

This is an Open Access document downloaded from ORCA, Cardiff University's institutional repository: <https://orca.cardiff.ac.uk/id/eprint/146992/>

This is the author's version of a work that was submitted to / accepted for publication.

Citation for final published version:

Kendall, K. M. , Van Assche, E., Andlauer, T. F. M., Choi, K. W., Luykx, J. J., Schulte, E. C. and Lu, Y. 2021. The genetic basis of major depression. *Psychological Medicine* 51 (13) , pp. 2217-2230. 10.1017/S0033291721000441

Publishers page: <http://dx.doi.org/10.1017/S0033291721000441>

Please note:

Changes made as a result of publishing processes such as copy-editing, formatting and page numbers may not be reflected in this version. For the definitive version of this publication, please refer to the published source. You are advised to consult the publisher's version if you wish to cite this paper.

This version is being made available in accordance with publisher policies. See <http://orca.cf.ac.uk/policies.html> for usage policies. Copyright and moral rights for publications made available in ORCA are retained by the copyright holders.



# Psychological Medicine

## The Genetic Basis of Major Depression

--Manuscript Draft--

<b>Manuscript Number:</b>	PSM-D-20-01822R1
<b>Full Title:</b>	The Genetic Basis of Major Depression
<b>Article Type:</b>	Invited Review Article
<b>Corresponding Author:</b>	Yi Lu, PhD Karolinska Institutet Stockholm, SWEDEN
<b>Corresponding Author Secondary Information:</b>	
<b>Corresponding Author's Institution:</b>	Karolinska Institutet
<b>Corresponding Author's Secondary Institution:</b>	
<b>First Author:</b>	Kimberley M Kendall, PhD
<b>First Author Secondary Information:</b>	
<b>Order of Authors:</b>	Kimberley M Kendall, PhD
	Evelien Van Assche, PhD
	Till Andlauer, PhD
	Karmel W Choi, PhD
	Jurjen J Luykx, PhD
	Eva C Schulte, PhD
	Yi Lu, PhD
<b>Order of Authors Secondary Information:</b>	
<b>Manuscript Region of Origin:</b>	SWEDEN
<b>Abstract:</b>	Major depressive disorder (MDD) is a common, debilitating, phenotypically heterogeneous disorder with heritability ranges from 30% to 50%. Compared to other psychiatric disorders, its high prevalence, moderate heritability, and strong polygenicity have posed major challenges for gene-mapping in MDD. Studies of common genetic variation in MDD, driven by large international collaborations such as the Psychiatric Genomics Consortium, have confirmed the highly polygenic nature of the disorder and implicated over 100 genetic risk loci to date. Rare copy number variants associated with MDD risk were also recently identified. The goal of this review is to present a broad picture of our current understanding of the epidemiology, genetic epidemiology, molecular genetics, and gene-environment interplay in MDD. Insights into the impact of genetic factors on the aetiology of this complex disorder hold great promise for improving clinical care.

## The Genetic Basis of Major Depression

KM Kendall<sup>1</sup>, E Van Assche<sup>2</sup>, TFM Andlauer<sup>3</sup>, KW Choi<sup>4, 5, 6</sup>, JJ Luykx<sup>7,8,9</sup>, EC Schulte<sup>10,11</sup>, Y Lu<sup>12</sup>

1. MRC Centre for Neuropsychiatric Genetics and Genomics, Cardiff University, UK
2. Department of Psychiatry, University of Muenster, Muenster, Germany
3. Department of Neurology, Klinikum rechts der Isar, School of Medicine, Technical University of Munich, Germany
4. Department of Psychiatry, Massachusetts General Hospital, Boston, MA 02114, USA
5. Psychiatric and Neurodevelopmental Genetics Unit, Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA 02114, USA
6. Department of Epidemiology, Harvard TH Chan School of Public Health, Boston, MA 02115, USA
7. Department of Psychiatry, UMC Utrecht Brain Center, University Medical Center Utrecht, Utrecht University, Utrecht, The Netherlands
8. Department of Translational Neuroscience, UMC Utrecht Brain Center, University Medical Center Utrecht, Utrecht University, Utrecht, The Netherlands
9. Outpatient Second Opinion Clinic, GGNet Mental Health, Warnsveld, The Netherlands
10. Institute of Psychiatric Phenomics and Genomics (IPPG), University Hospital, LMU Munich, Munich, Germany
11. Department of Psychiatry and Psychotherapy, University Hospital, LMU Munich, Munich, Germany
12. Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Sweden.

Corresponding author:

Dr Yi Lu ([lu.yi@ki.se](mailto:lu.yi@ki.se))

Department of Medical Epidemiology and Biostatistics

Karolinska Institutet

Sweden

Nobels väg 12A Solna, Sweden 171 77

ORCID:

K Kendall 0000-0002-6755-6121

E Van Assche 0000-0003-3472-3031

T Andlauer 0000-0002-2917-5889

K Choi 0000-0002-3914-2431

JJ Luykx 0000-0002-6439-2774

EC Schulte 0000-0003-3105-5672

Y Lu 0000-0001-9933-3654

**Current word count: 5579**

**Abstract:**

Major depressive disorder (MDD) is a common, debilitating, phenotypically heterogeneous disorder with heritability ranges from 30% to 50%. Compared to other psychiatric disorders, its high prevalence, moderate heritability, and strong polygenicity have posed major challenges for gene-mapping in MDD. Studies of common genetic variation in MDD, driven by large international collaborations such as the Psychiatric Genomics Consortium, have confirmed the highly polygenic nature of the disorder and implicated over 100 genetic risk loci to date. Rare copy number variants associated with MDD risk were also recently identified. The goal of this review is to present a broad picture of our current understanding of the epidemiology, genetic epidemiology, molecular genetics, and gene-environment interplay in MDD. Insights into the impact of genetic factors on the aetiology of this complex disorder hold great promise for improving clinical care.

## Introduction

Major depressive disorder (MDD) is a common mood disorder that substantially and negatively impacts both quality and length of life (Kessler and Bromet, 2013), and represents the leading cause of disability worldwide (GBD 2017 DALYs and HALE Collaborators, 2018). This review summarises our existing knowledge about the genetic basis of depression. Specifically, it details how research in this area has informed our understanding of the biology underlying depression and how this knowledge may be of use to improve patient care in the future.

## Definition

The term 'depression' most often refers to the clinical diagnosis of MDD, as defined by the International Classification of Diseases (ICD) and the Diagnostic and Statistical Manual of Mental Disorders (DSM) diagnostic criteria. It is characterised by the core symptoms of i) depressed mood (e.g., feelings of sadness, irritability, or emptiness), and ii) anhedonia (e.g., decreased ability to feel pleasure or to experience interest in usual activities). Individuals with depression also experience a wide range of other symptoms which can be grouped into cognitive (e.g., concentration difficulties and indecisiveness), emotional (e.g. feelings of worthlessness and hopelessness), and vegetative symptoms (e.g., significant appetite or weight changes, insomnia or hypersomnia) (Malhi and Mann, 2018).

MDD is a heterogeneous disorder, which has important implications for clinical practice and research (Flint and Kendler, 2014; Cai *et al.*, 2020a). A clinical diagnosis of MDD may be reached in many different ways using established criteria (Fried *et al.*, 2020). This phenotypic heterogeneity in clinical practice is mirrored in the broad range of case definition strategies for genetic research. Genetic studies have adopted multiple methods of classifying depression -

from self-reported clinical diagnosis or electronic health record (EHR) diagnosis codes to structured diagnostic interviews (Smith *et al.*, 2013; Howard *et al.*, 2019). Despite the shortcomings of dichotomizing a heterogeneous disorder like MDD, a categorical definition of depression has been widely used in research. Apart from such a diagnosis-based definition, the severity of depressive symptoms is often studied as a quantitative trait, so the full variation of depressive symptoms can be taken into account. These phenotypic definitions can have a considerable impact on genetic findings (Cai *et al.*, 2020b) and their potential interpretation and translation to the clinic. In line with previous studies (Wray *et al.*, 2018), we will use ‘MDD’ where the study subjects met the clinical criteria for major depressive disorder, and ‘major depression’ where the subjects were ascertained using alternative strategies, but nevertheless likely to have met the diagnostic criteria for MDD.

## Epidemiology

The 12-month prevalence of MDD is estimated to be ~6% and the lifetime prevalence ~20%, although these figures vary considerably across countries (e.g., the 12-month prevalence ranged from 2.2% in Japan to 10.4% in Brazil) (Bromet *et al.*, 2011). Established epidemiological risk factors for MDD include age, female sex, separated/divorced marital status, physical health problems, and childhood trauma (Moskvina *et al.*, 2007; Bromet *et al.*, 2011; Van Assche *et al.*, 2017). The peak age of onset of MDD extends from mid to late adolescence to the early 40s, with the median in the mid-20s (Kessler and Bromet, 2013). Women are twice as likely as men to be diagnosed with MDD (Bromet *et al.*, 2011). Sex-specific differences in clinical presentation (Rice *et al.*, 2015), environmental risk factors, and differences in the brain’s exposure to sex hormones have been suggested as possible explanations. MDD is associated with chronic physical health problems including cardiovascular diseases, chronic pain disorders, and cancers

(Kessler and Bromet, 2013). The nature of the association between MDD and physical ill health is complex. There is evidence for a bidirectional causal association, as well as common antecedents such as social deprivation increasing the risk of both MDD and physical ill health (Kessler and Bromet, 2013).

### Genetic epidemiology

Early evidence from family-based studies clearly demonstrated a familial aggregation of MDD, with an increased risk in first-degree relatives of MDD patients (odds ratio, OR=2.84, 95% CI =2.31-3.49) (Sullivan *et al.*, 2000). However, families share both genetic and environmental factors and quantifying their relative contributions to MDD liability has been central to genetic epidemiological studies.

Classic quantitative genetic modelling partitions phenotypic variance into genetic and environmental components. The pedigree heritability is then estimated as the proportion of phenotypic variance attributable to additive genetic effects (“narrow-sense heritability”) (Lynch and Walsh, 1998; Visscher *et al.*, 2008). Such modelling is typically performed in twin or adoption studies. Twin studies, comparing monozygotic (MZ) and dizygotic (DZ) twins, have consistently demonstrated that MDD is moderately heritable ([Table 1](#)) and an early meta-analysis estimated the heritability at 37% (95% CI = 31-42%) (Sullivan *et al.*, 2000). Subsequent twin studies showed similar estimates, ranging between 29-49%, with higher estimates among treated or recurrent MDD cases – potentially due to greater clinical severity in the study samples (Kendler *et al.*, 2006; Polderman *et al.*, 2015; Kendler *et al.*, 2018a). These studies also suggested that the liability variance attributable to the shared family environment is low, ranging from 0 to 11% (Sullivan *et al.*, 2000; Kendler *et al.*, 2006; Polderman *et al.*, 2015;



Kendler *et al.*, 2018a). Adoption studies, comparing the transmission patterns among adoptive and intact families, are better able to separate shared genetics from other aspects of a shared family environment than twin studies (Kendler *et al.*, 2018b). Early adoption studies in MDD, however, were based on small samples and produced conflicting results (Sullivan *et al.*, 2000). Contrary to findings from twin studies, a recent population-based adoption study reported a much lower heritability (approximately 16%) and a higher contribution of the family environment (Kendler *et al.*, 2018b). In addition, the overestimation of the MDD heritability in twin studies has been raised as a potential reason for the “missing heritability” – i.e. the discrepancy between the pedigree heritability and the much lower phenotypic variance explained by genomic loci identified in genome-wide association studies (GWAS) (Manolio *et al.*, 2009). Recent studies utilised national registers and EHR to reconstruct extended familial relationships, and therefore provide estimates of pedigree heritability beyond classic twin design (Wray and Gottesman, 2012; Kendler *et al.*, 2018a; Polubriaginof *et al.*, 2018). These estimates, however, were very similar to those from twin studies ([Table 1](#)), suggesting that the heritability of MDD has not been overestimated.

One of the most intensively studied questions regarding the heritability of MDD is whether sex differences in the genetic architecture of the disorder exist. A higher heritability in women than in men might contribute to the higher rates of MDD observed in women. Early meta-analyses failed to detect either quantitative or qualitative evidence for sex differences in MDD heritability (Sullivan *et al.*, 2000). However, more recent studies demonstrated a significantly higher MDD heritability in women than in men ([Table 1](#)), although the genetic correlation between the sexes was high (0.63-0.89) (Kendler *et al.*, 2006, 2018a). These results suggest that

most, but not all, genetic risk is shared between the sexes (Kendler *et al.*, 2018a), consistent with the highly polygenic nature of MDD.

Research efforts have attempted to identify genetically more homogeneous MDD subtypes (Flint and Kendler, 2014). Parents with early-onset and recurrent MDD confer the highest risk for MDD in offspring (Birmaher *et al.*, 1996; Sullivan *et al.*, 2000). Other clinical features, including greater clinical severity, comorbid anxiety disorder, and the use of antidepressant and electroconvulsive therapy, also confer higher MDD risk for relatives (Kendler *et al.*, 2018a). Similarly, the Psychiatric Genomics Consortium (PGC) MDD GWAS demonstrated that individuals from many of these subtypes carry higher polygenic risk scores (PRS) for MDD (Power *et al.*, 2017; Wray *et al.*, 2018). Studies examining the heritability of these subtypes are, however, sparse. The heritability among recurrent MDD cases (41%) may be significantly higher than in those who had single-episode (28%) (Fernandez-Pujals *et al.*, 2015), and postpartum depression has been shown to have a higher heritability (44-54%) than non-postpartum depression (32%) (Viktorin *et al.*, 2016). Ascertaining these more heritable forms of MDD has led to some of early GWAS findings (CONVERGE consortium, 2015).

### Candidate gene studies

Candidate gene studies of MDD, first performed more than 40 years ago (Beckman *et al.*, 1978), focused on genes thought to be involved in the pathophysiology of the disorder and in drug mechanisms – as reviewed in (Shadrina *et al.*, 2018). More than 1500 publications have assessed variants in over 200 candidate genes with often conflicting results (Flint and Kendler, 2014). Most of these studies were vastly underpowered, lacked correction for population stratification, and the reported significance levels did not exceed what would be expected by

chance (Flint and Kendler, 2014). Based on the observed OR and minor allele frequencies (MAF), all published candidate gene variants should be detectable in the current GWAS at genome-wide significance (Flint and Kendler, 2014). Yet, this has not been the case (Bosker *et al.*, 2011; Wray *et al.*, 2012, 2018; Howard *et al.*, 2019). In large-scale data from the UK Biobank (UKBB) and PGC (n=62,138-443,264 per subsample), systematic analyses of the most commonly studied polymorphisms mapping to 18 candidate genes found no evidence for significant associations with major depression, except at the gene-level for *DRD2* (dopamine receptor D2) (Border *et al.*, 2019a).

Candidate gene studies were the mainstay of genetic research in MDD until the emergence of GWAS and some are still among the most highly cited works in the field (McIntosh *et al.*, 2019). However, a lack of robust association has led to the strong recommendation that candidate gene studies should no longer be performed (Duncan and Keller, 2011; Farrell *et al.*, 2015; Border *et al.*, 2019a, 2019b).

### The role of rare genetic variants

Historical linkage studies in families with MDD failed to identify causal variants but implicated regions on chromosomes 2, 3, 11, and 15 in disorder risk (Zubenko *et al.*, 2003; Holmans *et al.*, 2004, 2007; Camp *et al.*, 2005, p. 2005; Levinson *et al.*, 2007; Middeldorp *et al.*, 2008; Schol-Gelok *et al.*, 2010; Breen *et al.*, 2011). A highly polygenic architecture, as is the case for MDD, can generate significant linkage study results which could be erroneously interpreted as large-effect loci (Hemani *et al.*, 2013). Indeed, no significant overlap was observed between the reported linkage regions and association loci from current well-powered GWAS (McIntosh *et*

*al.*, 2019). A major contribution of single highly penetrant variants within these regions to the genetic architecture of MDD is therefore unlikely.

In the past decade, association testing of rare and low-frequency variants has progressed to whole-exome (WES) and whole-genome sequencing (WGS) studies. Rare (MAF <0.5%) or low-frequency (MAF range >0.5% and <5%) variants are expected to contribute to the genetic basis of MDD (Lee *et al.*, 2014; Yu *et al.*, 2018). However, the largest WGS study of MDD to date, by the CONVERGE consortium (5,303 cases and 5,337 controls), failed to find evidence for the association of individual rare or low-frequency variants with MDD (CONVERGE consortium, 2015). Using the same data set, it was subsequently shown that individuals with MDD had a significantly higher overall burden of singleton exonic variants predicted to be deleterious (Peterson *et al.*, 2017). This suggests that larger sample sizes are required to identify associated rare variants with moderate effect sizes (Lee *et al.*, 2014; Border *et al.*, 2019b). An approach to mitigate against the need for vast sample sizes is to examine MDD-related subphenotypes (summarised in [Table 2](#)). The majority of these studies identified single rare variants or gene-level burden scores as associated with at least one of the assessed subphenotypes. In most cases, these results await replication in independent and well-powered datasets.

Rare copy number variants (CNVs), deletions or duplications of >1kb of DNA, play a well-established role in risk of neurodevelopmental and psychiatric conditions such as autism spectrum disorder, schizophrenia, and intellectual disability (International Schizophrenia Consortium, 2008; Stefansson *et al.*, 2008; Cooper *et al.*, 2011; Coe *et al.*, 2014), but have received less attention in MDD. Until recently, studies of CNVs in depression have been underpowered and yielded inconsistent and sometimes conflicting results (Glessner *et al.*,

2010; Degenhardt *et al.*, 2012; Perlis *et al.*, 2012; Rucker *et al.*, 2013, 2016; O’Dushlaine *et al.*, 2014; Tansey *et al.*, 2014; Yu *et al.*, 2017).

Two recent large-scale efforts have assessed the contribution of rare CNVs to the genetic architecture of MDD. In a meta-analysis of four cohorts (5,780 cases, 6,626 controls), short deletions <100 kb were significantly enriched in individuals with MDD (Zhang *et al.*, 2019). The enriched deletions mapped primarily to intergenic regions and, within those, to enhancers, suggesting that rare CNVs may increase risk of MDD by altering gene expression. A study of 407,074 individuals from the UKBB (23,979 cases) implicated 53 rare CNVs associated with neurodevelopmental conditions in risk of major depression (OR=1.3–1.5) (Kendall *et al.*, 2019). Three CNV loci—duplications at 1q21.1, the Prader-Willi Syndrome locus at 15q11-13, and at 16p11.2—were individually associated with self-reported depression. Outside of this group, there was no evidence for a residual burden of rare CNVs >100 kb in major depression. However, smaller CNVs were not tested in the study by Kendall and colleagues. Rare CNVs therefore likely contribute to MDD risk, but the exact nature of the association remains unclear.

### Genome-wide association studies

Many MDD GWAS have been conducted during the past decade, predominantly on subjects of European ancestry ([Table 3](#)). Therefore, European-ancestry studies are described first, followed by a review of GWAS in more diverse populations.

Initial MDD GWAS of common variants (i.e., MAF  $\geq 1\%$ ) were limited by small sample sizes and failed to identify genome-wide significant loci ( $p$ -value threshold  $< 5 \times 10^{-8}$ ) ([Table 3](#)). However, the first suggestive GWAS result implicated the gene encoding the presynaptic active zone

component *Piccolo (PCLO)* (Sullivan *et al.*, 2009) which has been confirmed as a risk gene in subsequent studies (Hek *et al.*, 2010; Wray *et al.*, 2018; Howard *et al.*, 2019). In 2011, the first genome-wide significant locus was reported - *SLC6A15* (a neuron-specific neutral amino acid transporter), although this was in a recessive model (Kohli *et al.*, 2011). This association did not reach genome-wide significance with additive models in larger follow-up GWAS (Wray *et al.*, 2018; Howard *et al.*, 2019). Recognising the importance of combining data across studies to increase statistical power, the PGC published its first MDD GWAS meta-analysis in 2013, although no significant locus was identified (Ripke *et al.*, 2013).

Since then, large-scale GWAS of major depression have been conducted with data from the direct-to-consumer company 23andMe and large-scale biobank efforts such as UKBB ([Table 3](#)). In 2016, a GWAS by 23andMe using self-reported depression diagnosis identified 15 genome-wide significant loci (Hyde *et al.*, 2016), confirming the estimate that >75,000 cases would be required for genome-wide significant results (Levinson *et al.*, 2014). A second large GWAS meta-analysis from the PGC described 44 associated genetic loci, pooling cohorts with different phenotype definitions (Wray *et al.*, 2018). The three datasets used in the latter studies were eventually combined in an analysis yielding 102 genome-wide significant variants, 87 of which replicated (Howard *et al.*, 2019). The latest GWAS to date meta-analysed these datasets with data from the Million Veteran Program (MVP), yielding 178 associated loci (Levey *et al.*, 2020).

Collectively, these GWAS confirmed that major depression is highly polygenic, with ORs of individual variants under 1.05 (Wray *et al.*, 2018; Howard *et al.*, 2019), as suggested previously (Nishino *et al.*, 2018). The two loci with the overall most robust support for an association map to the genes *NEGR1* and *OLFM4*, both with an OR of 1.04 and a MAF of ~40% (Wray *et al.*, 2018;

Howard *et al.*, 2019). The neuronal growth regulator 1 (*NEGR1*) cell adhesion molecule modulates neurite outgrowth, synapse formation, and synaptic function (Wray *et al.*, 2018; Noh *et al.*, 2019). The secreted glycoprotein Olfactomedin 4 (*OLFM4*) has an established function in inflammation, innate immunity, and cancer (Liu and Rodgers, 2016; Wray *et al.*, 2018). Several other associated loci have functions in neurite outgrowth, synaptic function and plasticity, as well as in immunity and inflammation (Wray *et al.*, 2018). To establish the functions of the proposed risk SNPs (single nucleotide polymorphisms) and their links to specific genes, analyses of cell-type-specific gene expression and epigenetic data are necessary (Zhong *et al.*, 2019; Arloth *et al.*, 2020; Chan *et al.*, 2020; Li *et al.*, 2020; Wang *et al.*, 2020).

Given the small effect sizes of single SNPs, gene-level and pathway analyses that integrate the effects of many variants across genes and pathways may be more suitable for studying the aetiology of MDD. Gene-level association analyses have implicated genes encoding pre- and post-synaptic proteins, especially receptor units, e.g. voltage-gated calcium channels (VGCC), the dopamine receptor *DRD2*, and glutamate receptors (Wray *et al.*, 2018). Identified by both SNP- and gene-level analyses, the splicing regulator *RBFOX1* may play an important role in the aetiology of major depression (Wray *et al.*, 2018; Noh *et al.*, 2019). Gene-set analyses implicated pathways with roles in neuronal differentiation and projection and, most prominently, synaptic formation and function, including gene sets linked to VGCC and synaptic RNA-binding proteins like *FMRP* and *CELF4* (Wray *et al.*, 2018; Noh *et al.*, 2019; Levey *et al.*, 2020).

Despite a large number of identified genomic loci, the variance explained by major depression PRS is still very low (Nagelkerke's pseudo- $R^2$  1.5-3.2% on the liability scale) (Howard *et al.*,

2019). This prediction accuracy was lower than achieved for, e.g. schizophrenia, where Nagelkerke's pseudo- $R^2$  is 7.7% on the liability scale (Schizophrenia Working Group of the PGC *et al.*, 2020). At much smaller sample sizes, the GWAS on schizophrenia also identified many more associated variants (*Figure 1*). The identification of MDD risk variants has been more challenging than for schizophrenia because of its higher prevalence, lower heritability, and more heterogeneous samples (Wray *et al.*, 2012; Levinson *et al.*, 2014; Ormel *et al.*, 2019). Schizophrenia GWAS saw a critical inflection point at approximately 15,000 cases and are currently discovering a new locus for, on average, every 230 new cases (Levinson *et al.*, 2014). Such an inflection point was observed for depression GWAS only at approximately 75,000 cases, as predicted (Wray *et al.*, 2012; Levinson *et al.*, 2014). GWAS using liberal phenotype definitions of major depression now identify a new locus for every 1500 cases (*Figure 1*).

Importantly, phenotype definition matters. Given the strong genetic correlations across cohorts, recent GWAS meta-analyses combined cohorts with different ascertainment strategies to maximize the overall sample size. However, a comparison of genetic analyses using different phenotype definitions in UKBB revealed that GWAS of broad phenotype definitions might identify associated loci that are not specific to MDD, but, more likely, shared with other psychiatric disorders (Cai *et al.*, 2020b). In order to identify robustly associated MDD variants, a balanced strategy of maximising sample size while focusing on more stringent phenotype definitions has been suggested (Davies *et al.*, 2019; Schwabe *et al.*, 2019; Byrne *et al.*, 2020). Quantitative definitions have also been analysed in GWAS, and a continuum between depressive symptoms and MDD was suggested (Direk *et al.*, 2017). Several variants were reported to be associated with depressive symptoms (*Table 3*), with limited replication success. Of these symptom-associated variants, rs62100776 (Okbay *et al.*, 2016) maps to the gene *DCC*,



coding for a neuronal cell adhesion molecule, which was a genome-wide significant locus in the 2018 PGC case-control GWAS (Wray *et al.*, 2018). Further noteworthy GWAS were conducted on MDD stratified by the age at onset (Power *et al.*, 2017), sex (Hall *et al.*, 2018), or exposure to adversity (Peterson *et al.*, 2018), or combined GWAS of different depression phenotypes (Amare *et al.*, 2020).

The studies discussed so far focused on patients of European ancestry. However, complex disorders show population-specific landscapes of genetic risk (Martin *et al.*, 2017). Accordingly, GWAS in non-European populations have also been conducted ([Table 3](#)). The CONVERGE study on female Han Chinese recurrent in-patient MDD cases identified two significant loci near the *SIRT1* and in the *LHPP* gene, and replicated them in the same population (CONVERGE consortium, 2015). Neither variant was significantly associated in GWAS on European samples (Wray *et al.*, 2018; Howard *et al.*, 2019), possibly due to large differences in allele frequency across ancestries – both variants exhibit markedly lower allele frequencies in Europeans. Follow-up studies estimated the genetic correlation between East Asian and European MDD cohorts at 0.33-0.41, while the average genetic correlations between European cohorts was 0.76 (Bigdeli *et al.*, 2017; Wray *et al.*, 2018), and identified trans-ancestral risk variants (Li *et al.*, 2018). In a study of Hispanic/Latino individuals, two out of 22 European depression GWAS loci were replicated (Dunn *et al.*, 2018). The MVP study identified no genome-wide significant loci in over 20,000 African-ancestry patients with major depression ([Table 3](#)), and only one of 206 analysed European risk SNPs formally replicated in the African sample (Levey *et al.*, 2020). However, a trans-ancestry meta-analysis conducted within the same study increased the number of risk loci from 178 in Europeans alone to 183. These findings highlight the need for

greater geographical, ethnic, and economic diversity in depression genetic studies (McIntosh *et al.*, 2019).

