Depression is one of the most prevalent, disabling, and costly mental health conditions in the United States, with lifetime prevalence estimates of 11.7% among adolescents and 16.6% among adults. It is projected to be the leading cause of disease burden worldwide by 2030. Although the impact of depression can be minimized or prevented through early detection, treatment, and ongoing care, numerous individual and structural barriers—including stigma, lack of health insurance, and other barriers to accessing mental health services—prevent many from seeking help. Indeed, only slightly more than half of all people who experience depression seek treatment, and those who do tend to drop out prematurely or receive poor quality care.

Existing treatments for depression are modestly effective; only about one-fifth of adults receiving cognitive-behavioral therapy or psychodynamic therapy alone, and one-third of adults receiving antidepressant medication alone, will experience remission after an initial course of treatment. In children and adolescents, the efficacy of existing treatments is also limited. Moreover, nearly three-quarters of people with depression will experience a relapse at some point in their lives. These findings underscore the urgent need to prioritize prevention alongside treatment.

A deeper understanding of the etiology of depression, including its genetic and environmental determinants as well as their interplay (e.g., gene-environment interaction), will have implications for preventing depression and informing its clinical treatment. Numerous environmental risk factors for depression have been established, including poverty, negative family relationships and parental divorce, child...
maltreatment, and other stressful life events more generally. Although the risk of depression is elevated in the immediate aftermath of experiencing these environmental adversities, the effects of adversity can persist over the life course.

A robust literature implicates genetic factors in the etiology of depression and other psychiatric disorders. Depression is known to run in families; people with major depressive disorder (MDD) are three times more likely than those without the disorder to have a first-degree relative who also has depression. Twin studies, which allow for the simultaneous quantification of genetic and environmental influences, suggest that depression is moderately heritable. Specifically, twin studies have estimated that approximately 40% of the variation in the population risk of depression is attributable to genetic variation.

In recent years, the combination of advances in our understanding of human genomic variation (e.g., Human Genome Project, HapMap Project, 1,000 Genomes Project) and cost-effective genotyping techniques have led to extraordinary growth in molecular genetic studies of depression and other “complex” psychiatric phenotypes. These studies typically examine whether specific alleles (i.e., alternative forms of DNA sequence at a specific locus) or genotypes (i.e., the combination of alleles at a given locus) are associated with the phenotype of interest. Until recently, genetic studies of depression focused largely on candidate genes—that is, genes that are hypothesized to be implicated in the neurobiology of depression. Some of the most commonly studied candidate genes have been those regulating serotonin (5-HT) and dopamine (DA) neurotransmission, given the suspected involvement of these neurotransmitters in the pathophysiology of depression and their role as targets of antidepressant drugs. Unfortunately, most candidate gene studies have been underpowered, and replication of findings has been rare. More recently, the availability of DNA microarrays has enabled genome-wide association studies (GWAS) that do not rely on prior hypotheses. The GWAS approach allows for a hypothesis-free analysis of a million or more variants across the entire genome. The ultimate goal of GWAS is to improve diagnosis, prevention, and treatment through a nuanced understanding of the genetic underpinnings of the disease.

In this article we review recent findings from genetic association studies and G×E studies related to depression, and outline areas for future research. Several excellent reviews of this literature aimed at the genetic research community have already been published (see, e.g., references 28–33). We aim to provide a review for a broad audience, who may be unfamiliar with genetic concepts and methods. We have organized this review into three major sections. In the first, we describe recent findings from GWAS of depression. We begin with GWAS, rather than older methods (i.e., linkage and candidate gene association studies), since these older methods have already been extensively covered by prior reviews. We also do not review studies on genetic markers of antidepressant treatment response, or pharmacogenomics, since our focus is on the genetic determinants of illness risk. In the second section, we review findings from G×E studies, which aim to simultaneously examine the respective roles of genetic variants and environmental exposures in the etiology of depression. As described below, G×E studies have the potential to help identify genetic variants associated with both the risk of, and resilience against, depression—which are revealed only in specific subgroups of the population that have experienced a given environment. In the third section, we address the challenges that face genetic studies of depression and describe emerging strategies that may be useful for overcoming these challenges. We encourage readers who may be unfamiliar with basic genetic concepts to refer to two articles by Attia and colleagues and the resources listed in Text Box 1.

### FINDINGS FROM GENOME-WIDE ASSOCIATION STUDIES

GWAS have been one of the most widely used methods for identifying risk loci in the past decade. In a typical GWAS, one million or more common variants known as single nucleotide polymorphisms (SNPs) are examined for their association to disease. Common variants are generally defined as those alleles that are carried by at least 5% of the population. GWAS are typically conducted using a case-control design in which allele frequencies are compared between cases (with a disease) and controls (without the disease). Compared to candidate gene studies, GWAS provide a hypothesis free, or “unbiased,” approach to detecting susceptibility loci. To account for the large number of tests conducted, however, the threshold for declaring genome-wide significance in a GWAS is a p-value of less than $5 \times 10^{-8}$, which is equivalent to a p-value of .05 that has been corrected for a million independent tests ($p < .00000005$). Because common variant effects are typically modest, large samples (in the order of 10,000 or more cases and controls) are usually needed to have sufficient power to detect such effects at this statistical threshold.
According to the National Human Genome Research Institute GWAS catalog, more than 2000 GWAS have been published to date.41 A total of 14 GWAS have been conducted for either MDD or depressive symptoms. In addition, one GWAS focusing on age at onset of MDD has been conducted. These 15 studies were identified by conducting a systematic search of PubMed for articles published before October 2013. We searched the PubMed database using the following MESH terms: (depression OR depressive disorder OR depressive disorder, major OR depressive disorder, treatment-resistant) AND (genome-wide association study). We also searched for articles by examining the reference pages of review articles, meta-analyses, and other empirical articles published since 2005. As shown in Table 1, all of these studies were based on samples of European ancestry and represent a combination of population- and clinic-based samples.

The first GWAS of depression was published in 2009 and included 1738 cases and 1802 controls. Although no SNPs reached genome-wide significance, 11 of the top 200 SNPs were found in a 167 kilobase (kb) region overlapping the gene PCLO (piccolo presynaptic crytomatrix protein), which is involved in establishing active synaptic zones and synaptic vesicle tracking.42 In several subsequent studies,43,48,57 investigators found mixed evidence regarding the association of PCLO SNPs and MDD.

In the first study to report a genome-wide significant association for depression, Kohli and colleagues49 found support for a recessive effect of a SNP (rs1545843) in the gene SLC6A15 (solute carrier family 6, neutral amino acid transporter, member 15), which is involved in transporting neutral amino acids. The authors provided additional evidence in support of this association by demonstrating that risk alleles were correlated with reduced SLC6A15 expression in hippocampal tissue (taken from individuals undergoing surgery for epilepsy) and reduced hippocampal volume and neuronal integrity (as determined by neuroimaging). Mice susceptible to chronic stress were also found to have reduced hippocampal SLC6A15 expression. This locus, however, has not emerged as a prominent finding in subsequent depression GWAS (described below).

