Investigating Genetic Heterogeneity in Major Depression Through Item-level Genetic Analyses of the PHQ-9

Jackson G. Thorp[a], Andries T. Marees[a,b], Jue-Sheng Ong[c], Jiyuan An[c], Stuart MacGregor[c], Eske M. Derks[a,d]

[a] Translational Neurogenomics Laboratory, QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia

[b] Amsterdam UMC, University of Amsterdam, Department of Psychiatry, Amsterdam, The Netherlands

[c] Statistical Genetics Laboratory, QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia

[d] School of Biomedical Sciences, Faculty of Health, Queensland University of Technology, Brisbane, Queensland, Australia

Correspondence:

Jackson Thorp
Translational Neurogenomics Laboratory
300 Herston Road
Brisbane, QLD 4006
Jackson.Thorp@qimrberghofer.edu.au

Eske Derks
Translational Neurogenomics Laboratory
300 Herston Road
Brisbane, QLD 4006
Eske.Derks@qimrberghofer.edu.au

Running head:

GENETIC HETEROGENEITY IN MAJOR DEPRESSION

Abstract

Background: Major Depressive Disorder (MDD) is a clinically heterogeneous disorder. Previous large-scale genetic studies of MDD have explored genetic risk factors of MDD case-control status or aggregated sums of depressive symptoms, ignoring possible clinical or genetic heterogeneity.

Aim: In this study, we present the results of symptom-level genetic analyses and compare SNP-based heritability (h^2_{SNP}) and genetic correlations across major depression symptoms. We further investigate genetic correlations with a range of psychiatric disorders and other associated traits.

Methods: We have analysed data from the UK biobank and included 148,752 subjects of white British ancestry with genotype data who completed nine items of a self-rated measure of depression: the Patient Health Questionnaire (PHQ-9). Genome-Wide Association analyses were conducted for nine symptoms and two composite measures. LD score regression analysis was used to calculate SNP-based heritability (h^2_{SNP}) and genetic correlations (r_g) across symptoms and to investigate genetic correlations with 25 external phenotypes. Confirmatory factor analyses were applied to test whether one, two, or three-factor models best fit the pattern of genetic correlations across the nine symptoms.

Results: We identified 9 novel genome-wide significant genomic loci, with no overlap in loci across depression symptoms. h^2_{SNP} ranged from 3% (suicidal ideation) to 11% (fatigue). Genetic correlations range from 0.54 to 0.96 (all $p < 1.39\times10^{-3}$) with 30 of 36 correlations being significantly smaller than 1. A 3-factor model provided the best fit to the genetic correlation matrix, with factors representing “psychological”, “neurovegetative”, and “psychomotor / concentration” symptoms. The genetic correlations with external phenotypes showed large variation across the nine symptoms.

Discussion: Patterns of h^2_{SNP} and genetic correlations differed across the nine symptoms of depression. Our findings suggest that the large phenotypic heterogeneity observed for MDD is recapitulated at a genetic level. Future studies should investigate how genetic heterogeneity in MDD influences the efficacy of clinical interventions.

Introduction

Clinical depression is a markedly complex and debilitating mental disorder characterised by sad, irritable or empty mood, diminished pleasure, and cognitive and somatic impairment[1]. The heritability of major depressive disorder (MDD) is estimated to be ~37% from twin studies[2] with common Single Nucleotide Polymorphisms (SNPs) explaining around 9% of the variation in liability[3]. MDD has substantial comorbidity with other psychiatric and substance use disorders and is related to a wide range of personality, socioeconomic, and human traits[4]. There is substantial overlap in the genetic risk factors of MDD and other psychiatric disorders[3], including significant genetic correlations (r_g) with schizophrenia ($r_g = 0.34$), bipolar disorder ($r_g = 0.32$), autism spectrum disorders ($r_g = 0.44$) and ADHD ($r_g = 0.42$). MDD has notably high genetic overlap with anxiety disorders ($r_g = 0.80$) and neuroticism ($r_g = 0.70$), which may reflect the overlap in diagnostic criteria between the three traits. Initial efforts to identify genetic variants associated with major depression were unsuccessful, despite successes with other psychiatric diseases and traits. While a Genome Wide Association Study (GWAS) of schizophrenia (9,394 cases), for example, detected seven genome-wide significant associations [5], a mega-analysis of MDD (9240 cases)[6] and a meta-analysis of depressive symptoms (N = 34,549)[7] found no significant associations. By 2014, 108 independent genetic loci for schizophrenia had been identified[8], and not a single one for depression. The struggle to identify significant genetic variants was likely related to low statistical power due to the clinical heterogeneity of MDD[9].

Depression is a polygenic disorder, influenced by the combination of small effects from many genetic variants which can only be detected in studies with large sample sizes[10]. Due to the relatively high prevalence of depression (~15% vs. <1% for schizophrenia), power is lower than for other diseases with similar numbers of cases but lower prevalence[11]. Also, depression is less

heritable than other psychiatric disorders (~37% vs. ~80% for schizophrenia[12]) and therefore larger sample sizes are required to obtain similar statistical power to detect significant effects. In the last two yours, increasing sample size has proved to be effective with the number of genome-wide significant variants increasing steadily with sample size. Hyde, et al. [13] identified 15 genome-wide significant loci associated with self-reported depression (N = 307,354). Another 17 loci were identified across three broad depression phenotypes (N = 322,580)[14]. The largest GWAS of major depression to date (N = 480,359) identified 44 significant loci[3].

These genetic studies ignored possible clinical heterogeneity in MDD, despite clinical presentations and symptoms of MDD being diverse. The Diagnostic and Statistical Manual of Mental Disorders 5th edition (DSM-5) defines major depression by the following symptoms: (1) depressed mood, (2) diminished interest or pleasure in activities (anhedonia), (3) decrease or increase in weight or appetite, (4) insomnia or hypersomnia, (5) psychomotor agitation or retardation, (6) fatigue or loss of energy, (7) feelings of worthlessness or excessive or inappropriate guilt, (8) diminished ability to think or concentrate, or indecisiveness, and (9) recurrent thoughts of death or recurrent suicidal ideation[15]. For a diagnosis of MDD five or more of these symptoms need to be present during a two week period, with at least one symptom being depressed mood or anhedonia. Østergaard, et al. [16] highlighted that there are 227 possible combinations of symptoms meeting DSM-5 criteria, indicating MDD is an extremely heterogeneous disorder. Further, individual symptoms have been found to differ substantially in their association with psychosocial impairment, influence from environmental and personality risk factors, and biological correlates[17]. GWASs of depression have typically focused on MDD case-control status or aggregated sums of depressive symptoms. By combining different symptoms into a single clinical measure, it is implicitly assumed that individual symptoms of

depression are genetically similar. However, the extreme heterogeneity of depression and numerous clinical presentations of the disorder suggest that different biological mechanisms could underlie the diverse subtypes of depression. Supporting this notion, depression symptoms have been found to differ substantially in heritability (h^2 range, 0 – 35%); with somatic and cognitive symptoms being most heritable[18]. Further, the diagnostic criteria of MDD were found to reflect three underlying genetic factors (cognitive / psychomotor symptoms, mood symptoms, and neurovegetative symptoms) rather than a single factor of genetic risk in a twin study[19]. Nagel, et al. [20] found substantial genetic heterogeneity in neuroticism, a personality trait with extensive phenotypic and genetic overlap with MDD[21], by conducting genetic analyses on the individual items used to measure neuroticism.

