Molecular Genetic Analysis Subdivided by Adversity Exposure Suggests Etiologic Heterogeneity in Major Depression

Roseann E. Peterson, Ph.D., Na Cai, Ph.D., Andy W. Dahl, Ph.D., Tim B. Bigdeli, Ph.D., Alexis C. Edwards, Ph.D., Bradley T. Webb, Ph.D., Silviu-Alin Bacanu, Ph.D., Noah Zaitlen, Ph.D., Jonathan Flint, M.D., Kenneth S. Kendler, M.D.

Objective: The extent to which major depression is the outcome of a single biological mechanism or represents a final common pathway of multiple disease processes remains uncertain. Genetic approaches can potentially identify etiologic heterogeneity in major depression by classifying patients on the basis of their experience of major adverse events.

Method: Data are from the China, Oxford, and VCU Experimental Research on Genetic Epidemiology (CONVERGE) project, a study of Han Chinese women with recurrent major depression aimed at identifying genetic risk factors for major depression in a rigorously ascertained cohort carefully assessed for key environmental risk factors (N=9,599). To detect etiologic heterogeneity, genome-wide association studies, heritability analyses, and gene-by-environment interaction analyses were performed.

Results: Genome-wide association studies stratified by exposure to adversity revealed three novel loci associated with major depression only in study participants with no history of adversity. Significant gene-by-environment interactions were seen between adversity and genotype at all three loci, and 13.2% of major depression liability can be attributed to genome-wide interaction with adversity exposure. The genetic risk in major depression for participants who reported major adverse life events (27%) was partially shared with that in participants who did not (73%; genetic correlation=+0.64). Together with results from simulation studies, these findings suggest etiologic heterogeneity within major depression as a function of environmental exposures.

Conclusions: The genetic contributions to major depression may differ between women with and those without major adverse life events. These results have implications for the molecular dissection of major depression and other complex psychiatric and biomedical diseases.

Am J Psychiatry 2018; 175:545–554; doi: 10.1176/appi.ajp.2017.17060621

The heterogeneity of major depression, demonstrated by variable symptom presentation, course of illness, and treatment response, has hindered our understanding of its etiology (1, 2). To counter the problem of heterogeneity, researchers have attempted to study homogeneous subtypes (e.g., atypical depression, early age at onset) (3, 4). Indeed, the decades-long debate about the number of distinct depressive subtypes remains unresolved (1). Of particular interest is that a large literature suggests that major depression can be usefully divided into a stress-responsive subtype (e.g., reactive depression) and a subtype with no apparent environmental precipitants (e.g., endogenous) (5–9).

In this study, we examined whether genetic approaches can identify etiologically heterogeneous depressive subtypes. We explored whether the two main classes of known causal factors for major depression, genes and environment, represent partially distinct pathways to major depression. Genetic effects on major depression are well established from twin studies (10) and genome-wide association data (11–13). Molecular genetic analysis reveals that major depression, like other complex diseases, has a polygenic architecture with multiple loci of small effect (14, 15).

Adversity exposure increases risk for major depression (16) with a dose-response relationship between severity of stressors and disease risk (17–19). Results of co-twin control studies suggest that this association is largely causal (18, 20). However, adversity is neither necessary nor sufficient to produce major depression, and it has been difficult to identify clinical features distinguishing cases of major depression with and without environmental precipitants (21–23).

Genetic risk factors for major depression not only alter average risk but also influence sensitivity to depressogenic effects of environmental adversities, particularly childhood maltreatment and adult life events (24–26). For example,

See related features: **Editorial** *by Dr. Yehuda and Dr. Flory (p. 491),* **AJP Audio** *(online), and* **Video** *by Dr. Pine (online)*

exposure to severe stressful life events increases risk for major depression more strongly in those with high genetic liability compared with those with low genetic liability (27). Despite strong effects of environmental exposures on major depression risk, there have been limited efforts incorporating these factors into large-scale molecular genetic studies. Furthermore, the major depression-associated genetic loci identified to date account for only a small portion of the variance in disease liability (12–14), underscoring the importance of continued research on detecting heterogeneity as a mechanism for identifying etiologically relevant determinants.

Therefore, we investigated whether genetic approaches can demonstrate etiologic heterogeneity among major depression cases by classifying individuals on the basis of adversity exposure. Using data from the China, Oxford, and VCU Experimental Research on Genetic Epidemiology (CONVERGE) project (13), a study of Han Chinese women with recurrent major depression (N=9,599) aimed at identifying genetic risk factors in a rigorously ascertained cohort assessed for key environmental risk factors, we explored whether major depression with and without major environmental adversities may represent, from a genetic perspective, partially distinct subtypes.

METHOD

Sample Collection

Patients with recurrent major depression (case subjects) were recruited from 58 provincial mental health centers and psychiatric departments of medical hospitals in 45 cities and 23 provinces of China. Control subjects were recruited from multiple locations, including general hospitals and local community centers; all were screened and did not meet criteria for major depression, schizophrenia, or bipolar illness. Study participants were Han Chinese women with four Han grandparents. Case subjects were ages 30–60 and had at least two episodes of major depression meeting DSM-IV criteria, with the first episode between ages 14 and 50. The study was approved by the ethical review boards of Oxford University and participating hospitals. All participants provided written informed consent. Details on sample collection, phenotypes, and sequencing have been reported elsewhere (13, 14, 28).

Measures of Adversity

A binary measure of adversity was derived from self-reported stressful life events and childhood sexual abuse (see the Methods section in the data supplement that accompanies the online edition of this article). The stressful life events questionnaire, which was adapted from a previous study (29), assessed 16 traumatic events and the respondent's age when they occurred (see Table S1 in the online data supplement). Stressful life events for case subjects were included only if they preceded depression onset. The childhood sexual abuse questionnaire was a shortened version of a scale (30) that queried whether, before age 16, any older person involved the respondent in unwanted sexual incidents, including sexual invitation, fondling, and intercourse. Participants were included in this study if they had non-missing data on stressful life events and childhood sexual abuse questionnaires and were considered "adversity-exposed" if they endorsed any childhood sexual abuse or had high aggregate stressful life event scores (3 standard deviations above the mean). Since life events vary in severity, our score was constructed by weighting each event by its estimated effect size on major depression and summing across events. We note that this approach may cause biases in inference in general, but that is unlikely here (see the Methods section in the data supplement). We thereby grouped subjects into adversity-exposed and unexposed subgroups.