### Genetic overlap between MDD and other traits

Several lines of evidence, including cross-disorder analyses, support the notion that MDD shares genetic underpinnings with many other psychiatric phenotypes. A genetic overlap has also been demonstrated between MDD and non-psychiatric phenotypes. Common techniques for studying cross-disorder genetic overlap include genetic correlation ( $r_g$ ) estimation and PRS analyses (Grotzinger *et al.*, 2019).

Genetic correlation quantifies the genetic relationship between two traits, reflecting their underlying modes of pleiotropy – i.e. the same genetic variant influencing both traits (van Rheenen *et al.*, 2019; Baselmans *et al.*, 2020). Mathematically, it is derived as the genetic covariance between the two traits scaled by the square roots of their genetic variances (van Rheenen *et al.*, 2019). Two methods are commonly used to estimate  $r_g$  – i) bivariate restricted maximum likelihood (REML) which uses individual-level genotype data as the input (Lee *et al.*, 2012), and ii) linkage disequilibrium score regression (LDSC) (Bulik-Sullivan *et al.*, 2015) and its recent extension, high-definition likelihood (HDL) method (Ning *et al.*, 2020), which uses GWAS summary statistics.  $r_g$  ranges between  $-1$  and  $1$ , with  $1$  meaning that shared liability is caused by the same risk SNPs and  $-1$  meaning that identical SNPs increase the risk for one trait while decreasing risk for the other. It is important to note that  $r_g$  estimates cannot infer causality.

Many studies have examined  $r_g$  between MDD and both psychiatric and non-psychiatric disorders or traits. The strongest  $r_g$  detected for MDD are with depressive symptoms measured

as a continuous trait (0.63-1.0) (Direk *et al.*, 2017; Turley *et al.*, 2018), confirming the concept of a continuum between subthreshold depressive symptoms and clinical diagnoses of MDD. Similarly, strong  $r_g$  have been found for personality traits such as neuroticism ( $r_g = 0.70-0.75$ ) and subjective well-being (-0.75) (Okbay *et al.*, 2016; Howard *et al.*, 2019). Major depression was moderately correlated with several other psychiatric disorders, e.g. autism spectrum disorder (0.45), attention-deficit/hyperactivity disorder (0.44), bipolar disorder (BD) (0.36), and schizophrenia (0.34) (Cross-Disorder Group of the PGC, 2019). Further studies reported a weaker  $r_g$  of major depression with type-I (0.26-0.31) than with type-II BD (0.61-0.69) (Stahl *et al.*, 2019; Coleman *et al.*, 2020). Moreover, a positive  $r_g$  was reported between the body-mass index (BMI) and atypical depression with increased appetite/weight, which was not present for those without changes in appetite (Milaneschi *et al.*, 2017). An investigation of genetic correlations between neurological and psychiatric disorders found little evidence of shared genetic underpinnings, except for a significant but small positive  $r_g$  between MDD and migraine (The Brainstorm Consortium *et al.*, 2018). Negative genetic correlations have been detected for major depression with educational attainment and cognitive performance (each  $\sim -0.15$ ) (Howard *et al.*, 2019).

Analyses of PRS is another method used to assess the polygenic overlap between MDD and other traits. PRS use SNP effect sizes from a training GWAS to compute the cumulative genetic risk burden of each individual in a target sample. PRS are calculated as the weighted sum of the allele count multiplied by its effect size across all selected SNPs (Andlauer and Nöthen, 2020; Lewis and Vassos, 2020). Therefore, they capture the combined effects of many SNPs and are used to examine the polygenic liability for a disorder or trait or to examine the genetic overlap across traits (Wray *et al.*, 2014). Although the generation of PRS requires large training GWAS, it

can be applied to much smaller target sets. For examples, a MDD PRS was predictive of population-based obsessive-compulsive symptoms (Zilhão *et al.*, 2018). Clinically, PRS may be used for diagnosis, prognosis, and treatment prediction (Schijven *et al.*, 2020). However, PRS for schizophrenia have, so far, shown greater potential for such applications, given their higher variance explained (Schizophrenia Working Group of the PGC *et al.*, 2020). Importantly, PRS were used to demonstrate that MDD subtypes may diverge in their genetic overlaps with other disorders. For example, early-onset MDD cases show a stronger association with PRS for schizophrenia and BD than late-onset MDD patients do (Power *et al.*, 2017). In the future, PRS might also be used to identify aetiologically more homogeneous MDD subtypes. For example, subgrouping on the basis of treatments such as electroconvulsive therapy might be informative (Foo *et al.*, 2019) and provide further insights into the underlying genetic differences of the severity of the disorder. Finally, MDD PRS may also be used to study PRS-environment (PRS×E) interactions, to support treatment choice, and to refine the penetrance of high-risk variants (Lewis and Vassos, 2020).

### Gene environment interaction

In addition to the heritable component of MDD, environmental factors such as stressful life events (e.g. childhood trauma) play a major role in the aetiology of MDD, both independently and in combination with genetic factors. One active area of research has been the study of gene-environment interactions (G×E). Early G×E efforts examined specific candidate genes with putative biological significance for MDD, e.g. the 5-HTTLPR polymorphism of the serotonin transporter gene, and found evidence of interaction with stressful life events (Caspi *et al.*, 2003). However, these initial reports showed inconsistent replicability (Duncan and Keller, 2011), and recent meta-analytic evidence indicated no robust evidence of candidate G×E

effects on MDD (Van der Auwera *et al.*, 2018; Border *et al.*, 2019a). Studies using genome-wide data have reported varied G×E results. First, stratified GWAS results from the CONVERGE study have suggested that MDD may have a stronger genetic basis when it occurs in the absence of environmental adversity, with genome-wide significant hits for MDD identified only among women not exposed to adversity (Peterson *et al.*, 2018). Data from the UKBB, however, have suggested the opposite, with a higher SNP-based heritability for major depression observed among individuals reporting lifetime trauma exposure (Coleman *et al.*, 2020). Second, PRS have been used to aggregate genome-wide effects for MDD and examine their interactions with stressful life events. Again, this revealed discrepant findings where MDD PRS were more strongly associated with MDD among individuals both exposed (Peyrot *et al.*, 2014) and unexposed (Mullins *et al.*, 2016) to childhood trauma. A later but still modestly powered meta-analysis (n=5,765) identified limited evidence for the interaction of PRS and childhood trauma on MDD (Peyrot *et al.*, 2018), which was echoed in epidemiological and clinical cohorts (n=184-1,450) (Halldorsdottir *et al.*, 2019). However, large-scale studies like the UKBB have provided the opportunity to re-examine PRS×E effects with greater power. One study showed stronger PRS effects on major depression among individuals exposed to childhood trauma and socioeconomic adversity (Shen *et al.*, 2020), though not other types of exposures such as adulthood stressful life events. These findings support the notion that genetic influences could vary by specific environments. Third, genome-wide G×E analyses have identified few variants with genome-wide significant G×E effects (Dunn *et al.*, 2016; Arnau-Soler *et al.*, 2019), although these studies have been largely underpowered and results have not been further replicated. Together, current evidence suggests that genome-wide influences on MDD may vary by stressful life exposures but that the detection of G×E may be sensitive to the research design (e.g. how depression and adversity are defined).

The role of stressful life events, specifically childhood trauma, was the predominant focus of G×E in MDD, but other environmental exposures have also received attention. For example, studies examining social support in combination with MDD PRS in vulnerable populations have consistently shown that social support appears to reduce risk of depression. This effect remained even after accounting for genetic vulnerability for depression, but did not vary by or modify the genetic effects (Choi *et al.*, 2020a; Stringa *et al.*, 2020). Similar observations of independent effects have been made for physical activity and major depression, suggesting that even among individuals at high genetic vulnerability for depression, physical activity levels may still be protective (Choi *et al.*, 2020c). Another genetically informed approach to estimating environmental effects on MDD is Mendelian randomisation (MR). MR methods have emerged as an alternative strategy to examine causal inference, obviating typical sources of confounding (Smith and Ebrahim, 2003). Here, strongly associated genetic variants are selected as statistical instruments — or proxies — to estimate the effect of an exposure of interest (e.g., substance use) on an outcome (e.g., MDD) under specific assumptions. MR studies have supported the role of educational attainment in reducing the depression risk and of BMI in increasing the risk (Wray *et al.*, 2018). Studies have also suggested a potential role of lifestyle exposures in shaping the risk of major depression, such as social connections and sedentary media use (Choi *et al.*, 2020b). Establishing stronger causal evidence for different environmental factors with updated methods may help to clarify inconsistencies across G×E findings (Ni *et al.*, 2019). An additional source of potential bias and inconsistency in G×E research are underlying gene-environment correlations, as the presence of gene-environment correlations may lead to false G×E (Purcell and Sham, 2002; Ni *et al.*, 2019).

## Discussion

In summary, MDD is a common, debilitating, phenotypically heterogeneous disorder. Its heritability ranges between 30-50%. MDD with early-onset, recurrence, comorbid anxiety, greater severity, and postpartum depression may be more heritable subtypes. Compared to other psychiatric disorders, its high prevalence, moderate heritability, and strong polygenicity have posed major challenges for gene-mapping. Recently, rare CNVs have been implicated in risk of MDD. Studies of common variation in MDD, driven forward by large international collaborations such as the PGC, confirmed the highly polygenic nature of the disorder and implicated over 100 loci in disorder risk. These studies identified genes involved in neuronal growth, synaptic function, and inflammation in the pathophysiology of the disorder. Finally, gene-environment interaction and Mendelian randomisation studies have implicated childhood trauma, socioeconomic adversity, social support, physical activity, and their interactions with genetic factors, in risk of MDD.

Evidence on the genetic basis of MDD has multiple potential applications including the development of pharmacological treatments, informing discussions regarding disorder risk, and treatment choice. A greater understanding of the biological processes underlying MDD may facilitate the development of new, potentially more effective medications. MDD GWAS have implicated biological processes underlying the disorder including neurotransmission and synaptic structure as well as genetically informed pharmacological modes of action (Howard *et al.*, 2019). MDD is a very heterogeneous disorder and genetic research will likely benefit from the study of more homogeneous disorder subtypes. For example, research on patients with atypical features of increased appetite or weight suggested that this patient subgroup might

benefit from treatments specifically targeting immunometabolic pathways (Milaneschi *et al.*, 2017).

An individual's genetic make-up likely impacts the disease course – from prevention to diagnosis, treatment, and prognosis (Lewis and Vassos, 2020). Whilst the application of MDD PRS to predict individual risk is currently limited by its low discriminative accuracy, in the future PRS might be used for risk stratification at the population level (Andlauer and Nöthen, 2020). This is supported by recent evidence that MDD PRS – well-established to predict prevalent depression – also predict incident cases in the general population (Musliner *et al.*, 2019). For other complex disorders like coronary artery disease and type-2 diabetes, PRS-based risk stratification is already powerful (Khera *et al.*, 2018). Further development of PRS to allow the identification of individuals at heightened genetic risk, coupled with targeted prevention and intervention measures, holds potential to mitigate depression risk (McIntosh *et al.*, 2019). Knowledge of the genetic architecture of MDD and its links with other psychiatric and non-psychiatric disorders might inform genetic counselling and clinical care. For example, identified genetic correlations with physical health phenotypes might facilitate the development of clinical guidelines for physical health monitoring in individuals with MDD.

MDD is a multifactorial disorder. Other levels of genetic variation, such as DNA methylation (Barbu *et al.*, 2020; Van Assche *et al.*, 2017b) and subsequent gene-expression differences (Arloth *et al.*, 2015; Schiweck *et al.*, 2020), and also the telomere length (Lin *et al.*, 2016), influence depression risk and severity. Importantly, epigenetic mechanisms serve as an interface between the individual and its environment (Provencal and Binder, 2015). A better understanding of the role of epigenetic variation and an improved integration of multi-omics



data may enhance our understanding of the aetiology and treatment of MDD. In addition, this review focuses on genetic factors influencing the risk or severity of depression. We did not discuss how genetic variants affect response to treatment, especially the metabolism of antidepressant drugs via cytochrome P450 genes. This important topic was recently reviewed by Bousman and colleagues (Bousman *et al.* 2021).

Here, we highlighted the state of the art, the promises, and the pitfalls of genetic research in MDD. The field would benefit from further exploration of the stratification properties of genetics – both for the understanding of its aetiology, including the role of environmental factors, and for the development of targeted treatment opportunities. Rich datasets with high-quality and deep phenotyping are crucial for studying heterogeneous disorders like MDD. Combining deep, harmonized phenotypes with molecular data will greatly aid the genetic stratification of MDD. Improving our insights into the impact of genetic factors on MDD holds great promise for improving clinical care, and this intense and fast-developing field has a bright future.

### **Acknowledgments**

We thank Andrew McIntosh, Cathryn Lewis, and Naomi Wray for critical comments on our manuscript.

## Tables and Figures

Table 1. MDD heritability estimates from selected publications

<b>Key paper</b>	<b>Type</b>	<b><math>h^2</math> (95% CI)</b>	<b><math>c^2</math> (95% CI)</b>	<b><math>e^2</math> (95% CI)</b>
<b>Twin data</b>				
(Sullivan <i>et al.</i> , 2000)	Meta-analysis	0.37 (0.31-0.42)	0 (0-0.05)	0.63 (0.58-0.67)
(Kendler <i>et al.</i> , 2006)	Largest twin study (15.5K pairs)	male: 0.29 (0.19-0.38) female: 0.42 (0.36-0.47)	NA <sup>a</sup>	0.71 (0.62-0.81) 0.58 (0.53-0.64)
(Polderman <i>et al.</i> , 2015)	Meta-analysis of 50 years of twin studies	depressive episode: 0.34 (0.31-0.38) recurrent MDD: 0.45 (0.36-0.55)	0.11 (0.08-0.13) 0.03 (0-0.08)	0.55 0.52
(Kendler <i>et al.</i> , 2018a)	Twins identified from SWE register; Treated cases	male: 0.41 (0.19-0.49) female: 0.49 (0.31-0.56)	0 (0-0.13) 0.02 (0-0.17)	0.59 (0.51-0.66) 0.49 (0.44-0.55)
<b>EHR or register-based data</b>				
(Wray and Gottesman, 2012)	DEN register, parent-offspring	0.32 (0.30-0.34)	NA <sup>b</sup>	NA <sup>b</sup>
(Kendler <i>et al.</i> , 2018a)	SWE register, siblings (full vs half, reared together vs apart)	male: 0.36 (0.31-0.38) female: 0.51 (0.51-0.53)	0.05 (0.05-0.05) 0.02 (0.02-0.02)	0.59 (0.58-0.61) 0.47 (0.46-0.49)
(Polubriaginof <i>et al.</i> , 2018)	US EHR, <i>inferred</i> extended pedigree	depressive episode: 0.25 (0.17-0.30) recurrent MDD: 0.36 (0.21-0.51)	NA <sup>a</sup> NA <sup>a</sup>	0.75 (0.70-0.83) 0.64 (0.49-0.79)

$h^2$  narrow-sense heritability;  $c^2$  shared environment;  $e^2$  unique environment. SWE: Sweden; DEN: Denmark. EHR: Electronic Health Records.

- a. Missing estimate of  $c^2$  as a result of AE models (i.e., quantitative genetic models with only additive genetics and unique environment components) in the original study.
- b. Only the heritability was derived from the summary level data.

Table 2. Whole-exome sequencing and exome chip studies assessing the role of rare variants in subphenotypes of depression.

Phenotype	Method	Sample size	Ancestry	Implicated gene or variant	Implicated pathway	Reference
Population isolate	WES, exome chip	1999 (discovery) 1604 (replication)	European	<i>NKPD1</i>	de novo synthesis of sphingolipids	(Amin <i>et al.</i> , 2017a)
Depressive symptoms in families & population isolate	Linkage, haplotype analysis, WES, exome chip	1336 WES 1527 exome chip	European	<i>RCL1</i> p.Leu186Phe (rs115482041)	interlaminar astrocytes	(Amin <i>et al.</i> , 2018)
Familial & early-onset MDD	targeted resequencing of 1742 synaptic genes	259 cases 334 controls	European	Ca <sub>v</sub> 2-adaptor gene set	calcium signaling, actin polymerisation, spine formation	(Pirooznia <i>et al.</i> , 2016)
Familial & early-onset MDD	WES	12	European	<i>ZNF34</i>	brain TFs	(Subaran <i>et al.</i> , 2016)
Sex differences in MDD	WES and WGS	1000 cases (70.7% female) 72 controls	Asian	<i>PDE4A</i> (rs201432982), <i>FDX1L</i> (rs62640397, rs79442975), <i>MYO15B</i> (rs820182, rs820148) more frequent in female cases	none	(Kang <i>et al.</i> , 2020)
Suicide in MDD	WES specifically with 5`/3`-UTRs and promoters	23 cases 21 controls (post-mortem brain samples)	European	<i>COL6A6</i> variants in 17% of suicide completers; 3 calcium channel genes highlighted ( <i>CACNA1B</i> , <i>-1C</i> , <i>-2D4</i> ) 86 non-synonymous variants in 42 genes found in at least two suicide victims but not controls	TGF-β pathway	(Tombácz <i>et al.</i> , 2017)
MRI grey matter volume changes in MDD	WES	77 cases 245 controls	Asian	<i>CSMD1</i> , <i>CNTNAP5</i>	neuroactive ligand receptor interacting pathway	(Zhang <i>et al.</i> , 2020)

Depressive symptoms, general population	WES	3612	European	<i>LIPG</i> p.Asn396Ser (rs77960347)	steroid and cholesterol biosynthesis	(Amin <i>et al.</i> , 2017b)
Depressive symptoms in context of environmental stress	exome chip, WGS	203 cases, 196 controls (discovery) 473 cases, 497 controls (replication)	Hispanic (discovery), European (replication)	<i>PHF21B</i>	9 pathways including innate immune response, glutamate signalling, sensory perception	(Wong <i>et al.</i> , 2017)
Treatment resistance	GWAS, WES, exome chip	1209 cases (discovery) 1128 cases (replication)	European	response predictor in resistance vs. response to treatment with serotonergic antidepressants using rare variants	cell survival and proliferation, neurodegeneration, immune response	(Fabbri <i>et al.</i> , 2020)

MDD=major depressive disorder; WES=whole-exome sequencing; WGS=whole-genome sequencing;  
 GWAS=genome-wide association study; TFs=transcription factors; UTR=untranslated region; TGF- $\beta$ =Transforming growth factor beta; MRI=magnetic resonance imaging

Table 3. Published genome-wide association studies on depression.

Year	Discovery: Nca/Nco or Ntot	Discovery: Loci	Replication: Cases/Controls	Replicated loci	Definition of depression	Ancestry	Reference
2008	1738/1802	0	6079/5893	0* <sup>1</sup>	Structured interview: MDD	European	(Sullivan <i>et al.</i> , 2009) <sup>‡</sup>
2008	1359/1782	0	NA	NA	Structured interview: MDD	European	(Muglia <i>et al.</i> , 2010) <sup>‡</sup>
2009	3957/3428	0	NA	NA	Structured interview: MDD	European	(Shyn <i>et al.</i> , 2011) <sup>‡</sup>
2010	1636/1594	0	1418/1918	0	Structured interview: MDD	European	(Lewis <i>et al.</i> , 2010) <sup>‡</sup>
2010	604/1364	0	409/541	0	Structured interview: MDD	European	(Rietschel <i>et al.</i> , 2010) <sup>‡</sup>
2010	1020/1636	0	NA	NA	Structured interview: MDD	European	(Shi <i>et al.</i> , 2011) <sup>‡</sup>
2010	2431/3673	0	3332/3228	0	Structured interview: MDD	European	(Wray <i>et al.</i> , 2012) <sup>‡</sup>
2011	353/366	0	3735/10635	1* <sup>2</sup>	Structured interview: MDD	European	(Kohli <i>et al.</i> , 2011) <sup>‡</sup>
2012	9240/9519	0	6783/50695	0	Structured interview: MDD	European	(Ripke <i>et al.</i> , 2013) <sup>‡</sup>
2013	34549	0	16709	1* <sup>3</sup>	Depressive symptoms	European	(Hek <i>et al.</i> , 2013)
2015	5303/5337	2	3231/3186	2	Structured interview: Women with recurrent MDD	Han Chinese	(CONVERGE consortium, 2015)
2015	16948	0	NA	NA	Depressive symptoms	Trans-ancestry	(Ware <i>et al.</i> , 2015)

2015	1443	1* <sup>4</sup>	NA	NA	Depressive symptoms	Hispanic American	(Ware <i>et al.</i> , 2015)
2016	32528	1	6813	0	Depressive symptoms	European	(Demirkan <i>et al.</i> , 2016)
2016	7179	0	NA	NA	Depressive symptoms	African American	(Dunn <i>et al.</i> , 2016)
2016	3138	0	NA	NA	Depressive symptoms	Hispanic/ Latino	(Dunn <i>et al.</i> , 2016)
2016	161460	2	368890	2	Depressive symptoms	European	(Okbay <i>et al.</i> , 2016)
2016	3869/9519	1	6107/124230	1	Structured interview: Adult-onset MDD	European	(Power <i>et al.</i> , 2017)
2016	75607/231747 and 84847/241266* <sup>4</sup>	7 and 4* <sup>5</sup>	45773/106354	15* <sup>6</sup>	Self-reported diagnosis	European	(Hyde <i>et al.</i> , 2016) <sup>‡</sup>
2016	70017	1	38328	1	Mixed (including depressive symptoms)	European	(Direk <i>et al.</i> , 2017)
2017	14729/14435	1	4504/7007	0	Structured interview: MDD	European & Han Chinese	(Bigdeli <i>et al.</i> , 2017)
2017	5347/1871	1	NA	NA	Structured interview: MDD with decreased vs increased appetite/ weight	European	(Milaneschi <i>et al.</i> , 2017)
2018	12310	0	7948	0	Depressive symptoms	Hispanic/ Latino	(Dunn <i>et al.</i> , 2018)

2018	90150/246603	10	2659/17237	6* <sup>7</sup>	Mixed	European & Han Chinese	(Li <i>et al.</i> , 2018)
2018	10851/32211	0	NA	NA	Mixed	European	(Hall <i>et al.</i> , 2018) <sup>‡</sup>
2018	3852/16034	1	NA	NA	Mixed: Men only	European	(Hall <i>et al.</i> , 2018)
2018	3139/3832	3	NA	NA	Structured interview: Women without exposure to adversity	Han Chinese	(Peterson <i>et al.</i> , 2018)
2018	113769/208811	14	75607/231747	14* <sup>3</sup>	Broad depression	European	(Howard <i>et al.</i> , 2018) <sup>‡</sup>
2018	30603/143916	2	75607/231747	2* <sup>3</sup>	Probable MDD	European	(Howard <i>et al.</i> , 2018) <sup>‡</sup>
2018	8276/209308	1	75607/231747	1* <sup>3</sup>	ICD-coded MDD	European	(Howard <i>et al.</i> , 2018)
2018	135458/344901	44	NA	NA	Mixed	European	(Wray <i>et al.</i> , 2018) <sup>‡</sup>
2019	Combination of Direk <i>et al.</i> and Hyde <i>et al.</i>	3	Howard <i>et al.</i> 2018: broad depression	3	Broad and self-reported major depression	European	(Amare <i>et al.</i> , 2020)
2019	Combination of Direk <i>et al.</i> and CONVERGE	1	Howard <i>et al.</i> 2018: broad depression	0	Mixed: Broad depression and recurrent MDD	European & Han Chinese	(Amare <i>et al.</i> , 2020)
2019	246363/561190	101* <sup>8</sup>	414574/892299	87	Mixed	European	(Howard <i>et al.</i> , 2019) <sup>‡</sup>
2019	164933/408383* 8	63	NA	NA	Mixed	European	(Coleman <i>et al.</i> , 2020) <sup>‡</sup>
2020	113262/219360	24	NA	NA	Help-seeking: seen doctor	European	(Cai <i>et al.</i> , 2020b)

2020	36286/297126	5	NA	NA	Help-seeking: seen psychiatrist	European	(Cai <i>et al.</i> , 2020b) <sup>‡</sup>
2020	16301/50870	1	NA	NA	Structured interview: MDD	European	(Cai <i>et al.</i> , 2020b)
2020	83810/166405	10	NA	NA	ICD-based	European	(Levey <i>et al.</i> , 2020) <sup>‡</sup>
2020	330173/824094	178	NA	NA	Mixed	European	(Levey <i>et al.</i> , 2020) <sup>‡</sup>
2020	25843/33757	0	NA	NA	ICD-based	African American	(Levey <i>et al.</i> , 2020)
2020	366434/847433	183	NA	NA	Mixed	European and African American	(Levey <i>et al.</i> , 2020)

*Year* indicates the year in which the study was first published (including online ahead of print and, where no published version was available, preprint), which may deviate from the year the article was published in print. *Nca/Nco* for number of cases and controls in case-control design; *Ntot* as total sample size for quantitative trait.

<sup>‡</sup> Studies used for Figure 1.

\*<sup>1</sup> The top locus replicated in subsequent studies (Hek *et al.* 2010, Wray *et al.* 2018, Howard *et al.* 2019).

\*<sup>2</sup> Significant in the meta-analysis of discovery and replication cohorts in a recessive model. This association did not replicate in subsequent GWAS in additive models. Replication has, to our knowledge, not been attempted using a recessive model.

\*<sup>3</sup> Significant in the meta-analysis of discovery and replication cohorts.

\*<sup>4</sup> This association did not replicate in the follow-up Hispanic ancestry GWAS meta-analysis Dunn *et al.* 2018.

\*<sup>4</sup> Meta-analysis of 23andMe and PGC MDD cohorts.

\*<sup>5</sup> Replication of top results from the meta-analysis of 23andMe and PGC MDD cohorts.

\*<sup>6</sup> Only SNPs not reaching genome-wide significance in the discovery-stage were analysed in the replication phase. Six loci (of which two were not significant in the discovery GWAS) covered by these selected SNPs reached significance in the meta-analysis of discovery and replication cohorts.

\*<sup>7</sup> From a meta-analysis of overlapping subjects (PGC and UKBB), case and control numbers may thus be inflated.



\*<sup>8</sup> The study identified 102 independent variants at 101 loci.