To date, it is not known to what extent genetic risk factors overlap in individual symptoms of MDD. The aim of the present study is to examine and assess the extent of genetic heterogeneity in major depression. We conduct genetic analyses on individual symptoms of depression in 148,752 participants within the UK Biobank, as measured by the nine items of the Patient Health Questionnaire (PHQ-9)[22], a depression measure which directly maps onto the DSM-5 criteria. In order to examine genetic heterogeneity in depression we (1) conduct symptom-level GWA analyses and then compare genetic associations and SNP-based heritability across symptoms; (2) calculate phenotypic and genetic correlations across depression symptoms and determine their underlying genetic factor structure; and (3) calculate genetic correlations between individual symptoms and a range of psychiatric disorders and human complex traits.

Methods

UK Biobank Cohort

UK Biobank (UKBB) is a major health data resource containing phenotypic information on a wide range of health-related measures and characteristics in over 500,000 participants from the United Kingdom general population[23]. Participants were recruited between 2006 and 2010 and provided written informed consent. A total of 157,365 participants completed the PHQ-9, as part of a UKBB mental health follow-up questionnaire administered online in 2016.

Sample selection

First, participants were included in the present study if they were of white British ancestry, identified through self-reported ethnicity and genetic principal components. Participants who self-reported as not white British, but for whom the first two genetic principal components indicated them to be genetically similar to those of white British ancestry were also included in order to maximise sample size (these commonly were participants who reported to be of Irish ancestry). Second, Participants were excluded if they were identified with schizophrenia and / or other psychotic disorders, bipolar disorder, cyclothymic disorder, or dissociative identity disorder, based on self-reported symptoms or diagnosis, reported prescription of an antipsychotic medication, and/or ICD-10 (The International Classification of Diseases, Tenth Revision) codes from linked hospital admission records. Third, only participants who provided a response for all nine items of the PHQ-9 were included (list-wise deletion represented a less than 2% reduction in sample size). This resulted in a final sample size of 148,752 (see Supplementary Figure 13 for flow diagram of sample selection).

PHQ-9

The PHQ-9 is a commonly used self-administered measure of depression containing nine items that map directly onto the nine DSM diagnostic criteria for major depression[22]. Each PHQ-9 item assesses the frequency of that symptom over the past two weeks, rated on a four-point ordinal scale: (0) Not at all, (1) Several days, (2) More than half the days, (3) Nearly every day. (See Supplementary Table 1 for the nine symptoms of major depression, PHQ-9 items, and DSM-5 diagnostic criteria).

The PHQ-9 is a psychometrically valid and reliable measure of depression[24]. Test-retest reliability was high (r = .84, over a span of 48 hours) and internal consistency was excellent with Cronbach's alphas (α) of .89 and .86 in primary care and obstetrics-gynaecology samples, respectively. The authors also reported good criterion and construct validity. The PHQ-9 was validated against professional diagnoses of MDD, resulting in 88% sensitivity and 88% specificity (at a PHQ-9 sum-score of $\geq$ 10); and scores correlated highly with similar constructs, such as the 20-item Short-Form General Health Survey (SF-20)[25] mental health scale (r = .73). Internal consistency of the PHQ-9 in the UK Biobank sample in the current study was high (Cronbach's α = .83).

Depression Item Phenotypes

Each of the nine PHQ-9 items is considered a separate phenotype in the genetic analyses. The ordinal scale of measurement of these items complicates interpretation of the SNP-based heritability estimates (amount of phenotypic variance in the item explained by SNPs). SNP-based heritability is an important concept in genetics, essential to understanding the magnitude of the genetic influence on a particular trait[26]. To enable a direct comparison across each of the PHQ-9 items, each ordinal phenotype was transformed to a binary phenotype for heritability

estimation. The nine items were dichotomised such that an item was considered to be endorsed if the item score was one or greater (several days, more than half the days, or nearly every day), and not endorsed if the score was zero (not at all). A cut-off score of one was used in order to maximise the number of subjects who endorsed an item and hence statistical power, a strategy that has provided greater benefit in GWASs of depression over ensuring a seamless phenotype[3,11,14,27]. In addition to the nine ordinal items and nine binary items, a *sum-score* (sum of all ordinal item scores; ranging from 0 to 27) and *binary sum-score* (number of binary items endorsed; ranging from 0 to 9) were included as phenotypes. We will present the results from the binary items and the two sum-scores while results for ordinal items are provided in supplementary.

Genome-Wide Association Analyses

A total of 20 GWA analyses were conducted (nine ordinal scale depression items, nine binary items, plus the sum-score and binary sum-score phenotypes) using BOLT-LMM[28]. Associations between SNPs and a phenotype are tested using a linear mixed model in order to correct for population structure and cryptic relatedness. While BOLT-LMM is based on a quantitative trait model, it can be used to analyse binary traits by treating them as continuous and applying a transformation. Ordinal items are treated as continuous. An issue when analysing binary traits in BOLT-LMM is the inflated type 1 error rates for rare SNPs when the number of cases and controls are very unbalanced[29]. In practice, all of the traits we consider here have a case proportion which is large enough (3%) for this not to be a problem[30].

Analyses were limited to autosomal SNPs with high imputation quality score (INFO score ≥ 0.80) and a minor allele frequency of 1% or higher, resulting in 9,413,637 SNPs being tested for association. Sex and age were included as covariates. GWAS results were annotated

using the FUMA GWAS platform[31]. The conventional genome-wide significance threshold of $p < 5\times10^{-8}$ was applied. Due to the exploratory nature of the analyses and the high correlation between the 20 phenotypes, we decided not to correct for multiple testing of the 20 phenotypes as this would lead to increased type-II error rate and reduced power.

Significant SNPs were clumped into blocks high in linkage disequilibrium (the non-random association of alleles at a specific locus; LD) using a threshold of $r^2 < 0.10$ (correlation between allele frequencies of two SNPs; as calculated by PLINK). Independent significant SNPs were defined as the SNP with the lowest p-value within an LD block. Genomic risk loci (distinct, fixed positions on a chromosome) were identified by merging independent SNPs if $r^2 \geq 0.10$ and their LD blocks are physically close to each other at a distance of 1,000 kb.

LDSC analyses

Estimates of the variance in each phenotype attributable to the additive effects of all SNPs (SNP-based heritability; h^2_{SNP}) were calculated via single-trait LD Score Regression using GWAS summary statistics from our analyses[32] (see Supplementary methods). In order to interpret h^2_{SNP} for binary items estimates are converted to a normally distributed liability scale, because liability scale heritability is independent of prevalence and can be compared across different phenotypes and populations[33]. The population prevalence of PHQ-9 items was estimated from our UK Biobank sample (population prevalence = sample prevalence; see Table 1). We applied a Bonferroni corrected significance threshold for the 11 h^2_{SNP} estimates ($p < 4.55\times10^{-3}$).

Cross-trait LD Score Regression[34] was used to estimate genetic correlations (rg) between each of the nine binary items. We applied a Bonferroni corrected significance threshold for these 36 r_g tests ($p < 1.39\times10^{-3}$). Additionally, we also calculated pairwise genetic correlations

between our phenotypes (9 depression items and sum-scores) and 25 other psychiatric, substance use, socioeconomic and human traits with publicly available GWAS summary statistics (see Supplementary Table 2). Multiple testing was corrected for by adjusting p values based on false discovery rate (FDR) across all tests.

Hierarchical Cluster Analysis

A hierarchical cluster analysis was conducted to examine the underlying genetic structure between depression items. Implemented in the hclust function in R[35], items are grouped into similar clusters based on a measure of dissimilarity between each pair of items and the results are presented in a cluster dendrogram. Dissimilarity was defined as one minus the genetic correlation ($1 - r_g$).

