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

Supplemental Material

Supplemental Methods

Other studies referenced in this paper

We have summarised the other studies we reference in our paper, summary statistics of many of which are available on the Psychiatric Genomics Consortium website, in Supplemental Table S1. For each of the studies, we use only those SNPs with imputation INFO score above 0.9 and MAF > 5% in estimation of heritability, partitioned heritability, and genetic correlation.

Sample filtering

Of all 502,637 samples in UKBiobank full release, we performed the following QC steps to select the samples for use in our analyses. We first removed samples that were not included in the UKBiobank full release PCA analysis, which includes samples that were indicated as "het.missing.outliers" ("Indicates samples identified as outliers in heterozygosity and missing rates, which indicates poor-quality genotypes for these samples"), "excess.relatives" ("Indicates samples which have more than 10 putative third-degree relatives in the kinship table"), and whose "Submitted.Gender" were different from "Inferred.Gender", bringing the sample size down to 407,219. We checked that this remaining sample contains only one out of any pair or group of related individuals with relatedness > 0.05.

We then selected samples indicated to be "in.white.British.ancestry.subset" ("Indicates samples who self-reported 'White British' and have very similar genetic ancestry based on a principal components analysis of the genotypes"), bringing the sample size down further to 337,545. We then removed 337 samples indicated as having "putative.sex.chromosome.aneuploidy" ("Indicates samples identified as putatively carrying sex chromosome configurations that are not either XX or XY"). Finally, we removed 7 samples who have withdrawn their consent for use of their genetic data in analyses, arriving at our final set of 337,198 samples passing QC[1].

Of these samples, 37,041 were part of UK Biobank Lung Exome Variant Evaluation (UKBiLEVE), a study for chronic obstructive pulmonary disease (COPD). We retain all samples in UKBiLEVE, but as they are genotyped using a custom array optimized for coverage over regions implicated in lung health and disease[2], we consistently use genotyping array as a covariate in all our analyses.

For our analyses on different definitions of depression in UKBiobank, we further excluded 2,499 samples that indicated as having a history of substance abuse ("alcohol dependency": 1408, "opiod dependency": 1409 or "other substance abuse/dependency": 1410), manic or psychotic conditions (in the "Non-cancer illness code,

self-reported" (data field 20002) section of the verbal interview, or who are classified under "Bipolar I Disorder" or "Bipolar II Disorder" in the "Bipolar and major depression status" (data field 20126) derived data field, giving a total of 334,699 samples for our study on definitions of MDD in UKBiobank.

DSM-based definition of MDD

We derive a DSM-V[3] based *in silico* diagnosis of lifetime MDD using diagnostic questions on the online mental health follow-up (data category 138).

To qualify as a case of lifetime MDD according to DSM-V criteria, we first require individuals participating in UKBiobank to answer "Yes" to either the questions "Have you ever had a time in your life when you felt sad, blue, or depressed for two weeks or more in a row?" (data field 20446, DSM criterion A1 in Supplemental Table S2) or "Have you ever had a time in your life lasting two weeks or more when you lost interest in most things like hobbies, work, or activities that usually give you pleasure?" (data field 20441, DSM criterion A2 in Supplemental Table S2). Both are cardinal symptoms for DSM-V defined MDD.

We then require them to have 3 or 4 among the criteria A3-A9 shown in Supplemental Table S2. We note that this questionnaire does not contain the DSM criterion A5, which requires "Psychomotor agitation or retardation nearly every day (observable by others, not merely subjective feelings of restlessness or being slowed down)", and hence assays only 8 out of the 9 symptoms for DSM-V MDD. Nonetheless, we require individuals to have a total of at least 5 symptoms, including at least one out of the two cardinal symptoms (A1 and A2) such that they have at least 5 out of 9 symptoms for DSM-V defined MDD, to fulfil the DSM-V A criterion: *Five (or more) of the following symptoms have been present during the same 2-week period and represent a change from previous functioning; at least one of the symptoms is either (1) depressed mood or (2) loss of interest or pleasure*.

Finally, we require those who fulfil the symptoms criteria to answer "Yes" to the question "Think about your roles at the time of this episode, including study / employment, childcare and housework, leisure pursuits. How much did these problems interfere with your life or activities?" (data field 20440), to fulfil the DSM-V B criterion: *The symptoms cause clinically significant distress or impairment in social, occupational, or other important areas of functioning.*

As we exclude individuals who report a history of substance abuse, as well as manic or psychotic conditions (Methods), we further fulfil criteria C, D and E of DSM-V for MDD:

C. The (MDD) episode is not attributable to the physiological effects of a substance or to another medical condition.

D. The occurrence of the major depressive episode is not better explained by schizoaffective disorder, schizophrenia, schizophreniform disorder, delusional disorder, or other specified and unspecified schizophrenia spectrum and other psychotic disorders.

E. There has never been a manic episode or a hypomanic episode" respectively.

Among the diagnostic questions on the online mental health follow-up (data category 138), there are also questions on whether one had recent (2 weeks leading up to the assessment) symptoms of MDD. Using the questions on current MDD shown in Supplemental Table S2 and the same principles as stated above, for lifetime MDD, we derived a current MDD phenotype, which we combine with the lifetime MDD phenotype to arrive at the LifetimeMDD phenotype, representing DSM-based definition of MDD in our analysis. Further, using the question on lifetime number of depressed episodes in the online mental health follow-up (data field 20442), we designated those cases of LIfetimeMDD who indicated they had two or more episodes as cases for recurrent DSM-based MDD (MDDRecur). We show the overlaps between definitions of MDD in UKBiobank in Supplemental Figure S1.

Prevalence of definitions of MDD in UKBiobank

In our analyses of definitions of depression in UKBiobank, we make an important assumption that the prevalence of each definition of depression in the population UKBiobank is sampling from is exactly represented by the prevalence in the dataset. This cannot be true, as not all participants in UKBiobank were asked questions from all categories. Questions for the Symptom-based definition DepAll were asked in only 10 out of 22 assessment centers in UKBiobank (Supplemental Table S3). In addition, there may be self-report biases and uneven sampling of different demographics.

Using the 2011 UK census data[4], we find that discrepancy between sample and population prevalence for all definitions of depression we examine in UKBiobank are small when corrected for regional populations, age and sex (Supplemental Tables S4-6). As such, in all of our analyses, we use the observed sample prevalence for each definition of depression in UKBiobank as the population prevalence.

Control of population structure in heritability estimates

Questions forming the criteria for different definitions of MDD are not all answered by the same individuals in UKBiobank, potentially leading to different levels of cryptic relatedness and population structure among individuals making up cases and controls in different definitions of MDD. For example, questions on cardinal symptoms of MDD in the touchscreen questionnaire, necessary for meeting the criteria for being a case in DepAll, is answered by only those who went to 10 out of 22 assessment centres (Supplemental Table S3).

To ensure we control for population structure adequately for all definitions of MDD in UKBiobank regardless of which assessment centres the individuals making up cases and controls are from, we assessed the per-chromosome heritabilities of all definitions of MDD estimated jointly and separately with BOLT-REML, using a total of 334,681 genotyped SNPs with MAF > 5% and HWE P value > 10e-6 across all chromosomes as model SNPs. Difference between the slopes between per chromosome heritability and length of chromosome (approximated with number of SNPs used as model SNPs in each chromosome) in the two models (joint and separate) is telling of population structure that induce long-range LD between chromosomes, as heritability attributable to SNPs on one chromosome can be "leaked" into the estimates of a different chromosome due to long-range LD in presence of population structure, when heritabilities are estimated separately. We find that difference between the two slopes are minimal when using either PCs calculated using all White-British samples in UKBiobank, which we used as covariates in heritability estimates and GWAS on the definitions of MDD in UKBiobank (Methods). We show the results in Supplemental Figure S2a; we therefore use all-samples PCs in all the heritability estimates and GWAS we report, except for stratified analyses where numbers of samples in each stratum are substantially smaller than the whole UKBiobank White-British cohort where we used definition and stratum specific PCs.

As recommended by the UKBiobank[1], we removed the major histocompatibility complex (MHC) region as well as known structural variants (SVs)[5] before selecting SNPs for the computation of PCs (Methods) which we use for all association and heritability analyses. As a result, there is little control over population structure over these regions in both association testing and heritability estimation. This may lead to false positive associations as well as inflated heritability estimates.

We therefore indicated whether each significant association is in MHC and SV regions in Supplemental Table S6, and use hollow points to represent SNPs in manhattan plots (Supplemental Figure S6) - the validity of these associations can be followed up with sequencing of regions involved, and does not fall in the scope of this paper. To check if lack of population structure control over SVs and MHC regions affect heritability estimates, we estimate heritability using LDSC using all SNPs (LDSC-AllSNPs), excluding SNPs in the MHC region on chromosome 6:25-35MB (LDSC-noMHC), and excluding SNPs in the both MHC region and SVs [5] (LDSC-noMHCSVs). In Supplemental Figure S2b we show that while heritability decreases from LDSC-AllSNPs estimates when we remove MHC or/and SVs, the decreases are not significant in any of the definitions of MDD

in UKBiobank, and the trend between definitions of MDD remain the same. As such, we conclude that excluding MHC and SVs in the calculation of PCs is unlikely to cause significant biases in estimation of heritabilities of particular or all of the definitions of MDD.

Measure of lifetime trauma

A binary measure of lifetime trauma was derived from self-reported experience of traumatic events from the online mental health follow up questionnaire (data category 145), in order to identify individuals exposed to severe environmental adversities. The online mental health questionnaire for traumatic events consist of 16 questions on childhood, adulthood and lifetime trauma. We have scored individuals as having experienced each traumatic event as "1" and those who have not experienced each traumatic event as "0", as shown in Supplemental Table S7. Of the 16 questions, we have included 12 for derivation of an aggregate "lifetime trauma" measure, the remaining 4 capture the same traumatic events the other questions we have included do, and are hence redundant.

We then asked how much each traumatic experience contribute to risk of developing LifetimeMDD, the DSM-based definition of MDD in UKBiobank, by jointly modelling their contribution along with age, sex, social deprivation (in townsend deprivation index), years of education, experience of any traumatic life event in the past 2 years (data field 6145), neuroticism, and total number of traumatic experiences one had in logistic regression. The odds ratios (ORs) of each traumatic experience on LifetimeMDD is shown in Supplemental Table S7. Since traumatic events vary in severity, a lifetime trauma score was constructed by weighting each item by their effect size on LifetimeMDD, and summing across all 12 items. As not all questions on traumatic experience are answered by all individuals who took part in the online mental health follow-up, we removed individuals who did not answer more than 3 questions on traumatic experience. We obtained lifetime trauma score for 109,699 individuals, with mean score 1.90 (standard deviation = 2.11). We categorize 33,619 individuals with scores above 2.5 as having experienced significant lifetime trauma (who report a mean of 3.17 traumatic experiences) and those with scores below 2.5 as not having experienced significant lifetime trauma (who report a mean of 0.66 traumatic experiences).

We note the following caveats in our use of self-reported answers to questionnaires to infer level of lifetime trauma. First, in weighting traumatic events based on their effects on LifetimeMDD in creating the measure of lifetime trauma, one can potentially incur biases in genetic associations to LifetimeMDD at genetics variants that contribute to both lifetime trauma and LifetimeMDD when one stratifies individuals in LifetimeMDD by lifetime trauma. However, this bias will likely be small, as the weighted lifetime trauma score is highly correlated with number of traumatic experiences individuals self-report ($r2=0.92$, $p < 10^{-16}$), and this self-report is assumed

to be independent of LifetimeMDD (see second caveat). Moreover, even if present, this bias will not extend to other definitions of MDD in UKBiobank acquired through independent questionnaires.

Second, as both questionnaires on traumatic events and DSM-based MDD symptoms (used for specifying LifetimeMDD) are on the same online mental health follow up assessment, how one answers one questionnaire may affect how one answers the other, potentially incurring errors in retrospective recall. This is an unavoidable problem in UKBiobank given the structure of the questionnaire, and as such here we assume that the errors incurred are negligible. In addition, as the online mental health follow-up is conducted a year or two after one takes part in the initial assessment through the touchscreen questionnaire, there should be little to no effects of retrospective recall errors on all other definitions of MDD in UKBiobank which we derive from answers to questions in the touchscreen questionnaire.

Epidemiological analysis on risk factors for MDD

We assess the effects of known risk factors for MDD on the different definitions of MDD in UKBiobank; for each binary or quantitative risk factor, we asked for its odds ratio (OR) for each definition of MDD using logistic regression correcting for UKBiobank collection centers (data field 54, as factors), as well as years of education (data field 845, quantitative), age (data field 21022, quantitative) and sex (data field 31, binary) unless one of three factors is being tested. The binary risk factors we tested are:

1. Age: data field 21022; dividing individuals into those < 60 years old and those >= 60
2. Sex: data field 31; dividing individuals into males and females
3. lifetime trauma: data category 145 for "Traumatic events reported within the on-line mental health questionnaire"; we calculated a weighted measure of traumatic events, dividing individuals into those who have weighted measure = 1 and those with weighted measure = 0.

