Archival Report

Biological Psychiatry

The Australian Genetics of Depression Study: New Risk Loci and Dissecting Heterogeneity Between Subtypes

Brittany L. Mitchell, Adrian I. Campos, David C. Whiteman, Catherine M. Olsen, Scott D. Gordon, Adam J. Walker, Olivia M. Dean, Michael Berk, Ian B. Hickie, Sarah E. Medland, Naomi R. Wray, Nicholas G. Martin, and Enda M. Byrne

ABSTRACT

BACKGROUND: Major depressive disorder (MDD) is a common and highly heterogeneous psychiatric disorder, but little is known about the genetic characterization of this heterogeneity. Understanding the genetic etiology of MDD can be challenging because large sample sizes are needed for gene discovery—often achieved with a trade-off in the depth of phenotyping.

METHODS: The Australian Genetics of Depression Study is the largest stand-alone depression cohort with both genetic data and in-depth phenotyping and comprises a total of 15,792 participants of European ancestry, 92% of whom met diagnostic criteria for MDD. We leveraged the unique nature of this cohort to conduct a meta-analysis with the largest publicly available depression genome-wide association study to date and subsequently used polygenic scores to investigate genetic heterogeneity across various clinical subtypes of MDD.

RESULTS: We increased the number of known genome-wide significant variants associated with depression from 103 to 126 and found evidence of association of novel genes implicated in neuronal development. We found that a polygenic score for depression explained 5.7% of variance in MDD liability in our sample. Finally, we found strong support for genetic heterogeneity in depression with differential associations of multiple psychiatric and comorbid traits with age of onset, longitudinal course, and various subtypes of MDD.

CONCLUSIONS: Until now, this degree of detailed phenotyping in such a large sample of MDD cases has not been possible. Along with the discovery of novel loci, we provide support for differential pathways to illness models that recognize the overlap with other common psychiatric disorders as well as pathophysiological differences.

https://doi.org/10.1016/j.biopsych.2021.10.021

Major depressive disorder (MDD) is the most common psychiatric illness and the leading cause of disability worldwide. Recent large-scale genome-wide association studies (GWASs) have seen major breakthroughs in our understanding of the genetic etiology of psychiatric disorders, including MDD (1–4). However, the identified risk loci explain only a small fraction of the overall heritability of MDD, reflecting the small effect sizes of common genetic variants and its underlying heterogeneity as a diagnostic construct. Thus, there is a need for even larger studies to further elucidate the biological pathways contributing to depression risk.

To date, achieving the sample sizes required to detect genome-wide significant loci for depression has required pooling of data from many studies around the world. Consequently, there have been large differences between studies in terms of populations studied and the specific criteria used for assessing depression case status. Furthermore, information such as age of onset, individual symptoms, longitudinal course, or response to treatment are not often available, thus limiting the ability to evaluate intrinsic heterogeneity and the relative contribution of genetics thereto.

The extent to which MDD reflects one or multiple underlying disorders with differences in age of onset, symptom profiles, longitudinal course, comorbidities, or response to treatment is heavily debated (5,6). A number of subtypes of MDD have been proposed based on differences in presentation, including symptom-based subtypes such as atypical depression—characterized by neurovegetative symptoms such as hypersomnia, increased appetite, and weight gain—and timing-based subtypes such as early- and late-onset depression and seasonal affective disorder (SAD) (7). Previous studies utilizing polygenic scores (PGSs) support a role for genetic variation in differences between subtypes, especially age of onset. PGSs for MDD are enriched in those with earlier onset (1,8,9), and others have found evidence of enrichment for PGSs for schizophrenia (SCZ), bipolar disorder (BIP) (10), and attention-deficit/hyperactivity disorder (ADHD) in those with earlier onset (11). However, as psychiatric disorders are genetically correlated with one another, it is not clear to what extent the association between each PGS is capturing a unique component of risk that is not already captured by the

SEE COMMENTARY ON PAGE 177

ISSN: 0006-3223

MDD-PGS. The MDD-PGS is also increased in those with recurrent compared with single-episode depression (1,8). Similar differences in PGSs have been found for atypical depression compared to typical depression. Genetic risk factors for body mass index (BMI) and C-reactive protein are differentially associated with atypical depression (12,13), and an association between the SCZ-PGS and typical depression has been reported (12). Likewise, a tendency for seasonal changes in mood and behavior, the key symptom of SAD, has been linked to polygenic risk of SCZ (14).

Another key source of heterogeneity in MDD is comorbidity with anxiety disorders. Approximately 60% of individuals with MDD also meet criteria for a comorbid anxiety disorder, and the disorders share symptoms (15). Both twin and molecular genetic studies suggest a high degree of genetic correlation between MDD and anxiety disorders (4,16,17). However, the extent to which there are differences in genetic variation between MDD with and without comorbid anxiety is not yet known.

This report first aims to identify potential novel genetic loci for depression by conducting a GWAS of lifetime MDD (including only participants meeting the DSM-5 criteria for MDD as cases) and meta-analyzing the results with the largest publicly available depression GWAS meta-analysis. Second, we conducted a series of analyses using PGSs from 10 traits previously shown to be genetically correlated with MDD to investigate differences in genetic risk between subtypes of depression according to age of onset, recurrence, comorbid anxiety, atypical features, and SAD. This degree of detailed clinical phenotyping in such a large sample of MDD cases has not been previously possible.

METHODS AND MATERIALS

The Australian Genetics of Depression Study

The recruitment and sample characteristics of the Australian Genetics of Depression Study (AGDS) have been described in detail elsewhere (18). Briefly, more than 21,000 participants were recruited (15,792 currently genotyped) through a dual recruitment approach whereby participants were recruited either through Australian government prescription records or through a media campaign. Once enrolled in the study, participants completed online questionnaires comprising an obligatory core module that assessed MDD diagnosis as well as a range of satellite modules assessing a variety of phenotypes. Participants were assessed on DSM-5 criteria using the Composite International Diagnostic Interview short form diagnostic questionnaire (19).

Participants meeting DSM-5 criteria for MDD at some point in their lifetime (lifetime MDD) and who did not report a diagnosis of SCZ, BIP, or ADHD (n = 13,104) were assigned to a number of subtypes: early/late onset, recurrent depression, atypical depression, MDD with comorbid anxiety, and SAD. Demographic information for the AGDS cohort is reported in Table 1. Criteria used to define subtypes are described in Supplemental Methods in Supplement 1 and Table S1 in Supplement 2.

Clinical Trial Cohort

To maximize the sample size for GWASs, we also included 214 individuals (66% female; mean age = 51.3, SD = 12.5, range = 21–79; mean BMI = 27.7, SD = 5.6) from a mental health cohort provided by Deakin University's IMPACT Institute. All individuals in this cohort met DSM–IV criteria for MDD (20). Of these individuals, 73% exhibited moderate to severe symptoms at baseline (score ≥ 25 on the Montgomery–Åsberg Depression Rating Scale) (21). All blood samples included in this cohort were collected from individuals who had provided informed consent for unspecified further use of their samples. As participants in this cohort did not complete the AGDS questionnaire, subtypes could not be determined, and therefore, for simplicity, these samples were excluded from PGS analyses.

