Minimal phenotyping yields GWAS hits of low specificity for major depression

Na Cai, Ph.D.[1,2], Kenneth S. Kendler, M.D.[3], Jonathan Flint, M.D.[4]

1. Wellcome Sanger Institute, Wellcome Genome Campus, Cambridgeshire, UK
2. European Bioinformatics Institute (EMBL-EBI), Wellcome Genome Campus, Cambridgeshire, UK
3. Virginia Institute for Psychiatric and Behavioral Genetics, Department of Psychiatry, Virginia Commonwealth University, Richmond, Virginia, USA
4. Center for Neurobehavioral Genetics, Semel Institute for Neuroscience and Human Behavior, University of California, Los Angeles, Los Angeles, California, USA

Corresponding author:
Na Cai, Ph.D.
Wellcome Sanger Institute
Wellcome Genome Campus
Hinxton CB10 1SA
Cambridgeshire, UK
Email: nc10@sanger.ac.uk
Telephone: +44 (0)7821 070220

Abstract

Background

Minimal phenotyping refers to the reliance on self-reported responses to one or two questions for disease case identification instead of full diagnostic criteria. This strategy has been applied in genome-wide association studies (GWAS) on major depressive disorder (MDD), leading to lowering of phenotyping costs, increasing sample sizes, and more GWAS hits. It assumes that any increase in diagnostic noise, and its impact on the nature of GWAS loci thus identified, can be mitigated by the large increase in sample size.

Methods

We assess the impact of using different definitions of MDD in 337,198 White-British, unrelated samples in the UKBiobank with GWAS, analyses of heritability and genetic correlation, and comparison with previously published statistics on MDD and other psychiatric conditions. Definitions of MDD include seeking medical help for anxiety or depression, reporting MDD or its cardinal symptoms, and meeting full DSM criteria for MDD.

Findings

Heritability of depression defined by minimal phenotyping (<15%) is lower than DSM-based MDD (26%), and the former shares as much genetic liability with DSM-MDD (0·81) as it does a non-MDD help-seeking condition (0·84). While minimal phenotyping-based depression and DSM-based MDD show similar shared genetic liability with other conditions like neuroticism, a greater proportion of the genome contribute to the former (77%) than the latter (64%). Enrichment of heritability in CNS specific genes is found in both minimal phenotyping definitions of depression and other psychiatric conditions, but not DSM-MDD. GWAS loci identified in the minimal phenotyping definitions of depression, even when found in DSM-based MDD, show similar effects in other psychiatric conditions.

Interpretation

Using minimal phenotyping strategy for GWAS, when applied to MDD, carries significant potential risks. It primarily identifies non-specific genetic factors shared between MDD and other psychiatric conditions, biases our view of genetic architecture of MDD, and limits our ability to identify pathways specific to MDD.

Funding

Support for authors are detailed in Acknowledgements.

Research in context

Evidence before this study

The number of genetic loci associated with MDD published has increased from two in 2015 to 44 in 2018 due to the large increase in sample size afforded by minimal phenotyping approaches. Downstream analyses on these association statistics have shown that genetic factors contributing to MDD are enriched in central nervous system (CNS) specific genes, correlated with other psychiatric conditions, and account for only a small proportion of total liability to MDD. Most importantly, the increase in GWAS loci and reasonable functional interpretations from secondary analyses have led to the conclusion that increasing sample size, even at the expense of diagnostic accuracy or symptomatic homogeneity, is the key to identifying genetic loci associated with MDD.

Added value of this study

This study examines the consequences of using minimal phenotyping as a sample collection strategy for MDD on the results of genome-wide association studies conducted with the collected samples. Our results and simulations suggest that the increase in GWAS hits, the lower heritabilities, the genetic correlation with other psychiatric conditions, and the CNS-specific enrichment of heritability may be result of false-positive cases and misidentification of other conditions as MDD. With new evidence we conclude that genetic factors associated with minimal phenotyping definitions of depression are largely not specific to MDD, but shared with other psychiatric conditions. Using results from previously published GWAS on depression based on minimal phenotyping strategies, we are able to generalize our results to be representative of this strategy.

Implications of all available evidence

Minimal phenotyping definitions of MDD primarily identify loci shared with other psychiatric conditions due to an overly broad collection strategy that increase sample size at the expense of specificity. Secondary analyses using association statistics from GWAS conducted using this strategy should be mindful they are more useful for elucidating shared molecular pathways with other conditions than pathology specific to MDD. Taken at face value, they can misinform our prioritization of genes to be investigated for function in experimental or clinical studies, and incur great costs. With increasing accessibility and utilization of large biobanks and commercially collected datasets, our results are a timely warning of the consequences of lowering diagnostic specificity in genetic research. Future GWAS studies should resist the temptation of focusing solely on maximizing the number of identified GWAS loci, and put resources and collaborative efforts into building of large databases of detailed clinical diagnosis for identifying loci specific to MDD.

Introduction

There is now little doubt that a key requisite for the robust identification of genetic risk loci underlying psychiatric disease is the use of an appropriately large sample. However, while the costs of genotyping and sequencing continue to fall, the cost of phenotyping remains high[1], limiting sample collection. One solution for reducing the burden of case identification is to utilize minimal phenotyping, such as clinical information from a discharge register[2], or reliance on subjects' self-reported symptoms, help-seeking, diagnoses or medication, which can be collected much more cheaply a psychiatric assessment including all diagnostic criteria. We refer to this strategy as "minimal phenotyping", as it both minimizes phenotyping costs and reduces data collected from a full set of diagnostic criteria to a single or few self-reported answers.

However, apart from the detection of more and more GWAS loci[3-5] (Supplemental Table S1), the consequences of sacrificing symptomatic information on phenotype quality and the nature of loci identified has rarely been investigated. These may be particularly important for major depressive disorder (MDD) because of its phenotypic and likely etiological heterogeneity[6], high degree of comorbidity with other psychiatric diseases[7], and substantial discrepancies between self-assessment using symptom scales and diagnoses made with full diagnostic criteria[8]. While a majority of the population self-identify as having one or two depressive symptoms at any one time, only between 9 and 20% of the population have sufficient symptoms to meet criteria for lifetime occurrence of MDD[8-10]. Similarly, self-report of diagnosis or prescribed treatment, as employed by 23andMe, can be affected by the low rate of help-seeking among those who meet MDD diagnostic criteria (~50%)[11-13], false positives outnumbering true positives in primary care diagnoses unassisted with diagnostic criteria (by ~50%)[12], and high rates of anti-depressant prescription for a wide range of conditions[13-15]. As such, a cohort of MDD cases obtained through the use of self-report of either the illness or a prescribed treatment may yield a sample unrepresentative of the clinical disorder, but enriched in those with non-specific sub-clinical depressive symptoms and depression secondary to a comorbid disease.

By comparing the genetic architecture of minimal phenotyping definitions of depression with those using full diagnostic criteria for MDD in UKBiobank[16], a community-based survey of half a million men and women, we assess the implications of a minimal phenotyping strategy for GWAS in MDD. We find that MDD defined by minimal phenotyping has a large non-specific component, and if GWAS loci from these definitions are chosen for follow-up molecular characterization, they may not be informative about biology specific to MDD.

Results

Definitions of depression in UKBiobank

We identified four ways that MDD can be defined in UKBiobank. First, using self-reports of seeking medical attention for depression or related conditions, we identified a "Help-seeking" definition of MDD (referred to as "broad depression" in a previous GWAS[3]). Second, participants were diagnosed with "Symptom-based" MDD if, in addition to meeting the help seeking criteria described above, they reported ever experiencing one or more of the two cardinal features of depression (low mood or anhedonia) for at least two weeks (this is the "probable MDD" diagnosis available in UKBiobank[17]). 30% of cases in the Help-seeking based definition reported having no cardinal symptoms of depression and did not meet the "Symptom" based definition. We call these "No-MDD" help-seeking conditions. Third, we define a "Self-Report" form of depression based on participants' self-reports of all past and current medical conditions to trained nurses. Finally, a "DSM-based" diagnosis of lifetime MDD was derived from subjects who answered an online "Mental Health Follow-up" questionnaire[18] which included all DSM-V criteria for MDD (Supplemental Methods, Supplemental Table S2). We consider the first three categories to be various forms of "minimal phenotyping", while the fourth utilized full diagnostic criteria including the need for episode associated impairment or distress. Figure 1 outlines the different diagnostic categories, and the numbers that each contains.