Confirmatory Factor Analyses

Confirmatory factor analyses (CFA) were conducted based on genetic covariances between items, in order to quantitatively assess the genetic factor structure of the PHQ-9 identified in the cluster analysis. The fit of a one-factor baseline model and two and three-factor models identified in the cluster analysis were compared.

χ^2 likelihood ratio tests are very sensitive to large samples and often produce spurious positive results[36]. Given the very large sample size in the present study, model fit was evaluated with a range of alternative fit indices. These indices (and their commonly used thresholds for acceptable model fit) include: NFI ($\geq .95$), AGFI ($\geq .95$), RMSEA ($\leq .06$), and SRMR ($\leq .06$)[37]. Models were compared using AIC and BIC indices, which take into account both model fit and complexity. The most parsimonious model is the model with the lowest AIC and BIC values.

Results

Descriptive Statistics

The final sample (N = 148,752) was 56% female, ranging in age from 38 to 72 years old (M = 55.93, SD = 7.73). The distribution of responses to all PHQ-9 items (on the ordinal scale) are displayed in Supplementary Table 3. The distribution of item scores varied considerably across items; sleep problems and fatigue had the highest endorsement rates while suicidal ideation and psychomotor changes had the lowest rates. Sum-scores ranged from 0 to 27, with a mean of 2.71 (SD = 3.61). Endorsement rates of binary depression items are shown in Table 1. The number of symptoms endorsed ranged from zero to nine, with a mean of 2.02 (SD = 2.20).

GWA Analyses

Genome-wide association analyses of the 9 binary depression items plus sum-score phenotypes identified a total of 326 genome-wide significant SNPs ($p < 5\times10^{-8}$), tagged by 13 independent SNPs. Two lead SNPs were significant in more than one phenotype, such that across all phenotypes there are 11 unique, independent genome-wide significant SNPs. These SNPs mapped onto nine genomic risk loci (see Table 2 for results, Supplementary Figures 1-10 for QQ plots and Manhattan plots of all phenotypes; and Supplementary Table 4 for the ordinal item GWAS results).

Heritability Estimates

Estimates of the proportion of phenotypic variance in each item attributable to the additive effects of all SNPs (SNP-based heritability; h^2_{SNP}) varied considerably across the nine items (see Figure 1 and Supplementary Table 5). All estimates were significant after Bonferroni correction ($p < 4.55\times10^{-3}$). The amount of variance explained by common SNPs ranged from 3% of variance in suicidal ideation up to 11% of the variance in fatigue (mean h^2_{SNP} across the nine

depression items was 7%). h^2_{SNP} estimates for the sum-score and no. symptoms phenotypes were 6% and 7%, respectively.

Inter-item Phenotypic and Genetic Correlations

Spearman correlations between all pairs of PHQ-9 depression items showed that all items were positively correlated with each other phenotypically and remained significant after Bonferroni correction for 36 tests ($p < 1.39\times10^{-3}$). Coefficients ranged from .19 to .69, with the strongest association between anhedonia and depressed mood, the two core symptoms of MDD (see Figure 2).

Summary statistics from the GWASs of the nine binary items were used to calculate genetic correlations (r_g) between items. All correlations were significant after correcting for multiple testing ($p < 1.39\times10^{-3}$) and were in the same direction (see Figure 2). Estimated r_g's ranged from .54 (suicidal ideation / psychomotor changes; *s.e* = .15) to .96 (psychomotor changes / concentration problems; *s.e* = .11), with a mean r_g of .77. Thirty out of the 36 genetic correlations were significantly less than one (95% CI did not include one), indicating substantial genetic heterogeneity across the PHQ-9 items (partly unique genetic risk factors contribute to the majority of pairs of depressive symptoms; see Figure 2 and Supplementary Table 6). Some of the genetic correlations that were not significantly different from 1 were relatively low, but have large standard errors which explains their overlap with 1.

A very similar pattern of genetic correlations emerged for the ordinal items (r_g range: .55 to .96), such that the Pearson correlation between the set of binary item r_g's and ordinal item r_g's was high, $r = .90$, $p < .001$ (see Supplementary Figure 11).

The Pearson correlation between the genetic correlations and phenotypic correlations was moderate, $r = .48$, $p = .003$, suggesting phenotypic correlations do not map one to one with genetic correlations (see Supplementary Figure 12).

Genetic Clustering Analysis

A hierarchical clustering analysis based on genetic covariance between the nine depression items revealed two main genetic clusters: the first cluster including anhedonia, depressed mood, suicidal ideation, and low self-esteem (psychological symptoms); and the second cluster including psychomotor changes, concentration problems, fatigue, appetite change, and sleep problems (somatic symptoms; see Figure 3). Further exploration of the cluster dendrogram suggests the somatic symptoms cluster could again be split into two clusters: "neurovegetative" symptoms (fatigue, appetite change, and sleep problems); and "psychomotor / concentration" symptoms.

Confirmatory Factor Analyses

CFA of the genetic factor structure found that all three models provided good fit to the data (see Supplementary Table 7). Comparison of models based on AIC and BIC values found that the three-factor model was the most parsimonious model compared to the one-factor model and the two-factor model (substantially lower AIC and BIC values). These results suggest the PHQ-9 is reflected genetically by three factors, comprising "psychological", "neurovegetative", and "psychomotor / concentration" symptoms.

Genetic Correlations with External Traits

Genetic correlations of the nine depression items, sum-score and binary sum-score with 25 other psychiatric, substance use, socioeconomic and human traits are displayed in Figure 4.

299 Correlations significant after correcting for false discovery rate (FDR) are indicated by non-
300 white squares (see Supplementary Table 8).

301 Individual depression items correlated as expected with closely related traits, supporting
302 the validity of the individual symptom phenotypes in the present study. For example, appetite
303 change had a substantially stronger positive genetic correlation with body mass index (r_g = .61)
304 than the other eight depression symptoms (r_g's range between .10 to .29); and sleep problems
305 had a strong, positive correlation with insomnia (r_g = .71). All symptoms were negatively
306 correlated with subjective well-being (r_g range = -.54 to -.91), with suicidal ideation having the
307 strongest association. Furthermore, all items positively correlated (and showed a similar pattern)
308 with the other MDD and overall depression phenotypes.

309 Genetic overlap with other psychiatric disorders and traits differed substantially across
310 depression symptoms, such as with anxiety disorders (r_g range = .50 to .93), neuroticism (r_g
311 range = .49 to .85), schizophrenia (r_g range = .09 to .32), and insomnia (r_g range = .31 to .71).
312 Furthermore, bipolar disorder was significantly correlated with 4 out of 9 depression items only
313 (sleep problems, low self-esteem, concentration problems, and psychomotor changes). Anorexia
314 nervosa overlapped with just three items, with genetic correlations even being in different
315 directions (low self-esteem r_g = .28, psychomotor changes r_g = .27, and appetite change r_g = -.26).
316

Discussion

In the present study, we investigated genetic heterogeneity in major depression by conducting genetic analyses on individual symptoms of MDD in 148,752 participants from the UK Biobank. We identified nine genomic risk loci across the nine MDD symptoms and sum-score phenotypes, all have not been associated with major depression in previous GWASs[3,7,13,14,27,38-43]. Our results revealed substantial genetic heterogeneity in depression symptoms with no overlap in significant loci across PHQ items. Though we acknowledge that the lack of overlap may be due to low statistical power to detect all true associations, we highlight some notable examples where a specific symptom of depression is linked to a gene that was previously found to be associated with a strongly related phenotype. For the item "sleep problems", we found SNPs that implicate *PAX8* (based on proximity), a transcription factor related to thyroid follicular cell development and expression of thyroid-specific genes, replicating previous studies linking this gene to sleep duration[44-46]. In addition, SNPs associated with "depressed mood" influenced the expression of *KLHDC8B* (protein coding gene involved in cytokinesis). This gene has been previously linked to depressed affect, a sub-cluster of neuroticism that is strongly related to depression[47]. Neither of these genes were implicated in the largest GWASs of overall depression[3,14], illustrating the importance of exploring genetic associations for specific symptoms of depression.

