

No Support for Historical Candidate Gene or Candidate Gene-by-Interaction Hypotheses for Major Depression Across Multiple Large Samples

Richard Border, M.A., Emma C. Johnson, Ph.D., Luke M. Evans, Ph.D., Andrew Smolen, Ph.D., Noah Berley, Patrick F. Sullivan, M.D., Matthew C. Keller, Ph.D.

Objective: Interest in candidate gene and candidate gene-by-environment interaction hypotheses regarding major depressive disorder remains strong despite controversy surrounding the validity of previous findings. In response to this controversy, the present investigation empirically identified 18 candidate genes for depression that have been studied 10 or more times and examined evidence for their relevance to depression phenotypes.

Methods: Utilizing data from large population-based and case-control samples (Ns ranging from 62,138 to 443,264 across subsamples), the authors conducted a series of pre-registered analyses examining candidate gene polymorphism main effects, polymorphism-by-environment interactions, and gene-level effects across a number of operational definitions of depression (e.g., lifetime diagnosis, current severity, episode recurrence) and environmental moderators (e.g., sexual or physical abuse during childhood, socioeconomic adversity).

Results: No clear evidence was found for any candidate gene polymorphism associations with depression phenotypes or any polymorphism-by-environment moderator effects. As a set, depression candidate genes were no more associated with depression phenotypes than noncandidate genes. The authors demonstrate that phenotypic measurement error is unlikely to account for these null findings.

Conclusions: The study results do not support previous depression candidate gene findings, in which large genetic effects are frequently reported in samples orders of magnitude smaller than those examined here. Instead, the results suggest that early hypotheses about depression candidate genes were incorrect and that the large number of associations reported in the depression candidate gene literature are likely to be false positives.

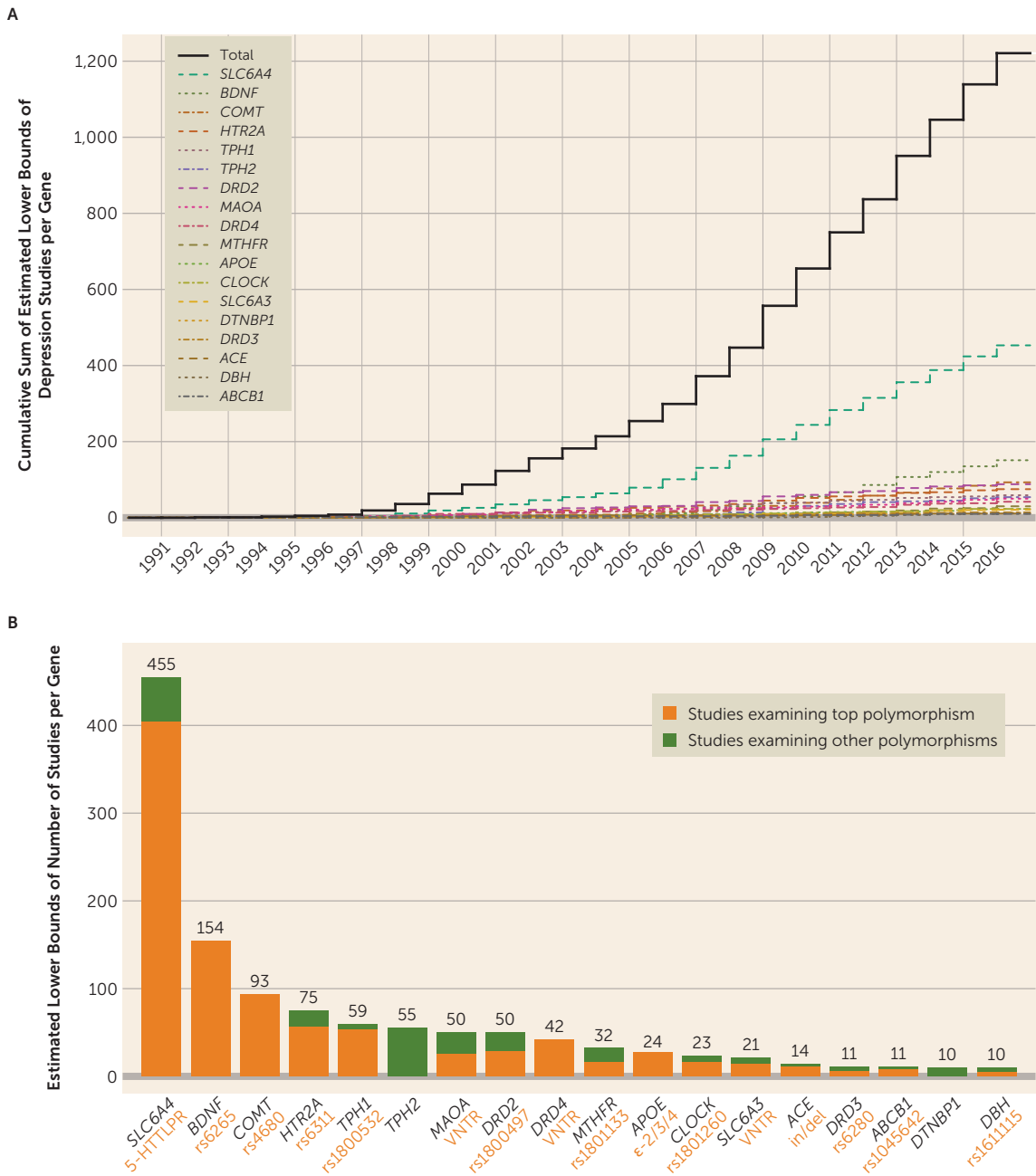
AJP in Advance (doi: 10.1176/appi.ajp.2018.18070881)

Major depressive disorder (hereafter referred to as “depression”) is moderately heritable (twin-based heritability, ~37%) (1), but its genetic architecture is complex, and identifying specific polymorphisms underlying depression susceptibility has been challenging. With the ability to genotype particular genetic variants and optimism about the potential public health impact of identifying reliable biomarkers for depression (2), early research focused on the effects of specific candidate polymorphisms in genes hypothesized to underlie depression liability. These genes were chosen on the basis of hypotheses regarding the biological underpinnings of depression. The 5-HTTLPR variable number tandem repeat (VNTR) polymorphism in the promoter region of the serotonin transporter gene *SLC6A4*, the most commonly studied polymorphism in relation to depression (Figure 1; see also Table S1.1 in the online supplement), serves as a prototypical example: Given the theorized importance of the serotonergic system in the etiology of depression, a logical target for early association studies was a

common, large (and hence relatively easy to genotype), and potentially functional repeat polymorphism in a serotonergic gene (3–5). Early investigations, although by necessity focused on a small number of variants (low-cost genome-wide arrays were not yet available), reported promising positive associations. However, replication attempts produced inconsistent results (6–8).

To critics of candidate gene findings, replication failures suggested that the initial findings were artifactual (9–11). However, at least two alternative explanations could account for the inability to replicate early findings and the inconsistent results across studies. First, in the early 2000s, Caspi et al. (12) posited that previous inconsistencies might reflect the effects of candidate polymorphisms that were dependent on environment exposures (gene-by-environment interaction [$G \times E$] effects). In what would become one of the most highly cited (>8,000 citations as of July 2018) and influential papers in psychiatric genetics, Caspi et al. (13) reported that the impact of the 5-HTTLPR repeat

FIGURE 1. Estimated lower bounds of studies per candidate gene^a



^a Panel A shows cumulative sums of the estimated number of depression candidate gene studies identified by our algorithm per year per gene from 1991 through 2016. Estimates reflect the number of correctly classified studies among identified studies, excluding studies not detected by our protocol, and thus comprise lower bounds for the true number of studies per gene. Panel B shows the 18 candidate genes studied ≥ 10 times between 1991 and 2016. The estimated number of studies focused on the top polymorphism (see Table S1.1 in the online supplement) is displayed relative to the other identified studies within each gene. No top polymorphisms were identified for *DTNBP1* or *TPH2* (see section S1 of the online supplement).

polymorphism in *SLC6A4* on depression was moderated by exposure to stressful life events, such that the positive association between stressful life events and depression was stronger in individuals carrying the “short” allele. This early work led many researchers to shift their attention to G×E hypotheses, focusing on the same polymorphisms first investigated for main effects (8). Second, in an alternative but complementary line of reasoning, other researchers

suggested that polymorphisms in the same candidate genes, other than those studied previously, were likely to explain depression risk, given the genes’ putative biological relevance (14). These lines of inquiry are well represented in the literature of the past 25 years. Thousands of investigations of depression or depression endophenotypes have examined 1) the direct effects of the most studied polymorphisms within candidate genes, 2) the moderation of their effects by

environmental stressors, or 3) the effects of alternative polymorphisms within the same candidate genes. The popularity of these lines of inquiry has not diminished over time (Figure 1; see also Figures S1.4 and S1.5 of the online supplement), and many studies have reported statistically significant associations.

