Trivedi MH. The Inventory of Depressive Symptomatology (IDS): psychometric properties. *Psychol Med* 1996;26:477-86.
13. Marks IM, Matthews A. Brief standard self-rating for phobic patients. *Behaviour Research and Therapy* 1979;17:263-67.
14. Beck AT, Epstein N, Brown G, Steer RA. An inventory for measuring clinical anxiety: psychometric properties. *J Consult Clin Psychol* 1988;56:893-7.
15. Costa PT, McCrae RR. The NEO Personality Inventory Manual. Odessa, FL: Psychological Assessment Resources; 1985.

16. Beck AT, Ward CH, Mendelson M, Mock J, Erbaugh J. An inventory for measuring depression. *Archives of General Psychiatry* 1961;4:561-571.
17. Achenbach TM. Young Adult Self Report. Burlington, VT: University of Vermont, Dept of Psychiatry; 1990.
18. Wilde G. Neurotische labiliteit gemeten volgens de vragenlijstmethode (The questionnaire method as a means of measuring neurotic instability). Amsterdam: Van Rossen; 1970.
19. Speilberger C, Gorsuch R, Luchene R. The State-Trait Anxiety Inventory: Test Manual for Form X. Palo Alto, CA: Consulting Psychologists Press; 1970.
20. Middeldorp CM, Birley AJ, Cath DC, Gillespie NA, Willemsen G, Statham DJ, de Geus EJ, Andrews JG, van Dyck R, Beem AL, Sullivan PF, Martin NG, Boomsma DI. Familial clustering of major depression and anxiety disorders in Australian and Dutch twin sibships. *Twin Res Hum Genet* 2005;8:609-15.
21. Bierut LJ, Madden PA, Breslau N, Johnson EO, Hatsukami D, Pomerleau OF, Swan GE, Rutter J, Bertelsen S, Fox L, Fugman D, Goate AM, Hinrichs AL, Konvicka K, Martin NG, Montgomery GW, Saccone NL, Saccone SF, Wang JC, Chase GA, Rice JP, Ballinger DG. Novel genes identified in a high-density genome wide association study for nicotine dependence. *Hum Mol Genet* 2007;16:24-35.
22. Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, Gibbs RA, Belmont JW, Boudreau A, Hardenbol P, Leal SM, Pasternak S, Wheeler DA, Willis TD, Yu F, Yang H, Zeng C, Gao Y, Hu H, Hu W, Li C, Lin W, Liu S, Pan H, Tang X, Wang J, Wang W, Yu J, Zhang B, Zhang Q, Zhao H, Zhou J, Gabriel SB, Barry R, Blumenstiel B, Camargo A, Defelice M, Faggart M, Goyette M, Gupta S, Moore J, Nguyen H, Onofrio RC, Parkin M, Roy J, Stahl E, Winchester E, Ziaugra L, Altshuler D, Shen Y, Yao Z, Huang W, Chu X, He Y, Jin L, Liu Y, Sun W, Wang H, Wang Y, Xiong X, Xu L, Waye MM, Tsui SK, Xue H, Wong JT, Galver LM, Fan JB, Gunderson K, Murray SS, Oliphant AR, Chee MS, Montpetit A, Chagnon F, Ferretti V, Leboeuf M, Olivier JF, Phillips MS, Roumy S, Sallee C, Verner A, Hudson TJ, Kwok PY, Cai D, Koboldt DC, Miller RD, Pawlikowska L, Taillon-Miller P, Xiao M, Tsui LC, Mak W, Song YQ, Tam PK, Nakamura Y, Kawaguchi T, Kitamoto T, Morizono T, Nagashima A, Ohnishi Y, Sekine A, Tanaka T, Tsunoda T, Deloukas P, Bird CP, Delgado M, Dermitzakis ET, Gwilliam R, Hunt S, Morrison J, Powell D, Stranger BE, Whittaker P, Bentley DR, Daly MJ, de Bakker PI, Barrett J, Chretien YR, Maller J, McCarroll S, Patterson N, Pe'er I, Price A, Purcell S, Richter DJ, Sabeti P, Saxena R, Schaffner SF, Sham PC, Varilly P, Stein LD, Krishnan L, Smith AV, Tello-Ruiz MK, Thorisson GA, Chakravarti A, Chen PE, Cutler DJ, Kashuk CS, Lin S, Abecasis GR, Guan W, Li Y,