Table 1. Sample Size and Mean Value for Demographic and Depression Variables for Individuals Who Met MDD Criteria in the AGDS Cohort

	Sample Size	Mean (SD) [Range]
Age, Years	13,104	44.1 (15.1) [18–90]
Females	9836	41.4 (14.8) [18–90]
Males	3268	49 (14.9) [18–89]
BMI	13,104	28.5 (7.2) [17.1–58.7]
Age of Onset, Years	12,538	22.5 (11.4) [7.2–79]
Number of Episodes	12,512	7.9 (4.5) [1–13+]
Subtypes[a]		
Early/late onset	10,203	–
Recurrent depression	11,014	–
Atypical depression	8032	–
MDD with comorbid anxiety	2851	–
Seasonal affective disorder	761	–

AGDS, Australian Genetics of Depression Study; BMI, body mass index; MDD, major depressive disorder.

[a]Subtype sample sizes reflect number of cases that meet criteria. Subtypes were not mutually exclusive.

Control Cohort

As our control sample, we used a large volunteer community sample from Queensland, Australia, that was invited for study participation through random draw from the electoral roll (QSkin cohort) (22). Participants completed a lifestyle questionnaire, which included a disease checklist about previous diagnoses of psychiatric disorders. QSkin study participants who reported never having been diagnosed with any psychiatric disorders were included as control subjects (n = 12,684; 51% female; mean age = 61.3, SD = 7.9, range = 38–87.8; mean BMI = 27, SD = 0.04). All protocols and questionnaires for both the AGDS and QSkin cohorts were approved by the QIMR Berghofer Medical Research Institute Human Research Ethics Committee.

Participants from all three cohorts were genotyped using the Illumina Global Screening Array v2. Quality control of genotype data was conducted as described in Supplemental Methods in Supplement 1 before imputation using the Haplotype Reference Consortium 1.1 reference panel.

Genome-wide Association Analysis and Genetic Correlations

Our primary GWAS consisted of only participants who met the DSM-5 criteria for MDD and had not participated in previous studies that contributed to the Psychiatric Genomics Consortium (PGC). This led to a final sample size of 13,104 cases

from AGDS and 214 cases from the Deakin University samples (total N = 13,318 cases, 12,684 controls). The GWAS analysis was conducted using the SAIGE v. 0.39 (23) with the first 10 ancestry principal components as covariates.

Although common in genetic studies of MDD (24), screened controls excluding those with any psychiatric disorder can bias heritability estimates and genetic correlations. Additionally, as previously reported (18), a bias with regard to the educational attainment (EA) of volunteer participants in the AGDS and QSkin cohorts was observed, resulting in the over-representation of participants with a higher EA in the AGDS. Therefore, the GWAS analysis was repeated with 1) controls screened only for depression rather than all psychiatric disorders (n = 13,696) and 2) the highest level of education achieved included as a covariate, to evaluate their influence on the results.

Results were filtered on a minor allele frequency $>$ 0.01 and R^2 imputation quality metric $>$ 0.6. The proportion of liability variance for MDD explained by common single nucleotide polymorphisms (SNPs) was estimated using GREML (Supplemental Methods in Supplement 1) (25). Genome-wide genetic correlations between the AGDS GWAS and traits of interest were assessed using bivariate linkage disequilibrium (LD) score regression (26) and LD Hub (27).

Meta-analysis

We conducted an inverse-variance weighted meta-analysis of depression using the unadjusted AGDS GWAS with results from the largest published depression GWAS from the PGC [hereafter referred to as Howard *et al.* (2)] using METAL. Annotation of the meta-analysis results is described in Supplemental Methods in Supplement 1.

Polygenic Scores

We generated PGSs for depression using the Howard *et al.* summary statistics (including 23andMe) (2). PGSs for a further nine traits that have been shown to have high genetic correlation with MDD were generated (28). These include four psychiatric disorders using the most recent published studies of SCZ (29), BIP (30), ADHD (31) and anxiety (16); BMI (32), which has been previously found to be associated with atypical depression; the personality trait neuroticism associated with depression; and insomnia, which is a key symptom difference between typical and atypical depression. Finally, we evaluated PGSs for EA (33) and socioeconomic status (SES) as measured by the Townsend Deprivation Index (TDI) in UK Biobank (34) (Table S2 in Supplement 2). Recent studies have highlighted that a considerable proportion of the genetic variance shared between psychiatric disorders is also shared with markers of SES such as EA and TDI (35). This shared heritability complicates biological interpretation of studies evaluating genetic heterogeneity, as it has been demonstrated that both the SNP-based heritability and genetic correlations of psychiatric disorders are reduced when accounting for shared heritability with SES and EA (36,37). We therefore sought to investigate the associations of PGSs for these SES-related traits and their influence on the associations of PGSs from the other traits. To avoid bias due to potential sample overlap, PGSs for MDD and EA were constructed using leaving-one-out summary statistics. All GWASs used to construct the PGSs were obtained from European populations.

PGSs were constructed using SBayesR, a Bayesian method that assumes that SNP effects are drawn from a mixture of four zero-mean normal distributions with different variances (38) (Supplemental Methods in Supplement 1).

The association between each of the PGSs with MDD in AGDS/QSkin and subtypes of MDD within AGDS cases was estimated using logistic regression with age at study enrollment, sex, and 10 principal components included as covariates. Initially, the marginal effect of each PGS was estimated, and then all of the PGSs were fitted together in a multiple regression to investigate their associations when adjusting for the other PGSs. Only unrelated individuals were included in the PGS analyses. A total of 10 PGSs were tested for association with five subtypes of depression. A Bonferroni-corrected threshold of .05/(10 $\times$ 5) = .001 was used to establish statistical significance of associations.

RESULTS

Genome-wide Association Analysis

After combining with the clinical trial sample, a total of 13,318 cases and 12,684 controls were included in the GWAS for MDD. The genomic inflation factor (λ) was 1.08, and the LD-score intercept was 1.025 (SE = 0.008), indicating little evidence for residual population stratification (Figure S1 in Supplement 1). The most significant hit was located in the RNA binding protein fox-1 homolog 1 gene (*RBFOX1*) located on chromosome 16 (rs113726301:G:A; odds ratio = 1.14; p = 3.6 $\times$ 10^{-8}) (Figure S2 in Supplement 1), which has been robustly implicated in depression, neurodevelopment, and aggression in previous studies (39–41). The potential role of *RBFOX1* as a candidate gene for MDD is discussed in detail in the Supplementary Note of Wray *et al.* (1).