The quantitative risk factors we tested are:

1. neuroticism: data field 20127 for "12 neurotic behaviour domains as reported from fields 1920, 1930, 1940, 1950, 1960, 1970, 1980, 1990, 2000, 2010, 2020 and 2030 from the touchscreen questionnaire at baseline"
2. social deprivation: data field 189 for townsend deprivation index (TDI) calculated for per individual just before baseline assessment)
3. years of education: data field 845; for age one stopped continuous full-time education, from which we infer years of continuous education received by each individual assuming starting age of 6.

For quantitative risk factors we report ORs for per SD increase in the measure. The UK Biobank cohort contains more women than men (the female to male ratio is 1.16 to 1), with more younger women than men (there are 6% more women than men younger than 60, compare to 2% older than 60) and is more educated among the young than the old (those with more than 10 years of education form 42.5% of those younger than 60, compare to 28.4% among those older than 60). Thus, in each analysis we included age, sex and number of years of education as covariates unless they are the risk factors tested (together with collection center).

Comparison of methods of heritability and genetic correlation estimation

Many methods have been developed to estimate narrow sense heritability from either whole-genome SNP data[6-9] or summary statistics of association tests [10-12] across the whole genome, and the sensitivity of heritability estimates to assumptions in different models employed by the different methods have been discussed and reviewed extensively in the past few years[13,14].

For all the heritability estimates in our analysis we use the phenotype-correlation-genotype-correlation (PCGC) approach, an adaptation of Haseman-Elston regression specifically suited for analysing case-control data in the presence of covariates of large effects, such as sex and age, and has been shown to produce unbiased estimates of heritability and genetic correlation for case-control phenotypes[6,12]. We use the PCGCs[6,12] framework, generating summary statistics for definitions of MDD in UKBiobank in our study which can be reused for estimation of genetic correlation with other phenotypes from other cohorts processed the same way.

We note, however, that previously published estimates of heritability of MDD[15-19], are mostly generated using LDSC[10] if they were estimated from GWAS summary statistics, or either GCTA[9], LDAK[8], or BOLT-REML[7] if they were estimated from individual level genotypes. All of them estimate heritability of case-control traits on the "observed scale" under a quantitative trait framework, and apply a correction factor to results to convert them to a "liability scale" to account for the ascertainment of cases in case-control studies. The same applies when estimating genetic correlation.

In Supplemental Table S8, we show the heritability estimates of each definition of MDD in UKBiobank using PCGCs, LDSC, and BOLT-REML. For both PCGCs and LDSC, we use 8,968,716 imputed SNPs with MAF > 5% in our analysis (Methods), while for BOLT-REML we use 334,681 genotyped SNPs with MAF > 5% and HWE P value > 10e-6 as model SNPs using which it estimates heritability. For both results from LDSC and BOLT-REML, we use LD scores calculated on all imputed SNPs in UKBiobank using LDSC as reference, and convert the heritability estimates on the observed scale to liability scale specifying the prevalence of each definition of MDD in UKBiobank as shown in Supplemental Table S8.

For PCGCs, we show in addition to heritability the genetic variance and its standard error, which is variance attributable to genetics over total phenotypic variance (referred to as "marginal heritability" in Weisbrod et al 2018[12]), the comparable measure to heritability from the other two methods. Heritability (h2, referred to as "conditional heritability") in PCGCs is genetic variance divided by the total liability variance, which also includes variance introduced by fixed effects (in this case, genotype PCs and genotyping array). The similarity between heritability and genetic variance estimates from PCGCs shows that the fixed effect covariates do not contribute much to variance in any of the definitions of MDD.

Results from the three methods show the same trend - the DSM-based definitions LifetimeMDD and MDDRecur show the highest heritability estimates while the help-seeking based definitions GPpsy and Psypsy, as well as their no-MDD components GPNoDep and PsyNoDep, show the lowest heritability estimates. To put the heritability estimates of definition of MDD used in our study in context with previous published estimates from LDSC, we show in Supplemental Figure S3a that the DSM-based definitions LifetimeMDD and MDDRecur have heritabilities closer to LDSC heritability estimates of MDD in CONVERGE[15] and PGC1-MDD[18], while help-seeking based definitions GPpsy and Psypsy have estimates closer to that in similarly defined, minimal phenotyping based MDD in 23andMe[17].

The difference between the methods show the same expected trend discussed in the literature on heritability estimation. While LDSC and BOLT-REML estimates are highly similar to each other, PCGCs estimates are higher than both of them, consistent with expectations of downward bias in heritability estimation of case-control traits using LDSC and REML due to a variety of factors including handling of covariates and use of external MAF references in PCGCs.

Finally, we compare the estimation of genetic correlation between definitions of MDD using PCGCs and LDSC. In Supplemental Figure S3b we show the genetic correlation between LifetimeMDD, the DSM-based definition, with all other definitions of MDD in UKBiobank estimated using LDSC (Figure 3b shows the PCGCs estimates). The trend LDSC estimates show is similar to that from PCGCs, though both point estimates are higher across the board and standard errors are larger than that from PCGCs, making most estimates not significantly different from 1. This is consistent with the expectation that LDSC may over-estimate genetic correlations due to mishandling of covariates and overlap of samples[12].

Overestimates of genetic correlation between definitions of MDD (not significantly different from 1) may send the misleading message that the genetic architectures in different definitions of MDD, even those on opposite ends of the spectrum in terms of case criteria, are not significantly different from each other. This obscures important genetic differences between different definitions of MDD.

Stratification of definitions of MDD by environmental risk factors

We stratified samples from definitions of MDD in UKBiobank by environmental risk factors which we assessed the effects for, in the following ways. We estimate heritability in each strata of samples using PCGCs (Methods), and all results are shown in Supplemental Table S9.

1. by age: data field 21022; dividing individuals into those < 60 years old and those >= 60
2. by sex: data field 31; dividing individuals into males and females
3. by lifetime trauma: data category 145 for "Traumatic events reported within the on-line mental health questionnaire"; we calculated a weighted measure of traumatic events, dividing individuals into those who have weighted measure >= 2.5 and those with weighted measure < 2.5
4. by neuroticism: data field 20127 for neuroticism score; dividing individuals into those who score >= 5 and those who score < 5
5. by social deprivation: data field 189 for Townsend deprivation index (TDI); dividing individuals into those with TDI <0 and those with TDI >=0
6. by years of education: data field 845; dividing individuals into those with years of education >= 11 and < 11).

Using the heritability estimates and standard errors from PCGCs, we performed t-tests on the heritabilities from the strata in each stratification, and ask whether they are significantly different from 0 at $p < 0.05$, using a two-tailed test and correcting for 49 stratifications. None of the difference between strata, from any definition of MDD and stratification, is significant, though the method is underpowered as compared to other methods that requires the use of individual level genotypes[20].

Simulations of effects of misdiagnosis and misclassification on heritability estimates

If we assume DSM-based definitions of MDD in UKBiobank (LifetimeMDD), with its set of genetic factors contributing a particular fraction of liability to the disease (heritability), as the gold standard, the question is whether the other definitions of MDD (from minimal phenotyping) share the same genetic factors and their total contribution to the liability to the disease. If yes, do they simply contain milder forms of the disease (those with lower liability) that do not qualify for DSM-based definition of MDD, or do they contain misdiagnosis of those without the disease as cases? If no, are they cases of a genetically correlated disease that may be misclassified as MDD?

We perform simulations to show the effects of misdiagnosis and misclassification on heritability estimates, and

in turn, whether they may be the cause of lower heritability estimates in the minimal phenotyping definitions of MDD. We adopt the theoretical framework of the liability threshold model, where every individual has a normally distributed liability value for a trait such that case subjects of the trait are individuals whose liability exceeds a given cutoff[21]. The cutoff for the liability was determined as the 1-Ki percentile of the simulated liabilities, where K is what we set as the prevalence of cases. To simulate a biobank sample (which is assumed to be representative of the population), we do not ascertain for cases.

Using array genotypes at 344,184 SNPs (LD < 0.5, MAF > 5%, HWE P value > 10e-6) in 25,000 random individuals from UKBiobank who are White-British (Methods), we used LDAK[8,22] to simulate pairs of traits $y_{i,1}$ and $y_{i,2}$ with genetic correlation rG_i, where for each i in $i \in \{1..10\}$: 5000 causal SNPs are picked uniformly at random to contribute a total heritability for $y_{i,1}$ and $y_{i,2}$ of $h^2_{i,1} = h^2_{i,2} = h^2_i$ using the model $Y = \sum \beta_j X_j + e$, where β_j is the effect size of X_j, the jth causal SNP, and e is Gaussian-distributed noise. Effect size at each SNP X_j is sampled under the Model $\beta_j \sim \mathrm{N}\left(0, \left[f_j(1-f_j)\right]-1\right)$, where f_j is the MAF of X_j. The LDAK command we use is as follows:

ldak5.linux --make-phenos traitpair$i.h2$h2i.rg$rGi --bfile $bfile --ignore-weights YES --power -1 --num-causals 5000 --num-phenos 2 --her $h2i --bivar $rGi

Where $i \in \{1..10\}$, $h^2_i \in \{0.2, 0.4, 0.6, 0.8\}$, $rG_i \in \{0, 0.2, 0.4, 0.6, 0.8, 0.95\}$

In the first set of simulations, we show that including individuals with lower liabilities to a disease (trait) as cases (increasing prevalence of cases K_i through lowering the liability cutoff) would not have an effect on liability scale heritability estimates of the trait. We estimate the liability scale heritability of each trait using the --pcgc option in LDAK appropriate for binary traits, accounting for the prevalence of cases K_i used. As we both do not ascertain for cases, and do not simulate covariates with effects on the traits, heritability estimates from the simulations do not suffer from downward estimates in REML presented previously for case-control traits[12] when converted to the liability scale[6,21]. As such, using both --reml with --prevalence and --pcgc options give the same liability scale heritability estimates on the simulated traits. Supplemental Figure S5a shows the heritability estimates of traits with heritabilities $h^2_i \in \{0.2, 0.4, 0.6, 0.8\}$ for $i \in \{1..10\}$, where we shift the prevalence K_i from 0.1 to 0.5, and recover the simulated heritabilities exactly regardless of K_i. While this is well-known, it refutes the misconception that minimal phenotyping definitions of MDD have lower heritabilities than DSM-based definitions because they contain individuals with lower liabilities to MDD.

In the second set of simulations, we show that misdiagnosis of controls as cases lowers heritability. At a constant liability threshold corresponding to $K_i = 0.2$, if we identify all individuals above liability threshold correctly as cases, but identify random controls as cases such that the percentage of misdiagnosis cases ranges from 0% to

50% of all cases (thereby increasing apparent case prevalence), then liability scale heritability (corrected for the apparent prevalence) decrease as a result (Supplemental Figure S4b). This decrease is consistent with the high prevalence and low heritability we observe in minimal phenotyping based MDD in UKBiobank and 23andMe[17] (Figure 3a, Supplemental Figure S3a). We note that our simulations do not misidentify true cases as controls (in other words, we have a sensitivity of 100% for true cases, while having a false discovery rate of cases ranging from 0% to 50%); in realistic settings, it is possible sensitivity to true cases will decrease as collection criteria becomes less stringent. If we lowered the sensitivity, the heritability estimates are likely decrease further.

However, the above scenario must not be the only explanation for the lower heritability in minimal phenotyping based MDD. As genetic correlation between definitions of MDD should not differ if all genetic factors are shared and effects were only lower in some definitions due to noise (Figure 3b). Hence, we hypothesize minimal phenotyping definitions of MDD may in fact contain cases of other conditions that are misclassified as MDD, and simulate misclassification of cases from a genetically correlated disease to find out if it may result in a decrease in heritability consistent with the lowered heritability estimate in minimal phenotyping definitions of MDD.