SNP-based heritability analyses revealed that individual depression symptoms were differentially heritable (h^2_{SNP} ranging from 3 to 11%), suggesting that depression symptoms differ in their relative proportions of common SNP contributions. Notably, items within the "neurovegetative" symptom cluster were most highly heritable, consistent with a previous report

that found somatic symptoms (such as sleep problems and appetite changes) to have a stronger heritable basis [18].

Genetic correlations between depression symptoms ranged from moderate ($r_g < .60$) to high ($r_g > .90$), suggesting that while some symptoms have high genetic overlap, a substantial amount of genetic variation is not shared between symptoms. This indicates extensive genetic heterogeneity in major depression, in line with the finding that depression represents multiple dimensions of genetic risk[19] and previous associations between individual symptoms and specific polymorphisms[48].

The underlying genetic structure between symptoms was best explained by three genetic clusters. This suggests there are risk factors specific to clusters which could indicate underlying biology specific to either "neurovegetative", "psychological" or "psychomotor / concentration" symptoms of depression. This is consistent with symptoms differing in their biological correlates, with neurovegetative symptoms such as weight gain, increased appetite, and sleep problems being associated with higher levels of inflammation markers[49,50]. These clusters were not in full agreement with the three genetic factors found by Kendler, et al. [19] based on an analysis of twin data. As an example, in the Kendler study, suicidal ideation loaded onto the same factor as psychomotor changes and concentration problems, while we find that suicidal ideation clusters with symptoms of depressed mood, anhedonia, and low self-esteem that together form a "psychological symptoms" factor. However results are not easily comparable given that they derived factors from a twin study (and therefore captured rare genetic variants as well as common SNPs), used a subset of eight symptoms (appetite changes did not load onto any factor), and symptom phenotypes came from structured clinical interview rather than a self-report measure such as the PHQ-9.

Results from genetic correlations between items and a range of external traits lead us to note three general observations. First, genetic correlations with external traits differed substantially between symptoms providing evidence for genetic heterogeneity in major depression. In agreement with previous findings for major depression[3,14], all symptoms overlapped with anxiety, schizophrenia, ADHD, insomnia, neuroticism, and subjective well-being, however the proportion of overlap varied considerably across symptoms. For example, anxiety disorders had a substantially higher genetic correlation with "suicidal ideation" than the other items. This supports a strong phenotypic association, with over 70% of people with a history of suicide attempt having an anxiety disorder, compared to ~33% in the general population[51]. Second, some traits (such as bipolar disorder, cannabis lifetime use, cigarettes per day, and intelligence) were genetically correlated with a subset of items only. Bipolar disorder for example, was genetically correlated with only four items (low self-esteem, concentration, psychomotor changes, and sleep problems), suggesting that the moderate genetic overlap between bipolar and major depression[52] is predominately driven by these selected symptoms. This highlights how insight into the genetic architecture between traits can be gained from conducting symptom-level analyses. Third, we found traits that were genetically correlated with individual items, but not with the sum-score phenotypes. Anorexia nervosa did not overlap with aggregate measures of depression symptoms as operationalized in the sum-score phenotypes, in agreement with Howard, et al. [14] who similarly found no genetic overlap between anorexia and their three overall depression phenotypes. Yet, anorexia nervosa was genetically correlated with appetite change, low self-esteem, and psychomotor changes. Interestingly, the overlap with appetite change was in the opposite direction compared to the other two items. This finding

emphasises the importance of analysing individual symptoms of a disorder, as important information is ignored by relying on sum-scores or overall phenotypes.

Limitations

The findings and conclusions of this study should be interpreted in view of some key limitations. First, despite having the largest sample available to date, the current study is still underpowered to detect significant SNPs. Given the relatively high prevalence and phenotypic heterogeneity of depression, much larger sample sizes are needed compared to other psychiatric disorders[11]. To not reduce power further we did not correct for multiple testing (of 11 GWA analyses) and hence our GWAS results require independent replication. Second, depression items were analysed in isolation, regardless of the overall MDD status of the participant. For example, a participant could strongly endorse the symptom fatigue, yet have no other signs of depression and hence the endorsement of fatigue is unrelated to major depression. Nevertheless, it is possible that fatigue, regardless of the context it occurs in, possesses the same underlying genetic basis. Third, we used a PHQ-9 cut-off score of 1 to dichotomise items in order to maximise the number of cases and improve statistical power. A PHQ-9 item score of one does not meet the diagnostic criteria for endorsement, hence the phenotypes may represent a predisposition to rather than full endorsement of the particular symptom. Fourth, our results may be affected by ascertainment bias due to healthy volunteerism within the UKBB. As such our sample could represent a truncated version of the population's genetic distribution for symptoms (people on far end of liability scale may be less likely to participate), hence resulting in reduced number of cases for some symptoms or reduced variation between cases and controls.

Implications

The recent success in the discovery of genetic variants associated with depression has been driven by ever increasing sample sizes, an approach that has been favoured over reducing phenotypic heterogeneity. Consequently, GWASs have been conducted on a diverse range of depression-related phenotypes that often include a small subset of symptoms, generally with the view that the increase in sample size can overcome the lack of clinical precision. While this has indeed been proven to be effective at increasing the number of significant variants identified, our finding of symptom-level genetic heterogeneity raises questions about this approach. Using broad diagnostic phenotypes ignores the unique genetic factors associated with specific symptoms of depression that would likely provide useful information to further unravel the genetic architecture of MDD. Further, our finding of genetic heterogeneity across MDD symptoms implicates that patients with MDD show variation in disease pathogenesis. This variation may be linked to response to clinical interventions, such that patients presenting with specific symptom patterns (e.g., characterized primarily by neurovegetative symptoms) may be expected to respond differently.

Conclusion

Our results provide convincing evidence that major depression is a genetically heterogeneous disorder, and highlight the utility of analysing the genetics of individual items or symptoms of a psychiatric disorder. Insights into the genetic aetiology and underlying biology of MDD will be maximised by combining large-scale genetic studies of broad clinical definitions with follow-up studies of more refined phenotypic measures of specific diagnostic subtypes.

Table and Figure legends

Table 1: Sample sizes and prevalence of all binary PHQ-9 items.

Table 2: GWAS Results for binary PHQ-9 items and sum-score phenotypes.

Figure 1. SNP-based heritability estimates and 95% confidence intervals (95% CI) for the nine depression items and sum-score phenotypes.

Figure 2. Inter-item genetic and phenotypic correlations.

Figure 3. Cluster dendrogram of inter-item genetic correlations.

Figure 4. Genetic correlations between PHQ-9 items and a range of other complex traits (psychiatric, substance use, and socioeconomic phenotypes) based on publicly available summary statistics.

Tables

Table 1: Sample sizes and prevalence rates of all binary PHQ-9 items.