Genome-Wide Association

Genome-wide associations between 4,313,801 imputed autosomal single-nucleotide polymorphisms (SNPs) with minor allele frequency (MAF) >5%, imputation information >0.95, p value for violation of Hardy-Weinberg equilibrium $>10^{-6}$, and major depression was performed in the whole cohort, in those unexposed to adversity (the unexposed group), and in those exposed to adversity (the exposed group) using linear mixed modeling (BOLT-LMM, version 2.2) (31). To calibrate the BOLT-LMM statistic, we calculated linkage disequilibrium (LD) scores of each SNP using LD Score Regression, version 1.0.0 (32). The kinship matrix used was constructed from 413,669 LD pruned SNPs (LD <0.8). PLINK, version 1.9 (33, 34) was used for logistic regression to obtain odds ratios for top variants identified from BOLT. SNPs with p values smaller than 5×10^{-8} were selected for gene-by-environment (G×E) interaction tests. Regional association plots were constructed using LocusZoom, version 0.4.8 (35).

Polygenic Risk Scores

Polygenic risk scores for CONVERGE have previously been constructed by two methods (14, 36). First, using a random independent 50–50 split (sample 1, sample 2) we estimated sample 1 SNP effects using the best linear unbiased prediction method implemented in GCTA (Genome-Wide Complex Trait Analysis) and tested polygenic risk scores constructed using the profile option in PLINK using SNP best linear unbiased prediction solutions as weights in sample 2 and vice versa (CONVERGE-trained polygenic risk score) (14). Second, using summary statistics from the Psychiatric Genomics Consortium meta-analysis of European studies of major depression, recurrent depression polygenic risk scores were constructed from SNP weights based on a p value threshold of 0.2 (36).

Interaction Test

Gene-by-environment interaction effects were tested at loci identified from genome-wide association studies (GWAS). Interaction was tested on both the multiplicative (logistic regression) and additive scales (blm package in R) (37–39), including 10 principal components as covariates to control for population stratification.

Random-Effect Meta-Analysis

Heterogeneity of SNP effects on depression in the adversity-exposed and unexposed groups was tested using random-effect

models that identify both main and heterogeneity effects and Cochran's Q test, all implemented in METASOFT, version 2.0.1 (40, 41).

Heritability Estimation

SNP-based heritability (h^2_{SNP}) was estimated using GCTA, version 1.26.0 (42) with a genetic relatedness matrix (GRM) constructed from 413,669 LD pruned SNPs and 10 principal components used as covariates in the full cohort, the adversity-exposed group, and the unexposed group. Using the bivariate option (43), the genetic correlation (ρ) was estimated for major depression between the adversity-exposed and unexposed groups, for adversity between the major depression group and the control group, and between the major depression group and the adversity group. GCTA was used to estimate the proportion of variance in major depression due to aggregate additive G×E interaction between adversity and all GRM SNPs. Details for ascertainment adjustment (K) and alternative h^2_{SNP} estimation using LDAK (44) and PCGC (45) are provided in the Methods section in the online data supplement.

Simulations of Etiologic Heterogeneity

We used simulations to mirror genetic approaches to discern features of heterogeneity and to demonstrate that stratification of samples by adversity is an appropriate means for uncovering heterogeneous genetic effects (see the Methods section in the data supplement). Three scenarios were applied: in scenario A, SNP effect and adversity exposure contribute additively to liability (no etiologic heterogeneity); in scenario B, SNP effect is only present in the adversity unexposed (reflecting etiologic heterogeneity); and in scenario C, h^2_{SNP} was estimated under the presence and absence of etiologic heterogeneity by replacing the single causal SNP with polygenic contributions.

For each simulation, SNP effects were tested under four logistic regression models:

Model 1: ignoring effects of adversity
Model 2: controlling for effects of adversity by incorporating it as a covariate
Model 3: including an interaction between SNP and adversity
Model 4: analyzing adversity-exposed and unexposed groups separately

For scenarios A and B, we simulated 1,000 independent replicates of a SNP effect (matching SNPs associated in the unexposed group) on a disease with a prevalence of 5% in a cohort with adversity exposure, prevalence, and sample sizes matching those of the CONVERGE cohort. For scenario C, the single causal SNP was replaced with polygenic contributions from 10,000 simulated independent SNPs.

RESULTS

Association Between Adversity and Major Depression

Adversity was significantly associated with major depression, confirming previous analyses (46, 47). Individuals in the depression group experienced significantly more life events than did those in the control group ($p=2.66\times10^{-81}$) (see Figure S1 and Table S1 in the online data supplement). Childhood sexual abuse was significantly associated with major depression (10.3% case subjects compared with 2.5% control subjects; odds ratio=2.98, $p=2.6\times10^{-19}$), with effects increasing with greater abuse severity (see Figure S2 in the data supplement). Together, stressful life events and childhood sexual abuse accounted for 11.6% of the phenotypic variance in major depression. Twenty-seven percent of the sample (1,646 case subjects, 982 control subjects) was adversity exposed, and 73% (3,139 case subjects, 3,832 control subjects) was not. Table S2 in the data supplement lists rates of key clinical features by adversity exposure. Adversity-exposed individuals endorsed higher levels of neuroticism and younger age at onset and were more likely to have comorbid dysthymia and anxiety disorders.