To assess the impact that EA differences between our cases and controls may have on the association results and downstream analyses, we repeated the GWAS including EA as a covariate and found that our association signals, including those of known depression-associated variants, were significantly attenuated (rs113726301:G:A; odds ratio = 1.13; p = 1.03 $\times$ 10^{-6}) (Table 2; Figure S3 in Supplement 1). This is likely driven by the fact that our volunteer depression cases had higher EA levels than our electoral roll control sample (18). There is no evidence of association of the *RBFOX1* variant with EA in the largest EA meta-analysis (33) (rs113726301:G:A; β = 0.002; p = .14).

SNP-Based Heritability and Genetic Correlation Estimates

Assuming a population lifetime prevalence of 0.15 for MDD, the estimated SNP-based heritability was 0.24 (SE = 0.01) using GREML and 0.25 (SE = 0.03) using LD score regression. This is similar to the estimates for MDD in UK Biobank and larger than estimates from minimal phenotyping approaches and from the PGC meta-analysis (h^2_{SNP} = 0.05) (5).

The genetic correlations between AGDS and previous depression GWASs were highly significant, with an r_g = 0.88, p = 6.6 $\times$ 10^{-67} with Howard (2) and r_g = 0.92, p = 2.4 $\times$ 10^{-27}

Table 2. Comparison of Effect Sizes and Significance Values for the Top Three Independent SNPs Identified in AGDS GWAS With and Without Educational Attainment as a Covariate

SNP	CHR	Frequency	β	SE	*p* Value	β_{covar}	SE_{covar}	*p* $Value_{covar}$
rs111429920	16	0.23	0.13	0.02	2.9×10^{-8}	0.12	0.03	1.2×10^{-6}
rs2904659	8	0.37	0.11	0.02	7.5×10^{-8}	0.11	0.02	1.6×10^{-7}
rs188753797	5	0.02	−0.48	0.09	1.5×10^{-7}	−0.40	0.10	8.5×10^{-5}

AGDS, Australian Genetics of Depression Study; CHR, chromosome; covar, covariate; GWAS, genome-wide association study; SNP, single nucleotide polymorphism.

with Wray (1). We further investigated the genetic correlations with other relevant traits using LD Hub. Consistent with previous studies, MDD showed strong positive genetic correlations with depressive symptoms and neuroticism and significant genetic correlations with BIP and SCZ (Figure 1). The genetic correlations with the EA-adjusted results were largely similar to those observed with our unadjusted results for all non–cognitive-related traits (Figure 1). SNP-based heritability and genetic correlation estimates were not significantly different when including controls were screened only for depression rather than all psychiatric disorders (Figure S4 in Supplement 1).

Novel Loci Discovered for MDD

Meta-analysis of the largest publicly available GWAS of depression (2) and AGDS was based on 7,617,771 variants with matching alleles in common between the two datasets. After conditional and joint analysis, the number of independent genome-wide significant variants increased from 103 to 126 (Figure 2; Table S3 in Supplement 2).

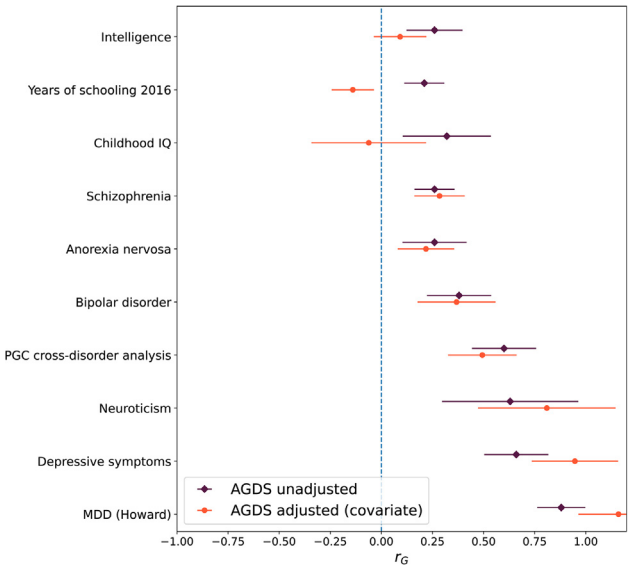

Figure 1. Genetic correlations between MDD and phenotypically relevant traits estimated from genome-wide association study summary statistics using linkage disequilibrium score regression. AGDS genome-wide association study results from models with and without a covariate for educational attainment. AGDS, Australian Genetics of Depression Study; MDD, major depressive disorder; PGC, Psychiatric Genomics Consortium. r_G, genetic correlation.

We found 232 genes associated with MDD using a gene-based test conducted in MAGMA (Table S4 in Supplement 2). Of these, 43 genes were uniquely identified in our meta-analysis results (as compared with those in Howard). We found one novel gene-set putatively associated with MDD in our meta-analysis: GO_synaptic_membrane, $p = 5.5 \times 10^{-7}$.

Polygenic Scores

Predicting MDD Within the AGDS. Nine of the 10 PGSs were significantly higher in cases than controls, with the exception being BMI. The MDD-PGS had the largest effect with an MDD odds ratio of 1.74 per SD of PGS (Table 3). The variance explained on the liability scale by the MDD-PGS (assuming a population prevalence of 0.15) was 5.7% (SE = 0.3%). The odds of those in the highest decile of PGS having MDD compared with the lowest decile was 7.3 (95% CI 6.3–8.4) (Figure 3A). We observed no significant differences in risk between male and female participants, despite all levels of depressive disorders, including MDD, being twice as prevalent in females. The anxiety-PGS was the next best predictor of MDD, with an odds ratio of 1.28 (95% CI 1.2–1.4) per SD of PGS. The EA-PGS was also associated with MDD such that a propensity for higher educational achievement is associated with increased risk of MDD. However, this presumably reflects differences in the ascertainment of cases and controls for our study (18).

When all of the PGSs were fitted together in a multiple logistic regression model, the effect sizes of the PGSs for the nine significant traits were attenuated but remained significantly associated with the MDD case status (Table 3). The variance explained on the liability scale by all the PGSs jointly was 7.8% (SE = 0.3%).

PGS by Age of Onset

Earlier age of onset was associated with PGSs for MDD ($p = 2.0 \times 10^{-4}$) and ADHD ($p = 8.7 \times 10^{-4}$) (Figures 3B and 4). When all of the PGSs were fitted in the same model, the SCZ-PGS and EA-PGS were significantly associated with earlier age of onset after multiple testing correction (Table S5 in Supplement 2).