Item	N endorsed	N not endorsed	Prevalence %
Anhedonia	27,288	121,464	18.3
Depressed mood	32,263	116,489	21.7
Sleep problems	72,302	76,450	48.6
Fatigue	73,924	74,828	49.7
Appetite changes	26,757	121,995	18.0
Low self-esteem	28,206	120,546	19.0
Concentration problems	26,229	122,523	17.6
Psychomotor changes	7,914	140,838	5.3
Suicidal ideation	6,064	142,688	4.1

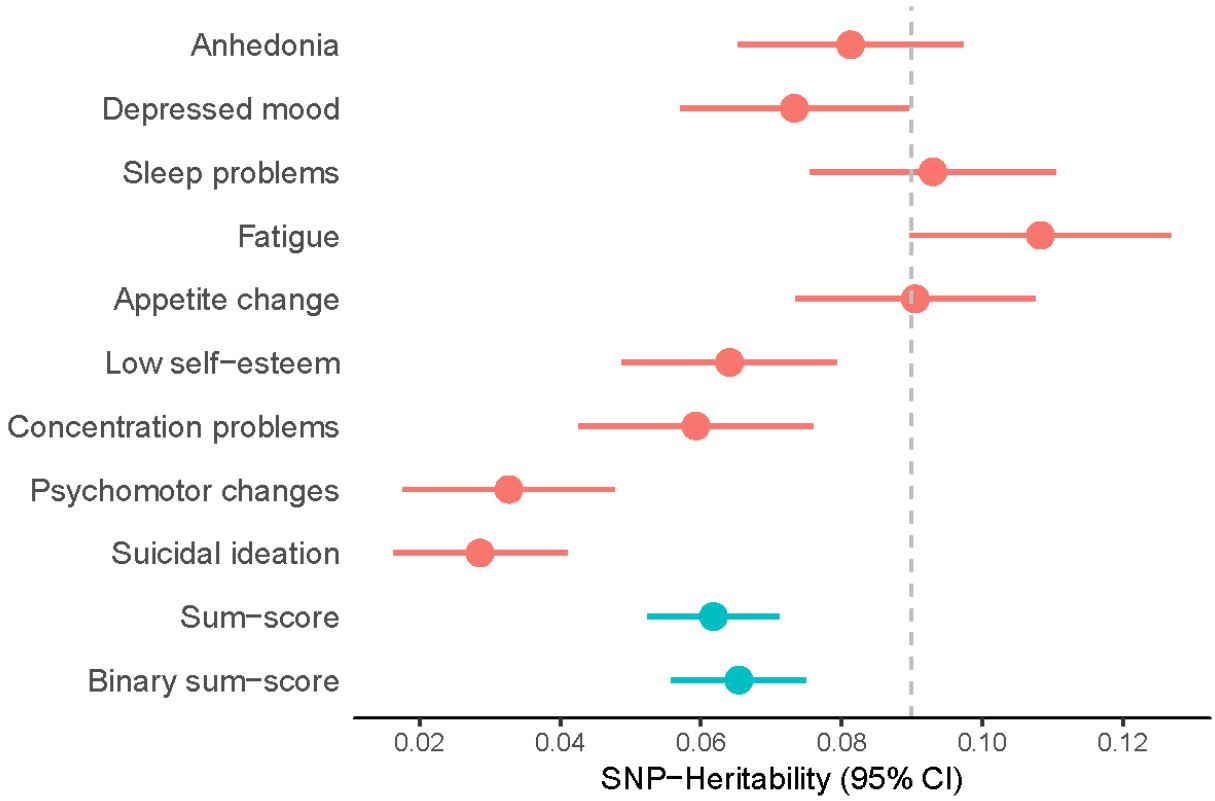
Table 2: GWAS Results for binary PHQ-9 items and sum-score phenotypes

Gen. locus	SNP id	Chr	BP	A1	A2	Freq A1	Phenotype	β	SE	p-value	nSNPs	nearest gene	eQTL genes
1	rs2279681	1	201861016	C	G	0.657	Sum-score	0.093	0.014	5.60E-11	25	*SHISA4*	SHISA4, LMOD1
							Binary sum-score	0.023	0.004	8.60E-09	6	*SHISA4*	
2	rs62158169	2	114081827	C	T	0.784	Sleep problems	0.015	0.002	1.40E-10	22	*PAX8*	FOXD4L1, CBWD2
3	rs137997194	3	48824937	A	G	0.955	Depressed mood	-0.023	0.004	7.70E-10	15	*PRKAR2A*	AMT, KLHDC8B
	rs143756010	3	49312248	C	T	0.956	Binary sum-score	-0.053	0.009	7.40E-09	15	*C3orf62*	AMT, NICN1, RNF123
4	rs12492113	3	50521402	G	A	0.875	Binary sum-score	-0.037	0.006	1.20E-10	114	*CACNA2D2*	C3orf18, CACNA2D2, CYB561D2, DOCK3, HEMK1, HYAL1, LSMEM2, RBM6, MANF, MAPKAPK3
							Sum-score	-0.124	0.020	1.30E-09	76	*CACNA2D2*	
							Appetite changes	-0.013	0.002	8.80E-09	28	*CACNA2D2*	
5	rs13127129	4	30501860	A	G	0.499	Sleep problems	0.010	0.002	4.90E-08	1	*PCDH7*	-
6	rs7073667	10	107809043	T	C	0.452	Sleep problems	0.010	0.002	4.30E-08	12	*RP11-298H24.1*	-
7	rs140920627	11	41577268	C	A	0.988	Anhedonia	-0.037	0.007	3.70E-08	1	*RP11-124G5.3*	-
8	rs840161	12	57323523	A	G	0.366	Binary sum-score	-0.021	0.004	3.90E-08	1	*SDR9C7*	-
9	rs2335859	19	7871837	T	G	0.602	Anhedonia	-0.009	0.001	9.80E-09	10	*EXOSC3P2*	-

Note: Table displays SNPs significant at $p < 5\times10^{-8}$ and independent at $r^2 < 0.10$. Genomic risk loci (Gen. locus) are defined by $r^2 < 0.10$, window size 1,000kb. Chromosome (Chr), location in base pairs (BP) on Hg19, effect allele (A1), allele 2 (A2), frequency of effect allele (Freq A1), effect size beta (β), standard error of beta (SE), p-value, and number of SNPs clumped under lead SNP (nSNPs) are shown. Proximity (nearest gene) and eQTL (eQTL genes) mapping results are given. eQTL mapping limited to significant (FDR < .05) cis-eQTLs from GTEx v7[53] and the CommonMind Consortium (CMC)[54]. SNP rs137997194 and rs143756010 are tagging the same signal, but was the SNP with lowest p-value in "depressed mood" and "binary sum-score", respectively.