Genome-Wide Association of Major Depression in Groups With and Without Adversity

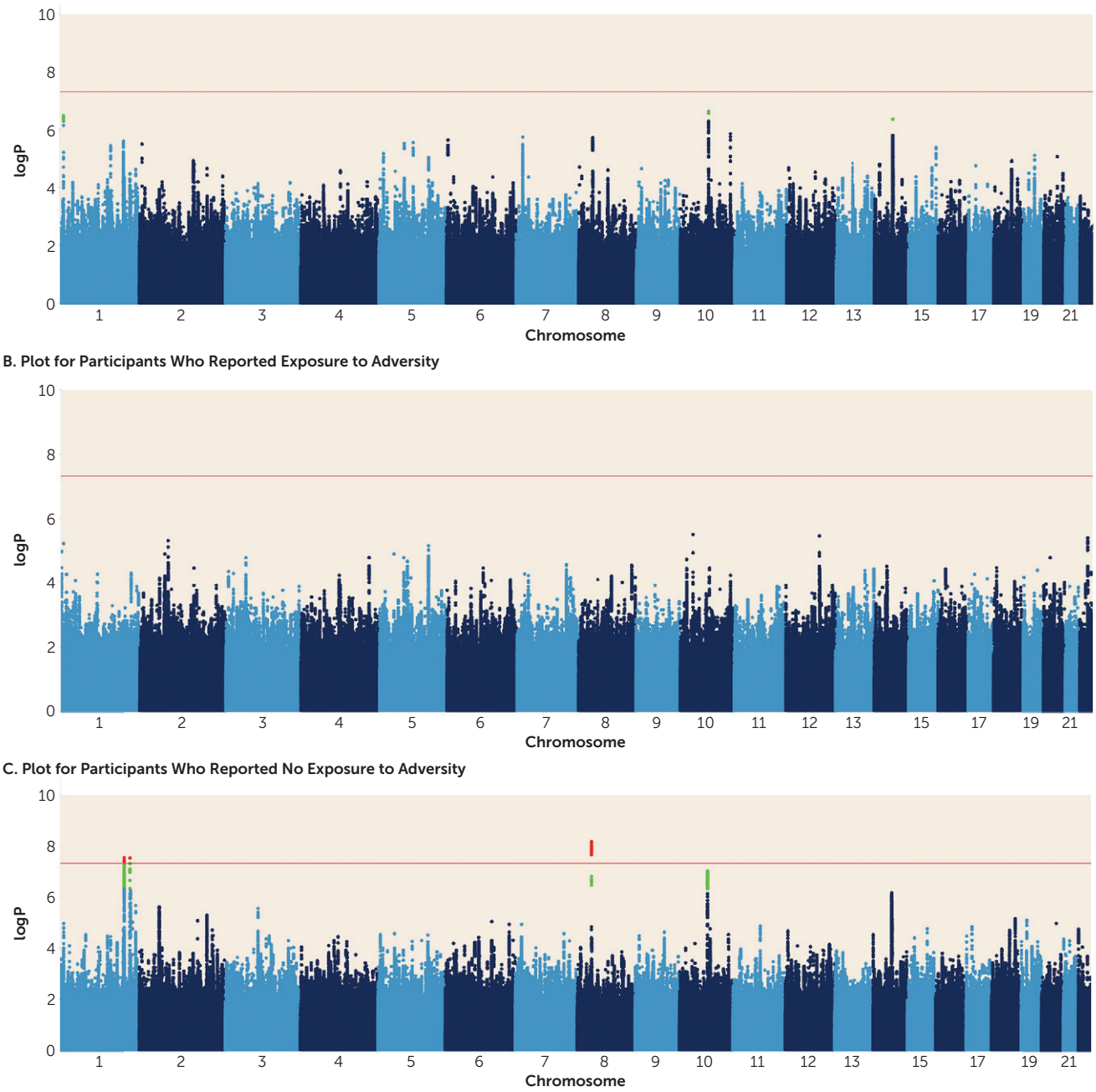
Figure 1 shows Manhattan plots for the GWAS of major depression in CONVERGE participants for whom complete information on adversity was available (N=9,599), those who reported adversity (N=2,628), and those without adversity (N=6,971). The genomic control factors (λ) were 1.047, 1, and 1.047, respectively; the adjusted measures to that of 1,000 cases and 1,000 controls (λ_{1000}) were 1.01, 1, and 1.014, respectively (see Figure S3 in the data supplement).

In the adversity-exposed subset (Figure 1B), no locus exceeded $p<5.0\times10^{-8}$. In the subset without adversity, neither of the two previously reported loci on chromosome 10 exceeded $p<5.0\times10^{-8}$ (rs12415800, $p=3.2\times10^{-6}$; rs35936514, $p=8.7\times10^{-5}$), likely because of reduced power, as odds ratios were not significantly different from the full cohort odds ratios (see Figure S6 and Table S10 in the data supplement). However, three novel loci were detected (Table 1, Figure 1C): on chromosome 1 near *LPGAT1* (lysophosphatidylglycerol acyltransferase 1) (rs7526682, chr1:211973950, MAF=13.3%, $p=3.0\times10^{-8}$, odds ratio=1.31) (Figure 2A); on chromosome 1 in *C1ORF95* (rs11577545, chr1:226799083, MAF=21.5%, $p=3.1\times10^{-8}$, odds ratio=1.25) (Figure 2B); and on chromosome 8 at the 3′ end of *SLC25A37* (Mitoferrin-1) (rs950893, chr8:23450510, MAF=28.0%, $p=6.9\times10^{-9}$, odds ratio=0.79) (Figure 2C).

Comparison of these newly identified loci with the Psychiatric Genomics Consortium mega-analysis of European studies (11) revealed an association between rs950893 on chromosome 8 and major depression (p=0.009), in the same direction as observed in CONVERGE. In contrast, the chromosome 1 loci (rs7526682, rs11577545) were not associated in the Psychiatric Genomics Consortium study (p=0.37, p=0.81), although results were in the same direction (see Figure S4 in the data supplement).

We performed four further analyses on the three newly identified SNPs in the unexposed group. First, to determine whether the results were due to stochastic effects, we randomly removed samples equal in size to the adversity-exposed

FIGURE 1. Manhattan Plots of Loci Associated With Major Depression[a]



[a] The figure shows Manhattan plots of major depression for all subjects for whom information on exposure to adversity was available (panel A), major depression in a subgroup reporting exposure to adversity (panel B), and major depression in a subgroup reporting no exposure to adversity (panel C). In each plot, the $-\log_{10}$ p values of imputed SNPs associated with major depression by leave-one-chromosome-out linear mixed model association in BOLT-LMM are shown on the y-axis. The horizontal axis indicates position on each chromosome, and chromosomes are numbered below the axis.

group 10,000 times and obtained empirical distributions of odds ratios at these SNPs for major depression (see the Methods section in the data supplement). Our results were unlikely to have arisen by chance, as all SNPs showed significant deviation in odds ratio from the full cohort (rs7526682, 99.9th percentile of the empirical distribution of odds ratios; rs11577545, 100th percentile; rs950893, 0.2th percentile) (see Figure S5 in the data supplement). In comparison, the two previously reported SNPs on chromosome 10 (rs12415800, rs35936514) were not significant (see Figure S6 in the data supplement).