There are three features to these definitions of MDD we recognize as important for the comparison of their genetic architecture. First, for each diagnostic category we used controls who were asked the relevant questions but failed to meet criteria. As such, cases from one category can be controls in another, resulting in substantial overlap in both cases and controls between categories (Supplemental Figure S1), which impacts assessment of genetic architecture and genetic correlation between them[19]. Second, not all participants in UKBiobank were asked questions from all categories. For example, questions for the Symptom-based definition DepAll were asked in only 10 out of 22 assessment centers in UKBiobank (Supplemental Table S3), resulting in differences in population structure between definitions (Supplemental Methods, Supplemental Figure S2). Third, different definitions of depression have different prevalences (from 0·078 to 0·341, Supplemental Table S4), which though unlikely biased due to population structure (Supplemental Tables S5-6), need to be corrected for when assessing genetic architecture.

Minimal phenotyping definitions of depression are epidemiologically different from DSM-based MDD

We assessed whether risk factors were similar between definitions of depression[20]. Figure 2a-f shows the mean effect (odds ratio, OR) with confidence intervals of each of the following known or putative risk factors on each definition of depression: sex[21,22], age[23], educational attainment[24-26], socio-economic status[27], neuroticism[28,29], and cumulative traumatic life events preceding assessment[30,31] (Supplemental Methods, Supplemental Table S7).

Estimates of the risk factor effect sizes differed substantially, and often highly significantly, as shown by the confidence intervals in the figure. This is most marked for traumatic experience, sex, and age, much less so for neuroticism and socio-economic status. We examined how the relationship between risk factors classifies depression diagnoses and found that minimal phenotyping definitions of depression cluster separately from DSM-based MDD (Figure 2g).

Minimal definitions of depression are not just milder or noisier version of DSM-MDD

We next addressed the question of the genetic relationship between the different MDD definitions. We made three observations. First, we found that depression defined by minimal phenotyping strategies have lower heritabilities than more strictly defined definitions (Figure 3a). Self-report and help-seeking based definitions have heritabilities of 15% or less (SelfRepDep heritability = 10·96%, se = 0·85%; Psypsy heritability = 12·57%, se = 1·18%; GPpsy heritability = 14·34%, se = 0·81%). Increasing the number of diagnostic criteria used for case definition to include the cardinal symptoms increases the estimated heritability moderately (DepAll heritability = 18·49%, se = 1·50%). By contrast, the DSM-based definition of MDD (LifetimeMDD) has a much higher heritability of 26·21% (se = 2·15%); imposing the further criterion of recurrence brings the heritability up to 31·95% (se = 2·56%). For comparison, we provide heritability estimates from previous studies of MDD[4,32,33] (Supplemental Figure S3) and find they fit squarely into the trend we observe: the less strict the criteria used to diagnose MDD, the lower the heritability. All heritability estimates were estimated on the liability scale using PCGCs[34], a method specifically suited for case-control analyses with covariates[19]. We ensured that the trend we observe holds regardless of method used[19,35-37] (Supplemental Methods, Supplemental Table S8), and was not affected by regions of high linkage-disequilibrium or complexity[38] (Supplemental Methods, Supplemental Figure S2).

Second, we found that shared genetic liability between minimal and strictly defined MDD is likely not specific to MDD. The genetic correlation between the help-seeking definition of depression GPpsy and DSM-based MDD LifetimeMDD is 0·81 (se = 0·03), significantly different than unity (Figure 3c). One interpretation of this finding is that the correlation represents a large shared genetic liability to MDD, and minimal phenotyping definitions such as GPpsy are therefore of interest to researchers trying to find the biological basis of depression[4,5]. However, of the proportion of genetic liability of LifetimeMDD that can be explained by that of GPpsy (approximately $0{\cdot}81^2$=66%), much of it may be shared with the No-MDD definition that *excluded* MDD symptoms (GPNoDep), the genetic liability of which explains approximately 0·70 of genetic liability of GPpsy (genetic correlation = 0·84, se = 0·05), and 0·34 of that of LifetimeMDD (genetic correlation = 0·58, se = 0·08). In other words, the genetic correlation between minimal and strictly defined MDD includes a large proportion of non-specific liability to mental ill-health, rather than specifically to MDD.

Finally, we excluded, or limited the role of a number of additional factors for the lower heritabilities of minimal phenotyping definitions of MDD. First, the inclusion of milder cases of MDD, equivalent to lowering the liability threshold under the liability threshold model[39], does not reduce the overall heritability (Supplemental Methods, Supplemental Figure S4). Second, minimal phenotyping definitions do not simply have a higher environmental contribution to MDD than the stricter definitions. When we assessed heritability in MDD cases with high and low exposure to environmental risk factors[36,40] we found that minimal phenotyping definitions of depression (GPpsy, SelfRepDep) show no significant difference between exposures, similar to or lower than DSM-based definitions (LifetimeMDD and MDDRecur) (Supplemental Methods, Supplemental Table S9). Finally, we show through simulations that while the lower heritabilities of minimal phenotyping definitions of depression may be in part because they are "noisier" versions of DSM-based MDD, misclassifications of those with other conditions as depressed can also contribute to the lower heritabilities (Supplemental Figure S4, Supplemental Figure S5). The latter is more consistent with the high genetic correlations between minimal phenotyping definitions of MDD with GPNoDep, the help-seeking no-MDD definition where cases do not have cardinal symptoms for MDD.

Minimal definitions of depression greater more genetic factors with other conditions

We next examined genetic correlations between different definitions of MDD and comorbid diseases. We used cross-trait LDSC[35] to estimate genetic correlations with neuroticism and smoking (Supplemental Figure S6, Supplemental Tables S10-11) in UKBiobank, as well as with all psychiatric conditions in the Psychiatric Genomics Consortium (PGC)[41] including PGC1-MDD[33] and depression defined in 23andMe[4] (Supplemental Table S1). While genetic correlation estimates in LDSC between case-control traits can be biased when strong covariates are involved[19], we were unable to mitigate this as we did not have access to case-control aware summary statistics appropriate for analysis in PCGCs. Figure 4a and Supplemental Table S13 displays little difference in estimates for different definitions of MDD in UKBiobank in their genetic correlation with other psychiatric conditions or previous studies of MDD, consistent with previous reports[42].

However, the observed genetic correlations result from different genetic architectures. We estimated local genetic correlations and their percentage contribution to total genetic correlation using rho-HESS[43] (Methods, Figure 4b) and show that while just 63·5%, 34·1% and 41·2% of the genome (indexed by percentage of total number of independent loci) explains 90% of the genetic correlation between DSM-based MDD (LifetimeMDD) and neuroticism, BIP and SCZ respectively, 77·2%, 43·6% and 45·3% of the genome is needed to explain the same percentage of genetic correlation between help-seeking definitions of depression (GPpsy) and the same conditions (Figure 4c). In other words, genetic liability of minimal phenotyping definitions of depression that can be accounted for by genetic liability of other psychiatric conditions is distributed across a greater proportion of loci across genome (each exerting a smaller effect), than that of DSM-based MDD.

Finally, we compared enrichment of heritability in genes expressed in each GTEx tissue[44] using LDSC-SEG[45]. As shown in Figure 5, help-seeking definition of depression (GPpsy) show significant enrichment of heritability in central nervous system (CNS) tissues, replicated by results from minimal phenotyping definition of MDD in 23andMe[4], while DSM-based MDD (LifetimeMDD), as well as PGC1-MDD, do not. Conversely, the no-MDD help-seeking definition GPNoDep, as well as other disorders in PGC[41] such as schizophrenia[46] (SCZ) and bipolar disorder[47] (BIP) all show enrichment of heritability in CNS, as do neuroticism and smoking. As such, previously reported[5] CNS enrichment in minimal definitions of MDD is potentially driven by loci shared with other psychiatric conditions like SCZ and BIP, as well as neuroticism and smoking, rather than being specific to MDD.

GWAS hits from minimal phenotyping are not specific to MDD

We next examined the specificity of action of individual genetic loci found in GWAS of each definition of MDD. We found that the help-seeking definitions gave the greatest number of genome-wide significant hits (27 from GPpsy and Psypsy) in GWAS. Are these loci relevant and specific to MDD, or are they non-specific and shared with other conditions such as SCZ, neuroticism and smoking? We consider loci that show significant and similar directions of effects in DSM-based definition of MDD (LifetimeMDD) as relevant to MDD, and those that show significant and similar directions of effects in other conditions as non-specific to MDD.

