Supplementary Material

Table of Contents

Table S1. Genetic marker and quality control data for each sample

Sample	N of Subjects[1] FS / CC	Subject Call Rate	SNP Call Rate	N of SNPs Analyzed [2]	Pre-imputation MAF/HWE	SNP Platform	Software Association / Imputation	Post-imputation R^2/MAF/HWE	Lambda[3] FS/CC
MGS (EA)	2009 / 1336	99%	99%	5.9M	0.01 / 10^{-6}	Affy 6.0	ProbAbel / IMPUTE2	0.3 / 0.01 / 10^{-6}	1.00 / 1.01
PsyCoLaus	2887 / 1955	90%	90%	5.1M	0.01 / 10^{-7}	Affy 500K	in-house / MACH	0.3 / 0.01 / 10^{-6}	1.01 / 1.01
RS	7832 / 5379	97.5%	98%	5.7M	0.01 / 10^{-6}	Illumina 550K, Illumina 610	GRIMP / MACH	0.3 / 0.01 / 10^{-6}	1.02 / 0.99
SHIP	2026 / 1379	92%	80%	6.1M	NA / 10^{-4}	Affy 6.0	Quicktest, R / IMPUTE2	0.3 / 0.01 / 10^{-4}	1.01 / 1.02
QIMR	2277 / 2156	95%	95%	4.5M	0.01 / 10^{-6}	Illumina 317K, 370K, 610K	R / MACH	0.3 / 0.01 / 10^{-6}	1.00 / 1.01
TRAILS	1155 / 614	95%	95%	5.9M	0.01 / 10^{-4}	Illumina Cyto SNP12 v2	SNPtest, Quicktest / IMPUTE2	0.3 / 0.01 / 10^{-3}	1.00 / 1.02
NTR / NESDA	NA / 4491	90%	95%	4.6M	0.01 / 10^{-5}	Affy 6.0, Perlegen-Affy 500K / Illumina 370K, 600K, 1M Omni.	R / MACH	0.3 / 0.005 / 10^{-5}	NA / 1.02

FS, factor score; CC, case-control; MAF, minor allele frequency; HWE, Hardy-Weinberg equilibrium; QC, quality control.

All imputation are based on 1000 Genome reference data (Build 37/HG 19).

For NTR/NESDA, SNPs that failed Mendelian error cutoff were also excluded from the analyses, explaining the lower number of SNPs

[1] Number of subjects used for association analyses for each sample. [2] Number of common SNPs at MAF > 5%. [3] Genomic inflation factor of SNP main effect model after adjusting for covariates (sex, age, principal components).

SAMPLE DESCRIPTIONS

Nine samples containing information on up to five different anxiety disorder (AD) phenotypes from seven independent cohorts were included in the meta-analysis. We note several distinctions among the samples. Three cohorts come from the Netherlands. Only NESDA was clinically ascertained; the remaining were community samples. NESDA/NTR is a hybrid case/control sample combined from two independent Dutch studies initially to investigate the genetics of major depressive disorder (MDD) (1); additional AD cases were added for the current study. Genotyping, QC and imputation for these studies were performed in the same lab. The Rotterdam Study (RS) sample is divided into three cohorts (RS1, RS2, RS3), which were analyzed identically but genotyped on different arrays and separately imputed. MGS is the control sample collected in the US for a schizophrenia case-control study. The QIMR sample is an Australian twin/family sample by design from which we selected one representative member from each family for consistency of analysis with the other samples. SHIP and PsyCoLaus were psychiatric sub-studies of medical research cohorts ascertained in Germany and Switzerland, respectively. There is wide variation in the subjects' ages across samples, ranging from late adolescence (TRAILS) to older adults (RS). We note that lifetime DSM-based AD diagnostic assessments were available for all cohorts except for RS, in which only one-year prevalence was assessed. See the following sections for more detailed descriptions of each sample.

MGS

Subjects

Data were derived from the "control" sample originally part of a large schizophrenia study (Molecular Genetics of Schizophrenia (MGS)). The sample consisted of unrelated subjects selected by random digit dialing from approximately 60 000 US households. They were screened for psychotic and bipolar disorders for use as a comparison group for genetic association studies of these more severe psychiatric phenotypes, but they were not excluded for other common psychiatric disorders seen in the general

population. The full MGS control sample is described in detail elsewhere (2). The data were obtained with permission from dbGaP (Database of Genotypes and Phenotypes, http://www.ncbi.nlm.nih.gov/gap, Study Accessions: phs000021.v3.p2 ("GAIN") and phs000167.v1.p1 "nonGAIN"). Data for the European American (EA) subjects were combined from both the GAIN (n=1442) and nonGAIN (n=1367) datasets. This study was approved by review boards at participating institutions. All subjects gave informed consent to participate in the study after being provided with and receiving an explanation of study protocols and objectives.

Phenotypic Measures

All MGS control subjects completed an online psychiatric screening interview that included the lifetime version of the Composite International Diagnostic Interview, Short Form (CIDI-SF) (3). For those subjects with requisite response data, we applied DSM-based algorithms to the CIDI-SF responses to obtain the following six lifetime clinical phenotypes: MDD, GAD, panic attacks, agoraphobia, social phobia, and specific phobia. We note that only panic attacks, and not panic disorder, could be identified due to limitations in the items included in that section of the CIDI-SF.

Genotype Data

As previously described in detail (4), samples were genotyped at the Broad Institute using the Affymetrix 6.0 array. To improve the overall signal quality, we recalled the genotypes from intensity data using BEAGLECALL (5). There were 2612 EA subjects and 680K autosomal SNPs available after QC procedures performed in PLINK. After phasing with SHAPEIT, genome-wide imputation with 1000 Genomes Project Phase I integrated reference set v3 (March 2012) was conducted using IMPUTE2 (6) under default parameters. Association analyses were conducted with imputed dosage data in ProbAbel (7).

PsyCoLaus

The sample was randomly selected from the residents of the city of Lausanne (Switzerland) in 2003 according to the civil register (CoLaus) (8). Sixty-seven percent of the 35 to 66 year-old subjects who

underwent the physical exam between 2003 and 2006 also accepted the psychiatric evaluation, which resulted in a sample of 3717 individuals (PsyCoLaus), of whom 92% were Caucasians (9). GWAS genotyping data from the Affymetrix 500K SNP array were available for 3419 Caucasian participants of PsyCoLaus. The psychiatric assessment included the semi-structured Diagnostic Interview for Genetic Studies ((DIGS (10), French version (11)). The DIGS was completed with a section on GAD using the questions from the Schedule for Affective Disorders and Schizophrenia - Lifetime and Anxiety disorder version (SADS-LA (12), French version (13)). Similarly, the brief phobia chapter of the DIGS was replaced by the corresponding more extensive chapters from the SADS-LA. Diagnoses were made according to DSM-IV criteria. This research was approved by the local institutional review board. All participants received a detailed description of the goal and funding of the study and signed a written consent.

Rotterdam Study

The Rotterdam Study (RS) is an ongoing cohort study since 1989-1990 among the inhabitants of a district in Rotterdam (Ommoord), and aims to investigate the determinants of chronic diseases in the elderly. In 1990-1993, 7983 persons aged 55 years and older were recruited and re-examined every 3-4 years (RS1). In year 2000-2001, the cohort was expanded by 3,011 additional persons aged 55 and over (RS2). In year 2006-2008, in a second expansion wave, an additional 3932 persons, aged 45 and older, were recruited (RS3) (14). Data on both anxiety and depressive disorders was present for 6800 persons.

Assessment of psychiatric disorders

ADs were diagnosed as part of the home interview. Trained lay interviewers conducted a slightly adapted version of the Munich version of the Composite International Diagnostic Interview (M-CIDI) to assess the following ADs with a computerized diagnostic algorithm according to DSM-IV criteria: GAD, panic disorder with or without history of agoraphobia, agoraphobia, social phobia and specific phobia. The assessments of ADs were implemented to measure one-year prevalence as life time prevalence in the elderly cannot be assessed reliably (15). Obsessive compulsive disorder and post-traumatic stress

disorder were not assessed. The M-CIDI was specifically designed to obtain DSM-IV diagnoses of mental disorders, and test–retest reliability for ADs is good (16) .

