Molecular Psychiatry (2016) 21, 1485

www.nature.com/mp

ERRATUM

Meta-analysis of genome-wide association studies of anxiety disorders

T Otowa, K Hek, M Lee, EM Byrne, SS Mirza, MG Nivard, T Bigdeli, SH Aggen, D Adkins, A Wolen, A Fanous, MC Keller, E Castelao, Z Kutalik, SV der Auwera, G Homuth, M Nauck, A Teumer, Y Milaneschi, J-J Hottenga, N Direk, A Hofman, A Uitterlinden, CL Mulder, AK Henders, SE Medland, S Gordon, AC Heath, PAF Madden, ML Pergadia, PJ van der Most, IM Nolte, FVA van Oort, CA Hartman, AJ Oldehinkel, M Preisig, HJ Grabe, CM Middeldorp, BWJH Penninx, D Boomsma, NG Martin, G Montgomery, BS Maher, EJ van den Oord, NR Wray, H Tiemeier and JM Hettema

Molecular Psychiatry (2016) **21,** 1485; doi:10.1038/mp.2016.11; published online 9 February 2016

Correction to: *Molecular Psychiatry* (2015); advance online publication 12 January 2016; doi:10.1038/mp.2015.197

Following publication of the above article, the authors noticed that the fifteenth author's name was presented incorrectly. The author's name should have appeared as S Van der Auwera. The publisher regrets the error.

Molecular Psychiatry (2016) 21, 1391–1399

www.nature.com/mp

ORIGINAL ARTICLE

Meta-analysis of genome-wide association studies of anxiety disorders

T Otowa[1,2,35], K Hek[3,4,35], M Lee[1,35], EM Byrne[5,6,35], SS Mirza[3,35], MG Nivard[7], T Bigdeli[1], SH Aggen[1], D Adkins[8], A Wolen[9], A Fanous[10,11], MC Keller[12,13], E Castelao[14], Z Kutalik[15,16,17], SV der Auwera[18], G Homuth[19], M Nauck[20], A Teumer[21], Y Milaneschi[22,23], J-J Hottenga[22,23], N Direk[3], A Hofman[3], A Uitterlinden[3,24], CL Mulder[25], AK Henders[5,6], SE Medland[5], S Gordon[5], AC Heath[26], PAF Madden[26], ML Pergadia[26,27], PJ van der Most[28], IM Nolte[28], FVA van Oort[29], CA Hartman[30], AJ Oldehinkel[30], M Preisig[14], HJ Grabe[18,31], CM Middeldorp[7,32], BWJH Penninx[22,23], D Boomsma[7,23], NG Martin[5], G Montgomery[5], BS Maher[33], EJ van den Oord[8], NR Wray[5,6], H Tiemeier[3,25,34] and JM Hettema[1]

Anxiety disorders (ADs), namely generalized AD, panic disorder and phobias, are common, etiologically complex conditions with a partially genetic basis. Despite differing on diagnostic definitions based on clinical presentation, ADs likely represent various expressions of an underlying common diathesis of abnormal regulation of basic threat–response systems. We conducted genome-wide association analyses in nine samples of European ancestry from seven large, independent studies. To identify genetic variants contributing to genetic susceptibility shared across interview-generated DSM-based ADs, we applied two phenotypic approaches: (1) comparisons between categorical AD cases and supernormal controls, and (2) quantitative phenotypic factor scores (FS) derived from a multivariate analysis combining information across the clinical phenotypes. We used logistic and linear regression, respectively, to analyze the association between these phenotypes and genome-wide single nucleotide polymorphisms. Meta-analysis for each phenotype combined results across the nine samples for over 18 000 unrelated individuals. Each meta-analysis identified a different genome-wide significant region, with the following markers showing the strongest association: for case–control contrasts, rs1709393 located in an uncharacterized non-coding RNA locus on chromosomal band 3q12.3 ($P = 1.65 \times 10^{-8}$); for FS, rs1067327 within *CAMKMT* encoding the calmodulin-lysine *N*-methyltransferase on chromosomal band 2p21 ($P = 2.86 \times 10^{-9}$). Independent replication and further exploration of these findings are needed to more fully understand the role of these variants in risk and expression of ADs.

Molecular Psychiatry (2016) **21,** 1391–1399; doi:10.1038/mp.2015.197; published online 12 January 2016

INTRODUCTION

Anxiety disorders (ADs), namely generalized AD (GAD), panic disorder (PD) and phobias, are relatively common, often disabling conditions with lifetime prevalence of over 20%.[1] Family and twin studies suggest both genetic and environmental factors underlying their etiology, with moderate levels of familial aggregation (odds ratio 3–6) and heritability (30–50%).[2] As with most complex genetic traits, many linkage and candidate gene association studies of ADs have been conducted, with little success in robustly identifying their susceptibility genes.[3,4]

[1]Department of Psychiatry, Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University, Richmond, VA, USA; [2]Department of Neuropsychiatry, Graduate School of Medicine, University of Tokyo, Tokyo, Japan; [3]Department of Epidemiology, Erasmus Medical Center, Rotterdam, The Netherlands; [4]Department of Psychiatry, Epidemiological and Social Psychiatric Research institute, Erasmus Medical Center, Rotterdam, The Netherlands; [5]QIMR Berghofer Institute of Medical Research, Brisbane, QLD, Australia; [6]Queensland Brain Institute, University of Queensland, Brisbane, QLD, Australia; [7]Biological Psychology, VU University Amsterdam, Amsterdam, The Netherlands; [8]Department of Pharmacy, Center for Biomarker Research and Personalized Medicine, Virginia Commonwealth University, Richmond, VA, USA; [9]Center for Clinical and Translational Research, Virginia Commonwealth University, Richmond, VA, USA; [10]Mental Health Service Line, Washington VA Medical Center, Washington DC, USA; [11]Department of Psychiatry, Georgetown University School of Medicine, Washington DC, USA; [12]Department of Psychology and Neuroscience, University of Colorado, Boulder, CO, USA; [13]Institute for Behavioral Genetics, University of Colorado, Boulder, CO, USA; [14]Department of Psychiatry, Lausanne University Hospital, Lausanne, Switzerland; [15]Department of Medical Genetics, University of Lausanne, Lausanne, Switzerland; [16]Institute of Social and Preventive Medicine, Centre Hospitalier Universitaire Vaudois, Lausanne, Switzerland; [17]Swiss Institute of Bioinformatics, Lausanne, Switzerland; [18]Department of Psychiatry and Psychotherapy, University Medicine Greifswald, Greifswald, Germany; [19]Interfaculty Institute for Genetics and Functional Genomics, University Medicine and Ernst Moritz Arndt-University Greifswald, Greifswald, Germany; [20]Institute of Clinical Chemistry and Laboratory Medicine, University Medicine Greifswald, Greifswald, Germany; [21]Institute for Community Medicine, University Medicine Greifswald, Greifswald, Germany; [22]Department of Psychiatry, VU University Medical Center/GGZ inGeest, Amsterdam, The Netherlands; [23]EMGO+ Institute for Health and Care Research, VU University Medical Centre, Amsterdam, The Netherlands; [24]Department of Internal Medicine, Erasmus Medical Center, Rotterdam, The Netherlands; [25]Department of Psychiatry, Erasmus Medical Center, Rotterdam, The Netherlands; [26]Department of Psychiatry, Washington University School of Medicine, St Louis, MO, USA; [27]Florida Atlantic University, Boca Raton, FL, USA; [28]Department of Genetic Epidemiology, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands; [29]Department of Child and Adolescent Psychiatry/Psychology, Erasmus Medical Center, Sophia Children's Hospital, Rotterdam, The Netherlands; [30]Department of Psychiatry, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands; [31]Department of Psychiatry and Psychotherapy, Helios Hospital Stralsund, Stralsund, Germany; [32]Department of Child and Adolescent Psychiatry, GGZ inGeest/VU University Medical Center, Amsterdam, The Netherlands; [33]Department of Mental Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA and [34]Department of Child and Adolescent Psychiatry, Erasmus University Medical Center, Rotterdam, The Netherlands. Correspondence: Dr JM Hettema, Department of Psychiatry, Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University, PO Box 980126, Richmond, VA 23298-0126, USA.
E-mail: jhettema@vcu.edu
[35]These authors contributed equally to this work.
Received 26 May 2015; revised 16 October 2015; accepted 5 November 2015; published online 12 January 2016