Of the 27 loci from minimal phenotyping definitions, 20 show significant effects (at P<0·05 after multiple testing correction for 23 loci) on LifetimeMDD, but all 20 also show significant effects in neuroticism, smoking, SCZ, or the no-MDD help-seeking condition (GPNoDep, Supplemental Table S14). Five loci show replication only in neuroticism, and ten show replication the no-MDD help-seeking condition GPNoDep. Figure 6 shows that the effects of all 27 loci on neuroticism mirror their effects on GPpsy. Hence, while using minimal phenotyping in GWAS is useful in identifying loci relevant to MDD, the loci identified are not specific to MDD.

We find the same pattern of results when we use loci identified from a minimal phenotyping strategy in an independent study. The consumer genetics company 23andMe mapped loci that contribute to the risk of MDD defined by self-reports of receiving a doctor's diagnosis and treatment of depression[4]. Of the 17 loci, ten replicated in GPpsy (at P<0·05, after multiple testing correction for 17 loci), of which six show significant effects in neuroticism, smoking or SCZ (Supplemental Figure S10, Supplemental Table S15) and are therefore non-specific to MDD. None of the 17 loci has a significant effect on LifetimeMDD alone. This demonstrates that minimal phenotyping approaches consistently identify genetic loci non-specific to MDD in GWAS.

Discussion

Though it is believed that the statistical power gain from using larger sample size compensates for inaccuracies in minimal phenotyping[4,5,48,49], our study demonstrate that this assumption does not always hold. Using a range of definitions of MDD in UKBiobank, from self-reported help-seeking to a full assessment of the DSM-V criteria for MDD, we made three major observations. First, the heritabilities of depression defined by minimal phenotyping strategies are lower than MDD defined by full DSM-V criteria. Second, although there is substantial genetic correlation between definitions, there remain significant differences, indicating the presence of genetic effects unique to each definition. Of the genetic liability shared between depression defined by minimal phenotyping and DSM-based MDD, a large proportion is not specific to MDD, as indexed by sharing with a diagnosis that excludes core depressive symptoms. Third, a larger proportion of genetic loci contributing to minimal phenotyping definitions of depression are shared with other psychiatric conditions than those contributing to DSM-based MDD, and the former show similar enrichment of heritability in CNS-specific genes and shared effects at GWAS loci with other psychiatric conditions. In other words, a large proportion of genetic factors identified in minimal phenotyping definitions of depression are not specific to MDD.

Previous studies have reported that sub-clinical depression lies on the same genetic liability continuum as MDD, has similar heritability as MDD, and a genetic correlation of 1 with MDD, therefore assaying the former is a valid way to interrogate genetics of MDD[50,51]. However, we show that none of these are satisfied by depression defined with minimal phenotyping – it does not simply include those with lower genetic liability to MDD as cases. Instead, the discrepancies can be explained by broadly classifying those with depressive symptoms as MDD due to their apparent high degree of sharing in genetic liability (0.7)[20], without taking into account the even higher degree of sharing between depressive symptoms and other traits such as neuroticism (0.79-0.94), especially if both were assayed at a single time point[52]. A recent study of three large twin cohorts specifically asked if a combination of MDD, depressive symptoms and neuroticism is able to capture all genetic liability of MDD[53], and shows that 65% of the genetic effects contributing to MDD are specific. In other words, broadly defined depression (inclusive of MDD, depressive symptoms and neuroticism) can index only around one-third of the genetic liability to MDD.

One view of these findings is that the shared loci between represent biological features of mental ill health that could be of use for understanding the biology of psychiatric disorders and hence be of use in their treatment[5,48]. A recent report on genetic analyses of subjective well-being, depressive symptoms and neuroticism identify high degree of sharing between genetic liabilities of the three measures, and use the "quasi-replication" of GWAS loci between depressive symptoms and neuroticism as validation of their functional significance[54]. An alternative view is that these loci reflect the ways in which depressive symptoms can develop as secondary effects, including through susceptibility to adverse life events[55], personality types[28], and use or exposure to psychoactive agents like cigarette smoking[56,57]. While useful for understanding the basis of mental ill health, they do not point specifically to the genetic basis of MDD.

As sample sizes in association studies increase to hundreds of thousands of subjects, the challenge is no longer in increasing the number of genetic loci associated with MDD, but identifying those that are relevant to the disease that warrant follow-up analyses of the mechanism by which they exert their effects. Ultimately, we wish to understand the etiology of the disease and identify targets for developing treatments. A reliance on minimal phenotyping will bias both GWAS and secondary analysis findings towards those genetic loci shared between MDD and other conditions, limiting the potential for disease-specific discoveries.

Methods

Control for population structure

We performed principal component analysis (PCA) on directly genotyped SNPs from samples in UKBiobank and used PCs as covariates in all our analyses to control for population structure. From the array genotype data, we first removed all samples who did not pass QC, leaving 337,198 White-British, unrelated samples. We then removed SNPs not included in the phasing and imputation and retained those with minor allele frequencies (MAF) >= 0·1%, and P value for violation of Hardy-Weinberg equilibrium > 10^{-6}, leaving 593,300 SNPs. We then removed 20,567 SNPs that are in known structural variants (SVs) and the major histocompatibility complex (MHC)[38] as recommended by UKBiobank[58], leaving 572,733 SNPs that we consistently use for all analyses. Of these, 334,702 are common (MAF > 5%), and from these common SNPs we further filtered based on missingness <0·02 and pairwise LD r^2 < 0·1 with SNPs in a sliding window of 1000 SNPs to obtain 68,619 LD-pruned SNPs for computing PCs using flashPCA[59]. We obtained 20 PCs, their eigenvalues, loadings and variance explained, and consistently use these PCs as covariates for all our genetic analyses. We note that control over population structure over the SVs and MHC is minimal, and we explore the impact of this in Supplemental Methods, Supplemental Table S4-6 and Supplemental Figure S2.

Imputed genotype filtering

We performed stringent filtering on imputed variants (version 2) used for GWAS in this study, removing variants not among the 33,619,058 variant sites in the Haplotype Reference Consortium[60] (HRC) panel, then removing all insertions and deletions (INDELs) and multi-allelic SNPs. We hard-called genotypes from imputed dosages at 8,968,715 biallelic SNPs with imputation INFO score greater than 0·9, MAF greater than 0·1%, and P value for violation of Hardy-Weinberg equilibrium > 10^{-6}, with a genotype probability threshold of 0·9 (anything below would be considered missing). Of these, 5,276,842 SNPs are common (MAF > 5%). We consistently use these SNPs for all analyses in this study.

Genome-wide associations

To obtain and access the difference between odds ratios of associations in different definitions of depression in UKBiobank, as well as smoking (data field 20160) and neuroticism (data field 20127), we perform logistic regression (or linear regression with --standard-beta for neuroticism) on all 5,276,842 common SNPs (MAF > 5% in all 337,198 White-British, unrelated samples) in PLINK (version 1·9)[61] with 20 PCs and genotyping array as covariates. We report all associations with P values smaller than 5×10^{-8} as genome-wide significant. We indicated the SNPs in SVs and the MHC in tables of top hits as well as all manhattan plots as hollow points instead of solid points due to lack of control for population structure in these regions, and do not interpret potential causal effects in these regions (Supplemental Tables S10-11,13, Supplemental Figure S6).

Estimation of heritability and genetic correlation among definitions of MDD

All estimates of heritability we refer to in the main text of this paper are computed with the phenotype-correlation-genotype-correlation (PCGC)[62] approach implemented with PCGCs[19], using 5,276,842 common SNPs (MAF > 5% in all 337,198 White-British, unrelated samples). We used LD scores at SNPs computed with LDSC[35] in 10,000 random samples drawn from the White-British samples in UKBiobank as LD reference, and MAF at all 5,276,842 common SNPs in all 337,198 White-British samples as MAF reference. As covariates we used genotyping array and 20 PCs computed using samples in each definition of MDD with flashPCA[59], and supplied the eigenvalues of the 20 PCs to PCGCs as deflation factors to correct for deflation in sum of squared kinship coefficients when regressing PCs out of genotype vectors. Where we stratified each definition of MDD in UKBiobank into two strata by risk factors such as sex (Supplemental Methods), we computed specific PCs for each definition and strata. We use summary statistics generated by PCGCs for each definition of MDD in UKBiobank to estimate its heritability, and summary statistics of pairs of definitions of MDD to estimate their genetic correlation. We detail the methods and results from other approaches we used to compare our estimates from PCGCs with in Supplemental Methods and Supplemental Table S8.