Second, we tested for statistical interaction between adversity and the minor allele at each locus and compared findings to results including adversity as a covariate. For the two previously reported SNPs (rs12415800, rs35936514), the interaction terms were not significant. The three newly

TABLE 1. Top SNP Associations With Major Depression in the Full Cohort of Women With Depression and the Subgroups Exposed or Unexposed to Adversity[a]

Model and Measure	Chromosome, rs ID, Position, and Major/Minor Alleles				
	Chr1, rs7526682, 211973950, C/G	Chr1, rs11577545, 226799083, C/T	Chr8, rs950893, 23450510, A/G	Chr10, rs12415800, 69624180, G/A	Chr10, rs35936514, 126244970, C/T
Linear mixed model (BOLT-LMM)					
Full cohort					
MAF	0.13	0.22	0.28	0.46	0.26
p	4.1×10^{-5}	3.2×10^{-5}	1.9×10^{-6}	5.1×10^{-7}	1.8×10^{-6}
z score	4.10	4.16	−4.76	5.02	−4.78
No adversity					
MAF	0.13	0.21	0.28	0.46	0.26
p	**3.0×10^{-8}**	**3.1×10^{-8}**	**6.9×10^{-9}**	3.2×10^{-6}	8.7×10^{-5}
z score	5.54	5.54	−5.79	4.66	−3.93
Adversity					
MAF	0.13	0.23	0.27	0.45	0.27
p	3.8×10^{-1}	9.4×10^{-2}	8.2×10^{-2}	6.0×10^{-2}	2.50×10^{-3}
z score	−0.89	−1.67	0.82	1.88	−3.03
Logistic regression (PLINK)					
Full cohort					
p	3.7×10^{-5}	9.0×10^{-5}	6.9×10^{-7}	9.6×10^{-7}	1.5×10^{-6}
Odds ratio	1.19	1.15	0.85	1.15	0.85
CI	1.10–1.30	1.07–1.23	0.80–0.91	1.09–1.22	0.80–0.91
No adversity					
p	**4.6×10^{-8}**	8.0×10^{-8}	**2.1×10^{-9}**	2.9×10^{-6}	8.8×10^{-5}
Odds ratio	1.31	1.25	0.79	1.17	0.86
CI	1.19–1.45	1.15–1.36	0.74–0.86	1.10–1.26	0.80–0.93
Adversity					
p	3.7×10^{-1}	9.7×10^{-2}	4.6×10^{-1}	5.1×10^{-2}	2.0×10^{-3}
Odds ratio	0.93	0.90	1.05	1.12	0.82
CI	0.78–1.10	0.79–1.02	0.92–1.19	1.00–1.25	0.72–0.93

[a] The table reports the test statistics at the SNPs associated with major depression in the full cohort, in the subgroup unexposed to adversity, and in the adversity-exposed group; the minor allele at each SNP is the tested allele. Results from leave-one-chromosome-out linear mixed model association testing in BOLT-LMM and logistic regression with 10 principal components as covariates in PLINK are shown. SNPs showing genome-wide significant associations are shown in boldface. Two SNPs (rs12415800 and rs35936514) showing genome-wide significant association with major depression in our previous analysis of all samples in CONVERGE (including those without the self-reported adversity measure) are included for comparison.

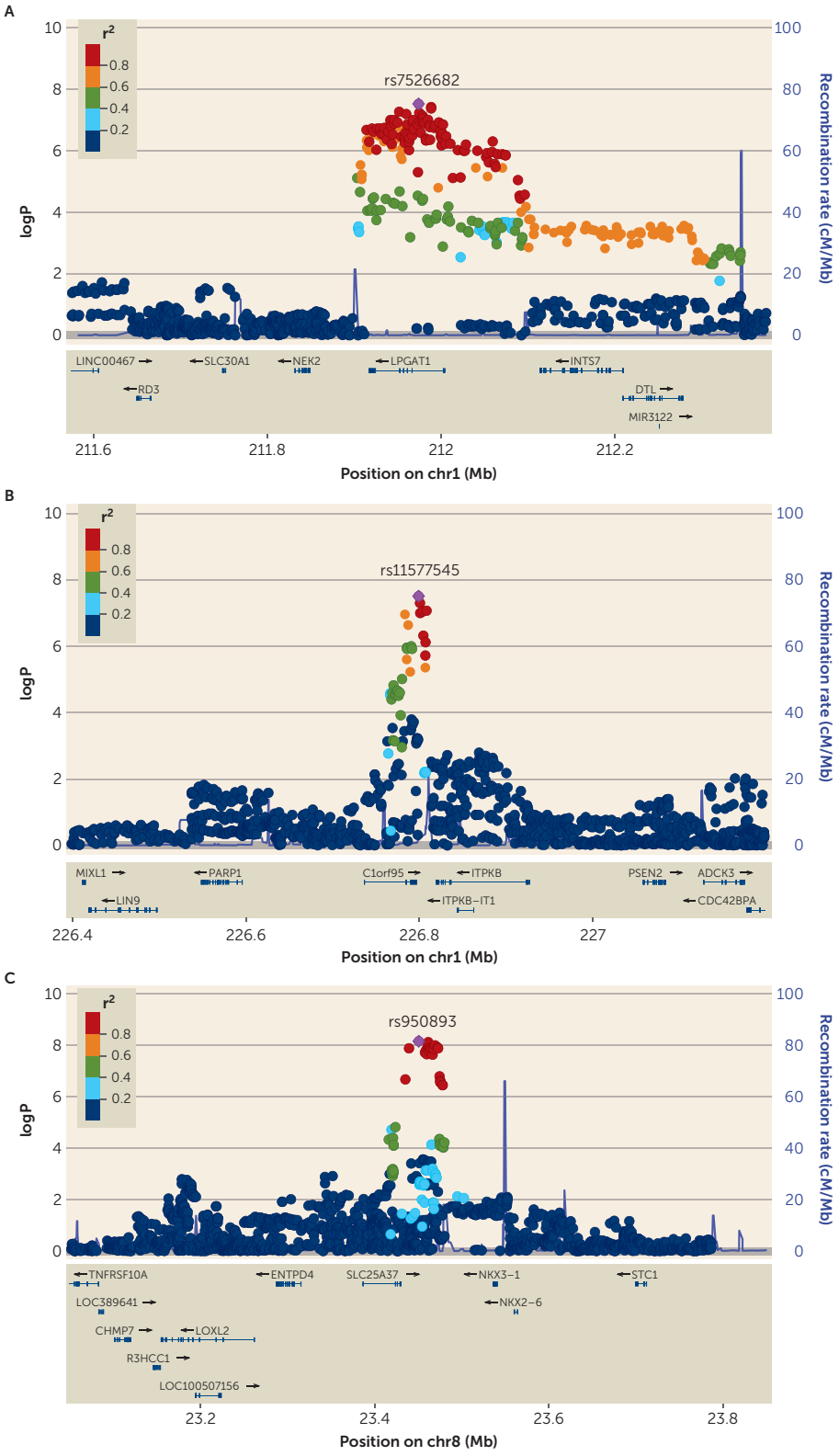
identified SNPs, however, all had significant multiplicative and additive interaction terms (Table 2; see also Table S3 and Figure S7 in the data supplement).