As in the case of other complex traits,58,59 one of the major lessons from these early GWAS of depression was that the effect of most SNPs is small in magnitude (allelic odds ratios of around 1.3 or less) and that considerably larger samples would therefore be needed to identify genetic loci associated with depression. To enhance the power of psychiatric GWAS studies, the Psychiatric Genomics Consortium was established in 2007 as an international collaborative effort to define the spectrum of risk variants across psychiatric disorders (http://www.med.unc.edu/pgc/). One of the consortium’s major goals is to conduct mega-analyses for MDD as well as autism, attention-deficit/hyperactivity disorder, bipolar disorder, and schizophrenia.60–62 A mega-analysis pools individual-level phenotype and genotype data from across many studies; this approach differs from a meta-analysis, where the summary statistics produced by each study are analyzed. In 2012, the consortium published the results of a GWAS mega-analysis of MDD comprising 9240 cases and 9519 controls across nine primary samples, all of European ancestry.52 Although this sample was the largest to date, no SNP reached genome-wide significance. The most significant SNPs in the discovery sample were rs11579964 (p = 1.0 × 10−7), a variant closest to several genes (CNIH4, NVL, WDR26), and rs7647854 (p = 6.5 × 10−7), a variant closest to C3orf70 and EHHADH. These findings were not supported, however, in a large, independent replication sample.

GWAS of depressive symptoms have also been largely unrevealing. The first GWAS of depressive symptoms did not find any SNPs reaching genome-wide significance.54 One modestly associated (p = 1.59 × 10−6) SNP (rs7582472) did show evidence of replication in two independent cohorts. However, this SNP was more than 300kb away from two genes, and neither gene showed significant association to depression in a gene-based analysis. A second study of depressed mood, while finding no genome-wide significant SNP, did find in the meta-analysis that an intronic SNP (rs12912233) in RORA (retinoic-related orphan receptor alpha gene) was modestly associated (p = 6.3 × 10−7).48 Although this result is interesting because another RORA SNP has been linked through GWAS to posttraumatic stress disorder,63 it awaits replication. In the largest study—a meta-analysis comprising 17 population-based studies (n = 34,549 individuals) as the discovery sample—no SNP reached genome-wide significance.55 The strongest association was for rs8020095 (p = 1.05 × 10−7), located in the gene GPHN. When the discovery and replication samples were combined into one meta-analysis of 22 studies with 51,258 respondents, one region (indexed by the SNP rs40465) was associated with depressive symptoms at genome-wide levels of significance.52 This variant is in a gene desert, an area of the genome where there are long regions without protein-coding sequences and whose biological function is unknown.

Another major lesson from depression GWAS has been that popular candidate genes have generally not shown evidence of association. Prior to the GWAS era, meta-analyses of candidate gene studies concluded that there was nominally significant evidence (at p < 0.05) for six candidate genes in depression: APOE, DRD4, GNB3, MTHFR, SLC6A3, and SLC6A4.64,65 To date, however, none of these genes, or any of the more than 100 frequently examined candidate genes, has shown evidence of significant association in the published GWAS of depression. Replication of candidate genes in GWAS is challenging since several widely studied candidate gene markers, including the serotonin transporter 5-HTTLPR variable tandem number repeat, are not directly captured by a typical GWAS platform. Some groups have developed techniques to impute or derive best-guess estimates of these genetic markers using available SNP data,66,67 though these efforts have not yet been widely adopted. Nonetheless, the evidence for many candidate genes has not been compelling.

Copyright @ 2014 President and Fellows of Harvard College. Unauthorized reproduction of this article is prohibited.
Table 1