Estimation of genetic correlation between definitions of MDD and other conditions

We obtain summary statistics for other psychiatric conditions from previous GWAS studies as detailed on Supplemental Table S1. We also performed GWAS on smoking and neuroticism in UKBiobank (Supplemental Table S10-11, Supplemental Figure S6) for their association summary statistics. We estimated the genetic correlation between definitions of MDD in UKBiobank with each of these conditions with LDSC[63], with a LD reference panel generated with EUR individuals from 1000 Genomes[64], at SNPs with association statistics in both conditions in each instance. To obtain regional genetic correlation, we partitioned the genome into 1703 independent loci[65] and estimated regional genetic correlation at each locus with rho-HESS[43] using a LD reference panel generated with EUR individuals from 1000 Genomes[64]. To estimate the percentage of loci needed to explain 90% of the total genetic correlation for both LifetimeMDD and GPpsy with all conditions, we ranked all independent loci by their absolute genetic correlation (such that a locus with large negative genetic correlation counts just as much as that with a large positive genetic correlation), and asked genetic correlation contributions from how many loci would sum up to 90% of the total genetic correlation.

Enrichment of heritability in genes specifically expressed in tissues

We estimate the enrichment of heritability in genes specifically expressed in 44 tissues in Genotype–Tissue Expression (GTEx)[44] project using the partitioned heritability framework in LDSC-SEG[45], and a LD reference panel generated with EUR individuals from 1000 Genomes[64]. We first obtain tissue specific gene expression annotations in GTEx tissues from LDSC-SEG, then estimated the enrichment of heritability in annotations that corresponded to each of the tissues together with 52 annotations in the baseline model[66] which we also obtain from LDSC-SEG. We report the P value of the one-sided test of enrichment of heritability (positive regression coefficient for the tissue-specific annotation conditioning on other baseline annotations) in genes specifically expressed in each tissue against the baseline.

Figures

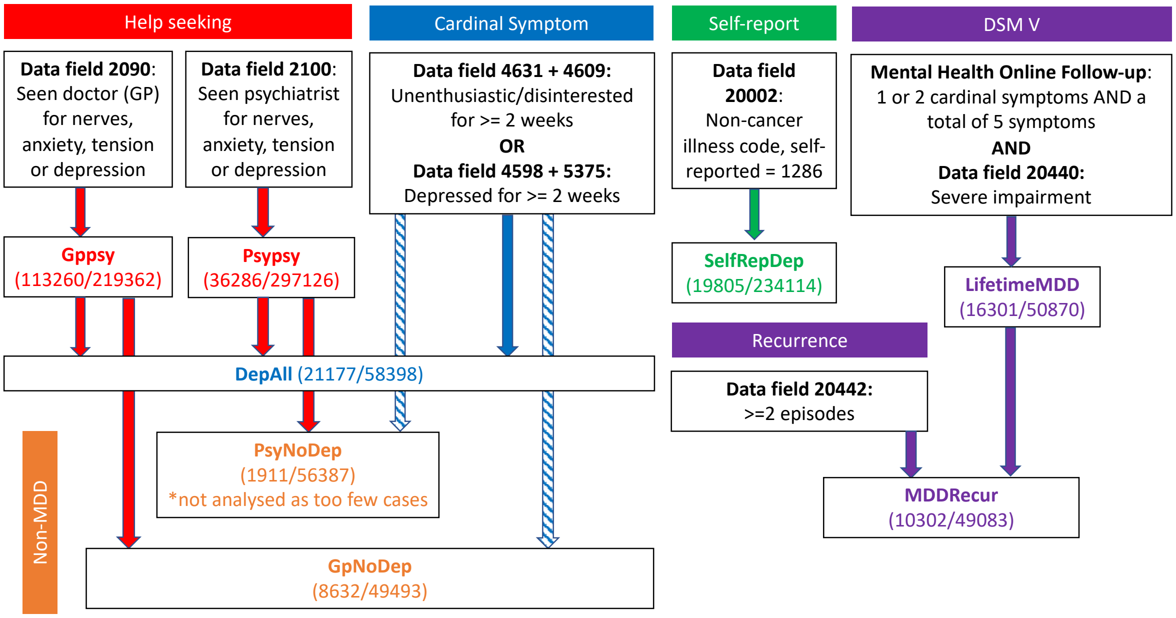
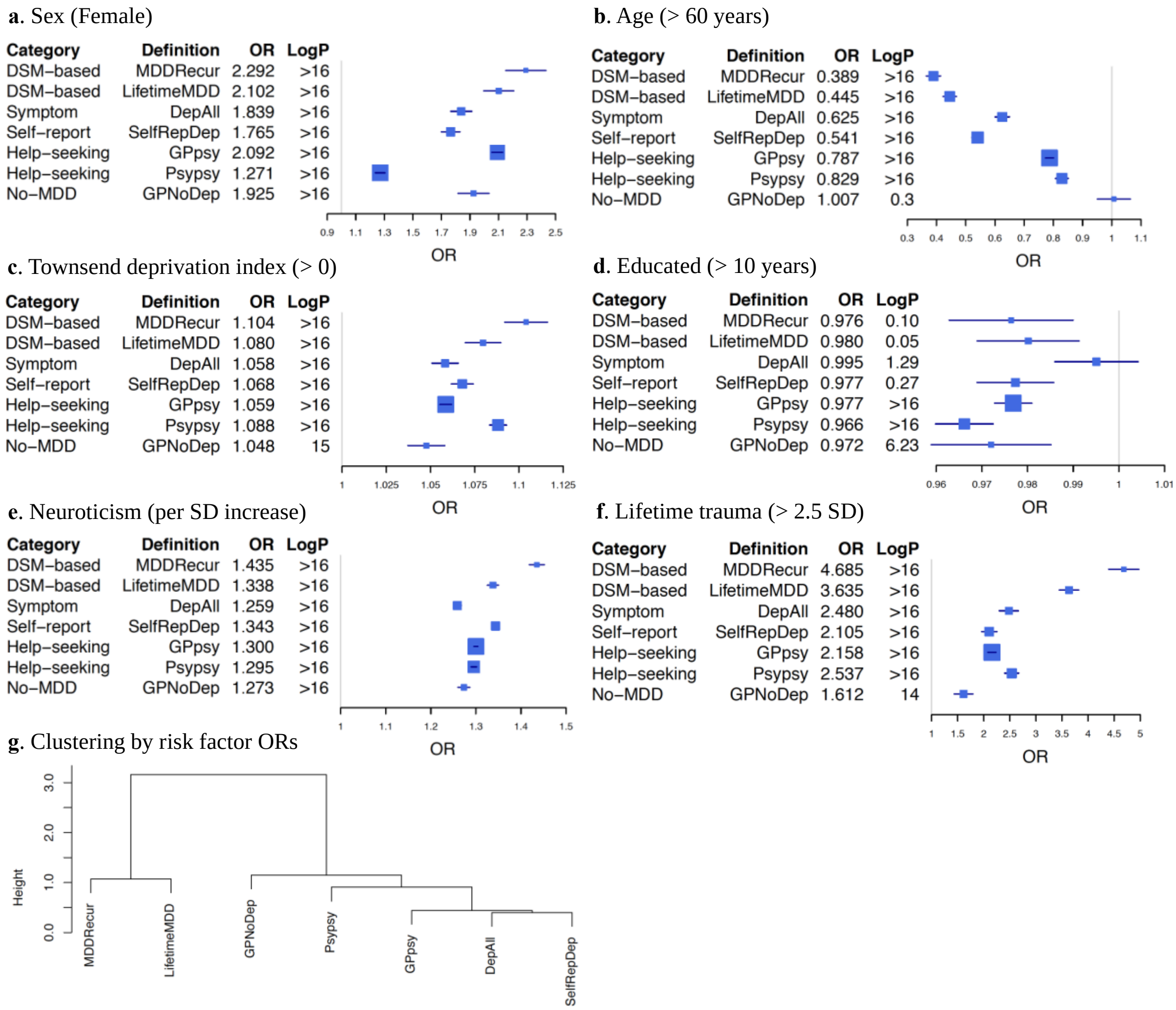

Help seeking
Cardinal Symptom
Self-report
DSM V
Data field 2090: Seen doctor (GP) for nerves, anxiety, tension or depression
Data field 2100: Seen psychiatrist for nerves, anxiety, tension or depression
Data field 4631 + 4609: Unenthusiastic/disinterested for >= 2 weeks
OR
Data field 4598 + 5375: Depressed for >= 2 weeks
Data field 20002: Non-cancer illness code, self-reported = 1286
Mental Health Online Follow-up: 1 or 2 cardinal symptoms AND a total of 5 symptoms
AND
Data field 20440: Severe impairment
Gppsy (113260/219362)
Psypsy (36286/297126)
SelfRepDep (19805/234114)
LifetimeMDD (16301/50870)
Recurrence
DepAll (21177/58398)
Data field 20442: >=2 episodes
PsyNoDep (1911/56387)
*not analysed as too few cases
Non-MDD
MDDRecur (10302/49083)
GpNoDep (8632/49493)