**Figure 1: Identified GWAS loci in relation to the number of cases.**

The relationship of identified loci to the number of cases based on published studies on schizophrenia, bipolar disorder (BD), and major depression (MD). To simplify the presentation, studies with different depression definitions have been combined into the MD category. The plot shows the number of cases and independent significant loci in the discovery-stage analyses (without replication). Schizophrenia GWAS saw a critical inflection point at approximately 15,000 cases and are currently discovering a new locus for, on average, every 230 new cases (Levinson *et al.*, 2014). BD currently discovers a new locus for every 400 new cases. Such an inflection point was observed for MD GWAS only at approximately 75,000 cases, as predicted (Wray *et al.*, 2012; Levinson *et al.*, 2014), and MD GWAS using liberal phenotype definitions now identify a new locus every 1500 cases. Slopes starting at the inflection point were estimated using linear regression. The MD studies used for the plot are provided in Table 3.

## References

- Amare, A. T., Vaez, A., Hsu, Y. H., Direk, N., Kamali, Z., Howard, D. M., ... & Hartman, C. A. (2019). Bivariate genome-wide association analyses of the broad depression phenotype combined with major depressive disorder, bipolar disorder or schizophrenia reveal eight novel genetic loci for depression. *Molecular psychiatry*, 1-10.
- Amin, N., Belonogova, N. M., Jovanova, O., Brouwer, R. W., van Rooij, J. G., van den Hout, M. C., ... & van Duijn, C. M. (2017a) Nonsynonymous variation in NKPD1 increases depressive symptoms in European populations. *Biological Psychiatry*, 81(8), 702-707.
- Amin, N., Jovanova, O., Adams, H. H. H., Dehghan, A., Kavousi, M., Vernooij, M. W., ... & Van Duijn, C. M. (2017b). Exome-sequencing in a large population-based study reveals a rare Asn396Ser variant in the LIPG gene associated with depressive symptoms. *Molecular psychiatry*, 22(4), 537-543.
- Amin, N., De Vrij, F. M. S., Baghdadi, M., Brouwer, R. W. W., Van Rooij, J. G. J., Jovanova, O., ... & Van Duijn, C. M. (2018). A rare missense variant in RCL1 segregates with depression in extended families. *Molecular psychiatry*, 23(5), 1120-1126.
- Andlauer, T. F., & Nöthen, M. M. (2020). Polygenic scores for psychiatric disease: from research tool to clinical application. *Medizinische Genetik*, 32(1), 39-45.
- Arloth, J., Bogdan, R., Weber, P., Frishman, G., Menke, A., Wagner, K. V., ... & Hariri, A. R. (2015). Genetic differences in the immediate transcriptome response to stress predict risk-related brain function and psychiatric disorders. *Neuron*, 86(5), 1189-1202..
- Arloth, J., Eraslan, G., Andlauer, T. F., Martins, J., Iurato, S., Kühnel, B., ... & Mueller, N. S. (2020). DeepWAS: Multivariate genotype-phenotype associations by directly integrating regulatory information using deep learning. *PLoS computational biology*, 16(2), e1007616.
- Arnau-Soler, A., Macdonald-Dunlop, E., Adams, M. J., Clarke, T. K., MacIntyre, D. J., Milburn, K., ... & Thomson, P. A. (2019). Genome-wide by environment interaction studies of depressive symptoms and psychosocial stress in UK Biobank and Generation Scotland. *Translational psychiatry*, 9(1), 1-13.
- Barbu, M. C., Shen, X., Walker, R. M., Howard, D. M., Evans, K. L., Whalley, H. C., ... & McIntosh, A. M. (2020). Epigenetic prediction of major depressive disorder. *Molecular Psychiatry*, 1-12.
- Baselmans, B. M., Yengo, L., van Rheenen, W., & Wray, N. R. (2020). Risk in relatives, heritability, SNP-based heritability and genetic correlations in psychiatric disorders: a review. *Biological Psychiatry*.
- Beckman, G., Beckman, L., Cedergren, B., Perris, C., & Strandman, E. (1978). Serum protein and red cell enzyme polymorphisms in affective disorders. *Human heredity*, 28(1), 41-47.
- Bigdeli, T. B., Ripke, S., Peterson, R. E., Trzaskowski, M., Bacanu, S. A., Abdellaoui, A., ... & Kendler, K. S. (2017). Genetic effects influencing risk for major depressive disorder in China and Europe. *Translational psychiatry*, 7(3), e1074-e1074.
- Birmaher, B., Ryan, N. D., Williamson, D. E., Brent, D. A., Kaufman, J., Dahl, R. E., ... & Nelson, B. (1996). Childhood and adolescent depression: a review of the past 10 years. Part I. *Journal of the American Academy of Child & Adolescent Psychiatry*, 35(11), 1427-1439.
- Border, R., Johnson, E. C., Evans, L. M., Smolen, A., Berley, N., Sullivan, P. F., & Keller, M. C. (2019a). No support for historical candidate gene or candidate gene-by-interaction hypotheses for major depression across multiple large samples. *American Journal of Psychiatry*, 176(5), 376-387.
- Border, R., Smolen, A., Corley, R. P., Stallings, M. C., Brown, S. A., Conger, R. D., ... & Evans, L. M. (2019b). Imputation of behavioral candidate gene repeat variants in 486,551 publicly-available UK Biobank individuals. *European Journal of Human Genetics*, 27(6), 963.

- Bosker, F. J., Hartman, C. A., Nolte, I. M., Prins, B. P., Terpstra, P., Posthuma, D., ... & Nolen, W. A. (2011). Poor replication of candidate genes for major depressive disorder using genome-wide association data. *Molecular psychiatry*, *16*(5), 516-532.
- Bousman, C. A., Bengesser, S. A., Aitchison, K. J., Amare, A. T., Aschauer, H., Baune, B. T., ... & Müller, D. J. (2020). Review and Consensus on Pharmacogenomic Testing in Psychiatry. *Pharmacopsychiatry*.
- Brainstorm Consortium. (2018). Analysis of shared heritability in common disorders of the brain. *Science (New York, NY)*, *360*(6395).
- Breen, G., Webb, B. T., Butler, A. W., Van Den Oord, E. J., Tozzi, F., Craddock, N., ... & McGuffin, P. (2011). A genome-wide significant linkage for severe depression on chromosome 3: the depression network study. *American Journal of Psychiatry*, *168*(8), 840-847.
- Bromet, E., Andrade, L. H., Hwang, I., Sampson, N. A., Alonso, J., De Girolamo, G., ... & Kessler, R. C. (2011). Cross-national epidemiology of DSM-IV major depressive episode. *BMC medicine*, *9*(1), 90.
- Bulik-Sullivan, B., Finucane, H. K., Anttila, V., Gusev, A., Day, F. R., Loh, P. R., ... & Psychiatric Genomics Consortium. (2015). An atlas of genetic correlations across human diseases and traits. *Nature genetics*, *47*(11), 1236.
- Byrne, E. M., Kirk, K. M., Medland, S. E., McGrath, J. J., Colodro-Conde, L., Parker, R., ... & Martin, N. G. (2020). Cohort profile: the Australian genetics of depression study. *BMJ open*, *10*(5), e032580.
- Cai, N., Choi, K. W., & Fried, E. I. (2020a). Reviewing the genetics of heterogeneity in depression: operationalizations, manifestations and etiologies. *Human molecular genetics*, *29*(R1), R10-R18.
- Cai, N., Revez, J. A., Adams, M. J., Andlauer, T. F., Breen, G., Byrne, E. M., ... & Flint, J. (2020b). Minimal phenotyping yields genome-wide association signals of low specificity for major depression. *Nature Genetics*, *52*(4), 437-447.
- Camp, N. J., Lowry, M. R., Richards, R. L., Plenk, A. M., Carter, C., Hensel, C. H., ... & Cannon-Albright, L. A. (2005). Genome-wide linkage analyses of extended Utah pedigrees identifies loci that influence recurrent, early-onset major depression and anxiety disorders. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*, *135*(1), 85-93.
- Caspi, A., Sugden, K., Moffitt, T. E., Taylor, A., Craig, I. W., Harrington, H., ... & Poulton, R. (2003). Influence of life stress on depression: moderation by a polymorphism in the 5-HTT gene. *Science*, *301*(5631), 386-389.
- Chan, R. F., Turecki, G., Shabalin, A. A., Guintivano, J., Zhao, M., Xie, L. Y., ... & van den Oord, E. J. (2020). Cell Type-Specific Methylome-wide Association Studies Implicate Neurotrophin and Innate Immune Signaling in Major Depressive Disorder. *Biological psychiatry*, *87*(5), 431-442.
- Choi, K. W., Chen, C. Y., Ursano, R. J., Sun, X., Jain, S., Kessler, R. C., ... & Smoller, J. W. (2020a). Prospective study of polygenic risk, protective factors, and incident depression following combat deployment in US Army soldiers. *Psychological Medicine*, *50*(5), 737-745.
- Choi, K. W., Stein, M. B., Nishimi, K. M., Ge, T., Coleman, J. R., Chen, C. Y., ... & Smoller, J. W. (2020b). An exposure-wide and Mendelian randomization approach to identifying modifiable factors for the prevention of depression. *American Journal of Psychiatry*, *177*(10), 944-954.
- Choi, K. W., Zheutlin, A. B., Karlson, R. A., Wang, M. J., Dunn, E. C., Stein, M. B., ... & Smoller, J. W. (2020c). Physical activity offsets genetic risk for incident depression assessed via electronic health records in a biobank cohort study. *Depression and anxiety*, *37*(2), 106-114.
- Coe, B. P., Witherspoon, K., Rosenfeld, J. A., Van Bon, B. W., Vulto-van Silfhout, A. T., Bosco, P., ... & Eichler, E. E. (2014). Refining analyses of copy number variation identifies specific genes associated with developmental delay. *Nature genetics*, *46*(10), 1063-1071.

- Coleman, J. R., Gaspar, H. A., Bryois, J., Byrne, E. M., Forstner, A. J., Holmans, P. A., ... & Lawson, W. B. (2020). The genetics of the mood disorder Spectrum: genome-wide association analyses of more than 185,000 cases and 439,000 controls. *Biological psychiatry*, *88*(2), 169-184.
- CONVERGE consortium (2015) (2015). Sparse whole-genome sequencing identifies two loci for major depressive disorder. *Nature*, *523*(7562), 588-591.
- Cooper, G. M., Coe, B. P., Girirajan, S., Rosenfeld, J. A., Vu, T. H., Baker, C., ... & Eichler, E. E. (2011). A copy number variation morbidity map of developmental delay. *Nature genetics*, *43*(9), 838-846.
- Cross-Disorder Group of the PGC (2019). Genomic relationships, novel loci, and pleiotropic mechanisms across eight psychiatric disorders. *Cell*, *179*(7), 1469-1482.
- Davies, M. R., Kalsi, G., Armour, C., Jones, I. R., McIntosh, A. M., Smith, D. J., ... & Breen, G. (2019). The Genetic Links to Anxiety and Depression (GLAD) Study: online recruitment into the largest recontactable study of depression and anxiety. *Behaviour research and therapy*, *123*, 103503.
- Degenhardt, F., Priebe, L., Herms, S., Mattheisen, M., Mühleisen, T. W., Meier, S., ... & Cichon, S. (2012). Association between copy number variants in 16p11. 2 and major depressive disorder in a German case-control sample. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*, *159*(3), 263-273.
- Demirkan, A., Lahti, J., Direk, N., Viktorin, A., Lunetta, K. L., Terracciano, A., ... & Luik, A. I. (2016). Somatic, positive and negative domains of the Center for Epidemiological Studies Depression (CES-D) scale: a meta-analysis of genome-wide association studies. *Psychological medicine*, *46*(8), 1613.
- Direk, N., Williams, S., Smith, J. A., Ripke, S., Air, T., Amare, A. T., ... & Sullivan, P. F. (2017). An analysis of two genome-wide association meta-analyses identifies a new locus for broad depression phenotype. *Biological psychiatry*, *82*(5), 322-329.
- Duncan, L. E., & Keller, M. C. (2011). A critical review of the first 10 years of candidate gene-by-environment interaction research in psychiatry. *American Journal of Psychiatry*, *168*(10), 1041-1049.
- Dunn, E. C., Sofer, T., Wang, M. J., Soare, T. W., Gallo, L. C., Gogarten, S. M., ... & Smoller, J. W. (2018). Genome-wide association study of depressive symptoms in the Hispanic Community Health Study/Study of Latinos. *Journal of psychiatric research*, *99*, 167-176.
- Dunn, E. C., Wiste, A., Radmanesh, F., Almli, L. M., Gogarten, S. M., Sofer, T., ... & Smoller, J. W. (2016). Genome-wide association study (GWAS) and genome-wide by environment interaction study (GWEIS) of depressive symptoms in African American and Hispanic/Latina women. *Depression and anxiety*, *33*(4), 265-280.
- Fabbri, C., Kasper, S., Kautzky, A., Zohar, J., Souery, D., Montgomery, S., ... & Serretti, A. (2020). A polygenic predictor of treatment-resistant depression using whole exome sequencing and genome-wide genotyping. *Translational psychiatry*, *10*(1), 1-12.
- Farrell, M. S., Werge, T., Sklar, P., Owen, M. J., Ophoff, R. A., O'Donovan, M. C., ... & Sullivan, P. F. (2015). Evaluating historical candidate genes for schizophrenia. *Molecular psychiatry*, *20*(5), 555-562.
- Fernandez-Pujals, A. M., Adams, M. J., Thomson, P., McKechnie, A. G., Blackwood, D. H., Smith, B. H., ... & McIntosh, A. M. (2015). Epidemiology and heritability of major depressive disorder, stratified by age of onset, sex, and illness course in Generation Scotland: Scottish Family Health Study (GS: SFHS). *PloS one*, *10*(11), e0142197.
- Flint, J., & Kendler, K. S. (2014). The genetics of major depression. *Neuron*, *81*(3), 484-503.
- Foo, J. C., Streit, F., Frank, J., Witt, S. H., Treutlein, J., Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium, ... & Kranaster, L. (2019). Evidence for increased genetic risk load for major depression in patients assigned to electroconvulsive therapy. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*, *180*(1), 35-45.

- Fried, E. I., Coomans, F., & Lorenzo-Luaces, L. (2020). The 341 737 ways of qualifying for the melancholic specifier. *The Lancet Psychiatry*, 7(6), 479-480.
- Fried, E. I., & Nesse, R. M. (2015). Depression is not a consistent syndrome: an investigation of unique symptom patterns in the STAR\* D study. *Journal of affective disorders*, 172, 96-102.
- GBD 2017 DALYs and HALE Collaborators (2018). Global, regional, and national disability-adjusted life-years (DALYs) for 359 diseases and injuries and healthy life expectancy (HALE) for 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. *The Lancet*, 392(10159), 1859-1922.
- Glessner, J. T., Wang, K., Sleiman, P. M., Zhang, H., Kim, C. E., Flory, J. H., ... & Hakonarson, H. (2010). Duplication of the SLIT3 locus on 5q35. 1 predisposes to major depressive disorder. *PLoS one*, 5(12), e15463.
- Grotzinger, A. D., Rhemtulla, M., de Vlaming, R., Ritchie, S. J., Mallard, T. T., Hill, W. D., ... & Tucker-Drob, E. M. (2019). Genomic structural equation modelling provides insights into the multivariate genetic architecture of complex traits. *Nature human behaviour*, 3(5), 513-525.
- Hall, L. S., Adams, M. J., Arnau-Soler, A., Clarke, T. K., Howard, D. M., Zeng, Y., ... & McIntosh, A. M. (2018). Genome-wide meta-analyses of stratified depression in Generation Scotland and UK Biobank. *Translational psychiatry*, 8(1), 1-12.
- Halldorsdottir, T., Piechaczek, C., Soares de Matos, A. P., Czamara, D., Pehl, V., Wagenbuechler, P., ... & Binder, E. B. (2019). Polygenic risk: predicting depression outcomes in clinical and epidemiological cohorts of youths. *American Journal of Psychiatry*, 176(8), 615-625.
- Hek, K., Demirkan, A., Lahti, J., Terracciano, A., Teumer, A., Cornelis, M. C., ... & Murabito, J. (2013). A genome-wide association study of depressive symptoms. *Biological psychiatry*, 73(7), 667-678.
- Hek, K., Mulder, C. L., Luijendijk, H. J., van Duijn, C. M., Hofman, A., Uitterlinden, A. G., & Tiemeier, H. (2010). The PCLO gene and depressive disorders: replication in a population-based study. *Human Molecular Genetics*, 19(4), 731-734.
- Hemani, G., Yang, J., Vinkhuyzen, A., Powell, J. E., Willemsen, G., Hottenga, J. J., ... & Visscher, P. M. (2013). Inference of the genetic architecture underlying BMI and height with the use of 20,240 sibling pairs. *The American Journal of Human Genetics*, 93(5), 865-875.
- Holmans, P., Weissman, M. M., Zubenko, G. S., Scheftner, W. A., Crowe, R. R., DePaulo Jr., MD, J. R., ... & Levinson, D. F. (2007). Genetics of recurrent early-onset major depression (GenRED): final genome scan report. *American Journal of Psychiatry*, 164(2), 248-258.
- Holmans, P., Zubenko, G. S., Crowe, R. R., DePaulo Jr, J. R., Scheftner, W. A., Weissman, M. M., ... & Levinson, D. F. (2004). Genomewide significant linkage to recurrent, early-onset major depressive disorder on chromosome 15q. *The American Journal of Human Genetics*, 74(6), 1154-1167.
- Howard, D. M., Adams, M. J., Clarke, T. K., Hafferty, J. D., Gibson, J., Shireli, M., ... & McIntosh, A. M. (2019). Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. *Nature neuroscience*, 22(3), 343-352.
- Howard, D. M., Adams, M. J., Shireli, M., Clarke, T. K., Marioni, R. E., Davies, G., ... & McIntosh, A. M. (2018). Genome-wide association study of depression phenotypes in UK Biobank identifies variants in excitatory synaptic pathways. *Nature communications*, 9(1), 1-10.
- Hyde, C. L., Nagle, M. W., Tian, C., Chen, X., Paciga, S. A., Wendland, J. R., ... & Winslow, A. R. (2016). Identification of 15 genetic loci associated with risk of major depression in individuals of European descent. *Nature genetics*, 48(9), 1031-1036.

- International Schizophrenia Consortium. (2008). Rare chromosomal deletions and duplications increase risk of schizophrenia. *Nature*, *455*(7210), 237.
- Kang, H. J., Park, Y., Yoo, K. H., Kim, K. T., Kim, E. S., Kim, J. W., ... & Kim, J. M. (2020). Sex differences in the genetic architecture of depression. *Scientific reports*, *10*(1), 1-12.
- Kendall, K. M., Rees, E., Bracher-Smith, M., Legge, S., Riglin, L., Zammit, S., ... & Walters, J. T. R. (2019). Association of rare copy number variants with risk of depression. *JAMA psychiatry*, *76*(8), 818-825.
- Kendler, K. S., Gatz, M., Gardner, C. O., & Pedersen, N. L. (2006). A Swedish national twin study of lifetime major depression. *American Journal of Psychiatry*, *163*(1), 109-114.
- Kendler, K. S., Ohlsson, H., Lichtenstein, P., Sundquist, J., & Sundquist, K. (2018a). The genetic epidemiology of treated major depression in Sweden. *American Journal of Psychiatry*, *175*(11), 1137-1144.
- Kendler, K. S., Ohlsson, H., Sundquist, K., & Sundquist, J. (2018b). Sources of parent-offspring resemblance for major depression in a national Swedish extended adoption study. *Jama Psychiatry*, *75*(2), 194-200.
- Kessler, R. C., & Bromet, E. J. (2013). The epidemiology of depression across cultures. *Annual review of public health*, *34*, 119-138.
- Khera, A. V., Chaffin, M., Aragam, K. G., Haas, M. E., Roselli, C., Choi, S. H., ... & Kathiresan, S. (2018). Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nature genetics*, *50*(9), 1219-1224.
- Kohli, M. A., Lucae, S., Saemann, P. G., Schmidt, M. V., Demirkan, A., Hek, K., ... & Binder, E. B. (2011). The neuronal transporter gene SLC6A15 confers risk to major depression. *Neuron*, *70*(2), 252-265.
- Lee, S., Abecasis, G. R., Boehnke, M., & Lin, X. (2014). Rare-variant association analysis: study designs and statistical tests. *The American Journal of Human Genetics*, *95*(1), 5-23.
- Lee, S. H., Yang, J., Goddard, M. E., Visscher, P. M., & Wray, N. R. (2012). Estimation of pleiotropy between complex diseases using single-nucleotide polymorphism-derived genomic relationships and restricted maximum likelihood. *Bioinformatics*, *28*(19), 2540-2542.
- Levey, D. F., Stein, M. B., Wendt, F. R., Pathak, G. A., Zhou, H., Aslan, M., ... & Gelernter, J. (2020). GWAS of Depression Phenotypes in the Million Veteran Program and Meta-analysis in More than 1.2 Million Participants Yields 178 Independent Risk Loci. *medRxiv*.
- Levinson, D. F., Evgrafov, O. V., Knowles, J. A., Potash, J. B., Weissman, M. M., Scheftner, W. A., ... & Holmans, P. (2007). Genetics of recurrent early-onset major depression (GenRED): significant linkage on chromosome 15q25-q26 after fine mapping with single nucleotide polymorphism markers. *American Journal of Psychiatry*, *164*(2), 259-264.
- Levinson, D. F., Mostafavi, S., Milaneschi, Y., Rivera, M., Ripke, S., Wray, N. R., & Sullivan, P. F. (2014). Genetic studies of major depressive disorder: why are there no genome-wide association study findings and what can we do about it?. *Biological psychiatry*, *76*(7), 510-512.
- Lewis, C. M., Ng, M. Y., Butler, A. W., Cohen-Woods, S., Uher, R., Pirlo, K., ... & McGuffin, P. (2010). Genome-wide association study of major recurrent depression in the UK population. *American Journal of Psychiatry*, *167*(8), 949-957.
- Lewis, C. M., & Vassos, E. (2020). Polygenic risk scores: from research tools to clinical instruments. *Genome Medicine*, *12*, 1-11.
- Li, S., Li, Y., Li, X., Liu, J., Huo, Y., Wang, J., ... & Luo, X. J. (2020). Regulatory mechanisms of major depressive disorder risk variants. *Molecular Psychiatry*, 1-20.

- Li, X., Luo, Z., Gu, C., Hall, L. S., McIntosh, A. M., Zeng, Y., ... & Luo, X. J. (2018). Common variants on 6q16. 2, 12q24. 31 and 16p13. 3 are associated with major depressive disorder. *Neuropsychopharmacology*, *43*(10), 2146-2153.
- Lin, P. Y., Huang, Y. C., & Hung, C. F. (2016). Shortened telomere length in patients with depression: a meta-analytic study. *Journal of psychiatric research*, *76*, 84-93.
- Liu, W., & Rodgers, G. P. (2016). Olfactomedin 4 expression and functions in innate immunity, inflammation, and cancer. *Cancer and Metastasis Reviews*, *35*(2), 201-212.
- Lynch, M., & Walsh, B. (1998). *Genetics and analysis of quantitative traits* (Vol. 1, pp. 535-557). Sunderland, MA: Sinauer.
- Malhi GS, Mann JJ (2018) Depression. *The Lancet* *392*, 2299–2312.
- Manolio, T. A., Collins, F. S., Cox, N. J., Goldstein, D. B., Hindorff, L. A., Hunter, D. J., ... & Visscher, P. M. (2009). Finding the missing heritability of complex diseases. *Nature*, *461*(7265), 747-753.
- Martin, A. R., Gignoux, C. R., Walters, R. K., Wojcik, G. L., Neale, B. M., Gravel, S., ... & Kenny, E. E. (2017). Human demographic history impacts genetic risk prediction across diverse populations. *The American Journal of Human Genetics*, *100*(4), 635-649.
- McIntosh, A. M., Sullivan, P. F., & Lewis, C. M. (2019). Uncovering the genetic architecture of major depression. *Neuron*, *102*(1), 91-103.
- Middeldorp, C. M., Hottenga, J. J., Slagboom, P. E., Sullivan, P. F., de Geus, E. J., Posthuma, D., ... & Boomsma, D. I. (2008). Linkage on chromosome 14 in a genome-wide linkage study of a broad anxiety phenotype. *Molecular psychiatry*, *13*(1), 84-89.
- Milaneschi, Y., Lamers, F., Peyrot, W. J., Baune, B. T., Breen, G., Dehghan, A., ... & Penninx, B. W. (2017). Genetic association of major depression with atypical features and obesity-related immunometabolic dysregulations. *Jama Psychiatry*, *74*(12), 1214-1225.
- Moskvina, V., Farmer, A., Swainson, V., O'Leary, J., Gunasinghe, C., Owen, M., ... & Korszun, A. (2007). Interrelationship of childhood trauma, neuroticism, and depressive phenotype. *Depression and anxiety*, *24*(3), 163-168.
- Muglia, P., Tozzi, F., Galwey, N. W., Francks, C., Upmanyu, R., Kong, X. Q., ... & Roses, A. D. (2010). Genome-wide association study of recurrent major depressive disorder in two European case-control cohorts. *Molecular psychiatry*, *15*(6), 589-601.
- Mullins, N., Power, R. A., Fisher, H. L., Hanscombe, K. B., Euesden, J., Iniesta, R., ... & Lewis, C. M. (2016). Polygenic interactions with environmental adversity in the aetiology of major depressive disorder. *Psychological medicine*, *46*(4), 759-770.
- Musliner, K. L., Mortensen, P. B., McGrath, J. J., Suppli, N. P., Hougaard, D. M., Bybjerg-Grauholm, J., ... & Agerbo, E. (2019). Association of polygenic liabilities for major depression, bipolar disorder, and schizophrenia with risk for depression in the Danish population. *JAMA psychiatry*, *76*(5), 516-525.
- Ni, G., van der Werf, J., Zhou, X., Hyppönen, E., Wray, N. R., & Lee, S. H. (2019). Genotype-covariate correlation and interaction disentangled by a whole-genome multivariate reaction norm model. *Nature communications*, *10*(1), 1-15.
- Ning, Z., Pawitan, Y., & Shen, X. (2020). High-definition likelihood inference of genetic correlations across human complex traits. *Nature genetics*, *52*(8), 859-864.