Third, we investigated differences in variant effects in the adversity-exposed and unexposed groups using random-effect meta-analysis. Table S4 in the data supplement shows significant effect size heterogeneity at the three new loci (Q-tests: rs7526682, $p=3.13\times10^{-4}$; rs11577545, $p=9.42\times10^{-6}$; rs950893, $p=1.82\times10^{-4}$) and significant random-effect tests for heterogeneity ($p=1.02\times10^{-7}$, $p=1.07\times10^{-7}$, and $p=2.34\times10^{-8}$, respectively). This method detected significant heterogeneity of SNP effects across the adversity exposure groups for the three newly identified loci. Major depression case–only and control-only association of adversity also demonstrated effect size differences at these variants (see Table S8 in the data supplement).

Fourth, we performed simulations to determine whether the difference in the estimated SNP effects between the adversity-exposed and unexposed groups implicates heterogeneity. The average logistic regression results for scenario A (no heterogeneity) are displayed in the left panel of Table 3. Three results are noteworthy. First, in model 4 (analyzing adversity groups separately), the p values are orders of magnitude less significant than in model 1 (adversity ignored) and model 2 (adversity as covariate). Second, as no heterogeneity is simulated, the p value difference between the two groups in model 4 must only reflect power differences, not heterogeneity. Crucially, this shows that disparate p values between groups alone do not indicate heterogeneity. Third, the G×E interaction test in model 3 is well calibrated and shows no evidence of (false) inflation. These features of homogeneous SNP effects are all evident for both loci on chromosome 10.

Next, we modified the baseline simulation by making the SNP causal only in the adversity-unexposed group (scenario B) and performed the same tests (Table 3, right panel). The presence of heterogeneity induces three novel features: the genetic effect sizes for each group estimated in model 4 are now different; the unexposed group test in model 4 is more powerful than the test in model 1 (ignoring adversity), despite an attendant reduction in sample size; and the G×E interaction

FIGURE 2. Genes at Three Loci Associated With Major Depression in the Subgroup Unexposed to Adversity[a]



[a] Loci are shown in panel A on chromosome 1, 212 Mb, over the gene encoding lysophosphatidylglycerol acyltransferase 1 (*LPGAT1*) (peak SNP=rs7526682), in panel B on chromosome 1, 226 Mb, over the gene *C1ORF95* (peak SNP=rs11577545), and in panel C on chromosome 8 at 23.4 Mb, over the 3′ end of the mitoferrin gene *SLC25A37* (peak SNP=rs950893). The $-\log_{10}$ p values of imputed SNPs associated with major depression by logistic regression are shown on the left y-axis together with the recombination rates (NCBI Build GRCh37), represented by light blue lines, with scales on the right y-axis. Genes within the regions are shown in the bottom panels. The horizontal axis gives the chromosomal position in megabases (Mb). The index SNPs are shown as purple diamonds labeled by their marker names; linkage disequilibrium (hg19 1000 Genomes ASN panel November 2014) with the remaining SNPs is indicated by different colors.

TABLE 2. Tests for Gene-by-Environment Interaction Between Adversity and Genetic Variants Associated With Major Depression[a]

	Model 1: Interaction				Model 2: Covariate		
Test	Odds Ratio	95% CI	p	R^2	Odds Ratio	95% CI	p
Chr1: rs7526682_G	1.29	1.17–1.43	1.94×10^{-7}	0.0024	1.20	1.10–1.31	2.50×10^{-5}
Adversity	2.21	1.99–2.46	2.22×10^{-48}	0.0316	2.03	1.85–2.23	5.21×10^{-51}
Adversity:rs7526682	0.73	0.60–0.89	**0.0016**	0.0013	—	—	—
Chr1: rs11577545_T	1.28	1.17–1.39	$\mathbf{1.48\times10^{-8}}$	0.0018	1.14	1.06–1.23	0.0003
Adversity	2.41	2.15–2.71	4.38×10^{-49}	0.0311	2.02	1.84–2.22	3.11×10^{-50}
Adversity:rs11577545	0.67	0.57–0.79	$\mathbf{9.31\times10^{-7}}$	0.0032	—	—	—
Chr8: rs950893_G	0.80	0.74–0.86	$\mathbf{5.05\times10^{-9}}$	0.0029	0.86	0.81–0.92	3.28×10^{-6}
Adversity	1.74	1.54–1.97	6.47×10^{-19}	0.0311	2.02	1.84–2.22	3.04×10^{-50}
Adversity:rs950893	1.31	1.13–1.52	**0.0003**	0.0018	—	—	—
Chr10: rs12415800_A	1.17	1.10–1.26	2.77×10^{-6}	0.0035	1.16	1.10–1.23	2.87×10^{-7}
Adversity	2.10	1.81–2.45	2.76×10^{-22}	0.0315	2.03	1.85–2.23	1.00×10^{-50}
Adversity:rs12415800	0.96	0.84–1.10	0.5630	<0.0001	—	—	—
Chr10: rs35936514_T	0.85	0.79–0.92	3.61×10^{-5}	0.0037	0.84	0.79–0.90	1.61×10^{-7}
Adversity	2.09	1.85–2.36	2.80×10^{-32}	0.0315	2.03	1.85–2.23	7.43×10^{-51}
Adversity:rs35936514	0.95	0.82–1.10	0.5010	<0.0001	—	—	—

[a] The table reports odds ratios, 95% confidence intervals, and p values of the minor allele of each single-nucleotide polymorphism (SNP) association with major depression in the full cohort in logistic regression, with an interaction term between SNP and self-reported adversity (adversity:SNP) term included in model 1, and without it in model 2. R^2 is the predictive value of each model term reported in terms of Nagelkerke's pseudo R^2. All analyses were performed using 10 principal components as covariates; boldface indicates significant genetic effect ($p<5.0\times10^{-8}$) or gene-by-environment interaction ($p<0.005$).

test in model 3 is statistically significant. These simulation results all distinguish the loci on chromosomes 1 and 8 from those on chromosome 10.

The Genetic Basis for Major Depression in Adversity-Exposed and Unexposed Groups May Differ

To determine the presence of heterogeneity on aggregate genetic effects, estimates of the additive SNP contribution (h^2_{SNP}) on the liability scale, after correction for sample ascertainment, were compared between adversity-exposed and unexposed depressive subgroups. Without etiologic heterogeneity, h^2_{SNP} in subgroups should be similar to that in the entire sample. However, given genetic heterogeneity, h^2_{SNP} may be larger in both subgroups than in the entire sample.