Figure1: Definitions of depression in UKBiobank | This figure shows the different definitions of MDD in UKBiobank and the colour codings we use consistently in this paper: Red for help-seeking based definitions derived from Touchscreen Questionnaire; blue for symptom based definitions derived from Touchscreen Questionnaire; green for self-report based definition derived from Verbal Interview; purple for DSM V based definitions derived from Online Mental Health Followup; orange for the No-MDD definition where cases are those who are cases in help-seeking definitions, but do not have cardinal symptoms for MDD. Data fields containing relevant information necessary for building these definitions of depression are specified; solid arrows emerging from a data field mean an answer of “Yes" to the question specified by the data field, while textured arrows mean an answer of “No"; number of cases and controls for each definition are shown in the form of cases/controls.



a. Sex (Female)

Category	Definition	OR	LogP
DSM–based	MDDRecur	2.292	>16
DSM–based	LifetimeMDD	2.102	>16
Symptom	DepAll	1.839	>16
Self–report	SelfRepDep	1.765	>16
Help–seeking	GPpsy	2.092	>16
Help–seeking	Psypsy	1.271	>16
No–MDD	GPNoDep	1.925	>16

b. Age (> 60 years)

Category	Definition	OR	LogP
DSM–based	MDDRecur	0.389	>16
DSM–based	LifetimeMDD	0.445	>16
Symptom	DepAll	0.625	>16
Self–report	SelfRepDep	0.541	>16
Help–seeking	GPpsy	0.787	>16
Help–seeking	Psypsy	0.829	>16
No–MDD	GPNoDep	1.007	0.3

c. Townsend deprivation index (> 0)

Category	Definition	OR	LogP
DSM–based	MDDRecur	1.104	>16
DSM–based	LifetimeMDD	1.080	>16
Symptom	DepAll	1.058	>16
Self–report	SelfRepDep	1.068	>16
Help–seeking	GPpsy	1.059	>16
Help–seeking	Psypsy	1.088	>16
No–MDD	GPNoDep	1.048	15

d. Educated (> 10 years)

Category	Definition	OR	LogP
DSM–based	MDDRecur	0.976	0.10
DSM–based	LifetimeMDD	0.980	0.05
Symptom	DepAll	0.995	1.29
Self–report	SelfRepDep	0.977	0.27
Help–seeking	GPpsy	0.977	>16
Help–seeking	Psypsy	0.966	>16
No–MDD	GPNoDep	0.972	6.23

e. Neuroticism (per SD increase)

Category	Definition	OR	LogP
DSM–based	MDDRecur	1.435	>16
DSM–based	LifetimeMDD	1.338	>16
Symptom	DepAll	1.259	>16
Self–report	SelfRepDep	1.343	>16
Help–seeking	GPpsy	1.300	>16
Help–seeking	Psypsy	1.295	>16
No–MDD	GPNoDep	1.273	>16

f. Lifetime trauma (> 2.5 SD)

Category	Definition	OR	LogP
DSM–based	MDDRecur	4.685	>16
DSM–based	LifetimeMDD	3.635	>16
Symptom	DepAll	2.480	>16
Self–report	SelfRepDep	2.105	>16
Help–seeking	GPpsy	2.158	>16
Help–seeking	Psypsy	2.537	>16
No–MDD	GPNoDep	1.612	14

g. Clustering by risk factor ORs

Figure2: Relationship between definitions of depression and environmental risk factors | a-f) These figures show forest plots of odds ratios (OR) and -log10 P values (LogP) between known environmental risk factors and different types (Category) of definitions of depression in UKBiobank (Definition) from logistic regression, using UKBiobank assessment centre, age, sex and years of education as covariates to control for potential geographical and demographic differences between environmental risk factors, except when they are being tested. Lifetime trauma measure was derived from Online Mental Health Followup (Supplemental Methods, Supplemental Table S7); Townsend deprivation index, years of education, sex, age, and neuroticism were derived from Touchscreen Questionnaire (Supplemental Methods). g) This figure shows a hierarchical clustering of definitions of depression in UKBiobank using ORs with environmental risk factors performed using the hclust function in R, "Height" refers to the Euclidean distance between MDD definitions at the ORs of all six risk factors.

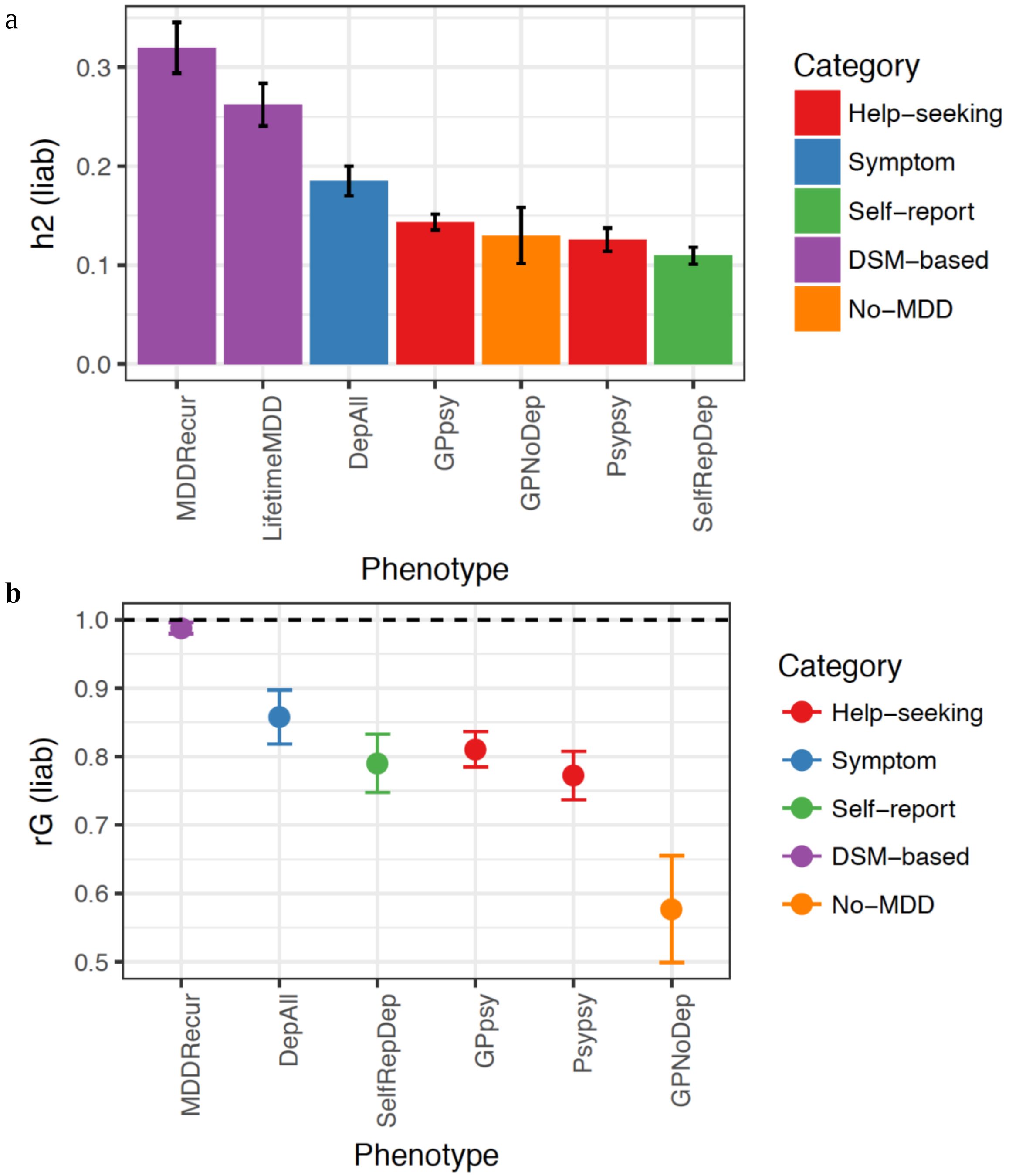


Figure3: Heritability and genetic correlation estimates among definitions of MDD in UKBiobank | a) This figure shows the heritability estimates from PCGCs[19] on each of the definitions of MDD in UKBiobank (Methods). Heritability "h2(liab)" as shown on the figure has been converted to liability scale[39,62] using the observed prevalence of each definition of depression in UKBiobank as both population and sample prevalences (Supplemental Table S4). b) This figure shows the genetic correlation between DSM-based LifetimeMDD and all other definitions of MDD in UKBiobank, estimated using PCGCs.