- Nishino, J., Ochi, H., Kochi, Y., Tsunoda, T., & Matsui, S. (2018). Sample size for successful genome-wide association study of major depressive disorder. *Frontiers in genetics, 9*, 227.
- Noh, K., Lee, H., Choi, T. Y., Joo, Y., Kim, S. J., Kim, H., ... & Lee, S. J. (2019). Negr1 controls adult hippocampal neurogenesis and affective behaviors. *Molecular psychiatry, 24*(8), 1189-1205.
- O'Dushlaine, C., Ripke, S., Ruderfer, D. M., Hamilton, S. P., Fava, M., Iosifescu, D. V., ... & Perlis, R. H. (2014). Rare copy number variation in treatment-resistant major depressive disorder. *Biological psychiatry, 76*(7), 536-541.
- Okbay, A., Baselmans, B. M., De Neve, J. E., Turley, P., Nivard, M. G., Fontana, M. A., ... & Rich, S. S. (2016). Genetic variants associated with subjective well-being, depressive symptoms, and neuroticism identified through genome-wide analyses. *Nature genetics, 48*(6), 624-633.
- Ormel, J., Hartman, C. A., & Snieder, H. (2019). The genetics of depression: successful genome-wide association studies introduce new challenges. *Translational psychiatry, 9*(1), 1-10.
- Perlis, R. H., Ruderfer, D., Hamilton, S. P., & Ernst, C. (2012). Copy number variation in subjects with major depressive disorder who attempted suicide.
- Peterson, R. E., Cai, N., Dahl, A. W., Bigdeli, T. B., Edwards, A. C., Webb, B. T., ... & Kendler, K. S. (2018). Molecular genetic analysis subdivided by adversity exposure suggests etiologic heterogeneity in major depression. *American Journal of Psychiatry, 175*(6), 545-554.
- Peyrot, W. J., Milaneschi, Y., Abdellaoui, A., Sullivan, P. F., Hottenga, J. J., Boomsma, D. I., & Penninx, B. W. (2014). Effect of polygenic risk scores on depression in childhood trauma. *The British Journal of Psychiatry, 205*(2), 113-119.
- Peyrot, W. J., Van der Auwera, S., Milaneschi, Y., Dolan, C. V., Madden, P. A., Sullivan, P. F., ... & Yang, J. (2018). Does childhood trauma moderate polygenic risk for depression? A meta-analysis of 5765 subjects from the psychiatric genomics consortium. *Biological psychiatry, 84*(2), 138-147.
- Pirooznia, M., Wang, T., Avramopoulos, D., Potash, J. B., Zandi, P. P., & Goes, F. S. (2016). High-throughput sequencing of the synaptome in major depressive disorder. *Molecular psychiatry, 21*(5), 650-655.
- Polderman, T. J., Benyamin, B., De Leeuw, C. A., Sullivan, P. F., Van Bochoven, A., Visscher, P. M., & Posthuma, D. (2015). Meta-analysis of the heritability of human traits based on fifty years of twin studies. *Nature genetics, 47*(7), 702-709.
- Polubriaginof, F. C., Vanguri, R., Quinnes, K., Belbin, G. M., Yahi, A., Salmasian, H., ... & Tatonetti, N. P. (2018). Disease heritability inferred from familial relationships reported in medical records. *Cell, 173*(7), 1692-1704.
- Power, R. A., Tansey, K. E., Buttenschøn, H. N., Cohen-Woods, S., Bigdeli, T., Hall, L. S., ... & GERAD1 Consortium. (2017). Genome-wide association for major depression through age at onset stratification: major depressive disorder working group of the psychiatric genomics consortium. *Biological Psychiatry, 81*(4), 325-335.
- Provencal, N., & Binder, E. B. (2015). The neurobiological effects of stress as contributors to psychiatric disorders: focus on epigenetics. *Current opinion in neurobiology, 30*, 31-37.
- Purcell, S., & Sham, P. (2002). Variance components models for gene-environment interaction in quantitative trait locus linkage analysis. *Twin Research and Human Genetics, 5*(6), 572-576.
- Van Rheenen, W., Peyrot, W. J., Schork, A. J., Lee, S. H., & Wray, N. R. (2019). Genetic correlations of polygenic disease traits: from theory to practice. *Nature Reviews Genetics, 20*(10), 567-581.
- Rice, S. M., Fallon, B. J., Aucote, H. M., Möller-Leimkühler, A., Treeby, M. S., & Amminger, G. P. (2015). Longitudinal sex differences of externalising and internalising depression symptom trajectories: Implications for assessment of depression in men from an online study. *International Journal of Social Psychiatry, 61*(3), 236-240.



- Rietschel, M., Mattheisen, M., Frank, J., Treutlein, J., Degenhardt, F., Breuer, R., ... & Cichon, S. (2010). Genome-wide association-, replication-, and neuroimaging study implicates HOMER1 in the etiology of major depression. *Biological psychiatry*, *68*(6), 578-585.
- Ripke, S., Wray, N. R., Lewis, C. M., Hamilton, S. P., Weissman, M. M., Breen, G., ... & Major Depressive Disorder Working Group of the Psychiatric GWAS Consortium. (2013). A mega-analysis of genome-wide association studies for major depressive disorder. *Molecular psychiatry*, *18*(4), 497.
- Rucker, J. J., Breen, G., Pinto, D., Pedroso, I., Lewis, C. M., Cohen-Woods, S., ... & McGuffin, P. (2013). Genome-wide association analysis of copy number variation in recurrent depressive disorder. *Molecular psychiatry*, *18*(2), 183-189.
- Rucker, J. J., Tansey, K. E., Rivera, M., Pinto, D., Cohen-Woods, S., Uher, R., ... & Breen, G. (2016). Phenotypic association analyses with copy number variation in recurrent depressive disorder. *Biological psychiatry*, *79*(4), 329-336.
- Schijven, D., Veldink, J. H., & Luykx, J. J. (2020). Genetic cross-disorder analysis in psychiatry: from methodology to clinical utility. *The British Journal of Psychiatry*, *216*(5), 246-249.
- Schiweck, C., Claes, S., Van Oudenhove, L., Lafit, G., Vaessen, T., de Beeck, G. O., ... & Vrieze, E. (2020). Childhood trauma, suicide risk and inflammatory phenotypes of depression: insights from monocyte gene expression. *Translational psychiatry*, *10*(1), 1-12.
- Schizophrenia Working Group of the PGC, Ripke S, Walters JT, O'Donovan MC (2020). Mapping genomic loci prioritises genes and implicates synaptic biology in schizophrenia. *MedRxiv*.
- Schol-Gelok, S., Janssens, A. C. J., Tiemeier, H., Liu, F., Lopez-Leon, S., Zorkoltseva, I. V., ... & van Duijn, C. M. (2010). A genome-wide screen for depression in two independent Dutch populations. *Biological psychiatry*, *68*(2), 187-196.
- Schwabe, I., Milaneschi, Y., Gerring, Z., Sullivan, P. F., Schulte, E., Suppli, N. P., ... & Middeldorp, C. M. (2019). Unraveling the genetic architecture of major depressive disorder: merits and pitfalls of the approaches used in genome-wide association studies. *Psychological medicine*, *49*(16), 2646-2656.
- Shadrina, M., Bondarenko, E. A., & Slominsky, P. A. (2018). Genetics factors in major depression disease. *Frontiers in psychiatry*, *9*, 334.
- Shen, X., Howard, D. M., Adams, M. J., Hill, W. D., Clarke, T. K., Deary, I. J., ... & McIntosh, A. M. (2020). A phenome-wide association and Mendelian Randomisation study of polygenic risk for depression in UK Biobank. *Nature communications*, *11*(1), 1-16.
- Shi, J., Potash, J. B., Knowles, J. A., Weissman, M. M., Coryell, W., Scheftner, W. A., ... & Levinson, D. F. (2011). Genome-wide association study of recurrent early-onset major depressive disorder. *Molecular psychiatry*, *16*(2), 193-201.
- Shyn, S. I., Shi, J., Kraft, J. B., Potash, J. B., Knowles, J. A., Weissman, M. M., ... & Hamilton, S. P. (2011). Novel loci for major depression identified by genome-wide association study of Sequenced Treatment Alternatives to Relieve Depression and meta-analysis of three studies. *Molecular psychiatry*, *16*(2), 202-215.
- Smith, D. J., Nicholl, B. I., Cullen, B., Martin, D., Ul-Haq, Z., Evans, J., ... & Pell, J. P. (2013). Prevalence and characteristics of probable major depression and bipolar disorder within UK biobank: cross-sectional study of 172,751 participants. *PLoS one*, *8*(11), e75362.
- Davey Smith, G., & Ebrahim, S. (2003). 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease?. *International journal of epidemiology*, *32*(1), 1-22.
- Stahl, E. A., Breen, G., Forstner, A. J., McQuillin, A., Ripke, S., Trubetskoy, V., ... & Reif, A. (2019). Genome-wide association study identifies 30 loci associated with bipolar disorder. *Nature genetics*, *51*(5), 793-803.

- Stefansson, H., Rujescu, D., Cichon, S., Pietiläinen, O. P., Ingason, A., Steinberg, S., ... & Stefansson, K. (2008). Large recurrent microdeletions associated with schizophrenia. *nature*, *455*(7210), 232-236.
- Stringa, N., Milaneschi, Y., van Schoor, N. M., Suanet, B., van der Lee, S., Holstege, H., ... & Huisman, M. (2020). Genetic liability for depression, social factors and their interaction effect in depressive symptoms and depression over time in older adults. *The American Journal of Geriatric Psychiatry*.
- Subaran, R. L., Odgerel, Z., Swaminathan, R., Glatt, C. E., & Weissman, M. M. (2016). Novel variants in ZNF34 and other brain-expressed transcription factors are shared among early-onset MDD relatives. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*, *171*(3), 333-341.
- Sullivan, P. F., Neale, M. C., & Kendler, K. S. (2000). Genetic epidemiology of major depression: review and meta-analysis. *American Journal of Psychiatry*, *157*(10), 1552-1562.
- Sullivan, P. F., de Geus, E. J., Willemsen, G., James, M. R., Smit, J. H., Zandbelt, T., ... & Penninx, B. W. (2009). Genome-wide association for major depressive disorder: a possible role for the presynaptic protein piccolo. *Molecular psychiatry*, *14*(4), 359-375.
- Tansey, K. E., Rucker, J. J., Kavanagh, D. H., Guipponi, M., Perroud, N., Bondolfi, G., ... & Uher, R. (2014). Copy number variants and therapeutic response to antidepressant medication in major depressive disorder. *The pharmacogenomics journal*, *14*(4), 395-399.
- Tombácz, D., Maróti, Z., Kalmár, T., Csabai, Z., Balázs, Z., Takahashi, S., ... & Boldogkői, Z. (2017). High-coverage whole-exome sequencing identifies candidate genes for suicide in victims with major depressive disorder. *Scientific reports*, *7*(1), 1-11.
- Turley, P., Walters, R. K., Maghzian, O., Okbay, A., Lee, J. J., Fontana, M. A., ... & Benjamin, D. J. (2018). Multi-trait analysis of genome-wide association summary statistics using MTAG. *Nature genetics*, *50*(2), 229-237.
- Van Assche, E., Moons, T., Cinar, O., Viechtbauer, W., Oldehinkel, A. J., Van Leeuwen, K., ... & van Winkel, R. (2017a). Gene-based interaction analysis shows GABAergic genes interacting with parenting in adolescent depressive symptoms. *Journal of Child Psychology and Psychiatry*, *58*(12), 1301-1309.
- Van Assche, E., Vangeel, E., Freson, K., Van Leeuwen, K., Verschueren, K., Colpin, H., ... & Claes, S. (2017b). Epigenome-wide analysis of methylation and perceived parenting in adolescents and its correlation with depressive symptoms over time. *European Neuropsychopharmacology*, *27*, S449-S450.
- Van der Auwera, S., Peyrot, W. J., Milaneschi, Y., Hertel, J., Baune, B., Breen, G., ... & Grabe, H. (2018). Genome-wide gene-environment interaction in depression: A systematic evaluation of candidate genes: The childhood trauma working-group of PGC-MDD. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*, *177*(1), 40-49.
- Viktorin, A., Meltzer-Brody, S., Kuja-Halkola, R., Sullivan, P. F., Landén, M., Lichtenstein, P., & Magnusson, P. K. (2016). Heritability of perinatal depression and genetic overlap with nonperinatal depression. *American Journal of Psychiatry*, *173*(2), 158-165.
- Visscher, P. M., Hill, W. G., & Wray, N. R. (2008). Heritability in the genomics era—concepts and misconceptions. *Nature reviews genetics*, *9*(4), 255-266.
- Wang, X., Cheng, W., Zhu, J., Yin, H., Chang, S., Yue, W., & Yu, H. (2020). Integrating genome-wide association study and expression quantitative trait loci data identifies NEGR1 as a causal risk gene of major depression disorder. *Journal of Affective Disorders*, *265*, 679-686.
- Ware, E. B., Mukherjee, B., Sun, Y. V., Diez-Roux, A. V., Kardia, S. L., & Smith, J. A. (2015). Comparative genome-wide association studies of a depressive symptom phenotype in a repeated measures setting by race/ethnicity in the multi-ethnic study of atherosclerosis. *BMC genetics*, *16*(1), 118.

- Wong, M. L., Arcos-Burgos, M., Liu, S., Velez, J. I., Yu, C., Baune, B. T., ... & Licinio, J. (2017). The PHF21B gene is associated with major depression and modulates the stress response. *Molecular psychiatry*, 22(7), 1015-1025.
- Wray, N. R., & Gottesman, I. I. (2012). Using summary data from the danish national registers to estimate heritabilities for schizophrenia, bipolar disorder, and major depressive disorder. *Frontiers in genetics*, 3, 118.
- Wray, N. R., Lee, S. H., Mehta, D., Vinkhuyzen, A. A., Dudbridge, F., & Middeldorp, C. M. (2014). Research review: polygenic methods and their application to psychiatric traits. *Journal of child psychology and psychiatry*, 55(10), 1068-1087.
- Wray, N. R., Pergadia, M. L., Blackwood, D. H. R., Penninx, B. W. J. H., Gordon, S. D., Nyholt, D. R., ... & Sullivan, P. F. (2012). Genome-wide association study of major depressive disorder: new results, meta-analysis, and lessons learned. *Molecular psychiatry*, 17(1), 36-48.
- Wray, N. R., Ripke, S., Mattheisen, M., Trzaskowski, M., Byrne, E. M., Abdellaoui, A., ... & Viktorin, A. (2018). Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nature genetics*, 50(5), 668-681.
- Yu, C., Arcos-Burgos, M., Baune, B. T., Arolt, V., Dannlowski, U., Wong, M. L., & Licinio, J. (2018). Low-frequency and rare variants may contribute to elucidate the genetics of major depressive disorder. *Translational psychiatry*, 8(1), 1-8.
- Yu, C., Baune, B. T., Wong, M. L., & Licinio, J. (2017). Investigation of copy number variation in subjects with major depression based on whole-genome sequencing data. *Journal of Affective Disorders*, 220, 38-42.
- Zhang, X., Abdellaoui, A., Rucker, J., de Jong, S., Potash, J. B., Weissman, M. M., ... & Levinson, D. F. (2019). Genome-Wide burden of rare short deletions is enriched in major depressive disorder in four cohorts. *Biological psychiatry*, 85(12), 1065-1073.
- Zhang, Y., Li, M., Wang, Q., Hsu, J. S., Deng, W., Ma, X., ... & Li, T. (2020). A joint study of whole exome sequencing and structural MRI analysis in major depressive disorder. *Psychological Medicine*, 50(3), 384-395.
- Zhong, J., Li, S., Zeng, W., Li, X., Gu, C., Liu, J., & Luo, X. J. (2019). Integration of GWAS and brain eQTL identifies FLOT1 as a risk gene for major depressive disorder. *Neuropsychopharmacology*, 44(9), 1542-1551.
- Zilhão, N. R., Abdellaoui, A., Smit, D. J. A., Cath, D. C., Hottenga, J. J., & Boomsma, D. I. (2018). Polygenic prediction of obsessive compulsive symptoms. *Mol Psychiatry*, 23, 168-9.
- Zubenko, G. S., Maher, B., Hughes III, H. B., Zubenko, W. N., Stiffler, J. S., Kaplan, B. B., & Marazita, M. L. (2003). Genome-wide linkage survey for genetic loci that influence the development of depressive disorders in families with recurrent, early-onset, major depression. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*, 123(1), 1-18.

## The Genetic Basis of Major Depression

KM Kendall<sup>1</sup>, E Van Assche<sup>2</sup>, TFM Andlauer<sup>3</sup>, KW Choi<sup>4, 5, 6</sup>, JJ Luykx<sup>7,8,9</sup>, EC Schulte<sup>10,11</sup>, Y Lu<sup>12</sup>

1. MRC Centre for Neuropsychiatric Genetics and Genomics, Cardiff University, UK
2. Department of Psychiatry, University of Muenster, Muenster, Germany
3. Department of Neurology, Klinikum rechts der Isar, School of Medicine, Technical University of Munich, Germany
4. Department of Psychiatry, Massachusetts General Hospital, Boston, MA 02114, USA
5. Psychiatric and Neurodevelopmental Genetics Unit, Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA 02114, USA
6. Department of Epidemiology, Harvard TH Chan School of Public Health, Boston, MA 02115, USA
7. Department of Psychiatry, UMC Utrecht Brain Center, University Medical Center Utrecht, Utrecht University, Utrecht, The Netherlands
8. Department of Translational Neuroscience, UMC Utrecht Brain Center, University Medical Center Utrecht, Utrecht University, Utrecht, The Netherlands
9. Outpatient Second Opinion Clinic, GGNet Mental Health, Warnsveld, The Netherlands
10. Institute of Psychiatric Phenomics and Genomics (IPPG), University Hospital, LMU Munich, Munich, Germany
11. Department of Psychiatry and Psychotherapy, University Hospital, LMU Munich, Munich, Germany
12. Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Sweden.

Corresponding author:

Dr Yi Lu ([lu.yi@ki.se](mailto:lu.yi@ki.se))

Department of Medical Epidemiology and Biostatistics

Karolinska Institutet

Sweden

Nobels väg 12A Solna, Sweden 171 77

ORCID:

K Kendall 0000-0002-6755-6121

E Van Assche 0000-0003-3472-3031

T Andlauer 0000-0002-2917-5889

K Choi 0000-0002-3914-2431

JJ Luykx 0000-0002-6439-2774

EC Schulte 0000-0003-3105-5672

Y Lu 0000-0001-9933-3654

**Current word count: ~~5579~~5138**

**Abstract:**

Major ~~depression-depressive disorder~~ (~~MDMDD~~) is a common, debilitating, phenotypically heterogeneous disorder with a heritability ~~of~~ ranges from 30% to -50%. Compared to other psychiatric disorders, its high prevalence, ~~relatively low~~moderate heritability, and strong polygenicity have posed major challenges for gene-mapping in ~~MDMDD~~. Studies of common genetic variation in ~~MDMDD~~, driven by large international collaborations such as the Psychiatric Genomics Consortium, have confirmed the highly polygenic nature of the disorder and implicated over 100 genetic risk loci to date. Rare copy number variants associated with ~~MD~~ MDD risk ~~were have~~ also recently ~~been~~ identified. The goal of this review is to present a broad picture of our current understanding of the epidemiology, genetic epidemiology, molecular genetics, and gene-environment interplay in ~~MDMDD~~. Insights into the impact of genetic factors on the aetiology of this complex disorder hold great promise for improving clinical care.

## Introduction

~~Depression~~ Major depressive disorder (MDD) is a common mood disorder that substantially and negatively impacts both quality and length of life (Kessler and Bromet, 2013), and represents the leading cause of disability worldwide (GBD 2017 DALYs and HALE Collaborators, 2018). This review summarises our existing knowledge about the genetic basis of depression. Specifically, it details how research in this area has informed our understanding of the biology underlying depression and how this knowledge may be used-of use to improve patient care in the future.

## Definition

The term 'depression' most often refers to the clinical diagnosis of MDD, as defined by the International Classification of Diseases (ICD) and the Diagnostic and Statistical Manual of Mental Disorders (DSM) diagnostic criteria. ~~Depression-It~~ is characterised by the core symptoms of i) depressed mood (e.g., feelings of sadness, irritability, or emptiness), and ii) anhedonia (e.g., decreased ability to feel pleasure or to experience interest in usual activities). Individuals with depression also experience a wide range of other symptoms which can be grouped into cognitive (e.g., concentration difficulties and indecisiveness), emotional (e.g. feelings of worthlessness and hopelessness), and ~~somatic domains~~ vegetative symptoms (e.g., significant appetite or weight changes, insomnia or hypersomnia) (Malhi and Mann, 2018).

~~Depression-MDD~~ is a heterogeneous disorder, which has important implications for clinical practice and research (Flint and Kendler, 2014; Cai *et al.*, 2020a). ~~Even within one classification system, multiple combinations of symptoms can lead to a diagnosis. A study of depression symptoms in participants of the STAR\*D trial found over a thousand unique symptom profiles with the most common only being reported by 1.8% of the sample (Fried and Nesse, 2015).~~

~~Furthermore, the term depression is used inconsistently, and its~~ clinical diagnosis of MDD may be reached in many different ways using established criteria (Fried *et al.*, 2020). This phenotypic heterogeneity in clinical practice is mirrored in the broad range of case definition strategies for genetic research. Genetic studies have adopted multiple methods of classifying depression - from self-reported clinical diagnosis or electronic health record (EHR) diagnosis codes to structured diagnostic interviews (Smith *et al.*, 2013; Howard *et al.*, 2019). Despite the shortcomings of dichotomizing a heterogeneous disorder like MDD, a categorical definition of depression has been widely used in research. Apart from such a diagnosis-based definition, the severity of depressive symptoms is often studied as a quantitative trait, so the full variation of depressive symptoms can be taken into account. These phenotypic definitions can have a considerable impact on genetic findings (Cai *et al.*, 2020b) and their potential interpretation and translation to the clinic. In line with previous studies (Wray *et al.*, 2018)~~The Psychiatric Genomics Consortium (PGC) has adopted the term Major Depression (MD) to reflect the degree of uncertainty present when analysing data from a combination of studies with varying classification methods. Throughout this review,~~ we will use 'MDD' where the study subjects met the clinical criteria for major depressive disorder, and 'major depression' where the subjects were ascertained using alternative strategies, but nevertheless likely to have met the diagnostic criteria for MDD.

## Epidemiology

The 12-month prevalence of ~~MD~~ MDD is estimated to be ~6% and the lifetime prevalence ~20%, although these figures vary considerably across countries (e.g., the 12-month prevalence ranged from 2.2% in Japan to 10.4% in Brazil) (Bromet *et al.*, 2011). Established epidemiological risk factors for ~~MD~~ MDD include age, female sex, separated/divorced marital status, physical



health problems, and childhood trauma (Moskvina *et al.*, 2007; Bromet *et al.*, 2011; Van Assche *et al.*, 2017). The peak age of onset of ~~MD~~MDD extends from mid to late adolescence to the early 40s, with the median in the ~~mid-20s~~mid-20s (Kessler and Bromet, 2013). Women are twice as likely as men to be diagnosed with ~~MD~~MDD (Bromet *et al.*, 2011). Sex-specific differences in clinical presentation (Rice *et al.*, 2015), environmental risk factors, and differences in the brain's exposure to sex hormones have been suggested as possible explanations. ~~MD~~MDD is associated with chronic physical health problems including cardiovascular diseases, chronic pain disorders, and cancers (Kessler and Bromet, 2013). The nature of the association between ~~MD~~MDD and physical ill health is complex. There is evidence for a bidirectional causal association, as well as common antecedents such as social deprivation increasing the risk of both ~~MD~~MDD and physical ill health (Kessler and Bromet, 2013).

### Genetic epidemiology

Early evidence from family-based studies clearly demonstrated a familial aggregation of ~~MD~~MDD, with an increased risk in first-degree relatives of ~~MD~~MDD patients (odds ratio, OR=2.84, 95% CI =2.31-3.49) (Sullivan *et al.*, 2000). However, families share both genetic and environmental factors and quantifying their relative contributions to ~~MD~~MDD liability has been central to genetic epidemiological studies.

Classic quantitative genetic modelling partitions phenotypic variance into genetic and environmental components. The pedigree heritability is then estimated as the proportion of phenotypic variance attributable to additive genetic effects ("narrow-sense heritability") (Lynch and Walsh, 1998; Visscher *et al.*, 2008). Such modelling is typically performed in twin or adoption studies. Twin studies, comparing monozygotic (MZ) and dizygotic (DZ) twins, have

consistently demonstrated that MD-MDD is moderately heritable ([Table 1](#)) and an early meta-analysis estimated the heritability at 37% (95% CI = 31-42%) (Sullivan *et al.*, 2000). Subsequent twin studies showed similar estimates, ranging between 29-49%, with higher estimates among treated or recurrent MD-MDD cases – potentially due to greater clinical severity in the study samples (Kendler *et al.*, 2006; Polderman *et al.*, 2015; Kendler *et al.*, 2018a). These studies also suggested that the liability variance attributable to the shared family environment is low, ranging from 0 to 11% (Sullivan *et al.*, 2000; Kendler *et al.*, 2006; Polderman *et al.*, 2015; Kendler *et al.*, 2018a). Adoption studies, ~~which~~ comparing the transmission patterns among adoptive and intact families, are better able to separate shared genetics from other aspects of a shared family environment than twin studies (Kendler *et al.*, 2018b). Early adoption studies in MD-MDD, however, were based on small samples and produced conflicting results (Sullivan *et al.*, 2000). Contrary to findings from twin studies, a recent population-based adoption study reported a much lower heritability (approximately 16%) and a higher contribution of the family environment (Kendler *et al.*, 2018b). In addition, the overestimation of the MD-MDD heritability in twin studies has been raised as a potential reason for the “missing heritability” – i.e. the discrepancy between the pedigree heritability and the much lower phenotypic variance explained by genomic loci identified in genome-wide association studies (GWAS) (Manolio *et al.*, 2009). Recent studies utilised national registers and EHR to reconstruct extended familial relationships, and therefore provide estimates of pedigree heritability beyond classic twin design (Wray and Gottesman, 2012; Kendler *et al.*, 2018a; Polubriaginof *et al.*, 2018). These estimates, however, were very similar to those from twin studies ([Table 1](#)), suggesting that the heritability of MD-MDD has not been overestimated.

One of the most intensively studied questions regarding the heritability of MD-MDD is whether sex differences in the genetic architecture of the disorder exist. A higher heritability in women than in men might contribute to the higher rates of MD-MDD observed in women. Early meta-analyses failed to detect either quantitative or qualitative evidence for sex differences in MD-MDD heritability (Sullivan *et al.*, 2000). However, more recent studies demonstrated a significantly higher MD-MDD heritability in women than in men (*Table 1*), although the genetic correlation between the sexes was high (0.63-0.89) (Kendler *et al.*, 2006, 2018a). These results suggest that most, but not all, genetic risk is shared between the sexes (Kendler *et al.*, 2018a), consistent with the highly polygenic nature of MD-MDD.