Munro HM, Qin ZS, Thomas DJ, McVean G, Auton A, Bottolo L, Cardin N, Eyheramendy S, Freeman C, Marchini J, Myers S, Spencer C, Stephens M, Donnelly P, Cardon LR, Clarke G, Evans DM, Morris AP, Weir BS, Mullikin JC, Sherry ST, Feolo M, Skol A, Zhang H, Matsuda I, Fukushima Y, Macer DR, Suda E, Rotimi CN, Adebamowo CA, Ajayi I, Aniagwu T, Marshall PA, Nkwodimmah C, Royal CD, Leppert MF, Dixon M, Peiffer A, Qiu R, Kent A, Kato K, Niikawa N, Adewole IF, Knoppers BM, Foster MW, Clayton EW, Watkin J, Muzny D, Nazareth L, Sodergren E, Weinstock GM, Yakub I, Birren BW, Wilson RK, Fulton LL, Rogers J, Burton J, Carter NP, Clee CM, Griffiths M, Jones MC, McLay K, Plumb RW, Ross MT, Sims SK, Willey DL, Chen Z, Han H, Kang L, Godbout M, Wallenburg JC, L'Archeveque P, Bellemare G, Saeki K, An D, Fu H, Li Q, Wang Z, Wang R, Holden AL, Brooks LD, McEwen JE, Guyer MS, Wang VO, Peterson JL, Shi M, Spiegel J, Sung LM, Zacharia LF, Collins FS, Kennedy K, Jamieson R, Stewart J. A second generation human haplotype map of over 3.1 million SNPs. *Nature* 2007;449:851-61.

23. Hinds DA, Stuve LL, Nilsen GB, Halperin E, Eskin E, Ballinger DG, Frazer KA, Cox DR. Whole-genome patterns of common DNA variation in three human populations. *Science* 2005;307:1072-9.
24. Carlson CS, Eberle MA, Rieder MJ, Yi Q, Kruglyak L, Nickerson DA. Selecting a maximally informative set of single-nucleotide polymorphisms for association analyses using linkage disequilibrium. *Am J Hum Genet* 2004;74:106-20.
25. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira M, Bender D, Maller J, de Bakker P, Daly M, Sham P. PLINK: a toolset for whole-genome association and population-based linkage analysis. *American Journal of Human Genetics* 2007;81:559-75.
26. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics* 2006;38:904-9.
27. Devlin B, Roeder K. Genomic control for association studies. *Biometrics* 1999;55:997-1004.
28. WTCCC. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 2007;447:661-78.
29. Saxena R, Voight BF, Lyssenko V, Burtt NP, de Bakker PI, Chen H, Roix JJ, Kathiresan S, Hirschhorn JN, Daly MJ, Hughes TE, Groop L, Altshuler D, Almgren P, Florez JC, Meyer J, Ardlie K, Bengtsson K, Isomaa B, Lettre G, Lindblad U, Lyon HN, Melander O, Newton-Cheh C, Nilsson P, Orho-Melander M, Rastam L, Speliotes EK, Taskinen MR, Tuomi T, Guiducci C, Berglund A, Carlson J, Gianniny L, Hackett R, Hall L, Holmkvist J,

Laurila E, Sjogren M, Sterner M, Surti A, Svensson M, Tewhey R, Blumenstiel B, Parkin M, Defelice M, Barry R, Brodeur W, Camarata J, Chia N, Fava M, Gibbons J, Handsaker B, Healy C, Nguyen K, Gates C, Sougnez C, Gage D, Nizzari M, Gabriel SB, Chirn GW, Ma Q, Parikh H, Richardson D, Ricke D, Purcell S. Genome-Wide Association Analysis Identifies Loci for Type 2 Diabetes and Triglyceride Levels. *Science* 2007;316:1331-6.