Perhaps surprisingly given the continued interest in studying these historical depression candidate genes and the large number of associations documented in the candidate gene literature, many researchers have expressed skepticism about the validity of such findings (11, 15–17). There are several reasons for this. First, genome-wide association studies (GWASs), which agnostically examine associations at millions of common single-nucleotide polymorphisms (SNPs) across the genome in large samples, have consistently found that individual SNPs exert small effects on genetically complex traits such as depression (18–20). For example, in the most recent GWAS of depression, which utilized a sample of 135,458 case subjects and 344,901 control subjects, the strongest individual signal detected (rs12552; odds ratio=1.044, $p=6.07 \times 10^{-19}$) would require a sample of approximately 34,100 individuals to be detected with 80% power at an alpha level of 0.05, assuming a balanced case-control design (18). In contrast, the median study sample size in a review of 103 candidate $G \times E$ studies published between 2000 and 2009 was 345, with 65% of studies reporting positive results (15). Thus, given the small sample sizes typically employed, candidate gene research has likely been severely underpowered (21, 22). This, in turn, may suggest that the false discovery rate for the many positive reports in the candidate gene literature is high. Consistent with this possibility, targeted, well-powered genetic association studies of depression and other psychiatric phenotypes in large samples have not supported candidate gene hypotheses (18, 23–27). For example, a preregistered collaborative meta-analysis of the interaction of stressful life events and 5-HTTLPR genotype in a sample of 38,802 individuals failed to support the original finding of Caspi et al. (28), although we note that this variant and several other candidate VNTRs have not previously been examined in a GWAS context (29, 30). The absence of previous large-sample investigations of VNTR hypotheses is noteworthy, as VNTRs comprise several of the earliest candidate polymorphisms to be examined in the context of behavioral research; concerns about variability in VNTR genotyping procedures and analytic methods over time have further complicated the interpretation of the literature (31). Additionally, a number of researchers have suggested that incorrect analytic methods and inadequate control for population stratification characterize the majority of published candidate gene studies (21, 32–34), and other researchers have questioned the clinical utility of focusing on individual polymorphisms or polymorphism-by-environment interactions (35). Finally, there is evidence of systematic publication bias in the candidate gene literature; in the aforementioned review of all candidate $G \times E$ studies published between 2000 and 2009, 96% percent of novel

findings were significant, compared with only 27% of replication attempts, and replication attempts reporting null findings had larger sample sizes than those presenting positive findings (15). In response to such skepticism, candidate gene proponents have argued that lack of replication of candidate gene associations in large-sample studies may reflect poor or limited phenotyping (36–38), exclusion of non-SNP polymorphisms such as VNTRs (14, 30), the “multiple-testing burden” associated with genome-wide scans (36), and failure to account for environmental moderators (36, 37, 39).

The present study is the most comprehensive and well-powered investigation of historical candidate polymorphism and candidate gene hypotheses in depression to date. We focus on three lines of inquiry concerning how historical candidate genes may affect depression liability: 1) main effects of the most commonly studied candidate polymorphisms, 2) moderation of the effects of these polymorphisms by environmental exposures, and 3) main effects of common SNPs across each of the candidate genes.

We first empirically identified 18 commonly studied candidate genes represented in at least 10 peer-reviewed depression-focused journal articles between 1991 and 2016 from the body of publications indexed in PubMed. Within these candidate genes, we identified the most commonly studied polymorphisms, as well as their canonical risk alleles, at which point our primary analysis plan was preregistered. Using multiple large samples (N s ranging from 62,138 to 443,264 across subsamples; total $N=621,214$ individuals), we examined multiple measures of depression (e.g., lifetime diagnostic status, symptom severity among individuals reporting mood disturbances, lifetime number of depressive episodes) (Table 1), employing multiple statistical frameworks (e.g., main effects of polymorphisms and genes, interaction effects on both the additive and multiplicative scales) and, in $G \times E$ analyses, considering multiple indices of environmental exposure (e.g., traumatic events in childhood or adulthood). Previous large-sample studies of depression have largely focused on genetic main effects on depression diagnosis in the context of SNP data across the genome. In contrast, we examined several alternative depression phenotypes, analyzed both main effects and interactions with multiple potential moderators, included the most studied polymorphisms, including VNTRs (Figure 1), and employed a liberal significance threshold. We also quantified the extent to which phenotypic measurement error may have biased our results. The unifying question underlying this “multiverse” analytic approach (44) was the following: Do the large data sets of the whole-genome-data era support any previous depression candidate gene hypotheses?

METHODS

Identification of Genes and Polymorphisms

Using the Biopython bioinformatics package (45), we identified 18 candidate genes studied for their associations with depression phenotypes at least 10 times from within the body

TABLE 1. Depression and environmental moderator phenotypes

Phenotype	Description	Sample Size
Depression phenotypes^a		
Estimated lifetime depression diagnosis	Binary indicator of lifetime DSM-5 depression diagnosis assessed in the UK Biobank online mental health follow-up questionnaire. To meet criteria, participants had to endorse at least four of eight DSM-5 depression symptoms (motor agitation/retardation was not assessed), as well as duration, frequency, and impairment criteria.	N=115,458 (control subjects: 85,513; case subjects: 29,945)
Current depression severity	Sum score of all nine DSM-5 depression symptom severities (using a 4-point Likert scale to index the severity of each symptom) over the 2 weeks preceding to assessment. Assessed in the UK Biobank online mental health follow-up questionnaire.	N=115,463 (mean=2.502, SD=3.347)
Conditional lifetime symptom count	Sum of symptom indicators for eight of nine lifetime DSM-5 depression symptoms (motor agitation/retardation was not assessed) among individuals endorsing lifetime incidence of a period of at least 2 weeks characterized by anhedonia and/or depressed mood (questionnaire skip patterns necessitated this precondition). Assessed in the UK Biobank online mental health follow-up questionnaire.	N=62,138 (mean=4.746, SD=1.745)
Lifetime episode count	Ordinal measure of incidence/recurrence of a period of at least 2 weeks characterized by anhedonia and/or depressed mood indicating zero episodes, a single episode, or recurrent episodes. Assessed in the UK Biobank online mental health follow-up questionnaire.	N=115,457 (zero: 55,388; single: 30,724; recurrent: 26,345)
Touchscreen probable lifetime diagnosis, ordinal classification	Ordinal measure of depression diagnostic status based on a selection of items from the Patient Health Questionnaire (40), the Structured Clinical Interview for DSM-IV Axis I Disorders—Research Version (41), and items assessing treatment-seeking behavior specific to the UK Biobank touchscreen interview, as described in Smith et al. (42). Categories included no depression, single depressive episode, recurrent episodes (moderate), and recurrent episodes (severe), in that order. Assessed as part of the UK Biobank initial touchscreen interview.	N=91,121 (control subjects: 66,605; one episode: 6,209; ≥2 moderate episodes: 11,634; ≥2 severe episodes: 6,633)
Touchscreen probable lifetime diagnosis	Dichotomized coding of the touchscreen probable life diagnosis ordinal classification, contrasting no depression with the three diagnosis categories.	N=91,121 (control subjects: 66,605; case subjects: 84,516)
Severe recurrent depression	Binary indicator of case/control status for depression, excluding case and control subjects with mild to moderate depressive symptoms. Control subjects were individuals who did not endorse incidence of a period of at least 2 weeks characterized by anhedonia and/or depressed mood. Case subjects were individuals who met criteria for estimated lifetime depression diagnosis, endorsed at least five of the eight measured DSM-5 symptoms, and experienced recurrent depressive episodes. Assessed in the UK Biobank online mental health follow-up questionnaire.	N=64,432 (control subjects: 53,218; case subjects: 14,214)
PGC lifetime depression diagnosis	Binary indicator of lifetime depression diagnosis as measured in the PGC2 depression GWAS (18). The present study utilized data from the full expanded cohort meta-analysis, excepting UK-based cohorts (UK Biobank and Generation Scotland).	N=443,264 (control subjects: 323,063; case subjects: 120,201)
Moderator phenotypes^b		
Childhood trauma	Binary indicator of sexual and/or physical abuse during childhood. Assessed in the UK Biobank online mental health follow-up questionnaire.	N=157,146 (unexposed: 118,800; exposed: 38,346)
Adulthood trauma	Binary indicator of any of the following traumatic events during adulthood: physical assault, sexual assault, witness to sudden/violent death, diagnosis of a life-threatening illness, involvement in a life-threatening accident, and exposure to combat or war zone conditions. Assessed in the UK Biobank online mental health follow-up questionnaire.	N=157,223 (unexposed: 64,286; exposed: 92,937)
Recent trauma	Binary indicator of whether any of the above events occurred in the year preceding assessment.	N=157,220 (unexposed: 142,008; exposed: 15,212)
Stressor-induced depression	Binary indicator of whether a period of depressed mood or anhedonia was a possible consequence of a traumatic event among individuals endorsing lifetime incidence of a period of at least 2 weeks characterized by anhedonia and/or depressed mood (questionnaire skip patterns necessitated this precondition). Assessed in the UK Biobank online mental health follow-up questionnaire.	N=88,585 (unrelated to stressor: 23,746; stressor-induced: 64,839)
Townsend deprivation index	Measure of socioeconomic adversity (43), with higher values indicating greater adversity. Standardized to have zero mean and unit standard deviation. Assessed during the UK Biobank initial touchscreen interview.	N=187,094