Research efforts have attempted to identify genetically more homogeneous MD-MDD subtypes (Flint and Kendler, 2014). Parents with early-onset and recurrent MD-MDD confer the highest risk for MD-MDD in offspring (Birmaher *et al.*, 1996; Sullivan *et al.*, 2000). Other clinical features, including greater clinical severity, comorbid anxiety disorder, and the use of antidepressant and electroconvulsive therapy, also confer higher MD-MDD risk for relatives (Kendler *et al.*, 2018a). Similarly, the Psychiatric Genomics Consortium (PGC) MD-MDD GWAS have demonstrated that individuals from many of these subtypes carry higher polygenic risk scores (PRS) for MD-MDD (Power *et al.*, 2017; Wray *et al.*, 2018). Studies examining the heritability of these subtypes are, however, sparse. The heritability among recurrent MD-MDD cases (41%) may be significantly higher than in those who had single-episode MD (28%) (Fernandez-Pujals *et al.*, 2015), and postpartum depression has been shown to have a higher heritability (44-54%) than non-postpartum depression (32%) (Viktorin *et al.*, 2016). Ascertaining these more heritable forms of MD-MDD has led to some of early GWAS findings (CONVERGE consortium, 2015).

## Candidate gene studies

Candidate gene studies of ~~MD~~MDD, first performed more than 40 years ago (Beckman *et al.*, 1978), focused on genes thought to be involved in the pathophysiology of the disorder and in drug mechanisms – as reviewed in (Shadrina *et al.*, 2018). More than 1500 publications have assessed variants in over 200 candidate genes with often conflicting results (Flint and Kendler, 2014). Most of these studies were vastly underpowered, lacked correction for population stratification, and the reported significance levels did not exceed what would be expected by chance (Flint and Kendler, 2014). Based on the observed ~~odds ratios (OR)~~ and minor allele frequencies (MAF), all published candidate gene variants should be detectable in the current GWAS at genome-wide significance (Flint and Kendler, 2014). Yet, this has not been the case (Bosker *et al.*, 2011; Wray *et al.*, 2012, 2018; Howard *et al.*, 2019). In large-scale data from the UK Biobank (UKBB) and PGC (n=62,138-443,264 per subsample), systematic analyses of the most commonly studied polymorphisms mapping to 18 candidate genes found no evidence for significant associations with ~~MD~~major depression, except at the gene-level for *DRD2* (dopamine receptor D2) (Border *et al.*, 2019a).

Candidate gene studies were the mainstay of genetic research in ~~MD~~MDD until the emergence of GWAS and some are still among the most highly cited works in the field (McIntosh *et al.*, 2019). However, a lack of robust association has led to the strong recommendation that candidate gene studies should no longer be performed (Duncan and Keller, 2011; Farrell *et al.*, 2015; Border *et al.*, 2019a, 2019b).

## The role of rare genetic variants

Historical linkage studies in families with ~~MD~~-MDD failed to identify causal variants but implicated regions on chromosomes 2, 3, 11, and 15 in disorder risk (Zubenko *et al.*, 2003; Holmans *et al.*, 2004, 2007; Camp *et al.*, 2005, p. 2005; Levinson *et al.*, 2007; Middeldorp *et al.*, 2008; Schol-Gelok *et al.*, 2010; Breen *et al.*, 2011). A highly polygenic architecture, as is the case for ~~MD~~MDD, can generate significant linkage study results which could be erroneously interpreted as large-effect loci (Hemani *et al.*, 2013). Indeed, no significant overlap was observed between the reported linkage regions and association loci from current well-powered GWAS (McIntosh *et al.*, 2019). A major contribution of single highly penetrant variants within these regions to the genetic architecture of ~~MD~~-MDD is therefore unlikely.

In the past decade, association testing of rare and low-frequency variants has progressed to whole-exome (WES) and whole-genome sequencing (WGS) studies. Rare (MAF <0.5%) or low-frequency (MAF range >0.5% and <5%) variants are expected to contribute to the genetic basis of ~~MD~~-MDD (Lee *et al.*, 2014; Yu *et al.*, 2018). However, the largest WGS study of ~~MD~~-MDD to date, by the CONVERGE consortium (5,303 cases and 5,337 controls), failed to find evidence for the association of individual rare or low-frequency variants with ~~MD~~-MDD (CONVERGE consortium, 2015). Using the same data set, it was subsequently shown that individuals with MDD had a significantly higher overall burden of singleton exonic variants predicted to be deleterious (Peterson *et al.*, 2017). This suggests that larger sample sizes are required to identify associated rare variants with moderate effect sizes (Lee *et al.*, 2014; Border *et al.*, 2019b). An approach to mitigate against the need for vast sample sizes is to examine ~~MD~~MDD-related subphenotypes (summarised in [Table 2](#)). The majority of these studies identified single rare variants or gene-level burden scores as associated with at least one of the assessed

subphenotypes. In most cases, these results await replication in independent and well-powered datasets.

Rare copy number variants (CNVs), deletions or duplications of >1kb of DNA, play a well-established role in risk of neurodevelopmental and psychiatric conditions such as autism spectrum disorder, schizophrenia, and intellectual disability ([International Schizophrenia Consortium, 2008](#); Stefansson *et al.*, 2008; Cooper *et al.*, 2011; Coe *et al.*, 2014), but have received less attention in ~~MD~~MDD. Until recently, studies of CNVs in depression have been underpowered and yielded inconsistent and sometimes conflicting results (Glessner *et al.*, 2010; Degenhardt *et al.*, 2012; Perlis *et al.*, 2012; Rucker *et al.*, 2013, 2016; O'Dushlaine *et al.*, 2014; Tansey *et al.*, 2014; Yu *et al.*, 2017).

Two recent large-scale efforts have assessed the contribution of rare CNVs to the genetic architecture of ~~MD~~MDD. In a meta-analysis of four cohorts (5,780 cases, 6,626 controls), short deletions <100 kb were significantly enriched in individuals with ~~MD~~MDD (Zhang *et al.*, 2019). The enriched deletions mapped primarily to intergenic regions and, within those, to enhancers, suggesting that rare CNVs may increase risk of ~~MD~~MDD by altering gene expression. A study of 407,074 individuals from the UKBB (23,979 cases) implicated 53 rare CNVs associated with neurodevelopmental conditions in risk of ~~MD~~major depression (~~OR~~OR=1.3–1.5) (Kendall *et al.*, 2019). Three CNV loci—duplications at 1q21.1, the Prader-Willi Syndrome locus at 15q11-13, and at 16p11.2—were individually associated with self-reported depression. Outside of this group, there was no evidence for a residual burden of rare CNVs >100 kb in ~~MD~~major depression. However, smaller CNVs were not tested in the study by Kendall and colleagues.

Rare CNVs therefore likely contribute to ~~MD~~-MDD risk, but the exact nature of the association remains unclear.

### Genome-wide association studies

Many MDD GWAS have been conducted during the past decade, predominantly on subjects of European ancestry (Table 3). Therefore, European-ancestry studies are described first, followed by a review of GWAS in more diverse populations.

Initial ~~MD~~-MDD GWAS of common variants (i.e., MAF  $\geq 1\%$ ) were limited by small sample sizes and failed to identify genome-wide significant loci ( $p$ -value threshold  $< 5 \times 10^{-8}$ ) (Table 3).

However, the first suggestive GWAS result implicated the gene encoding the presynaptic active zone component Piccolo (*PCLO*) (Sullivan *et al.*, 2009) ~~and this~~which has been confirmed as a risk gene in subsequent studies (Hek *et al.*, 2010; Wray *et al.*, 2018; Howard *et al.*, 2019). In 2011, the first genome-wide significant locus was reported - *SLC6A15* (a neuron-specific neutral amino acid transporter), although this was in a recessive model (Kohli *et al.*, 2011). This association did not reach genome-wide significance with additive models in larger follow-up GWAS (Wray *et al.*, 2018; Howard *et al.*, 2019). Recognising the importance of combining data across studies to increase statistical power, the PGC published its first ~~MD~~-MDD GWAS meta-analysis in 2013, although no significant locus was identified (Ripke *et al.*, 2013).

Since then, large-scale ~~MD~~-GWAS of major depression have been conducted with data from the direct-to-consumer company 23andMe and large-scale biobank efforts such as UKBB (Table 3).

In 2016, a GWAS by 23andMe using self-reported depression diagnosis identified 15 genome-wide significant loci (Hyde *et al.*, 2016), confirming the estimate that  $> 75,000$  cases would be

required for genome-wide significant results (Levinson *et al.*, 2014). A second large GWAS meta-analysis from the PGC described 44 ~~MD~~-associated genetic loci, pooling cohorts with different phenotype definitions (Wray *et al.*, 2018). The three datasets used in the latter studies were eventually combined in an analysis yielding 102 genome-wide significant variants, 87 of which replicated (Howard *et al.*, 2019). The latest GWAS to date meta-analysed these datasets with data from the Million Veteran Program (MVP), yielding 178 associated loci (Levey *et al.*, 2020).

Collectively, these GWAS confirmed that ~~MD~~-major depression is highly polygenic, with ORs of individual variants under 1.05 (Wray *et al.*, 2018; Howard *et al.*, 2019), as suggested previously (Nishino *et al.*, 2018). The two loci with the overall most robust support for an association map to the genes *NEGR1* and *OLFM4*, both with an OR of 1.04 and a MAF of ~40% (Wray *et al.*, 2018; Howard *et al.*, 2019). The neuronal growth regulator 1 (*NEGR1*) cell adhesion molecule modulates neurite outgrowth, synapse formation, and synaptic function (Wray *et al.*, 2018; Noh *et al.*, 2019). The secreted glycoprotein Olfactomedin 4 (*OLFM4*) has an established function in inflammation, innate immunity, and cancer (Liu and Rodgers, 2016; Wray *et al.*, 2018). Several other ~~MD~~-associated loci have functions in neurite outgrowth, synaptic function and plasticity, as well as in immunity and inflammation (Wray *et al.*, 2018). To establish the functions of the proposed risk SNPs (single nucleotide polymorphisms) and their links to specific genes, analyses of cell-type-specific gene expression and epigenetic data are necessary (Zhong *et al.*, 2019; Arloth *et al.*, 2020; Chan *et al.*, 2020; Li *et al.*, 2020; Wang *et al.*, 2020).

Given the small effect sizes of single SNPs, gene-level and pathway analyses that integrate the effects of many variants across genes and pathways may be more suitable for studying the



aetiology of MDD. Gene-level association analyses have implicated genes encoding pre- and post-synaptic proteins, especially receptor units, e.g. voltage-gated calcium channels (VGCC), the dopamine receptor *DRD2*, and glutamate receptors (Wray *et al.*, 2018). Identified by both SNP- and gene-level analyses, the splicing regulator *RBFOX1* may play an important role in the aetiology of MD-major depression (Wray *et al.*, 2018; Noh *et al.*, 2019). Gene-set analyses implicated pathways with roles in neuronal differentiation and projection and, most prominently, synaptic formation and function, including gene sets linked to VGCC and synaptic RNA-binding proteins like *FMRP* and *CELF4* (Wray *et al.*, 2018; Noh *et al.*, 2019; Levey *et al.*, 2020).

Despite a large number of identified genomic loci, the variance explained by MD-major depression PRS is still very low (Nagelkerke's pseudo- $R^2$  1.5-3.2% on the liability scale) (Howard *et al.*, 2019). This prediction accuracy was lower than achieved for, e.g. schizophrenia, where Nagelkerke's pseudo- $R^2$  is 7.7% on the liability scale (Schizophrenia Working Group of the PGC *et al.*, 2020). At much smaller sample sizes, the GWAS on schizophrenia also identified many more associated variants (*Figure 1*). The identification of MDD risk variants has been more challenging than for schizophrenia because of its higher prevalence, lower heritability, and more heterogeneous samples (Wray *et al.*, 2012; Levinson *et al.*, 2014; Ormel *et al.*, 2019). Schizophrenia GWAS saw a critical inflection point at approximately 15,000 cases and are currently discovering a new locus for, on average, every 230 new cases (Levinson *et al.*, 2014). Such an inflection point was observed for MD-depression GWAS only at approximately 75,000 cases, as predicted (Wray *et al.*, 2012; Levinson *et al.*, 2014). MD-GWAS using liberal phenotype definitions of major depression now identify a new locus for every 1500 cases (*Figure 1*).

Importantly, phenotype definition matters. Given the strong genetic correlations across cohorts, recent GWAS meta-analyses combined cohorts with different ascertainment strategies to maximize the overall sample size. However, a comparison of genetic analyses using different phenotype definitions in UKBB revealed that GWAS of broad phenotype definitions might identify associated loci that are not specific to MD-MDD, but, more likely, shared with other psychiatric disorders (Cai *et al.*, 2020b). In order to identify robustly associated MD-MDD variants, a balanced strategy of maximising sample size while focusing on more stringent phenotype definitions has been suggested (Davies *et al.*, 2019; Schwabe *et al.*, 2019; Byrne *et al.*, 2020). Quantitative definitions ~~of MD~~ have also been analysed in GWAS, and a continuum between depressive symptoms and MD-MDD was suggested (Direk *et al.*, 2017). Several variants were reported to be associated with depressive symptoms (*Table 3*), with limited replication success. Of these symptom-associated variants, rs62100776 (Okbay *et al.*, 2016) maps to the gene *DCC*, coding for a neuronal cell adhesion molecule, which was a genome-wide significant locus in the 2018 PGC case-control GWAS (Wray *et al.*, 2018). Further noteworthy GWAS were conducted on MD-MDD stratified by the age at onset (Power *et al.*, 2017), sex (Hall *et al.*, 2018), or exposure to adversity (Peterson *et al.*, 2018), or combined GWAS of different depression MD-phenotypes (Amare *et al.*, 2020).

The studies discussed so far focused on patients of European ancestry. However, complex disorders show population-specific landscapes of genetic risk (Martin *et al.*, 2017). Accordingly, GWAS in non-European populations have also been conducted (*Table 3*). The CONVERGE study on female Han Chinese recurrent in-patient MD-MDD cases identified two significant loci near the *SIRT1* and in the *LHPP* gene, and replicated them in the same population (CONVERGE consortium, 2015). Neither variant was significantly associated in GWAS on European samples

(Wray *et al.*, 2018; Howard *et al.*, 2019), possibly due to large differences in allele frequency across ancestries – both variants exhibit markedly lower allele frequencies in Europeans.

Follow-up studies estimated the genetic correlation between East Asian and European ~~MD~~ MDD cohorts at 0.33-0.41, while the average genetic correlations between European cohorts was 0.76 (Bigdeli *et al.*, 2017; Wray *et al.*, 2018), and identified trans-ancestral risk variants (Li *et al.*, 2018). In a study of Hispanic/Latino individuals, two out of ~~25~~22 European ~~MD~~ MDDdepression risk-GWAS loci were replicated (Dunn *et al.*, 2018). The MVP study identified no genome-wide significant loci in over 20,000 African-ancestry ~~MD~~-patients with major depression (Table 3), and only one of 206 analysed European risk SNPs formally replicated in the African sample (Levey *et al.*, 2020). However, a trans-ancestry meta-analysis conducted within the same study increased the number of risk loci from 178 in Europeans alone to 183. These findings highlight the need for greater geographical, ethnic, and economic diversity in ~~MD~~depression genetic studies (McIntosh *et al.*, 2019).

### Genetic overlap between ~~MD~~MDD and other traits

Several lines of evidence, including cross-disorder analyses, support the notion that ~~MD~~MDD shares genetic underpinnings with many other psychiatric phenotypes. A genetic overlap has also been demonstrated between ~~MD~~MDD and non-psychiatric phenotypes. Common techniques for studying cross-disorder genetic overlap include genetic correlation ( $r_g$ ) estimation and PRS analyses (Grotzinger *et al.*, 2019).

Genetic correlation quantifies the genetic relationship between two traits, reflecting their underlying modes of pleiotropy – i.e. the same genetic variant influencing both traits (van Rheenen *et al.*, 2019; Baselmans *et al.*, 2020). Mathematically, it is derived as the genetic

covariance between the two traits scaled by the square roots of their genetic variances (van Rheenen *et al.*, 2019). Two methods are commonly used to estimate  $r_g$  – i) bivariate restricted maximum likelihood (REML) which uses individual-level genotype data as the input (Lee *et al.*, 2012), and ii) linkage disequilibrium score regression (LDSC) (Bulik-Sullivan *et al.*, 2015) and its recent extension, high-definition likelihood (HDL) method (Ning *et al.*, 2020), which uses GWAS summary statistics.  $r_g$  ranges between  $-1$  and  $1$ , with  $1$  meaning that shared liability is caused by the same risk SNPs and  $-1$  meaning that identical SNPs increase the risk for one trait while decreasing risk for the other. It is important to note that  $r_g$  estimates cannot infer causality.

Many studies have examined  $r_g$  between MD-MDD and both psychiatric and non-psychiatric disorders or traits. The strongest  $r_g$  detected for MD-MDD are with depressive symptoms measured as a continuous trait (0.63-1.0) (Direk *et al.*, 2017; Turley *et al.*, 2018), confirming the concept of a continuum between subthreshold depressive symptoms and clinical diagnoses of MD-MDD. Similarly, strong  $r_g$  have been found for personality traits such as neuroticism ( $r_g = 0.70-0.75$ ) and subjective well-being ( $-0.75$ ) (Okbay *et al.*, 2016; Howard *et al.*, 2019). MD-MDD major depression was moderately correlated with several other psychiatric disorders, e.g. autism spectrum disorder (0.45), attention-deficit/hyperactivity disorder (0.44), bipolar disorder (BD) (0.36), and schizophrenia (0.34) (Cross-Disorder Group of the PGC, 2019). Further studies reported a weaker  $r_g$  of MD-MDD major depression with type-I (0.26-0.31) than with type-II BD (0.61-0.69) (Stahl *et al.*, 2019; Coleman *et al.*, 2020). Moreover, a positive  $r_g$  was reported between the body-mass index (BMI) and atypical MD-depression with increased appetite/weight, which was not present for MD-those without changes in appetite (Milaneschi *et al.*, 2017). An investigation of genetic correlations between neurological and psychiatric disorders found little evidence of shared genetic underpinnings, except for a significant but

small positive  $r_g$  between MD-MDD and migraine (The Brainstorm Consortium *et al.*, 2018).

Negative genetic correlations have been detected for MD-major depression with educational attainment and cognitive performance (each  $\sim -0.15$ ) (Howard *et al.*, 2019).

Analyses of PRS is another method used to assess the polygenic overlap between MD-MDD and other traits. PRS use SNP effect sizes from a training GWAS to compute the cumulative genetic risk burden of each individual in a target sample. PRS are calculated as the weighted sum of the allele count multiplied by its effect size across all selected SNPs (Andlauer and Nöthen, 2020; Lewis and Vassos, 2020). Therefore, they capture the combined effects of many SNPs and are used to examine the polygenic liability for a disorder or trait or to examine the genetic overlap across traits (Wray *et al.*, 2014). Although the generation of PRS requires large training GWAS, it can be applied to much smaller target sets. For examples, a MD-MDD PRS was predictive of population-based obsessive-compulsive symptoms (Zilhão *et al.*, 2018). Clinically, PRS may be used for diagnosis, prognosis, and treatment prediction (Schijven *et al.*, 2020). However, PRS for schizophrenia have, so far, shown greater potential for such applications, given their higher variance explained (Schizophrenia Working Group of the PGC *et al.*, 2020). Importantly, PRS were used to demonstrate that MD-MDD subtypes may diverge in their genetic overlaps with other disorders. For example, early-onset MD-MDD cases show a stronger association with PRS for schizophrenia and BD than late-onset MD-MDD patients do (Power *et al.*, 2017). In the future, PRS might also be used to identify aetiologically more homogeneous MD-MDD subtypes. For example, subgrouping on the basis of treatments such as electroconvulsive therapy might be informative (Foo *et al.*, 2019) and provide further insights into the underlying genetic differences of the severity of the disorder. Finally, MD-MDD PRS may also be used to study PRS-

environment (PRS×E) interactions, to support treatment choice, and to refine the penetrance of high-risk variants (Lewis and Vassos, 2020).

### Gene environment interaction

In addition to the heritable component of MDMDD, environmental factors such as stressful life events (e.g. childhood trauma) play a major role in the aetiology of MDMDD, both independently and in combination with genetic factors. One active area of research has been the study of gene-environment interactions (G×E). Early G×E efforts examined specific candidate genes with putative biological significance for MDMDD, e.g. the 5-HTTLPR polymorphism of the serotonin transporter gene, and found evidence of interaction with stressful life events (Caspi *et al.*, 2003). However, these initial reports showed inconsistent replicability (Duncan and Keller, 2011), and recent meta-analytic evidence indicated no robust evidence of candidate G×E effects on MD-MDD (Van der Auwera *et al.*, 2018; Border *et al.*, 2019a). Studies using genome-wide data have reported varied G×E results. First, stratified GWAS results from the CONVERGE study have suggested that MD-MDD may have a stronger genetic basis when it occurs in the absence of environmental adversity, with genome-wide significant hits for MD-MDD identified only among women not exposed to adversity (Peterson *et al.*, 2018). Data from the UKBB, however, have suggested the opposite, with a higher SNP-based heritability for major depression-MD observed among individuals reporting lifetime trauma exposure (Coleman *et al.*, 2020). Second, PRS have been used to aggregate genome-wide effects for MD-MDD and examine their interactions with stressful life events. Again, this revealed discrepant findings where MD-MDD PRS were more strongly associated with MD-MDD among individuals both exposed (Peyrot *et al.*, 2014) and unexposed (Mullins *et al.*, 2016) to childhood trauma. A later but still modestly powered meta-analysis (n=5,765) identified limited

evidence for the interaction of PRS and childhood trauma on MD-MDD (Peyrot *et al.*, 2018), which was echoed in epidemiological and clinical cohorts (n=184-1,450) (Halldorsdottir *et al.*, 2019). However, large-scale studies like the UKBB have provided the opportunity to re-examine PRS×E effects with greater power. One study showed stronger PRS effects on MD-major depression among individuals exposed to childhood trauma and socioeconomic adversity (Shen *et al.*, 2020), though not other types of exposures such as adulthood stressful life events. These findings support the notion that genetic influences could vary by specific environments. Third, genome-wide G×E analyses have identified few variants with genome-wide significant G×E effects (Dunn *et al.*, 2016; Arnau-Soler *et al.*, 2019), although these studies have been largely underpowered and results have not been further replicated. Together, current evidence suggests that genome-wide influences on MD-MDD may vary by stressful life exposures but that the detection of G×E may be sensitive to the research design (e.g. how depression and adversity are defined).

The role of stressful life events, specifically childhood trauma, was the predominant focus of G×E in MD-MDD, but other environmental exposures have also received attention. For example, studies examining social support in combination with MD-MDD PRS in vulnerable populations have consistently shown that social support appears to reduce risk of depression. This effect remained even after accounting for genetic vulnerability for depression, but did not vary by or modify the genetic effects (Choi *et al.*, 2020a; Stringa *et al.*, 2020). Similar observations of independent effects have been made for physical activity and MD-major depression, suggesting that even among individuals at high genetic vulnerability for MD-depression, physical activity levels may still be protective (Choi *et al.*, 2020c). Another genetically informed approach to estimating environmental effects on MD-MDD is Mendelian randomisation (MR), ~~which uses~~

~~genetic instruments to identify potential causal relationships between environmental exposures and MD risk~~ MR methods have emerged as an alternative strategy to examine causal inference, obviating typical sources of confounding (Smith and Ebrahim, 2003). Here, strongly associated genetic variants are selected as statistical instruments — or proxies — to estimate the effect of an exposure of interest (e.g., substance use) on an outcome (e.g., MDD) under specific assumptions. MR studies have supported the role of educational attainment in reducing the depression risk and of BMI in increasing the risk (Wray et al., 2018). ~~These S~~ studies have also suggested a potential role of lifestyle exposures in shaping the risk of ~~MD~~ major depression, such as social connections and sedentary media use (Choi et al., 2020b). Establishing stronger causal evidence for different environmental factors with updated methods may help to clarify inconsistencies across G×E findings (Ni et al., 2019). An additional source of potential bias and inconsistency in G×E research are underlying gene-environment correlations, as the presence of gene-environment correlations may lead to false G×E (Purcell and Sham, 2002; Ni et al., 2019).

## Discussion

In summary, ~~MD~~ MDD is a common, debilitating, phenotypically heterogeneous disorder. It ~~elts~~ with a heritability of ranges between 30-50%. ~~MD~~ MDD with early-onset, recurrence, comorbid anxiety, greater severity, and postpartum depression may be more heritable subtypes.

Compared to other psychiatric disorders, its high prevalence, ~~relatively low~~ moderate heritability, and strong polygenicity have posed major challenges for gene-mapping. Recently, rare CNVs have been implicated in risk of ~~MD~~ MDD. Studies of common variation in ~~MD~~ MDD, driven forward by large international collaborations such as the PGC, confirmed the highly polygenic nature of the disorder and implicated over 100 loci in disorder risk. These studies identified genes involved in neuronal growth, synaptic function, and inflammation in the



pathophysiology of the disorder. Finally, gene-environment interaction and Mendelian randomisation studies have implicated childhood trauma, socioeconomic adversity, social support, physical activity, and their interactions with genetic factors, in risk of [MD-MDD](#).

Evidence on the genetic basis of [MD-MDD](#) has multiple potential applications including the development of pharmacological treatments, informing discussions regarding disorder risk, and treatment choice. A greater understanding of the biological processes underlying [MD-MDD](#) may facilitate the development of new, potentially more effective medications. [MD-MDD](#) GWAS have implicated biological processes underlying the disorder including neurotransmission and synaptic structure as well as genetically informed pharmacological modes of action (Howard *et al.*, 2019). [MD-MDD](#) is a very heterogeneous disorder and genetic research will likely benefit from the study of more homogeneous disorder subtypes. [For example, research on patients with atypical features of increased appetite or weight suggested that this patient subgroup might benefit from treatments specifically targeting immunometabolic pathways \(Milaneschi \*et al.\*, 2017\).](#)

An individual's genetic make-up likely impacts the disease course – from prevention to diagnosis, treatment, and prognosis (Lewis and Vassos, 2020). Whilst the application of [MD-MDD](#) PRS to predict individual risk is currently limited by its low discriminative accuracy, in the future [MD-PRS](#) might be used for risk stratification at the population level (Andlauer and Nöthen, 2020). [This is supported by recent evidence that MDD PRS – well-established to predict prevalent depression – also predict incident cases in the general population \(Musliner \*et al.\*, 2019\).](#) [For other complex disorders like coronary artery disease and type-2 diabetes, PRS-based risk stratification is already powerful \(Khera \*et al.\*, 2018\).](#) Further development of PRS to allow

the identification of individuals at heightened genetic risk, coupled with targeted prevention and intervention measures, holds potential to mitigate depression risk (McIntosh *et al.*, 2019). Knowledge of the genetic architecture of MD-MDD and its links with other psychiatric and non-psychiatric disorders might inform genetic counselling and clinical care. For example, identified genetic correlations with physical health phenotypes might facilitate the development of clinical guidelines for physical health monitoring in individuals with MDMDD.