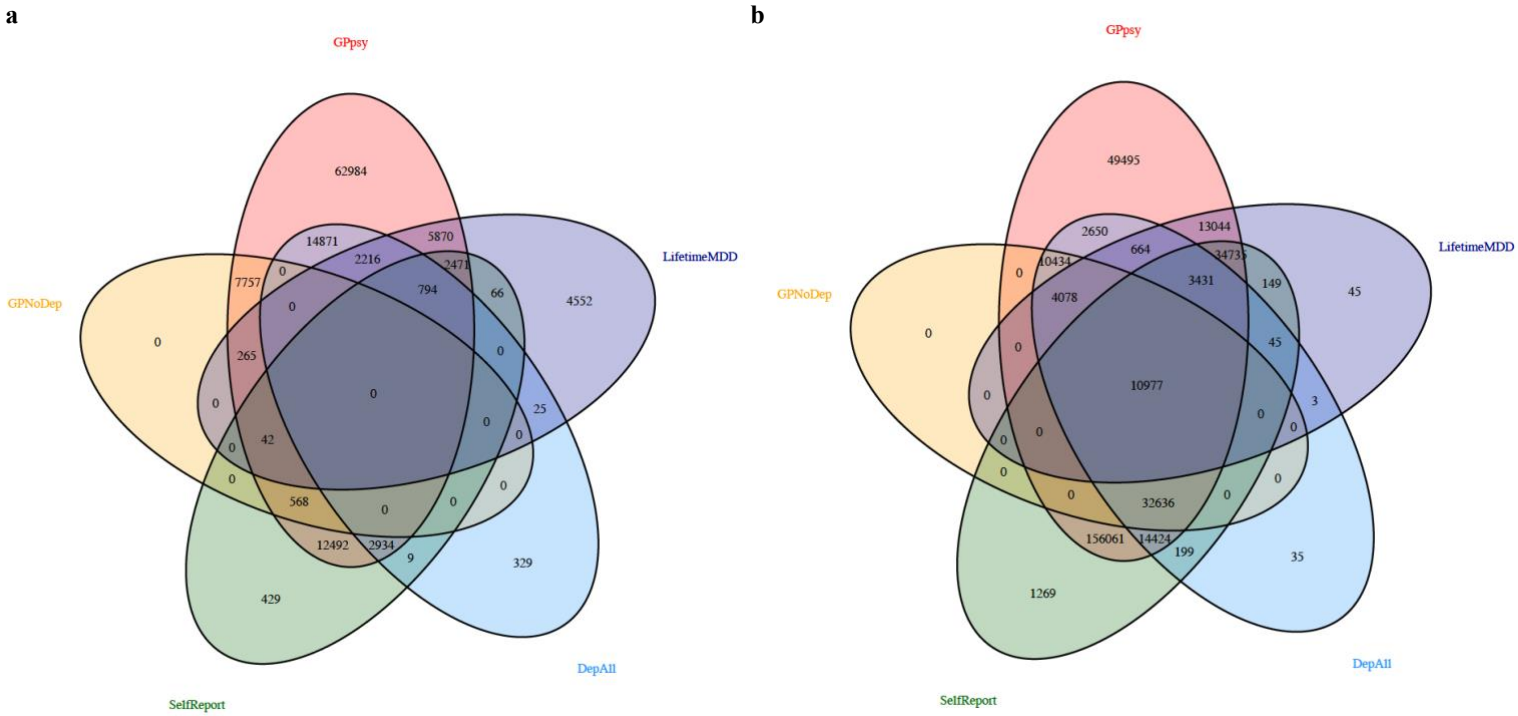
We first correctly identify all individuals above liability threshold for prevalence $K_{i,1} = 0.2$ in $y_{i,1}$ ($h^2_{i,1} = 0.4$) as cases of $y_{i,1}$. Then, we misclassify 10% to 50% of cases of genetically correlated $y_{i,2}$ ($h^2_{i,2} = 0.4$, prevalence $K_{i,2} = 0.2$) as cases of $y_{i,1}$, hence increasing apparent case prevalence of $y_{i,1}$. The two traits have genetic correlation $rG_i \in \{0,0.2,0.4,0.6,0.8,0.95\}$. As shown in Supplemental Figure S4c, we find that increasing the percentage of misidentification decreases liability scale heritability of $y_{i,1}$ (corrected for the apparent prevalence) if rG_i is small, but can inflate liability scale heritability if rG_i is large (>=0.6). This inflation decreases with increase in heritabilities of both traits (Supplemental Figure S5). This analysis shows that as long as two traits are completely heritable, and there are environmental contributions, misclassifying cases of one as the cases of the other would lead to erroneous estimates of heritability, even at very high genetic correlations.

As genetic correlation between definitions of MDD and other psychiatric conditions are low (maximum genetic correlation with neuroticism at a mean of 0.67 among all definitions of MDD, Supplemental Table S12, Figure 4), and genetic correlation between DSM-definition of MDD (LifetimeMDD) and those without MDD (GPNoDep, making up 30% cases in minimal phenotyping definition of MDD GPpsy) is 0.58 (se=0.078), most of the potential misclassification could have come from conditions with genetic correlation with MDD lower than 0.6 and led to downward bias of heritability observed in minimal definitions of MDD in UKBiobank.

In summary, our simulations are consistent with a model where the lower heritability in minimal phenotyping based MDD is not due to lowering liability threshold for case definition in MDD, but a consequence of

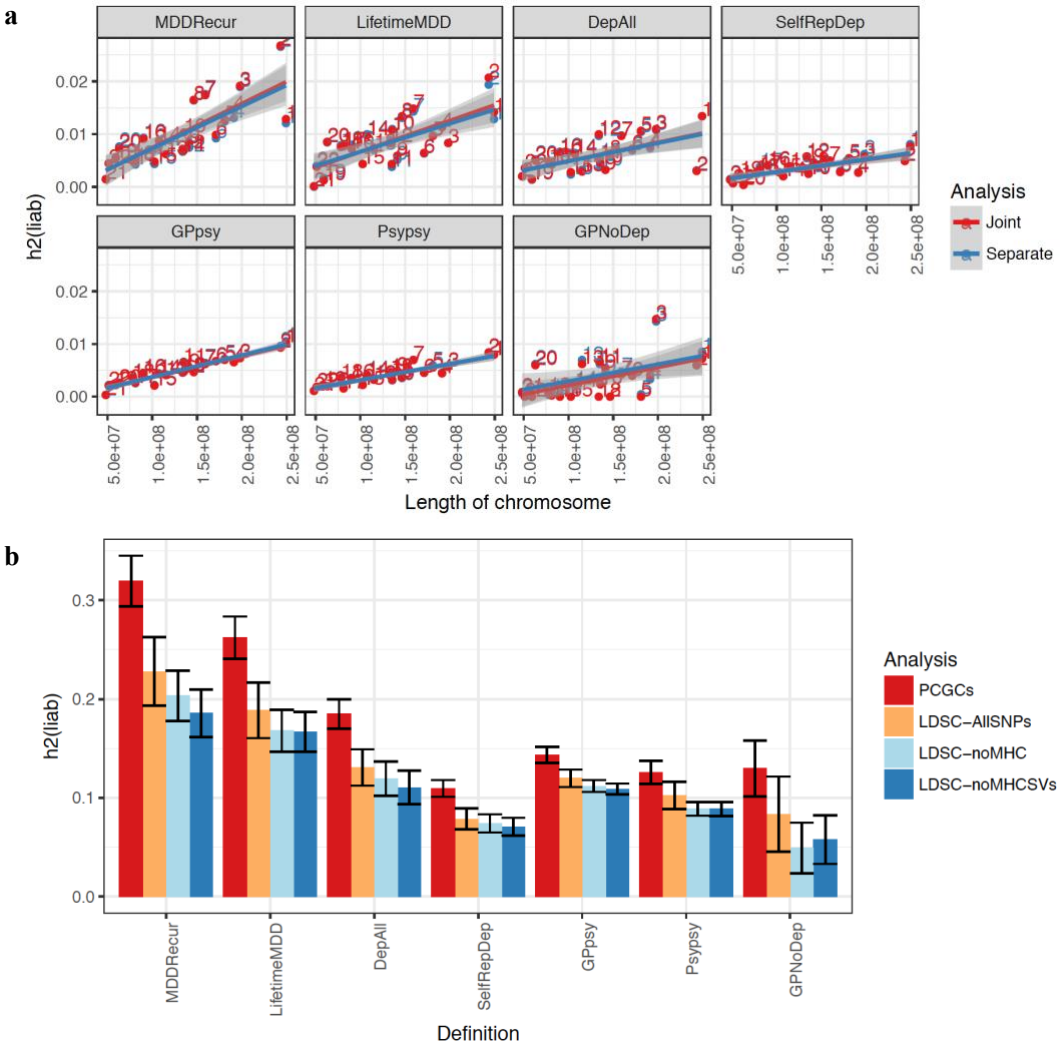
potentially both misdiagnosis of those who do not have MDD as cases, and misclassification of those with other conditions as cases (and the misidentification of true cases as controls, not simulated).

Supplemental Figures



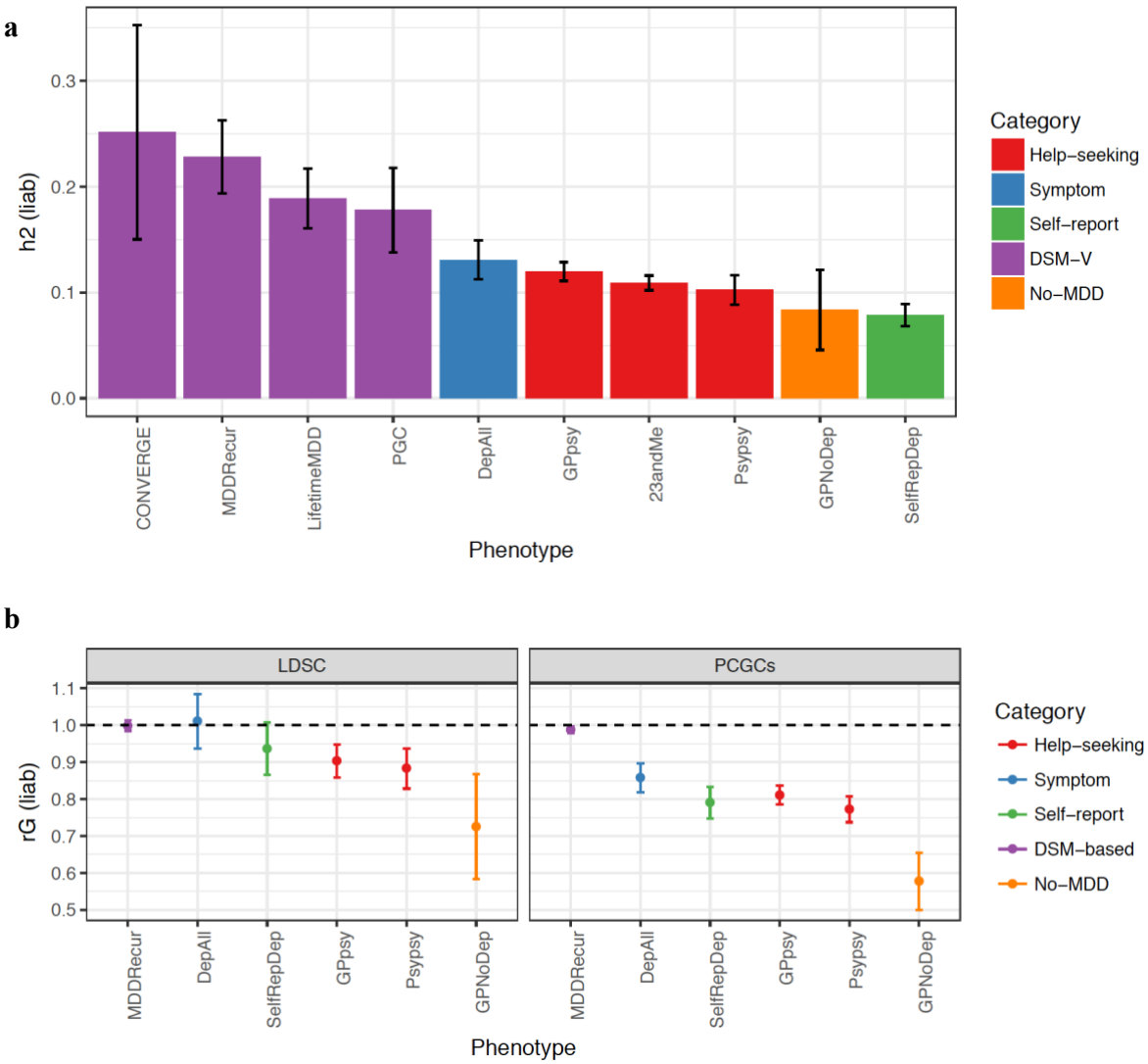
Supplemental Figure S1: Overlap between definitions of MDD in UKBiobank

a) This figure shows the number of overlapping cases between help-seeking (GPpsy in red), symptom (DepAll, in blue), self-report (SelfRepDep, in green), DSM-V (LifetimeMDD, in purple) based definitions of MDD, as well as the help-seeking, no-MDD definition that excludes MDD (GPNoDep, in orange). As not all individuals answered all questions necessary to assess whether they are a case or control in any of the definitions of MDD, we also show in b) the number of overlapping individuals (both cases and controls) who answered the question necessary to qualify as either cases or controls in each of the definitions of MDD (refer to main text and Figure 1 for details).