^a Depression phenotypes are described in further detail in section S3.1 and visually summarized in Figure S3.1 in the online supplement.^b Moderator phenotypes are described in further detail in section S3.2 and visually summarized in Figure S3.2 in the online supplement. All moderators were only measured in the UK Biobank.

of peer-reviewed biomedical literature indexed in PubMed. We used regular expressions to find articles potentially corresponding to each gene and hand-verified the number of correctly classified articles for each gene in order to estimate hypergeometric confidence intervals for the true number of correctly classified studies (for additional details, see section S1 of the online supplement). We identified single polymorphisms comprising a large proportion of study foci for 16 of the 18 candidate genes. Figure 1 lists the most studied candidate genes and polymorphisms within them, as well as probabilistic estimates of the minimum number of times each has been studied with respect to depression and the number of studies per gene per year (confidence intervals are presented in Table S1.1 in the online supplement).

Samples

UK Biobank samples. A large portion of the data used in our analysis was collected by the UK Biobank, a population sample of 502,682 individuals collected at 22 centers across the United Kingdom between 2006 and 2010 (46). Within this group, we analyzed several depression phenotypes and moderators among 177,950 unrelated (pairwise genome-wide relatedness, <0.05) European-ancestry individuals for whom relevant depression measures were collected. We analyzed two partially overlapping subsets of these individuals: 91,121 individuals for whom selected items from the initial touchscreen interview were available and 115,458 individuals who completed a series of online mental health questionnaires, 62,138 of whom endorsed a 2-week period characterized by anhedonia or depressed mood at some point during their lives. DNA was extracted from whole blood and genotyped using the Affymetrix UK Biobank Axiom array or the Affymetrix UK BiLEVE Axiom array and imputed to the Haplotype Reference Consortium by the UK Biobank (47). Further details on genotyping and sampling procedures are available online (48) and in section S2 of the online supplement. Because VNTRs were not genotyped in the UK Biobank data set, we used two independent whole-genome SNP data sets (the Family Transition Project [49] and the Genetics of Antisocial Drug Dependence [50, 51]) that also measured these repeat polymorphisms as reference panels in order to impute highly studied VNTRs within *DRD4*, *MAOA*, *SLC6A3*, and *SLC6A4* in the UK Biobank. The estimated out-of-sample imputed genotype match rates were ≥ 0.919 for all four VNTRs (mean $R^2=0.868$; details are provided in reference 29).

Psychiatric Genomics Consortium sample. To investigate candidate gene polymorphism main effect hypotheses, we also used data from the most recent GWAS on depression conducted by the Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium (PGC), which is described in detail in Wray et al. (18). Lack of access to raw genotypes for a large number of the PGC cohorts precluded imputation of VNTRs in the PGC sample. To minimize sample overlap with the UK Biobank, U.K.-based cohorts

were excluded from the PGC data set, resulting in GWAS summary statistics for a total of 443,264 individuals (120,201 case subjects and 323,063 control subjects) (for further details, see section S2 of the online supplement).

Phenotypes

Table 1 describes all phenotypes examined in the present investigation, and additional information is provided in section S3 of the online supplement. Correlations between depression outcomes and Cohen's kappa estimates for diagnosis phenotypes are presented in Tables S3.1 and S3.2 in the online supplement. Marker-based heritabilities of, and genetic correlations between, depression outcomes were estimated via linkage disequilibrium (LD) score regression (52) and are presented in Tables S3.3 and S3.4 and Figure S3.3 in the online supplement (for further details, see section S4.4 of the online supplement).

Analyses

All analyses were preregistered through the Open Science Framework and are available at <https://osf.io/akgvz/>. Statistical models are described in detail in section S4 of the online supplement, and departures from the preregistered analyses are documented in section S5.

Polymorphism-wise analyses. We analyzed associations between outcomes and each of the top 16 candidate polymorphisms using a generalized linear model framework (link functions are listed in Table S4.1 in the online supplement). For two of the genes, *TPH2* and *DTNBP1*, no particular polymorphism was investigated in a preponderance of studies (see Figures S1.2 and S1.3 in the online supplement), so these genes were not included in the polymorphism-wise analyses. Covariates included genotyping batch, testing center, sex, age, age squared, and the first 10 European-ancestry principal components. Sixteen polymorphism-by-environment effects were tested on both the additive and multiplicative scales for each of the 16 polymorphisms; each model tested is listed in Table S4.1 in the online supplement. For interaction tests, we included all covariate-by-polymorphism and covariate-by-moderator terms to control for the potential confounding influences of covariates on the interaction (53). We also tested interaction models that controlled only for covariate main effects, which is insufficient but common in the candidate gene literature (33). Across all outcomes, we employed a preregistered significance threshold of $\alpha_{\text{poly}}=0.05/16=3.13 \times 10^{-3}$, corresponding to a Bonferroni correction across the top 16 candidate polymorphisms. This threshold is liberal because it does not account for the multiple ways each polymorphism was analyzed or the multiple outcomes it was assessed with respect to. Further details are provided in section S4.1 of the online supplement.

Gene-wise and gene-set analyses. We used the National Center for Biotechnology Information (NCBI) Build 37 gene

locations to annotate SNPs to genes, allowing SNPs within a 25-kb window of the gene start and end points to be mapped to each gene. We used MAGMA, version 1.05b (54), to perform gene-wise and gene-set analyses for the top 18 candidate genes separately in the UK Biobank and PGC data sets. Gene-wise tests summarize the degree of association between a phenotype and polymorphisms within a given gene; in contrast, gene-set tests examine the association between a phenotype and a set of genes rather than individual genes.

We conducted gene-wise association analyses for each gene and outcome using the MAGMA default gene-level association statistic (sum $-\log$ p-based statistics and principal components regression, for tests based on summary statistics and individual-level genotypes, respectively) and using a liberal significance threshold of $\alpha_{\text{gene}}=0.05/18=2.78\times 10^{-3}$ to correct for multiple tests across the 18 candidate genes. We used summary statistics from the PGC2 depression GWAS (18) (excluding UK-based cohorts) as input for the PGC analyses, whereas individual-level genotypes were available for the UK Biobank. The gene-level association statistics were in turn used to perform “competitive” gene-set tests that compared enrichment of depression phenotype-associated loci between our set of 18 candidate genes and all other genes not in the gene set, controlling for potentially confounding gene characteristics. Further analyses, which compared the 18 candidate genes to negative control sets of genes involved in type 2 diabetes, height, or synaptic processes, are described in section S4.2 of the online supplement, and results are reported in section S11.