Table 1. Contributing cohorts

Cohort	*Country*	*Mean age (s.d.) at interview*	N *phenotyped*	N *cases (ANX = 2)*	N *controls (ANX = 0)*	N *analyzed (FS/CC)*
MGS controls[45]	USA	50.0 (16.4)	2659	757	1059	2009/1336
PsyCoLaus[46]	Switzerland	50.9 (8.8)	3575	1044	1351	2887/1955
RS[47]	Netherlands	66.5 (10.8)	9686	1112	5459	7832/5379
SHIP[48]	Germany	55.4 (13.9)	2279	581	890	2026/1379
QIMR[49]	Australia	41.3 (11.5)	6147	1611	2544	2277/2156
TRAILS[50]	Netherlands	18.7 (0.7)	1584	390	472	1155/614
NESDA[51]/NTR[52,53]	Netherlands	45.6 (14.2)/44.6 (12.7)	4491	1521	2970	NA/4491
Total			31 060	7016	14 745	18 186/17 310

Abbreviations: ANX, anxiety scoring variable; CC, case–control; FS, factor score; MGS, molecular genetics of Schizophrenia; NESDA/NTR, The Netherlands study of depression and anxiety/Netherlands twin registry; QIMR, Queensland institute of medical research; RS, Rotterdam Study; SHIP, study of health in pomerania; TRAILS, tracking adolescents' individual lives survey.

Genome-wide association studies (GWAS) have proven to be a successful method for the identification of common genetic variants that increase susceptibility to complex disease. Recently, GWAS of specific anxiety and related disorders such as PD,[5,6] post-traumatic stress disorder,[7–9] obsessive compulsive disorder,[10,11] and phobias[12] have been published. However, these have been limited by small sample sizes and resulting low overall power to detect significant associations.

Despite differing on diagnostic definitions based on clinical presentation, ADs likely represent various expressions of an underlying common diathesis of abnormal regulation of basic threat–response systems.[13] ADs exhibit strong lifetime comorbidity with each other,[14] with genetic epidemiologic studies pointing to shared genetic risk factors between them.[15,16] Since clinical descriptions do not reflect underlying genetic architecture, traditional studies focused on individual ADs may not represent an effective study design for such phenotypes. A more informative approach would coordinate data from clusters of disorders with shared genetic risk factors.[17] One such strategy is to model a latent anxiety liability factor indexing ADs with substantial genetic overlap. Also, for common disorders like ADs, disease states can be interpreted as extremes of continuous liability dimensions, as has been done for somatic illnesses like obesity and hypertension. Therefore, quantitative trait approaches, assuming a continuous liability distribution, can be used to construct informative latent psychiatric phenotypes.[18] Analyzing AD phenotypes in a coordinated manner may represent a powerful approach for identifying susceptibility genes for ADs. This strategy has yielded some success, as demonstrated by prior reports from our group.[19]

In the current study, we conducted genome-wide association analyses in nine large, independent samples. To identify genetic variants contributing to genetic susceptibility shared across the ADs, we applied two phenotypic approaches: (1) categorical case–control (CC) comparisons based on having any AD diagnosis, and (2) quantitative phenotypic factor scores (FS) derived from a multivariate analysis combining information across the clinical phenotypes. We performed a meta-analysis for each phenotype across the 9 samples for over 18 000 unrelated individuals using ~6.5 million imputed single nucleotide polymorphisms (SNPs). This represents the largest genetic study to date of any of the ADs and the first of this magnitude to explicitly incorporate comorbidity structure directly into prediction of SNP effects.

MATERIALS AND METHODS

Overview

We conducted parallel GWAS in nine samples of European ancestry and combined the results via meta-analysis. We applied two phenotypic strategies aimed at capturing common (pleiotropic) genetic effects shared across the five core ADs: GAD, PD, social phobia, agoraphobia and specific phobias. We conducted two types of analyses in each sample based on complementary approaches to modeling the comorbidity and common genetic risk across the ADs: (1) CC comparisons, in which cases were designated as having 'any AD' versus supernormal controls, and (2) quantitative FS estimated for every subject in the sample using confirmatory factor analysis.

Samples

Nine samples containing AD phenotypes from seven independent studies participating in the Anxiety NeuroGenetics STudy (ANGST) Consortium were included in the meta-analysis. Standardized assessment instruments were used to generate DSM-based AD diagnoses, with some exceptions. The samples were genotyped on various SNP arrays according to their original study designs. Genotype calling, quality control, imputation and association analyses were performed at each site under similar standard protocols. SNP imputation was conducted within each sample using IMPUTE2 (ref. 20) or MACH[21] software utilizing the full 1000 Genomes Project reference data (March 2012, release v3). Genomic locations were based on NCBI build 37/UCSC hg 19 data. After imputation, SNPs with minor allele frequency (MAF) < 0.01, poor imputation quality < 0.30 and Hardy–Weinberg equilibrium P-value $< 10^{-6}$ were removed. See Supplement for study descriptions and Supplementary Table S1 for details of genotyping and quality control procedures. Table 1 summarizes basic statistics by cohort.

Genome-wide association analyses

To identify genetic variants contributing to genetic susceptibility shared across the ADs, we applied and compared two complementary phenotypic approaches: (1) categorical CC comparisons, and (2) quantitative phenotypic FS. For CC comparisons, AD cases were assigned to subjects meeting criteria for any lifetime AD (ANX = 2) while control subjects were 'supernormal', that is, having few or no clinical anxiety symptoms (ANX = 0); those with subsyndromal ADs (ANX = 1) were excluded from the CC analyses. For FS analyses, first exploratory factor analyses were conducted using Mplus (version 4)[22] separately in each sample, finding evidence for a single common factor model by screen plots. This was followed by confirmatory factor analyses that estimated a single FS for each subject from this common AD liability factor. (See Supplement for details of phenotype construction.) Association analyses were then performed in each study independently with imputed SNP dosages under an additive genetic model using logistic regression for CC phenotype and linear regression for quantitative FS phenotype. As covariates, we used sex and age at interview, as they were significant predictors of the phenotypes. Ancestry principal components were estimated for each sample and included on a sample-by-sample basis depending on their correlation with the outcome phenotypes. The quantile–quantile plot was used to evaluate overall significance of the association test results and the genomic control factor λ.

Meta-analysis of GWAS

We performed an inverse-variance weighted, fixed-effects meta-analysis with all GWAS samples using METAL[23] (nine samples using CC phenotypes and eight using FS phenotypes). For each SNP, a pooled effect size, s.e. and

P-value were computed. SNPs with low minor allele frequency (< 0.05) were excluded, resulting in a final meta-analytic data set of ~6.5M SNPs. Cochran's *Q* statistics and corresponding I^2 statistics were used as heterogeneity metrics. Cochran's *Q* statistic was computed by summing the squared deviations of each study's estimate by weighting each study's contribution in the same manner as in the meta-analyses. I^2 measured the amount of heterogeneity that is not due to chance.