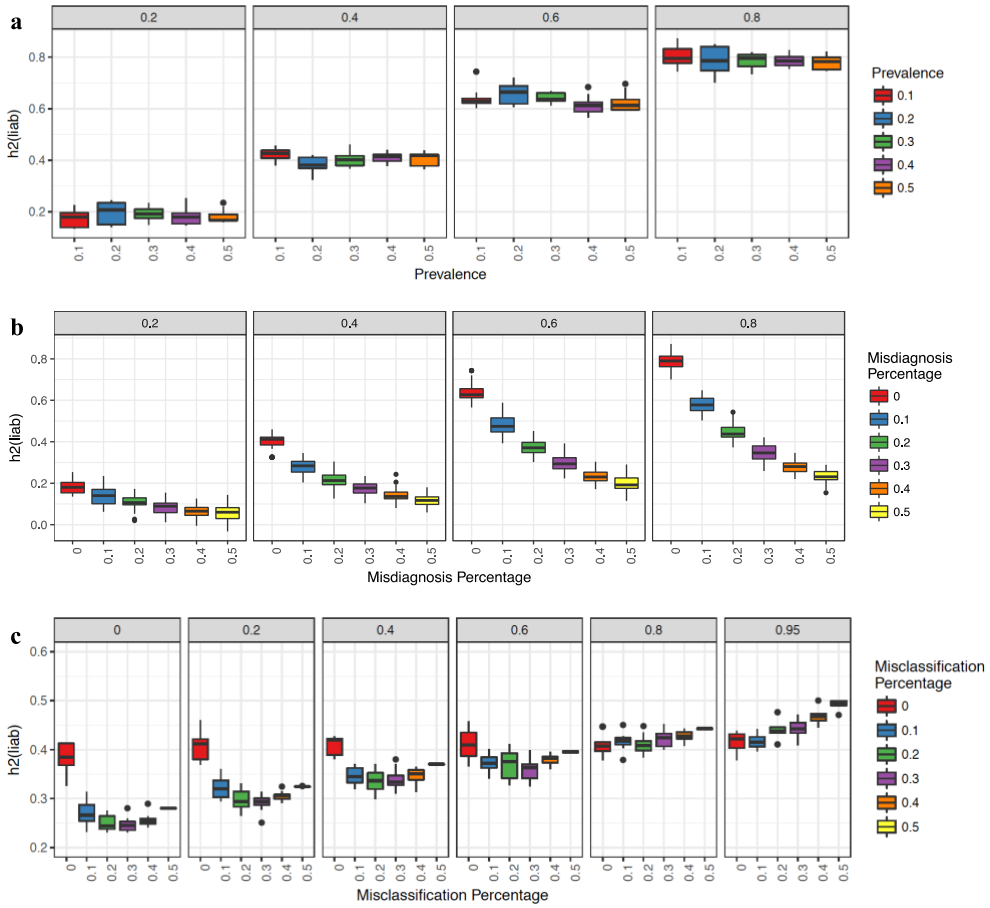
Supplemental Figure S2: Effects of population structure and MHC/SVs

a) This figure shows for each definition of MDD in UKBiobank the estimates of heritability (on the liability scale) contributed by each chromosome (obtained using BOLT-REML, see Supplemental Methods) estimated both jointly (in red) and separately (in blue), plotted against the lengths of the chromosomes, using both PCs obtained from the whole White-British cohort in UKBiobank. b) This figure shows for each definition of MDD in UKBiobank estimtates of heritability (on the liability scale) from PCGCs (using all SNPs > 5% MAF, in red), which specifically models the binary structure of disease traits and contribution from covariates accordingly, and that from LDSC performed on summary statistics from logistic regression between with covariates all SNPs in GWAS and each definition of MDD (what is commonly used in analysis): LDSC-AllSNPs (using all SNPs > 5% MAF, in orange), as well as LDSC-noMHC (using all SNPs > 5% MAF except those in the MHC region on chromosome 6:25-35MB, in light blue), and LDSC-noMHCSVs (using all SNPs > 5% MAF except those in the MHC region and SVs[5], in dark blue).



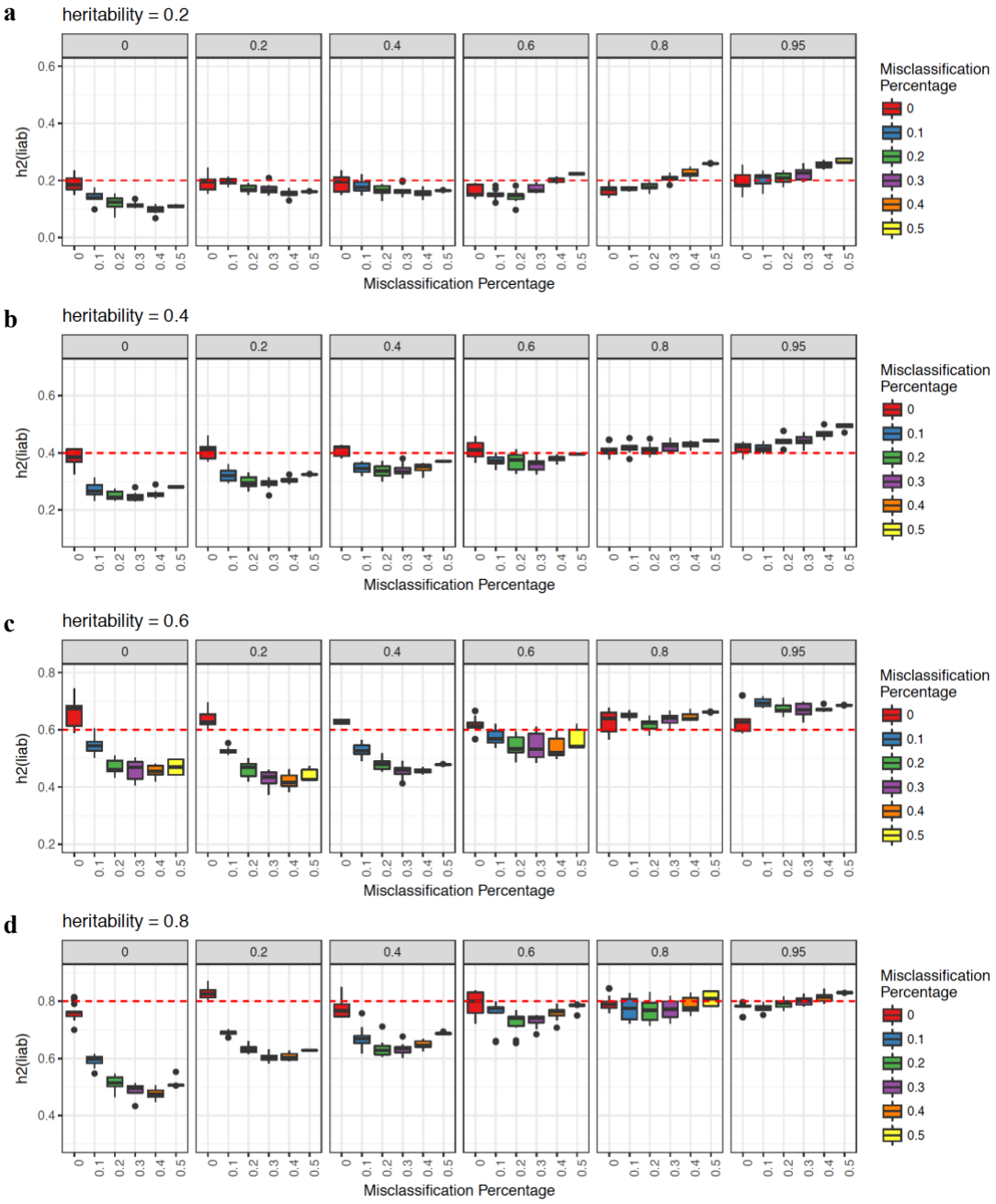
Supplemental Figure S3: LDSC estimates of heritability and genetic correlation

a) This figure shows the heritability estimates of definitions of MDD in UKBiobank from LDSC using logistic regression summary statistics on all SNPs > 5% MAF (Methods). In the figure we also show LDSC estimates of heritabilities of previous studies of MDD including CONVERGE[15], PGC1-MDD[18], and 23andMe[17], where we show DSM-based definitions in UKBiobank (in purple) show similar estimates to CONVERGE, symptom-based definitions in UKBiobank (in blue) are similar to PGC1-MDD, while help-seeking based (in red) definitions in UKBiobank are similar to 23andMe. b) This figure shows the genetic correlation between all definitions of MDD in UKBiobank against a DSM-based definition of MDD (LifetimeMDD) estimated in LDSC and PCGCs (duplicated from Figure 3b for comparison).



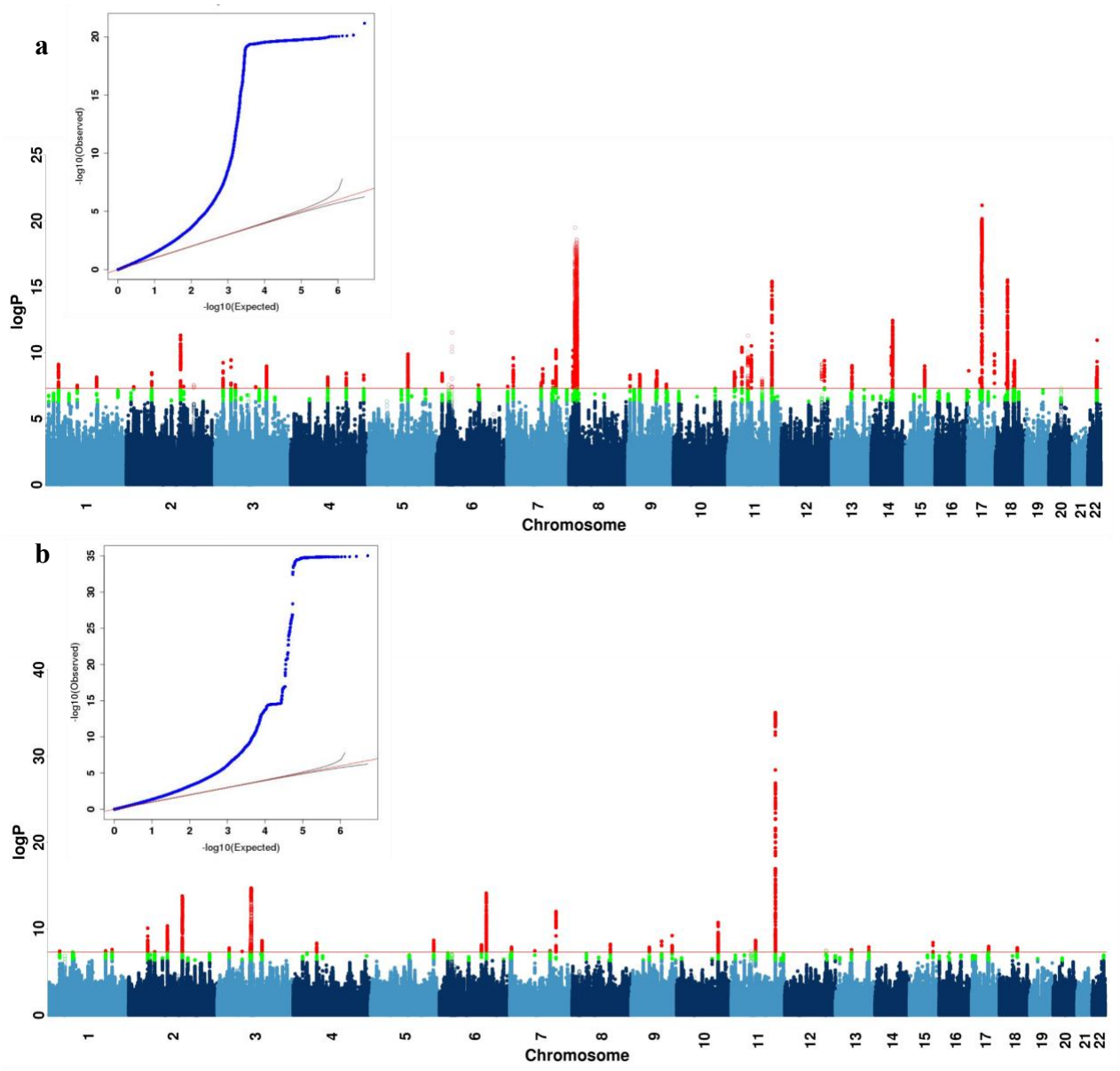
Supplemental Figure S4: Simulations of misdiagnosis and misclassification

a) This figure shows the estimated heritability (using --pcgc option with --prevalence K in LDAK, plotted on the y axis) of binary traits y_i, where $i \in \{1..10\}$ with simulated heritabilities $h_i^2 \in \{0.2, 0.4, 0.6, 0.8\}$ (in the grey bars above each panel) plotted against the prevalence $K_i \in \{0.1, 0.2, 0.3, 0.4, 0.5\}$ of cases (defined as those individuals with liabilities greater than 1-Kth percentile). b) This figure shows the estimated heritability (using --pcgc option with --prevalence K_i in LDAK, plotted on the y axis) of binary traits y_i, where $i \in \{1..10\}$ with simulated heritabilities $h_i^2 \in \{0.2, 0.4, 0.6, 0.8\}$ (in the grey bars above each panel), for each of which all cases (at prevalence K_i =0.2) are correctly identified as cases, while varying numbers of random controls are misdiagnosed as cases (plotted on the x axis as percentage of all identified cases). c) This figure shows the estimated heritability (using --pcgc option with --prevalence K in LDAK, plotted on the y axis) of binary traits $y_{i,1}$, where $i \in \{1..10\}$ with simulated heritability $h_{i,1}^2 = 0.4$, for each of which all cases at prevalence $K_{i,1} = 0.2$ are correctly identified as cases, while varying numbers of misclassifications from cases of a genetically correlated binary trait $y_{i,2}$, where $i \in \{1..10\}$ of equal heritability and prevalence as cases of $y_{i,1}$. Genetic correlations between $y_{i,1}$ and $y_{i,2}$ $rG_i \in \{0,0.2,0.4,0.6,0.8,0.95\}$ are shown in the grey bars above each panel.