RESULTS

Polymorphism-Level Analyses

Table 2 lists the most significant result for each of the most-studied candidate gene polymorphisms for the main effect across the eight outcomes investigated (eight main effect tests per polymorphism) and the interaction effect across five moderators measured in the UK Biobank (32 interaction tests per polymorphism [see section S4.1 of the online supplement]). Given the number of tests conducted, there was little evidence that any effect was larger than what would be expected by chance under the null hypothesis. Only for *COMT* rs4680 on current depression severity was there was evidence of a small main effect that surpassed our liberal threshold of significance, such that the incident rate of current depression severity scores decreased by a factor of 0.983 per copy of the G allele (odds ratio 95% CI=0.967–0.999; $p=0.002$) (Figure 2). Detecting an effect of this size at an alpha level of 0.05 with 80% power would require a sample of over 100,000 individuals (see section S4.3 of the online supplement). Similarly, across all polymorphisms, outcomes, and exposures, on both the additive and multiplicative scales, no polymorphism-by-exposure moderation effects attained significance at α_{poly} . Failing to include all covariate-by-polymorphism and covariate-by-moderator terms as covariates, as is common in the $G\times E$ literature (33), inflated

product term test statistics on average but did not result in any additional significant effects (see section S10 of the online supplement). Complete results for all outcomes are provided in sections S7–S10 of the online supplement.

Despite the lack of evidence for $G\times E$ effects, all moderators exhibited large significant effects on all outcomes in the expected directions (see section S6 of the online supplement). For example, experiencing childhood trauma increased odds for estimated lifetime depression diagnosis by a factor of 1.655 ($z=32.048$, $p=2.33\times 10^{-225}$) and experiencing a traumatic event in the past 2 years increased incidence rate of current depression severity index by a factor of 1.431 ($z=27.004$, $p=1.32\times 10^{-160}$).

Gene-Level Analyses

Across all candidate genes and outcomes, only *DRD2* showed a significant gene-wise effect ($\alpha_{\text{gene}}=0.05/18=2.78\times 10^{-3}$), and only on PGC lifetime depression diagnosis, using both the sum $-\log$ p statistic ($p=5.14\times 10^{-7}$) and the minimum p-value statistic ($p=2.74\times 10^{-3}$; see Figure 3 for gene-wise effects on estimated lifetime depression diagnosis and current depression severity, section S11.1 of the online supplement for all gene-wise results, and section S4.2 for comparison of methods). The former estimate, based on the sum $-\log$ p statistic, was also significant at the more stringent genome-wide level ($\alpha_{\text{GW}}=0.05/19,165=2.61\times 10^{-6}$). *DRD2* did not exhibit a significant effect on any of the UK Biobank outcomes despite its high genetic correlations with the UK Biobank depression phenotypes (see Table S3.3 and Figure S3.3 in the online supplement). Investigating the effects of the 18 genes together as a set revealed no associations with depression above what would be expected by chance under the null hypothesis; the set of 18 depression candidate genes did not show stronger associations with any depression phenotype compared with all other genes at an alpha of 0.05 (see section S11.2 of the online supplement).

Attempted Replication of Top 16 Loci Implicated by PGC GWAS Results

In order to contextualize the lack of replication of the 16 candidate genetic polymorphisms, we sought to replicate the top 16 independent genome-wide significant loci implicated for PGC lifetime diagnosis by examining their associations with estimated lifetime diagnosis in the independent UK Biobank sample (for details, see section S4.5 of the online supplement). Three loci attained significance at α_{poly} (0.05/16) (rs12552, rs12658032, and rs11135349; see section S12 of the online supplement), which is consistent with the low power to detect small associations; the median power for the 16 loci was 0.143, and the 95% confidence interval for number of replications we would expect given power estimates was 2 to 7 (see Figure S4.6 in the online supplement).

Sensitivity of Results to Measurement Error

One possible reason candidate gene polymorphism associations detected in small samples are not replicated in large

TABLE 2. Minimum p value effect across eight main effect models and 32 interaction effect models per polymorphism^a

Polymorphism	MAF	Outcome: Additive Effect	β	Min p	Outcome: Interaction Effect	Moderator	Scale	β	Min p
<i>SLC6A4</i> ; 5-HTTLPR ^{b,c}	0.499	Current depression severity	0.008	0.138	Lifetime episode count	TDI	Primary	0.019	0.041
<i>BDNF</i> ; rs6265	0.188	Severe recurrent depression	0.018	0.325	Estimated lifetime depression diagnosis	TDI	Alternate	0.007	0.008
<i>COMT</i> ; rs4680	0.483	Current depression severity	-0.017	0.002 ^d	Conditional lifetime symptom count	Stressor-induced depression ^e	Alternate	0.048	0.040
<i>HTR2A</i> ; rs6311	0.402	Estimated lifetime depression diagnosis	0.020	0.045	Estimated lifetime depression diagnosis	Childhood trauma	Alternate	0.008	0.072
<i>TPH1</i> ; rs1800532	0.391	Current depression severity	-0.012	0.036	Conditional lifetime symptom count	Childhood trauma	Primary	-0.045	0.049
<i>DRD4</i> ; VNTR ^b	0.223	Touchscreen probable lifetime diagnosis (ordinal)	0.022	0.079	Severe recurrent depression	TDI	Primary	0.011	0.094
<i>DRD2</i> ; rs1800497	0.201	PGC lifetime diagnosis	-0.019	0.006	Conditional lifetime symptom count	Stressor-induced depression ^e	Alternate	-0.044	0.134
<i>MAOA</i> ; VNTR ^b	^f	Severe recurrent depression	0.023	0.073	Conditional lifetime symptom count	TDI	Primary	-0.024	0.014
<i>APOE</i> ; rs429358/ rs7412 ^b	0.148	Lifetime episode count	0.019	0.091	Current depression severity	Recent trauma	Alternate	-0.182	0.009
<i>MTHFR</i> ; rs1801133	0.334	Current depression severity	-0.012	0.034	Estimated lifetime depression diagnosis	Adulthood trauma	Alternate	-0.007	0.054
<i>CLOCK</i> ; rs1801260	0.268	Touchscreen probable lifetime diagnosis	0.030	0.013	Severe recurrent depression	TDI	Primary	0.014	0.012
<i>SLC6A3</i> ; VNTR ^b	0.255	Touchscreen probable lifetime diagnosis	0.019	0.114	Estimated lifetime depression diagnosis	Childhood trauma	Alternate	-0.008	0.099
<i>ACE</i> ; in/del	0.474	Touchscreen probable lifetime diagnosis	0.016	0.143	Lifetime episode count	TDI	Primary	0.015	0.107
<i>ABCB1</i> ; rs1045642	0.456	PGC lifetime diagnosis	-0.006	0.164	Current depression severity	Recent trauma	Alternate	-0.108	0.027
<i>DRD3</i> ; rs6280	0.336	Current depression severity	-0.010	0.078	Current depression severity	Recent trauma	Alternate	-0.111	0.031
<i>DBH</i> ; rs1611115	0.205	Estimated lifetime depression diagnosis	-0.014	0.236	Severe recurrent depression	Adulthood trauma	Alternate	-0.005	0.087

^a MAF=minor allele frequency in the subset of the UK Biobank sample for whom estimated lifetime depression diagnosis was available; PGC=Psychiatric Genomics Consortium; TDI=Townsend deprivation index. "Touchscreen" refers to the initial computerized touchscreen interview in the UK Biobank. The p values are the minimum for each polymorphism across outcomes/moderators for additive and interaction effects (on additive and multiplicative scales), respectively. Interaction tests were not conducted in the PGC sample because moderators were unavailable for that sample. Only one effect was significant after a liberal correction for number of polymorphisms (but not for outcomes or moderators; $\alpha_{\text{poly}}=0.05/16=3.125 \times 10^{-3}$). Details of each model are provided in section S4 of the online supplement, with all interaction models listed in Table S4.1; complete results are presented in sections S7–S9 of the online supplement.

^b VNTRs and the triallelic *APOE* polymorphism were unavailable for the PGC samples, and thus these variants were examined only across the seven UK Biobank outcomes.

^c Allele frequency reflects the low-activity VNTR/rs25531 haplotype (5).

^d Significant at $\alpha_{\text{poly}}=3.125 \times 10^{-3}$.

^e Variant-by-stressor-induced depression estimates reflect differences in the magnitude of variant/outcome associations between individuals reporting that their depression was induced by a stressful event and those reporting otherwise.