Depressive disorders were diagnosed during a home interview (17). During a first home interview, participants were screened for symptoms of depression with the Center for Epidemiological Studies Depression (CES-D) scale. Screen-positive persons (CES-D-score≥16) were invited for a semi-structured clinical interview with the Schedules for Clinical Assessment of Neuropsychiatry (SCAN) (18). This interview was conducted by a trained clinician at the participant's home one week to two months (median time interval: two weeks) after the screening procedure and the anxiety interview. We were able to use the SCAN in this population-based setting because depression can be screened for with high sensitivity (19). With a computerized DSM-IV based diagnostic algorithm, major depression, minor depression and dysthymia during the past month were diagnosed.

Genotyping and Imputation

Genotyping was performed on 550K and 610K arrays (Illumina, Inc., San Diego, California) in 11 496 participants of the Rotterdam Study. The genotyped dataset was restricted to persons who reported that they were from European descent. Ethnic outliers were further excluded using identity-by-state distances >4SD. Duplicates and first-degree or second-degree relatives were excluded using identity-by-state probabilities >97% as well as samples with gender mismatch and excess autosomal heterozygosity. Variants with call rate <95.0%, failing missingness test, Hardy-Weinberg equilibrium (HWE) *p* value $<10^{-6}$, and minor allele frequency (MAF) <1% were also removed. MACH 1.0 software (20) was used to impute to ~30M SNPs based on the 1000 genomes. SNPs included in imputation met the thresholds MAF ≥ 1%, HWE $>10^{-6}$, and SNP call rate ≥ 98.0%.

SHIP

We analyzed data from the Study of Health in Pomerania (SHIP) (21) comprising adult German residents in northeastern Germany. A two-stage stratified cluster sample of adults aged 20-79 years (baseline) was

randomly drawn from local registries. Between 1997 and 2001, 4308 Caucasian subjects participated at baseline. From 2008 to 2012, the third phase of data collection (SHIP-2, *n*=2333) was carried out. In parallel the SHIP-LEGEND study (Life Events and Gene-Environment Interaction in Depression) was carried out for the detailed assessment of life-events and mental disorders (*n*=2400). The ethics committee of the University of Greifswald approved SHIP.

Interview and psychometric data

Sociodemographic factors and medical history were assessed by a computer-assisted face-to-face interview. The diagnosis of mental disorders was assessed using the Munich-Composite International Diagnostic Interview (M-CIDI; (22)) in SHIP-LEGEND. The M-CIDI is a standardized fully structured instrument for assessing psychiatric disorders over the lifespan according to DSM-IV criteria.

Genotyping and Imputation

The SHIP sample was genotyped using the Affymetrix Human SNP Array 6.0. The overall genotyping efficiency of the GWA was 98.55 %. Standard imputation procedures using the software IMPUTE v2.2.2 and all 1000 Genomes data (phase 1 version 3; March 2012).

QIMR

Diagnostic information was collected as part of four studies with overlapping participants conducted at the Queensland Institute of Medical Research (QIMR) in Brisbane, Australia. All studies were approved by the Human Research Ethics Council of QIMR. Between 1980 and 1982, a self-report Health and Lifestyle Questionnaire was sent to all twin pairs in the Australian Twin Registry who were over the age of 18, and hence had been born prior to 1965 (Cohort 1). A second wave of recruitment from the Australian Twin Registry was conducted in 1990 to those twins born between 1964 and 1970 (Cohort 2).

From 1993 to 1995, a telephone-based semi-structured interview designed to assess the physical, psychological and social manifestations of alcoholism and related disorders (Semi-Structured Assessment for the Genetics of Alcoholism - SSAGA) was administered to a subset of the twins from

Cohort 1 (2456 twin pairs and 771 individuals). The SSAGA is based on previously validated questionnaires such as the CIDI and SCID and included diagnostic items for DSM-III-R (subsequently adapted to DSM-IV) for panic disorder, agoraphobia, social phobia and major depressive disorder (MDD). The phone interviews were conducted by trained interviewers. Further information is provided in Mosing et al (23). Between 1996 and 2000, the same instrument was administered to twins in Cohort 2. A total of 2765 twin pairs and 688 single twins participated in the study.

The third study was conducted with the purposes of identifying regions of the genome linked to anxiety and depression using linkage analysis. Twin pairs highly concordant and highly discordant for neuroticism as measured by the Eysenck Personality Questionnaire were selected for participation. In addition, first degree relatives of some twins were included such as parents, children and siblings were included. Full details of the recruitment strategy for this study are given in Kirk et al (24). A shortened version of the CIDI was administered and the sections included allowed for diagnosis based on DSM-IV criteria of: social phobia, agoraphobia, generalized anxiety disorder, panic disorder, obsessive compulsive disorder, and MDD. A total of 2470 individuals participated in the study.

The fourth study was conducted between 2003 and 2007 and combines multiple separate studies focused on the genetics of alcohol and nicotine addiction. All participants were administered the SSAGA instrument as described above. The individual studies are described in Table 2 of Hansell et al (25). 6925 individuals completed the sections from which diagnoses of MDD and ADs could be made.

Across each of the studies, participants were assigned a 0 for MDD and each anxiety disorder if they reported no symptoms of the disorder, a 1 if they reported some symptoms, but did not meet the full DSM-IV criteria for that disorder, and 2 if they met the full disorder criteria. A total of 19 787 individuals had diagnostic information for at least one anxiety disorder or MDD across the 4 studies. Of these, 9386 individuals from 4818 families had genome-wide genotypes. In cases where a participant met the criteria for a disorder in one study, but did not in another, they were treated as being a full case

(given a score of 2). For consistency with the other studies, one representative individual was selected from each family for association analysis.

Participants were genotyped on one of several Illumina genotyping platforms (Illumina 317K, 370K, or 610K). A full description of the quality control and imputation procedure is given elsewhere (26). Briefly, strict quality control was applied to each genotyping project separately (removal of SNPs with missingness > 0.05, HWE p-value < 0.0001, individuals with missingness > 0.02, SNPs with a large number of Mendelian errors of transmission). Principal components were used to identify and remove ancestry outliers. A set of SNPs common to all of the genotyping projects after QC was used for imputation. Phasing for imputation was performed using MACH. Data was imputed to 1000 Genomes Project reference data. Association analyses were performed in R using age, sex, and ancestry PCs.

TRAILS

TRAILS (TRacking Adolescents' Individual Lives Survey) is a prospective population cohort study of Dutch adolescents with bi- or triennial measurements from age 11 to up until adulthood (27). Five assessment waves have been completed to date, which ran from March 2001 to July 2002 (T1), September 2003 to December 2004 (T2), September 2005 to August 2007 (T3), October 2008 to September 2010 (T4), and January 2012 to December 2013 (T5). The study was approved by the Dutch Central Committee on Research Involving Human Subjects. Participants were treated in compliance with the Declaration of Helsinki, and all measurements were carried out with their adequate understanding and written consent. Data for the present study were collected during the fourth assessment wave. At T1, 2230 (pre)adolescents were enrolled in the study (response rate 76%, mean age 11.1, SD 0.6, 51% girls (28). At T4, 84% (N = 1881, mean age 19.1, SD 0.6, 52% girls) participated again, of whom 1584 (84%) received a diagnostic interview. Of these, 614 were included in the case-control analyses, and 1155 in the factor score analyses. DSM-IV generalized anxiety disorder, social phobia, specific phobia, panic disorder, agoraphobia, and major depressive disorder were assessed with the Composite International

Diagnostic Interview (CIDI), version 3.0. Genome-wide genotyping was performed by the Illumina Cyto SNP12 v2 array. This data was imputed using IMPUTE2 (1000 Genomes, March 2012 release) and association analysis was performed with SNPTEST v2.4.1.