Quantile–quantile and Manhattan plots were examined, and false discovery rate *q*-values were calculated based on the *P*-values from the meta-analyses. *Q*-values provide a balance between type I and Type II errors and can be interpreted as the probability that a marker identified as significant is a false discovery.[24]

Cross-validation

To examine overall consistency of association between datasets, we employed a leave-one-out procedure for internal cross-validation. At each step, we meta-analyzed eight of the nine CC GWAS samples as the 'training' set (seven of the eight samples were used for FS), the results of which were then tested in the respective remaining target sample ('testing' set). The top associated SNPs in the training set ($P_{training} < 1 \times 10^{-5}$, pruned to $r^2 < 0.4$ within a 500-kb window) were used to test the replicability ($P_{testing} < 0.05$) and consistency of the direction of their effects with the top associated SNPs identified in each testing set. One thousand random permutations of phenotype allocation to an individual's genome-wide genotypes were performed in each training–testing set pair, totaling 9,000 and 8,000 permutations in CC and FS, respectively. Across all sets, we compared the aggregate numbers of replicated SNPs and SNPs with the same direction of effect against the numbers expected by chance.

Gene-based tests

The SNP-based *P*-values derived from the meta-analyses were applied to gene-based association testing using KGG software (http://statgenpro.psychiatry.hku.hk/limx/kgg/).[25] No prioritization or pre-selection of genes was performed. Gene-based tests in KGG combine univariate association statistics to evaluate the cumulative evidence of association in a gene using extended Simes test (GATES).[26] SNPs were mapped onto 23 931 genes according to the gene coordinate information from NCBI, and SNPs within 10 kb of each gene were assigned to that gene. We considered genes with $P < 2 \times 10^{-6}$ (=.05/23 931) as significant and those with *q*-value < 0.1 as interesting.[27]

Secondary analyses

We conducted several secondary analyses, the details of which are described in the Supplement.

1. SNP-based heritability: Genomic-relatedness-matrix restricted maximum likelihood (GREML), as implemented in the software program GCTA,[28] was conducted in our largest cohort (RS) to estimate the total amount of variance explained by all analyzed SNPs. This was supplemented by a similar procedure in the full meta-analytic sample using linkage disequilibrium (LD) score regression.[29]
2. Polygenic risk profile analyses: Given the observed high comorbidity between ADs and other psychiatric syndromes, genomic profile risk scores[30] were estimated to test the additive joint effects of multiple variants between our AD GWAS data as target samples and summary data from Psychiatric Genomics Consortium phase 1 schizophrenia (SCZ), bipolar disorder (BIP) and major depressive disorder as discovery samples.

RESULTS

GWAS meta-analysis

We performed an inverse-variance weighted, fixed-effects meta-analysis with all discovery GWAS data including ~6.5M common SNPs after applying post-imputation quality control to each study. The genomic inflation factor λ ranged from 0.990 to 1.038 for all studies. The quantile–quantile plots of the meta-analyses for the CC and FS phenotypes are presented in Figure 1. Meta-analytic inflation factors were 1.03 and 1.02, suggesting little effect of population stratification. Manhattan plots are presented in Figure 2. Table 2 lists the LD-independent, genome-wide significant SNPs and associated regions. For the CC model, the strongest association was observed at rs1709393 located in an intron of an uncharacterized non-coding RNA locus *LOC152225* on chromosome 3q12.3 ($P = 1.65 \times 10^{-8}$; $Q = 0.027$). Allelic frequencies were very similar across studies and ranged between 0.55 and 0.60. The most significant SNP in the FS model was rs1067327 on chromosome 2p21 within *CAMKMT* encoding the calmodulin-lysine *N*-methyltransferase ($P = 2.86 \times 10^{-9}$; $Q = 0.0017$) with LD extending into several adjacent genes. Allelic frequencies were consistent across studies, ranging from 0.32 to 0.36. Both of these SNPs were imputed with very high quality across studies ($R^2 > 0.93$). As indicted in the forest plots (Supplementary Figure S1), no heterogeneity of effects was observed for either SNP. Figure 3 displays the regional SNP plots for these two genome-wide significant loci.

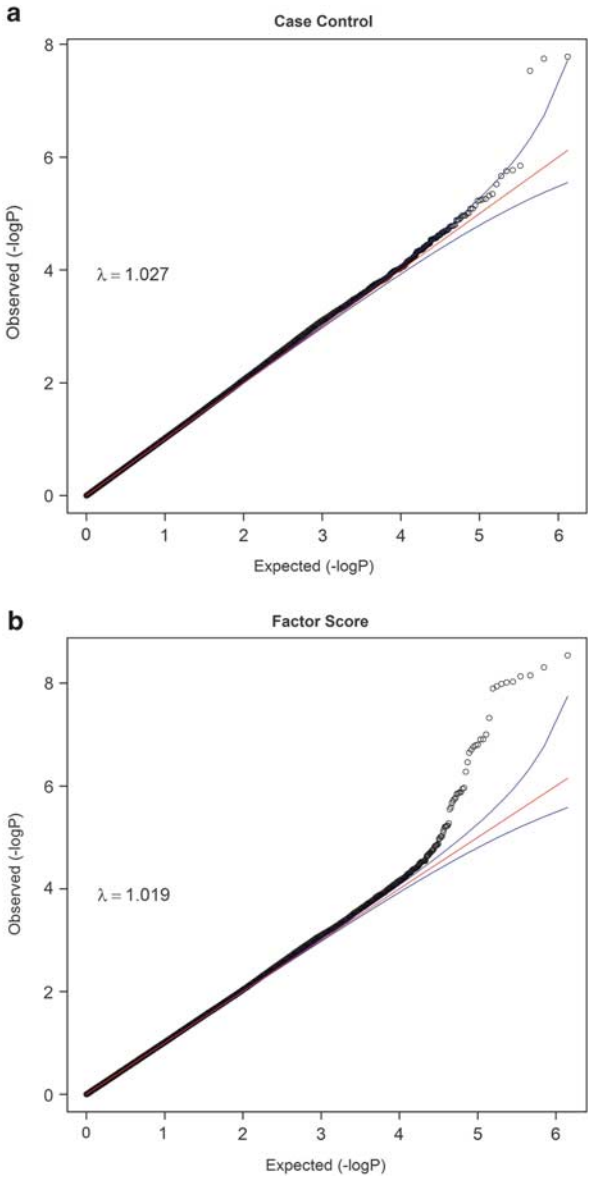

Figure 1. Quantile–quantile plots of meta-analysis results for (**a**) case–control and (**b**) factor score phenotypes. Observed association results of $-\log_{10}P$, after LD pruning at r^2 of 0.4, are plotted against the expected distribution under the null hypothesis of no association. LD, linkage disequilibrium.