MDD is a multifactorial disorder. Other levels of genetic variation, such as DNA methylation (Barbu *et al.*, 2020; Van Assche *et al.*, 2017b) and subsequent gene-expression differences (Arloth *et al.*, 2015; Schiweck *et al.*, 2020), and also the telomere length (Lin *et al.*, 2016), influence depression risk and severity. Importantly, epigenetic mechanisms serve as an interface between the individual and its environment (Provencal and Binder, 2015). A better understanding of the role of epigenetic variation and an improved integration of multi-omics data may enhance our understanding of the aetiology and treatment of MDD. In addition, this review focuses on genetic factors influencing the risk or severity of depression. We did not discuss how genetic variants affect response to treatment, especially the metabolisation of antidepressant drugs via cytochrome P450 genes. This important topic was recently reviewed by Bousman and colleagues (Bousman *et al.* 2021).

Here, we highlighted the state of the art, the promises, and the pitfalls of genetic research in MDMDD. The field would benefit from further exploration of the stratification properties of genetics – both for the understanding of its aetiology, including the role of environmental factors, and for the development of targeted treatment opportunities. Rich datasets with high-quality and deep phenotyping are crucial for studying heterogeneous disorders like MDD. Combining deep, harmonized phenotypes with molecular data will greatly aid the genetic

stratification of MDD. Improving our insights into the impact of genetic factors on ~~MD~~-MDD

holds great promise for improving clinical care, and this intense and fast-developing field has a bright future.

### **Acknowledgments**

We thank Andrew McIntosh, Cathryn Lewis, and Naomi Wray for critical comments on our manuscript.

## Tables and Figures

Table 1. MDD heritability estimates from selected publications

<b>Key paper</b>	<b>Type</b>	<b><math>h^2</math> (95% CI)</b>	<b><math>c^2</math> (95% CI)</b>	<b><math>e^2</math> (95% CI)</b>
<b>Twin data</b>				
(Sullivan <i>et al.</i> , 2000)	Meta-analysis	0.37 (0.31-0.42)	0 (0-0.05)	0.63 (0.58-0.67)
(Kendler <i>et al.</i> , 2006)	Largest twin study (15.5K pairs)	male: 0.29 (0.19-0.38) female: 0.42 (0.36-0.47)	NA <sup>a</sup>	0.71 (0.62-0.81) 0.58 (0.53-0.64)
(Polderman <i>et al.</i> , 2015)	Meta-analysis of 50 years of twin studies	depressive episode: 0.34 (0.31-0.38) recurrent MDD: 0.45 (0.36-0.55)	0.11 (0.08-0.13) 0.03 (0-0.08)	0.55 0.52
(Kendler <i>et al.</i> , 2018a)	Twins identified from SWE register; Treated cases	male: 0.41 (0.19-0.49) female: 0.49 (0.31-0.56)	0 (0-0.13) 0.02 (0-0.17)	0.59 (0.51-0.66) 0.49 (0.44-0.55)
<b>EHR or register-based data</b>				
(Wray and Gottesman, 2012)	DEN register, parent-offspring	0.32 (0.30-0.34)	NA <sup>b</sup>	NA <sup>b</sup>
(Kendler <i>et al.</i> , 2018a)	SWE register, siblings (full vs half, reared together vs apart)	male: 0.36 (0.31-0.38) female: 0.51 (0.51-0.53)	0.05 (0.05-0.05) 0.02 (0.02-0.02)	0.59 (0.58-0.61) 0.47 (0.46-0.49)
(Polubriaginof <i>et al.</i> , 2018)	US EHR, <i>inferred</i> extended pedigree	depressive episode: 0.25 (0.17-0.30) recurrent MDD: 0.36 (0.21-0.51)	NA <sup>a</sup> NA <sup>a</sup>	0.75 (0.70-0.83) 0.64 (0.49-0.79)

$h^2$  narrow-sense heritability;  $c^2$  shared environment;  $e^2$  unique environment. SWE: Sweden; DEN: Denmark. EHR: Electronic Health Records.

- a. Missing estimate of  $c^2$  as a result of AE models (i.e., quantitative genetic models with only additive genetics and unique environment components) in the original study.
- b. Only the heritability was derived from the summary level data.

Table 2. Whole-exome sequencing and exome chip studies assessing the role of rare variants in subphenotypes of depression.

Phenotype	Method	Sample size	Ancestry	Implicated gene or variant	Implicated pathway	Reference
Population isolate	WES, exome chip	1999 (discovery) 1604 (replication)	European	<i>NKPD1</i>	de novo synthesis of sphingolipids	(Amin <i>et al.</i> , 2017a)
<a href="#">Depressive symptoms in families &amp; Familial MD</a> & population isolate	Linkage, haplotype analysis, WES, exome chip	1336 WES 1527 exome chip	European	<i>RCL1</i> p.Leu186Phe (rs115482041)	interlaminar astrocytes	(Amin <i>et al.</i> , 2018)
Familial & early-onset MDD	targeted resequencing of 1742 synaptic genes	259 cases 334 controls	European	Cav <sub>v2</sub> -adapter gene set	calcium signaling, actin polymerisation, spine formation	(Pirooznia <i>et al.</i> , 2016)
Familial & early-onset MDD	WES	12	European	<i>ZNF34</i>	brain TFs	(Subaran <i>et al.</i> , 2016)
Sex differences in MDD	WES and WGS	1000 cases (70.7% female) 72 controls	Asian	<i>PDE4A</i> (rs201432982), <i>FDX1L</i> (rs62640397, rs79442975), <i>MYO15B</i> (rs820182, rs820148) more frequent in female cases	none	(Kang <i>et al.</i> , 2020)
Suicide in MDD	WES specifically with 5'/3'-UTRs and promoters	23 cases 21 controls (post-mortem brain samples)	European	<i>COL6A6</i> variants in 17% of suicide completers; 3 calcium channel genes highlighted ( <i>CACNA1B</i> , <i>-1C</i> , <i>-2D4</i> ) 86 non-synonymous variants in 42 genes found in at least two suicide victims but not controls	TGF- $\beta$ pathway	(Tombácz <i>et al.</i> , 2017)
MRI grey matter volume changes in MDD	WES	77 cases 245 controls	Asian	<i>CSMD1</i> , <i>CNTNAP5</i>	neuroactive ligand receptor interacting pathway	(Zhang <i>et al.</i> , 2020)

Depressive symptoms, general population	WES	3612	European	<i>LIPG</i> p.Asn396Ser (rs77960347)	steroid and cholesterol biosynthesis	(Amin <i>et al.</i> , 2017b)
Depressive symptoms in context of environmental stress	exome chip, WGS	203 cases, 196 controls (discovery) 473 cases, 497 controls (replication)	Hispanic (discovery), European (replication)	<i>PHF21B</i>	9 pathways including innate immune response, glutamate signalling, sensory perception	(Wong <i>et al.</i> , 2017)
Treatment resistance	GWAS, WES, exome chip	1209 cases (discovery) 1128 cases (replication)	European	response predictor in resistance vs. response to treatment with serotonergic antidepressants using rare variants	cell survival and proliferation, neurodegeneration, immune response	(Fabbri <i>et al.</i> , 2020)

MDD=major depressive disorder; WES=whole-exome sequencing; WGS=whole-genome sequencing;  
 GWAS=genome-wide association study; TFs=transcription factors; UTR=untranslated region; TGF- $\beta$ =Transforming growth factor beta; MRI=magnetic resonance imaging

Table 3. Published genome-wide association studies on depression.

Year	Discovery: Nca/Nco or Ntot	Discovery: Loci	Replication: Cases/Controls	Replicated loci	<del>MD-d</del> Definition of depression	Ancestry	Reference
2008	1738/1802	0	6079/5893	0*1	<del>Structured</del> Structured interview: <u>MDD</u>	European	(Sullivan <i>et al.</i> , 2009) <sup>‡</sup>
2008	1359/1782	0	NA	NA	<u>Structured interview:</u> <del>MDD</del> Structured interview	European	(Muglia <i>et al.</i> , 2010) <sup>‡</sup>
2009	3957/3428	0	NA	NA	<u>Structured interview:</u> <del>MDD</del> Structured interview	European	(Shyn <i>et al.</i> , 2011) <sup>‡</sup>
2010	1636/1594	0	1418/1918	0	<u>Structured interview:</u> <del>MDD</del> Structured interview	European	(Lewis <i>et al.</i> , 2010) <sup>‡</sup>
2010	604/1364	0	409/541	0	<u>Structured interview:</u> <del>MDD</del> Structured interview	European	(Rietschel <i>et al.</i> , 2010) <sup>‡</sup>
2010	1020/1636	0	NA	NA	<u>Structured interview:</u> <del>MDD</del> Structured interview	European	(Shi <i>et al.</i> , 2011) <sup>‡</sup>
2010	2431/3673	0	3332/3228	0	<u>Structured interview:</u> <del>MDD</del> Structured interview	European	(Wray <i>et al.</i> , 2012) <sup>‡</sup>
2011	353/366	0	3735/10635	1*2	<u>Structured interview:</u> <del>MDD</del> Structured interview	European	(Kohli <i>et al.</i> , 2011) <sup>‡</sup>

2012	9240/9519	0	6783/50695	0	Structured interview: MDDStructured interview	European	( <a href="#">Ripke et al., 2013</a> ) <sup>‡</sup>
2013	34549	0	16709	1* <sup>3</sup>	Depressive symptoms	European	( <a href="#">Hek et al., 2013</a> )
2015	5303/5337	2	3231/3186	2	Structured interview: Women with recurrent MDD	Han Chinese	(CONVERGE consortium, 2015)
2015	16948	0	NA	NA	Depressive symptoms	Trans-ancestry	( <a href="#">Ware et al., 2015</a> )
2015	1443	1* <sup>4</sup>	NA	NA	Depressive symptoms	Hispanic American	( <a href="#">Ware et al., 2015</a> )
2016	32528	1	6813	0	Depressive symptoms	European	( <a href="#">Demirkan et al., 2016</a> )
2016	7179	0	NA	NA	Depressive symptoms	African American	( <a href="#">Dunn et al., 2016</a> )
2016	3138	0	NA	NA	Depressive symptoms	Hispanic/ Latino	( <a href="#">Dunn et al., 2016</a> )
2016	161460	2	368890	2	Depressive symptoms	European	( <a href="#">Okbay et al., 2016</a> )
2016	3869/9519	1	6107/124230	1	Structured interview: Adult- onset MDD	European	( <a href="#">Power et al., 2017</a> )
2016	75607/231747 and 84847/241266* <sup>4</sup>	7 and 4* <sup>5</sup>	45773/106354	15* <sup>6</sup>	Self-reported diagnosis	European	( <a href="#">Hyde et al., 2016</a> ) <sup>‡</sup>



2016	70017	1	38328	1	Mixed (including depressive symptoms)	European	(Direk <i>et al.</i> , 2017)
2017	14729/14435	1	4504/7007	0	Structured interview: <u>MDD</u>	European & Han Chinese	(Bigdeli <i>et al.</i> , 2017)
2017	5347/1871	1	NA	NA	Structured interview: MDD with decreased vs increased appetite/ weight	European	(Milaneschi <i>et al.</i> , 2017)
2018	12310	0	7948	0	Depressive symptoms	Hispanic/ Latino	(Dunn <i>et al.</i> , 2018)
2018	90150/246603	10	2659/17237	6*7	Mixed	European & Han Chinese	(Li <i>et al.</i> , 2018)
2018	10851/32211	0	NA	NA	Mixed	European	(Hall <i>et al.</i> , 2018) <sup>‡</sup>
2018	3852/16034	1	NA	NA	Mixed: Men only	European	(Hall <i>et al.</i> , 2018)
2018	3139/3832	3	NA	NA	Structured interview: Women without exposure to adversity	Han Chinese	(Peterson <i>et al.</i> , 2018)
2018	113769/208811	14	75607/231747	14* <sup>3</sup>	Broad depression	European	(Howard <i>et al.</i> , 2018) <sup>‡</sup>
2018	30603/143916	2	75607/231747	2* <sup>3</sup>	Probable MDD	European	(Howard <i>et al.</i> , 2018) <sup>‡</sup>
2018	8276/209308	1	75607/231747	1* <sup>3</sup>	ICD-coded MDD	European	(Howard <i>et al.</i> , 2018)
2018	135458/344901	44	NA	NA	Mixed	European	(Wray <i>et al.</i> , 2018) <sup>‡</sup>

2019	Combination of Direk et al. and Hyde et al.	3	Howard et al. 2018: broad depression	3	Broad and self-reported <del>MDD</del> major depression	European	(Amare <i>et al.</i> , 2020)
2019	Combination of Direk et al. and CONVERGE	1	Howard et al. 2018: broad depression	0	Mixed: Broad depression and recurrent MDD	European & Han Chinese	(Amare <i>et al.</i> , 2020)
2019	246363/561190	101* <sup>8</sup>	414574/892299	87	Mixed	European	(Howard <i>et al.</i> , 2019) <sup>‡</sup>
2019	164933/408383* 8	63	NA	NA	Mixed	European	(Coleman <i>et al.</i> , 2020) <sup>‡</sup>
2020	113262/219360	24	NA	NA	Help-seeking: seen doctor	European	(Cai <i>et al.</i> , 2020b)
2020	36286/297126	5	NA	NA	Help-seeking: seen psychiatrist	European	(Cai <i>et al.</i> , 2020b) <sup>‡</sup>
2020	16301/50870	1	NA	NA	Structured <del>interview</del> interview: MDD	European	(Cai <i>et al.</i> , 2020b)
2020	83810/166405	10	NA	NA	ICD-based	European	(Levey <i>et al.</i> , 2020) <sup>‡</sup>
2020	330173/824094	178	NA	NA	Mixed	European	(Levey <i>et al.</i> , 2020) <sup>‡</sup>
2020	25843/33757	0	NA	NA	ICD-based	African American	(Levey <i>et al.</i> , 2020)
2020	366434/847433	183	NA	NA	Mixed	European and African American	(Levey <i>et al.</i> , 2020)

Year indicates the year in which the study was first published (including online ahead of print and, where no published version as available, preprint), which may deviate from the year the article was published

in print.  $Nca/Nco$  for number of cases and controls in case-control design;  $Ntot$  as total sample size for quantitative trait.

‡ Studies used for Figure 1.

\*<sup>1</sup> The top locus replicated in subsequent studies (Hek et al. 2010, Wray et al. 2018, Howard et al. 2019).

\*<sup>2</sup> Significant in the meta-analysis of discovery and replication cohorts in a recessive model. This association did not replicate in subsequent GWAS in additive models. Replication has, to our knowledge, not been attempted using a recessive model.

\*<sup>3</sup> Significant in the meta-analysis of discovery and replication cohorts.

\*<sup>4</sup> This association did not replicate in the follow-up Hispanic ancestry GWAS meta-analysis Dunn *et al.* 2018.

\*<sup>4</sup> Meta-analysis of 23andMe and PGC MDD cohorts.

\*<sup>5</sup> Replication of top results from the meta-analysis of 23andMe and PGC MDD cohorts.

\*<sup>6</sup> Only SNPs not reaching genome-wide significance in the discovery-stage were analysed in the replication phase. Six loci (of which two were not significant in the discovery GWAS) covered by these selected SNPs reached significance in the meta-analysis of discovery and replication cohorts.

\*<sup>7</sup> From a meta-analysis of overlapping subjects (PGC and UKBB), case and control numbers may thus be inflated.

\*<sup>8</sup> The study identified 102 independent variants at 101 loci.

### Figure 1: Identified GWAS loci in relation to the number of cases.

The relationship of identified loci to the number of cases based on published studies on

schizophrenia, bipolar disorder (BD), and major depression (MDD). To simplify the presentation,

studies on both MDD and MD with different depression definitions have been combined into

the MD category. The plot shows the number of cases and independent significant loci in the

discovery-stage analyses (without replication). Schizophrenia GWAS saw a critical inflection

point at approximately 15,000 cases and are currently discovering a new locus for, on average, every 230 new cases (Levinson *et al.*, 2014). BD currently discovers a new locus for every 400 new cases. Such an inflection point was observed for MD GWAS only at approximately 75,000 cases, as predicted (Wray *et al.*, 2012; Levinson *et al.*, 2014), and MD GWAS using liberal phenotype definitions now identify a new locus every 1500 cases. Slopes starting at the inflection point were estimated using linear regression. The MD studies used for the plot are provided in Table 3.

## References

- [Amare, A. T., Vaez, A., Hsu, Y. H., Direk, N., Kamali, Z., Howard, D. M., ... & Hartman, C. A. \(2019\). Bivariate genome-wide association analyses of the broad depression phenotype combined with major depressive disorder, bipolar disorder or schizophrenia reveal eight novel genetic loci for depression. \*Molecular psychiatry\*, 1-10.](#)
- [Amin, N., Belonogova, N. M., Jovanova, O., Brouwer, R. W., van Rooij, J. G., van den Hout, M. C., ... & van Duijn, C. M. \(2017a\) Nonsynonymous variation in NKPD1 increases depressive symptoms in European populations. \*Biological Psychiatry\*, 81\(8\), 702-707.](#)
- [Amin, N., Jovanova, O., Adams, H. H. H., Dehghan, A., Kavousi, M., Vernooij, M. W., ... & Van Duijn, C. M. \(2017b\). Exome-sequencing in a large population-based study reveals a rare Asn396Ser variant in the LIPG gene associated with depressive symptoms. \*Molecular psychiatry\*, 22\(4\), 537-543.](#)
- [Amin, N., De Vrij, F. M. S., Baghdadi, M., Brouwer, R. W. W., Van Rooij, J. G. J., Jovanova, O., ... & Van Duijn, C. M. \(2018\). A rare missense variant in RCL1 segregates with depression in extended families. \*Molecular psychiatry\*, 23\(5\), 1120-1126.](#)
- [Andlauer, T. F., & Nöthen, M. M. \(2020\). Polygenic scores for psychiatric disease: from research tool to clinical application. \*Medizinische Genetik\*, 32\(1\), 39-45.](#)
- [Arloth, J., Bogdan, R., Weber, P., Frishman, G., Menke, A., Wagner, K. V., ... & Hariri, A. R. \(2015\). Genetic differences in the immediate transcriptome response to stress predict risk-related brain function and psychiatric disorders. \*Neuron\*, 86\(5\), 1189-1202..](#)
- [Arloth, J., Eraslan, G., Andlauer, T. F., Martins, J., Iurato, S., Kühnel, B., ... & Mueller, N. S. \(2020\). DeepWAS: Multivariate genotype-phenotype associations by directly integrating regulatory information using deep learning. \*PLoS computational biology\*, 16\(2\), e1007616.](#)
- [Arnau-Soler, A., Macdonald-Dunlop, E., Adams, M. J., Clarke, T. K., MacIntyre, D. J., Milburn, K., ... & Thomson, P. A. \(2019\). Genome-wide by environment interaction studies of depressive symptoms and psychosocial stress in UK Biobank and Generation Scotland. \*Translational psychiatry\*, 9\(1\), 1-13.](#)
- [Barbu, M. C., Shen, X., Walker, R. M., Howard, D. M., Evans, K. L., Whalley, H. C., ... & McIntosh, A. M. \(2020\). Epigenetic prediction of major depressive disorder. \*Molecular Psychiatry\*, 1-12.](#)
- [Baselmans, B. M., Yengo, L., van Rheenen, W., & Wray, N. R. \(2020\). Risk in relatives, heritability, SNP-based heritability and genetic correlations in psychiatric disorders: a review. \*Biological Psychiatry\*.](#)
- [Beckman, G., Beckman, L., Cedergren, B., Perris, C., & Strandman, E. \(1978\). Serum protein and red cell enzyme polymorphisms in affective disorders. \*Human heredity\*, 28\(1\), 41-47.](#)
- [Bigdeli, T. B., Ripke, S., Peterson, R. E., Trzaskowski, M., Bacanu, S. A., Abdellaoui, A., ... & Kendler, K. S. \(2017\). Genetic effects influencing risk for major depressive disorder in China and Europe. \*Translational psychiatry\*, 7\(3\), e1074-e1074.](#)
- [Birmaher, B., Ryan, N. D., Williamson, D. E., Brent, D. A., Kaufman, J., Dahl, R. E., ... & Nelson, B. \(1996\). Childhood and adolescent depression: a review of the past 10 years. Part I. \*Journal of the American Academy of Child & Adolescent Psychiatry\*, 35\(11\), 1427-1439.](#)
- [Border, R., Johnson, E. C., Evans, L. M., Smolen, A., Berley, N., Sullivan, P. F., & Keller, M. C. \(2019a\). No support for historical candidate gene or candidate gene-by-interaction hypotheses for major depression across multiple large samples. \*American Journal of Psychiatry\*, 176\(5\), 376-387.](#)
- [Border, R., Smolen, A., Corley, R. P., Stallings, M. C., Brown, S. A., Conger, R. D., ... & Evans, L. M. \(2019b\). Imputation of behavioral candidate gene repeat variants in 486,551 publicly-available UK Biobank individuals. \*European Journal of Human Genetics\*, 27\(6\), 963.](#)

- [Bosker, F. J., Hartman, C. A., Nolte, I. M., Prins, B. P., Terpstra, P., Posthuma, D., ... & Nolen, W. A. \(2011\). Poor replication of candidate genes for major depressive disorder using genome-wide association data. \*Molecular psychiatry\*, 16\(5\), 516-532.](#)
- [Bousman, C. A., Bengesser, S. A., Aitchison, K. J., Amare, A. T., Aschauer, H., Baune, B. T., ... & Müller, D. J. \(2020\). Review and Consensus on Pharmacogenomic Testing in Psychiatry. \*Pharmacopsychiatry\*.](#)
- [Brainstorm Consortium. \(2018\). Analysis of shared heritability in common disorders of the brain. \*Science \(New York, NY\)\*, 360\(6395\).](#)
- [Breen, G., Webb, B. T., Butler, A. W., Van Den Oord, E. J., Tozzi, F., Craddock, N., ... & McGuffin, P. \(2011\). A genome-wide significant linkage for severe depression on chromosome 3: the depression network study. \*American Journal of Psychiatry\*, 168\(8\), 840-847.](#)
- [Bromet, E., Andrade, L. H., Hwang, I., Sampson, N. A., Alonso, J., De Girolamo, G., ... & Kessler, R. C. \(2011\). Cross-national epidemiology of DSM-IV major depressive episode. \*BMC medicine\*, 9\(1\), 90.](#)
- [Bulik-Sullivan, B., Finucane, H. K., Anttila, V., Gusev, A., Day, F. R., Loh, P. R., ... & Psychiatric Genomics Consortium. \(2015\). An atlas of genetic correlations across human diseases and traits. \*Nature genetics\*, 47\(11\), 1236.](#)
- [Byrne, E. M., Kirk, K. M., Medland, S. E., McGrath, J. J., Colodro-Conde, L., Parker, R., ... & Martin, N. G. \(2020\). Cohort profile: the Australian genetics of depression study. \*BMJ open\*, 10\(5\), e032580.](#)
- [Cai, N., Choi, K. W., & Fried, E. I. \(2020a\). Reviewing the genetics of heterogeneity in depression: operationalizations, manifestations and etiologies. \*Human molecular genetics\*, 29\(R1\), R10-R18.](#)
- [Cai, N., Revez, J. A., Adams, M. J., Andlauer, T. F., Breen, G., Byrne, E. M., ... & Flint, J. \(2020b\). Minimal phenotyping yields genome-wide association signals of low specificity for major depression. \*Nature Genetics\*, 52\(4\), 437-447.](#)
- [Camp, N. J., Lowry, M. R., Richards, R. L., Plenk, A. M., Carter, C., Hensel, C. H., ... & Cannon-Albright, L. A. \(2005\). Genome-wide linkage analyses of extended Utah pedigrees identifies loci that influence recurrent, early-onset major depression and anxiety disorders. \*American Journal of Medical Genetics Part B: Neuropsychiatric Genetics\*, 135\(1\), 85-93.](#)
- [Caspi, A., Sugden, K., Moffitt, T. E., Taylor, A., Craig, I. W., Harrington, H., ... & Poulton, R. \(2003\). Influence of life stress on depression: moderation by a polymorphism in the 5-HTT gene. \*Science\*, 301\(5631\), 386-389.](#)
- [Chan, R. F., Turecki, G., Shabalin, A. A., Guintivano, J., Zhao, M., Xie, L. Y., ... & van den Oord, E. J. \(2020\). Cell Type-Specific Methylome-wide Association Studies Implicate Neurotrophin and Innate Immune Signaling in Major Depressive Disorder. \*Biological psychiatry\*, 87\(5\), 431-442.](#)
- [Choi, K. W., Chen, C. Y., Ursano, R. J., Sun, X., Jain, S., Kessler, R. C., ... & Smoller, J. W. \(2020a\). Prospective study of polygenic risk, protective factors, and incident depression following combat deployment in US Army soldiers. \*Psychological Medicine\*, 50\(5\), 737-745.](#)
- [Choi, K. W., Stein, M. B., Nishimi, K. M., Ge, T., Coleman, J. R., Chen, C. Y., ... & Smoller, J. W. \(2020b\). An exposure-wide and Mendelian randomization approach to identifying modifiable factors for the prevention of depression. \*American Journal of Psychiatry\*, 177\(10\), 944-954.](#)
- [Choi, K. W., Zheutlin, A. B., Karlson, R. A., Wang, M. J., Dunn, E. C., Stein, M. B., ... & Smoller, J. W. \(2020c\). Physical activity offsets genetic risk for incident depression assessed via electronic health records in a biobank cohort study. \*Depression and anxiety\*, 37\(2\), 106-114.](#)
- [Coe, B. P., Witherspoon, K., Rosenfeld, J. A., Van Bon, B. W., Vulto-van Silfhout, A. T., Bosco, P., ... & Eichler, E. E. \(2014\). Refining analyses of copy number variation identifies specific genes associated with developmental delay. \*Nature genetics\*, 46\(10\), 1063-1071.](#)