NESDA and NTR

The two parent projects that supplied 1521 cases and 2970 controls for this GWAS are large-scale longitudinal studies, the Netherlands Study of Depression and Anxiety (NESDA) and the Netherlands Twin Register (NTR) (1;29;30). Recruitment of participants for NESDA took place from 2004-2007, with three additional follow-up assessments since then. Ascertainment was from outpatient specialist mental health facilities, primary care practices and the general population with some additional cases from the NTR. Inclusion criteria were age 18-65 years, self-reported northwestern European ancestry a lifetime diagnosis of the following DSM-IV anxiety disorders as diagnosed via the Composite International Diagnostic Interview (CIDI) (version 2.1) (31) during one of the NESDA or NTR assessments: generalized anxiety disorder, social phobia, panic disorder and/or agoraphobia. Persons who were not fluent in Dutch and those with a primary diagnosis of a psychotic disorder, obsessive compulsive disorder, bipolar disorder, or severe substance use dependence were excluded.

Control subjects were mainly from the NTR. Longitudinal phenotyping includes assessment of depressive symptoms (via multiple instruments), anxiety, neuroticism, and personality measures. Inclusion for controls required a low score on the trait version of the STAI (State-Trait Anxiety Inventory) or on a composite measure of neuroticism, anxiety and depression (32-35). A subsample of the NTR controls was also screened via a CIDI interview. A small subset of controls was from NESDA and had no lifetime diagnosis of depressive or anxiety disorders as assessed by the CIDI. NESDA and NTR studies were approved by the Central Ethics Committee on Research Involving Human Subjects of the VU University Medical Center, Amsterdam. All subjects provided written informed consent.

Genotyping, Imputation and Genome-Wide Association analysis

Whole blood and/or buccal DNA samples were collected for various NTR and NESDA projects (see (29;30;36;37)). Genotyping was performed on multiple platforms: Affymetrix 6.0, Affymetrix Perlegen 5.0, Illumina 370, Illumina 660, Illumina Omni 1M. Genotype calls were made with platform specific software, i.e., Birdseed, Genotyper and Beadstudio.

QC was then performed within and between the different platforms and all genotypes were lifted over to build HG37 of the human genome. Genotypes that did not properly map to HG37 were removed as well as SNPs with a minor allele frequency below 1%, an allele frequency difference with the reference set above 20%, HWE < 0.00001, or a call rate below 95%.

IBD was calculated for all pairs; samples where IBD did not match the expected pedigree were removed. Cross platform concordance was calculated for samples that were genotyped on multiple platforms; samples showing a concordance < 99% were removed. Imputation was conducted in a two stage approach. First, the genotype platform specific SNPs were imputed using the MaCH (20) software suite. Next, the reference-set SNPs were imputed using Minimach.

The genome-wide association analyses were performed on the dosage genotype data that were transformed into a single additive dosage score per SNP and imported into R. Related subjects or those from non-Western European ancestry were not included in the analyses. Three case-control logistic regression analyses were run, all including 3 principal components and a dichotomous variable coding for study of origin as covariates: 1) without any additional covariates, 2) with sex and age included as covariates, 3) with sex, age and a SNP-by-sex interaction term as covariate. SNPs were selected with an $R^2 > 0.30$, MAF > 0.01 and HWE > 10^{-5}.

CONSTRUCTION OF PHENOTYPIC OUTCOMES FOR ASSOCIATION TESTING

Each study assessed DSM-based criteria for the following six lifetime clinical phenotypes: GAD, panic disorder, agoraphobia, social phobia, specific phobia, and major depressive disorder (MDD). We attempted to apply similar standardized phenotypic criteria in each study, with some exceptions due to limitations of the diagnostic instrument used for some participants in QIMR (see above description). Besides attempting to identify subjects meeting full symptomatic criteria for these disorders ("cases", score = 2), we also sought to differentiate subjects who were highly symptomatic but did not meet full criteria ("sub-syndromal", score = 1) versus those with few or no prior symptoms ("unaffected controls", score = 0). This was operationalized by either (i) keeping the full symptomatic criteria and removing the diagnostic requirements of distress / impairment or (ii) reducing the symptomatic severity or duration. This strategy provides an ordinal, rather than purely binary, scale for input to the factor analyses described below. It also identifies more extreme comparison groups for use in case-control analyses, since diagnostic thresholds are defined for clinical purposes and may not sufficiently differentiate subjects by the risk alleles they carry.

Given prior evidence supporting shared genetic liability across AD phenotypes, we applied and compared two broad phenotypic approaches combining information across the assessed ADs: (1) categorical case-control (CC) comparisons, and (2) quantitative phenotypic factor scores (FS). For the former, we classified AD cases as those scoring "2" for Any AD (full affection status) and controls as "0" (supernormal). Subjects who were affected at a sub-syndromic level (score=1) were excluded from this approach because they may share liability alleles with full cases. We note that the NESDA/NTR sample was designed to approximate the 2/0 designations of our scoring strategy: all of the NESDA subjects we used had at least one AD with or without comorbid depressive disorder, and the NTR controls were selected to possess low liability for internalizing psychopathology. For both approaches, we removed

subjects that had reported a pure mood disorder from all samples. Main manuscript Table 1 lists the resulting numbers of ANX cases and controls for each of the samples.

We also performed multivariate phenotypic analyses in the form of the common factor model (CFM) to estimate an overall latent AD factor score (FS) for each subject as we had done in a previous study (38). This was done for all samples except NESDA/NTR, since that used a case-control design derived from two separate groups. For the remaining community-based samples, the full spectrum of anxiety risk severity and expression is represented in the subjects. We entered the five ordinal AD clinical phenotypes into exploratory factor analyses in Mplus (version 4) (39). Scree plots indicated one dominant component for the inter-correlations among the five ordinal AD variables. Fitting a single CFM returned fit indexes in the conventionally recommended acceptable range. Examination of the pattern of factor loadings across samples suggested a reasonable degree of consistency in how ADs related to and defined a single common AD liability (Supplementary Table S2). Confirmatory factor analysis was applied to estimate a phenotypic FS for use as a quantitative phenotype in association analyses.

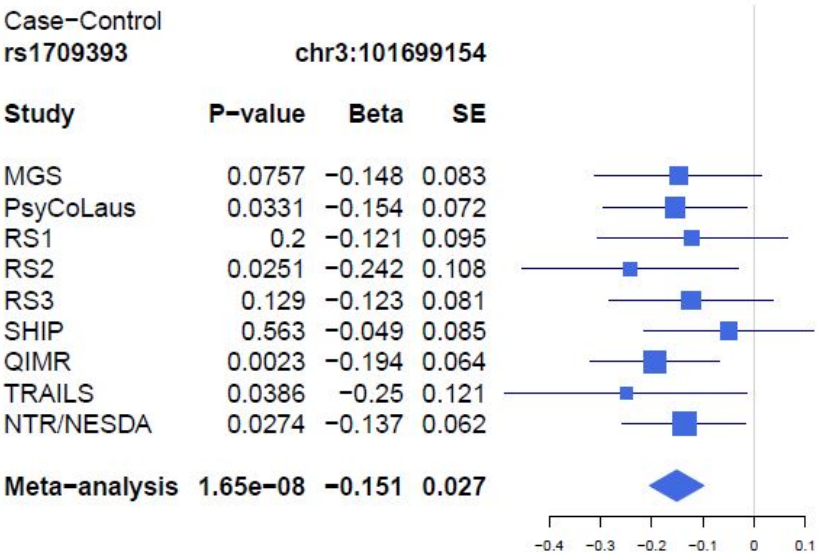
Table S2. Factor loadings by sample for 1-factor solutions

Sample	GAD	PANIC	AGORA	SOC	SP
MGS	0.67	0.60	0.88	0.73	0.59
PsyColaus	0.40	0.79	0.77	0.38	0.39
RS	0.52	0.55	0.81	0.64	0.48
SHIP	0.60	0.74	0.82	0.64	0.53
QIMR	0.54	0.81	0.97	0.55	0.40
TRAILS	0.60	0.59	0.73	0.54	0.51

GAD = generalized anxiety disorder; PANIC = panic disorder; AGORA = agoraphobia; SOC = social phobia; SP = specific phobia.

Figure S1. Forest plots for association effects of the top SNPs by study and meta-analysis.
(A) Association of rs1709393 with case-control phenotype. (B) Association of rs1067327 with factor score phenotype.