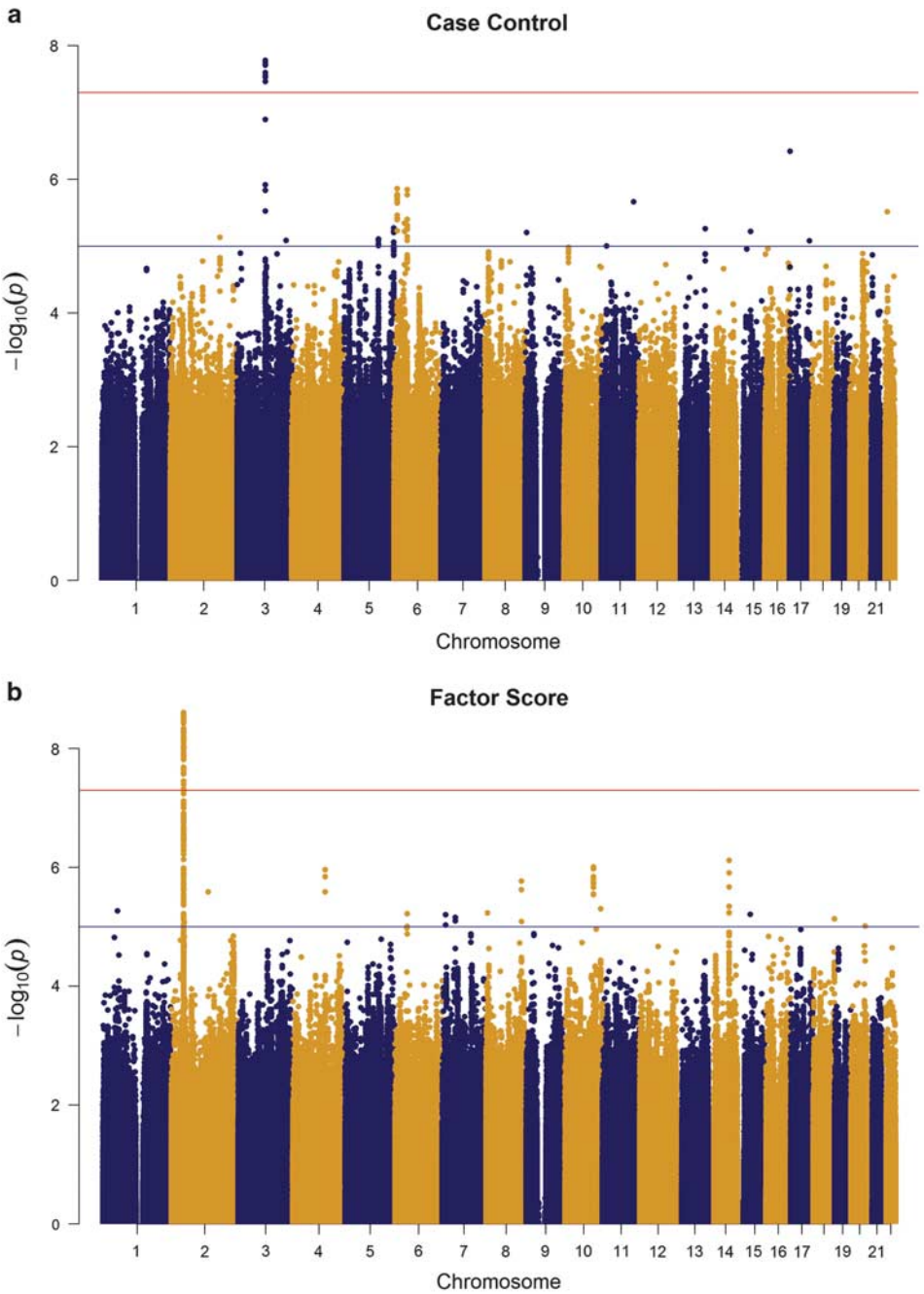

Figure 2. Manhattan plots of meta-analysis results for (**a**) case–control and (**b**) factor score phenotypes. Red horizontal line indicates the genome-wide significant *P*-value 5×10^{-8}; blue line indicates the suggestive *P*-value $= 1 \times 10^{-5}$.

Given that the CC and FS phenotypic approaches provide conceptually different but otherwise complementary information, we estimated the overlap in their association signals. These phenotypes were highly correlated in the different cohorts (0.88–0.94). Overall rank-based correlations between the CC and FS association effects were 0.61. The degree of correlation increased with decreasing *P*-value threshold, ranging from 0.275 to 0.899 (Supplementary Table S3). The most significant SNPs all have the same direction of effect (top 1000 SNPs in CC and top 1500 SNPs in FS); indeed, among the ~30% of total SNPs with opposite sign, none had even suggestively significant association ($P < 1 \times 10^{-5}$). However, among ~1.4M independent SNPS (pruned at $r^2 = 0.4$), significantly more with $P < 10^{-5}$ were identified for the FS phenotype than for the CC phenotype: 42 verses 18 (test *P*-value = 0.0034).

Cross-validation

In the leave-one-out cross-validation analyses, the replication rate was significantly higher than expected by chance (Table 3). In CC, 18 of 173 tested SNPs across all leave-one out analyses replicated in the left-out testing sets (permutation $P = 0.001$) and the proportion of SNPs with the same direction of effect was 59.5% (sign test $P = 0.005$). Of 315 tested SNPs in FS, 43 SNPs replicated

Table 2. Top association results for meta-analysis of SNP main effects for case–control and factor score phenotypes

SNP	*A12*[a]	*Frq*$_{mean}$	*Chr*	*Position*[b]	*Association with original phenotype*				*Cross phenotype*[e]	
					Effect[c] *(95% CI)*	P-*value*	Q-*value*	*Direction*[d]	P-*value*	*Effect*[c]
Case–control										
rs1709393	TC	0.58	3	101684480–101692234	0.860 (0.816–0.906)	1.65e-8	0.027	−−−−−−−	0.0342	−0.010
Factor score										
rs1067327	CG	0.34	2	44588941–44678648	0.028 (0.019–0.038)	2.86e-9	0.0017	++++++++	0.0002	1.123

Abbreviations: Chr, chromosome; CI, 95% confidence interval; Frq, frequency of allele 1; SNP, single nucleotide polymorphism. [a]First allele is the reference allele, for which the effect is reported. [b]Position denotes the associated region surrounding the top SNP containing one or more genome-wide significant SNPs in LD ($r^2 > 0.4$) with the top SNP. [c]Effect size represents odds ratio for case–control and regression coefficient for factor score. [d]Direction of association is provided for each study in the following order: RS1, RS2, RS3, NTR/NESDA, MGS, PsyColaus, SHIP, QIMR and TRAILS for case–control; RS1, RS2, RS3, MGS, PsyColaus, SHIP, QIMR and TRAILS for factor score. Plus (+) indicates the association between the SNP and the corresponding anxiety phenotype is positive. Minus (−) indicates a negative direction of association. [e]Shown are the association results of the top SNPs in the other phenotype, that is, factor score results of rs1709393 and case–control results of rs1067327.

(permutation $P < 0.001$) and 77.8% had the same direction of effect (sign test $P < 0.001$). Supplementary Figure S2 displays Manhattan plots of the training set meta-analyses conducted after leaving out each sample.

Gene-based tests

In the CC model, *LOC152225* on 3q12.3 surpassed genome-wide significance ($P = 1.19 \times 10^{-6}$; $Q = 0.028$). In the FS model, three genes exceeded genome-wide significance: *PREPL*, *CAMKMT* and *SLC3A1* on chromosome 2 (Table 4). Supplementary Figure S3 depicts the Manhattan plots for these gene-based analyses.

Secondary analyses

SNP-based heritability. This was estimated by GREML using GCTA in the Rotterdam sample as 0.106 (s.e. = 0.06, $P = 0.05$) for FS phenotype and 0.138 (s.e. = 0.18, $P = 0.2$) for CC phenotype on the liability scale assuming 10% AD population prevalence. Within the margin of error, these were consistent with LD score regression using summary statistics in the full meta-analysis sample, with SNP heritability estimated as 0.072 (s.e. = 0.028) for FS phenotype and 0.095 (s.e. = 0.037) for CC phenotype (see Supplement for details).