<table>
<thead>
<tr>
<th>Study</th>
<th>Discovery sample</th>
<th>Definition of depression</th>
<th>GWAS discovery results</th>
<th>Replication sample</th>
<th>GWAS replication results</th>
<th>Other genetic findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sullivan et al. (2009)42</td>
<td>1738 cases (mainly drawn from primary care screening and outpatient mental health clinics) 1802 controls (drawn from twin registry; selected to be at low risk of depression) European ancestry</td>
<td>Lifetime diagnosis of DSM-IV MDD based on CIDI</td>
<td>No SNP reached GWS Of the top 200 SNPs, 11 were within a 167 kb region overlapping PCLO Top hits: rs2715148 (p = 7.7 × 10^-7) rs2522833 (p = 1.2 × 10^-6) Secondary analyses revealed findings were largely driven by women and participants with recurrent, early onset MDD</td>
<td>5 independent samples: 6979 cases 5893 controls</td>
<td>No SNP reached significance after correction The replication sample most similar to the discovery had a nominal effect: rs2522833 6.4 × 10^-8</td>
<td>Examined 75 candidate genes and found significant effect for NOS1 (p = .0006)</td>
</tr>
<tr>
<td>Rietsche et al. (2010)43</td>
<td>597 cases (hospital based) 1295 controls (population based) European ancestry</td>
<td>Lifetime MDD based on SCI, medical records, or family history</td>
<td>No SNP reached GWS Top 3 hits: rs2765501 (p = 1.66 × 10^-7) in CD5L rs7713917 (p = 5.87 × 10^-5) located 20 kb upstream of HOMER1 rs9943849 (p = 6.22 × 10^-5) located 14 kb upstream of CPM Did not replicate previous findings for FCLO</td>
<td>409 cases 541 controls</td>
<td>The two top hits showed a trend toward significance in replication sample: rs7713917 (p = 7.61 × 10^-3) rs9943849 (p = 1.59 × 10^-2) Meta-analysis with discovery and replication samples: rs9943849 (p = 3.24 × 10^-5) located upstream of CPM rs7713917 (p = 1.48 × 10^-5) located in HOMER1</td>
<td>Examined genes that contained SNPs with p-values &lt;.001 and found the most significant pathway related to cell signaling (included GRM5 and HOMER1 After correcting for gene size, found 6 of 22 genes that had a probability &gt;10% of being selected by chance (ACTN2, CYP17A1, DLCl, GRM5, MAP2K4, MAP24K, RSU1) Conducted fMRI follow-up analyses with top variants to examine several intermediate phenotypes and found some differences by genotype group in working memory</td>
</tr>
<tr>
<td>Terracciano et al. (2010)44</td>
<td>3972 participants European ancestry</td>
<td>Depression inventory from the revised NEO-PI</td>
<td>No SNP reached GWS Top hit: rs349475 (p = 2.4 × 10^-7) in CDH18</td>
<td>839 community-dwelling respondents from the Baltimore Longitudinal Study of Aging European ancestry</td>
<td>No SNP reached GWS Top hit: rs4885589 (p = 2.4 × 10^-4)</td>
<td>Conducted meta-analysis in combined sample (n = 4811) Top hit in meta-analysis: rs12912233 (p = 6.3 × 10^-7) (an intronic SNP in RORA) Described rs17864092 (p = 5.5 × 10^-5) in GRM8 as the most biologically plausible top hit</td>
</tr>
<tr>
<td>Study</td>
<td>Discovery sample</td>
<td>Definition of depression</td>
<td>GWAS discovery results</td>
<td>Replication sample</td>
<td>GWAS replication results</td>
<td>Other genetic findings</td>
</tr>
<tr>
<td>------------------</td>
<td>------------------</td>
<td>-----------------------------------------------------------------------------------------</td>
<td>------------------------</td>
<td>---------------------</td>
<td>-------------------------</td>
<td>---------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Muglia et al.</td>
<td>926 recurrent MDD cases (clinical sample) 866 controls (age and gender matched from community sample) European ancestry</td>
<td>Recurrent MDD (at least 2 separate episodes of severe intensity) as assessed by SCAN Cases allowed to have comorbid anxiety (no OCD or PTSD); 26% did</td>
<td>No SNP reached GWS</td>
<td>492 recurrent cases (clinical sample) 1052 controls (from community survey) European ancestry</td>
<td>No SNP reached GWS in replication sample Little agreement between discovery and replication sample with respect to the most significant SNP associations identified Conducted meta-analysis: no SNP reached GWS; most significant SNP was for rs4238010 (nearest gene was CCND2; p = 5.8 × 10^{-6})</td>
<td>Gene-based analysis found the most significant results were similar to the SNP-level analysis; strongest adjusted gene-based association (p = .009) was for SMG7 Examined whether top 104 genes (p &lt; 10^{-4}) from meta-analysis were found in a previous study to associate with mood disorders or schizophrenia; none of these genes was associated with depression in previous studies; some genes (ADCY9, GRM7) were related to bipolar disorder/schizophrenia Candidate gene analysis with 15 candidates found none was GWS; strongest effect in GRM7 (rs1485171; p = .0001) Functional annotation analysis (i.e., generated pathways) of the most significant (p &lt; .0005) SNPs from meta-analysis found this subset of genes was enriched (p &lt; .05) in four pathways: (1) synaptic long-term depression, (2) cAMP-mediated signaling, (3) G-protein-coupled receptor signaling, and (4) glutamate receptor signaling</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Study</td>
<td>Discovery sample</td>
<td>Definition of depression</td>
<td>GWAS discovery results</td>
<td>Replication sample</td>
<td>GWAS replication results</td>
<td>Other genetic findings</td>
</tr>
<tr>
<td>-----------------------</td>
<td>----------------------------------------------------------------------------------</td>
<td>----------------------------------------------------------------</td>
<td>---------------------------------------------------------------------------------------</td>
<td>--------------------</td>
<td>------------------------------------------------------</td>
<td>------------------------------------------------------------</td>
</tr>
<tr>
<td>Lewis et al. (2010)</td>
<td>1636 cases (from 3 clinic-based groups) 1594 controls (primary care screening; screened negative for depression or anxiety) European ancestry</td>
<td>Recurrent depression of at least moderate severity defined by SCAN</td>
<td>No SNP reached GWS Four genotyped SNPs showed suggestive evidence: 2 SNPs in BICCI: rs9416742 (p = 1.3 × 10^{-7}) and rs999845 (p = 3.1 × 10^{-7}); rs2698195 (p = 3.54 × 10^{-6}) located closest to KIAA1841; rs8050326 (p = 4.07 × 10^{-6}) located closest to (over 100 kb from) IRF8</td>
<td>Conducted meta-analysis 1418 cases (recurrent depression from clinical sample) 1918 controls (from population-based study)</td>
<td>No SNP attained GWS in meta-analysis Top hit: rs606149 (p = 2.57 × 10^{-6}) near LOC647167 No replication for BICCI</td>
<td>Analyzed haplotypes in discovery at 2 SNPs in BICCI (rs7903712 and rs9416742) and found association (p = 9.04 × 10^{-7}) Conducted sex-specific analyses for top 20 SNPs in discovery and found GWS association for women with rs9416742 in BICCI (p = 1.8 × 10^{-7}); in women, 4 additional SNPs had suggestive evidence (rs8067196, rs2930553, rs13079811, rs9873900); in men, rs6989226 (near TUSC3) had suggestive evidence (p = 1.81 × 10^{-6}) Analyzed 84 candidate genes and found strongest evidence for rs13050655 in PDE9A (p = 3.58 × 10^{-7})</td>
</tr>
<tr>
<td>Shi et al. (2011)</td>
<td>1020 cases (with recurrent, early-onset MDD; clinical sample) 1636 controls (no lifetime MDD; population based) European ancestry</td>
<td>MDD based on DIGS and consensus by 2 independent reviewers</td>
<td>No SNP reached GWS Top hit: rs7077540 (p = 1.83 × 10^{-7})</td>
<td>Compared results to other meta-analyses described in Shyn et al. (2011)</td>
<td>No SNP reached GWS Strongest support for rs17144465 (p = 8.38 × 10^{-7})</td>
<td>Examined 41 candidate genes using single SNP tests and found the lowest p-value (p = 6.7 × 10^{-5}) for CACNA1C Also conducted an aggregate test of whether p-values in gene were more significant than expected by chance and found no significant findings</td>
</tr>
<tr>
<td>Aragam et al. (2011)</td>
<td>1726 population-based cases 1630 population-based controls European ancestry</td>
<td>MDD diagnosis based on CIDI</td>
<td>No SNP reached GWS Top hit: rs1558477 (p = 2.63 × 10^{-7}) in ADCYAP1R1 4 SNPs in PCLO had marginal effects: rs2796184 (p = 1.38 × 10^{-4}) rs2522833 (p = 2.46 × 10^{-4}) rs2522840 (p = 4.38 × 10^{-4}) rs2107828 (p = 1.48 × 10^{-5})</td>
<td>None</td>
<td>No replication for BICCI</td>
<td>Also conducted subgroup analyses stratified by sex Best SNP for males: rs9352274 (p = 2.26 × 10^{-6}) in LGSN Best SNP for females: rs2715148 (p = 5.64 × 10^{-7}) in PCLO Also tested for interactions by gender and found best SNP in rs12692709 (p = 5.75 × 10^{-6})</td>
</tr>
<tr>
<td>Study</td>
<td>Discovery sample</td>
<td>Definition of depression</td>
<td>GWAS discovery results</td>
<td>Replication sample</td>
<td>GWAS replication results</td>
<td>Other genetic findings</td>
</tr>
<tr>
<td>-----------------------</td>
<td>------------------</td>
<td>--------------------------</td>
<td>-----------------------</td>
<td>--------------------</td>
<td>-------------------------</td>
<td>-------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Kohli et al. (2011)</td>
<td>353 cases (inpatients in tertiary clinic) 366 matched controls (no lifetime MDD; community sample) European ancestry</td>
<td>Met DSM-IV criteria for first depressive episode or recurrent depressive disorder and had HAM-D score ≥ 14</td>
<td>One SNP reached GWS: rs1545843 (p = 5.53 × 10−8)</td>
<td>6 independent samples of different racial/ethnic background</td>
<td>Nominally significant association in 4 of the 5 initial replication samples with recessive model GWS replication for rs1545843 (p = 4.37 × 10−8) in meta-analysis, after adjustment for multiple testing. Further replicated findings for rs154843 in second replication sample (p = .008)</td>
<td>Further validated findings in analyses of (1) associations to pre-mortem human hippocampus and lymphoblastoid cell line expression profiles, (2) imaging, and (3) hippocampal expression in mouse model of chronic social stress</td>
</tr>
<tr>
<td>Shyn et al. (2011)</td>
<td>1221 cases (from STAR*D; outpatient clinics) 1636 controls (population based; no lifetime history of MDD) European ancestry</td>
<td>Diagnosis of MDD by clinician rating and HAM-D score ≥ 14 by independent raters</td>
<td>No SNP reached GWS in analyses of genotyped SNPs Top hit: rs2462886 (p = 1.73 × 10−6) located in gene desert</td>
<td>Meta-analysis with 2 additional datasets (GenRED and GAIN) (n = 3957 cases; n = 4328 controls) Examined broad (all cases) and narrow phenotype (recurrent depression with onset before age 31)</td>
<td>No SNP (imputed or genotyped) reached GWS Strongest evidence in broad meta-analysis was for intronic SNPs in ATP6V1B2 (rs1106634; p = 6.78 × 10−7), SP4 (rs17144465; p = 7.68 × 10−7), and GRM7 (rs9870680; p = 1.11 × 10−6) Best SNP in narrow analysis was in stratified analysis of males only (rs11710109; p = 5.64 × 10−5)</td>
<td>Examined 41 candidate genes in discovery, but none were supported; best finding was for rs3788477, a SNP intronic to SYN3 (p = 1.64 × 10−7); no other SNP achieved p &lt; 10−5 Also examined in meta-analysis; aggregate analysis did not suggest an excess of low p-values among these candidates</td>
</tr>
<tr>
<td>Wray et al. (2012)</td>
<td>2431 cases 3673 controls (population based; family members without disease; drawn from five different sites) European ancestry</td>
<td>Lifetime diagnosis of MDD by CIDI or other interview instrument</td>
<td>No SNP reached GWS Top hits in total sample: rs3732293 (p = 1.5 × 10−5) rs7226852 (p = 1.5 × 10−5)</td>
<td>Compared results to other published studies and meta-analyses</td>
<td>No SNPs reached GWS</td>
<td>Tested 183 candidate genes and found none reached significance after correction in the discovery sample, other published studies, or meta-analysis</td>
</tr>
<tr>
<td>Study</td>
<td>Discovery sample</td>
<td>Definition of depression</td>
<td>GWAS discovery results</td>
<td>Replication sample</td>
<td>GWAS replication results</td>
<td>Other genetic findings</td>
</tr>
<tr>
<td>-------</td>
<td>------------------</td>
<td>--------------------------</td>
<td>-----------------------</td>
<td>--------------------</td>
<td>-------------------------</td>
<td>-----------------------</td>
</tr>
<tr>
<td><strong>Psychiatric GWAS Consortium (2012)</strong>&lt;sup&gt;52&lt;/sup&gt;</td>
<td>9240 cases and 9519 controls (mostly population based; no lifetime history of depression) Came from 9 primary samples European ancestry</td>
<td>Lifetime MDD established using structured diagnostic instruments from direct interviews or clinician-administered checklists</td>
<td>No SNP reached GWS in mega-analysis Top hits: rs11579964 (p = 1.0 × 10&lt;sup&gt;−7&lt;/sup&gt;) rs7647854 (p = 6.5 × 10&lt;sup&gt;−7&lt;/sup&gt;)</td>
<td>6783 cases and 50,695 controls (7 independent samples from discovery) European ancestry</td>
<td>Tried to replicate 554 SNPs with p &lt; .001 Did not find SNPs that replicated in the same direction as discovery analysis more frequently than chance No SNP achieved GWS for a joint analysis of the discovery and replication samples; top hit was for rs1969253 (p = 4.8 × 10&lt;sup&gt;−6&lt;/sup&gt;) located in DVL3</td>
<td>Conducted a cross-disorder meta-analysis and a set of secondary analyses (by sex; recurrent, early age at onset; and subtype) Direction of effects was generally consistent between discovery and replication for analyses restricted to women and for recurrent MDD, but no SNP reached GWS Only in MDD/bipolar disorder cross-disorder analysis did 15 SNPs exceed GWS; top hit was for rs2535629 (p = 5.9 × 10&lt;sup&gt;−6&lt;/sup&gt;) Conducted a polygene analysis using discovery-phase samples and found SNPs explained 0.6% of the variance in case-control status (p &lt; 10&lt;sup&gt;−6&lt;/sup&gt;)</td>
</tr>
<tr>
<td><strong>Power et al. (2012)</strong>&lt;sup&gt;53&lt;/sup&gt;</td>
<td>Time-to-event analysis for age at onset 1480 cases and 1584 controls, both from the UK cohorts in the RADIANT study (see Lewis et al. [2010])&lt;sup&gt;56&lt;/sup&gt; Additional analyses used all RADIANT participants (n = 2746) European ancestry</td>
<td>Age at onset to recurrent depression of at least moderate severity defined by SCAN</td>
<td>No SNP achieved GWS in any analysis Top hit (in case-control analysis): rs2273289 in PLOD1 (p = 1.29 × 10&lt;sup&gt;−7&lt;/sup&gt;)</td>
<td>2 clinical cohorts based in Germany</td>
<td>None of the previously identified suggestive loci replicated</td>
<td>Also performed a GCTA analysis and found that 55% of the variance in age at onset was explained by common SNPs Sex-specific analyses found suggestive evidence for 36 SNPs</td>
</tr>
<tr>
<td>Study</td>
<td>Discovery sample</td>
<td>Definition of depression</td>
<td>GWAS discovery results</td>
<td>Replication sample</td>
<td>GWAS replication results</td>
<td>Other genetic findings</td>
</tr>
<tr>
<td>-----------------------</td>
<td>---------------------------------------------------------------------------------</td>
<td>----------------------------------------------------------------------------------------</td>
<td>---------------------------------------------------------------------------------------</td>
<td>-----------------------------------------------------------------------------------</td>
<td>------------------------------------------------------------------------------------------</td>
<td>---------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Luciano et al.</td>
<td>5 population-based cohorts (n = 45,255); European ancestry</td>
<td>Depressive symptoms measured by BDI or HADS</td>
<td>No SNP reached GWS in meta-analysis</td>
<td>1 population-based German cohort using the POMS, and the Netherlands twin register using BDI</td>
<td>One SNP (rs7582472), whose closest gene was 300 kb away, replicated in the German (p = .01) and Netherlands cohorts (p = .006)</td>
<td>None of the other 4 SNPs replicated in either sample</td>
</tr>
<tr>
<td>Hek et al. (2013)</td>
<td>17 population-based studies (n = 34,549 individuals); European ancestry</td>
<td>Depressive symptoms measured by 10-, 11-, or 20-item CESD</td>
<td>No SNP reached GWS Top hit: rs8020095 (p = 1.05 × 10^{-7})</td>
<td>5 population-based studies (n = 16,709) Focused on 7 SNPs</td>
<td>No SNP reached significance after correction; best SNP was rs161645 (p = 9.19 × 10^{-5})</td>
<td>Performed combined meta-analysis of 22 studies (n = 51,258) and found rs40465 reached GWS (p = 4.78 × 10^{-8}) Conducted pathway analysis with 104 genes to identify and classify biological processes among SNPs with p-values &lt;10^{-3} found neurotransmitter secretion (p = 9.94 × 10^{-5}), vitamin transport (p = .014), and synaptic transmission (p = .037) processes were overrepresented among top SNPs Examined 17 candidate SNPs based on previous findings and found none that replicated</td>
</tr>
<tr>
<td>Power et al. (2013)</td>
<td>805 case-control pairs matched first on ancestry and second on exposure to stressful life events from the RADIANT study see Lewis et al. (2010) European ancestry</td>
<td>Recurrent depression of at least moderate severity defined by SCAN</td>
<td>No SNPs achieved GWS or suggestive evidence (p &lt; 5 × 10^{-6})</td>
<td>None</td>
<td>None</td>
<td>None</td>
</tr>
</tbody>
</table>