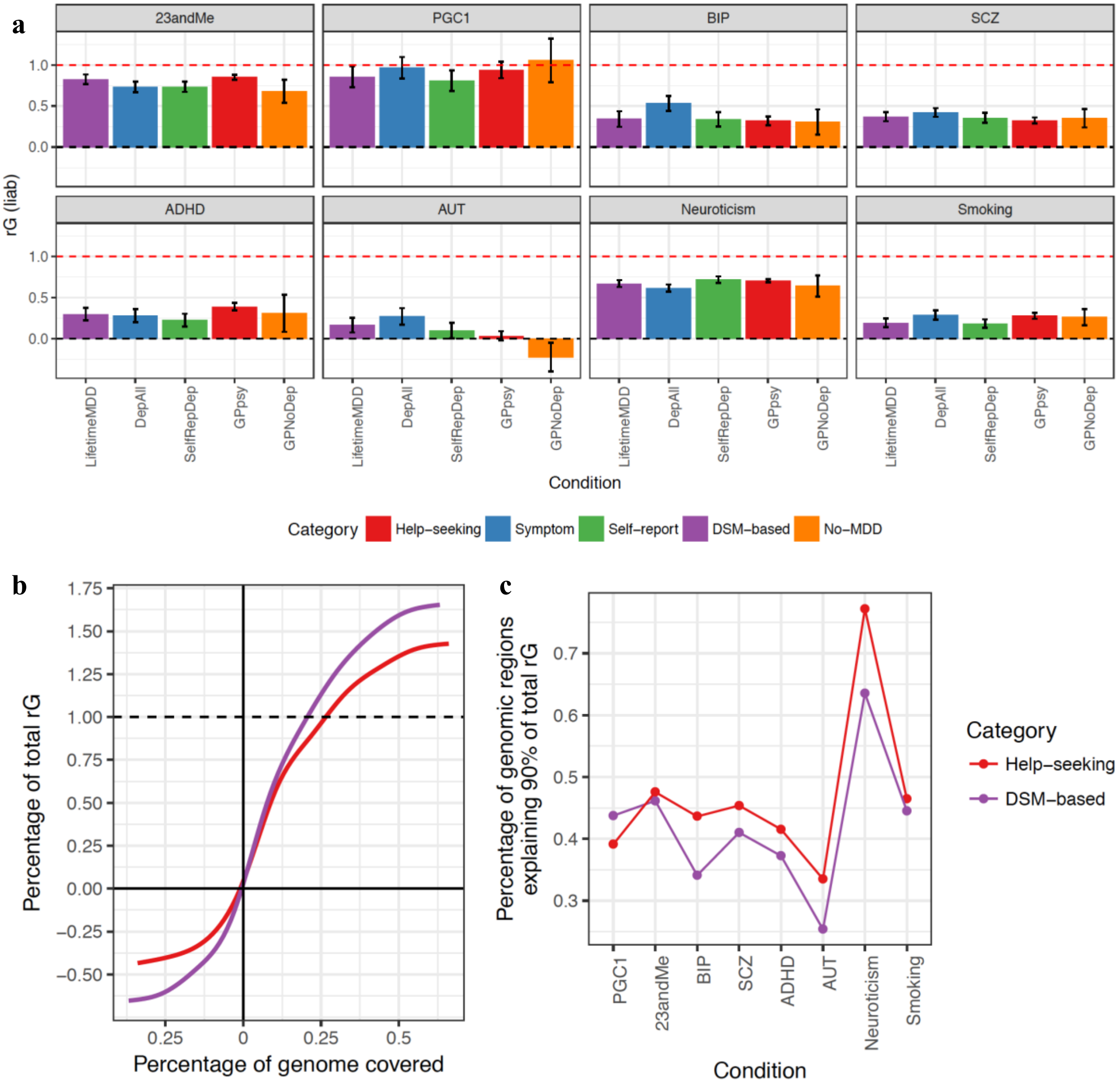


Figure4: Genetic correlation between definitions of MDD and other psychiatric conditions | a) This figure shows the genetic correlation "rG (liab)" estimated by cross-trait LDSC[63] on the liability scale between definitions of MDD in UKbiobank with other psychiatric conditions in both UKBiobank (smoking and neuroticism) and PGC[41] (Supplemental Table S1), including schizophrenia[46] (SCZ) and bipolar disorder[47] (BIP) (Supplemental Table S1). b) This figure shows the cumulative fraction of regional genetic correlation "rG" (out of sum of regional genetic correlation across all loci) between definitions of MDD in UKBiobank with SCZ in 1703 indepedent loci in the genome[65] estimated using rho-HESS[43], plotted against percentage of independent loci. DSM-based LifetimeMDD is shown in purple while help-seeking based GPpsy is shown in red. The steeper the curve, the smaller the number of loci explain the total genetic correlation. The dotted black line represents

100% of the sum of genetic correlation between each definition of MDD in UKBiobank with SCZ. The cumulative sums of positive regional genetic correlations (right of y axis) go beyond 100% – this is mirrored by the negative regional genetic correlation (left of y axis) that go below 0%. We rank all 1703 loci by their magnitude of genetic correlation, and ask what fraction of loci sums up to 90% of total genetic correlation. c) This figure shows the percentage of loci summing up to 90% of total genetic correlation “rG” between either LifetimeMDD (in purple) or GPpsy (in red) with all psychiatric conditions tested. The higher the percentage, the higher the number of genetic loci contributing to 90% of total genetic correlation.

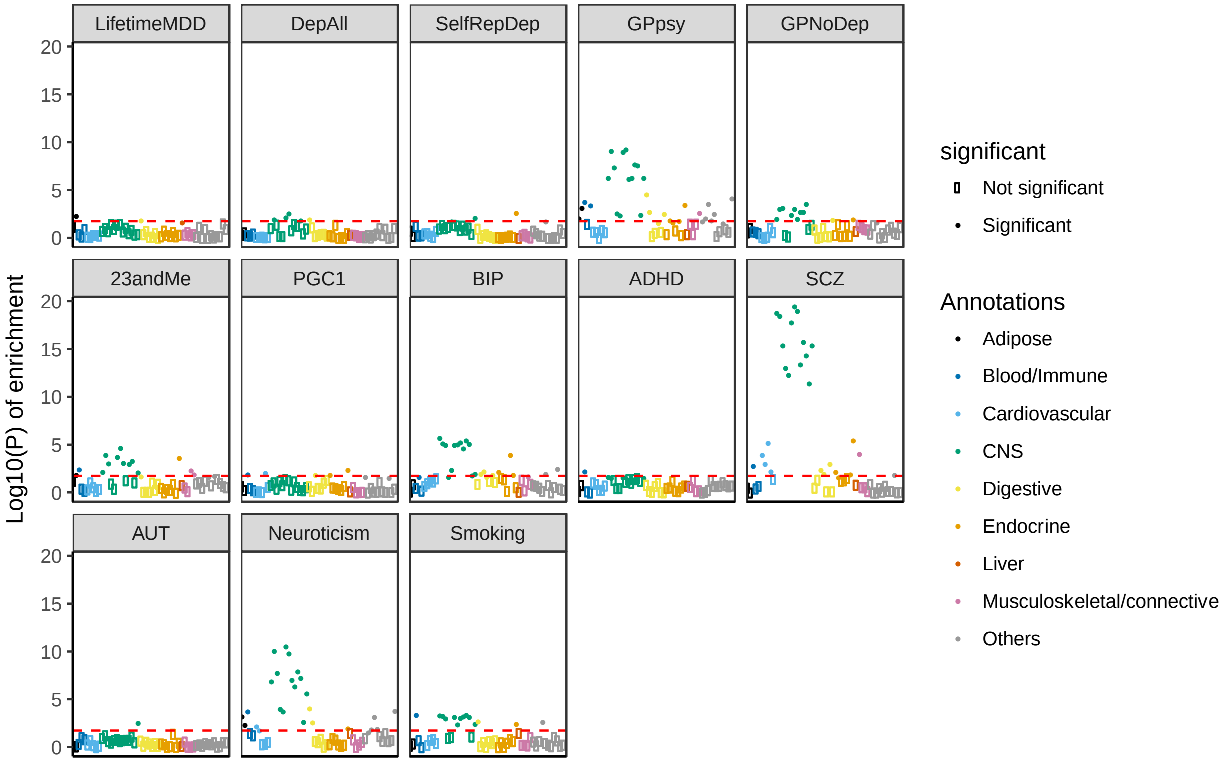


Figure5: Tissue-specific gene expression enrichment in definitions of MDD | This figure shows the -log10(P) of enrichment in heritability in genes specifically expressed in 44 GTEx tissues[45], estimated using partitioned heritability in LDSC[35]; help-seeking based definitions of MDD GPpsy, as well as its constituent no-MDD phenotype GPNoDep, show enrichment of heritability in genes specifically expressed in CNS tissues, similar to an independent cohort of help-seeking based MDD (23andMe[4]) and other psychiatric conditions such as bipolar disorder (BIP)[47], schizophrenia (SCZ)[46], autism (AUT), and personality dimension neuroticism.