Using age of onset of 30 years as the cutoff for early onset is somewhat arbitrary. However, we obtained similar results when using onset prior to age 18 as the cutoff, with the associations between the SCZ-PGS and TDI-PGS being statistically significant (Table S6 in Supplement 2). To further investigate the relationship between PGS for psychiatric disorders and age of onset, ordered binned categories of age of onset were generated, and the mean PGS *z* score (using the entire sample of cases and controls to estimate the mean and SD) in each

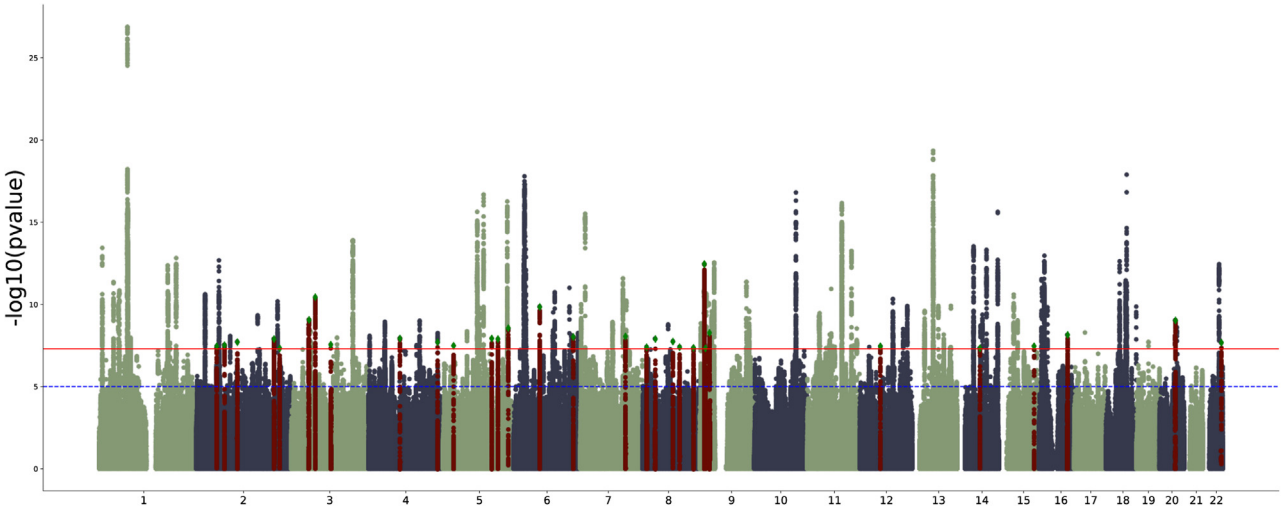

Figure 2. Manhattan plot depicting associations from a meta-analysis of Australian Genetics of Depression Study with Howard *et al.* (2); 126 independent loci reaching genome-wide significance of which 23 are novel (highlighted in red).

category calculated and plotted (Figure 3B). Sample sizes per category are shown in Table S7 in Supplement 2.

We found that the mean MDD-PGS was highest in those with the earliest age of onset (<10 years) with a steady decline in every 5-year bin until ages 25–29 when it plateaus before dropping in the 60+ category. Until age 60, the mean MDD-PGS was above the mean of the entire sample. However, for SCZ and ADHD, the mean PGS declined with later onset, and cases with an age of onset after 35 have lower PGS than the population average. By contrast, BIP-PGS remained above the population mean for all ages of onset <60 (11).

Longitudinal Course of MDD

There was an approximately linear relationship between the number of reported episodes and the MDD-PGS (Figure 3C; Table S8 in Supplement 2), indicating an association with more recurrent disorders. We compared the PGSs for all 10 traits in those reporting one or two episodes (n = 1498) and those with more than two episodes (n = 11,348). Recurrence was significantly associated with the MDD, ADHD, neuroticism, and TDI PGSs (Figure 4). When all of the PGSs were fitted together, the MDD ($p = 1.2 \times 10^{-4}$) and TDI ($p = 1.2 \times 10^{-4}$) PGSs showed independent significant associations.

MDD Clinical Subtypes in AGDS

We investigated for evidence of genetic heterogeneity between clinical subtypes of MDD including depression with comorbid anxiety, atypical depression, and SAD (Figure 4; Table S9 in Supplement 2). The MDD-PGS was associated with increased likelihood of having a comorbid anxiety disorder ($p = 4.2 \times 10^{-11}$) but not with atypical depression (p = .45) or SAD (p = .23). It was notable that when fitting all of the PGSs in the same model, there was no evidence of an effect of the

Table 3. PGS Prediction of MDD in AGDS and QSkin Cohorts Individually (Marginal) and When All Included in a Joint Model (Multiple Regression)

	Marginal			Multiple Regression		
PGS	β	SE	p Value	β	SE	p Value
MDD	0.56	0.02	1.5×10^{-227}	0.46	0.02	1.3×10^{-190}
Anxiety	0.25	0.02	1.1×10^{-52}	0.06	0.01	7.4×10^{-6}
SCZ	0.25	0.02	9.5×10^{-28}	0.05	0.02	2.3×10^{-3}
BIP	0.21	0.02	7.9×10^{-38}	0.07	0.01	1.0×10^{-6}
ADHD	0.10	0.02	4.5×10^{-10}	0.06	0.01	6.0×10^{-5}
Neuroticism	0.12	0.02	2.6×10^{-32}	0.05	0.02	1.7×10^{-3}
Insomnia	0.13	0.02	2.4×10^{-16}	0.04	0.01	2.3×10^{-3}
BMI	0.01	0.02	8.0×10^{-1}	0.02	0.01	1.1×10^{-1}
EA	0.14	0.02	1.7×10^{-16}	0.19	0.01	1.1×10^{-41}
TDI	0.11	0.02	2.4×10^{-10}	0.07	0.02	9.1×10^{-6}

Results from PGS prediction of MDD in AGDS and QSkin cohorts show that MDD-PGS has the largest predictive effect. The joint model indicates that many of these comorbid traits capture variability over-and-above that of the MDD-PGS.

ADHD, attention-deficit/hyperactivity disorder; AGDS, Australian Genetics of Depression Study; BIP, bipolar disorder; BMI, body mass index; EA, educational attainment; MDD, major depressive disorder; PGS, polygenic score; SCZ, schizophrenia; TDI, Townsend Deprivation Index.