- [Coleman, J. R., Gaspar, H. A., Bryois, J., Byrne, E. M., Forstner, A. J., Holmans, P. A., ... & Lawson, W. B. \(2020\). The genetics of the mood disorder Spectrum: genome-wide association analyses of more than 185,000 cases and 439,000 controls. \*Biological psychiatry\*, 88\(2\), 169-184.](#)
- [CONVERGE consortium \(2015\) \(2015\). Sparse whole-genome sequencing identifies two loci for major depressive disorder. \*Nature\*, 523\(7562\), 588-591.](#)
- [Cooper, G. M., Coe, B. P., Girirajan, S., Rosenfeld, J. A., Vu, T. H., Baker, C., ... & Eichler, E. E. \(2011\). A copy number variation morbidity map of developmental delay. \*Nature genetics\*, 43\(9\), 838-846.](#)
- [Cross-Disorder Group of the PGC \(2019\). Genomic relationships, novel loci, and pleiotropic mechanisms across eight psychiatric disorders. \*Cell\*, 179\(7\), 1469-1482.](#)
- [Davies, M. R., Kalsi, G., Armour, C., Jones, I. R., McIntosh, A. M., Smith, D. J., ... & Breen, G. \(2019\). The Genetic Links to Anxiety and Depression \(GLAD\) Study: online recruitment into the largest recontactable study of depression and anxiety. \*Behaviour research and therapy\*, 123, 103503.](#)
- [Degenhardt, F., Priebe, L., Herms, S., Mattheisen, M., Mühleisen, T. W., Meier, S., ... & Cichon, S. \(2012\). Association between copy number variants in 16p11. 2 and major depressive disorder in a German case-control sample. \*American Journal of Medical Genetics Part B: Neuropsychiatric Genetics\*, 159\(3\), 263-273.](#)
- [Demirkan, A., Lahti, J., Direk, N., Viktorin, A., Lunetta, K. L., Terracciano, A., ... & Luik, A. I. \(2016\). Somatic, positive and negative domains of the Center for Epidemiological Studies Depression \(CES-D\) scale: a meta-analysis of genome-wide association studies. \*Psychological medicine\*, 46\(8\), 1613.](#)
- [Direk, N., Williams, S., Smith, J. A., Ripke, S., Air, T., Amare, A. T., ... & Sullivan, P. F. \(2017\). An analysis of two genome-wide association meta-analyses identifies a new locus for broad depression phenotype. \*Biological psychiatry\*, 82\(5\), 322-329.](#)
- [Duncan, L. E., & Keller, M. C. \(2011\). A critical review of the first 10 years of candidate gene-by-environment interaction research in psychiatry. \*American Journal of Psychiatry\*, 168\(10\), 1041-1049.](#)
- [Dunn, E. C., Sofer, T., Wang, M. J., Soare, T. W., Gallo, L. C., Gogarten, S. M., ... & Smoller, J. W. \(2018\). Genome-wide association study of depressive symptoms in the Hispanic Community Health Study/Study of Latinos. \*Journal of psychiatric research\*, 99, 167-176.](#)
- [Dunn, E. C., Wiste, A., Radmanesh, F., Almlil, L. M., Gogarten, S. M., Sofer, T., ... & Smoller, J. W. \(2016\). Genome-wide association study \(GWAS\) and genome-wide by environment interaction study \(GWEIS\) of depressive symptoms in African American and Hispanic/Latina women. \*Depression and anxiety\*, 33\(4\), 265-280.](#)
- [Fabbri, C., Kasper, S., Kautzky, A., Zohar, J., Souery, D., Montgomery, S., ... & Serretti, A. \(2020\). A polygenic predictor of treatment-resistant depression using whole exome sequencing and genome-wide genotyping. \*Translational psychiatry\*, 10\(1\), 1-12.](#)
- [Farrell, M. S., Werge, T., Sklar, P., Owen, M. J., Ophoff, R. A., O'Donovan, M. C., ... & Sullivan, P. F. \(2015\). Evaluating historical candidate genes for schizophrenia. \*Molecular psychiatry\*, 20\(5\), 555-562.](#)
- [Fernandez-Pujals, A. M., Adams, M. J., Thomson, P., McKechnie, A. G., Blackwood, D. H., Smith, B. H., ... & McIntosh, A. M. \(2015\). Epidemiology and heritability of major depressive disorder, stratified by age of onset, sex, and illness course in Generation Scotland: Scottish Family Health Study \(GS: SFHS\). \*PloS one\*, 10\(11\), e0142197.](#)
- [Flint, J., & Kendler, K. S. \(2014\). The genetics of major depression. \*Neuron\*, 81\(3\), 484-503.](#)
- [Foo, J. C., Streit, F., Frank, J., Witt, S. H., Treutlein, J., Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium, ... & Kranaster, L. \(2019\). Evidence for increased genetic risk load for major depression in patients assigned to electroconvulsive therapy. \*American Journal of Medical Genetics Part B: Neuropsychiatric Genetics\*, 180\(1\), 35-45.](#)

- [Fried, E. I., Coomans, F., & Lorenzo-Luaces, L. \(2020\). The 341 737 ways of qualifying for the melancholic specifier. \*The Lancet Psychiatry\*, 7\(6\), 479-480.](#)
- [Fried, E. I., & Nesse, R. M. \(2015\). Depression is not a consistent syndrome: an investigation of unique symptom patterns in the STAR\\* D study. \*Journal of affective disorders\*, 172, 96-102.](#)
- [GBD 2017 DALYs and HALE Collaborators \(2018\). Global, regional, and national disability-adjusted life-years \(DALYs\) for 359 diseases and injuries and healthy life expectancy \(HALE\) for 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. \*The Lancet\*, 392\(10159\), 1859-1922.](#)
- [Glessner, J. T., Wang, K., Sleiman, P. M., Zhang, H., Kim, C. E., Flory, J. H., ... & Hakonarson, H. \(2010\). Duplication of the SLIT3 locus on 5q35.1 predisposes to major depressive disorder. \*PLoS one\*, 5\(12\), e15463.](#)
- [Grotzinger, A. D., Rhemtulla, M., de Vlaming, R., Ritchie, S. J., Mallard, T. T., Hill, W. D., ... & Tucker-Drob, E. M. \(2019\). Genomic structural equation modelling provides insights into the multivariate genetic architecture of complex traits. \*Nature human behaviour\*, 3\(5\), 513-525.](#)
- [Hall, L. S., Adams, M. J., Arnau-Soler, A., Clarke, T. K., Howard, D. M., Zeng, Y., ... & McIntosh, A. M. \(2018\). Genome-wide meta-analyses of stratified depression in Generation Scotland and UK Biobank. \*Translational psychiatry\*, 8\(1\), 1-12.](#)
- [Halldorsdottir, T., Piechaczek, C., Soares de Matos, A. P., Czamara, D., Pehl, V., Wagenbuechler, P., ... & Binder, E. B. \(2019\). Polygenic risk: predicting depression outcomes in clinical and epidemiological cohorts of youths. \*American Journal of Psychiatry\*, 176\(8\), 615-625.](#)
- [Hek, K., Demirkan, A., Lahti, J., Terracciano, A., Teumer, A., Cornelis, M. C., ... & Murabito, J. \(2013\). A genome-wide association study of depressive symptoms. \*Biological psychiatry\*, 73\(7\), 667-678.](#)
- [Hek, K., Mulder, C. L., Luijendijk, H. J., van Duijn, C. M., Hofman, A., Uitterlinden, A. G., & Tiemeier, H. \(2010\). The PCLO gene and depressive disorders: replication in a population-based study. \*Human Molecular Genetics\*, 19\(4\), 731-734.](#)
- [Hemani, G., Yang, J., Vinkhuyzen, A., Powell, J. E., Willemsen, G., Hottenga, J. J., ... & Visscher, P. M. \(2013\). Inference of the genetic architecture underlying BMI and height with the use of 20,240 sibling pairs. \*The American Journal of Human Genetics\*, 93\(5\), 865-875.](#)
- [Holmans, P., Weissman, M. M., Zubenko, G. S., Scheftner, W. A., Crowe, R. R., DePaulo Jr., MD, J. R., ... & Levinson, D. F. \(2007\). Genetics of recurrent early-onset major depression \(GenRED\): final genome scan report. \*American Journal of Psychiatry\*, 164\(2\), 248-258.](#)
- [Holmans, P., Zubenko, G. S., Crowe, R. R., DePaulo Jr, J. R., Scheftner, W. A., Weissman, M. M., ... & Levinson, D. F. \(2004\). Genomewide significant linkage to recurrent, early-onset major depressive disorder on chromosome 15q. \*The American Journal of Human Genetics\*, 74\(6\), 1154-1167.](#)
- [Howard, D. M., Adams, M. J., Clarke, T. K., Hafferty, J. D., Gibson, J., Shiralil, M., ... & McIntosh, A. M. \(2019\). Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. \*Nature neuroscience\*, 22\(3\), 343-352.](#)
- [Howard, D. M., Adams, M. J., Shiralil, M., Clarke, T. K., Marioni, R. E., Davies, G., ... & McIntosh, A. M. \(2018\). Genome-wide association study of depression phenotypes in UK Biobank identifies variants in excitatory synaptic pathways. \*Nature communications\*, 9\(1\), 1-10.](#)
- [Hyde, C. L., Nagle, M. W., Tian, C., Chen, X., Paciga, S. A., Wendland, J. R., ... & Winslow, A. R. \(2016\). Identification of 15 genetic loci associated with risk of major depression in individuals of European descent. \*Nature genetics\*, 48\(9\), 1031-1036.](#)



- [International Schizophrenia Consortium. \(2008\). Rare chromosomal deletions and duplications increase risk of schizophrenia. \*Nature\*, 455\(7210\), 237.](#)
- [Kang, H. J., Park, Y., Yoo, K. H., Kim, K. T., Kim, E. S., Kim, J. W., ... & Kim, J. M. \(2020\). Sex differences in the genetic architecture of depression. \*Scientific reports\*, 10\(1\), 1-12.](#)
- [Kendall, K. M., Rees, E., Bracher-Smith, M., Legge, S., Riglin, L., Zammit, S., ... & Walters, J. T. R. \(2019\). Association of rare copy number variants with risk of depression. \*JAMA psychiatry\*, 76\(8\), 818-825.](#)
- [Kendler, K. S., Gatz, M., Gardner, C. O., & Pedersen, N. L. \(2006\). A Swedish national twin study of lifetime major depression. \*American Journal of Psychiatry\*, 163\(1\), 109-114.](#)
- [Kendler, K. S., Ohlsson, H., Lichtenstein, P., Sundquist, J., & Sundquist, K. \(2018a\). The genetic epidemiology of treated major depression in Sweden. \*American Journal of Psychiatry\*, 175\(11\), 1137-1144.](#)
- [Kendler, K. S., Ohlsson, H., Sundquist, K., & Sundquist, J. \(2018b\). Sources of parent-offspring resemblance for major depression in a national Swedish extended adoption study. \*Jama Psychiatry\*, 75\(2\), 194-200.](#)
- [Kessler, R. C., & Bromet, E. J. \(2013\). The epidemiology of depression across cultures. \*Annual review of public health\*, 34, 119-138.](#)
- [Khera, A. V., Chaffin, M., Aragam, K. G., Haas, M. E., Roselli, C., Choi, S. H., ... & Kathiresan, S. \(2018\). Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. \*Nature genetics\*, 50\(9\), 1219-1224.](#)
- [Kohli, M. A., Lucae, S., Saemann, P. G., Schmidt, M. V., Demirkan, A., Hek, K., ... & Binder, E. B. \(2011\). The neuronal transporter gene SLC6A15 confers risk to major depression. \*Neuron\*, 70\(2\), 252-265.](#)
- [Lee, S., Abecasis, G. R., Boehnke, M., & Lin, X. \(2014\). Rare-variant association analysis: study designs and statistical tests. \*The American Journal of Human Genetics\*, 95\(1\), 5-23.](#)
- [Lee, S. H., Yang, J., Goddard, M. E., Visscher, P. M., & Wray, N. R. \(2012\). Estimation of pleiotropy between complex diseases using single-nucleotide polymorphism-derived genomic relationships and restricted maximum likelihood. \*Bioinformatics\*, 28\(19\), 2540-2542.](#)
- [Levey, D. F., Stein, M. B., Wendt, F. R., Pathak, G. A., Zhou, H., Aslan, M., ... & Gelernter, J. \(2020\). GWAS of Depression Phenotypes in the Million Veteran Program and Meta-analysis in More than 1.2 Million Participants Yields 178 Independent Risk Loci. \*medRxiv\*.](#)
- [Levinson, D. F., Evgrafov, O. V., Knowles, J. A., Potash, J. B., Weissman, M. M., Scheftner, W. A., ... & Holmans, P. \(2007\). Genetics of recurrent early-onset major depression \(GenRED\): significant linkage on chromosome 15q25-q26 after fine mapping with single nucleotide polymorphism markers. \*American Journal of Psychiatry\*, 164\(2\), 259-264.](#)
- [Levinson, D. F., Mostafavi, S., Milaneschi, Y., Rivera, M., Ripke, S., Wray, N. R., & Sullivan, P. F. \(2014\). Genetic studies of major depressive disorder: why are there no genome-wide association study findings and what can we do about it?. \*Biological psychiatry\*, 76\(7\), 510-512.](#)
- [Lewis, C. M., Ng, M. Y., Butler, A. W., Cohen-Woods, S., Uher, R., Pirlo, K., ... & McGuffin, P. \(2010\). Genome-wide association study of major recurrent depression in the UK population. \*American Journal of Psychiatry\*, 167\(8\), 949-957.](#)
- [Lewis, C. M., & Vassos, E. \(2020\). Polygenic risk scores: from research tools to clinical instruments. \*Genome Medicine\*, 12, 1-11.](#)
- [Li, S., Li, Y., Li, X., Liu, J., Huo, Y., Wang, J., ... & Luo, X. J. \(2020\). Regulatory mechanisms of major depressive disorder risk variants. \*Molecular Psychiatry\*, 1-20.](#)

[Li, X., Luo, Z., Gu, C., Hall, L. S., McIntosh, A. M., Zeng, Y., ... & Luo, X. J. \(2018\). Common variants on 6q16. 2, 12q24. 31 and 16p13. 3 are associated with major depressive disorder. \*Neuropsychopharmacology\*, 43\(10\), 2146-2153.](#)

[Lin, P. Y., Huang, Y. C., & Hung, C. F. \(2016\). Shortened telomere length in patients with depression: a meta-analytic study. \*Journal of psychiatric research\*, 76, 84-93.](#)

[Liu, W., & Rodgers, G. P. \(2016\). Olfactomedin 4 expression and functions in innate immunity, inflammation, and cancer. \*Cancer and Metastasis Reviews\*, 35\(2\), 201-212.](#)

[Lynch, M., & Walsh, B. \(1998\). \*Genetics and analysis of quantitative traits\* \(Vol. 1, pp. 535-557\). Sunderland, MA: Sinauer.](#)

[Malhi GS, Mann JJ \(2018\) Depression. \*The Lancet\* 392, 2299–2312.](#)

[Manolio, T. A., Collins, F. S., Cox, N. J., Goldstein, D. B., Hindorff, L. A., Hunter, D. J., ... & Visscher, P. M. \(2009\). Finding the missing heritability of complex diseases. \*Nature\*, 461\(7265\), 747-753.](#)

[Martin, A. R., Gignoux, C. R., Walters, R. K., Wojcik, G. L., Neale, B. M., Gravel, S., ... & Kenny, E. E. \(2017\). Human demographic history impacts genetic risk prediction across diverse populations. \*The American Journal of Human Genetics\*, 100\(4\), 635-649.](#)

[McIntosh, A. M., Sullivan, P. F., & Lewis, C. M. \(2019\). Uncovering the genetic architecture of major depression. \*Neuron\*, 102\(1\), 91-103.](#)

[Middeldorp, C. M., Hottenga, J. J., Slagboom, P. E., Sullivan, P. F., de Geus, E. J., Posthuma, D., ... & Boomsma, D. I. \(2008\). Linkage on chromosome 14 in a genome-wide linkage study of a broad anxiety phenotype. \*Molecular psychiatry\*, 13\(1\), 84-89.](#)

[Milaneschi, Y., Lamers, F., Peyrot, W. J., Baune, B. T., Breen, G., Dehghan, A., ... & Penninx, B. W. \(2017\). Genetic association of major depression with atypical features and obesity-related immunometabolic dysregulations. \*Jama Psychiatry\*, 74\(12\), 1214-1225.](#)

[Moskvina, V., Farmer, A., Swainson, V., O'Leary, J., Gunasinghe, C., Owen, M., ... & Korszun, A. \(2007\). Interrelationship of childhood trauma, neuroticism, and depressive phenotype. \*Depression and anxiety\*, 24\(3\), 163-168.](#)

[Muglia, P., Tozzi, F., Galwey, N. W., Francks, C., Upmanyu, R., Kong, X. Q., ... & Roses, A. D. \(2010\). Genome-wide association study of recurrent major depressive disorder in two European case-control cohorts. \*Molecular psychiatry\*, 15\(6\), 589-601.](#)

[Mullins, N., Power, R. A., Fisher, H. L., Hanscombe, K. B., Euesden, J., Iniesta, R., ... & Lewis, C. M. \(2016\). Polygenic interactions with environmental adversity in the aetiology of major depressive disorder. \*Psychological medicine\*, 46\(4\), 759-770.](#)

[Musliner, K. L., Mortensen, P. B., McGrath, J. J., Suppli, N. P., Hougaard, D. M., Bybjerg-Grauholm, J., ... & Agerbo, E. \(2019\). Association of polygenic liabilities for major depression, bipolar disorder, and schizophrenia with risk for depression in the Danish population. \*JAMA psychiatry\*, 76\(5\), 516-525.](#)

[Ni, G., van der Werf, J., Zhou, X., Hyppönen, E., Wray, N. R., & Lee, S. H. \(2019\). Genotype-covariate correlation and interaction disentangled by a whole-genome multivariate reaction norm model. \*Nature communications\*, 10\(1\), 1-15.](#)

[Ning, Z., Pawitan, Y., & Shen, X. \(2020\). High-definition likelihood inference of genetic correlations across human complex traits. \*Nature genetics\*, 52\(8\), 859-864.](#)

[Nishino, J., Ochi, H., Kochi, Y., Tsunoda, T., & Matsui, S. \(2018\). Sample size for successful genome-wide association study of major depressive disorder. \*Frontiers in genetics\*, 9, 227.](#)

[Noh, K., Lee, H., Choi, T. Y., Joo, Y., Kim, S. J., Kim, H., ... & Lee, S. J. \(2019\). Negr1 controls adult hippocampal neurogenesis and affective behaviors. \*Molecular psychiatry\*, 24\(8\), 1189-1205.](#)

[O'Dushlaine, C., Ripke, S., Ruderfer, D. M., Hamilton, S. P., Fava, M., Iosifescu, D. V., ... & Perlis, R. H. \(2014\). Rare copy number variation in treatment-resistant major depressive disorder. \*Biological psychiatry\*, 76\(7\), 536-541.](#)

[Okbay, A., Baselmans, B. M., De Neve, J. E., Turley, P., Nivard, M. G., Fontana, M. A., ... & Rich, S. S. \(2016\). Genetic variants associated with subjective well-being, depressive symptoms, and neuroticism identified through genome-wide analyses. \*Nature genetics\*, 48\(6\), 624-633.](#)

[Ormel, J., Hartman, C. A., & Snieder, H. \(2019\). The genetics of depression: successful genome-wide association studies introduce new challenges. \*Translational psychiatry\*, 9\(1\), 1-10.](#)

[Perlis, R. H., Ruderfer, D., Hamilton, S. P., & Ernst, C. \(2012\). Copy number variation in subjects with major depressive disorder who attempted suicide.](#)

[Peterson, R. E., Cai, N., Dahl, A. W., Bigdeli, T. B., Edwards, A. C., Webb, B. T., ... & Kendler, K. S. \(2018\). Molecular genetic analysis subdivided by adversity exposure suggests etiologic heterogeneity in major depression. \*American Journal of Psychiatry\*, 175\(6\), 545-554.](#)

[Peyrot, W. J., Milaneschi, Y., Abdellaoui, A., Sullivan, P. F., Hottenga, J. J., Boomsma, D. I., & Penninx, B. W. \(2014\). Effect of polygenic risk scores on depression in childhood trauma. \*The British Journal of Psychiatry\*, 205\(2\), 113-119.](#)

[Peyrot, W. J., Van der Auwera, S., Milaneschi, Y., Dolan, C. V., Madden, P. A., Sullivan, P. F., ... & Yang, J. \(2018\). Does childhood trauma moderate polygenic risk for depression? A meta-analysis of 5765 subjects from the psychiatric genomics consortium. \*Biological psychiatry\*, 84\(2\), 138-147.](#)

[Pirooznia, M., Wang, T., Avramopoulos, D., Potash, J. B., Zandi, P. P., & Goes, F. S. \(2016\). High-throughput sequencing of the synaptome in major depressive disorder. \*Molecular psychiatry\*, 21\(5\), 650-655.](#)

[Polderman, T. J., Benyamin, B., De Leeuw, C. A., Sullivan, P. F., Van Bochoven, A., Visscher, P. M., & Posthuma, D. \(2015\). Meta-analysis of the heritability of human traits based on fifty years of twin studies. \*Nature genetics\*, 47\(7\), 702-709.](#)

[Polubriaginof, F. C., Vanguri, R., Quinnes, K., Belbin, G. M., Yahi, A., Salmasian, H., ... & Tatonetti, N. P. \(2018\). Disease heritability inferred from familial relationships reported in medical records. \*Cell\*, 173\(7\), 1692-1704.](#)

[Power, R. A., Tansey, K. E., Buttenschøn, H. N., Cohen-Woods, S., Bigdeli, T., Hall, L. S., ... & GERAD1 Consortium. \(2017\). Genome-wide association for major depression through age at onset stratification: major depressive disorder working group of the psychiatric genomics consortium. \*Biological Psychiatry\*, 81\(4\), 325-335.](#)

[Provencal, N., & Binder, E. B. \(2015\). The neurobiological effects of stress as contributors to psychiatric disorders: focus on epigenetics. \*Current opinion in neurobiology\*, 30, 31-37.](#)

[Purcell, S., & Sham, P. \(2002\). Variance components models for gene-environment interaction in quantitative trait locus linkage analysis. \*Twin Research and Human Genetics\*, 5\(6\), 572-576.](#)

[Van Rheenen, W., Peyrot, W. J., Schork, A. J., Lee, S. H., & Wray, N. R. \(2019\). Genetic correlations of polygenic disease traits: from theory to practice. \*Nature Reviews Genetics\*, 20\(10\), 567-581.](#)

[Rice, S. M., Fallon, B. J., Aucote, H. M., Möller-Leimkühler, A., Treeby, M. S., & Amminger, G. P. \(2015\). Longitudinal sex differences of externalising and internalising depression symptom trajectories: Implications for assessment of depression in men from an online study. \*International Journal of Social Psychiatry\*, 61\(3\), 236-240.](#)

- [Rietschel, M., Mattheisen, M., Frank, J., Treutlein, J., Degenhardt, F., Breuer, R., ... & Cichon, S. \(2010\). Genome-wide association-, replication-, and neuroimaging study implicates HOMER1 in the etiology of major depression. \*Biological psychiatry\*, 68\(6\), 578-585.](#)
- [Ripke, S., Wray, N. R., Lewis, C. M., Hamilton, S. P., Weissman, M. M., Breen, G., ... & Major Depressive Disorder Working Group of the Psychiatric GWAS Consortium. \(2013\). A mega-analysis of genome-wide association studies for major depressive disorder. \*Molecular psychiatry\*, 18\(4\), 497.](#)
- [Rucker, J. J., Breen, G., Pinto, D., Pedroso, I., Lewis, C. M., Cohen-Woods, S., ... & McGuffin, P. \(2013\). Genome-wide association analysis of copy number variation in recurrent depressive disorder. \*Molecular psychiatry\*, 18\(2\), 183-189.](#)
- [Rucker, J. J., Tansey, K. E., Rivera, M., Pinto, D., Cohen-Woods, S., Uher, R., ... & Breen, G. \(2016\). Phenotypic association analyses with copy number variation in recurrent depressive disorder. \*Biological psychiatry\*, 79\(4\), 329-336.](#)
- [Schijven, D., Veldink, J. H., & Luykx, J. J. \(2020\). Genetic cross-disorder analysis in psychiatry: from methodology to clinical utility. \*The British Journal of Psychiatry\*, 216\(5\), 246-249.](#)
- [Schiwack, C., Claes, S., Van Oudenhove, L., Lafit, G., Vaessen, T., de Beeck, G. O., ... & Vrieze, E. \(2020\). Childhood trauma, suicide risk and inflammatory phenotypes of depression: insights from monocyte gene expression. \*Translational psychiatry\*, 10\(1\), 1-12.](#)
- [Schizophrenia Working Group of the PGC, Ripke S, Walters JT, O'Donovan MC \(2020\). Mapping genomic loci prioritises genes and implicates synaptic biology in schizophrenia. \*MedRxiv\*.](#)
- [Schol-Gelok, S., Janssens, A. C. J., Tiemeier, H., Liu, F., Lopez-Leon, S., Zorkoltseva, I. V., ... & van Duijn, C. M. \(2010\). A genome-wide screen for depression in two independent Dutch populations. \*Biological psychiatry\*, 68\(2\), 187-196.](#)
- [Schwabe, I., Milaneschi, Y., Gerring, Z., Sullivan, P. F., Schulte, E., Suppli, N. P., ... & Middeldorp, C. M. \(2019\). Unraveling the genetic architecture of major depressive disorder: merits and pitfalls of the approaches used in genome-wide association studies. \*Psychological medicine\*, 49\(16\), 2646-2656.](#)
- [Shadrina, M., Bondarenko, E. A., & Slominsky, P. A. \(2018\). Genetics factors in major depression disease. \*Frontiers in psychiatry\*, 9, 334.](#)
- [Shen, X., Howard, D. M., Adams, M. J., Hill, W. D., Clarke, T. K., Deary, I. J., ... & McIntosh, A. M. \(2020\). A phenome-wide association and Mendelian Randomisation study of polygenic risk for depression in UK Biobank. \*Nature communications\*, 11\(1\), 1-16.](#)
- [Shi, J., Potash, J. B., Knowles, J. A., Weissman, M. M., Coryell, W., Scheftner, W. A., ... & Levinson, D. F. \(2011\). Genome-wide association study of recurrent early-onset major depressive disorder. \*Molecular psychiatry\*, 16\(2\), 193-201.](#)
- [Shyn, S. I., Shi, J., Kraft, J. B., Potash, J. B., Knowles, J. A., Weissman, M. M., ... & Hamilton, S. P. \(2011\). Novel loci for major depression identified by genome-wide association study of Sequenced Treatment Alternatives to Relieve Depression and meta-analysis of three studies. \*Molecular psychiatry\*, 16\(2\), 202-215.](#)
- [Smith, D. J., Nicholl, B. I., Cullen, B., Martin, D., Ul-Haq, Z., Evans, J., ... & Pell, J. P. \(2013\). Prevalence and characteristics of probable major depression and bipolar disorder within UK biobank: cross-sectional study of 172,751 participants. \*PLoS one\*, 8\(11\), e75362.](#)
- [Davey Smith, G., & Ebrahim, S. \(2003\). 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease?. \*International journal of epidemiology\*, 32\(1\), 1-22.](#)
- [Stahl, E. A., Breen, G., Forstner, A. J., McQuillin, A., Ripke, S., Trubetskoy, V., ... & Reif, A. \(2019\). Genome-wide association study identifies 30 loci associated with bipolar disorder. \*Nature genetics\*, 51\(5\), 793-803.](#)