Although the h^2_{SNP} estimate of major depression in the adversity unexposed group (h^2_{SNP}=38.0%, SE=4.8%, $p=1.11\times10^{-16}$) was higher than in the exposed group (h^2_{SNP}=34.2%, SE=15.9%, p=0.013) and the overall combined major depression sample (h^2_{SNP}=30.5%, SE=3.7%, $p<10^{-16}$), they were not statistically different. Because differences in h^2_{SNP} estimation methods may affect estimates and their interpretations (44, 45, 48, 49), we accounted for LD in dense, imputed data using LDAK (44) and assessed underestimation from restricted maximum likelihood using PCGC. These results were consistent with results from GCTA (see Table S5 in the data supplement).

Second, the proportions of variance in major depression due to aggregate additive G×E interaction between adversity and all GRM SNPs, CONVERGE-trained polygenic risk score, and Psychiatric Genomics Consortium polygenic risk scores were estimated. The interaction component for the GRM-by-adversity term was significant (p=0.038), with the proportion of variance attributable to additive genetic (h^2_{SNP}) and G×E interaction components estimated at 23.3% (SE=5.8%) and 13.2% (SE=7.4%), respectively. However, none of the polygenic risk score-by-adversity interactions were significant (see Table S6 in the data supplement), perhaps because of limitations of a polygenic risk score–based approach (see the Methods section in the data supplement).

Third, the genetic correlation of major depression between the adversity-exposed and unexposed groups was estimated at +0.64 (SE=0.23). While less than unity, this is known so imprecisely that it is not significantly different from 1 (95% CI=0.19, 1.0).

Finally, we consider which models of genetic architecture are consistent with observed trends. The resulting h^2_{SNP} estimates from the overall cohort with and without adversity exposure are shown in Figure S9 in the data supplement. The two within-cohort heritabilities, along with genetic correlation and G×E estimates, are shown in Figure S10 in the data supplement. The results confirm our prior intuition: without heterogeneity, within-group heritabilities coincide with the overall average heritability, although the reduced sample sizes induce larger variance in the within-group estimators; however, as heterogeneity increases (or causal variant sharing decreases) the overall heritability decreases while the within-cohort heritabilities remain constant.

Exposure to Adversity May Have a Heterogeneous Genetic Basis

One interpretation of our findings is that the presence of adversity in one group attenuates the contribution of genetic effects. However, self-reported environmental measures are moderately heritable (reviewed in reference 50), and ~29% of the variance in the number of stressful life events has been attributed to SNPs (51). Here, the h^2_{SNP} of adversity was 18.2% for the overall sample (SE=6.2%, p=0.001, K=0.215), 25.7% for the depression group (SE=12.0%, p=0.013, K=0.344), and 44.2% for the control group (SE=14.7%, p=0.001, K=0.20). The genetic correlation of adversity between the depression

TABLE 3. Average Test Output From Four Types of Logistic Regression on 1,000 Simulated Data Sets[a]

	Without Heterogeneity			With Heterogeneity		
Regression Model	Z	Odds Ratio	p	Z	Odds Ratio	p
Model 1, g	5.36	1.18	8.22×10^{-8}	4.67	1.16	3.06×10^{-6}
Model 2, g	5.36	1.18	8.15×10^{-8}	4.99	1.17	6.05×10^{-7}
Model 2, s	13.95	1.90	2.94×10^{-44}	11.14	1.68	7.77×10^{-29}
Model 3, g	4.64	1.19	3.34×10^{-6}	5.85	1.24	4.80×10^{-9}
Model 3, s	10.2	1.92	1.93×10^{-24}	10.22	1.92	1.64×10^{-24}
Model 3, g:s	−0.15	0.99	0.877	−3.08	0.8	2.10×10^{-3}
Model 4, g, no adversity	4.64	1.19	3.44×10^{-6}	5.85	1.24	4.80×10^{-9}
Model 4, g, adversity	2.69	1.18	7.14×10^{-3}	−0.08	1.00	0.94

[a] Model 1 ignores adversity; model 2 controls for the additive effect of adversity; model 3 additionally incorporates an interaction between genotype and adversity; and model 4 analyzes adversity-exposed and unexposed groups separately. For each row, Z statistic, odds ratio, and p value are shown for SNP effect (g), adversity effect (s), or an interaction effect (g:s). The three columns under "Without Heterogeneity" list results for simulations of no etiologic heterogeneity between simulated phenotype in samples with and without adversity; the three columns under "With Heterogeneity" list these data for simulations with heterogeneity. Data were simulated using a liability threshold model with realistic ascertainment, effect sizes, and allele frequencies (see the Methods section in the online data supplement).

and control groups was +0.34 (SE=0.31) but was not statistically significant.

Assessment of G-E Correlation

Since G-E correlation can bias G×E results, we tested for G-E correlation by three methods. The estimated SNP-based genetic correlation between major depression and self-reported adversity was negligible, as ρ was −0.02 (SE=0.15) and not significantly different from 0 (p=0.45). Additional tests of G-E correlation examined association of major depression polygenic risk scores with adversity and were not significant (see Table S7 in the data supplement). An exploratory test of G-E correlation examined by genome-wide correlation of SNP odds ratios for adversity between the depression and control groups was also small (r=0.008) (see Table S9 and Figure S8 in the data supplement). These results do not support significant systematic G-E correlation in our sample between major depression and adversity exposure.

DISCUSSION

We applied molecular genetic methods to a large sample of carefully characterized depressed women to evaluate etiologic heterogeneity between those exposed and those unexposed to severe environmental adversities. These efforts yielded three major findings. First, classifying samples based on adversity exposure identified genetic loci with heterogeneous effects. We identified three novel loci on chromosomes 1 and 8 that confer risk of major depression only among individuals unexposed to adversity. The newly discovered locus on chromosome 8 is at the 3′ end of the SLC25A37 gene, which encodes an iron carrier localized in the mitochondrial inner membrane (52), adding further support for a mitochondrial role in major depression (13, 14, 53). Second, we found evidence for interaction between adversity and genotype at all three loci. Third, we provide modest evidence for heterogeneity at the whole-genome level: 13.2% of the variance in major depression liability arises from interaction between genome-wide SNP effects and adversity; genetic correlation for major depression between subgroups with and without adversity exposure was +0.64; and although confidence intervals overlapped, SNP-based heritability estimates of major depression in the unexposed subgroup was higher (~39%) than in the overall sample (30%). Furthermore, simulations reflecting etiologic heterogeneity are consistent with our results.