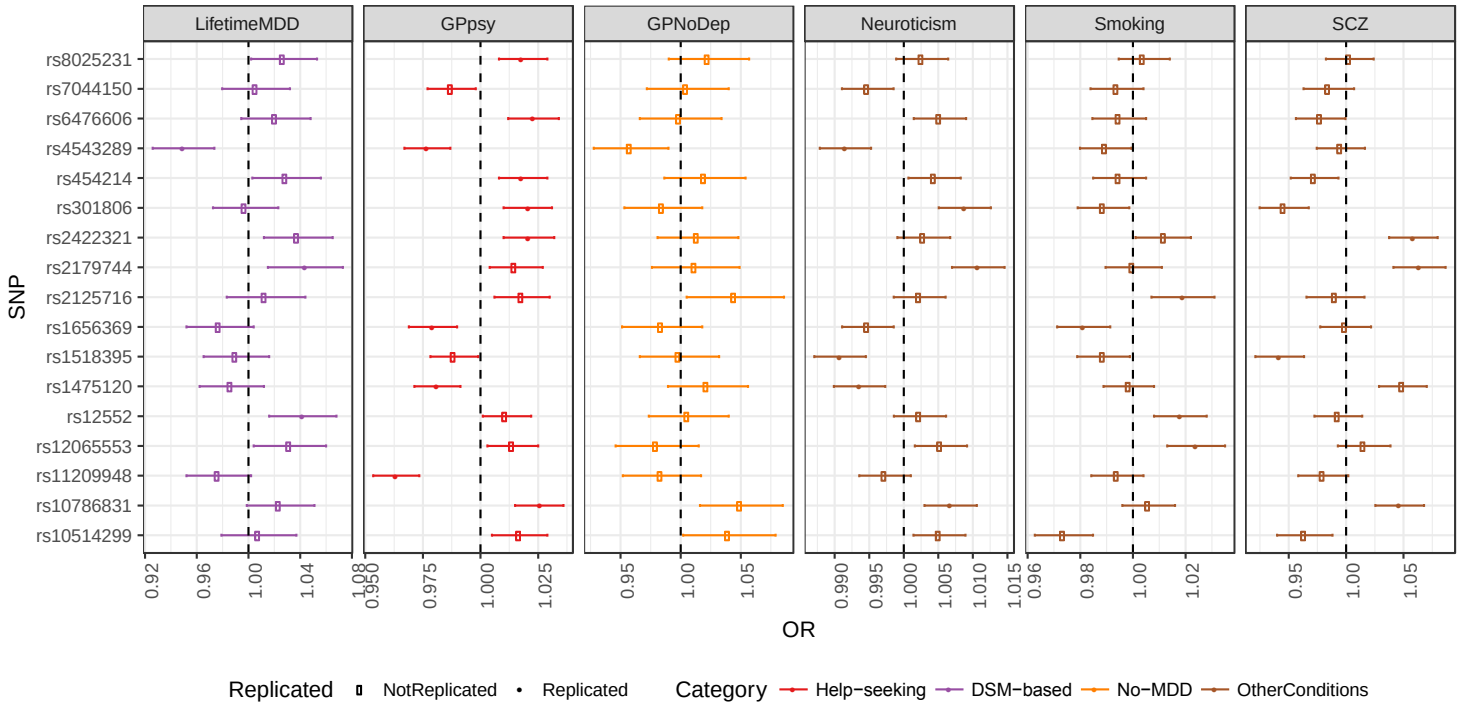
Supplemental Figure S5: Simulations of misclassification at different heritabilities

a-d) These figures shows the estimated heritability (using --pcgc option with --prevalence K in LDAK, plotted on the y axis) of binary traits ($y_{i,1}$, where i ∈ {1..10}) with simulated heritability 0.2, 0.4, 0.6, and 0.8, for each of which all cases (at prevalence $K_{i,1}$=0.2) are correctly identified as cases, while varying numbers of cases misclassified from a genetically correlated binary trait ($y_{i,2}$, where $i \in \{1..10\}$) of equal heritability and prevalence as cases of $y_{i,1}$. Genetic correlations between $y_{i,1}$and $y_{i,2}$ ($rG_i \in \{0, 0.2, 0.4, 0.6, 0.8, 0.95\}$) are shown in the grey bars above each panel.



Supplemental Figure S6: GWAS on neuroticism and smoking in UKBiobank

This figure shows the manhattan plot of neuroticism score (data field 20127, quantitative trait from 0 to 12) in 274,107 individuals and ever smoked status (data field 20160, binary trait of 0 for "No", and 1 for "Yes") in 336,066 individuals in UKBiobank using linear regression on all 8,968,716 common SNPs (MAF > 5% in all 337,198 White-British, unrelated samples) for all the above analyses in PLINK (version 1.9) [23] with 20 PCs and genotyping array as covariates. We report all associations with P values smaller than 5×10^{-8} as genome-wide significant (red). We indicated the SNPs in SVs and the MHC in all manhattan plots as hollow points instead of solid points due to lack of control for population structure in these regions, and show all top SNPs within peaks (1MB regions) in Supplemental Tables S10-11.



Supplemental Figure S7: GWAS hits from 23andMe are not specific to MDD

a) This figure shows how many out of at 17 loci significantly associated with help-seeking based definitions of MDD in 23andMe[17] show significant effects in DSM (LifetimeMDD, in purple), help-seeking (GPpsy, in red), and no-MDD help-seeking (GPNoDep, in orange) definitions of MDD, as well as risk factors and conditions other than MDD: neuroticism, smoking and SCZ[24] (all in brown). b) This figure shows the odds ratios (ORs) at 17 loci significantly associated with help-seeking based definitions of MDD in 23andMe[17], 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 17 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 are shown in Supplemental Table S15.

Supplemental Table

ABBREVIATION	PHENOTYPE	COLLECTION STRATEGY	STUDY TYPE	REFERENCE	N	SAMPLE PREV	POPULATION PREV	PREVALENCE REFERENCE
CONVERGE	**MDD**	Hospital based psychiatrist diagnosis	case-ascertained, screened-controls	CONVERGE Consortium, 2015[15]	10640	0.50	0.08	CONVERGE Consortium, 2015[15]
PGC1	**MDD**	Structured telephone interviews	case-ascertained, screened-controls (mega-analysis)	MDD Working Group of the Psychiatric Genomics Consortium, 2013[18]	18759	0.50	0.15	Kessler et al., 2003[25]
23andMe	**MDD**	Minimal phenotyping: self-report via questionaire	unascertained population cohort	Hyde et al., 2016[17]	307354	0.25	0.25	Hyde et al., 2016[17]
ADHD	**ADHD**	Hospital based psychiatrist diagnosis	case-ascertained, screened-controls (mega-analysis)	Demontis et al., 2017[26]	53293	0.36	0.05	Demontis et al., 2017[26]

BIP	**BIP**	Hospital based psychiatrist diagnosis	case-ascertained, 33% screened-controls (mega-analysis)	Bipolar Disorder Working Group of the Psychiatric Genomics Consortium, 2011[27]	16731	0.45	0.025	Merikangas et al., 2011[28]
SCZ	**SCZ**	Hospital based psychiatrist diagnosis /semi-structured interviews	case-ascertained, population-controls; trios (mega-analysis)	Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014[24]	150064	0.25	0.01	Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014[24]
AUT	**AUT**	Hospital based psychiatrist diagnosis	trios; case-ascertained, population-controls (mega-analysis)	Cross-Disorder Group of the Psychiatric Genomics Consortium, 2013[29]	10610	0.50	9.00E-04	Cross-Disorder Group of the Psychiatric Genomics Consortium, 2013[29]

Supplemental Table S1: Other studies referenced in this paper

This table shows the studies from which we obtain GWAS summary statistics for MDD and other psychiatric conditions that we reference in this paper. The columns are abbreviations of the studies used in this paper, the disease phenotype in question, their case collection strategy, their study type where known, the study reference, number of samples involved, case prevalence in sample, population prevalence of disease phenotypes, and reference for population prevalence of disease prevalence if different from the reference of the study itself.

DSM CRITERIA	DATA FIELD	QUESTION ON LIFETIME MDD	QUALIFYING ANSWERS	N_YES	N_NO
A1	20446	Have you ever had a time in your life when you felt sad, blue, or depressed for two weeks or more in a row?	Yes	59409	50018
A2	20441	Have you ever had a time in your life lasting two weeks or more when you lost interest in most things like hobbies, work, or activities that usually give you pleasure?	Yes	42565	66830
A3	20536	Did you gain or lose weight without trying, or did you stay about the same weight?	Gained weight, Lost weight, Both gained and lost some weight during the episode	30875	21324
A4	20532	Did your sleep change?	Yes	41955	10714
A6	20449	Did you feel more tired out or low on energy than is usual for you?	Yes	45034	10072
A7	20450	People sometimes feel down on themselves, no good, worthless. Did you feel this way?	Yes	28865	28472
A8	20435	Did you have a lot more trouble concentrating than usual?	Yes	42568	11592
A9	20437	Did you think a lot about death - either your own, someone else's or death in general?	Yes	29946	27856
DSM CRITERIA	**DATA FIELD**	**QUESTION ON CURRENT MDD: Over the last 2 weeks, how often have you been bothered by any of the following problems?**	**QUALIFYING ANSWERS**	**N_YES**	**N_NO**
A1	20510	Feeling down, depressed, or hopeless	Nearly every day	1449	108261

A2	20514	Little interest or pleasure in doing things	Nearly every day	1670	108040
A3	20511	Poor appetite or overeating	Nearly every day	2633	107077
A4	20517	Trouble falling or staying asleep, or sleeping too much	Nearly every day	8616	101094
A5	20518	Moving or speaking so slowly that other people could have noticed? Or the opposite - being so fidgety or restless that you have been moving around a lot more than usual	Nearly every day	650	109060
A6	20519	Feeling tired or having little energy	Nearly every day	5968	103742
A7	20507	Feeling bad about yourself or that you are a failure or have let yourself or your family down	Nearly every day	2166	107544
A8	20508	Trouble concentrating on things, such as reading the newspaper or watching television	Nearly every day	1769	107941
A9	20513	Thoughts that you would be better off dead or of hurting yourself in some way	Nearly every day	357	109353

Supplemental Table S2: DSM-based criteria for lifetime and current MDD

This table lists questions in the online mental health follow up questionnaire in UKBiobank forming the DSM-V diagnostic criteria for lifetime MDD[3] (Question on lifetime MDD) and current MDD (Question on current MDD), along with their data field in UKBiobank data showcase (datafield), the DSM criteria they correspond to, and the multiple-choice answer to each of the questions which indicates a subject as self-reporting having the sympom.

CENTRE	DEFINITION OF DEPRESSION (N_CASES/N_CONTROLS)						
	DSM-BASED		SYMPTOM	SELF-REPORT	HELP-SEEKING		NON-MDD
	LifetimeMDD	MDDRecur	DepAll	SelfRepDep	GPpsy	Psypsy	GPNoDep
Barts	399/1075	274/1040	NA	292/3623	1941/3795	1025/4728	NA
Birmingham	869/2333	563/2239	3198/7995	816/10902	5536/10139	1922/13782	1186/6769
Bristol	1613/5047	1001/4880	1649/4643	1534/21046	9817/20411	2960/27335	655/3972
Bury	826/2436	512/2347	NA	1506/15186	6971/13005	2118/17939	NA
Cardiff	554/1660	361/1599	NA	603/8663	4357/8239	1345/11285	NA
Croydon	801/2937	500/2842	2744/7922	811/10644	4706/9985	1791/12915	1074/6821
Edinburgh	568/2092	336/2033	NA	551/7368	3851/8325	1305/10889	NA
Glasgow	480/1605	309/1565	NA	771/8505	4550/7961	1553/10998	NA
Hounslow	743/3060	472/2954	2489/7862	619/9990	4401/9948	1741/12642	1055/6775
Leeds	1503/4312	932/4146	NA	2058/20840	10701/20079	3379/27457	NA
Liverpool	1080/3024	700/2916	2651/7765	1609/16421	7792/14444	2377/19930	1221/6500
Manchester	469/1282	321/1218	NA	821/6855	3244/5654	1091/7818	NA
Middlesborough	788/2175	504/2093	2934/8017	826/10477	5319/9595	1447/13520	1239/6727
Newcastle	1165/3363	746/3246	NA	1355/18166	9120/16690	2836/23035	NA
Nottingham	1208/3541	744/3420	676/1929	1325/17305	8178/15766	2515/21492	281/1639
Oxford	455/1695	274/1639	NA	656/6793	3236/6577	1148/8690	NA
Reading	999/4126	614/3995	NA	947/13561	6298/14662	1950/19033	NA
Sheffield	1125/3218	731/3081	4370/11248	1664/16219	7656/13696	2070/19329	1763/9436
Stockport_pilot	45/90	27/87	NA	33/247	108/207	30/287	NA
Stoke	485/1519	299/1474	NA	874/9823	4700/8913	1443/12210	NA
Swansea	93/228	59/221	360/762	107/1147	609/964	188/1385	127/632
Wrexham	33/52	23/48	106/255	27/333	169/307	52/427	31/222

Supplemental Table S3: Number of samples by assessment centre

This table lists the number of cases (N_cases) and controls (N_controls) collected for each of the definitions in UKBiobank in the categories DSM-based, Symptom, Self-report, Help-seeking, and no-MDD collected in each assessment centre. “NA” denotes no collection of answers to questions that form diagnostic criteria to the relevant definition of MDD in an assessment centre.