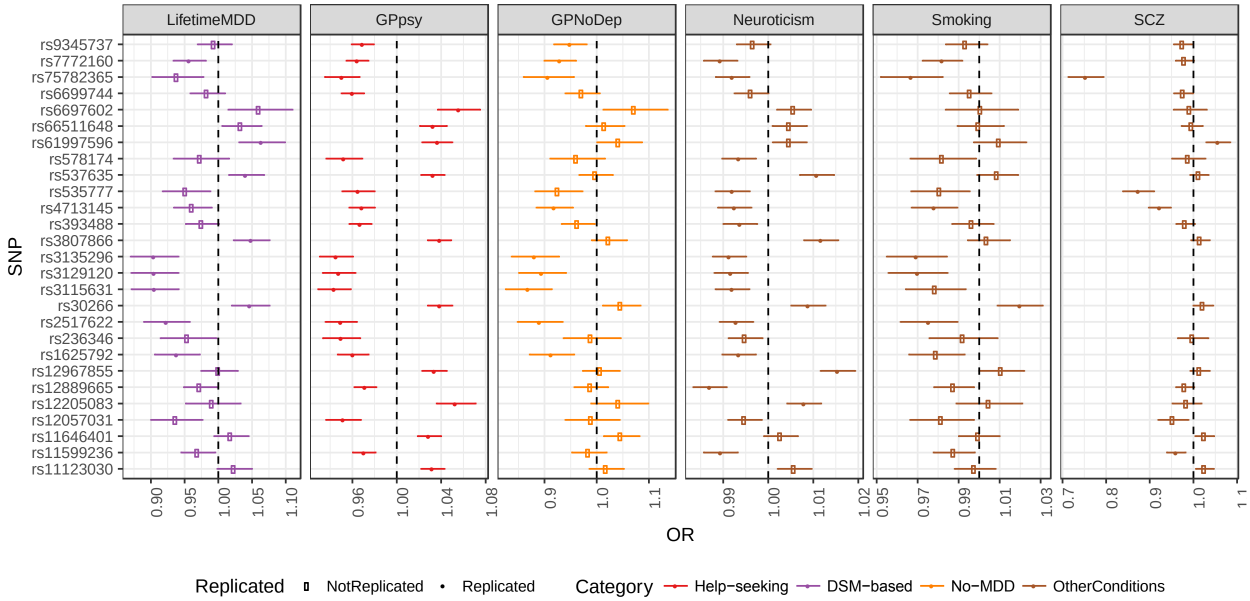


Figure6: GWAS hits from minimal phenotyping definition of MDD in UKBiobank are not specific to MDD | This figure shows the odds ratios (ORs) at 27 loci significantly associated with help-seeking based definitions of MDD in UKBiobank (GPpsy and Psypsy), in GWAS conducted on DSM (LifetimeMDD, in purple), help-seeking (GPpsy in red) and no-MDD (GPNoDep, in orange) based definitions of MDD, as well as conditions other than MDD: neuroticism, smoking and SCZ (all in brown). SNPs missing in each panel are not tested in the respective GWAS. For clarity of display, scales on different panels vary to accommodate the different magnitudes of ORs of SNPs in different conditions. ORs at all 27 loci are highly consistent across phenotypes, regardless of whether it is a definition or MDD or a risk factor or condition other than MDD. All results shown in Supplemental Table S14).

Author Contributions

N. C. and J. F. designed the study. N. C. performed the analyses. N. C. and J. F. obtained the data. N. C., K. S. K. and J. F. interpreted the results and wrote the manuscript.

Ethical approval:

This research was conducted under the ethical approval from the UK Biobank Resource under application no. 28709.

Acknowledgements

This research was conducted by using the UK Biobank Resource under application no. 28709. Na Cai is supported by the ESPOD Fellowship from the European Bioinformatics (EMBL-EBI) and Wellcome Sanger Institute.

Declaration of interests

The authors declare no conflicts of interest

Reference:

1 Lu, J. T., Campeau, P. M. & Lee, B. H. in *Obstetrical and Gynecological Survey* (2014).

2 Ripke, S. *et al.* Genome-wide association analysis identifies 13 new risk loci for schizophrenia. *Nature Genetics*, doi:10.1038/ng.2742 (2013).

3 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).

4 Hyde, C. L. *et al.* Identification of 15 genetic loci associated with risk of major depression in individuals of European descent. *Nature Genetics* **48**, 1031-1036, doi:10.1038/ng.3623 (2016).

5 Wray, N. R. *et al.* Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nature Genetics* **50**, 668-681, doi:10.1038/s41588-018-0090-3 (2018).

6 Flint, J. & Kendler, K. S. The genetics of major depression. *Neuron* **81**, 484-503, doi:10.1016/j.neuron.2014.01.027 (2014).

7 Kessler, R. C. *et al.* The epidemiology of major depressive disorder: results from the National Comorbidity Survey Replication (NCS-R). *JAMA* **289**, 3095-3105, doi:10.1001/jama.289.23.3095 (2003).

8 Boyd, J. H., Weissman, M. M., Thompson, W. D. & Myers, J. K. Screening for depression in a community sample. Understanding the discrepancies between depression symptom and diagnostic scales. *Arch Gen Psychiatry* **39**, 1195-1200 (1982).

9 Breslau, N. Depressive symptoms, major depression, and generalized anxiety: a comparison of self-reports on CES-D and results from diagnostic interviews. *Psychiatry Res* **15**, 219-229 (1985).

10 Weissman, M. M. & Myers, J. K. Rates and risks of depressive symptoms in a United States urban community. *Acta Psychiatr Scand* **57**, 219-231 (1978).

11 Berardi, D. *et al.* Increased recognition of depression in primary care. Comparison between primary-care physician and ICD-10 diagnosis of depression. *Psychother Psychosom* **74**, 225-230, doi:10.1159/000085146 (2005).

12 Mitchell, A. J., Vaze, A. & Rao, S. Clinical diagnosis of depression in primary care: a meta-analysis. *Lancet* **374**, 609-619, doi:10.1016/S0140-6736(09)60879-5 (2009).

13 Mojtabai, R. Clinician-identified depression in community settings: concordance with structured-interview diagnoses. *Psychother Psychosom* **82**, 161-169, doi:10.1159/000345968 (2013).

14 Druss, B. G. *et al.* Understanding mental health treatment in persons without mental diagnoses: results from the National Comorbidity Survey Replication. *Arch Gen Psychiatry* **64**, 1196-1203, doi:10.1001/archpsyc.64.10.1196 (2007).

15 Marcus, S. C. & Olfson, M. National trends in the treatment for depression from 1998 to 2007. *Arch Gen Psychiatry* **67**, 1265-1273, doi:10.1001/archgenpsychiatry.2010.151 (2010).

16 Sudlow, C. *et al.* UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med* **12**, e1001779, doi:10.1371/journal.pmed.1001779 (2015).

17 Smith, D. J. *et al.* Prevalence and characteristics of probable major depression and bipolar disorder within UK Biobank: Cross-sectional study of 172,751 participants. *PLoS ONE*, doi:10.1371/journal.pone.0075362 (2013).

18 Davis, K. A. S. *et al.* Mental health in UK Biobank: development, implementation and results from an online questionnaire completed by 157 366 participants. *BJPsych Open* **4**, 83-90, doi:10.1192/bjo.2018.12 (2018).

19 Weissbrod, O., Flint, J. & Rosset, S. Estimating SNP-Based Heritability and Genetic Correlation in Case-Control Studies Directly and with Summary Statistics. *The American Journal of Human Genetics* **103**, 89-99, doi:10.1016/j.ajhg.2018.06.002 (2018).

20 Foley, D. L., Neale, M. C. & Kendler, K. S. Genetic and environmental risk factors for depression assessed by subject-rated symptom check list versus structured clinical interview. *Psychol Med* **31**, 1413-1423 (2001).