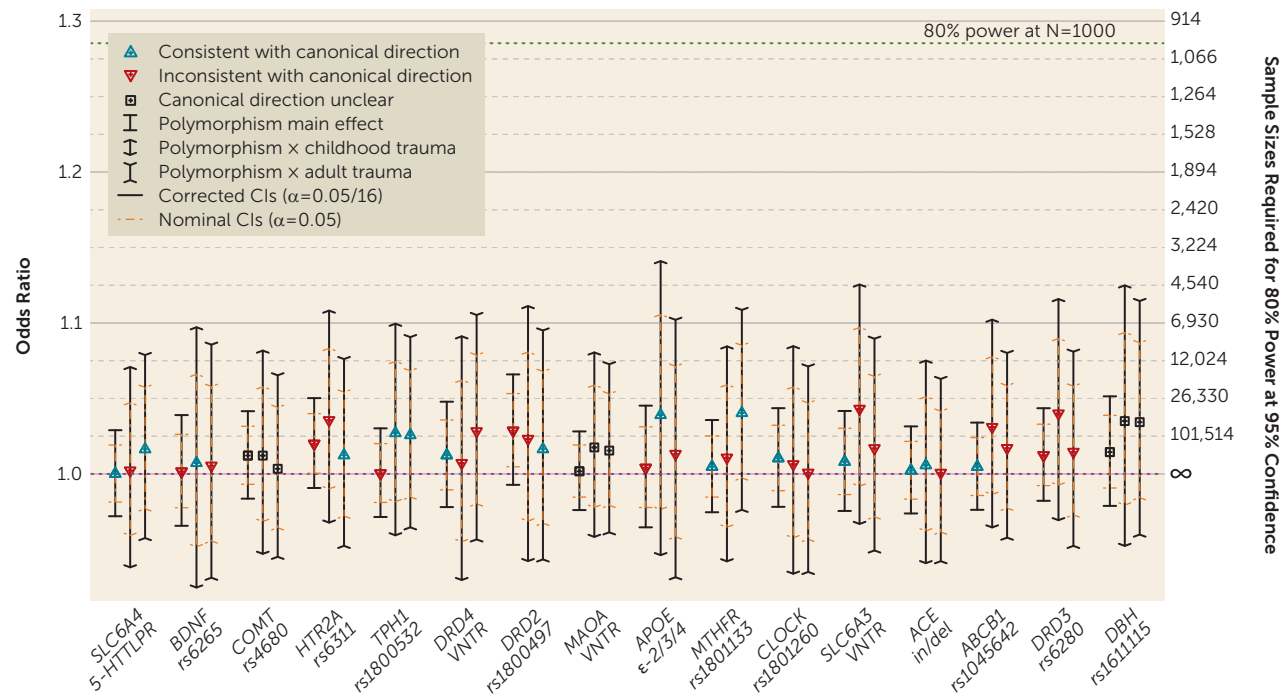
^f *MAOA* is located on the X chromosome; frequencies were 0.336 and 0.341 for females and males, respectively.

GWASs is the potentially worse phenotyping and higher measurement error in predictor or outcome variables in GWAS data sets. To investigate this possibility, we used a Monte Carlo procedure to quantify the extent to which measurement error may have affected the statistical power of our tests. As a lower bound on a candidate gene polymorphism study effect sizes, we used the minimally detectable log odds ratio for both main and interaction effects corresponding to 50% power at an alpha of 0.05 in a balanced

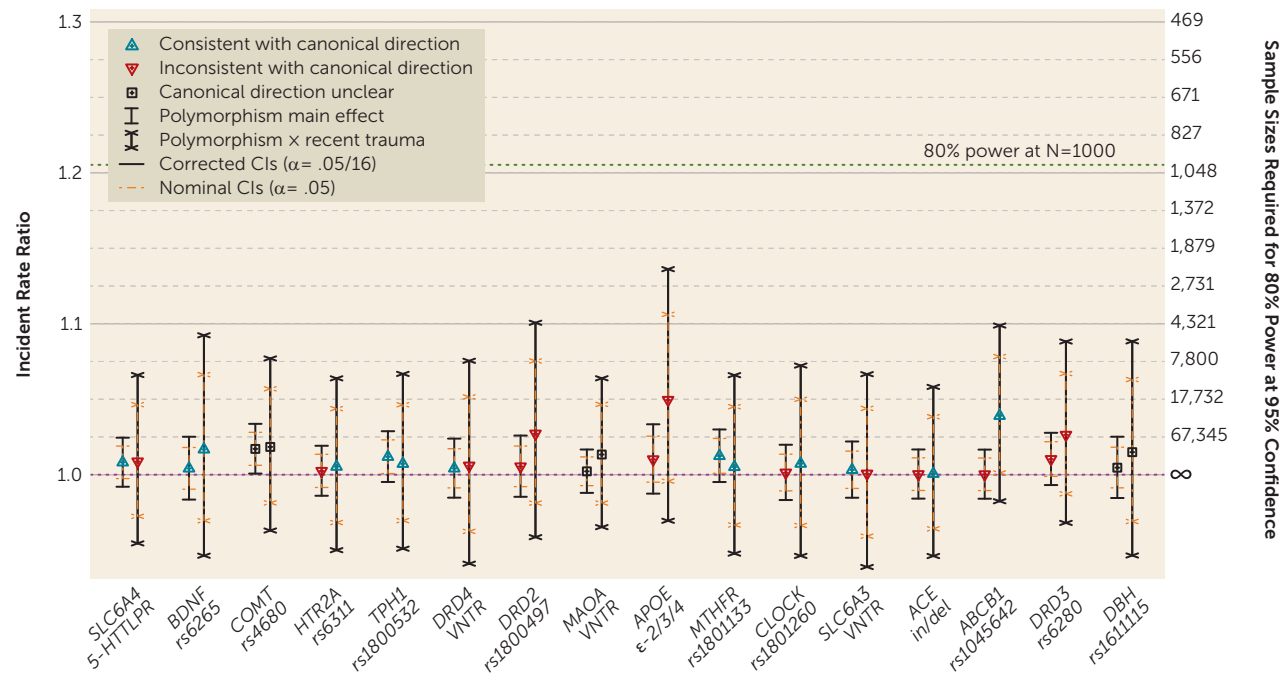
case-control study of 1,000 individuals and where the risk allele frequency was 0.5 (e.g., for main effects, genomic relative risk=1.16). Simulations demonstrated that we had ~100% power to detect such effects under multiple severe measurement error scenarios in a sample size typical of that in our UK Biobank analyses (~30,000 case subjects and ~85,000 control subjects; see section S4.3.3 of the online supplement). This was true even in the extreme scenario in which half of diagnoses and half of traumatic exposures were

FIGURE 2. Main effects and gene-by-environment effects of 16 candidate polymorphisms on estimated lifetime depression diagnosis and current depression severity in the UK Biobank sample^a

A. Estimated Lifetime Major Depression Diagnosis



B. Current Depression Severity



^a The graphs show effect size estimates for 16 candidate polymorphisms, presented in order of estimated number of studies from left to right, descending, on estimated lifetime depression diagnosis (panel A) and past-2-week depression symptom severity from the online mental health follow-up assessment (panel B) in the UK Biobank sample (N=115,257). Both polymorphism main effects and polymorphism-by-environment moderator interaction effects are presented for each outcome. Detailed descriptions of the variables and of the association and power analysis models are provided in sections S3 and S4, respectively, of the online supplement.

determined by coin toss (see Figure S4.5 in the online supplement).