30. Hemminger BM, Saelim B, Sullivan PF. TAMAL: An integrated approach to choosing SNPs for genetic studies of human complex traits. *Bioinformatics* 2006;22:626-7.
31. Wigginton JE, Cutler DJ, Abecasis GR. A note on exact tests of Hardy-Weinberg equilibrium. *Am J Hum Genet* 2005;76:887-3.
32. Wittke-Thompson JK, Pluzhnikov A, Cox NJ. Rational inferences about departures from Hardy-Weinberg equilibrium. *Am J Hum Genet* 2005;76:967-86.
33. Cox DG, Kraft P. Quantification of the power of Hardy-Weinberg equilibrium testing to detect genotyping error. *Hum Hered* 2006;61:10-4.
34. Zou GY, Donner A. The merits of testing Hardy-Weinberg equilibrium in the analysis of unmatched case-control data: a cautionary note. *Ann Hum Genet* 2006;70:923-33.
35. Leal SM. Detection of genotyping errors and pseudo-SNPs via deviations from Hardy-Weinberg equilibrium. *Genet Epidemiol* 2005;29:204-14.
36. Balding DJ. A tutorial on statistical methods for population association studies. *Nat Rev Genet* 2006;7:781-91.
37. Stirling W. Enhancements to aid interpretation of probability plots. *The Statistician* 1982;31:211-220.
38. Sullivan PF, Neale MC, Kendler KS. Genetic epidemiology of major depression: review and meta-analysis. *American Journal of Psychiatry* 2000;157:1552-1562.
39. Levinson DF, Zubenko GS, Crowe RR, DePaulo RJ, Scheftner WS, Weissman MM, Holmans P, Zubenko WN, Boutelle S, Murphy-Eberenz K, MacKinnon D, McInnis MG, Marta DH, Adams P, Sassoon S, Knowles JA, Thomas J, Chellis J. Genetics of recurrent early-onset depression (GenRED): design and preliminary clinical characteristics of a repository sample for genetic linkage studies. *Am J Med Genet B Neuropsychiatr Genet* 2003;119:118-30.
40. Ghosh A, Zou F, Wright FA. Estimating odds ratios in genome scans: an approximate conditional likelihood approach. *Am J Hum Genet* 2008;82:1064-74.
41. Gauderman WJ. Sample size requirements for association studies of gene-gene interaction. *Am J Epidemiol* 2002;155:478-84.

42. Sun L, Bull S. Reduction of selection bias in genomewide genetic studies by resampling. *Genetic Epidemiology* 2005;28:352-67.
43. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders. Fourth Edition ed. Washington, DC: American Psychiatric Association; 1994.
44. Spitzer RL, Williams JBW, Gibbon M, First MB. The Structured Clinical Interview for DSM-III-R (SCID) I: history, rationale, and description. *Archives of General Psychiatry* 1992;49:624-629.
45. Baune BT, Hohoff C, Berger K, Neumann A, Mortensen S, Roehrs T, Deckert J, Arolt V, Domschke K. Association of the COMT val158met variant with antidepressant treatment response in major depression. *Neuropsychopharmacology* 2008;33:924-32.
46. Baune BT, Hohoff C, Roehrs T, Deckert J, Arolt V, Domschke K. Serotonin receptor 1A-1019C/G variant: impact on antidepressant pharmacoresponse in melancholic depression? *Neurosci Lett* 2008;436:111-5.
47. Domschke K, Hohoff C, Mortensen LS, Roehrs T, Deckert J, Arolt V, Baune BT. Monoamine oxidase A variant influences antidepressant treatment response in female patients with Major Depression. *Prog Neuropsychopharmacol Biol Psychiatry* 2008;32:224-8.
48. Lucae S, Salyakina D, Barden N, Harvey M, Gagné B, Labbé M, Binder E, Uhr M, Paez-Pereda M, Sillaber I, Ising M, Brückl T, Lieb R, Holsboer F, Müller-Myhsok B. P2RX7, a gene coding for a purinergic ligand-gated ion channel, is associated with major depressive disorder. *Human Molecular Genetics* 2006;15:2438-45.
49. Wing JK, Babor T, Brugha T, Burke J, Cooper JE, Giel R, Jablenski A, Regier D, Sartorius N. SCAN. Schedules for Clinical Assessment in Neuropsychiatry. *Arch Gen Psychiatry* 1990;47:589-93.
50. Endicott J, Spitzer RL. A diagnostic interview: the Schedule for Affective Disorders and Schizophrenia. *Archives of General Psychiatry* 1978;35:837-844.
51. 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 DSM-IV. 1 ed: World Health Organisation, Geneva.; 2003.
52. Farmer A, Breen G, Brewster S, Craddock N, Gill M, Korszun A, Maier W, Middleton L, Mors O, Owen M, Perry J, Preisig M, Rietschel M, Reich T, Jones L, Jones I, McGuffin P. The Depression Network (DeNT) Study: methodology and sociodemographic characteristics of the first 470 affected sibling pairs from a large multi-site linkage genetic study. *BMC Psychiatry* 2004;4:42.