A. Effect of rs1709393 in case-control analysis



B. Effect of rs1067327 in factor score analysis

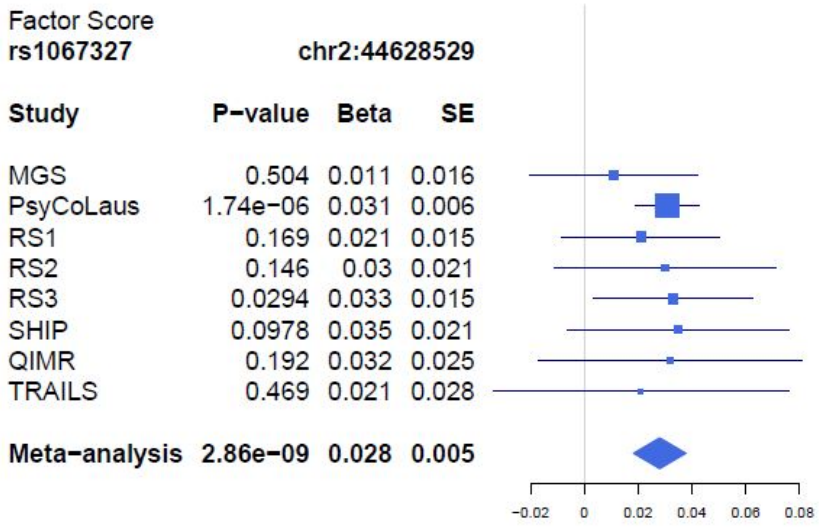
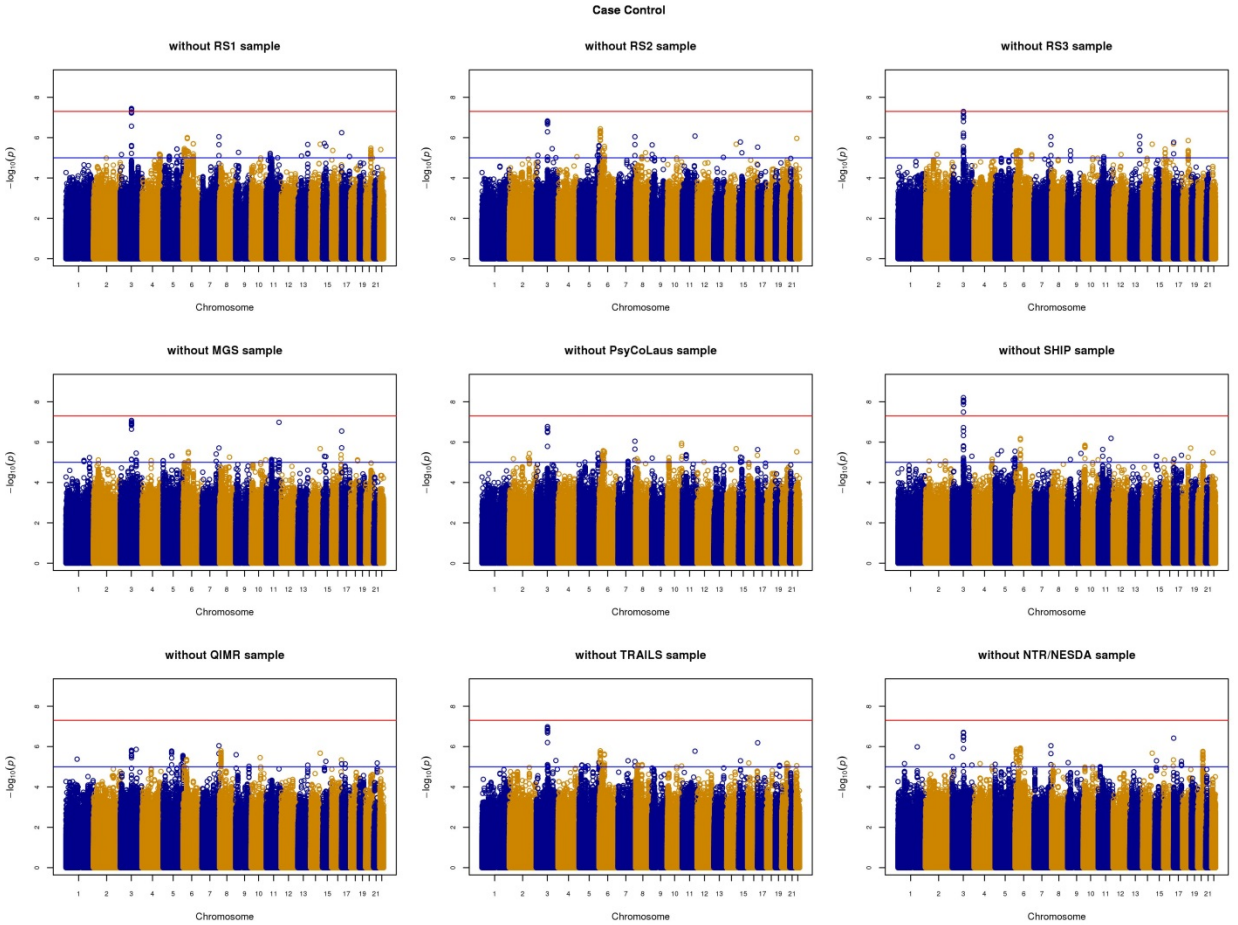


Figure S2. Manhattan plots of meta-analyses with N-1 samples after removing one testing sample in leave-one-out cross validation. Red horizontal line indicates the genome-wide significant P-value $5x10^{-8}$; blue line indicates the suggestive P = $1x10^{-5}$.



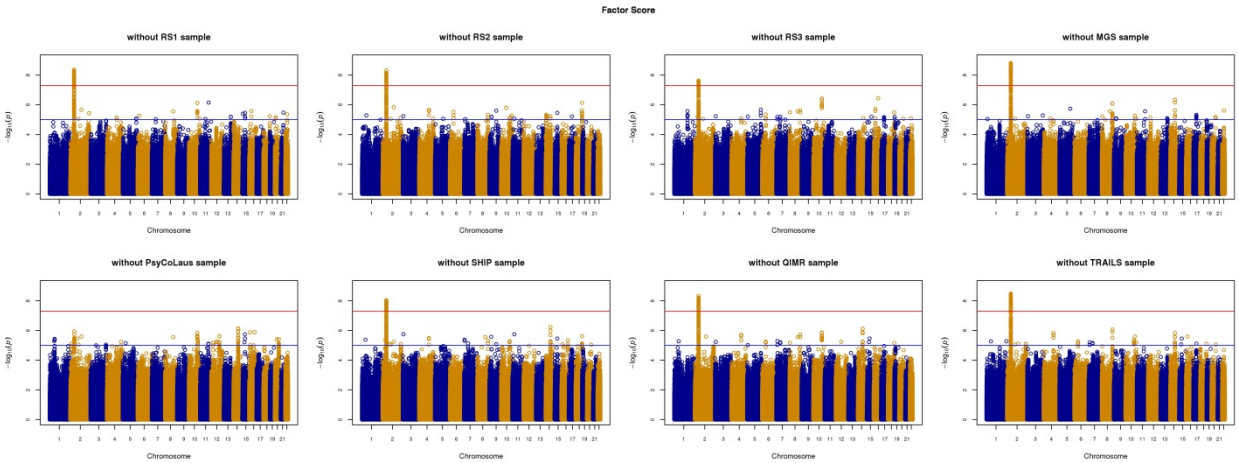
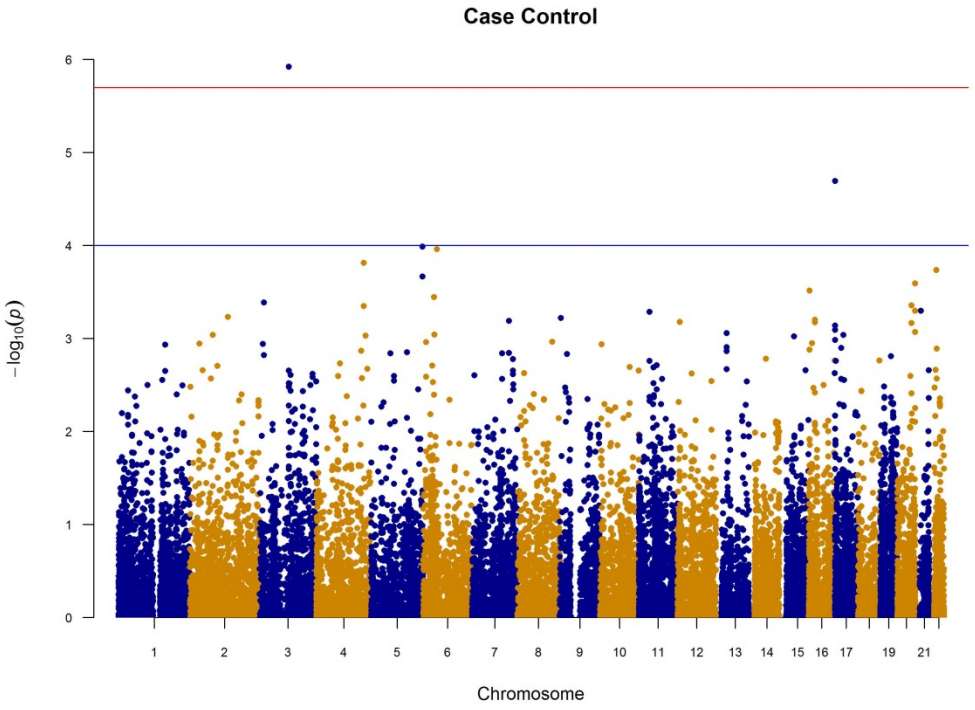