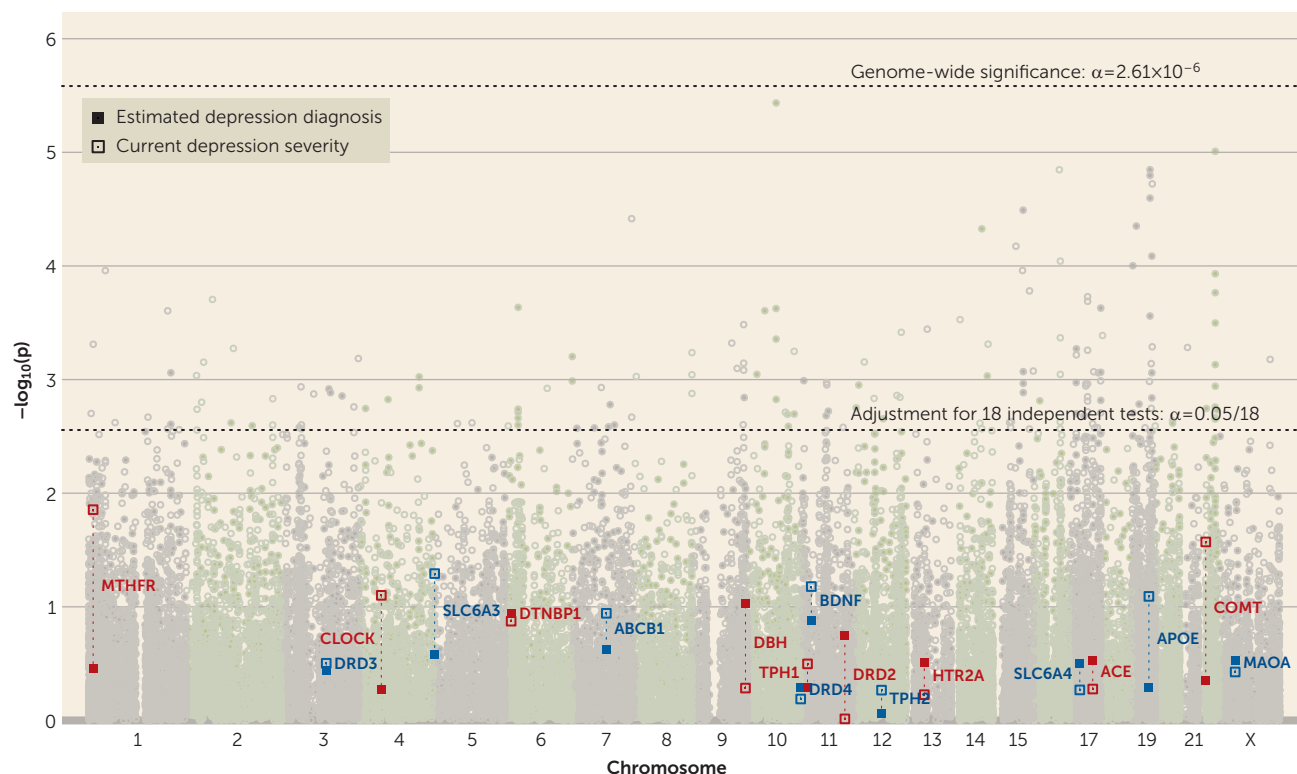
DISCUSSION

We examined multiple types of associations between 18 highly studied candidate genes for depression and multiple depression phenotypes. The study was very well powered compared with previous candidate gene studies, with N s ranging from 62,138 to 443,264 across subsamples. Despite the high statistical power, none of the most highly studied polymorphisms within these genes demonstrated substantial contributions to depression liability. Furthermore, we found no evidence to support moderation of polymorphism effects by exposure to traumatic events or socioeconomic adversity. We also found little evidence to support contributions of other common polymorphisms within these genes to depression liability, except *DRD2*, which showed a genome-wide significant gene-wise effect on depression diagnosis in the PGC sample but not on any outcomes in the UK Biobank sample. The reasons for the failure of *DRD2* to replicate in the UK Biobank are unclear, but it could be due to sampling variability, lower statistical power in the UK Biobank, or false positive or negative findings. Phenotypic heterogeneity, however, is an unlikely explanation, as genetic correlation estimates between depression phenotypes across samples were high (see Table S3.3 and Figure S3.3 in the online supplement)—for example, PGC lifetime depression diagnosis was strongly associated with estimated lifetime depression diagnosis from the UK Biobank online follow-up questionnaire ($\hat{h}^2_{LDSC}=0.085$, $SE=0.004$, and $\hat{h}^2_{LDSC}=0.057$, $SE=0.007$, respectively; $\hat{r}_g=0.855$, $SE=0.054$, $p=2.08\times10^{-57}$), which was in turn strongly associated with probable lifetime diagnosis from the UK Biobank initial touchscreen interview ($\hat{h}^2_{LDSC}=0.090$, $SE=0.008$; $\hat{r}_g=0.939$, $SE=0.082$, $p=2.83\times10^{-30}$). Finally, as a set, depression candidate genes were no more related to depression phenotypes than noncandidate genes. Our results stand in stark contrast to the candidate gene literature, where large, statistically significant effects are commonly reported for the specific polymorphisms in the 18 candidate genes we investigated here.

Several features of this investigation set it apart from previous candidate gene replication attempts, meta-analyses of candidate gene studies, and genome-wide studies that failed to support roles for depression candidate polymorphisms. First, this is the only study to have imputed and examined the effects of several highly studied VNTR polymorphisms in a large GWAS data set, including 5-HTTLPR in *SLC6A4*, which was examined in 38.14% of the depression candidate gene studies we identified (see reference 29 for imputation details). Second, we thoroughly examined several distinct depression phenotypes (e.g., diagnosis, depressive episode recurrence, symptom count among depressed individuals) to ensure that our results did not reflect a single operationalization of depression. Some researchers have attributed the poor replicability of candidate gene findings

to specificity of effects with respect to particular types of depression or stressors (e.g., prior versus subsequent depression onset with respect to stress exposure [38], recurrent versus single-episode depression [55], and financial versus other stress exposure [56]). We therefore examined all available depression and exposure phenotypes reflecting constructs of interest in the candidate gene literature. Results for all measures and modeling choices (e.g., multiplicative versus additive interactions), presented in detail in the supplement (see sections S7–S11 of the online supplement), were consistently null with respect to candidate gene hypotheses. Third, we employed exceedingly liberal significance thresholds (e.g., for polymorphism-wise analyses, $\alpha_{poly}=3.13\times10^{-3}$, as opposed to the standard $\alpha_{GWAS}=5\times10^{-8}$ utilized in GWASs) across all outcomes to ensure that no possible effect was missed, correcting only for the number of polymorphisms we examined. Our results therefore suggest that the zero or near-zero effect sizes of these candidate polymorphisms, rather than the multiple-testing burden imposed by genome-wide scans, account for the previous failures of large GWASs to detect candidate polymorphism effects. Finally, and perhaps most importantly, unlike meta-analyses that use previously published candidate gene findings, our results cannot be affected by selective publication or reporting practices that can inflate type I errors and lead to biased representations of evidence for candidate gene hypotheses.

Our study has several limitations. First, it is possible that we failed to identify a small number of candidate gene publications and that this resulted in the omission of some depression candidate genes examined in 10 or more publications. Nevertheless, the top nine of the 18 identified genes accounted for 86.59% of the estimated number of studies, and it is unlikely that we omitted any depression candidate genes with popularity approaching that of, for example, *SLC6A4* or *COMT*. Second, a subset of the UK Biobank sample was ascertained for smoking behaviors (the BiLEVE study [57]), and controlling for genotyping batch (which differentiates the two subsamples) has the potential to induce collider bias (58). However, only one of the 16 candidate gene polymorphisms demonstrated minor allele frequency (MAF) differences across these two subsamples ($rs6311$; $\chi^2=12.558$, $df=2$, $p=0.002$; MAF=0.402 in the BiLEVE sample, MAF=0.405 otherwise) and it is unlikely that ascertainment in the BiLEVE subsample unduly influenced association statistics. However, the potential influence of ascertainment in the BiLEVE subsample on interaction effect estimates, as well as other possible sources of selection-induced bias, remains unclear. Third, whereas some of phenotypes we examined closely matched standard diagnostic instruments (e.g., current depression severity was based on the widely used Patient Health Questionnaire–9 [59]), others were of undetermined reliability. For example, one of the nine DSM-5 depression symptoms (motor agitation/retardation) was omitted from the UK Biobank online mental health follow-up questionnaire, and our estimated lifetime depression diagnosis

FIGURE 3. Gene-wise statistics for effects of 18 candidate genes on primary depression outcomes in the UK Biobank sample^a

^a The plot shows gene-wise p values across the genome, highlighting the 18 candidate polymorphisms' effects on estimated depression diagnosis (filled points) and past-2-week depression symptom severity (unfilled points) from the online mental health follow-up assessment in the UK Biobank sample (N=115,257). Gene labels alternate colors to aid readability. Detailed descriptions of the variables and of the association models are provided in sections S3 and S4.2, respectively, of the online supplement.