53. McMahon FJ, Buervenich S, Charney D, Lipsky R, Rush AJ, Wilson AF, Sorant AJ, Papanicolaou GJ, Laje G, Fava M, Trivedi MH, Wisniewski SR, Manji H. Variation in the gene encoding the serotonin 2A receptor is associated with outcome of antidepressant treatment. *Am J Hum Genet* 2006;78:804-14.
54. Rush AJ, Fava M, Wisniewski SR, Lavori PW, Trivedi MH, Sackeim HA, Thase ME, Nierenberg AA, Quitkin FM, Kashner TM, Kupfer DJ, Rosenbaum JF, Alpert J, Stewart JW, McGrath PJ, Biggs MM, Shores-Wilson K, Lebowitz BD, Ritz L, Niederehe G. Sequenced treatment alternatives to relieve depression (STAR*D): rationale and design. *Control Clin Trials* 2004;25:119-42.
55. Sanders A, Duan J, Levinson DF, Shi J, He D, Hou C, Burrell GJ, RIce JP, Nertney DA, Olincy A, Rozic P, Buccola NG, Mowry BJ, Freedman R, Amin F, Black DW, Silverman JM, Byerley WF, Crowe RR, Cloninger CR, Martinez M, Gejman P. No significant association of fourteen candidate genes to schizophrenia in a large European-ancestry sample: Implications for the field. *American Journal of Psychiatry* 2008;165:497-506.
56. Wray N, James M, Mah S, Nelson M, Andrews G, Sullivan PF, Montgomery GW, Birley AJ, Braun A, Martin NG. Anxiety and comorbid measures associated with *PLXNA2*. *Archives of General Psychiatry* 2007;54:318-26.
57. Bucholz KK, Cadoret R, Cloninger CR, Dinwiddie SH, Hesselbrock VM, Nurnberger JI, Jr., Reich T, Schmidt I, Schuckit MA. A new, semi-structured psychiatric interview for use in genetic linkage studies: a report on the reliability of the SSAGA. *J Stud Alcohol* 1994;55:149-58.
58. Eysenck SBG, Eysenck HJ, Barrett P. A revised version of the psychoticism scale. *Personality and Individual Differences* 1985;6:21-29.
59. de Bakker PI, Yelensky R, Pe'er I, Gabriel SB, Daly MJ, Altshuler D. Efficiency and power in genetic association studies. *Nat Genet* 2005;37:1217-23.
60. Buetow KH, Edmonson M, MacDonald R, Clifford R, Yip P, Kelley J, Little DP, Strausberg R, Koester H, Cantor CR, Braun A. High-throughput development and characterization of a genomewide collection of gene-based single nucleotide polymorphism markers by chip-based matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Proc Natl Acad Sci U S A* 2001;98:581-4.
61. Livak KJ. Allelic discrimination using fluorogenic probes and the 5' nuclease assay. *Genet Anal* 1999;14:143-9.

62. Lopez-Leon S, Janssens AC, Gonzalez-Zuloeta Ladd AM, Del-Favero J, Claes SJ, Oostra BA, van Duijn CM. Meta-analyses of genetic studies on major depressive disorder. *Mol Psychiatry* 2007.