Figure S3. Manhattan plots of gene-based meta-analyses for case-control and factor score phenotypes. Red horizontal line indicates the significant association level of gene-based P-value ($0.05/23931=2x10^{-6}$); blue line indicates the suggestive association level of $P = 1x10^{-4}$.



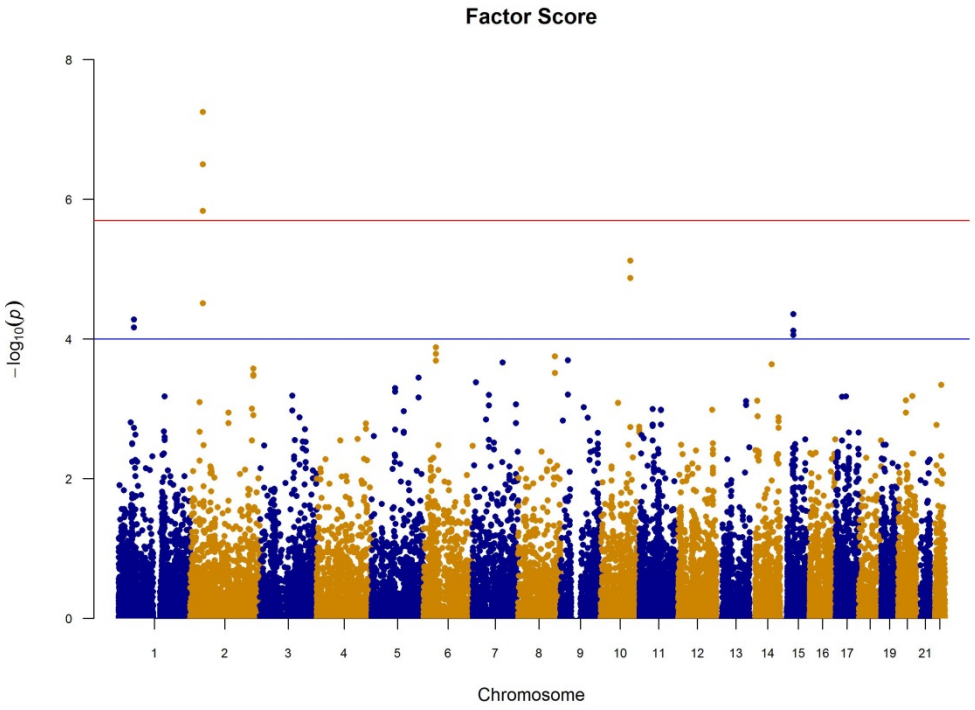


Table S3. Correlation between meta-analysis results for case-control and factor score phenotypes.

P-value threshold in case-control model	N of SNPs (%)	Correlation between effect size	N of SNPs (%) In same direction
P < 0.01	63 381 (1.2)	0.899	62 256 (98.2)
0.01 ≤ P < 0.10	517 581 (9.4)	0.827	480 612 (92.9)
0.10 ≤ P < 0.50	2 217 722 (40.2)	0.639	1 747 248 (78.8)
P ≥ 0.50	2 711 647 (49.2)	0.275	1 615 245 (59.6)
Overall	5 510 331 (100.0)	0.607	3 905 361 (70.9)

SECONDARY ANALYSES

SNP-based Heritability

Genomic-relatedness-matrix restricted maximum likelihood (GREML) was conducted to estimate the total amount of variance explained by all genome-wide SNPs for the combined Rotterdam Study cohort (RS1, RS2, RS3) since that study makes up the largest sample in our collaboration (N=3695 and 5362 individuals for CC and FS phenotypes, respectively). For RS1 and 2, genotyping was done using Illumina HumanHap550v3, whereas for RS3, genotyping was done using Illumina HumanHap 610. Genotyped data for all three cohorts were imputed against 1000 Genomes following stringent QC and combined into one imputed set. The GCTA software package (40) was used to quantify the heritability due to all SNPs after quality control. A genetic relatedness matrix (GRM) is calculated, which estimates the genetic similarity in unrelated individuals (41). Imputed data were filtered on R^2 of 0.5 before constructing the GRM for analyses and using a genetic-relationship cut-off of 0.025. Analyses were adjusted for age, sex and principle components. Results are presented in Table S5, with estimated SNP heritability of 0.071 for CC and 0.106 for FS phenotypes after adjusting for covariates. We note that the magnitude of the standard errors reflects the reduced sample sizes after more stringent QC needed for GREML analysis. When the estimated variance was transformed to the underlying liability scale using the RS sample values of 16% proportion of cases and 10% population disease prevalence, the adjusted CC heritability becomes 0.138 (SE=0.18). While results for FS showed marginal significance, the h^2_{SNP} estimates for CC were not significant (P=0.2).

Table S5. SNP heritability of anxiety disorders in RS sample as estimated by GREML †

Trait	N	σ^2 Genetic (SE)	σ^2 Residual (SE)	σ^2 Phenotypic (SE)	h^2_{SNP} (SE)‡	P-value
Factor Score	5362	0.033 (0.020)	0.283 (0.021)	0.316 (0.006)	0.106 (0.064)	0.05
Case-control	3695	0.010 (0.012)	0.120 (0.012)	0.129 (0.003)	0.071 (0.093)	0.2

†Adjusted for age, sex, cohort, and principal components

‡h^2_{SNP}: heritability (ratio of genetic variance to phenotypic variance)

σ^2: variance; SE: standard error

In addition, we applied LD score regression to the summary statistics to estimate SNP heritability for the combined meta-analytic samples. The LD score regression method is described in detail elsewhere (42). Briefly, when analyzing results from GWAS studies, it is common to find inflation in the distribution of test statistics when compared to the distribution under the null. This inflation can be due to a number of confounding factors such as population stratification, but such inflation is also seen under a scenario where many variants, each with small effects on the phenotype, contribute to the heritability (43). Confounders such as population stratification are not correlated with LD. Therefore, by regressing LD scores (the sum of the r^2 values of a SNP with all of the SNPs in the same region) against the test statistic, one can estimate the relative contribution of confounding factors and polygenecity to the test statistic inflation. The evaluation of (1 - intercept) gives an estimate of the contribution of confounding factors. Using this method, it was shown for a number of large GWAS meta-analyses that confounding factors play a very small role relative to polygenic inheritance in the observed inflation of the mean χ2 value for GWAS.

We used the European Hapmap 3 samples as a reference to calculate the LD scores. For the CC analysis, a population prevalence of 10% was used to transform the estimate of heritability attributable to the SNPs to the liability scale. A total of 995 869 SNPs in common between the meta-analysis results and Hapmap 3 samples were included in the analysis. Heritability on the liability scale was estimated to be 0.095 (SE =0.037) for CC model and 0.072 (SE = 0.028) for FS model. The intercepts from the regressions were 1.002 and 0.998, respectively, indicating that confounding factors do not influence the inflation in the mean χ^2 statistic. The results from GCTA in the RS sample and LD score regression in the entire meta-analytic sample are roughly consistent with each other. We note that these heritability estimates are substantially smaller than that those from twin studies of ADs (range 0.3-0.5) (44).