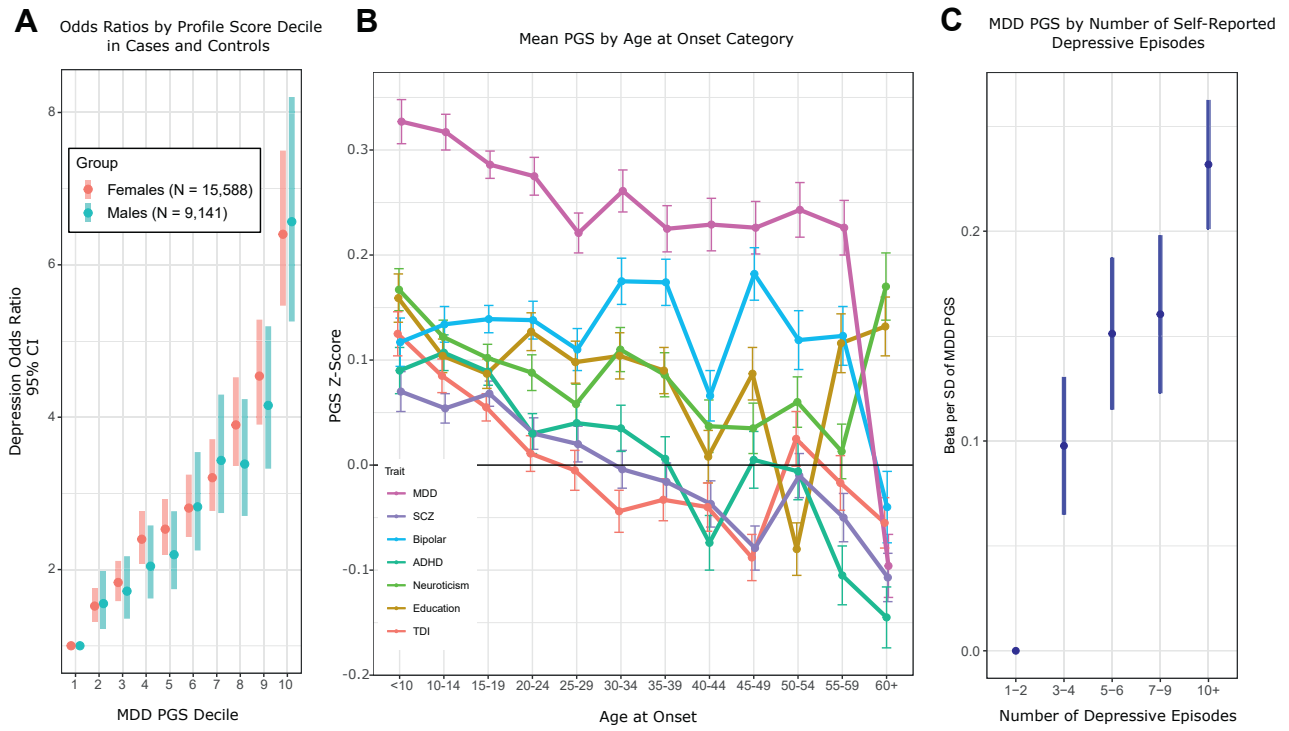

Figure 3. **(A)** MDD case odds within PGS decile relative to MDD case odds of the first PGS decile in the Australian Genetics of Depression Study. PGSs were calculated from the Howard MDD genome-wide association study using SBayesR. **(B)** Mean PGS *z* score per trait per age of onset of MDD category. Mean line indicates average PGS per age of onset category for the entire sample. **(C)** Association of MDD-PGS and number of reported depressive episodes in Australian Genetics of Depression Study. Bars indicate 95% CI. ADHD, attention-deficit/hyperactivity disorder; MDD, major depressive disorder; PGS, polygenic score; SCZ, schizophrenia; TDI, Townsend Deprivation Index.

anxiety-PGS on anxious depression (p = .36). Given the high genetic correlation between anxiety and depression, the current anxiety-PGS may not explain more variance in anxiety than was already captured by the MDD-PGS. Comorbid anxiety was also associated with higher SCZ-PGS ($p = 6.5 \times 10^{-4}$) and lower EA-PGS ($p = 2.6 \times 10^{-8}$).

In agreement with previous studies (42), the SCZ-PGS was associated with increased likelihood of typical depression ($p = 7.5 \times 10^{-5}$). The ADHD- and TDI-PGS were associated with atypical depression but were only marginally significant when all PGSs were included simultaneously (p = .03 and p = .02, respectively). This is consistent with the observation that atypical depression is characterized by vulnerability to weight gain, and, as has been previously found by others, the BMI-PGS is nominally associated with atypical depression (p = .03). However, when including all the other PGSs, the association with BMI was not significant. No significant findings were found with SAD.

DISCUSSION

We used the AGDS, one of the largest stand-alone depression cohorts with both genetic data and in-depth phenotyping, to conduct a GWAS of lifetime MDD and to investigate the role of genetic variation in heterogeneity in MDD. The MDD GWAS results showed very high genetic correlation with other previous MDD GWAS studies. Both phenotypic and genetic data indicate that the AGDS sample is enriched for severe, recurrent MDD. The estimated SNP-based heritability was 0.24, larger than estimates from minimal phenotyping in UK Biobank, suggesting that the sample has great utility for identifying more specific genetic associations with MDD (24).

We found one significant locus in the *RBFOX1* gene. This result is notable, given that previous GWASs of similar sample size have failed to produce any genome-wide significant hits, and the identification of a known depression-locus provides additional reassurance regarding the saliency of AGDS phenotyping (33,39–41,43,44). A meta-analysis between the AGDS GWAS and the most recently published GWAS of depression from the PGC (2) increased the number of genome-wide significant variants from 103 to 126, a marked increase in the number of known depression loci. Among the novel associations are an intronic variant in *TENM2*, which encodes teneurin, a protein known to play a key role in neuronal guidance during development (45), and an intronic variant in *NRG1* (neuregulin 1), a gene that has been implicated in risk of SCZ (46) and is associated with chronotype (47). The identification of 23 new loci in the meta-analysis underscores the value of AGDS to future GWAS efforts to identify loci associated with MDD and its subtypes.

We further utilized the deep phenotyping information to investigate heterogeneity in MDD. Our analyses indicated that the increased risk attributable to the MDD-PGS is the same in men and women. As the discovery GWAS includes both men