[Stefansson, H., Rujescu, D., Cichon, S., Pietiläinen, O. P., Ingason, A., Steinberg, S., ... & Stefansson, K. \(2008\). Large recurrent microdeletions associated with schizophrenia. \*nature\*, 455\(7210\), 232-236.](#)

[Stringa, N., Milaneschi, Y., van Schoor, N. M., Suanet, B., van der Lee, S., Holstege, H., ... & Huisman, M. \(2020\). Genetic liability for depression, social factors and their interaction effect in depressive symptoms and depression over time in older adults. \*The American Journal of Geriatric Psychiatry\*.](#)

[Subaran, R. L., Odgerel, Z., Swaminathan, R., Glatt, C. E., & Weissman, M. M. \(2016\). Novel variants in ZNF34 and other brain-expressed transcription factors are shared among early-onset MDD relatives. \*American Journal of Medical Genetics Part B: Neuropsychiatric Genetics\*, 171\(3\), 333-341.](#)

[Sullivan, P. F., Neale, M. C., & Kendler, K. S. \(2000\). Genetic epidemiology of major depression: review and meta-analysis. \*American Journal of Psychiatry\*, 157\(10\), 1552-1562.](#)

[Sullivan, P. F., de Geus, E. J., Willemsen, G., James, M. R., Smit, J. H., Zandbelt, T., ... & Penninx, B. W. \(2009\). Genome-wide association for major depressive disorder: a possible role for the presynaptic protein piccolo. \*Molecular psychiatry\*, 14\(4\), 359-375.](#)

[Tansey, K. E., Rucker, J. J., Kavanagh, D. H., Guipponi, M., Perroud, N., Bondolfi, G., ... & Uher, R. \(2014\). Copy number variants and therapeutic response to antidepressant medication in major depressive disorder. \*The pharmacogenomics journal\*, 14\(4\), 395-399.](#)

[Tombácz, D., Maróti, Z., Kalmár, T., Csabai, Z., Balázs, Z., Takahashi, S., ... & Boldogkői, Z. \(2017\). High-coverage whole-exome sequencing identifies candidate genes for suicide in victims with major depressive disorder. \*Scientific reports\*, 7\(1\), 1-11.](#)

[Turley, P., Walters, R. K., Maghzian, O., Okbay, A., Lee, J. J., Fontana, M. A., ... & Benjamin, D. J. \(2018\). Multi-trait analysis of genome-wide association summary statistics using MTAG. \*Nature genetics\*, 50\(2\), 229-237.](#)

[Van Assche, E., Moons, T., Cinar, O., Viechtbauer, W., Oldehinkel, A. J., Van Leeuwen, K., ... & van Winkel, R. \(2017a\). Gene-based interaction analysis shows GABAergic genes interacting with parenting in adolescent depressive symptoms. \*Journal of Child Psychology and Psychiatry\*, 58\(12\), 1301-1309.](#)

[Van Assche, E., Vangeel, E., Freson, K., Van Leeuwen, K., Verschueren, K., Colpin, H., ... & Claes, S. \(2017b\). Epigenome-wide analysis of methylation and perceived parenting in adolescents and its correlation with depressive symptoms over time. \*European Neuropsychopharmacology\*, 27, S449-S450.](#)

[Van der Auwera, S., Peyrot, W. J., Milaneschi, Y., Hertel, J., Baune, B., Breen, G., ... & Grabe, H. \(2018\). Genome-wide gene-environment interaction in depression: A systematic evaluation of candidate genes: The childhood trauma working-group of PGC-MDD. \*American Journal of Medical Genetics Part B: Neuropsychiatric Genetics\*, 177\(1\), 40-49.](#)

[Viktorin, A., Meltzer-Brody, S., Kuja-Halkola, R., Sullivan, P. F., Landén, M., Lichtenstein, P., & Magnusson, P. K. \(2016\). Heritability of perinatal depression and genetic overlap with nonperinatal depression. \*American Journal of Psychiatry\*, 173\(2\), 158-165.](#)

[Visscher, P. M., Hill, W. G., & Wray, N. R. \(2008\). Heritability in the genomics era—concepts and misconceptions. \*Nature reviews genetics\*, 9\(4\), 255-266.](#)

[Wang, X., Cheng, W., Zhu, J., Yin, H., Chang, S., Yue, W., & Yu, H. \(2020\). Integrating genome-wide association study and expression quantitative trait loci data identifies NEGR1 as a causal risk gene of major depression disorder. \*Journal of Affective Disorders\*, 265, 679-686.](#)

[Ware, E. B., Mukherjee, B., Sun, Y. V., Diez-Roux, A. V., Kardina, S. L., & Smith, J. A. \(2015\). Comparative genome-wide association studies of a depressive symptom phenotype in a repeated measures setting by race/ethnicity in the multi-ethnic study of atherosclerosis. \*BMC genetics\*, 16\(1\), 118.](#)

[Wong, M. L., Arcos-Burgos, M., Liu, S., Velez, J. I., Yu, C., Baune, B. T., ... & Licinio, J. \(2017\). The PHF21B gene is associated with major depression and modulates the stress response. \*Molecular psychiatry\*, 22\(7\), 1015-1025.](#)

[Wray, N. R., & Gottesman, I. I. \(2012\). Using summary data from the danish national registers to estimate heritabilities for schizophrenia, bipolar disorder, and major depressive disorder. \*Frontiers in genetics\*, 3, 118.](#)

[Wray, N. R., Lee, S. H., Mehta, D., Vinkhuyzen, A. A., Dudbridge, F., & Middeldorp, C. M. \(2014\). Research review: polygenic methods and their application to psychiatric traits. \*Journal of child psychology and psychiatry\*, 55\(10\), 1068-1087.](#)

[Wray, N. R., Pergadia, M. L., Blackwood, D. H. R., Penninx, B. W. J. H., Gordon, S. D., Nyholt, D. R., ... & Sullivan, P. F. \(2012\). Genome-wide association study of major depressive disorder: new results, meta-analysis, and lessons learned. \*Molecular psychiatry\*, 17\(1\), 36-48.](#)

[Wray, N. R., Ripke, S., Mattheisen, M., Trzaskowski, M., Byrne, E. M., Abdellaoui, A., ... & Viktorin, A. \(2018\). Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. \*Nature genetics\*, 50\(5\), 668-681.](#)

[Yu, C., Arcos-Burgos, M., Baune, B. T., Arolt, V., Dannlowski, U., Wong, M. L., & Licinio, J. \(2018\). Low-frequency and rare variants may contribute to elucidate the genetics of major depressive disorder. \*Translational psychiatry\*, 8\(1\), 1-8.](#)

[Yu, C., Baune, B. T., Wong, M. L., & Licinio, J. \(2017\). Investigation of copy number variation in subjects with major depression based on whole-genome sequencing data. \*Journal of Affective Disorders\*, 220, 38-42.](#)

[Zhang, X., Abdellaoui, A., Rucker, J., de Jong, S., Potash, J. B., Weissman, M. M., ... & Levinson, D. F. \(2019\). Genome-Wide burden of rare short deletions is enriched in major depressive disorder in four cohorts. \*Biological psychiatry\*, 85\(12\), 1065-1073.](#)

[Zhang, Y., Li, M., Wang, Q., Hsu, J. S., Deng, W., Ma, X., ... & Li, T. \(2020\). A joint study of whole exome sequencing and structural MRI analysis in major depressive disorder. \*Psychological Medicine\*, 50\(3\), 384-395.](#)

[Zhong, J., Li, S., Zeng, W., Li, X., Gu, C., Liu, J., & Luo, X. J. \(2019\). Integration of GWAS and brain eQTL identifies FLOT1 as a risk gene for major depressive disorder. \*Neuropsychopharmacology\*, 44\(9\), 1542-1551.](#)

[Zilhão, N. R., Abdellaoui, A., Smit, D. J. A., Cath, D. C., Hottenga, J. J., & Boomsma, D. I. \(2018\). Polygenic prediction of obsessive compulsive symptoms. \*Mol Psychiatry\*, 23, 168-9.](#)

[Zubenko, G. S., Maher, B., Hughes III, H. B., Zubenko, W. N., Stiffler, J. S., Kaplan, B. B., & Marazita, M. L. \(2003\). Genome-wide linkage survey for genetic loci that influence the development of depressive disorders in families with recurrent, early-onset, major depression. \*American Journal of Medical Genetics Part B: Neuropsychiatric Genetics\*, 123\(1\), 1-18.](#)

***Response to Reviewers – Psychological Medicine PSM-D-20-01822 – “The Genetic Basis of Major Depression” by Kendall et al.***

*Thank you for the careful review of our manuscript and the helpful feedback. In this response, we have addressed all issues raised, point-by-point. We have also carefully revised the manuscript and submitted a track-changed copy (MDDreview\_R1\_trackchanged) and a cleaned copy (MDDreview\_R1\_clean). We believe the manuscript is substantially improved with the constructive suggestions from the editor and the reviewers. We confirm this revised manuscript has met all the journal's requirements.*

**Reviewers' and editor's comments:**

Both reviewers were broadly positive but had some concerns. PM does not take a strong position on MD vs. MDD initials. But what is important is to define what you mean - especially to divide studies that apply full (typically DSM) criteria or not and to distinguish whether that is done in a clinical interview or be questionnaire and whether you are assessing current mood state or a lifetime history of episodes of MD. Both reviewers felt more attention to measurement issues was warranted. Other concerns were also raised. Your MS is already a bit long for us - typical upper limit 4500. So take a few more hundred words if needed, but not more than that in your revision.

*We agree with the editor and the reviewers that a clear definition is of vital importance, given that the construct definition is considered as one of the key sources of heterogeneity in MDD (Cai N, Choi KW, Fried EI. Human Molecular Genetics, 2020; PMID: 32568380). Therefore, we clarified the depression definitions in the Introduction (Page 5):*

*“In line with previous studies (Wray et al., 2018), we will use ‘MDD’ where the study subjects met the clinical criteria for major depressive disorder, and ‘major depression’ where the subjects were ascertained using alternative strategies, but nevertheless likely to have met the diagnostic criteria for MDD.”*

*Throughout the review, we have adjusted the use of the terms MDD and major depression to reflect the specific ascertainment strategies used in each cited study.*

*For more details, please also refer to our responses to Reviewer 1 comment 1 (i, iv, v), and Reviewer 2 comment 1.*

**Reviewer #1: Kendall et al., The Genetic Basis of Major Depression**

This is a well-written and comprehensive review article on genetics of depression, with the major focus on molecular genetic methods and findings. The authors have read the literature carefully and they generally review it with good perspective on the differences among methods and the strength of evidence for various findings. This would be a good place for a non-specialist to find an overview and a perspective of research to date, and for researchers to find references on the topics that are covered here.

*Thank you for this positive assessment.*

The following are mostly relatively minor comments that might help to improve the review. The first item is, in my view, a slightly more major point that I think should be corrected - recognizing that it is not easy to judge when it is necessary to use the term "MDD" vs. "MD," so perhaps the editor will weigh in on the journal's preference here.

*We have now carefully revised the definitions as suggested. Please refer to our response to the editor's comment for further details.*

1. The introductory paragraphs on defining depression are important but need improvement:  
(i) It should be stated that the definition here is for the major depressive disorder (MDD) category (DSM-V, ICD-10). The alternative of studying continuous depression questionnaire scores is discussed later but should be mentioned here.

*We agree and thank the reviewer for this suggestion. In response, we have added the following text to the Definition section (Page 4-5):*

*“The term ‘depression’ most often refers to the clinical diagnosis of MDD, as defined by the International Classification of Diseases (ICD) and the Diagnostic and Statistical Manual of Mental Disorders (DSM) diagnostic criteria. [...] Apart from such a dichotomous, diagnosis-based definition, the severity of depressive symptoms is often studied as a quantitative trait, so the full variation of depressive symptoms can be taken into account.”*

(ii) Consider mentioning that the MDD category is used by most studies because it predicts familial risk (relatives or twins), antidepressant treatment response and future course (which does not prove it is the best way to measure depression, but most genetic researchers have concluded that it is the best available way).

*We added a brief comment in the Definition section (Page 5):*

*“Despite the shortcomings of dichotomizing a heterogeneous disorder like MDD, a categorical definition of depression has been widely used in research.”*

*The specific examples mentioned by the reviewer were brought up in later sections.*

iii) The Cai et al. and PGC perspectives are not presented clearly. Correctly or not, both prefer the MDD category. PGC papers assert that because GWAS using "light phenotyping" methods (self-report, records) have high genetic (SNP) correlation with more rigorous interview-based studies, thus it is valid and more powerful to meta-analyze them. Cai asserts that examination of data for different strategies suggest that less rigorous methods risk missing genetic risk factors for the most severe forms of depressive illness.

*We thank the reviewer for raising this point, and agree that the relevant text can be improved so that the two perspectives are more clearly presented.*

*However, we prefer to discuss these perspectives fully in a later section (e.g., after we introduced the GWAS, Page 15), and limit the Introduction section to broadly presenting the different strategies.*

*We now added the PGC perspective on Page 14 where we discussed this topic:*

*“[Importantly, phenotype definition matters]. Given the strong genetic correlations across cohorts, recent GWAS meta-analyses combined cohorts with different ascertainment strategies to maximize the overall sample size.”*

(iv) PGC's "major depression" terminology is only partially explained. It does refer to the increased diagnostic uncertainty when combining diverse studies; but it also rests on the fact that all of the included strategies are targeting individuals likely to meet MDD criteria, as opposed to SSGAC which specifically includes studies based on symptom scores.

*We added the following clarification on Page 5:*



*“[...] and ‘major depression’ where the subjects were ascertained using alternative strategies, but nevertheless likely to have met the diagnostic criteria for MDD.”*

(v) The authors err in using "MD" when citing research on rigorously diagnosed MDD. "MDD" should be used when it was explicitly studied (i.e., for most of the literature cited here), "MD" or "proxy for MDD" when self-reported diagnoses or the like were used or were part of the range of definitions that were used; and "symptom score" or something comparable where appropriate; with a comment if the phenotyping strategy falls somewhere in between. The PGC's use of "MD" is an informal, descriptive term; it is not an historical, validated concept. An example of the negatives effects of this imprecision is a sentence on p 14, "A comparison of genetic analyses using different phenotype definitions in UKBB revealed that GWAS of broad phenotype definitions might identify associated loci that are not specific to MD (Cai et al., 2020b)." Cai's point is rather, that MD≠MDD, and that analyses of liberally-defined MD might miss critical loci associated with severe MDD. (I am agnostic personally about that conclusion, but it should be made more clear here.) On p 17, where it is stated that the high rg between MD and symptom score studies confirms "the concept of a continuum between subthreshold depressive symptoms and clinical diagnoses of MD." There is no such thing as a "clinical diagnosis of MD," only of MDD.

*We thank the reviewer for raising this important point. As the reviewer correctly pointed out, the majority of the cited studies used strict criteria to define MDD. It was inappropriate to combine all studies into a single category under the liberally-defined term MD. Now, we explicitly defined the two terms, MDD and major depression (see our response to the editor's comment) and used them correspondingly throughout the review.*

*In addition, we clarified the text on Page 14:*

*“[...] However, a comparison of genetic analyses using different phenotype definitions in UKBB revealed that GWAS of broad phenotype definitions might identify associated loci that are not specific to MDD, but, more likely, shared with other psychiatric disorders (Cai et al., 2020b).”*

*And on Page 16:*

*“[...] confirming the concept of a continuum between subthreshold depressive symptoms and clinical diagnoses of MDD.”*

(vi) The Fried and Hesse create a "straw man" regarding "unique combinations" of MDD symptoms. No combination is empirically unique or is a valid individual characteristic (i.e., endorsed symptoms show some variation across "test/re-test" interviews during an episode, or during different episodes, or during one interview that explicitly inquires about different episodes). But meeting MDD criteria has validated (if imperfect) significance.

*This reference is not closely related to the key topics of our review. Therefore, we decided to remove the respective sentence and the reference in order to shorten the manuscript (Page 4 in the Introduction).*

2. In the paragraph on CNVs, the Zhang et al. and Kendall et al. studies are described, but it is not made clear whether Kendall et al. actually tested the finding reported in Zhang et al. (increased burden for deletion < 100 kb) (I think they did not, but the authors of this review also wrote the excellent Kendall et al. CNV paper and could clarify.)

*Essentially, the analyses of Zhang et al. and Kendall et al. overlapped only partially. CNVs <100kb, which were significant in Zhang et al., were not tested in Kendall et al. Conversely,*

*the neurodevelopmental CNVs assessed in Kendall et al. were not specifically screened in Zhang et al. The relevant sentences were changed, which now read as follows (Page 11):*

*“Outside of this group, there was no evidence for a residual burden of rare CNVs >100 kb in major depression. However, smaller CNVs were not tested in the study by Kendall and colleagues.”*

3. It would be helpful to state at the beginning of the GWAS review that European-ancestry studies will be reviewed first. Otherwise it is a bit glaring that the CONVERGE GWAS paper is missing from the initial review of the history of significant MD/MDD GWAS findings.

*We agree with the reviewer and added the following sentence to the start of the GWAS section (Page 11):*

*“Many MDD GWAS have been conducted during the past decade, predominantly on subjects of European ancestry (Table 3). Therefore, European-ancestry studies are described first, followed by a review of GWAS in more diverse populations.”*

4. Mendelian Randomization is first introduced on p 20 in the context of GxE analysis. It would be helpful to introduce more broadly what the method does, and then discuss it in the specific GxE context. The MR analyses in the Wray et al. (2018) paper, specifically of obesity, are not mentioned, but should be.

*We thank the reviewer for this suggestion. We have now added a brief introduction to Mendelian randomization methods and discuss them in the context of studying environmental influences on depression on Page 20:*

*“[Another genetically informed approach to estimating environmental effects on MDD is Mendelian randomisation (MR).] MR methods have emerged as an alternative strategy to examine causal inference, obviating typical sources of confounding (Smith and Ebrahim, 2003). Here, strongly associated genetic variants are selected as statistical instruments — or proxies — to estimate the effect of an exposure of interest (e.g., substance use) on an outcome (e.g., MDD) under specific assumptions. MR studies have supported the role of educational attainment in reducing the depression risk and of BMI in increasing the risk (Wray et al., 2018). Studies have also supported the potential role of various lifestyle exposures in shaping the risk of MDD, such as social connections and sedentary media use (Choi et al., 2020b).”*

5. Parts of the Discussion read a bit too much like an advertisement for genetic studies - it is stated that the current state of this research does not adequately predict individual risk, and then all sorts of future benefits are described that would depend on much better individual predictions. It would be better to briefly note some of the uses that have been proposed (especially if supported by any empirical data), but not to oversell whether they will become reality. For example, I do not think that people with or without severe MDD need to know their PRS score to know that they might do well to eat better and exercise more. The problem isn't knowing that it might help, it is figuring out how to get oneself to actually do it, which is not a genetics problem.

*The reviewer's point is well-taken. Indeed, genetic risk prediction is in its infancy, with only 1.5-3% of the MDD liability currently explained by PRS. Thus, at this current stage, the outlook regarding future benefits is not clearly convincing. As the reviewer suggested, we have highlighted a few use cases in the Discussion, Pages 21-22:*

*“MDD is a very heterogeneous disorder and genetic research will likely benefit from the study of more homogeneous disorder subtypes. For example, research on patients with*

*atypical features of increased appetite or weight suggested that this patient subgroup might benefit from treatments specifically targeting immunometabolic pathways (PMID: 29049554).*

*[...] Whilst the application of MDD PRS to predict individual risk is currently limited by its low discriminative accuracy, in the future PRS might be used for risk stratification at the population level (Andlauer and Nöthen, 2020). This is supported by recent evidence that MDD PRS of MDD – well-established to predict prevalent depression – also predict incident cases in the general population (PMID: 30698613). For other complex disorders like coronary artery disease and type-2 diabetes, PRS-based risk stratification is already powerful (PMID: 30104762).”*

#### **Reviewer #2:**

Kendall and colleagues have submitted a very broad overview of the genetics of depression including sections on twin studies, candidate gene studies, rare variants, GWAS, PRS, and GxE. If the goal is to provide a cursory review, then they have done it by providing text on a wide range of topics. MDD/MD genetics is quite nuanced and more attention could be made on effects of phenotyping (depression symptoms, versus clinical interviews), subtypes (PPD, atypical depression), and population genetics (severe under-representation of diverse pops, transferability & precision medicine), and pharmacogenetics (cytochrome P450 enzyme genes & antidepressant drug metabolism). Could the authors also outline how this review significantly expands on previous reviews on the genetics of depression (Flint & Kendler 2014, Smoller 2016, McIntosh 2019, Ormel 2019, Cai Choi et al 2020)?

*We thank the reviewer for this suggestion.*

*First, we have extensively revised the definitions in light of the comments and suggestions by the editor and Reviewer 1. As outlined above, we have now specifically defined and adjusted the use of the terms ‘MDD’ and ‘major depression’.*

*We agree that phenotyping, subtypes, population genetics, and pharmacogenetics are important topics. In fact, we have reviewed some of these topics wherever relevant. For example, we have summarized the literature on MDD subtypes related to the topics of Genetic Epidemiology, Rare Genetic Variants, GWAS, and Genetic Overlap; studies on population genetics are presented in the GWAS section.*

*We further added the following sentence about phenotyping on Page 5 (depressive symptoms vs clinical interview):*

*“Apart from such a diagnosis-based definition of depression, the severity of depressive symptoms is often studied as a quantitative trait, so the full variation of depressive symptoms can be taken into account.”*

*And about pharmacogenetics on Page 22:*

*“In addition, this review focuses on genetic factors influencing the risk or severity of depression. We did not discuss how genetic variants affect response to treatment, especially the metabolization of antidepressant drugs via cytochrome P450 genes. This important topic was recently reviewed by Bousman and colleagues (PMID: 33147643).”*

*We also thank the reviewer for the question regarding how our review expands on previous reviews on the genetics of MDD. The intention of our review was not to provide an entirely novel perspective compared to previous literature. Instead, we aimed to provide a comprehensive, up-to-date overview of the current knowledge in the field, from the perspective of the PGC. To this end, we focused on, in our opinion, the key topics related to MDD genetics. Our review could be considered as a primer, aiming to answer the question “What is MDD and*

*what is known about its genetics?” As Reviewer 1 commented, this review “... would be a good place for a non-specialist to find an overview and a perspective of research to date, and for researchers to find references on the topics that are covered here.”*

A few minor comments:

\*Abstract - consider editing from (a heritability) to 'heritability ranges from... and is moderate' (not low)

*The sentence was modified in the Abstract (Page 3):*

*“[...] with heritability ranges from 30% to 50%. [...], moderate heritability, ...”*

*and in the Discussion (Page 20):*

*“The heritability ranges between 30-50%. [...] Compared to other psychiatric disorders, its high prevalence, moderate heritability, and strong polygenicity have posed major challenges for gene-mapping in MDD.”*

\*Definition - consider editing the name of 'somatic domain' - are sleep features considered somatic?

*On page 4, 'somatic domain' has been replaced by 'vegetative symptoms', a terminology often used in this context.*

\*Consider adding the range of prevalences & factors that impact it (ex: country) in the Epidemiology section

*We have added the range of prevalence across countries to the Epidemiology section (Page 5):*

*“The 12-month prevalence of MDD is estimated to be ~6% ..., although these figures vary considerably across countries (e.g., the 12-month prevalence ranged from 2.2% in Japan to 10.4% in Brazil).”*

*A range of factors including age and gender are likely to contribute to the cross-country differences (Bracke, et al. 2020; PMID: 32980172). These factors are also presented in the Epidemiology section on Pages 5-6.*

\*Error in CONVERGE citation - ultra rare variant burden scores WERE associated with MD (PMID: 28002544)

*We thank the reviewer for bringing to our attention that this point was not entirely clear. We changed the text as follows on Page 10:*

*“However, the largest WGS study of MDD to date, by the CONVERGE consortium (5,303 cases and 5,337 controls), failed to find evidence for the association of individual rare or low-frequency variants with MDD (CONVERGE consortium, 2015). Using the same data set, it was subsequently shown that individuals with MDD had a significantly higher overall burden of singleton exonic variants predicted to be deleterious (Peterson et al., 2017).”*

\*Consider adding a few sentences on other genetic approaches like methylation, telomere, mitochondrial genome

*We thank the reviewer for this suggestion. We added a brief paragraph on Page 22:*

*“MDD is a multifactorial disorder. Other levels of genetic variation, such as DNA methylation (Barbu et al., 2020; Van Assche et al., 2017b) and subsequent gene-expression differences (Arloth et al., 2015; Schiweck et al., 2020), and also the telomere length (Lin et al., 2016), influence depression risk and severity. Importantly, epigenetic mechanisms serve as an interface between the individual and its environment (Provencal and Binder, 2015). A better understanding of the role of epigenetic variation and an improved integration of multi-omics data may enhance our understanding of the aetiology and treatment of MDD.”*

\*Consider expanding discussion on what is needed for the field to move forward - areas in need of further research & implications

*We also thank the reviewer for the suggestion. We extended the relevant text on Page 23:*

*“[The field would benefit from further exploration of the stratification properties of genetics – both for the understanding of its aetiology, including the role of environmental factors, and for the development of targeted treatment opportunities.] Rich datasets with high-quality and deep phenotyping are crucial for studying heterogeneous disorders like MDD. Combining deep, harmonized phenotypes with molecular data will greatly aid the genetic stratification of MDD.”*