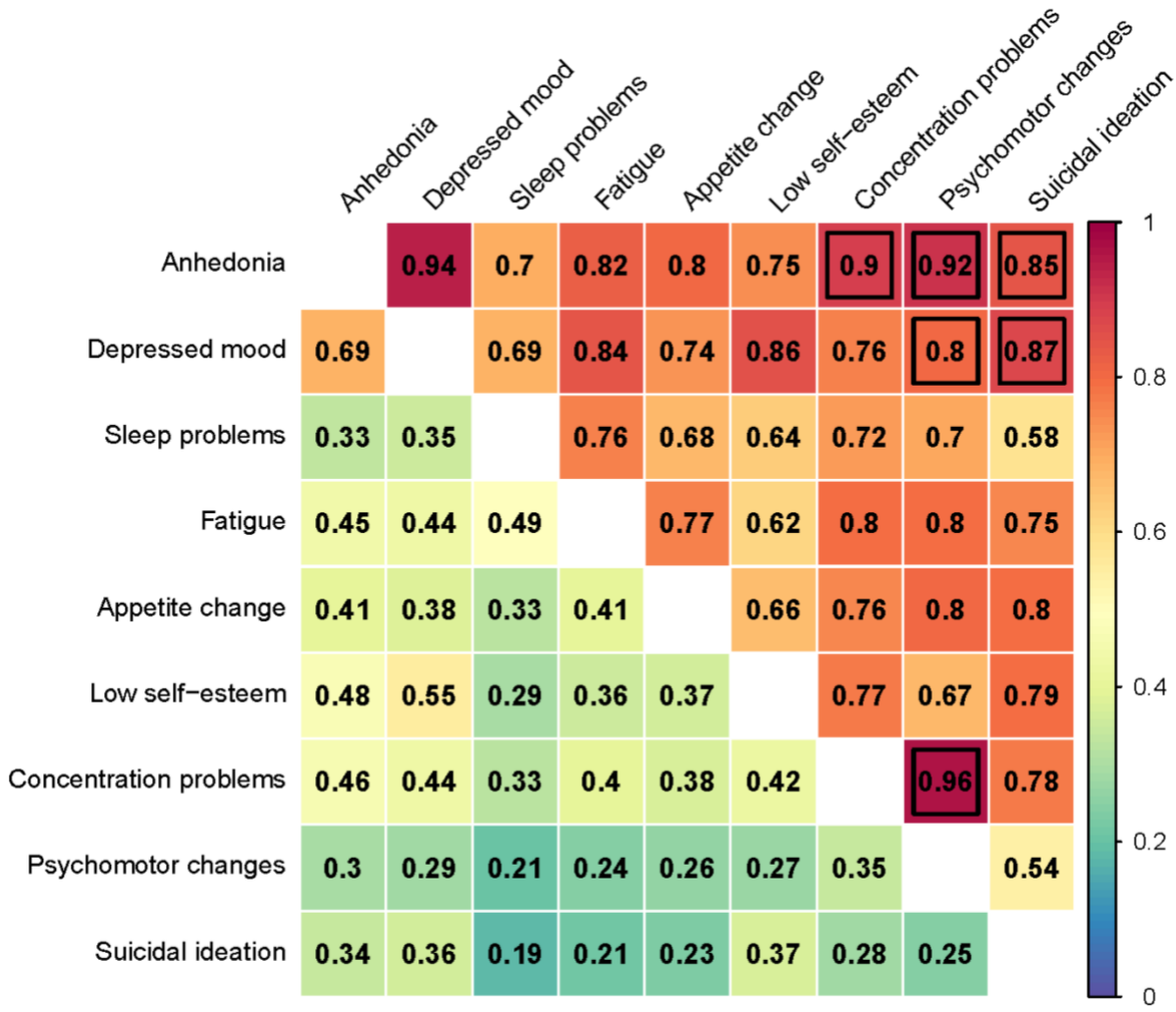
Figures

Figure 1. SNP-based heritability estimates and 95% confidence intervals (95% CI) for the nine depression items and sum-score phenotypes.



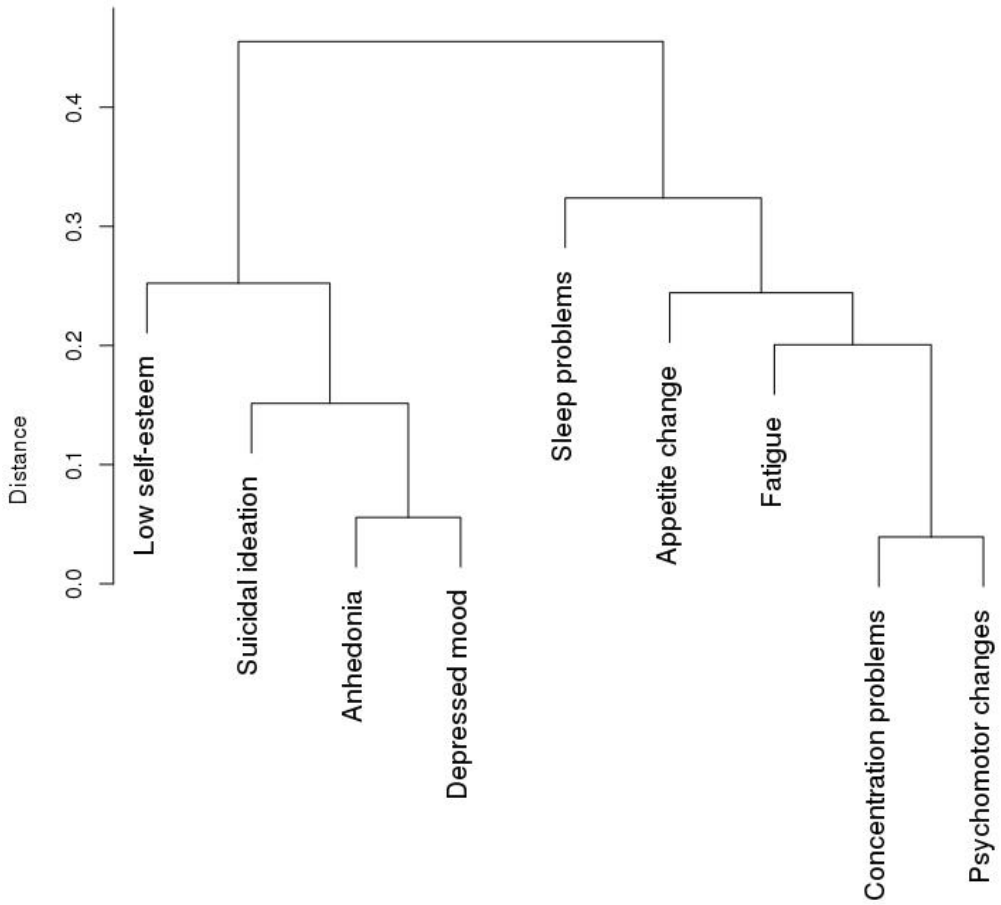
Note: heritability estimated via single-trait LD Score Regression. The dotted line indicates the SNP-based heritability of major depression (h^2_{SNP} = ~0.09)[3]. All estimates are significant after multiple testing correction ($p < 4.55\times10^{-3}$).

Figure 2. Inter-item genetic and phenotypic correlations.