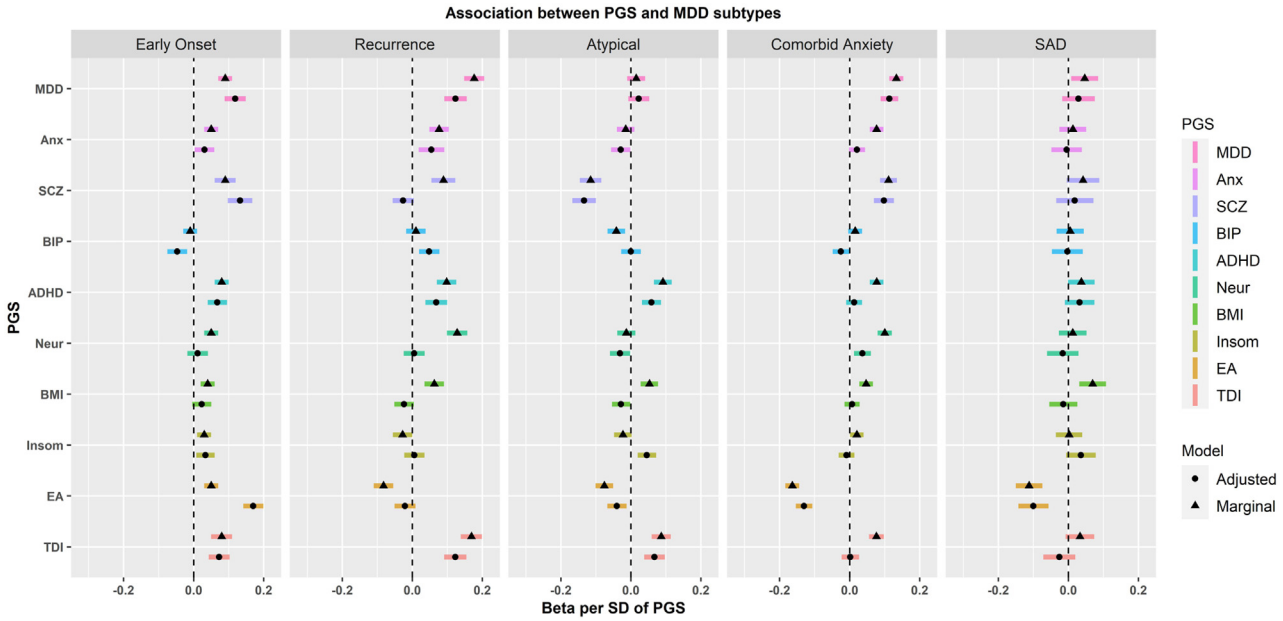

Figure 4. PGS association for MDD and nine related traits with five MDD subtypes previously defined in the literature. Results shown are of each PGS with each subtype individually (marginal) and when all of the PGSs were fitted together in a multiple logistic regression model (adjusted). Bars indicate 95% confidence intervals. ADHD, attention-deficit/hyperactivity disorder; Anx, anxiety; BIP, bipolar disorder; BMI, body mass index; EA, educational attainment; Insom, insomnia; MDD, major depressive disorder; Neur, neuroticism; PGS, polygenic score; SAD, seasonal affective disorder; SCZ, schizophrenia; TDI, Townsend Deprivation Index.

and women, it is possible that sex-specific effects are masked, and future sex-specific GWASs will reveal differences in genetic risk factors between the sexes. The MDD-PGS was associated with an earlier age of onset and scaled linearly with the number of reported episodes, indicating that those carrying more depression risk alleles are likely to report recurrent depression. These findings indicate that genotyping early-onset, recurrent cases will be most useful for gene-finding efforts.

Participants with earlier age of onset also had higher SCZ-, ADHD-, EA- and TDI-PGS. While these results further highlight a likely neurodevelopmental pathway to MDD (48) with onset early in life, they also highlight the role of social deprivation in influencing age of onset. Further study of the shared genetic component of risk between MDD and other neurodevelopmental disorders (e.g., SCZ, ADHD, autism) may yield further insights into the genetic etiology of early-onset depression. Furthermore, generating risk scores from multiple disorders will improve prediction of those individuals most at risk of any major mood or psychotic disorders early in life. An accurate identification of those individuals at greatest risk has major implications for the development and deployment of prevention and early intervention initiatives (49,50).

The association of early onset with increased EA-PGS is surprising, given that EA has been found to be protective for MDD (1). However, this is likely due to an ascertainment bias in the sample. Older participants had higher PGS for EA, implying that among older participants, those who are well-educated were more likely to participate than others in their age group (Figure S4 in Supplement 1). While many report an early age of onset, those with higher EA-PGS also report longer time since the most recent episode and better current mental health after adjusting for age, indicating that those with higher EA-PGS were more likely to have had depression early in life but to have been in remission longer. The PGS results are also reinforced by analysis of self-reported education in the sample. Earlier onset was significantly more frequently reported by those with a tertiary qualification than by those with secondary or below accounting for the effects of age and sex (odds ratio = 1.29, $p < 2.2 \times 10^{-16}$), providing further support that those with earlier onset were more likely to enroll in the study if they have higher levels of education.

The importance of SES and gene-environment correlation in MDD is reinforced by the finding of a positive association of the TDI-PGS with recurrence. However, even after controlling for the EA- and TDI-PGS, recurrence is associated with PGSs for MDD and ADHD, indicating that patients with recurrent depression carry a higher genetic load for multiple disorders, not all of which can be attributed to the effects of SES.

The findings of differential PGS associations across various clinical subtypes reinforce the hypothesis of considerable genetic heterogeneity in pathways to MDD. Of particular note was the association of the ADHD-PGS with four of the five clinical subtypes including a nominal association with atypical depression, a finding supported by a recent study in UK Biobank showing higher genetic correlation of atypical depression with ADHD (51). This suggests a strong neurodevelopmental aspect and one related to difficulty in establishing regular daily activity, sleep-wake, and other circadian rhythms. These findings provide preliminary support for differential pathways to illness models that recognize both the overlap with other common major mood

or psychotic disorders, and the likely pathophysiological differences. This has implications for future gene-mapping efforts for the identification of differential neurobiological pathways to MDD and, hence, for the potential development of new more personalized therapies for MDD.

Owing to the previous unavailability of large samples with deep phenotyping, it has, until now, been difficult to study genetic sources of heterogeneity in MDD (52). While AGDS provides an ideal resource for investigating this heterogeneity, the findings should be interpreted considering some potential limitations. First, while our results suggest that the sample is enriched for recurrent depression, it is possible that individuals with the most severe cases were unable to enroll in the study. Second, we were unable to assess whether each one of the participant's reported episodes met the DSM-5 criteria. Third, the absence of associations between various clinical MDD subtypes and certain polygenic scores may still reflect insufficient power to detect these effects owing to differences in sample sizes between the subtypes and of the discovery GWASs from which the PGSs are derived. Finally, our control cohort was screened on a single medical history question and not DSM criteria.

ACKNOWLEDGMENTS AND DISCLOSURES

The Australian Genetics of Depression Study (AGDS) was primarily funded by the National Health and Medical Research Council (NHMRC) of Australia Grant No. 1086683, and the QSkin study was funded by the NHMRC (Grant Nos. 1073898, 1058522 and 1123248). This study was further supported by NHMRC Grant Nos. 1145645, 1078901, 1113400 and 1087889 and National Institute of Mental Health; NHMRC R.D. Wright Career Development Fellowship (1145634 [to OMD]); NHMRC Senior Principal Research Fellowship (1059660 and 1156072 [to MB]); NHMRC investigator Grant Nos. 1172917, 1173790 and 1172990 (to SEM, NRW, and NGM); NHMRC Senior Principal Research Fellowship (1155413 [to DCW]); and Trisno Family Research Fellowship (to AJW).

We are indebted to all of the participants for giving their time to contribute to this study. We wish to thank all the people who helped in the conception, implementation, media campaign and data cleaning. We thank Richard Parker, Simone Cross, and Lenore Sullivan for their valuable work coordinating all the administrative and operational aspects of the AGDS project. We would also like to thank the research participants and employees of 23andMe for making this work possible.

The genome-wide summary statistics for the Howard *et al.* analysis of 23andMe, Inc., data were obtained under a data transfer agreement with QIMR Berghofer. To access the summary statistics for the meta-analysis of AGDS and the Howard *et al.* summary statistics (including 23andMe), a data transfer agreement is required from 23andMe (dataset-request@23andMe.com) before submitting a request to the authors.