CATEGORY	DEFINITION	NSAMPLES	PREV	CORRECTED_PREV
DSM-based	LifetimeMDD	67171	0.243	0.268
DSM-based	MDDRecur	59385	0.173	0.196
Symptom	DepAll	79576	0.266	0.279
Self-report	SelfRepDep	253926	0.078	0.084
Help-seeking	GPpsy	332629	0.341	0.343
Help-seeking	Psypsy	333419	0.109	0.111
Non-MDD	GPNoDep	58125	0.149	0.146

Supplemental Table S4: Prevalence of each definition of MDD

This table shows for each definition of MDD in UKBiobank the category we assign it to (DSM-based, Symptom, Self-report, Help-seeking, or no-MDD), the number of samples (both cases and controls), the case prevalence (prev), and the case prevalence when corrected for age group, sex and population in each assessment centre where samples are collected. Correction factors for age, sex and population in each assessment centre are shown in Supplemental Tables S5,6.

CENTRE	UKBIOBANK		CENSUS_2011		CORRECTION
	N_SAMPLE	FRACTION	POPULATION	POP_FRACTION	
Cardiff	12678	0.038	346090	0.044	1.151
Sheffield	21479	0.064	552698	0.070	1.085
Hounslow	14439	0.043	253957	0.032	0.742
Leeds	30964	0.093	751485	0.095	1.023
Croydon	14778	0.044	363378	0.046	1.037
Birmingham	15783	0.047	1073045	0.135	2.866
Bristol	30406	0.091	428234	0.054	0.594
Nottingham	24106	0.072	305680	0.039	0.535
Reading	21050	0.063	155698	0.020	0.312
Liverpool	22399	0.067	466415	0.059	0.878
Newcastle	25963	0.078	280177	0.035	0.455
Stoke	13705	0.041	249008	0.031	0.766
Middlesborough	15024	0.045	138412	0.017	0.388
Manchester	8949	0.027	503127	0.063	2.370
Bury	20138	0.060	185060	0.023	0.387
Glasgow	12589	0.038	593245	0.075	1.987
Oxford	9865	0.029	151906	0.019	0.649
Edinburgh	12224	0.037	476626	0.060	1.644
Barts	5779	0.017	7375	0.001	0.054
Wrexham	481	0.001	134844	0.017	11.819
Stockport_pilot	317	0.001	283275	0.036	37.675
Swansea	1583	0.005	239023	0.030	6.366

Supplemental Table S5: Correction for population in assessment centres

This stable shows the number of samples (N_sample) and the fraction of total sample size (fraction) from each assessment centre in UKBiobank, as well as the population in cities where each of the collection centres are located (population) and the fraction of the UK population they represent (pop_fraction) according to the UK Census 2011. Only cities where there are UKBiobank assessment centres are included in this table and considered when calculating fractions of the UK population. The correction factor (correction) is the ratio between the population fraction and sample fraction, and is multiplied to the case prevalence of each definition of MDD from each assessment centre when calculating the corrected prevalence for each definition of MDD across all assessment centres, along with correction factors for age and sex as shown in Supplemental Table S6.

AGE	UKBIOBANK			CENSUS_2011		CORRECTION	
	N_SAMPLE	% AGE	% FEMALE	% AGE	% FEMALE	AGE	FEMALE
39to44	31412	9.4	53.3	17.6	50.1	1.875	0.940
45to49	41684	12.5	55.3	17.7	50.3	1.421	0.909
50to54	49730	14.9	56.3	15.6	50.6	1.050	0.898
55to59	60517	18.1	55.4	13.8	50.6	0.763	0.913
60to64	84850	25.4	53.7	14.5	50.4	0.572	0.938
above65	66506	19.9	49.6	26.0	50.8	1.308	1.023

Supplemental Table S6: Correction for age and sex

This stable shows the number of samples (N_sample) and the fraction of total sample size of each age group (% age) as well as fraction of females (% female) per age group in UKBiobank, as well as the fraction of the UK population from each age group (% age) and who are female (% female) per age group according to the UK Census 2011. Only age groups of those ages of samples included in UKBiobank are shown in this table and considered when calculating fractions of each age group and sex. Correction factors (age, sex) are the ratios between the population fraction and sample fraction of each age group and sex, and are multiplied to the case prevalence of each definition of MDD along with correction factors for assessment centre population as shown in Supplemental Table S5 when calculating the corrected prevalence for each definition of MDD.

DATAFIELD	QUESTION ON CHILDHOOD TRAUMA "When I was growing up…"	RESPONSE AND SCORE (0/1)						OR for LifetimeMDD
		Prefer not to answer	Never true	Rarely true	Sometimes true	Often true	Very often true	
20487	I felt that someone in my family hated me	NA	0	0	0	1	1	1.799
20488	People in my family hit me so hard that it left me with bruises or marks	NA	0	0	1	1	1	1.215
20489	I felt loved	NA	1	1	1	0	0	1.664
20490	Someone molested me (sexually)	NA	0	1	1	1	1	1.174
20491	There was someone to take me to the doctor if I needed it	Question omitted						

DATAFIELD	QUESTION ON ADULTHOOD TRAUMA "Since I was sixteen…"	RESPONSE AND SCORE (0/1)						OR for LifetimeMDD
		Prefer not to answer	Never true	Rarely true	Sometimes true	Often true	Very often true	
20521	A partner or ex-partner repeatedly belittled me to the extent that I felt worthless	NA	0	0	1	1	1	2.388
20522	I have been in a confiding relationship	NA	1	1	0	0	0	1.039
20525	There was money to pay the rent or mortgage when I needed it	NA	1	1	1	0	0	1.311
20523	A partner or ex-partner deliberately hit me or used violence in any other way	Question omitted						
20524	A partner or ex-partner sexually interfered with me, or forced me to have sex against my wishes	Question omitted						

DATAFIELD	QUESTION FOR LIFETIME TRAUMA "In your life, have you...?"	RESPONSE AND SCORE (0/1)				OR for LifetimeMDD
		Prefer not to answer	Never	Yes, but not in the last 12 months	Yes, within the last 12 months	
20526	Been in a serious accident that you believed to be life-threatening at the time	NA	0	1	1	1.326
20527	Been involved in combat or exposed to a war-zone (either in the military or as a civilian)	Question omitted				
20528	Been diagnosed with a life-threatening illness	NA	0	1	1	1.275
20529	Been attacked, mugged, robbed, or been the victim of a physically violent crime	NA	0	1	1	1.145
20530	Witnessed a sudden violent death (eg. murder, suicide, aftermath of an accident)	NA	0	1	1	1.288
20531	Been a victim of a sexual assault, whether by a stranger or someone you knew	NA	0	1	1	1.339

Supplemental Table S7: Derivation of lifetime trauma score

This table lists questions in the online mental health follow up questionnaire in UKBiobank forming the derivation of lifetime trauma score used in this paper (Supplemental Methods), along with their data field in the UKBiobank data showcase. The questions come in three categories, "Questions on childhood trauma", "Questions on adulthood trauma", and "Question for lifetime trauma". Answers to questions that are of high similarity with another question are excluded from the calculation of lifetime trauma score (Question omitted). We show the multiple-choice answer to each of the question which indicates a subject as self-reporting having experienced the traumatic life event. We also show the odds ratio (OR) of having experienced each of the traumatic life event on DSM-based LifetimeMDD, when modelled jointly with all other traumatic life events, as well as age, sex, years of education, neuroticism, townsend deprivation index, experience of any traumatic life event in the past 2 years (data field 6145), and total number of traumatic life events reported. We then weight each of the traumatic life events by its OR and sum all weighted scores to arrive at a weighted lifetime trauma score for each individual.

PHENOTYPE				PCGCs				LDSC		BOLT-REML	
CATEGORY	**DEFINITION**	**PREV**	**N_SAMPLES**	**H2**	**SE**	**GENVAR**	**GENVAR_SE**	**H2**	**SE**	**H2**	**SE**
DSM-based	LifetimeMDD	0.243	67171	0.262	0.022	0.263	0.022	0.189	0.028	0.184	0.012
DSM-based	MDDRecur	0.173	59385	0.320	0.026	0.321	0.026	0.228	0.035	0.218	0.016
Symptom	DepAll	0.266	79576	0.185	0.015	0.185	0.015	0.131	0.018	0.132	0.010
Self-report	SelfRepDep	0.078	253926	0.110	0.009	0.110	0.009	0.079	0.010	0.077	0.006
Help-seeking	GPpsy	0.341	332629	0.143	0.008	0.144	0.008	0.120	0.009	0.109	0.003
Help-seeking	Psypsy	0.109	333419	0.128	0.012	0.126	0.012	0.103	0.014	0.090	0.004
Non-MDD	GPNoDep	0.149	58125	0.130	0.028	0.130	0.028	0.084	0.038	0.052	0.015

Supplemental Table S8: Heritability estimates of definitions of MDD from different methods

This table shows the heritability estimates of each definition of MDD calculated using three different methods, PCGCs, LDSC, and BOLD-REML, as detailed in Supplemental Methods. For each definition of MDD we show the sample size (N_sample) and prevalence (prev), and for each method we show the heritability estimate (h2) and its standard error (se). For PCGCs, we also show the genetic variance (genvar) and its standard error (genvar_se), the comparable measure to heritability from the other two methods (Supplemental Methods).

RISK FACTOR	PHENOTYPE		N_SAMPLE (PREV)	H2 (SE)	N_SAMPLE (PREV)	H2 (SE)	DIFFERENCE			
AGE	CATEGORY	DEFINITION	AGE >= 60		AGE < 60		ΔH2	DF	T-TEST STAT	T-TEST PVALUE
	DSM-based	LifetimeMDD	27671 (0.166)	0.298 (0.039)	39500 (0.297)	0.312 (0.027)	0.015	52168	-0.310	3.78E-01
	DSM-based	MDDRecur	24894 (0.106)	0.256 (0.048)	34491 (0.222)	0.387 (0.038)	0.132	51023	-2.142	1.61E-02
	Symptom	DepAll	38683 (0.224)	0.215 (0.027)	40893 (0.306)	0.232 (0.022)	0.017	76179	-0.503	3.07E-01
	Self-report	SelfRepDep	126331 (0.058)	0.141 (0.016)	127595 (0.098)	0.129 (0.013)	-0.012	243413	0.587	2.79E-01
	Help-seeking	GPpsy	150368 (0.318)	0.151 (0.014)	182261 (0.359)	0.153 (0.008)	0.002	233073	-0.143	4.43E-01
	Help-seeking	Psypsy	150840 (0.104)	0.150 (0.015)	182579 (0.113)	0.124 (0.012)	-0.026	301479	1.303	9.63E-02
	No-MDD	GPNoDep	29866 (0.152)	0.208 (0.044)	28259 (0.145)	0.153 (0.041)	-0.055	58049	0.920	1.79E-01
DEPRIVATION	CATEGORY	DEFINITION	TOWNSEND DEPRIVATION INDEX >= 0		TOWNSEND DEPRIVATION INDEX < 0		ΔH2	DF	T-TEST STAT	T-TEST PVALUE
	DSM-based	LifetimeMDD	13969 (0.304)	0.418 (0.059)	53120 (0.226)	0.252 (0.025)	-0.166	19489	2.596	4.71E-03
	DSM-based	MDDRecur	12308 (0.239)	0.434 (0.073)	47013 (0.156)	0.338 (0.032)	-0.096	17400	1.212	1.13E-01
	Symptom	DepAll	20435 (0.309)	0.325 (0.038)	59023 (0.251)	0.189 (0.018)	-0.136	30436	3.264	5.50E-04
	Self-report	SelfRepDep	65877 (0.099)	0.178 (0.023)	187750 (0.071)	0.119 (0.011)	-0.059	94849	2.297	1.08E-02
	Help-seeking	GPpsy	83671 (0.390)	0.149 (0.011)	248566 (0.324)	0.145 (0.008)	-0.004	182828	0.309	3.79E-01