21 Kendler, K. S., Gardner, C. O., Neale, M. C. & Prescott, C. A. Genetic risk factors for major depression in men and women: Similar or different heritabilities and same or partly distinct genes? *Psychological Medicine*, doi:10.1017/S0033291701003907 (2001).

22 Kendler, K. S., Gatz, M., Gardner, C. O. & Pedersen, N. L. A Swedish national twin study of lifetime major depression. *American Journal of Psychiatry*, doi:10.1176/appi.ajp.163.1.109 (2006).

23 Alexopoulos, G. S. *et al.* in *Archives of General Psychiatry* (1997).

24 Kessler, R. C. *et al.* Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. *Arch Gen Psychiatry* **62**, 593-602, doi:10.1001/archpsyc.62.6.593 (2005).

25 Kessler, R. C., Foster, C. L., Saunders, W. B. & Stang, P. E. Social consequences of psychiatric disorders, I: Educational attainment. *Am J Psychiatry* **152**, 1026-1032, doi:10.1176/ajp.152.7.1026 (1995).

26 Lorant, V. *et al.* Socioeconomic inequalities in depression: a meta-analysis. *Am J Epidemiol* **157**, 98-112 (2003).

27 Kessler, R. C. in *Journal of Affective Disorders* (2003).

28 Kendler, K. S., Gatz, M., Gardner, C. O. & Pedersen, N. L. Personality and major depression: a Swedish longitudinal, population-based twin study. *Arch Gen Psychiatry* **63**, 1113-1120, doi:10.1001/archpsyc.63.10.1113 (2006).

29 Kendler, K. S., Neale, M. C., Kessler, R. C., Heath, A. C. & Eaves, L. J. A longitudinal twin study of personality and major depression in women. *Arch Gen Psychiatry* **50**, 853-862 (1993).

30 Kessler, R. C. THE EFFECTS OF STRESSFUL LIFE EVENTS ON DEPRESSION. *Annual Review of Psychology*, doi:10.1146/annurev.psych.48.1.191 (1997).

31 Mazure, C. M. Life stressors as risk factors in depression. *Clinical Psychology: Science and Practice*, doi:10.1111/j.1468-2850.1998.tb00151.x (1998).

32 consortium, C. Sparse whole-genome sequencing identifies two loci for major depressive disorder. *Nature*, doi:10.1038/nature14659 (2015).

33 Major Depressive Disorder Working Group of the Psychiatric, G. C. *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).

34 Weissbrod, O., Flint, J. & Rosset, S. Estimating SNP-Based Heritability and Genetic Correlation in Case-Control Studies Directly and with Summary Statistics. *Am J Hum Genet* **103**, 89-99, doi:10.1016/j.ajhg.2018.06.002 (2018).

35 Bulik-Sullivan, B. K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nature Genetics*, doi:10.1038/ng.3211 (2015).

36 Loh, P.-R. *et al.* Contrasting genetic architectures of schizophrenia and other complex diseases using fast variance-components analysis. *Nature Genetics* **47**, 1385-1392, doi:10.1038/ng.3431 (2015).

37 Shi, H., Kichaev, G. & Pasaniuc, B. Contrasting the Genetic Architecture of 30 Complex Traits from Summary Association Data. *American Journal of Human Genetics*, doi:10.1016/j.ajhg.2016.05.013 (2016).

38 Price, A. L. *et al.* Long-Range LD Can Confound Genome Scans in Admixed Populations. *The American Journal of Human Genetics* **83**, 132-135, doi:10.1016/j.ajhg.2008.06.005 (2008).

39 Dempster, E. R. & Lerner, I. M. Heritability of Threshold Characters. *Genetics* **35**, 212-236 (1950).

40 Peterson, R. E. *et al.* Molecular genetic analysis subdivided by adversity exposure suggests etiologic heterogeneity in major depression. *American Journal of Psychiatry*, doi:10.1176/appi.ajp.2017.17060621 (2018).

41 Cross-Disorder Group of the Psychiatric Genomics, C. Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. *Lancet* **381**, 1371-1379, doi:10.1016/S0140-6736(12)62129-1 (2013).

42 Brainstorm, C. *et al.* Analysis of shared heritability in common disorders of the brain. *Science* **360**, doi:10.1126/science.aap8757 (2018).

43 Shi, H., Mancuso, N., Spendlove, S. & Pasaniuc, B. Local Genetic Correlation Gives Insights into the Shared Genetic Architecture of Complex Traits. *Am J Hum Genet* **101**, 737-751, doi:10.1016/j.ajhg.2017.09.022 (2017).

44 Consortium, G. T. The Genotype-Tissue Expression (GTEx) project. *Nat Genet* **45**, 580-585, doi:10.1038/ng.2653 (2013).

45 Finucane, H. K. *et al.* Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. *Nature Genetics*, doi:10.1038/s41588-018-0081-4 (2018).

46 Schizophrenia Working Group of the Psychiatric Genomics, C. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**, 421-427, doi:10.1038/nature13595 (2014).

47 Psychiatric, G. C. B. D. W. G. Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near ODZ4. *Nat Genet* **43**, 977-983, doi:10.1038/ng.943 (2011).

48 Mullins, N. & Lewis, C. M. Genetics of Depression: Progress at Last. *Curr Psychiatry Rep* **19**, 43, doi:10.1007/s11920-017-0803-9 (2017).

49 Sullivan, P. F. *et al.* Psychiatric Genomics: An Update and an Agenda. *Am J Psychiatry* **175**, 15-27, doi:10.1176/appi.ajp.2017.17030283 (2018).

50 Corfield, E. C., Yang, Y., Martin, N. G. & Nyholt, D. R. A continuum of genetic liability for minor and major depression. *Transl Psychiatry* **7**, e1131, doi:10.1038/tp.2017.99 (2017).

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

52 Duncan-Jones, P., Fergusson, D. M., Ormel, J. & Horwood, L. J. A model of stability and change in minor psychiatric symptoms: results from three longitudinal studies. *Psychol Med Monogr Suppl* **18**, 1-28 (1990).

53 Kendler, K. S., *et al.* Shared and Specific Genetic Risk Factors for Lifetime Major Depression, Depressive Symptoms and Neuroticism in Three Population-Based Twin Samples. *Psychological Medicine* (In Review).

54 Okbay, A. *et al.* Genetic variants associated with subjective well-being, depressive symptoms, and neuroticism identified through genome-wide analyses. *Nat Genet* **48**, 624-633, doi:10.1038/ng.3552 (2016).

55 Kendler, K. S. & Karkowski-Shuman, L. Stressful life events and genetic liability to major depression: genetic control of exposure to the environment? *Psychol Med* **27**, 539-547 (1997).

56 Fluharty, M., Taylor, A. E., Grabski, M. & Munafo, M. R. The Association of Cigarette Smoking With Depression and Anxiety: A Systematic Review. *Nicotine Tob Res* **19**, 3-13, doi:10.1093/ntr/ntw140 (2017).

57 Wootton, R. E. *et al.* Causal effects of lifetime smoking on risk for depression and schizophrenia: Evidence from a Mendelian randomisation study. *bioRxiv* (2018).

58 Bycroft, C. *et al.* Genome-wide genetic data on ~500,000 UK Biobank participants. *bioRxiv* (2017).

59 Abraham, G. & Inouye, M. Fast Principal Component Analysis of Large-Scale Genome-Wide Data. *PLoS ONE* **9**, e93766, doi:10.1371/journal.pone.0093766 (2014).

60 McCarthy, S. *et al.* A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet* **48**, 1279-1283, doi:10.1038/ng.3643 (2016).

61 Chang, C. C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience* **4**, 7, doi:10.1186/s13742-015-0047-8 (2015).

62 Golan, D., Lander, E. S. & Rosset, S. Measuring missing heritability: inferring the contribution of common variants. *Proc Natl Acad Sci U S A* **111**, E5272-5281, doi:10.1073/pnas.1419064111 (2014).

63 Bulik-Sullivan, B. *et al.* An atlas of genetic correlations across human diseases and traits. *Nat Genet* **47**, 1236-1241, doi:10.1038/ng.3406 (2015).

64 Genomes Project, C. *et al.* An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56-65, doi:10.1038/nature11632 (2012).

65 Berisa, T. & Pickrell, J. K. Approximately independent linkage disequilibrium blocks in human populations. *Bioinformatics*, doi:10.1093/bioinformatics/btv546 (2015).

66 Finucane, H. K. *et al.* Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nature Genetics* **47**, 1228-1235, doi:10.1038/ng.3404 (2015).