Polygenic Risk Profile Analyses

Genomic profile risk scores (GPRS) were calculated to test the additive joint effects of multiple variants following the method described elsewhere (45). Briefly, to generate GPRS for each individual in an independent "target" sample, the risk alleles and their effect sizes in a "discovery" sample were used for selected SNPs based on specific p-value thresholds (0.001, 0.01, 0.1, 0.5, and 1). The most strongly associated SNPs were selected into the profile set with a stringent LD threshold ($r^2 < 0.2$ across 500 kb). At each p-value threshold, the variance explained by the GPRS was evaluated by Nagelkerke's R^2 for CC and adjusted R^2 for FS through regression analyses after adjusting for the same covariates as in the association analyses. Since the this approach requires raw genotype data, we conducted these analyses using two individual AD GWAS data sets (QIMR and NTR/NESDA) as target samples and summary data from Psychiatric Genomics Consortium phase 1 (PGC1) schizophrenia (SCZ) (46), bipolar disorder (BIP) (47) and major depressive disorder (MDD) (48) as discovery samples. We note that the choice of AD samples we could include was limited by overlap with PGC data to which some of them contributed. Because of the QIMR PI's specific access to the PGC data, we were able to remove the QIMR data from the PGC samples and also conduct GPRS with those. The PGC summary data were obtained from the PGC download site: http://www.med.unc.edu/pgc/downloads. The results are shown in Table S6 for the CC phenotype. The results for FS phenotype were null. As indicated, GPRS from PGC-MDD explained a small but significant proportion of variance in ADs in QIMR (0.5%-0.7%). PGC-SCZ explained a somewhat smaller proportion of AD variance with marginal statistical significance. Finally, PGC-BIP explained 0.2% and 0.1% AD variance in QIMR and NESDA/NTR, respectively, significantly so only in the latter.

Table S6. Profile scoring results between PGC discovery samples and AD target samples (CC model)

AD Target Sample	Discovery Sample	P Threshold	Variance Explained	P-value
QIMR	PGC-MDD	p < 0.001	0.0002	0.595
		p < 0.01	0.0031	0.025
		p < 0.1	0.0070	0.001
		p < 0.5	0.0052	0.004
		All SNPs	0.0057	0.002
QIMR	PGC-SCZ	p < 0.001	0.0000	0.941
		p < 0.01	0.0007	0.300
		p < 0.1	0.0006	0.315
		p < 0.5	0.0023	0.053
		All SNPs	0.0024	0.048
QIMR	PGC-BIP	p < 0.001	0.0001	0.650
		p < 0.01	0.0000	0.914
		p < 0.1	0.0015	0.114
		p < 0.5	0.0019	0.080
		All SNPs	0.0021	0.065
NTR/NESDA	PGC-BIP	p < 0.001	0.0001	0.602
		p < 0.01	0.0008	0.038
		p < 0.1	0.0012	0.010
		p < 0.5	0.0012	0.011
		All SNPs	0.0011	0.016

Given the limited power available from single samples, we used the LD score regression method to estimate the polygenic SNP correlation between the combined AD meta-analysis CC data and PGC phenotypes. LD score regression has recently been extended to evaluate the genetic correlation between phenotypes (49). In this situation, the product of the z-scores of χ^2 values from two studies are regressed against LD scores, and the slope is an estimate of the genetic correlation between the two phenotypes. Sample overlap only affects the intercept; therefore, using summary statistics from two GWAS that include overlapping samples should not affect the estimate of the genetic correlation. In order to minimize the amount of overlapping samples between PGC disorders, we downloaded the results that were used in the PGC cross-disorder GWAS analysis (50) from http://pgc.unc.edu. Results, including population and sample prevalences used to estimate the genetic correlations, are shown in Table S7.

Table S7. Genetic correlation between ADs and PGC disorders as estimated by LD score regression.

Disorder	Total N	Sample Prevalence	Population Prevalence	Genetic correlation (SE)	P-value	Genetic Covariance (SE)
ADs	17 310	0.33	0.1	NA	NA	NA
MDD	16 610	0.55	0.15	0.68 (0.19)	0.004	0.08 (0.02)
SCZ	17 115	0.5	0.01	0.21 (0.14)	0.14	0.04 (0.03)
BIP	11 810	0.5	0.01	0.16 (0.15)	0.27	0.03 (0.03)

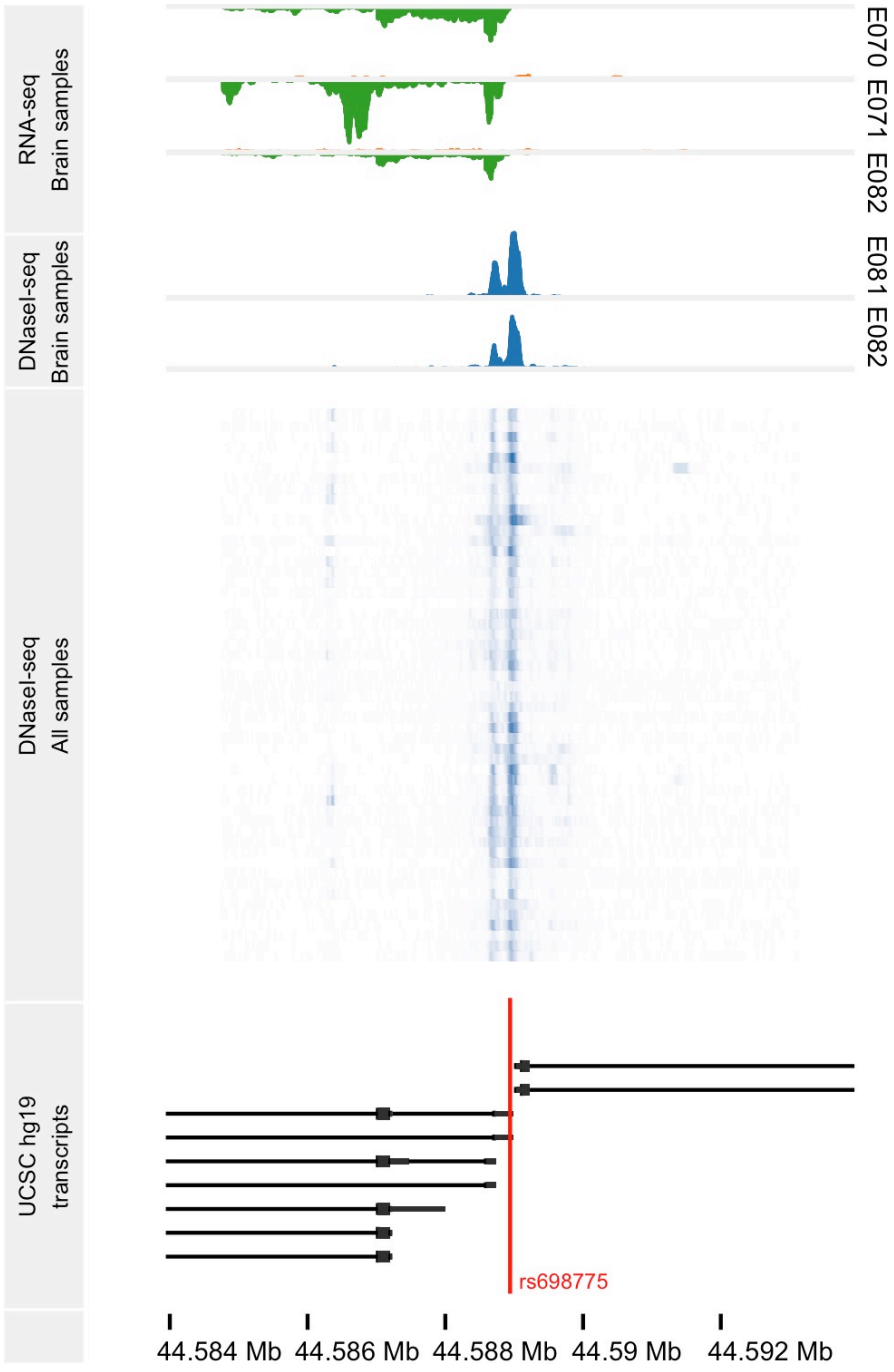
We find significant evidence for genetic overlap between ADs and MDD but not between ADs and BIP or SCZ. These results are consistent with the above results from genetic profile scoring analyses, where profile scores constructed from the results from PGC-MDD explained more variance in anxiety than did scores generated from PGC-BIP or PGC-SCZ.