Polygenic risk profile analyses. Genomic profile risk score from PGC-MDD explained a small but significant proportion of variance in CC ADs in QIMR (0.5–0.7%), while SCZ and BIP each explained a somewhat smaller proportion of this variance varying by sample. These results were supported by LD score regression performed in the meta-analysis sample, estimating significant genetic correlation between ADs and MDD ($r = 0.68$) but not between ADs and BIP or SCZ (see Supplement for details).

DISCUSSION

We conducted the largest and most comprehensive genetic study of the primary ADs to date. Specifically, we integrated phenotypic information on GAD, PD, agoraphobia, social phobia and specific phobias and combined this with genome-wide SNP data from 9 large samples totaling over 18 000 subjects. We conducted parallel GWAS in these samples and statistically combined the results via meta-analysis, with the aim of detecting common variants that play a role in shared AD susceptibility.

While only an approximate representation of the underlying complexity of AD genetic mechanisms, our integrated phenotypic approaches successfully identified novel genetic variants that significantly associate with these composite AD phenotypes. The results were generally the same whether analyzing individual SNPs or genes. In the CC model, we identified a novel genome-wide association within an uncharacterized non-coding RNA locus *LOC152225* on chromosome 3q12.3. We found no extant reports for this locus in PubMed or the NHGRI Catalogue of Published GWAS (www.genome.gov/gwastudies/). In the FS model, we detected genome-wide significant associations at SNPs in three genes within a large LD block on chromosome 2p21, each of which has reported expression in brain: (1) *SLC3A1* encoding the large subunit of a heterodimeric dibasic/neutral amino acid transporter (solute carrier family 3 (amino acid transporter heavy chain), member 1); (2) *PREPL* encoding a putative prolyl endopeptidase belonging to the prolyl oligopeptidase family; and (3) *CAMKMT* encoding a calmodulin-lysine *N*-methyltransferase. This region is well-known for two contiguous gene-deletion syndromes, the hypotonia–cystinuria syndrome and the more severe 2p21 deletion syndrome.[31] Deletion of *SLC3A1* results in the autosomal-recessive form of cystinuria,[32] while *PREPL* deletion causes hypotonia at birth, failure to thrive and growth hormone deficiency.[33] The evolutionarily conserved class I protein methyltransferase encoded by *CAMKMT* acts in the formation of trimethyllysine in calmodulin, which is involved in calcium-dependent signaling.[34] Interestingly, GWAS of SCZ and BIP have highlighted other genes encoding proteins involved in calcium-dependent signaling.[35] Although the most significant SNP, rs1067327, is located in an intron of *CAMKMT*, *in silico* analyses (Supplement) suggest rs698775 is the most likely functional candidate with a *cis* regulatory effect possibly specific to *PREPL*.

There is substantial phenotypic overlap between the CC and FS models used to capture the comorbidity and shared genetic risk among the ADs, and as expected there was a high degree of concordance in the association signals genome wide (Supplementary Table S3). The most significantly associated SNPs ($P < .0.05$) have very high correlation of association effects, suggesting they are tapping into strongly related AD risk factors. We note that, overall, the FS phenotype identified a larger number of associated SNPs than the CC model. This is likely due to several reasons: (1) this approach combines disorder information to capture individual differences on an underlying latent AD liability; (2) for high prevalence disorders, quantitative variables generally have greater power for genetic association than categorical variables;[36,37] (3) the FS models generally involve larger sample sizes since they also include the subjects with subthreshold ADs (score = 1); and (4) the FS model produces a phenotype that incorporates the observed relationship information (covariance) between the individual ADs. These findings support the use of quantitative phenotypic factor scores in future GWAS of comorbid psychiatric disorders assessed in the same individuals.

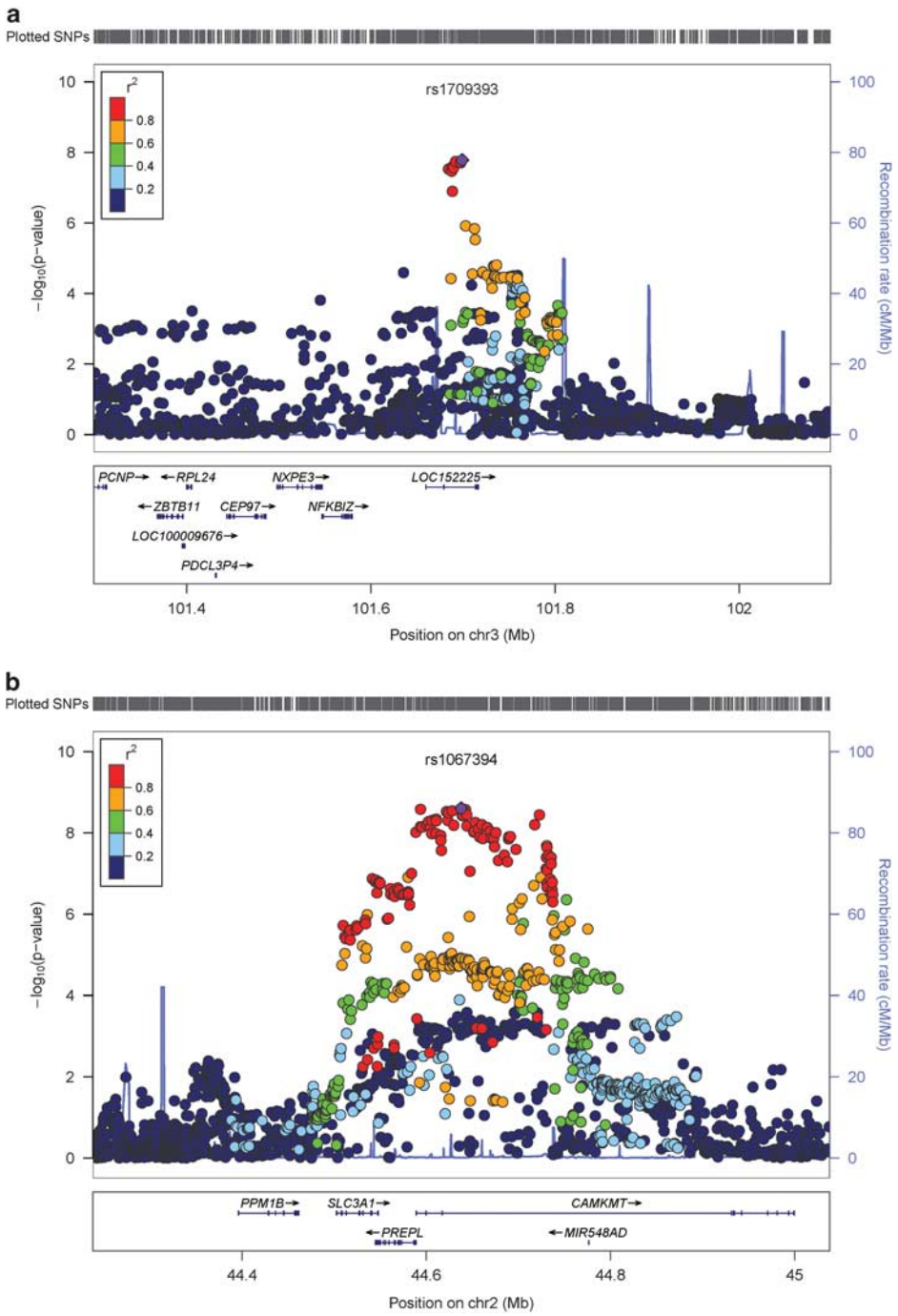

Figure 3. Regional plots around most significant SNPs in (**a**) case–control and (**b**) factor score model.