BDI, Beck Depression Inventory; CESD, Center for Epidemiological Studies of Depression Scale; CIDI, Composite International Diagnostic Interview; DIGS, Diagnostic Interview for Genetic Studies; GAIN, Genetic Association Information Network; GCTA, genome-wide complex trait analysis; GenRED, Genetics of Recurrent Early-Onset Depression; GWAS, genome-wide association studies; GWS, genome-wide significance; HADS, Hospital Anxiety and Depression Scale; HAM-D, Hamilton Depression Rating Scale; MDD, major depressive disorder; NEO-PI, NEO Personality Inventory; OCD, obsessive-compulsive disorder; POMS, Profile of Mood States; PTSD, posttraumatic stress disorder; SCAN, Schedules for Clinical Assessment in Neuropsychiatry; SNP, single-nucleotide polymorphism; STAR*D, Sequenced Treatments to Relieve Depression.
Another interesting observation from GWAS has been the failure to consider the role of environment. As we describe below, we believe that GWAS may be limited by not taking into account how genetic influences on depression may vary among individuals with certain environmental exposures. One exception is a study by Powers and colleagues, who used propensity-score matching to conduct a GWAS among case-control pairs matched on exposure to recent stressful life events. Although they did not formally test for G×E, the use of propensity-score matching enabled them to reduce sample heterogeneity and to compare cases to controls with similar levels of exposure. In their analysis, no SNPs were genome-wide significant or even suggestive (p < 5 × 10^{-6}); this finding was likely due to the very small sample size (n = 805).