In Silico Functional Analysis of Association Signals

Phenotype-associated variants are enriched within regions of open chromatin, characterized by DNaseI hypersensitivity (51). As such, we sought to narrow our list of association signals to potentially functional SNPs through the integration of DNaseI hypersensitive regions (DHRs). Epigenetic and RNA sequencing data were generated by the RoadMap Epigenomics Consortium (52) and downloaded from their public web portal ([http://egg2.wustl.edu/roadmap/web_portal]). DHRs were defined using the uniformly processed narrow peak data for consolidated epigenomes (http://egg2.wustl.edu/roadmap/data/byFileType/peaks/consolidated/narrowPeak).

We identified all AD-associated SNPs that overlap a DHR in at least one of the 39 primary human tissues and cell lines assayed by RoadMap. For FS model, of the 101 Chr 2 SNPs with p-values < $1x10^{-6}$, eleven overlapped a DHR present in four or fewer RoadMap samples. The one outlier SNP was rs698775 (Chr 2: 44,588,941), which is located within a DHR consistently identified in all 39 assayed RoadMap samples and located within the 5' UTR of *PREPL* (Supplementary Figure S4). Although rs698775 is also in close proximity to *CAMKMT*, examination of regional expression patterns in the three RoadMap samples for which RNA-seq data is available indicated that expression was only detectable along the negative strand, consistent with the hypothesis that rs698775's *cis* regulatory effect is specific to *PREPL*. No clear overlap was evident for the Chr 3 region of the CC model.

Figure S4. Candidate causal SNP for case-control ADs, rs698775 (red line), is located within the 5′ UTR of PREPL. Heatmap depicts enrichment of DNaseI hypersensitivity across 39 primary tissue and cell line tissues assayed by RoadMap. We specifically examined the behavior of this locus in brain-relevant samples assayed by RoadMap, which allowed us to verify this promoter is active in male and female fetal brain tissue (E081 and E082, respectively) and is associated with active expression of PREPL from the negative strand (green) in the germinal matrix (E070) and adult hippocampus (E071). The lack of detectable expression from the positive strand (orange) suggests rs698775's cis regulatory effect is specific to PREPL.



POWER ANALYSES

We conducted calculations to estimate the power to detect SNP main effects in our meta-analysis samples. Assuming an additive genetic model and complete linkage disequilibrium with the actual risk variant, power was calculated with the Genetic Power Calculator (42) at both genome-wide significant ($P < 5 \times 10^{-8}$) and suggestive thresholds ($P < 2 \times 10^{-5}$). These power calculations indicated that, in CC analyses, the sample had > 80% power to detect genetic variants with MAF ≥ 0.20 and OR ≥ 1.15 at the suggestive threshold and somewhat lower power for the genome-wide significant threshold. Using FS phenotype, the combined N=18 186 provides >70% power to detect common SNPs explaining 0.2% of the phenotypic variance at both genome-wide and suggestive p-values (Table S8).

Table S8. Power analyses for GWAS meta-analysis

a. Power (%) in 5761 cases and 11 765 controls for CC phenotype

	$P < 5 \times 10^{-8}$ (significant) MAF				$P < 2 \times 10^{-5}$ (suggestive) MAF			
OR	0.05	0.10	0.20	0.40	0.05	0.10	0.20	0.40
1.05	0	0	0	0	0	0	1	2
1.10	0	0	2	9	1	5	21	45
1.15	0	6	36	**74**	7	35	**80**	**97**
1.20	4	37	**89**	**99**	30	**80**	**99**	**100**

OR, odds ratio; MAF, minor allele frequency

†Assumptions: 25% lifetime prevalence of cases with any AD, complete linkage disequilibrium (D'=1)

b. Power (%) in combined sample N=18 186 for FS phenotype

	P threshold (significant)	
QTL heritability	5×10^{-8}	2×10^{-5}
0.0001	0	0
0.0003	0	3
0.0005	1	11
0.0010	12	50
0.0020	**72**	**96**

QTL, quantitative trait loci

†Assumptions: complete linkage disequilibrium (D'=1)

Reference List

(1) Boomsma DI, Willemsen G, Sullivan PF, Heutink P, Meijer P, Sondervan D, et al. Genome-wide association of major depression: description of samples for the GAIN Major Depressive Disorder Study: NTR and NESDA biobank projects. Eur J Hum Genet 2008 Mar;16(3):335-42.

(2) Sanders AR, Levinson DF, Duan J, Dennis JM, Li R, Kendler KS, et al. The Internet-based MGS2 control sample: self report of mental illness. Am J Psychiatry 2010 Jul;167(7):854-65.

(3) Kessler RC, Andrews G, Mroczek DK, Ustun B, Wittchen H-U. The World Health Organization Composite International Diagnostic Interview Short Form (CIDI-SF). International Journal of Methods in Psych Res 1998;7:171-85.

(4) Shi J, Levinson DF, Duan J, Sanders AR, Zheng Y, Pe'er I, et al. Common variants on chromosome 6p22.1 are associated with schizophrenia. Nature 2009 Aug 6;460(7256):753-7.

(5) Browning BL, Yu Z. Simultaneous genotype calling and haplotype phasing improves genotype accuracy and reduces false-positive associations for genome-wide association studies. Am J Hum Genet 2009 Dec;85(6):847-61.

(6) Howie B, Fuchsberger C, Stephens M, Marchini J, Abecasis GR. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. Nat Genet 2012 Aug;44(8):955-9.

(7) Aulchenko YS, Struchalin MV, van Duijn CM. ProbABEL package for genome-wide association analysis of imputed data. BMC Bioinformatics 2010;11:134.

(8) 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.

(9) Preisig M, Waeber G, Vollenweider P, Bovet P, Rothen S, Vandeleur C, et al. The PsyCoLaus study: methodology and characteristics of the sample of a population-based survey on psychiatric disorders and their association with genetic and cardiovascular risk factors. BMC Psychiatry 2009;9:9.

(10) Nurnberger JI, Jr., Blehar MC, Kaufmann CA, York-Cooler C, Simpson SG, Harkavy-Friedman J, et al. Diagnostic interview for genetic studies. Rationale, unique features, and training. NIMH Genetics Initiative. Arch Gen Psychiatry 1994 Nov;51(11):849-59.

(11) Preisig M, Fenton BT, Matthey ML, Berney A, Ferrero F. Diagnostic interview for genetic studies (DIGS): inter-rater and test-retest reliability of the French version. Eur Arch Psychiatry Clin Neurosci 1999;249(4):174-9.

(12) Endicott J, Spitzer RL. A diagnostic interview: the schedule for affective disorders and schizophrenia. Arch Gen Psychiatry 1978 Jul;35(7):837-44.

(13) Leboyer M, Maier W, Teherani M, Lichtermann D, d'Amato T, Franke P, et al. The reliability of the SADS-LA in a family study setting. Eur Arch Psychiatry Clin Neurosci 1991;241(3):165-9.

(14) Hofman A, Darwish MS, van Duijn CM, Franco OH, Goedegebure A, Ikram MA, et al. The Rotterdam Study: 2014 objectives and design update. Eur J Epidemiol 2013 Nov;28(11):889-926.

(15) Hek K, Tiemeier H, Newson RS, Luijendijk HJ, Hofman A, Mulder CL. Anxiety disorders and comorbid depression in community dwelling older adults. Int J Methods Psychiatr Res 2011 Sep;20(3):157-68.

(16) Wittchen HU, Lachner G, Wunderlich U, Pfister H. Test-retest reliability of the computerized DSM-IV version of the Munich-Composite International Diagnostic Interview (M-CIDI). Soc Psychiatry Psychiatr Epidemiol 1998 Nov;33(11):568-78.