phenotype required four or more of eight symptoms rather than the standard five or more of nine symptoms (in addition to episode duration and impairment criteria; see section S3.1 of the online supplement). However, enforcing stricter case-control criteria (i.e., comparing individuals who endorsed no 2-week period of either anhedonia or depressed mood throughout their lifetime to individuals reporting recurrent episodes, endorsing five or more of eight symptoms, and meeting duration and impairment criteria) failed to alter results (see sections S7–S9 of the online supplement), despite the fact that even this diminished sample size (N=67,304) was much larger than any previous candidate gene study we are aware of. Fourth, some of the phenotypes we examined were possibly measured with greater error than is typical in smaller candidate gene studies, an issue for which large studies are often criticized. For example, the prevalence of our measure of traumatic exposure in adulthood was uncommonly high (59.11%), and most of our retrospective measurements were likely corrupted by recall bias. However, as demonstrated in section S4.3.3 of the online supplement, even extreme measurement error cannot explain our failure to detect the relatively large effects necessary for detection in smaller samples. Furthermore, follow-up analyses demonstrated strong effects of all environmental moderators across all outcomes (see section S6 of the online supplement), suggesting that both moderators and depression phenotypes

were measured with sufficient accuracy to detect known environmental effects. It is exceedingly difficult to construct a plausible measurement error model that could, for example, comfortably reconcile the large effect estimate of childhood trauma on estimated lifetime diagnosis (odds ratio=1.655, $p=2.33 \times 10^{-225}$) and the negligible estimate for the 5-HTTLPR-by-childhood trauma interaction effect (odds ratio=0.988, $p=0.919$) with the existence of a substantial $G \times E$ interaction effect.

The genetic underpinnings of common complex traits such as depression appear to be far more complicated than originally hoped (60, 61), and large collaborative efforts have not supported the existence of common genetic variants with large effects on depression liability (18). In the context of our understanding of psychiatric genetics in the 1990s and early 2000s, the most studied candidate genes and the polymorphisms within them were defensible targets for association studies. However, our results demonstrate that historical depression candidate gene polymorphisms do not have detectable effects on depression phenotypes. Furthermore, the candidate genes themselves (with the possible exception of *DRD2*) were no more associated with depression phenotypes than genes chosen at random. The present study had >99.99% power at $\alpha_{GWA} = 5 \times 10^{-8}$ to detect a main effect of the magnitude commonly reported in candidate gene studies, even allowing for extreme measurement error in both

outcome and moderator phenotypes (see section S4.3 of the online supplement). Thus, it is extremely unlikely that we failed to detect any true associations between depression phenotypes and these candidate genes. The implication of our study, therefore, is that previous positive main effect or interaction effect findings for these 18 candidate genes with respect to depression were false positives. Our results mirror those of well-powered investigations of candidate gene hypotheses for other complex traits, including those of schizophrenia (16, 25) and white matter microstructure (19). The potential for self-correction is an essential strength of the scientific enterprise; it is with this mechanism in mind that we present these findings. In agreement with the recent recommendations of the National Institute of Mental Health Council Workgroup on Genomics (62), we conclude that it is time for depression research to abandon historical candidate gene and candidate gene-by-environment interaction hypotheses.

AUTHOR AND ARTICLE INFORMATION

The Institute for Behavioral Genetics (Border, Johnson, Evans, Smolen, Keller), the Department of Psychology and Neuroscience (Border, Berley, Keller), the Department of Applied Mathematics (Border), and the Department of Ecology and Evolutionary Biology (Evans), University of Colorado Boulder, Boulder; the Department of Psychiatry, Washington University School of Medicine, St. Louis (Johnson); the Department of Genetics and Psychiatry, University of North Carolina at Chapel Hill (Sullivan); and the Department of Medical Epidemiology and Biostatistics, Karolinska Institute, Stockholm (Sullivan).

Send correspondence to Mr. Border (richard.border@colorado.edu).

Mr. Border was supported by NIMH grant T32 MH016880 and the Institute for Behavioral Genetics. Dr. Sullivan was supported by NIMH grant U01 MH109528 and the Swedish Research Council (D0886501). Drs. Evans and Keller were supported by NIMH grant 2RO1 MH100141 and the Institute for Behavioral Genetics. This research was conducted using the UK Biobank Resource under application numbers 1665, 16651, and 24795. This work utilized the RMACC Summit supercomputer, which is supported by the National Science Foundation (awards ACI-1532235 and ACI-1532236), the University of Colorado Boulder, and Colorado State University. The Summit supercomputer is a joint effort of the University of Colorado Boulder and Colorado State University.

The authors thank SURFsara (www.surfsara.nl) for support in using the Lisa Compute Cluster. They also thank the research participants of the PGC and UK Biobank, and the employees of 23andMe for their contribution to this study.

Dr. Sullivan has received grant support from Lundbeck, served on advisory committees for Lundbeck and Pfizer, received consulting fees from Element Genomics, and received speaking fees from Roche; his spouse has received grant support from and served on a scientific advisory board for Shire and receives royalties from Pearson and Walker. The other authors report no financial relationships with commercial interests.

Received July 26, 2018; revision received October 24, 2018; accepted December 17, 2018.

REFERENCES

1. Sullivan PF, Neale MC, Kendler KS: Genetic epidemiology of major depression: review and meta-analysis. *Am J Psychiatry* 2000; 157: 1552–1562
2. McInnes LA, Freimer NB: Mapping genes for psychiatric disorders and behavioral traits. *Curr Opin Genet Dev* 1995; 5:376–381
3. Ramamoorthy S, Bauman AL, Moore KR, et al: Antidepressant- and cocaine-sensitive human serotonin transporter: molecular cloning, expression, and chromosomal localization. *Proc Natl Acad Sci USA* 1993; 90:2542–2546
4. Owens MJ, Nemeroff CB: Role of serotonin in the pathophysiology of depression: focus on the serotonin transporter. *Clin Chem* 1994; 40:288–295
5. Heils A, Teufel A, Petri S, et al: Allelic variation of human serotonin transporter gene expression. *J Neurochem* 1996; 66:2621–2624
6. Stoltenberg SF, Burmeister M: Recent progress in psychiatric genetics: some hope but no hype. *Hum Mol Genet* 2000; 9:927–935
7. Buckland PR: Genetic association studies of alcoholism: problems with the candidate gene approach. *Alcohol Alcohol* 2001; 36:99–103
8. Munafò MR: Candidate gene studies in the 21st century: meta-analysis, mediation, moderation. *Genes Brain Behav* 2006; 5(suppl 1):3–8
9. Lander ES, Schork NJ: Genetic dissection of complex traits. *Science* 1994; 265:2037–2048
10. Terwilliger JD, Weiss KM: Linkage disequilibrium mapping of complex disease: fantasy or reality? *Curr Opin Biotechnol* 1998; 9: 578–594
11. Colhoun HM, McKeigue PM, Davey Smith G: Problems of reporting genetic associations with complex outcomes. *Lancet* 2003; 361: 865–872
12. Caspi A, McClay J, Moffitt TE, et al: Role of genotype in the cycle of violence in maltreated children. *Science* 2002; 297:851–854
13. Caspi A, Sugden K, Moffitt TE, et al: Influence of life stress on depression: moderation by a polymorphism in the 5-HTT gene. *Science* 2003; 301:386–389
14. Niculescu AB 3rd: DISCOVERY in psychiatric genetics. *Mol Psychiatry* 2014; 19:145
15. Duncan LE, Keller MC: A critical review of the first 10 years of candidate gene-by-environment interaction research in psychiatry. *Am J Psychiatry* 2011; 168:1041–1049
16. Farrell MS, Werge T, Sklar P, et al: Evaluating historical candidate genes for schizophrenia. *Mol Psychiatry* 2015; 20:555–562
17. Munafò MR: Reliability and replicability of genetic association studies. *Addiction* 2009; 104:1439–1440
18. Wray NR, Ripke S, Mattheisen M, et al: Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nat Genet* 2018; 50:668–681
19. Thompson PM, Stein JL, Medland SE, et al: The ENIGMA Consortium: large-scale collaborative analyses of neuroimaging and genetic data. *Brain Imaging Behav* 2014; 8:153–182
20. Schizophrenia Working Group of the Psychiatric Genomics Consortium: Biological insights from 108 schizophrenia-associated genetic loci. *Nature* 2014; 511:421–427
21. Munafò MR, Gage SH: Improving the reliability and reporting of genetic association studies. *Drug Alcohol Depend* 2013; 132:411–413
22. Burton PR, Hansell AL, Fortier I, et al: Size matters: just how big is BIG? Quantifying realistic sample size requirements for human genome epidemiology. *Int J Epidemiol* 2009; 38:263–273
23. Bosker FJ, Hartman CA, Nolte IM, et al: Poor replication of candidate genes for major depressive disorder using genome-wide association data. *Mol Psychiatry* 2011; 16:516–532
24. Coleman JRI, Peyrot WJ, Purves KL, et al: Genome-wide gene-environment analyses of depression and reported lifetime traumatic experiences in UK Biobank. *bioRxiv* November 1, 2018 (<https://doi.org/10.1101/247353>)
25. Johnson EC, Border R, Melroy-Greif WE, et al: No evidence that schizophrenia candidate genes are more associated with schizophrenia than noncandidate genes. *Biol Psychiatry* 2017; 82:702–708
26. Jahanshad N, Ganjgahi H, Bralten J, et al: Do candidate genes affect the brain's white matter microstructure? Large-scale evaluation of 6,165 diffusion MRI scans. *bioRxiv* February 20, 2017 (<https://doi.org/10.1101/107987>)