**FINDINGS FROM GENE-ENVIRONMENT INTERACTION STUDIES**

The long-standing recognition that both genes (“nature”) and environments (“nurture”) contribute to the etiology of depression has motivated a great deal of interest in studying gene-environment interactions. G×E studies examine the degree to which genetic variants modify the association between environmental factors and depression (or similarly, the extent to which environmental factors modify the association between genes and depression). Typically, G×E studies have assumed a diathesis-stress model, where a genetic liability, also referred to as a diathesis, interacts with a stressful life event to give rise to depression. In this model, genes either exacerbate or buffer the effects of stress. More recently, the concept of G×E has been expanded to incorporate positive aspects of the environment, such as social support, psychosocial interventions, and other protective factors that reduce the risk of disease.

Emerging work has focused on differential susceptibility to the environment or on the extent to which genetic variation makes individuals more likely to respond adversely to negative environments but more positively to salutary environments.

Research on G×E in depression was essentially launched with a 2003 publication in Science. In that study Caspi and colleagues used data from a 26-year longitudinal study in New Zealand to test whether a functional length polymorphism in the promoter region (5-HTTLPR) of the serotonin transporter gene (SLC6A4) interacted with stressful life events to increase the risk of depression. Results of the Caspi study suggested that individuals with at least one short (s) allele (i.e., the “s/s” or “s/d” genotype of the biallelic coded version) had more depression in response to stressful life events when compared to subjects who were not s allele carriers. This result held regardless of how depression was measured, whether by level of depressive symptoms, depression diagnosis, incident depression, or suicidality. They also found that s allele carriers, compared to those without an s allele, had a greater probability of experiencing depression resulting from exposure to probable or severe childhood maltreatment. The Caspi article has become one of the most influential studies in the field, having been cited more than 5000 times.

Since the publication of Caspi and colleagues’ seminal research, numerous replication attempts have been made. Most of these have also focused on 5-HTTLPR, although other genetic variants have been studied, including variants in BDNF (brain-derived neurotropic factor), MAOA (monoamine oxidase A), FKBP5 (FK506 binding protein 51), CRHR1 (corticotropin-releasing hormone receptor 1), COMT (catechol-O-methyltransferase), and CREB1 (also known as CAMP or responsive element-binding protein 1). Many replication attempts have focused on stressful life events (either recent or in childhood) or on child maltreatment—namely, physical abuse, sexual abuse, or neglect. All of these “candidate” environments are appropriate to study in G×E research. Child maltreatment, for example, is one of the most potent environmental stressors in the etiology and course of depression and other types of psychopathology. Extant studies suggest child maltreatment at least doubles the risk of internalizing problems, including depression.

The large number of empirical studies trying to replicate Caspi’s G×E findings for depression have been summarized in several reviews focusing on G×E with 5-HTTLPR (see, e.g., references 70 and 79–86). These reviews ultimately fueled a heated debate on the plausibility of the Caspi findings. Including somewhat similar individual studies, review articles have drawn opposing conclusions about the support for G×E effects, with some studies finding consistent G×E effects and others failing to detect them. Meta-analyses have provided a quantitative summary of these studies but have also reached opposing conclusions. Specifically, the results of two meta-analyses, which found evidence against a consistent G×E effect, differed from a third meta-analysis, which concluded that the evidence was strong to support the 5-HTTLPR G×E. These conflicting results may be explained by differences in the selection of studies for inclusion in the meta-analyses. For example, the meta-analyses that used the most stringent inclusion criteria failed to support the G×E association. Some commentators have noted that an inverse relationship exists between the power of the replication studies and support for the 5-HTTLPR association—precisely the opposite of what one would expect if the association is valid. Moreover, the most direct replication attempt of the Caspi findings, which was not included in any prior meta-analysis, found no evidence in support of the G×E effect on depression. This longitudinal, birth cohort study followed a similar population (New Zealand residents) for a similar length of time (30 years) and used comparable phenotypic measures. The authors observed no interaction between stressful life events and 5-HTTLPR genotype, even after conducting 104 different regression models.