These results have several implications. First, they provide support for the long-debated typology that major depression patients can be meaningfully divided into those whose illness arises in reaction to environmental stressors and those whose disorder emerges "from within" (5–8). The genetic substrates of these two forms of major depression appear to be correlated but not identical, and some genetic factors may have subtype-specific effects.

Second, these findings provide insight into how effects of genes and environment combine to give rise to major depression. A leading hypothesis consistent with previous studies (24, 25, 27) is that certain genes have a stronger impact on risk for major depression in adversity-exposed than in unexposed individuals. We unexpectedly found for three loci an opposite pattern in which effects were *stronger* in cases without adversity exposure. While the CONVERGE sample may contain loci with an increased impact on adversity-exposed individuals, power to detect these is low, as only 27% of our sample reported severe adversity.

An appealing interpretation of our findings is that absent environmental stressors, a higher genetic loading is required to cause depression. This cannot, however, explain our findings, as it would predict a graded response at the three identified loci in exposed and unexposed individuals. Rather, our results suggest at least two classes of molecular variants that predispose to major depression: those whose effects are present in all cases and those whose effects depend on the history of adversity. In contrast to the three SNPs discovered by stratifying on adversity, results for the two previously reported SNPs on chromosome 10 (13) are consistent with a liability threshold model. Major depression may be a syndrome arising from several partially distinct etiologies. The design of CONVERGE enabled the combination of deep phenotypes and genotypes to detect differences in the genetic architecture of those with and without exposure to adversity. Other subtypes may be detectable in a similar manner.

Our findings counter the dominant paradigm in psychiatric molecular genetics research that increasing sample size should be the primary method for detecting more genetic loci

(12). Here, the newly detected loci were discovered in a sample size 30% smaller than the cohort that yielded the two previously reported loci (13). These results support the value of more detailed phenotyping, especially the assessment of environmental adversities. To characterize major depression etiology, future efforts may need careful assessment of both the phenotype and environmental exposures in large samples.

Three limitations to the study should be noted. First, our power to detect genetic variants with the expected small effect sizes is limited (see Table S10 in the data supplement). Second, we are unaware of Asian replication cohorts with genetic information and environmental adversities. Therefore, these results should be considered tentative, although their plausibility is supported by simulations. Third, our assessments of age at onset of depression and adversity exposure were retrospective. Although interviewers encouraged effortful responding, we cannot rule out recall biases. Despite these limitations, our results highlight the value of empirically driven approaches to addressing heterogeneity and provide a framework applicable to other complex psychiatric diseases to identify putative subtypes and etiologically relevant genetic variation.

AUTHOR AND ARTICLE INFORMATION

From the Virginia Institute for Psychiatric and Behavioral Genetics and the Department of Psychiatry, Virginia Commonwealth University, Richmond, Va.; the Wellcome Trust Sanger Institute, Cambridge, U.K.; the European Bioinformatics Institute (EMBL-EBI), Cambridge, U.K.; the Department of Medicine, University of California, San Francisco; the State University of New York Downstate Medical Center, Brooklyn, N.Y.; and the Center for Neurobehavioral Genetics, Semel Institute for Neuroscience and Human Behavior, University of California, Los Angeles.

Address correspondence to Dr. Kendler (kenneth.kendler@vcuhealth.org) or Dr. Flint (jflint@mednet.ucla.edu).

Drs. Peterson and Cai contributed equally to this work. Drs. Flint and Kendler are joint senior authors.

This work was funded by the Wellcome Trust (WT090532/Z/09/Z, WT083573/Z/07 /Z, WT089269/Z/09/Z) and by NIH grant MH100549. Dr. Peterson was supported by NIH T32 grant MH020030; Dr. Cai was supported by the ESPOD Fellowship from the European Bioinformatics Institute (EMBL-EBI) and Wellcome Trust Sanger Institute; Dr. Edwards was supported by NIH K01 grant AA021399; and Dr. Bacanu was supported by NIMH grants R21MH100560 and R21AA022717.

The authors, who are part of the CONVERGE (China, Oxford, and VCU Experimental Research on Genetic Epidemiology) consortium, gratefully acknowledge the support of all partners in hospitals across China, with special thanks to all the CONVERGE collaborators and patients who made this work possible.

The authors report no financial relationships with commercial interests.

Received June 6, 2017; revisions received Sept. 21 and Nov. 17, 2017; accepted Nov. 30, 2017; published online March 2, 2018.