Several secondary analyses support our findings. First, we applied cross-validation in the nine samples to examine the internal consistency of the results. We created sub-samples by iteratively removing the data of each of the individual samples and conducting meta-analysis with the remaining datasets. A highly significant proportion of the top results were consistently identified across these sub-analyses, suggesting the stability and validity of our findings. Next we estimated the genome-wide contribution via GREML and the complementary LD score regression approach, producing generally consistent estimates of SNP heritability across samples included and methods applied. Similar to GWAS studies of many phenotypes,[38] these estimates are substantially smaller than those predicted by twin studies of ADs. Finally, we tested the polygenic association between our results and those from other psychiatric disorders using GRPS, finding significant correlation of genetic risk between ADs and MDD but not between ADs and BIP or SCZ. The former result is consistent with large epidemiologic studies that report correlated genetic risk between ADs and MDD (see ref. 16 for review), as well as a prior overlap seen for depression and anxiety scales.[39]

A strength of this study is that we applied phenotypic strategies aimed at detecting genetic variants that play a central but non-specific role in AD susceptibility. This is counter to the approach taken in most psychiatric genetic studies, which generally apply

Table 3. Results of leave-one-out cross-validation analyses

Testing sample	*Case–control*			*Factor score*		
	N_{SNP}[a] $P_{training} < 10^{-5}$	N_{SNP}[b] $P_{testing} < 0.05$	N_{SNP} *(same direction)*	N_{SNP}[a] $P_{training} < 10^{-5}$	N_{SNP}[b] $P_{testing} < 0.05$	N_{SNP} *(same direction)*
MGS	19	3	10	39	0	25
GSK	26	3	12	22	4	11
RS1	24	0	13	42	6	36
RS2	15	3	9	45	0	33
RS3	22	0	10	46	21	35
SHIP	24	0	14	40	10	32
NTR/NESDA[c]	14	3	13	—	—	—
QIMR	13	3	9	35	0	32
TRAILS	16	3	13	46	2	41
OVERALL	173	18[d] ($P = 0.001$)	103[d] ($P = 0.005$)	315	43[d] ($P < 0.001$)	245[d] ($P < 0.001$)

[a]The number of SNPs associated at Ptraining $< 1x10^{-5}$ with $r^2 < 0.4$ in the leave-one-out meta-analysis using N-1 training samples after removing one testing sample at a time. [b]Of the SNPs with $P_{training} < 1x10^{-5}$ in the meta-analysis of the N-1 training samples, the number of replicated SNPs with one-sided $P_{testing} < 0.05$ in the left-out (testing) sample. [c]Only case–control phenotype is available for NTR/NESDA sample. [d]$P < 0.05$ thresholds applied for replication and sign tests; Replication and sign tests are conducted based on permutations with 9,000 iterations in case–control and 8,000 iterations in factor score phenotype under the null hypothesis of no association (1,000 iteration for each pair of training and testing sets).

Table 4. Top associated genes ($Q < 0.1$) using gene-based tests

Gene	P-*value*[a]	Q-*value*	*SNP*[b] *with lowest* P	*Lowest* P[c]	*Chr*	*Gene feature*
Case–control						
LOC152225	**1.19E – 06**	0.028	rs1709393	1.65E – 08	3	ncRNA
Factor score						
PREPL	**5.61E – 08**	0.001	rs786618	6.99E – 09	2	Intronic
CAMKMT	**3.15E – 07**	0.004	rs1067327	2.86E – 09	2	Intronic
SLC3A1	**1.45E – 06**	0.012	rs1142523	1.44E – 07	2	3′ UTR
LBX1	7.50E – 06	0.045	rs11190870	9.79E – 07	10	Downstream
LBX1-AS1	1.33E – 05	0.064	rs594791	1.05E – 06	10	Upstream

Abbreviations: ncRNA, non-coding RNA; UTR, untranslated region. [a]Gene-based *P*-value (bolded genome-wide significant $P < 2 \times 10^{-6}$). [b]Most significant SNP within the corresponding gene. [c]SNP-based *P*-value for the most significant SNP We used LD pruning at r^2 of 0.4 for gene-based tests. Boldface indicates genome-wide $P < 2 \times 10^{-6}$.

CC comparisons for specific clinical diagnoses, sometimes followed by adjunct cross-disorder analyses. However, it has long been recognized that clinical nosology poorly reflects etiological mechanisms, with both genetic and environmental risk factors showing non-specific effects across disorders. ADs, despite their heterogeneous clinical presentations, likely represent various expressions of an underlying common diathesis of abnormal regulation of basic threat–response systems.[13] Given the value of fear and anxiety for survival, there are likely sets of evolutionarily conserved genes that regulate these basic biological responses. This is supported by twin studies that identify factors of common genetic risk across ADs in addition to disorder-specific genetic factors. With this in mind, we applied and compared two strategies for combining information across clinical phenotypes. The first is a simple CC approach, comparing cases defined as having 'any AD' against supernormal controls. The second applied multivariate modeling of the covariation among the ADs using the common factor model to define a single continuous dimension of liability for which quantitative scores can be estimated for each subject. Our group has applied this approach in prior candidate gene association studies[19] and in a pilot GWAS in the MGS sample,[40] but this is the first such application in a large GWAS meta-analysis. We note that this strategy is consistent with NIMH's Research Domain Criteria (RDoC) initiative, which aims to serve as a framework for new approaches to research on mental disorders based on fundamental dimensions that cut across traditional disorder categories and more closely align with mechanisms that underlie psychopathology at various biological levels from genes to neural circuits.[41] Also important to note is that ADs not only share genetic risk factors among themselves but also with other internalizing phenotypes like MDD,[16] obsessive compulsive disorder[42] and personality traits like neuroticism and extroversion.[43] It will be important for future studies to examine this broader pleiotropic spectrum either through cross-disorder GWAS as previously conducted for other psychiatric conditions[44] or by including these additional traits directly in the phenotypic construction with the ADs. It is possible that, by including AD cases with comorbid MDD, the genetic overlap between these conditions has influenced our results.

Several potential limitations of this study should be noted. First, although the total sample size far exceeds those from prior AD genetic studies, it is still relatively underpowered to detect common genetic variants of small effect expected for the genetic architecture of such complex phenotypes.[37] Second, not all samples provided the same level of phenotypic coverage; in particular, some subjects in QIMR were missing diagnostic data for GAD or specific phobia. While this can produce bias, our forest plots, tests for heterogeneity, and internal validation analyses suggest that this likely did not bias our results. Third, consent agreements for some of the sites did not allow for sharing of

subject data, so GWAS analyses had to be conducted separately using a standardized procedure and combined via meta-analysis. While this has been shown to approximate the power obtained when using raw data via mega-analysis, we were limited in our ability to conduct additional *post hoc* analyses such as genomic profile risk score and GREML that require the use of raw GWAS data. Reassuringly, results obtained by applying LD score regression to summary statistics from the total meta-analysis sample were consistent with those using raw data from select individual samples. Fourth, the results apply only to subjects of European ancestry and might not generalize to individuals of other genetic and cultural backgrounds. Finally, we combined all data available at the time of this study into a single meta-analysis rather than divide into discovery and replication samples. This was necessary due to the large sample sizes required to detect small effects of genes involved in complex traits like ADs. Internal cross-validation supported the robustness of our results but do not substitute for replication in well-powered, independent samples. At this time, we are unaware of other large datasets that could be used for replication of our results.