By contrast, some have argued that support for the 5-HTTLPR G×E has been more consistent when childhood maltreatment is the exposure variable or when direct interview assessments (as opposed to self-report questionnaires)
have been used. These findings are important since there has been substantial variability in the characteristics of study populations, the measurements of depression and environmental exposures, and the analytic methods used across empirical studies to test for G×E in depression. Others have also tried to place these individual G×E studies in the context of the broader literature examining genetic variability and stress sensitivity in depression. They have appealed to the more consistent findings from animal studies showing that loss-of-function mutations in the serotonin gene have been associated with depressive-like behavior in rodents and that genetic variation in the serotonin transporter gene has been linked to depression among nonhuman primates. Proponents have noted that the results are more convincing when considered alongside both experimental imaging studies showing 5-HTTLPR variation in amygdala activity and treatment-response studies showing 5-HTTLPR variation in antidepressant treatment response. Overall, the validity of the influential 5-HTTLPR G×E finding remains unclear.

G×E studies focusing on other candidate genes, however, have found more consistent results. For example, studies examining FKBP5 and CHRNA1 have shown that variants in these genes moderate the effect of exposure to child maltreatment, childhood adversities, or negative life events on adult depression. These genes are interesting candidates because they regulate the stress response via the hypothalamic-pituitary-adrenal axis. Additional replications of these candidates would be helpful to further evaluate their role in shaping risk for depression. Evidence for other candidates, such as BDNF, has been mixed. For instance, a recent review found stronger evidence to support interactions with the BDNF Val66Met polymorphism and stressful life events compared to childhood adversity. As we later discuss, genome-wide approaches to G×E remain an important, but relatively unexplored, area.

CURRENT AND FUTURE DIRECTIONS FOR RESEARCH

The limited success of GWAS for depression is in contrast to other psychiatric disorders, where established risk variants are accumulating through GWAS. For example, at the time of writing, more than 100 loci have been associated with schizophrenia and bipolar disorder at stringent levels of statistical significance. Although no individual risk loci have been identified for depression, we know that such variants will be found with adequate sample sizes. For example, it is now possible through genome-wide complex trait analysis to estimate the common variant contribution to depression using genome-wide SNP data (these estimates are sometimes referred to as SNP chip heritability). These methods have yielded estimates of the common variant contribution to depression, ranging from a high of 32% to a low of 21%. It should be noted that these estimates are lower bound because SNP chip heritability reflects only the effect of common variation that is captured on genotyping arrays.

**Text Box 2**

Possible Explanations for the Lack of Success of GWAS and G×E Studies for Depression, and Strategies to Increase Gene Finding

<table>
<thead>
<tr>
<th>Explanations</th>
<th>Strategies to Address</th>
</tr>
</thead>
<tbody>
<tr>
<td>Depression has a different genetic architecture</td>
<td>• Increase sample size to improve power to detect associations of individually small effect loci</td>
</tr>
<tr>
<td></td>
<td>• Aggregate genetic signals into pathways or gene sets</td>
</tr>
<tr>
<td></td>
<td>• Examine rare variants and other types of structural variants (e.g., copy-number variants) in addition to common variants</td>
</tr>
<tr>
<td>Previous GWAS did not consider the role of environment</td>
<td>• Conduct GEWIS</td>
</tr>
<tr>
<td></td>
<td>• Test for G×E using candidate genes from GWAS</td>
</tr>
<tr>
<td>Depression is highly heterogeneous</td>
<td>• Examine depressive symptoms (quantitative phenotype) rather than only diagnoses of depression</td>
</tr>
<tr>
<td></td>
<td>• Use novel analytic methods (e.g., factor analysis, latent class analysis) to identify and refine distinct subtypes</td>
</tr>
<tr>
<td></td>
<td>• Focus on intermediate phenotypes or endophenotypes, consistent with RDoC</td>
</tr>
</tbody>
</table>

GEWIS, genome-environment wide interaction studies; GWAS, genome-wide association studies; G×E, gene-environment interaction; RDoC, National Institute of Mental Health Research Domain Criteria Initiative.

Thus, the field faces two major questions: what explains the lack success of GWAS and G×E studies for depression, and how can we best move forward? As described below (and summarized in Text Box 2), there are several likely explanations for the limited progress to date and several strategies that may help overcome these challenges.

**Genetic Architecture and the Need for Larger Studies**

The genetic architecture of depression is likely to be highly complex. Genetic architecture refers to the number of genetic loci associated with a phenotype, the effect size of each locus, and the manner in which these loci behave (e.g., whether they have additive or multiplicative effects). While all psychiatric disorders are thought to be polygenic, or influenced by multiple genes, the genetic basis of depression may reflect an even larger number of loci of individually small effect. Results from studies that have calculated polygenic risk scores (capturing aggregate effects of loci across the genome) support such a hypothesis. It is therefore likely that much larger samples than those examined to date will be needed to detect these individually small effects. Simulations suggest that to have comparable power to GWAS of schizophrenia or bipolar disorder, studies of depression will need to have sample sizes as much as five times larger than the sample sizes required for those disorders. Experience with GWAS for other psychiatric disorders has established that once a critical
sample size threshold is crossed, a larger and larger sample size yields more and more loci.

If depression is driven by many thousands of loci of weak effect, another strategy may be to combine genetic signals across many SNPs into functionally defined gene sets or pathways. Pathway approaches can be considerably more powerful than single-variant analyses, as the aggregation of weak signals from multiple causal variants may yield statistically significant evidence in support of a given gene or pathway. Thus, investigators have primarily examined pathways related to specific biological functions (e.g., axon guidance, cell functioning) as defined by human-curated bioinformatics resources, such as the Kyoto Encyclopedia of Genes and Genomes or Gene Ontology. Recent studies of candidate gene pathways have found evidence that genes involved in glutamatergic synaptic neurotransmission, among others, were significantly associated with depression. Evidence in support of gene sets or pathways also comes from several GWAS that we have described above (see Table 2). These studies found significant support for some pathways. One of the major drawbacks of gene-set analyses is that they require predefined sets of genes. Gene sets defined by current annotation databases, such as the Kyoto Encyclopedia or Gene Ontology, vary in their completeness; some pathways are more complete than others. Moreover, databases also vary in how they define gene sets. Thus, a given gene may belong to one pathway in one database and a second pathway in another. Although these challenges are substantial, we think that greater use of pathway-type analyses is needed.