27. Van der Auwera S, Peyrot WJ, Milaneschi Y, et al: Genome-wide gene-environment interaction in depression: a systematic evaluation of candidate genes: the Childhood Trauma Working-Group of PGC-MDD. *Am J Med Genet B Neuropsychiatr Genet* 2018; 177:40–49
28. Culverhouse RC, Saccone NL, Horton AC, et al: Collaborative meta-analysis finds no evidence of a strong interaction between stress and 5-HTTLPR genotype contributing to the development of depression. *Mol Psychiatry* 2018; 23:133–142
29. Border R, Smolen A, Corley R, et al: Imputation of behavioral candidate gene repeat variants in 486,551 publicly-available UK Biobank individuals. *Eur J Hum Genet*, February 5, 2019 (<https://doi.org/10.1038/s41431-019-0349-x>)
30. Brookes KJ: The VNTR in complex disorders: the forgotten polymorphisms? A functional way forward? *Genomics* 2013; 101:273–281
31. Wendland JR, Martin BJ, Kruse MR, et al: Simultaneous genotyping of four functional loci of human SLC6A4, with a reappraisal of 5-HTTLPR and rs25531. *Mol Psychiatry* 2006; 11:224–226
32. Dick DM, Agrawal A, Keller MC, et al: Candidate gene-environment interaction research: reflections and recommendations. *Perspect Psychol Sci* 2015; 10:37–59
33. Keller MC: Gene \times environment interaction studies have not properly controlled for potential confounders: the problem and the (simple) solution. *Biol Psychiatry* 2014; 75:18–24
34. Border R, Keller MC: Commentary: Fundamental problems with candidate gene-by-environment interaction studies: reflections on Moore and Thoenes (2016). *J Child Psychol Psychiatry* 2017; 58: 328–330
35. Munafò MR, Zammit S, Flint J: Limitations of gene \times environment interaction models in psychiatry. *J Child Psychol Psychiatry* 2014; 55:1092–1101
36. Assary E, Vincent JP, Keers R, et al: Gene-environment interaction and psychiatric disorders: review and future directions. *Semin Cell Dev Biol* 2018; 77:133–143
37. Moore SR: Commentary: What is the case for candidate gene approaches in the era of high-throughput genomics? A response to Border and Keller (2017). *J Child Psychol Psychiatry* 2017; 58: 331–334
38. Moffitt TE: Letter to Culverhouse. 2012. https://moffittcaspi.trinity.duke.edu/sites/moffittcaspi.trinity.duke.edu/files/file-attachments/Letter_to_Culverhouse_JUNE2012.pdf
39. Uher R: Gene-environment interactions in severe mental illness. *Front Psychiatry* 2014; 5:48
40. Spitzer RL, Kroenke K, Williams JB: Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study. *Primary Care Evaluation of Mental Disorders. Patient Health Questionnaire. JAMA* 1999; 282:1737–1744
41. First MB, Gibbon M, Spitzer RL, et al: User's Guide for the Structured Clinical Interview for DSM-IV Axis I Disorders—Research Version. New York, Biometrics Research Department, New York State Psychiatric Institute, 1996
42. Smith DJ, Nicholl BI, Cullen B, et al: Prevalence and characteristics of probable major depression and bipolar disorder within UK Biobank: cross-sectional study of 172,751 participants. *PLOS One* 2013; 8: e75362
43. Townsend P, Phillimore P, Beattie A: Health and Deprivation: Inequality and the North. Kent, UK, Croom Helm, 1988
44. Steegen S, Tuerlinckx F, Gelman A, et al: Increasing transparency through a multiverse analysis. *Perspect Psychol Sci* 2016; 11:702–712
45. Cock PJA, Antao T, Chang JT, et al: Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* 2009; 25:1422–1423
46. Sudlow C, Gallacher J, Allen N, et al: UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med* 2015; 12:e1001779
47. Loh P-R, Danecek P, Palamara PF, et al: Reference-based phasing using the Haplotype Reference Consortium panel. *Nat Genet* 2016; 48:1443–1448
48. Genotyping and Quality Control of UK Biobank, a Large-Scale, Extensively Phenotyped Prospective Resource: Information for Researchers: Interim Data Release, 2015. http://www.ukbiobank.ac.uk/wp-content/uploads/2014/04/UKBiobank_genotyping_QC_documentation-web.pdf
49. Conger RD, Schofield TJ, Neppl TK: Intergenerational continuity and discontinuity in harsh parenting. *Parent Sci Pract* 2012; 12: 222–231
50. Derringer J, Corley RP, Haberstick BC, et al: Genome-wide association study of behavioral disinhibition in a selected adolescent sample. *Behav Genet* 2015; 45:375–381
51. Stallings MC, Corley RP, Hewitt JK, et al: A genome-wide search for quantitative trait loci influencing substance dependence vulnerability in adolescence. *Drug Alcohol Depend* 2003; 70:295–307
52. Bulik-Sullivan B, Finucane HK, Anttila V, et al: An atlas of genetic correlations across human diseases and traits. *Nat Genet* 2015; 47: 1236–1241
53. Yzerbyt VY, Muller D, Judd CM: Adjusting researchers' approach to adjustment: on the use of covariates when testing interactions. *J Exp Soc Psychol* 2004; 40:424–431
54. de Leeuw CA, Mooij JM, Heskes T, et al: MAGMA: Generalized Gene-Set Analysis of GWAS Data. *PLOS Comput Biol* 2015; 11: e1004219
55. Uher R, Caspi A, Houts R, et al: Serotonin transporter gene moderates childhood maltreatment's effects on persistent but not single-episode depression: replications and implications for resolving inconsistent results. *J Affect Disord* 2011; 135:56–65
56. Gonda X, Eszlari N, Kovacs D, et al: Financial difficulties but not other types of recent negative life events show strong interactions with 5-HTTLPR genotype in the development of depressive symptoms. *Transl Psychiatry* 2016; 6:e798
57. Miller S, Wain L, Shrine N, et al: The UK BiLEVE study: the first genetic study in UK Biobank identifies novel regions associated with airway obstruction and smoking behaviour. Presented at the American Thoracic Society 2015 International Conference. Available in American Thoracic Society International Conference Abstracts, B31 Inflammation and COPD, 2015, p A2714. <https://www.atsjournals.org/doi/abs/10.1164/ajrccm-conference.2015.191.1.MeetingAbstracts.A2714>
58. Munafò MR, Tilling K, Taylor AE, et al: Collider scope: when selection bias can substantially influence observed associations. *Int J Epidemiol* 2018; 47:226–235
59. Kroenke K, Spitzer RL, Williams JB: The PHQ-9: validity of a brief depression severity measure. *J Gen Intern Med* 2001; 16:606–613
60. McClellan J, King M-C: Genetic heterogeneity in human disease. *Cell* 2010; 141:210–217
61. Kapur S, Phillips AG, Insel TR: Why has it taken so long for biological psychiatry to develop clinical tests and what to do about it? *Mol Psychiatry* 2012; 17:1174–1179
62. Hyman SE, Krystal JH: Report of the National Advisory Mental Health Council Workgroup on Genomics: Opportunities and Challenges of Psychiatric Genetics. Rockville, Md, NIMH, 2018. <https://www.nimh.nih.gov/about/advisory-boards-and-groups/namhc/reports/report-of-the-national-advisory-mental-health-council-workgroup-on-genomics.shtml>