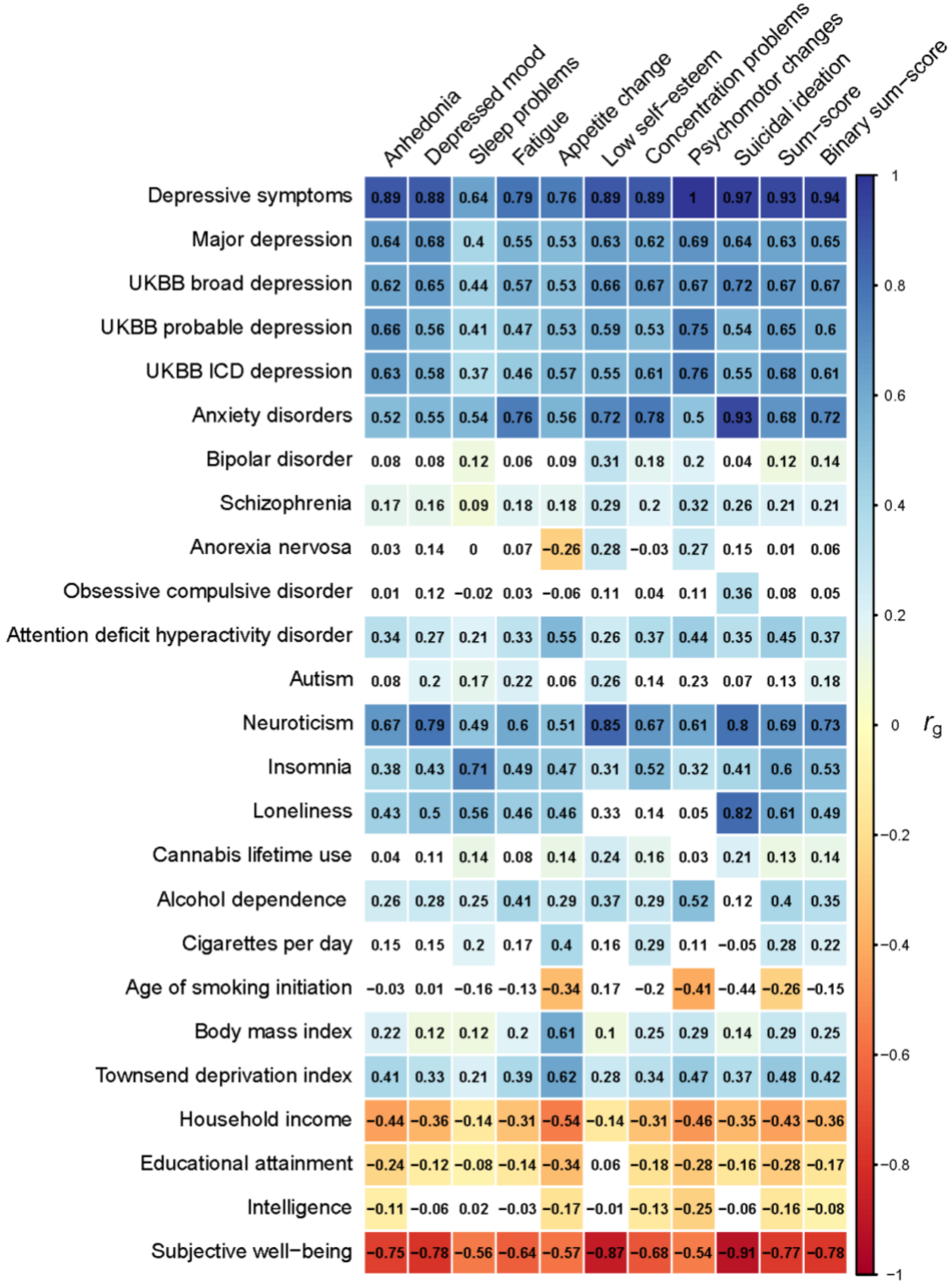
Note: Genetic correlations (r_g) above diagonal and phenotypic correlations (r_s) below diagonal. Genetic correlations estimated using cross-trait LD Score Regression. All correlations are significant at $p < 1.39\times10^{-3}$. Black squares indicate correlation is not significantly different from one (95% CI includes 1).

Figure 3. Cluster dendrogram of inter-item genetic correlations.



Note: distance is a measure of dissimilarity, defined as one minus the genetic correlation (1 - r_g).

Figure 4. Genetic correlations between PHQ-9 items and a range of other complex traits (psychiatric, substance use, and socioeconomic phenotypes) based on publicly available summary statistics.



Note: white squares indicate correlations that do not meet significance after correcting for FDR.

Acknowledgments

This work was conducted using the UK Biobank Resource (application number 25331). The UK Biobank was established by the Wellcome Trust medical charity, Medical Research Council (UK), Department of Health (UK), Scottish Government, and Northwest Regional Development Agency. It also had funding from the Welsh Assembly Government, British Heart Foundation, and Diabetes UK.

S.M. is supported by a National Health and Medical Research Council (NHMRC) Fellowship. A.T.M. is supported by the Foundation Volksbond Rotterdam.

Conflicts of interest

The authors declared no potential conflicts of interest with respect to the research, authorship and/or publication of this article.

References

1 Christian, O. *et al.* Major depressive disorder. *Nature Reviews Disease Primers* **2**, doi:10.1038/nrdp.2016.65 (2016).

2 Sullivan, P. F., Neale, M. C. & Kendler, K. S. Genetic epidemiology of major depression: review and meta-analysis. *The American journal of psychiatry* **157**, 1552-1562, doi:10.1176/appi.ajp.157.10.1552 (2000).

3 Wray, N. R. *et al.* Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nature Genetics*, doi:10.1038/s41588-018-0090-3 (2018).

4 Lee, S. H. *et al.* Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nature genetics* **45**, 984 (2013).

5 Ripke, S. *et al.* Genome-wide association study identifies five new schizophrenia loci. *Nature genetics* **43**, 969 (2011).

6 Ripke, S. *et al.* A mega-analysis of genome-wide association studies for major depressive disorder. *Mol Psychiatry* **18**, 497-511, doi:10.1038/mp.2012.21 (2013).

7 Hek, K. *et al.* A genome-wide association study of depressive symptoms. *Biol Psychiatry* **73**, 667-678, doi:10.1016/j.biopsych.2012.09.033 (2013).

8 Ripke, S. *et al.* Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**, 421 (2014).

9 Levinson, D. F. *et al.* 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**, 510-512, doi:10.1016/j.biopsych.2014.07.029 (2014).

10 Wray, N. R. *et al.* Research Review: Polygenic methods and their application to psychiatric traits. *Journal of Child Psychology and Psychiatry* **55**, 1068-1087, doi:10.1111/jcpp.12295 (2014).

11 Wray, N. R. *et al.* Genome-wide association study of major depressive disorder: new results, meta-analysis, and lessons learned. *Mol Psychiatry* **17**, 36-48, doi:10.1038/mp.2010.109 (2012).

12 Hilker, R. *et al.* Heritability of Schizophrenia and Schizophrenia Spectrum Based on the Nationwide Danish Twin Register. *Biol Psychiatry* **83**, 492-498, doi:10.1016/j.biopsych.2017.08.017 (2018).

13 Hyde, C. L. *et al.* Identification of 15 genetic loci associated with risk of major depression in individuals of European descent. *Nature Genetics* **48**, 1031, doi:10.1038/ng.3623
https://www.nature.com/articles/ng.3623#supplementary-information (2016).

14 Howard, D. M. *et al.* Genome-wide association study of depression phenotypes in UK Biobank identifies variants in excitatory synaptic pathways. *Nature Communications* **9**, 1470, doi:10.1038/s41467-018-03819-3 (2018).

15 American Psychiatric Association. *Diagnostic and statistical manual of mental disorders (DSM-5®)*. (American Psychiatric Pub, 2013).

16 Østergaard, S. D., Jensen, S. O. W. & Bech, P. The heterogeneity of the depressive syndrome: when numbers get serious. *Acta Psychiatrica Scandinavica* **124**, 495-496, doi:10.1111/j.1600-0447.2011.01744.x (2011).

17 Fried, E. I. & Nesse, R. M. Depression is not a consistent syndrome: An investigation of unique symptom patterns in the STAR*D study. *Journal of Affective Disorders* **172**, 96-102, doi:https://doi.org/10.1016/j.jad.2014.10.010 (2015).

18 Jang, K. L., Livesley, W. J., Taylor, S., Stein, M. B. & Moon, E. C. Heritability of individual depressive symptoms. *Journal of Affective Disorders* **80**, 125-133, doi:https://doi.org/10.1016/S0165-0327(03)00108-3 (2004).

19 Kendler, K. S., Aggen, S. H. & Neale, M. C. Evidence for multiple genetic factors underlying dsm-iv criteria for major depression. *JAMA Psychiatry* **70**, 599-607, doi:10.1001/jamapsychiatry.2013.751 (2013).

20 Nagel, M., Watanabe, K., Stringer, S., Posthuma, D. & van der Sluis, S. Item-level analyses reveal genetic heterogeneity in neuroticism. *Nature Communications* **9**, 905, doi:10.1038/s41467-018-03242-8 (2018).