(17) Luijendijk HJ, van den Berg JF, Dekker MJ, van Tuijl HR, Otte W, Smit F, et al. Incidence and recurrence of late-life depression. Arch Gen Psychiatry 2008 Dec;65(12):1394-401.

(18) Wing JK, Babor T, Brugha T, Burke J, Cooper JE, Giel R, et al. SCAN. Schedules for Clinical Assessment in Neuropsychiatry. Arch Gen Psychiatry 1990 Jun;47(6):589-93.

(19) Beekman AT, Deeg DJ, Van LJ, Braam AW, De Vries MZ, van TW. Criterion validity of the Center for Epidemiologic Studies Depression scale (CES-D): results from a community-based sample of older subjects in The Netherlands. Psychol Med 1997 Jan;27(1):231-5.

(20) Li Y, Willer CJ, Ding J, Scheet P, Abecasis GR. MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. Genet Epidemiol 2010 Dec;34(8):816-34.

(21) Volzke H, Alte D, Schmidt CO, Radke D, Lorbeer R, Friedrich N, et al. Cohort profile: the study of health in Pomerania. Int J Epidemiol 2011 Apr;40(2):294-307.

(22) Wittchen H-U, Pfister H. DIA-X Interviews (M-CIDI): Manual fur Screening-Verfahren und Interview:Interviewheft Langsschnittuntersuchung (DIA-X-Lifetime); Erganzungsheft (DIA-X-Lifetime); Interviewheft Querschnittuntersuchung (DIA-X 12 Monate); Erganzungsheft (DIA-X 12 Monate); PC-Programm zur Durchfuhrung des Interviews(Langs- und Querschnittuntersuchung); Auswertungsprogramm. Frankfurt, Germany: Swets & Zeitlinger; 1997.

(23) Mosing MA, Gordon SD, Medland SE, Statham DJ, Nelson EC, Heath AC, et al. Genetic and environmental influences on the co-morbidity between depression, panic disorder, agoraphobia, and social phobia: a twin study. Depress Anxiety 2009;26(11):1004-11.

(24) Kirk KM, Birley AJ, Statham DJ, Haddon B, Lake RI, Andrews JG, et al. Anxiety and depression in twin and sib pairs extremely discordant and concordant for neuroticism: prodromus to a linkage study. Twin Res 2000 Dec;3(4):299-309.

(25) Hansell NK, Agrawal A, Whitfield JB, Morley KI, Zhu G, Lind PA, et al. Long-term stability and heritability of telephone interview measures of alcohol consumption and dependence. Twin Res Hum Genet 2008 Jun;11(3):287-305.

(26) Medland SE, Nyholt DR, Painter JN, McEvoy BP, McRae AF, Zhu G, et al. Common variants in the trichohyalin gene are associated with straight hair in Europeans. Am J Hum Genet 2009 Nov;85(5):750-5.

(27) Oldehinkel AJ, Rosmalen JG, Buitelaar JK, Hoek HW, Ormel J, Raven D, et al. Cohort Profile Update: the TRacking Adolescents' Individual Lives Survey (TRAILS). Int J Epidemiol 2015 Feb;44(1):76-76n.

(28) de Winter AF, Oldehinkel AJ, Veenstra R, Brunnekreef JA, Verhulst FC, Ormel J. Evaluation of non-response bias in mental health determinants and outcomes in a large sample of pre-adolescents. Eur J Epidemiol 2005;20(2):173-81.

(29) Penninx BW, Beekman AT, Smit JH, Zitman FG, Nolen WA, Spinhoven P, et al. The Netherlands Study of Depression and Anxiety (NESDA): rationale, objectives and methods. Int J Methods Psychiatr Res 2008;17(3):121-40.

(30) Willemsen G, Vink JM, Abdellaoui A, den BA, van Beek JH, Draisma HH, et al. The Adult Netherlands Twin Register: twenty-five years of survey and biological data collection. Twin Res Hum Genet 2013 Feb;16(1):271-81.

(31) World Health Organization. Composite International Diagnostic Interview (version 2.1). Geneva: WHO; 1992.

(32) Achenbach TM, Rescorla LA. Manual for the ASEBA Adult Forms & Profiles. Burlington, VT: University of Vermont, Research Center for Children, Youth, & Families; 2003.

(33) Beck AT, Rial WY, Rickels K. Short form of depression inventory: cross-validation. Psychol Rep 1974 Jun;34(3):1184-6.

(34) Van der Ploeg H, Defares P.B., Spielberger C.D. Zelfbeoordelingsvragenslijst STAI, versie DY-1 en DY-2. Lisse: Swets & Zeitlinger; 1979.

(35) Wilde GJS. Neurotische labiliteit gemeten volgens de vragenlijstmethode (The questionnaire method as a means of measuring neurotic instability). Amsterdam: Van Rossen; 1970.

(36) Boomsma DI, Sullivan PF, de Geus EJ, Heutink P, Meijer P, Sondervan D, et al. Genome-wide association of major depression: description of samples for the GAIN major depressive disorder study: NTR and NESDA Biobank Projects. Eur J Hum Genet 2008;16:335-42.

(37) Scheet P, Ehli EA, Xiao X, van Beijsterveldt CE, Abdellaoui A, Althoff RR, et al. Twins, Tissue, and Time: An Assessment of SNPs and CNVs. Twin Res Hum Genet 2012 Sep 28;1-9.

(38) Hettema JM, Webb BT, Guo AY, Zhao Z, Maher BS, Chen X, et al. Prioritization and association analysis of murine-derived candidate genes in anxiety-spectrum disorders. Biol Psychiatry 2011 Nov 1;70(9):888-96.

(39) Muthen LK, Muthen BO. Mplus User's Guide. Sixth Edition, Los Angeles, CA ed. Los Angeles, CA: Muthen & Muthen; 2010.

(40) Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. Am J Hum Genet 2011 Jan 7;88(1):76-82.

(41) Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, Nyholt DR, et al. Common SNPs explain a large proportion of the heritability for human height. Nat Genet 2010 Jul;42(7):565-9.

(42) Bulik-Sullivan BK, Loh PR, Finucane HK, Ripke S, Yang J, Patterson N, et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. Nat Genet 2015 Mar;47(3):291-5.

(43) Yang J, Weedon MN, Purcell S, Lettre G, Estrada K, Willer CJ, et al. Genomic inflation factors under polygenic inheritance. Eur J Hum Genet 2011 Jul;19(7):807-12.

(44) Hettema JM, Neale MC, Kendler KS. A review and meta-analysis of the genetic epidemiology of anxiety disorders. Am J Psychiatry 2001 Oct;158(10):1568-78.

(45) Wray NR, Lee SH, Mehta D, Vinkhuyzen AA, Dudbridge F, Middeldorp CM. Research review: Polygenic methods and their application to psychiatric traits. J Child Psychol Psychiatry 2014 Oct;55(10):1068-87.

(46) Ripke S, O'Dushlaine C, Chambert K, Moran JL, Kahler AK, Akterin S, et al. Genome-wide association analysis identifies 13 new risk loci for schizophrenia. Nat Genet 2013 Oct;45(10):1150-9.

(47) Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near ODZ4. Nat Genet 2011 Oct;43(10):977-83.

(48) Ripke S, Wray NR, Lewis CM, Hamilton SP, Weissman MM, Breen G, et al. A mega-analysis of genome-wide association studies for major depressive disorder. Mol Psychiatry 2013 Apr;18(4):497-511.

(49) Bulik-Sullivan B, Finucane HK, Anttila V, Gusev A, Day FR, Loh PR, et al. An atlas of genetic correlations across human diseases and traits. Nat Genet 2015 Nov;47(11):1236-41.

(50) Cross-Disorder Group of the Psychiatric GWAS Consortium. Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. Lancet 2013 Apr 20;381(9875):1371-9.

(51) Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, et al. Systematic localization of common disease-associated variation in regulatory DNA. Science 2012 Sep 7;337(6099):1190-5.

(52) Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, Heravi-Moussavi A, et al. Integrative analysis of 111 reference human epigenomes. Nature 2015 Feb 19;518(7539):317-30.