In summary, this study has identified several potentially novel susceptibility loci that increase shared risk across the primary ADs. Future studies are needed to (1) further confirm these findings via independent replication, (2) increase the total sample size to enhance power to detect additional loci and (3) identify loci associated specifically with each particular AD not accounted for by the pleiotropic effects targeted in this study.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGMENTS

This overall project was supported by NIH grant R01MH87646 to JMH. TO was supported by a research fellowship from the Japan Society for the Promotion of Science (no. 21-8373). Samples and associated phenotype data for the MGS study were collected under the following grants: NIMH Schizophrenia Genetics Initiative U01s: MH046276 (CR Cloninger), MH46289 (C Kaufmann), and MH46318 (MT Tsuang); and MGS Part 1 (MGS1) and Part 2 (MGS2) R01s: MH67257 (NG Buccola), MH59588 (BJ Mowry), MH59571 (PV Gejman), MH59565(R Freedman), MH59587 (F Amin), MH60870 (WF Byerley), MH59566 (DW Black), MH59586 (JM Silverman), MH61675 (DF Levinson) and MH60879 (CR Cloninger). Analyses in the Rotterdam Study were supported by the Research Institute for Diseases in the Elderly (014-93-015; RIDE2), The Netherlands Genomics Initiative (NGI)/Netherlands Consortium for Healthy Ageing (NCHA) project No. 050-060-810. The work of HT is supported by Vidi (grant 017.106.370). The Rotterdam Study is funded by Erasmus Medical Center, Rotterdam, The Netherlands Organization for the Health Research and Development (ZonMw), the Ministry of Education, Culture and Science, and the Ministry for Health, Welfare and Sports. SHIP is part of the Community Medicine Research net of the University of Greifswald, Germany, which is funded by the Federal Ministry of Education and Research (grants no. 01ZZ9603, 01ZZ0103 and 01ZZ0403), the Ministry of Cultural Affairs and the Social Ministry of the Federal State of Mecklenburg-West Pomerania. Genome-wide data have been supported by the Federal Ministry of Education and Research (grant no. 03ZIK012) and a joint grant from Siemens Healthcare, Erlangen, Germany and the Federal State of Mecklenburg-West Pomerania. This work was also funded by the German Research Foundation (DFG: GR 1912/5-1). The University of Greifswald is a member of the Caché Campus program of the InterSystems GmbH. The QIMR samples are made available through the generous and willing participation of twins and their families registered at the Australian Twin Registry and through grant funding awarded from many grant funding bodies including the Australian National Health and Medical Research Council (241944, 339462, 389927, 389875, 389891, 389892, 389938, 442915, 442981, 496675, 496739, 552485, 552498, 613608), the FP-5 GenomEUtwin Project (QLG2-CT- 2002-01254), the US National Institutes of Health (NIH grants AA07535, AA10248, AA13320, AA13321, AA13326, AA14041, MH66206, DA12854, DA019951) and the Center for Inherited Disease Research, Baltimore. We thank Dixie Statham (sample collection); Leanne Wallace, Anthony Caracella and staff of the Molecular Epidemiology Laboratory (DNA processing); David Smyth, Harry Beeby and Daniel Park (IT support). Statistical analyses were partly conducted at the Genetic Cluster Computer (http://www.geneticcluster.org), which is financially supported by the Netherlands Scientific Organization (NWO 480-05-003). EMB (1053639), SEM, GM and NRW (613602, 1078901) are supported by the National Health and Medical Research Council Fellowship Scheme. The CoLaus|PsyCoLaus study was and is supported by research grants from GlaxoSmithKline, the Faculty of Biology and Medicine of Lausanne, and the Swiss National Science Foundation (grants 3200B0–105993, 3200B0-118308, 33CSCO-122661, 33CS30-139468 and 33CS30-148401). Participating centers of TRAILS include various departments of the University Medical Center and University of Groningen, the Erasmus University Medical Center Rotterdam, the University of Utrecht, the Radboud Medical Center Nijmegen and the Parnassia Bavo group, all in the Netherlands. TRAILS has been financially supported by various grants from the Netherlands Organization for Scientific Research NWO (Medical Research Council program grant GB-MW 940-38-011; ZonMW Brainpower grant 100-001-004; ZonMw Risk Behavior and Dependence grants 60-60600-97-118; ZonMw Culture and Health grant 261-98-710; Social Sciences Council medium-sized investment grants GB-MaGW 480-01-006 and GB-MaGW 480-07-001; Social Sciences Council project grants GB-MaGW 452-04-314 and GB-MaGW 452-06-004; NWO large-sized investment grant 175.010.2003.005; NWO Longitudinal Survey and Panel Funding 481-08-013), the Dutch Ministry of Justice (WODC), the European Science Foundation (EuroSTRESS project FP-006), Biobanking and Biomolecular Resources Research Infrastructure BBMRI-NL (CP 32), the participating universities, and Accare Center for Child and Adolescent Psychiatry. Statistical analyses were carried out on the Genetic Cluster Computer (http://www.geneticcluster.org), which is financially supported by the Netherlands Scientific Organization (NWO 480-05-003) along with a supplement from the Dutch Brain Foundation. Funding for the NESDA/NTR studies was obtained from the Netherlands Organization for Scientific Research (Geestkracht program grant 10-000-1002; 904-61-090, 985-10- 002, 904-61-193, 480-04-004, 400-05-717, 912-100-20; Spinozapremie 56-464-14192); the Center for Medical Systems Biology (CSMB, NWO Genomics), Biobanking and Biomolecular Resources Research Infrastructure (BBMRI-NL), VU University's Institutes for Health and Care Research (EMGO+) and Neuroscience Campus Amsterdam, European Research Council (ERC, 230374),National Institutes of Health (NIH, R01D0042157-01 A, MH081802, Grand Opportunity grants 1RC2 MH089951 and 1RC2 MH089995). Part of the genotyping and analyses were funded by the Genetic Association Information Network (GAIN) of the Foundation for the National Institutes of Health. Computing was supported by BiG Grid, the Dutch e-Science Grid, which is financially supported by NWO.

REFERENCES

1 Kessler RC, Berglund P, Demler O, Jin R, Merikangas KR, Walters EE. Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. *Arch Gen Psychiatry* 2005; **62**: 593–602.
2 Hettema JM, Neale MC, Kendler KS. A review and meta-analysis of the genetic epidemiology of anxiety disorders. *Am J Psychiatry* 2001; **158**: 1568–1578.
3 Smoller JW, Block SR, Young MM. Genetics of anxiety disorders: the complex road from DSM to DNA. *Depress Anxiety* 2009; **26**: 965–975.
4 Maron E, Hettema JM, Shlik J. Advances in molecular genetics of panic disorder. *Mol Psychiatry* 2010; **15**: 681–701.
5 Erhardt A, Czibere L, Roeske D, Lucae S, Unschuld PG, Ripke S *et al.* TMEM132D, a new candidate for anxiety phenotypes: evidence from human and mouse studies. *Mol Psychiatry* 2011; **16**: 647–663.
6 Otowa T, Kawamura Y, Nishida N, Sugaya N, Koike A, Yoshida E *et al.* Meta-analysis of genome-wide association studies for panic disorder in the Japanese population. *Transl Psychiatry* 2012; **2**: e186.
7 Logue MW, Baldwin C, Guffanti G, Melista E, Wolf EJ, Reardon AF *et al.* A genome-wide association study of post-traumatic stress disorder identifies the retinoid-related orphan receptor alpha (RORA) gene as a significant risk locus. *Mol Psychiatry* 2013; **18**: 937–942.
8 Xie P, Kranzler HR, Yang C, Zhao H, Farrer LA, Gelernter J. Genome-wide association study identifies new susceptibility loci for posttraumatic stress disorder. *Biol Psychiatry* 2013; **74**: 656–663.
9 Guffanti G, Galea S, Yan L, Roberts AL, Solovieff N, Aiello AE *et al.* Genome-wide association study implicates a novel RNA gene, the lincRNA AC068718.1, as a risk factor for post-traumatic stress disorder in women. *Psychoneuroendocrinology* 2013; **38**: 3029–3038.
10 Stewart SE, Yu D, Scharf JM, Neale BM, Fagerness JA, Mathews CA *et al.* Genome-wide association study of obsessive-compulsive disorder. *Mol Psychiatry* 2013; **18**: 788–798.
11 Mattheisen M, Samuels JF, Wang Y, Greenberg BD, Fyer AJ, McCracken JT *et al.* Genome-wide association study in obsessive-compulsive disorder: results from the OCGAS. *Mol Psychiatry* 2014; **20**: 337–344.
12 Walter S, Glymour MM, Koenen K, Liang L, Tchetgen Tchetgen EJ, Cornelis M *et al.* Performance of polygenic scores for predicting phobic anxiety. *PLoS One* 2013; **8**: e80326.
13 Craske MG, Rauch SL, Ursano R, Prenoveau J, Pine DS, Zinbarg RE. What is an anxiety disorder? *Depress Anxiety* 2009; **26**: 1066–1085.