IBH was a Commissioner in Australia's National Mental Health Commission from 2012 to 2018. He is the Co-Director, Health and Policy at the Brain and Mind Centre University of Sydney. The Brain and Mind Centre operates an early-intervention youth services at Camperdown under contract to headspace. IBH has previously led community-based and pharmaceutical industry-supported (Wyeth, Eli Lily, Servier, Pfizer, AstraZeneca) projects focused on the identification and better management of anxiety and depression. He has led investigator-initiated studies into the antidepressant and linked circadian impacts of agomelatine. These studies have been supported by Servier, the manufacturer of agomelatine. He is the Chief Scientific Advisor to, and an equity shareholder in, InnoWell. InnoWell has been formed by the University of Sydney and PwC to pursue the transformation of mental health services through the use of innovative technologies. All other authors report no biomedical financial interests or potential conflicts of interest.

ARTICLE INFORMATION

From QIMR Berghofer Medical Research Institute (BLM, AIC, DCW, CMO, SDG, SEM, NGM); School of Biomedical Sciences (BLM), Faculty of Health, Queensland University of Technology; School of Biomedical Sciences (AIC), Faculty of Medicine, The University of Queensland; Child Health Research Centre (EMB), The University of Queensland; Institute for Molecular Bioscience (AIC, NRW, EMB), The University of Queensland; and Queensland Brain Institute (NRW), The University of Queensland, Brisbane; Institute for Mental and Physical Health and Clinical Translation (AJW, OMD, MB), School of Medicine, Barwon Health, Deakin University, Geelong; Florey Institute for Neuroscience and Mental Health and the Department of Psychiatry (OMD, MB), The University of Melbourne, Melbourne; Orygen (MB), The National Centre of Excellence in Youth Mental Health, Centre for Youth Mental Health; and Brain and Mind Centre (IBH), The University of Sydney, Sydney, New South Wales, Australia.

Address correspondence to Enda M. Byrne, Ph.D., at Enda.byrne@uq.edu.au, or Brittany L. Mitchell, M.Sc., at Brittany.mitchell@qimrberghofer.edu.au.

Received Apr 9, 2021; revised Sep 21, 2021; accepted Oct 24, 2021.

Supplementary material cited in this article is available online at https://doi.org/10.1016/j.biopsych.2021.10.021.

REFERENCES

1. Wray NR, Ripke S, Mattheisen M, Trzaskowski M, Byrne EM, Abdellaoui A, *et al.* (2018): Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. Nat Genet 50:668–681.
2. Howard DM, Adams MJ, Clarke TK, Hafferty JD, Gibson J, Shirali M, *et al.* (2019): Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. Nat Neurosci 22:343–352.
3. CONVERGE consortium (2015): Sparse whole-genome sequencing identifies two loci for major depressive disorder. Nature 523:588–591.
4. Levey DF, Stein MB, Wendt FR, Pathak GA, Zhou H, Aslan M, *et al.* (2021): Bi-ancestral depression GWAS in the Million Veteran Program and meta-analysis in >1.2 million individuals highlight new therapeutic directions. Nat Neurosci 24:954–963.
5. Cai N, Revez JA, Adams MJ, Andlauer TFM, Breen G, Byrne EM, *et al.* (2020): Minimal phenotyping yields genome-wide association signals of low specificity for major depression. Nat Genet 52:437–447.
6. Cai N, Choi KW, Fried EI (2020): Reviewing the genetics of heterogeneity in depression: Operationalizations, manifestations and etiologies. Hum Mol Genet 29:R10–R18.
7. Harald B, Gordon P (2012): Meta-review of depressive subtyping models. J Affect Disord 139:126–140.
8. Musliner KL, Agerbo E, Vilhjálmsson BJ, Albiñana C, Als TD, Østergaard SD, Mortensen PB (2021): Polygenic liability and recurrence of depression in patients with first-onset depression treated in hospital-based settings. JAMA Psychiatry 78:792–795.
9. Halldorsdottir T, Piechaczek C, Soares de Matos AP, Czamara D, Pehl V, Wagenbuechler P, *et al.* (2019): Polygenic risk: Predicting depression outcomes in clinical and epidemiological cohorts of youths. Am J Psychiatry 176:615–625.
10. Musliner KL, Mortensen PB, McGrath JJ, Suppli NP, Hougaard DM, Bybjerg-Grauholm J, *et al.* (2019): Association of polygenic liabilities for major depression, bipolar disorder, and schizophrenia with risk for depression in the Danish population. JAMA Psychiatry 76:516–525.
11. Rice F, Riglin L, Thapar AK, Heron J, Anney R, O'Donovan MC, Thapar A (2019): Characterizing developmental trajectories and the role of neuropsychiatric genetic risk variants in early-onset depression. JAMA Psychiatry 76:306–313.
12. Milaneschi Y, Lamers F, Peyrot WJ, Baune BT, Breen G, Dehghan A, *et al.* (2017): Genetic association of major depression with atypical features and obesity-related immunometabolic dysregulations. JAMA Psychiatry 74:1214–1225.
13. Badini I, Coleman JRI, Hagenaars SP, Hotopf M, Breen G, Lewis CM, Fabbri C (2022): Depression with atypical neurovegetative symptoms