	Help-seeking	Psypsy	83835 (0.149)	0.154 (0.016)	249190 (0.095)	0.119 (0.011)	-0.034	172056	1.724	4.23E-02
	No-MDD	GPNoDep	14046 (0.168)	0.311 (0.071)	44004 (0.142)	0.143 (0.029)	-0.168	18912	2.203	1.38E-02
TRAUMA	**CATEGORY**	**DEFINITION**	**LIFETIME STRESS SCORE >= 2.5**		**LIFETIME STRESS SCORE < 2.5**		**ΔH2**	**DF**	**T-TEST STAT**	**T-TEST PVALUE**
	DSM-based	LifetimeMDD	19110 (0.424)	0.401 (0.047)	48057 (0.171)	0.232 (0.025)	-0.169	30581	3.185	7.24E-04
	DSM-based	MDDRecur	16139 (0.353)	0.468 (0.051)	47013 (0.156)	0.338 (0.032)	-0.130	35988	2.849	2.20E-03
	Symptom	DepAll	8448 (0.408)	0.344 (0.076)	20544 (0.219)	0.295 (0.049)	-0.048	15766	0.535	2.96E-01
	Self-report	SelfRepDep	26454 (0.117)	0.220 (0.047)	54827 (0.057)	0.252 (0.032)	0.032	51237	-0.550	2.91E-01
	Help-seeking	GPpsy	33451 (0.447)	0.233 (0.028)	75790 (0.277)	0.169 (0.015)	-0.064	53301	1.998	2.28E-02
	Help-seeking	Psypsy	33502 (0.160)	0.203 (0.033)	75929 (0.072)	0.193 (0.031)	-0.011	89536	0.238	4.06E-01
	No-MDD	GPNoDep	4980 (0.186)	0.759 (0.181)	16007 (0.119)	0.308 (0.065)	-0.452	6333	2.349	9.43E-03
NEUROTICISM	**CATEGORY**	**DEFINITION**	**NEUROTICISM >= 5**		**NEUROTICISM < 5**		**ΔH2**	**DF**	**T-TEST STAT**	**T-TEST PVALUE**
	DSM-based	LifetimeMDD	19043 (0.439)	0.381 (0.043)	38003 (0.136)	0.289 (0.030)	-0.092	37284	1.746	4.04E-02
	DSM-based	MDDRecur	16148 (0.370)	0.451 (0.054)	34367 (0.072)	0.315 (0.045)	-0.136	37543	1.948	2.57E-02
	Symptom	DepAll	25691 (0.420)	0.212 (0.030)	41428 (0.171)	0.183 (0.027)	-0.029	58976	0.711	2.39E-01
	Self-report	SelfRepDep	87606 (0.141)	0.106 (0.013)	117206 (0.027)	0.194 (0.032)	0.088	150998	-2.567	5.12E-03

	Help-seeking	GPpsy	110335 (0.513)	0.129 (0.008)	160959 (0.206)	0.115 (0.010)	-0.014	270728	1.067	1.43E-01
	Help-seeking	Psypsy	110374 (0.183)	0.118 (0.012)	161225 (0.052)	0.139 (0.015)	0.021	271596	-1.097	1.36E-01
	No-MDD	GPNoDep	14841 (0.257)	0.228 (0.054)	34261 (0.088)	0.120 (0.046)	-0.108	35650	1.526	6.35E-02
SEX	**CATEGORY**	**DEFINITION**	**MALES**		**FEMALES**		**ΔH2**	**DF**	**T-TEST STAT**	**T-TEST PVALUE**
	DSM-based	LifetimeMDD	32933 (0.167)	0.294 (0.039)	34238 (0.315)	0.328 (0.025)	0.033	57293	-0.719	2.36E-01
	DSM-based	MDDRecur	29760 (0.110)	0.360 (0.047)	29625 (0.237)	0.429 (0.033)	0.069	53839	-1.204	1.14E-01
	Symptom	DepAll	37728 (0.203)	0.203 (0.034)	41848 (0.323)	0.218 (0.021)	0.015	64013	-0.383	3.51E-01
	Self-report	SelfRepDep	117712 (0.058)	0.133 (0.015)	136214 (0.095)	0.125 (0.012)	-0.008	229531	0.432	3.33E-01
	Help-seeking	GPpsy	153917 (0.257)	0.149 (0.011)	178712 (0.413)	0.170 (0.009)	0.021	307791	-1.512	6.52E-02
	Help-seeking	Psypsy	154190 (0.097)	0.131 (0.016)	179229 (0.119)	0.145 (0.012)	0.014	293736	-0.699	2.42E-01
	No-MDD	GPNoDep	29968 (0.112)	0.188 (0.041)	28157 (0.187)	0.228 (0.069)	0.041	46298	-0.507	3.06E-01
EDUCATION	**CATEGORY**	**DEFINITION**	**EDUCATION >= 11 YEARS**		**EDUCATION < 11 YEARS**		**ΔH2**	**DF**	**T-TEST STAT**	**T-TEST PVALUE**
	DSM-based	LifetimeMDD	17452 (0.244)	0.356 (0.057)	19815 (0.241)	0.295 (0.042)	-0.061	6872	-2.259	1.20E-02
	DSM-based	MDDRecur	15446 (0.174)	0.424 (0.071)	17652 (0.174)	0.298 (0.056)	-0.126	6152	-1.607	5.40E-02
	Symptom	DepAll	19804 (0.261)	0.306 (0.044)	33136 (0.256)	0.188 (0.027)	-0.118	19841	-1.055	1.46E-01

	Self-report	SelfRepDep	59341 (0.078)	0.132 (0.027)	115285 (0.077)	0.126 (0.014)	-0.006	70255	-1.419	7.79E-02
	Help-seeking	GPpsy	79630 (0.332)	0.145 (0.013)	144159 (0.353)	0.144 (0.009)	-0.001	128193	-0.890	1.87E-01
	Help-seeking	Psypsy	79791 (0.100)	0.153 (0.024)	144659 (0.114)	NA (NA)	NA	135520	-0.997	1.59E-01
	No-MDD	GPNoDep	14570 (0.140)	0.372 (0.117)	24492 (0.164)	0.122 (0.043)	-0.251	14865	-1.277	1.01E-01

Supplemental Table S9: Heritability of definitions of MDD stratified by risk factors

This table shows the heritability estimates of each definition of MDD calculated using PCGCs in subgroups stratified by risk factors (age, deprivation, trauma, neuroticism, sex and education). For each stratum, we show the number of samples (N_sample), case prevalence (prev), heritability estimate (h2) and its standard error (se). We show the difference between heritabilities from the two strata per risk factor (Δh2), and calculate the degree of freedom (df), t-test statistic (t-test_stat), and significance of this difference (t-test_pvalue) using both heritability estimates, standard errors and sample sizes. None of the differences are significant in a two-tailed t-test at $p < 0{\cdot}05$ after multiple testing correction.

CHR	SNP	BP	A1	A0	A1FREQ	NMISS	OR	SE	P	MHC/SV
1	rs1324481	33892964	T	G	0.317	333782	0.9708	0.005	4.01E-08	
1	rs1697593	190453672	T	C	0.483	332979	1.028	0.005	3.70E-08	
1	rs10863714	208710593	A	G	0.4238	335411	1.029	0.005	2.59E-08	
2	rs1004787	45159091	G	A	0.4694	329067	0.9677	0.005	8.80E-11	
2	rs4671381	60496910	T	C	0.4366	334992	0.9728	0.005	4.77E-08	
2	rs67716713	104267572	A	C	0.4866	334796	0.9676	0.005	4.85E-11	
2	rs1427499	145400317	A	G	0.289	336066	1.043	0.006	1.67E-14	
3	rs11711099	25225163	C	T	0.4839	330036	1.029	0.005	1.75E-08	
3	rs591988	61847695	T	C	0.2663	334385	1.032	0.006	4.25E-08	
3	rs73121433	83154013	A	G	0.1632	328418	1.039	0.007	2.01E-08	
3	rs12714592	84387950	C	A	0.2734	335552	1.038	0.006	4.85E-11	SV
3	rs62250491	85616009	G	T	0.3713	333650	1.042	0.005	2.02E-15	SV
3	rs1995245	117776020	C	T	0.1497	333233	0.959	0.007	2.45E-09	
4	rs899632	57749347	C	T	0.3867	332773	0.9702	0.005	4.86E-09	
5	rs1549212	166996722	C	T	0.3744	333822	1.032	0.005	2.22E-09	
6	rs2892512	98744946	C	T	0.2684	334016	1.033	0.006	7.58E-09	
6	rs465646	111620758	G	A	0.1588	334411	1.055	0.007	7.74E-15	
7	rs13237637	3503207	C	G	0.4967	331983	0.9718	0.005	1.41E-08	
7	rs1174864	53127559	G	A	0.4483	334045	0.9726	0.005	3.68E-08	
7	rs2404324	99023461	G	A	0.1547	335670	0.9627	0.007	3.57E-08	
7	rs10244364	117529641	C	T	0.3201	330304	1.039	0.005	1.03E-12	
8	rs13258512	92777433	G	A	0.4221	331316	0.9709	0.005	6.40E-09	
9	rs56209921	38276428	T	C	0.163	335889	1.039	0.007	1.42E-08	
9	rs323740	102083530	G	C	0.4737	335303	0.9706	0.005	2.80E-09	
9	rs7870475	128134034	C	T	0.4743	334109	1.032	0.005	6.28E-10	
10	rs12244388	104640052	A	G	0.3357	335056	1.036	0.005	1.91E-11	
11	rs4756044	45961114	C	A	0.2525	334624	0.9689	0.006	4.10E-08	SV
11	rs10896972	59212884	G	T	0.3302	336010	0.9687	0.005	2.30E-09	

11	rs2155290	112851068	G	C	0.3826	335943	1.067	0.005	9.47E-36	
12	rs4766578	111904371	T	A	0.4958	335926	1.028	0.005	3.62E-08	SV
13	rs56081685	59454140	G	T	0.314	335337	0.9704	0.005	2.76E-08	
13	rs837333	101179012	C	T	0.4763	326768	1.029	0.005	1.34E-08	
15	rs150294	89931148	G	A	0.3996	329261	0.9701	0.005	4.04E-09	
17	rs7216173	51891405	A	T	0.2139	323640	1.036	0.006	1.13E-08	
18	rs1221983	49993266	T	G	0.394	335304	1.029	0.005	1.63E-08	

Supplemental Table S10: Genome-wide significant loci in GWAS for smoking

This table shows the genome-wide significant loci (top SNP in 1MB regions) in GWAS for smoking (data field 20160: "Ever smoked"). For each of the locus, we show the chromosome (CHR), rsid for the top SNP in 1MB window (SNP), position on the chromosome (BP), test and minor allele (A1), major allele (A0), allele frequency of the test allele (A1FREQ) in all White-British samples in UKBiobank, number of samples included in the logistic regression at the locus with no missing genotype, phenotype or covariate data (NMISS), odds ratio of the minor allele on the phenotype (OR), standard error of the OR (SE), p value of the association (P), and whether the locus is in the MHC region on chr6:25-35MB or in any SV regions as listed in Price et al 2008[5] (MHC/SV).

CHR	SNP	BP	A1	A0	A1FREQ	NMISS	BETA	SE	P	MHC/SV
1	rs1002656	37192741	C	T	0.2905	270132	0.01183	0.001924	7.84E-10	
1	rs4561025	91263757	A	C	0.3521	273426	0.01059	0.001912	3.04E-08	
1	rs12145171	171843249	C	A	0.4283	273985	-0.01105	0.00191	7.22E-09	
2	rs12466146	16077199	T	C	0.4219	271236	0.01055	0.00192	3.88E-08	
2	rs6743916	58704449	A	G	0.2927	271257	-0.01136	0.001921	3.34E-09	
2	rs2042555	148555489	A	G	0.418	273328	0.01323	0.001914	4.77E-12	
2	rs7567451	157053380	G	T	0.2672	272984	-0.01054	0.001915	3.67E-08	
2	rs10497655	185462041	C	T	0.3206	274114	-0.0106	0.001909	2.84E-08	SV
3	rs2278609	16924440	C	T	0.216	270342	0.01191	0.001923	5.86E-10	
3	rs1542212	35683935	G	T	0.3917	271082	0.01205	0.001921	3.54E-10	
3	rs57838764	50374568	C	T	0.1142	273935	0.01062	0.001911	2.75E-08	
3	rs836927	107201428	A	C	0.4257	265524	0.01067	0.001941	3.88E-08	
3	rs10935184	136153468	C	T	0.4093	274114	-0.01165	0.00191	1.06E-09	
4	rs13102162	90939567	G	A	0.3591	264987	-0.01124	0.001942	7.15E-09	
4	rs10032297	139013401	A	T	0.3958	269788	-0.01136	0.001925	3.54E-09	
4	rs7696796	183166469	A	G	0.2464	274114	0.01116	0.00191	5.05E-09	
5	rs4413518	107738001	A	G	0.2189	271136	0.01234	0.00192	1.29E-10	
6	rs2148254	11994762	G	A	0.4928	270744	-0.01133	0.001921	3.67E-09	
6	rs7772160	27412386	C	T	0.4783	274114	-0.01061	0.00191	2.72E-08	MHC
6	rs2269426	32076499	A	G	0.3594	274114	0.01334	0.001911	2.95E-12	MHC
6	rs6916891	98457595	T	G	0.1176	271957	0.01064	0.001917	2.89E-08	
7	rs56226325	2078981	T	C	0.1534	274114	-0.01052	0.00191	3.61E-08	
7	rs11509880	12261911	A	G	0.3278	273194	0.01211	0.001913	2.43E-10	
7	rs73167875	82939096	A	T	0.2022	269308	-0.01083	0.001926	1.90E-08	
7	rs274632	86269181	A	C	0.4229	273225	-0.01153	0.001915	1.71E-09	
7	rs6976111	117495667	A	C	0.2993	264510	0.01101	0.001943	1.45E-08	
7	rs13226841	126389408	C	T	0.4896	274064	0.0125	0.00191	6.03E-11	
8	rs7845515	4946228	A	G	0.2875	265208	0.01267	0.001941	6.69E-11	