14 Comorbidity of Mood and Anxiety Disorders. American Psychiatric Press: Washington DC, USA, 1990.
15 Hettema JM, Prescott CA, Myers JM, Neale MC, Kendler KS. The structure of genetic and environmental risk factors for anxiety disorders in men and women. *Arch Gen Psychiatry* 2005; **62**: 182–189.
16 Middeldorp CM, Cath DC, van Dyck R, Boomsma D. The co-morbidity of anxiety and depression in the perspective of genetic epidemiology. A review of twin and family studies. *Psychol Med* 2005; **35**: 611–624.
17 Kendler KS, Prescott CA, Myers J, Neale MC. The structure of genetic and environmental risk factors for common psychiatric and substance use disorders in men and women. *Arch Gen Psychiatry* 2003; **60**: 929–937.
18 Plomin R, Haworth CM, Davis OS. Common disorders are quantitative traits. *Nat Rev Genet* 2009; **10**: 872–878.
19 Hettema JM, An SS, Neale MC, Bukszar J, van den Oord EJ, Kendler KS *et al.* Association between glutamic acid decarboxylase genes and anxiety disorders, major depression, and neuroticism. *Mol Psychiatry* 2006; **11**: 752–762.
20 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; **44**: 955–959.
21 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; **34**: 816–834.
22 Muthen LK, Muthen BO. *Mplus User's Guide*. Sixth Edition, Muthen & Muthen: Los Angeles, CA, 2010.
23 Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genome-wide association scans. *Bioinformatics* 2010; **26**: 2190–2191.
24 Storey JD, Tibshirani R. Statistical significance for genomewide studies. *Proc Natl Acad Sci USA* 2003; **100**: 9440–9445.
25 Li MX, Sham PC, Cherny SS, Song YQ. A knowledge-based weighting framework to boost the power of genome-wide association studies. *PLoS One* 2010; **5**: e14480.
26 Li MX, Gui HS, Kwan JS, Sham PC. GATES: a rapid and powerful gene-based association test using extended Simes procedure. *Am J Hum Genet* 2011; **88**: 283–293.
27 van den Oord EJ, Sullivan PF. False discoveries and models for gene discovery. *Trends Genet* 2003; **19**: 537–542.
28 Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet* 2011; **88**: 76–82.
29 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; **47**: 291–295.
30 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; **55**: 1068–1087.
31 Martens K, Jaeken J, Matthijs G, Creemers JW. Multi-system disorder syndromes associated with cystinuria type I. *Curr Mol Med* 2008; **8**: 544–550.
32 Saravakos P, Kokkinou V, Giannatos E. Cystinuria: current diagnosis and management. *Urology* 2014; **83**: 693–699.
33 Martens K, Derua R, Meulemans S, Waelkens E, Jaeken J, Matthijs G *et al.* PREPL: a putative novel oligopeptidase propelled into the limelight. *Biol Chem* 2006; **387**: 879–883.
34 Magnani R, Dirk LM, Trievel RC, Houtz RL. Calmodulin methyltransferase is an evolutionarily conserved enzyme that trimethylates Lys-115 in calmodulin. *Nat Commun* 2010; **1**: 43.
35 Neale BM, Sklar P. Genetic analysis of schizophrenia and bipolar disorder reveals polygenicity but also suggests new directions for molecular interrogation. *Curr Opin Neurobiol* 2015; **30C**: 131–138.
36 Yang J, Wray NR, Visscher PM. Comparing apples and oranges: equating the power of case-control and quantitative trait association studies. *Genet Epidemiol* 2010; **34**: 254–257.
37 van der Sluis S, Posthuma D, Nivard MG, Verhage M, Dolan CV. Power in GWAS: lifting the curse of the clinical cut-off. *Mol Psychiatry* 2013; **18**: 2–3.
38 Visscher PM, Brown MA, McCarthy MI, Yang J. Five years of GWAS discovery. *Am J Hum Genet* 2012; **90**: 7–24.
39 Demirkan A, Penninx BW, Hek K, Wray NR, Amin N, Aulchenko YS *et al.* Genetic risk profiles for depression and anxiety in adult and elderly cohorts. *Mol Psychiatry* 2011; **16**: 773–783.
40 Otowa T, Maher BS, Aggen SH, McClay JL, van den Oord EJ, Hettema JM. Genome-wide and gene-based association studies of anxiety disorders in European and African American samples. *PLoS One* 2014; **9**: e112559.
41 Insel T, Cuthbert B, Garvey M, Heinssen R, Pine DS, Quinn K *et al.* Research domain criteria (RDoC): toward a new classification framework for research on mental disorders. *Am J Psychiatry* 2010; **167**: 748–751.
42 Bienvenu OJ, Samuels JF, Wuyek LA, Liang KY, Wang Y, Grados MA *et al.* Is obsessive-compulsive disorder an anxiety disorder, and what, if any, are spectrum conditions? A family study perspective. *Psychol Med* 2012; **42**: 1–13.
43 Bienvenu OJ, Hettema JM, Neale MC, Prescott CA, Kendler KS. Low extraversion and high neuroticism as indices of genetic and environmental risk for social phobia, agoraphobia, and animal phobia. *Am J Psychiatry* 2007; **164**: 1714–1721.
44 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; **381**: 1371–1379.
45 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; **167**: 854–865.
46 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.
47 Hofman A, Breteler MM, van Duijn CM, Krestin GP, Pols HA, Stricker BH *et al.* The Rotterdam Study: objectives and design update. *Eur J Epidemiol* 2007; **22**: 819–829.
48 Schmidt CO, Watzke AB, Schulz A, Baumeister SE, Freyberger HJ, Grabe HJ. [The lifetime prevalence of mental disorders in north-eastern Germany. What is the influence of earlier mental morbidity on survey participation and prevalence estimates? Results from the SHIP-study]. *Psychiatr Prax* 2013; **40**: 192–199.
49 Jardine R, Martin NG, Henderson AS. Genetic covariation between neuroticism and the symptoms of anxiety and depression. *Genet Epidemiol* 1984; **1**: 89–107.
50 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; **44**: 76–76n.
51 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**: 121–140.
52 Boomsma DI, Vink JM, Van Beijsterveldt TC, de Geus EJ, Beem AL, Mulder EJ *et al.* Netherlands Twin Register: a focus on longitudinal research. *Twin Res* 2002; **5**: 401–406.
53 Boomsma DI, de Geus EJ, Vink JM, Stubbe JH, Distel MA, Hottenga JJ *et al.* Netherlands Twin Register: from twins to twin families. *Twin Res Hum Genet* 2006; **9**: 849–857.

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