REFERENCES

1. Flint J, Kendler KS: The genetics of major depression. Neuron 2014; 81:484–503
2. Goldberg D: The heterogeneity of "major depression". World Psychiatry 2011; 10:226–228
3. Milaneschi Y, Lamers F, Peyrot WJ, et al: Polygenic dissection of major depression clinical heterogeneity. Mol Psychiatry 2016; 21:516–522
4. Power RA, Tansey KE, Buttenschøn HN, et al: Genome-wide association for major depression through age at onset stratification: Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium. Biol Psychiatry 2016; 81:325–335
5. Mendels J, Cochrane C: The nosology of depression: the endogenous-reactive concept. Am J Psychiatry 1968; 124:1–11
6. Coryell W, Winokur G, Maser JD, et al: Recurrently situational (reactive) depression: a study of course, phenomenology and familial psychopathology. J Affect Disord 1994; 31:203–210
7. Nelson JC, Charney DS: Primary affective disorder criteria and the endogenous-reactive distinction. Arch Gen Psychiatry 1980; 37:787–793
8. Nelson JC, Charney DS: The symptoms of major depressive illness. Am J Psychiatry 1981; 138:1–13
9. Pollitt J: The relationship between genetic and precipitating factors in depressive illness. Br J Psychiatry 1972; 121:67–70
10. Sullivan PF, Neale MC, Kendler KS: Genetic epidemiology of major depression: review and meta-analysis. Am J Psychiatry 2000; 157: 1552–1562
11. Ripke S, Wray NR, Lewis CM, et al: A mega-analysis of genome-wide association studies for major depressive disorder. Mol Psychiatry 2013; 18:497–511
12. Hyde CL, Nagle MW, Tian C, et al: Identification of 15 genetic loci associated with risk of major depression in individuals of European descent. Nat Genet 2016; 48:1031–1036
13. CONVERGE consortium: Sparse whole-genome sequencing identifies two loci for major depressive disorder. Nature 2015; 523:588–591
14. Peterson RE, Cai N, Bigdeli TB, et al: The genetic architecture of major depressive disorder in Han Chinese women. JAMA Psychiatry 2017; 74:162–168
15. Lubke GH, Hottenga JJ, Walters R, et al: Estimating the genetic variance of major depressive disorder due to all single nucleotide polymorphisms. Biol Psychiatry 2012; 72:707–709
16. Kessler RC: The effects of stressful life events on depression. Annu Rev Psychol 1997; 48:191–214
17. Brown GW, Harris T: Social Origins of Depression: A Study of Psychiatric Disorder in Women. London, Tavistock, 1978
18. Kendler KS, Bulik CM, Silberg J, et al: Childhood sexual abuse and adult psychiatric and substance use disorders in women: an epidemiological and cotwin control analysis. Arch Gen Psychiatry 2000; 57:953–959
19. Kendler KS, Karkowski LM, Prescott CA: Stressful life events and major depression: risk period, long-term contextual threat, and diagnostic specificity. J Nerv Ment Dis 1998; 186:661–669
20. Kendler KS, Karkowski LM, Prescott CA: Causal relationship between stressful life events and the onset of major depression. Am J Psychiatry 1999; 156:837–841
21. Cole J, McGuffin P, Farmer AE: The classification of depression: are we still confused? Br J Psychiatry 2008; 192:83–85
22. Farmer A, McGuffin P: The classification of the depressions: contemporary confusion revisited. Br J Psychiatry 1989; 155:437–443
23. Kendell RE: The classification of depressions: a review of contemporary confusion. Br J Psychiatry 1976; 129:15–28
24. Kendler KS, Kuhn J, Prescott CA: The interrelationship of neuroticism, sex, and stressful life events in the prediction of episodes of major depression. Am J Psychiatry 2004; 161:631–636
25. Heath AC, Eaves LJ, Martin NG: Interaction of marital status and genetic risk for symptoms of depression. Twin Res 1998; 1:119–122
26. Phelan J, Schwartz JE, Bromet EJ, et al: Work stress, family stress, and depression in professional and managerial employees. Psychol Med 1991; 21:999–1012
27. Kendler KS, Kessler RC, Walters EE, et al: Stressful life events, genetic liability, and onset of an episode of major depression in women. Am J Psychiatry 1995; 152:833–842
28. Cai N, Bigdeli TB, Kretzschmar WW, et al: 11,670 whole-genome sequences representative of the Han Chinese population from the CONVERGE project. Sci Data 2017; 4:170011
29. Kendler K, Prescott C: Genes, Environment, and Psychopathology: Understanding the Causes of Psychiatric and Substance Use Disorders. New York, Guilford Press, 2006

30. Martin J, Anderson J, Romans S, et al: Asking about child sexual abuse: methodological implications of a two stage survey. Child Abuse Negl 1993; 17:383–392
31. Loh PR, Tucker G, Bulik-Sullivan BK, et al: Efficient Bayesian mixed-model analysis increases association power in large cohorts. Nat Genet 2015; 47:284–290
32. Bulik-Sullivan BK, Loh PR, Finucane HK, et al: LD score regression distinguishes confounding from polygenicity in genome-wide association studies. Nat Genet 2015; 47:291–295
33. Chang CC, Chow CC, Tellier LC, et al: Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience 2015; 4:7
34. Purcell S, Neale B, Todd-Brown K, et al: PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 2007; 81:559–575
35. Pruim RJ, Welch RP, Sanna S, et al: LocusZoom: regional visualization of genome-wide association scan results. Bioinformatics 2010; 26:2336–2337
36. Bigdeli TB, Ripke S, Peterson RE, et al: Genetic effects influencing risk for major depressive disorder in China and Europe. Transl Psychiatry 2017; 7:e1074
37. Kendler KS, Gardner CO: Interpretation of interactions: guide for the perplexed. Br J Psychiatry 2010; 197:170–171
38. R Development Core Team. R: A language and environment for statistical computing; version 2.9.2. 2009
39. Kovalchik S, Varadhan R: Fitting additive binomial regression models with the R package blm. J Stat Softw 2013; 54:1–18
40. Han B, Eskin E: Random-effects model aimed at discovering associations in meta-analysis of genome-wide association studies. Am J Hum Genet 2011; 88:586–598
41. Kang EY, Han B, Furlotte N, et al: Meta-analysis identifies gene-by-environment interactions as demonstrated in a study of 4,965 mice. PLoS Genet 2014; 10:e1004022
42. Yang J, Lee SH, Goddard ME, et al: GCTA: a tool for genome-wide complex trait analysis. Am J Hum Genet 2011; 88:76–82
43. Lee SH, Yang J, Goddard ME, et al: Estimation of pleiotropy between complex diseases using single-nucleotide polymorphism-derived genomic relationships and restricted maximum likelihood. Bioinformatics 2012; 28:2540–2542
44. Speed D, Hemani G, Johnson MR, et al: Improved heritability estimation from genome-wide SNPs. Am J Hum Genet 2012; 91:1011–1021
45. Golan D, Lander ES, Rosset S: Measuring missing heritability: inferring the contribution of common variants. Proc Natl Acad Sci USA 2014; 111:E5272–E5281
46. Chen J, Cai Y, Cong E, et al: Childhood sexual abuse and the development of recurrent major depression in Chinese women. PLoS One 2014; 9:e87569
47. Tao M, Li Y, Xie D, et al: Examining the relationship between lifetime stressful life events and the onset of major depression in Chinese women. J Affect Disord 2011; 135:95–99
48. Lee SH: Implications of simplified linkage equilibrium SNP simulation. Proc Natl Acad Sci USA 2015; 112:E5449–E5451
49. Golan D, Rosset S, Lander ES: Reply to Lee: Downward bias in heritability estimation is not due to simplified linkage equilibrium SNP simulation. Proc Natl Acad Sci USA 2015; 112:E5452–E5453
50. Kendler KS, Baker JH: Genetic influences on measures of the environment: a systematic review. Psychol Med 2007; 37:615–626
51. Power RA, Wingenbach T, Cohen-Woods S, et al: Estimating the heritability of reporting stressful life events captured by common genetic variants. Psychol Med 2013; 43:1965–1971
52. Chen W, Paradkar PN, Li L, et al: Abcb10 physically interacts with mitoferrin-1 (Slc25a37) to enhance its stability and function in the erythroid mitochondria. Proc Natl Acad Sci USA 2009; 106:16263–16268
53. Cai N, Chang S, Li Y, et al: Molecular signatures of major depression. Curr Biol 2015; 25:1146–1156