21 John M. Hettema , M. D., Ph.D. ,, Michael C. Neale , P. D., John M. Myers , M. S., Carol A. Prescott , P. D. & Kenneth S. Kendler , M. D. A Population-Based Twin Study of the Relationship Between Neuroticism and Internalizing Disorders. *American Journal of Psychiatry* **163**, 857-864, doi:10.1176/ajp.2006.163.5.857 (2006).

22 Kroenke, K., Spitzer, R. L. & Williams, J. B. W. The PHQ‐9. *Journal of General Internal Medicine* **16**, 606-613, doi:doi:10.1046/j.1525-1497.2001.016009606.x (2001).

23 Bycroft, C. *et al.* The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203-209, doi:10.1038/s41586-018-0579-z (2018).

24 Kroenke, K., Spitzer, R. L. & Williams, J. B. W. The PHQ‐9: validity of a brief depression severity measure. *Journal of general internal medicine* **16**, 606-613 (2001).

25 Stewart, A. L., Hays, R. D. & Ware, J. E., Jr. The MOS short-form general health survey. Reliability and validity in a patient population. *Medical care* **26**, 724-735 (1988).

26 Yang, J., Zeng, J., Goddard, M. E., Wray, N. R. & Visscher, P. M. Concepts, estimation and interpretation of SNP-based heritability. *Nature Genetics* **49**, 1304, doi:10.1038/ng.3941
https://www.nature.com/articles/ng.3941#supplementary-information (2017).

27 Hall, L. S. *et al.* Genome-wide meta-analyses of stratified depression in Generation Scotland and UK Biobank. *Translational Psychiatry* **8**, 9, doi:10.1038/s41398-017-0034-1 (2018).

28 Loh, P.-R. *et al.* Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nature Genetics* **47**, 284, doi:10.1038/ng.3190
https://www.nature.com/articles/ng.3190#supplementary-information (2015).

29 Zhou, W. *et al.* Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies. *Nature Genetics* **50**, 1335-1341, doi:10.1038/s41588-018-0184-y (2018).

30 Loh, P.-R., Kichaev, G., Gazal, S., Schoech, A. P. & Price, A. L. Mixed-model association for biobank-scale datasets. *Nature Genetics* **50**, 906-908, doi:10.1038/s41588-018-0144-6 (2018).

31 Watanabe, K., Taskesen, E., Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nature communications* **8**, 1826 (2017).

32 Bulik-Sullivan, B. K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nature genetics* **47**, 291 (2015).

33 Lee, S. H., Wray, Naomi R., Goddard, Michael E. & Visscher, Peter M. Estimating Missing Heritability for Disease from Genome-wide Association Studies. *American Journal of Human Genetics* **88**, 294-305, doi:10.1016/j.ajhg.2011.02.002 (2011).

34 Bulik-Sullivan, B. K. *et al.* An atlas of genetic correlations across human diseases and traits. *Nature Genetics* **47**, 1236, doi:10.1038/ng.3406 https://www.nature.com/articles/ng.3406#supplementary-information (2015).

35 Müllner, D. fastcluster: Fast hierarchical, agglomerative clustering routines for R and Python. *Journal of Statistical Software* **53**, 1-18 (2013).

36 Martin-Löf, P. The Notion of Redundancy and Its Use as a Quantitative Measure of the Discrepancy between a Statistical Hypothesis and a Set of Observational Data [with Discussion]. *Scandinavian Journal of Statistics*, 3-18 (1974).

37 Kline, R. B. *Principles and practice of structural equation modeling*. 2nd edn, (Guilford, 2005).

38 Power, R. A. *et al.* Genome-wide Association for Major Depression Through Age at Onset Stratification: Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium(). *Biological Psychiatry* **81**, 325-335, doi:10.1016/j.biopsych.2016.05.010 (2017).

39 Okbay, A. *et al.* Genetic variants associated with subjective well-being, depressive symptoms, and neuroticism identified through genome-wide analyses. *Nature Genetics* **48**, 624, doi:10.1038/ng.3552 https://www.nature.com/articles/ng.3552#supplementary-information (2016).

40 Li, X. *et al.* Common variants on 6q16.2, 12q24.31 and 16p13.3 are associated with major depressive disorder. *Neuropsychopharmacology* **43**, 2146-2153, doi:10.1038/s41386-018-0078-9 (2018).

41 Kohli, Martin A. *et al.* The Neuronal Transporter Gene SLC6A15 Confers Risk to Major Depression. *Neuron* **70**, 252-265, doi:https://doi.org/10.1016/j.neuron.2011.04.005 (2011).

42 Direk, N. *et al.* An Analysis of Two Genome-wide Association Meta-analyses Identifies a New Locus for Broad Depression Phenotype. *Biological Psychiatry* **82**, 322-329, doi:https://doi.org/10.1016/j.biopsych.2016.11.013 (2017).

43 Cai, N. *et al.* Sparse whole-genome sequencing identifies two loci for major depressive disorder. *Nature* **523**, 588 (2015).

44 Gottlieb, D. J. *et al.* Novel loci associated with usual sleep duration: the CHARGE Consortium Genome-Wide Association Study. *Mol Psychiatry* **20**, 1232-1239, doi:10.1038/mp.2014.133 (2015).

45 Jones, S. E., Tyrrell, J. & Wood, A. R. Genome-Wide Association Analyses in 128,266 Individuals Identifies New Morningness and Sleep Duration Loci. **12**, e1006125, doi:10.1371/journal.pgen.1006125 (2016).

46 Lane, J. M. *et al.* Genome-wide association analyses of sleep disturbance traits identify new loci and highlight shared genetics with neuropsychiatric and metabolic traits. *Nature genetics* **49**, 274-281, doi:10.1038/ng.3749 (2017).

47 Nagel, M. *et al.* Meta-analysis of genome-wide association studies for neuroticism in 449,484 individuals identifies novel genetic loci and pathways. *Nature Genetics* **50**, 920-927, doi:10.1038/s41588-018-0151-7 (2018).

48 Myung, W. *et al.* Genetic association study of individual symptoms in depression. *Psychiatry Research* **198**, 400-406, doi:https://doi.org/10.1016/j.psychres.2011.12.037 (2012).

49 Motivala, S. J., Sarfatti, A., Olmos, L. & Irwin, M. R. Inflammatory markers and sleep disturbance in major depression. *Psychosomatic medicine* **67**, 187-194, doi:10.1097/01.psy.0000149259.72488.09 (2005).

50 Lamers, F. *et al.* Evidence for a differential role of HPA-axis function, inflammation and metabolic syndrome in melancholic versus atypical depression. *Mol Psychiatry* **18**, 692-699, doi:10.1038/mp.2012.144 (2013).

51 Nepon, J., Belik, S.-L., Bolton, J. & Sareen, J. The relationship between anxiety disorders and suicide attempts: findings from the National Epidemiologic Survey on Alcohol and Related Conditions. *Depression and Anxiety* **27**, 791-798, doi:doi:10.1002/da.20674 (2010).

52 Consortium, C.-D. G. o. t. P. G. Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. *The Lancet* **381**, 1371-1379 (2013).

53 Lonsdale, J. *et al.* The Genotype-Tissue Expression (GTEx) project. *Nature Genetics* **45**, 580, doi:10.1038/ng.2653
https://www.nature.com/articles/ng.2653#supplementary-information (2013).

54 Fromer, M. *et al.* Gene expression elucidates functional impact of polygenic risk for schizophrenia. *Nature Neuroscience* **19**, 1442, doi:10.1038/nn.4399
https://www.nature.com/articles/nn.4399#supplementary-information (2016).