Understudied Components of the Genetic Architecture of Depression

A related consideration is that GWAS are designed to capture common, but not rare, genetic variation. Rare variants can include genetic single-nucleotide variations (SNVs; present in <1% of the population) and rare copy-number variations (CNVs; that is, structural variations in DNA sequence that involve the duplication or deletion of thousands or more than a million base pairs). Such variants have now been shown to play a role in autism, schizophrenia, and bipolar disorder, but to date these components of the genetic architecture of depression have been largely unexplored.

Fortunately, advances in sequencing technology now provide an opportunity to address the role of rare SNVs. In recent years, the cost of direct DNA sequencing has dropped dramatically, and technological advances have facilitated the development of “high-throughput” sequencing. To date, these “next-generation sequencing technologies” have been largely applied to study variants in exons, which are the protein-coding regions of the genome, collectively known as the exome. Exons comprise about 30 megabases of DNA or 1% of the total genome. Although no exome-sequencing studies of depression have been reported at the time of writing, such studies are under way. Next-generation sequencing technologies can also be applied to the entire genome (whole genome sequencing), enabling researchers to explore a full range of genetic variants in both coding and noncoding regions of the genome.

The major strength of sequencing is that it captures variants that have been previously uncharacterized by candidate gene and GWAS methods and thus may provide new insights into the genetic underpinnings of depression. Like all techniques, however, sequencing approaches face a number of challenges. For example, despite enormous reductions in the cost of sequencing, well-powered studies remain very expensive. Whole-genome sequencing costs at least US$1000 per genome, whereas exome sequencing costs only several hundred dollars. Exome sequencing also assesses polymorphisms that, by definition, are rare and thus occur much less frequently than common variants. To have sufficient statistical power to identify an association between these rare variants and depression, very large sample sizes—in the order of 10,000 or more cases—are needed. In addition, rare-variant association methods are still largely under development.

Structural variation, including CNVs, is also a potential source of depression risk loci. CNVs can be inherited or spontaneous (de novo). De novo CNVs—those that are present in offspring but not in either parent—have been shown to be important risk factors for several neuropsychiatric disorders—namely autism, schizophrenia, and bipolar disorder. After conducting a systematic literature search of PubMed for articles published by December 2013 using the MESH terms for depression that were described above and the phrase “copy number var*,” we identified four studies that provide preliminary evidence implicating CNVs in depression. In the largest of these studies, Glessner and colleagues found 12 CNV regions that were exclusive to cases with MDD. The region with the highest frequency in cases was a locus on chromosome 5 (5q35.1) that overlapped the genes SLIT3, CCDC99, and DOCK2. The finding of a CNV overlapping the gene SLIT3 is interesting since SLIT3 is known to play a role in axon development and neurodevelopmental disorders.

One of the major strengths of studying CNVs is that the methods for association testing are similar, by and large, to examining common variants. Simultaneous examination of SNPs and CNVs in large samples may identify whether CNVs play a significant role in depression and what their importance is relative to common variants. One of the major drawbacks of association testing with CNVs is that catalogs of these variants do not exist with the same number or specificity as they do for SNPs. For example, the location, size, and boundary of CNVs in these publicly available resources have been relatively imprecise. Opportunities for misclassification of variants is consequently much higher for CNVs than for SNPs. Efforts are now under way to provide a more comprehensive catalog of CNVs (see, e.g., http://www.sanger.ac.uk/research/areas/humangenetics/cnv/). Moreover, until recently no genotyping array that could detect both SNPs and CNVs was commercially available. With the advent of
the PsychChip, a customized genotyping chip for psychiatric phenotypes, investigators will soon be able to simultaneously examine multiple genetic variants, including SNPs, CNVs, and rare variants. The importance of rare variants to depression risk remains to be seen. Large-scale studies will be needed to clarify their contribution.

Accounting for the Role of Gene-Environment Interaction

As noted previously, existing studies have not systematically addressed the possibility that a substantial proportion of the risk of depression is attributable to nonadditive effects, including G×E. Moreover, G×E studies to date have focused on a limited set of candidate genes and have typically been underpowered, creating a risk of both false-positive and false-negative results. It is well established that environmental factors, including exposure to stressful life events and child maltreatment, are important risk factors for depression, but we still know little about whether these environmental effects are moderated by genetic variation and, if so, which genetic variants are relevant.

One approach to filling this gap may come from genome-environment wide interaction studies (GEWIS), pronounced gee whiz.127,128 In a GEWIS, investigators test for statistical interaction or G×E, with the G defined as the genetic loci (e.g., SNPs) included in a GWAS and the E defined as a known environmental exposure. Unlike candidate gene G×E, GEWIS offers the opportunity to conduct a genetically unbiased search—that is, one in which prior genetic or biologic hypotheses are not required. In one type of GEWIS, investigators could focus on loci for which a main effect of a genetic variant has been established by GWAS. In this scenario, loci that have been identified by GWAS become candidates for G×E analysis but with the advantage over traditional candidate gene studies that the locus is already known to influence the phenotype of interest.

To our knowledge, no GEWIS of depression has been published to date. Although research on GEWIS of depression and other psychiatric phenotypes is lacking, a small but emerging body of research on other complex phenotypes suggests GEWIS can yield important new gains. For example, studies have identified significant genome-wide G×E interactions in cancer,129,130 diabetes131 and insulin resistance,132 Parkinson’s disease,133 pulmonary function,134 and nonsyndromic cleft palate.135 Interest in GEWIS is growing, but several challenges to conducting this type of study remain.127 The first is identifying the best methods to test for genome-wide G×E. Several methodological approaches have been developed (see, e.g., the reviews by Winham & Biernacka136 and Gauderman et al.137), though without any consensus as to which is the best. Selection of a specific analytic method depends largely on whether the goal is to leverage G×E to discover novel loci, or to characterize the joint effect of genetic variants and environmental factors.138

The second challenge is that the “environment” is somewhat indeterminate; it is unbounded in a way that the genome is not. Both children and adults are exposed to a range of experiences across the multiple social and physical contexts in which they are embedded (e.g., families, school, neighborhoods, workplaces); all of these experiences and exposures can contribute to health.139 One way to start is to focus on well-defined measures of environment where robust and consistent evidence supports a relationship between the exposure and depression. Such a list of measures could include in utero exposures (e.g., viruses, toxins, alcohol and drugs), social deprivation (e.g., poverty, child maltreatment), and enrichment (e.g., psychosocial interventions and treatments). However, even if we select the same environment, such as child maltreatment, we are still faced with multiple different types of maltreatment, multiple ages at which the maltreatment occurred, and multiple ways to measure maltreatment (e.g., self-report, administrative records, clinical interview).