shares genetic predisposition with immuno-metabolic traits and alcohol consumption. Psychol Med 52:726–736.
14. Byrne EM, Psychiatric Genetics Consortium Major Depressive Disorder Working Group, Raheja UK, Stephens SH, Heath AC, Madden PA, *et al.* (2015): Seasonality shows evidence for polygenic architecture and genetic correlation with schizophrenia and bipolar disorder. J Clin Psychiatry 76:128–134.
15. Pollack MH (2005): Comorbid anxiety and depression. J Clin Psychiatry 66(suppl 8):22–29.
16. Purves KL, Coleman JRI, Meier SM, Rayner C, Davis KAS, Cheesman R, *et al.* (2020): A major role for common genetic variation in anxiety disorders. Mol Psychiatry 25:3292–3303.
17. Kendler KS, Neale MC, Kessler RC, Heath AC, Eaves LJ (1992): Major depression and generalized anxiety disorder. Same genes, (partly) different environments? Arch Gen Psychiatry 49:716–722.
18. Byrne EM, Kirk KM, Medland SE, McGrath JJ, Colodro-Conde L, Parker R, *et al.* (2020): Cohort profile: The Australian genetics of depression study. BMJ, (Open) 10:e032580.
19. Kessler RC, Andrews G, Mroczek D, Ustun B, Wittchen HU (1998): The World Health Organization composite international diagnostic interview short-form (CIDI-SF). Int J Method Psychiat Res 7:171–185.
20. Sheehan DV, Lecrubier Y, Sheehan KH, Amorim P, Janavs J, Weiller E, *et al.* (1998): The Mini-International Neuropsychiatric Interview (M.I.N.I.): The development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. J Clin Psychiatry 59(suppl 20):22–33.. quiz 34–57.
21. Montgomery SA, Asberg M (1979): A new depression scale designed to be sensitive to change. Br J Psychiatry 134:382–389.
22. Olsen CM, Green AC, Neale RE, Webb PM, Cicero RA, Jackman LM, *et al.* (2012): Cohort profile: The QSkin Sun and Health Study. Int J Epidemiol 41:929–929i.
23. Zhou W, Nielsen JB, Fritsche LG, Dey R, Gabrielsen ME, Wolford BN, *et al.* (2018): Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies. Nat Genet 50:1335–1341.
24. Penninx BW, Beekman AT, Smit JH, Zitman FG, Nolen WA, Spinhoven P, *et al.* (2008): The Netherlands Study of Depression and Anxiety (NESDA): Rationale, objectives and methods. Int J Methods Psychiatr Res 17:121–140.
25. Yang J, Lee SH, Goddard ME, Visscher PM (2011): GCTA: A tool for genome-wide complex trait analysis. Am J Hum Genet 88:76–82.
26. Bulik-Sullivan BK, Loh PR, Finucane HK, Ripke S, Yang J, Schizophrenia Working Group of the Psychiatric Genomics Consortium, *et al.* (2015): LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. Nat Genet 47:291–295.
27. Zheng J, Erzurumluoglu AM, Elsworth BL, Kemp JP, Howe L, Haycock PC, *et al.* (2017): LD Hub: A centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. Bioinformatics 33:272–279.
28. Brainstorm Consortium, Anttila V, Bulik-Sullivan B, Finucane HK, Walters RK, Bras J, *et al.* (2018): Analysis of shared heritability in common disorders of the brain. Science 360:eaap8757.
29. Pardiñas AF, Holmans P, Pocklington AJ, Escott-Price V, Ripke S, Carrera N, *et al.* (2018): Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. Nat Genet 50:381–389.
30. Stahl EA, Breen G, Forstner AJ, McQuillin A, Ripke S, Trubetskoy V, *et al.* (2019): Genome-wide association study identifies 30 loci associated with bipolar disorder. Nat Genet 51:793–803.
31. Demontis D, Walters RK, Martin J, Mattheisen M, Als TD, Agerbo E, *et al.* (2019): Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. Nat Genet 51:63–75.
32. Locke AE, Kahali B, Berndt SI, Justice AE, Pers TH, Day FR, *et al.* (2015): Genetic studies of body mass index yield new insights for obesity biology. Nature 518:197–206.
33. Lee JJ, Wedow R, Okbay A, Kong E, Maghzian O, Zacher M, *et al.* (2018): Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. Nat Genet 50:1112–1121.
34. Jiang L, Zheng Z, Qi T, Kemper KE, Wray NR, Visscher PM, Yang J (2019): A resource-efficient tool for mixed model association analysis of large-scale data. Nat Genet 51:1749–1755.
35. Zhang Y, Qi G, Park JH, Chatterjee N (2018): Estimation of complex effect-size distributions using summary-level statistics from genome-wide association studies across 32 complex traits. Nat Genet 50:1318–1326.
36. Marees AT, Smit DJA, Abdellaoui A, Nivard MG, van den Brink W, Denys D, *et al.* (2021): Genetic correlates of socio-economic status influence the pattern of shared heritability across mental health traits. Nat Hum Behav 5:1065–1073.
37. Wendt FR, Pathak GA, Lencz T, Krystal JH, Gelernter J, Polimanti R (2021): Multivariate genome-wide analysis of education, socioeconomic status and brain phenome. Nat Hum Behav 5:482–496.
38. Lloyd-Jones LR, Zeng J, Sidorenko J, Yengo L, Moser G, Kemper KE, *et al.* (2019): Improved polygenic prediction by Bayesian multiple regression on summary statistics. Nat Commun 10:5086.
39. Fogel BL, Wexler E, Wahnich A, Friedrich T, Vijayendran C, Gao F, *et al.* (2012): RBFOX1 regulates both splicing and transcriptional networks in human neuronal development. Hum Mol Genet 21:4171–4186.
40. Zhao WW (2013): Intragenic deletion of RBFOX1 associated with neurodevelopmental/neuropsychiatric disorders and possibly other clinical presentations. Mol Cytogenet 6:26.
41. Davies MN, Verdi S, Burri A, Trzaskowski M, Lee M, Hettema JM, *et al.* (2015): Generalised anxiety disorder–A twin study of genetic architecture, genome-wide association and differential gene expression. PLoS One 10:e0134865.
42. Milaneschi Y, Lamers F, Peyrot WJ, Abdellaoui A, Willemsen G, Hottenga JJ, *et al.* (2016): Polygenic dissection of major depression clinical heterogeneity. Mol Psychiatry 21:516–522.
43. Fernàndez-Castillo N, Gan G, van Donkelaar MMJ, Vaht M, Weber H, Retz W, *et al.* (2020): RBFOX1, encoding a splicing regulator, is a candidate gene for aggressive behavior. Eur Neuropsychopharmacol 30:44–55.
44. Gehman LT, Stoilov P, Maguire J, Damianov A, Lin CH, Shiue L, *et al.* (2011): The splicing regulator Rbfox1 (A2BP1) controls neuronal excitation in the mammalian brain. Nat Genet 43:706–711.
45. DePew AT, Aimino MA, Mosca TJ (2019): The tenets of teneurin: Conserved mechanisms regulate diverse developmental processes in the Drosophila nervous system. Front Neurosci 13:27.
46. Mei L, Xiong WC (2008): Neuregulin 1 in neural development, synaptic plasticity and schizophrenia. Nat Rev Neurosci 9:437–452.
47. Tanizawa K, Chin K (2018): Genetic factors in sleep-disordered breathing. Respir Investig 56:111–119.
48. Gałecki P, Talarowska M (2018): Neurodevelopmental theory of depression. Prog Neuropsychopharmacol Biol Psychiatry 80:267–272.
49. Murray GK, Lin T, Austin J, McGrath JJ, Hickie IB, Wray NR (2021): Could polygenic risk scores be useful in psychiatry?: A review. JAMA Psychiatry 78:210–219.
50. Hoare E, Thorp A, Bartholomeusz-Raymond N, McCoy A, Berk M (2020): Opportunities in the Australian national education initiative for promoting mental health in schools. Lancet Child Adolesc Health 4(1):11–13.
51. Nguyen TD, Harder A, Xiong Y, Kowalec K, Hägg S, Cai N, *et al.* (2022): Genetic heterogeneity and subtypes of major depression. Mol Psychiatry 27:1667–1675.
52. Mitchell BL, Thorp JG, Wu Y, Campos I, *et al.* (2021): Polygenic risk scores derived from varying definitions of depression and risk of depression. JAMA Psychiatry 78:1152–1160.