The final, and perhaps the biggest, challenge is balancing the trade-off between large samples and precise measures of environmental exposure. Large samples are needed to detect G×E (larger even than those needed in standard GWAS). Large samples, however, often lack the depth and breadth that are necessary to capture data on environmental or phenotype measures. Although smaller samples frequently have rich and repeated measures, they are underpowered to establish robust associations. Smaller samples can be combined to increase statistical power, but challenges will arise in trying to harmonize measures of environment across these data sets. In other words, efforts to ensure an adequate sample size for each unique combination of risk factors and G×E strata can lead to watered-down environmental measures that lack any meaningful variability; a classic example would be an instance where respondents are simply classified as “exposed” or “non-exposed.” Longitudinal birth cohort studies, which can include prospective measures of environmental exposures along with detailed phenotype data and genome-wide data, may be one promising avenue for conducting GEWIS in the future. Moreover, the growing interest in the concept of the exposome, in environment-wide association studies (EWAS), and in ways to systematically identify relevant environmental factors (see, e.g., Wild140 and Patel et al.141) could yield new insights to guide GEWIS in the future.

The Phenotypic Complexity of Depression

Another obstacle to identifying susceptibility loci is that depression is a heterogeneous phenotype. Indeed, it is possible to meet DSM-IV or DSM-5 diagnostic criteria for a major depressive episode through at least 227 different symptom combinations.142 As currently described by DSM-5, MDD can manifest with or without (1) anxious distress, (2) mixed features, (3) melancholic features, (4) atypical features, (5) mood-congruent psychotic features, (6) mood-incongruent psychotic features, (7) catatonia, (8) peripartum onset, and (9) a seasonal pattern.143 These subtypes of MDD could reflect different genetic contributions. Consistent with such a
hypothesis, studies suggest that depression with a history of child maltreatment has a different onset, course, and response to treatment than a depression that arises among individuals without a history of abuse.\textsuperscript{144,145} Recent twin studies have also suggested that genetic liability to MDD reflects not one but three distinct symptom dimensions (psychomotor/cognitive, mood, and neurovegetative).\textsuperscript{146} Thus, GWAS that simply examine “depressed” cases versus controls may decrease the ratio of signal to noise by combining multiple disorder subtypes that vary in their genetic etiology. In light of evidence suggesting that there is no truly categorical threshold for depression cases\textsuperscript{147} and that different lifetime prevalence estimates of depression are found when comparing cross-sectional retrospective reports to cumulative evaluations based on multiple interviews,\textsuperscript{148} it is reasonable to posit that misclassification of individuals as cases or controls may be undermining the power of typical case-control GWAS.

We think several strategies can help reduce the heterogeneity in depression. First, examining the full range of variation in depression (e.g., depressive symptoms) rather than dichotomizing the phenotype (cases and controls) could be a statistically more powerful approach to identify variants associated with depression.\textsuperscript{149} This approach would be consistent with evidence that the diagnostic threshold for MDD has been artificially imposed on a continuum of depression risk.\textsuperscript{147} Second, more data-driven approaches to examine shared features or subtypes of depression through use of latent class analysis\textsuperscript{150} may also prove helpful. Prior studies applying such methods in both adolescents and adults have found distinct subtypes that differ based on severity, symptoms, and episode length.\textsuperscript{151,152} Examination of these subtypes in a genetic association study may help to identify variants that are common across, or unique to, specific subtypes. Third, another strategy would be to continue efforts to examine phenotypes thought to be more proximal to a genetic substrate than are clinically defined categories.\textsuperscript{153} Putative intermediate phenotypes or endophenotypes that are related to depression include emotion-based attention biases,\textsuperscript{134,155} impaired reward function,\textsuperscript{156} and deficits in domains of executive functioning, such as learning and memory.\textsuperscript{157} Investigation of endophenotypes is consistent with the National Institute on Mental Health Research Domain Criteria Initiative,\textsuperscript{158–161} which aims to provide a bottom-up characterization of psychopathology incorporating genomics, neural circuitry, and behavioral phenotypes. Endophenotypes have not yet been the subject of large-scale studies that might fully evaluate their relative power. One exception is the ENIGMA consortium, where a GWAS meta-analysis of structural magnetic resonance imaging phenotypes yielded a genome-wide significant association with hippocampal volume,\textsuperscript{162} one of the best-established biomarkers of depression risk. Nonetheless, this result required sample sizes in the thousands, challenging the view that endophenotype-based studies will be more powerful than studies of MDD itself.

**CONCLUSIONS**

Research on the genetic underpinnings of depression is at an exciting yet challenging crossroad. On the one hand, genotyping technologies have allowed for the characterization of individual and population-based genetic variation and have provided analytic tools to examine the individual and joint effects of genetic and environmental determinants. On the other hand, GWAS of depression have yet to see the same success that has been achieved with other psychiatric or medical disorders. Studies of G×E have thus far failed to provide clarity but have fueled plenty of debate. Some argue that positive findings reflect chance results among small, underpowered studies,\textsuperscript{84} whereas others see consistencies when focusing on studies that are methodologically comparable.

We have reviewed some of the potential explanations for the lack of success to date for GWAS and G×E studies of depression. Given the established heritability of depression, there is every reason to expect that increasingly well-powered studies will indeed identify risk loci. The genetic and phenotypic complexity of depression, however, may mean that such successes will require samples on the order of tens of thousands of participants. Efforts to parse the heterogeneity of depression and validate phenotypic subtypes may also be essential to facilitate gene identification. Further, as we have noted, potentially important areas to uncover the genetic basis of depression—specifically, rare variation and G×E—remain relatively unexplored on a large scale. It remains to be seen how much of the “missing heritability” of depression will be revealed through studies of these components.

Although the path forward to detect genetic risk loci for depression remains challenging, what is certain is that a deeper understanding of the etiology of depression is needed. Existing treatments for depression are based on decades-old biology, and genetic discoveries have already begun to identify promising targets for novel therapies in other disorders. Given the enormous burden of depression, identifying its genetic underpinnings may be essential to preventing the onset of this disorder and improving the lives of those who already suffer.

The authors thank Caitlin Clements, Patience Gallagher, Stephanie Kravitz, and Preetha Palasuberniam for their assistance in conducting the literature review for this article. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of Mental Health or the National Institutes of Health.

**REFERENCES**

Genetic Determinants of Depression


15. Widom CS, DuMont K, Czaja SJ. A prospective investigation of major depressive disorder and comorbidity in abused and neglected children grown up. Arch Gen Psychiatry 2007;64:49–56.


E. C. Dunn et al.


86. Wankerl M, Wust S, Otte C. Current developments and controversies: does the serotonin transporter gene-linked polymorphic
region (5-HTTLPR) modulate the association between stress and depression? Curr Opin Psychiatry 2010;23:582–7.