8	rs2921036	8363897	T	C	0.4898	269177	0.01775	0.001927	3.32E-20	SV
8	rs477860	9811765	T	C	0.2968	271822	-0.01439	0.001918	6.25E-14	SV
8	rs11250117	10972740	A	C	0.4667	272432	0.0172	0.001915	2.69E-19	SV
9	rs2380937	4145781	C	T	0.4011	272440	-0.01119	0.001916	5.23E-09	
9	rs4977844	23295899	C	G	0.3598	266189	0.01136	0.001937	4.48E-09	
9	rs7869969	96217447	G	A	0.3321	274057	-0.01087	0.00191	1.28E-08	
9	rs28377268	98225056	T	G	0.1071	272736	0.01142	0.001914	2.42E-09	
9	rs2094580	120490857	G	T	0.2825	270567	0.01071	0.001922	2.50E-08	
11	rs34796300	13315205	T	C	0.4254	272562	0.01137	0.001916	2.96E-09	
11	rs297343	16354653	T	G	0.3604	272965	0.01107	0.001914	7.23E-09	
11	rs2071754	31812582	C	T	0.201	274114	0.01262	0.00191	3.94E-11	
11	rs7107356	47676170	A	G	0.4943	273602	-0.01319	0.001911	5.21E-12	SV
11	rs10896636	57448032	G	C	0.3449	268952	0.01281	0.001928	3.02E-11	
11	rs674437	88689953	G	A	0.4843	273301	0.01098	0.001913	9.46E-09	SV
11	rs35738585	113386347	G	T	0.4334	271900	-0.01561	0.001918	3.95E-16	
12	rs11608355	109879292	C	T	0.3132	272760	0.01181	0.001914	6.75E-10	SV
12	rs3741475	117669914	A	G	0.1944	273597	0.01194	0.001911	4.14E-10	
13	rs2210903	69576975	G	A	0.3692	274114	-0.01168	0.00191	9.64E-10	
14	rs4140799	72170969	G	A	0.47	272113	0.01268	0.001917	3.68E-11	
14	rs10144845	75237770	C	T	0.317	274031	-0.01389	0.001911	3.60E-13	
15	rs12903563	78033735	T	C	0.4798	269275	0.01176	0.001927	1.03E-09	
17	rs12938775	2574821	G	A	0.4989	274114	0.0114	0.00191	2.36E-09	
17	rs35982947	38214275	C	A	0.3692	267464	-0.01109	0.001933	9.51E-09	
17	rs62062288	44096553	A	G	0.2181	267423	0.01863	0.001937	6.72E-22	
17	rs56084168	79084574	T	C	0.1478	273326	-0.0123	0.001913	1.27E-10	
18	rs9952522	31286129	C	T	0.4084	272613	-0.01074	0.001915	2.04E-08	
18	rs11665070	35152563	G	A	0.3313	272412	0.01565	0.001916	3.10E-16	
18	rs8097041	50898217	A	T	0.3644	273465	-0.01159	0.001953	2.95E-09	
18	rs56403421	52765283	C	A	0.3309	268579	0.01207	0.001931	4.15E-10	

22	rs1028321	39926929	A	G	0.3052	272405	-0.0111	0.001918	7.09E-09
22	rs11090045	41753603	A	G	0.3031	263768	0.01321	0.001945	1.13E-11

Supplemental Table S11: Genome-wide significant loci in GWAS for neuroticism

This table shows the genome-wide significant loci (top SNP in 1MB regions) in GWAS for neuroticism score (data field 20127: "Neuroticism score"). For each of the locus, we show the chromosome (CHR), rsid for the top SNP in 1MB window (SNP), position on the chromosome (BP), test and minor allele (A1), major allele (A0), allele frequency of the test allele (A1FREQ) in all White-British samples in UKBiobank, number of samples included in the linear regression at the locus with no missing genotype, phenotype or covariate data (NMISS), standardized effect size of the minor allele on the phenotype (BETA), standard error of the effect (SE), p value of the association (P), and whether the locus is in the MHC region on chr6:25-35MB or in any SV regions as listed in Price et al 2008[5] (MHC/SV).

CONDITION	CATEGORY	DEFINITION	LDSC				rho-HESS	
			INTERCEPT	RHO (SE)	RG (SE)	P	RHO (SE)	RG (SE)
ADHD	DSM-based	LifetimeMDD	0.022	0.048 (0.011)	0.297 (0.076)	8.93E-05	0.054 (0.016)	0.271 (0.026)
	Help-seeking	GPpsy	0.026	0.058 (0.007)	0.389 (0.045)	5.47E-18	0.043 (0.008)	0.272 (0.016)
	Symptom	DepAll	0.026	0.038 (0.011)	0.280 (0.081)	5.00E-04	0.044 (0.015)	0.263 (0.029)
	Self-report	SelfRepDep	0.017	0.017 (0.006)	0.227 (0.077)	3.20E-03	0.020 (0.009)	0.217 (0.030)
	No-MDD	GPNoDep	0.008	0.027 (0.010)	0.307 (0.224)	1.71E-01	0.022 (0.019)	0.206 (0.053)
BIP	DSM-based	LifetimeMDD	0.019	0.043 (0.013)	0.344 (0.095)	3.00E-04	0.038 (0.012)	0.286 (0.035)
	Help-seeking	GPpsy	0.026	0.033 (0.006)	0.320 (0.052)	5.19E-10	0.028 (0.006)	0.264 (0.021)
	Symptom	DepAll	0.001	0.052 (0.009)	0.534 (0.091)	4.66E-09	0.034 (0.011)	0.290 (0.036)
	Self-report	SelfRepDep	0.021	0.020 (0.005)	0.341 (0.088)	1.00E-04	0.019 (0.007)	0.291 (0.037)
	No-MDD	GPNoDep	0.003	0.021 (0.011)	0.305 (0.155)	4.87E-02	0.011 (0.014)	0.163 (0.070)
SCZ	DSM-based	LifetimeMDD	0.023	0.061 (0.010)	0.372 (0.055)	1.38E-11	0.059 (0.008)	0.247 (0.017)
	Help-seeking	GPpsy	0.028	0.046 (0.007)	0.326 (0.037)	5.01E-19	0.038 (0.004)	0.197 (0.010)
	Symptom	DepAll	0.010	0.056 (0.009)	0.421 (0.053)	2.08E-15	0.041 (0.007)	0.201 (0.017)
	Self-report	SelfRepDep	0.019	0.028 (0.005)	0.357 (0.059)	1.11E-09	0.025 (0.004)	0.226 (0.018)
	No-MDD	GPNoDep	0.006	0.031 (0.010)	0.351 (0.111)	1.60E-03	0.024 (0.009)	0.203 (0.035)
AUT	DSM-based	LifetimeMDD	0.010	0.030 (0.016)	0.164 (0.090)	6.66E-02	0.039 (0.040)	0.295 (0.077)
	Help-seeking	GPpsy	0.014	0.005 (0.009)	0.033 (0.057)	5.56E-01	0.016 (0.019)	0.147 (0.047)
	Symptom	DepAll	-0.002	0.041 (0.014)	0.272 (0.100)	6.30E-03	0.017 (0.039)	0.146 (0.080)
	Self-report	SelfRepDep	0.007	0.009 (0.009)	0.098 (0.096)	3.06E-01	0.014 (0.023)	0.227 (0.084)

	No-MDD	GPNoDep	0.011	-0.023 (0.017)	-0.225 (0.175)	1.98E-01	0.009 (0.048)	0.146 (0.168)
PGC1 (MDD)	DSM-based	LifetimeMDD	0.008	0.107 (0.013)	0.857 (0.127)	1.51E-11	0.076 (0.039)	0.796 (0.099)
	Help-seeking	GPpsy	0.013	0.103 (0.008)	0.940 (0.100)	6.52E-21	0.077 (0.018)	0.996 (0.073)
	Symptom	DepAll	0.003	0.103 (0.012)	0.968 (0.130)	1.12E-13	0.069 (0.037)	0.856 (0.104)
	Self-report	SelfRepDep	0.016	0.050 (0.008)	0.811 (0.124)	6.27E-11	0.048 (0.022)	1.110 (0.117)
	No-MDD	GPNoDep	-0.003	0.073 (0.013)	1.058 (0.265)	6.46E-05	0.040 (0.046)	0.809 (0.198)
23andMe (depression)	DSM-based	LifetimeMDD	0.020	0.066 (0.006)	0.826 (0.060)	6.41E-43	0.049 (0.010)	0.512 (0.025)
	Help-seeking	GPpsy	0.045	0.057 (0.003)	0.853 (0.030)	8.74E-179	0.044 (0.004)	0.571 (0.015)
	Symptom	DepAll	0.028	0.048 (0.005)	0.732 (0.064)	3.53E-30	0.041 (0.009)	0.503 (0.028)
	Self-report	SelfRepDep	0.030	0.028 (0.002)	0.735 (0.062)	2.24E-32	0.025 (0.005)	0.568 (0.032)
	No-MDD	GPNoDep	0.015	0.031 (0.005)	0.680 (0.141)	1.42E-06	0.026 (0.011)	0.533 (0.059)
neuroticism	DSM-based	LifetimeMDD	0.179	0.080 (0.007)	0.671 (0.041)	2.69E-60	0.156 (0.009)	1.150 (0.025)
	Help-seeking	GPpsy	0.384	0.072 (0.005)	0.705 (0.019)	1.88E-298	0.147 (0.004)	1.350 (0.015)
	Symptom	DepAll	0.161	0.063 (0.007)	0.617 (0.043)	2.22E-46	0.126 (0.009)	1.090 (0.028)
	Self-report	SelfRepDep	0.231	0.041 (0.004)	0.717 (0.040)	2.48E-71	0.096 (0.005)	1.570 (0.053)
	No-MDD	GPNoDep	0.109	0.046 (0.007)	0.642 (0.128)	5.69E-07	0.099 (0.011)	1.370 (0.092)
smoking	DSM-based	LifetimeMDD	0.037	0.019 (0.006)	0.192 (0.055)	5.00E-04	0.030 (0.009)	0.283 (0.022)
	Help-seeking	GPpsy	0.071	0.023 (0.004)	0.279 (0.037)	3.95E-14	0.031 (0.004)	0.365 (0.014)
	Symptom	DepAll	0.040	0.023 (0.005)	0.286 (0.057)	5.79E-07	0.031 (0.009)	0.348 (0.025)
	Self-report	SelfRepDep	0.022	0.008 (0.003)	0.182 (0.052)	5.00E-04	0.012 (0.005)	0.246 (0.026)
	No-MDD	GPNoDep	0.012	0.015 (0.007)	0.262 (0.100)	8.80E-03	0.015 (0.011)	0.274 (0.045)

Supplemental Table S12: Genetic correlation between definitions of MDD and other conditions

This table shows the genetic correlation between each definition of MDD in UKBiobank with other conditions: attention deficit hyperactive disorder (ADHD), bipolar disorder (BIP), schizophrenia (SCZ), autism (AUT), first meta-analysis of MDD by PGC (PGC1), minimal phenotyping based MDD study by 23andMe (23andMe), and neuroticism and smoking in UKBiobank. This table shows results from two different methods: LDSC and rho-HESS. For each of the methods we show the genetic covariance (rho) and its standard error (rho_se), as well as the genetic correlation (rG) and its standard error (rG_se); genetic correlation is genetic covariance divided by the product of the heritabilities of the pair of traits involved. For estimates from LDSC, we show the p value (P) for the genetic correlation being different from 0. For estimates from rho-HESS, both genetic covariance and genetic correlation are summed from regional genetic covariance and genetic correlations in 1703 independent genomic regions (Supplemental Methods). Correlation coefficient is high between genetic correlation estimated from both methods (r=0·71).

CATEGORY	PHENO	NSAMPLES	CHR	SNP	BP	A1	A0	A1FREQ	OR	SE	P	MHC/SV
DSM-MDD	LifetimeMDD	67171	6	rs926552	29548089	A	G	0.140	0.903	0.019	4.46E-08	MHC
Help-seeking	GPpsy	323344	1	rs6699744	72825144	A	T	0.388	0.960	0.005	6.55E-14	
Help-seeking	GPpsy	332112	1	rs6697602	177039372	G	C	0.083	1.056	0.009	6.36E-09	
Help-seeking	GPpsy	331205	2	rs11123030	124976163	T	C	0.491	1.032	0.005	1.13E-09	
Help-seeking	GPpsy	326346	3	rs66511648	117515519	C	T	0.284	1.033	0.006	2.42E-08	
Help-seeking	GPpsy	329391	5	rs30266	103972357	A	G	0.328	1.039	0.006	3.50E-12	
Help-seeking	GPpsy	331857	6	rs12205083	24275483	G	A	0.104	1.053	0.008	6.24E-10	
Help-seeking	GPpsy	332546	6	rs75782365	26408551	G	T	0.110	0.951	0.008	1.36E-09	MHC
Help-seeking	GPpsy	332629	6	rs7772160	27412386	C	T	0.478	0.965	0.005	3.27E-12	MHC
Help-seeking	GPpsy	332629	6	rs3135296	28795856	T	A	0.120	0.946	0.008	3.75E-12	MHC
Help-seeking	GPpsy	326902	6	rs3115631	29986324	A	T	0.126	0.944	0.008	2.86E-13	MHC
Help-seeking	GPpsy	332629	6	rs1625792	31306420	A	G	0.147	0.961	0.007	4.59E-08	MHC
Help-seeking	GPpsy	327948	6	rs236346	36832103	C	T	0.095	0.950	0.009	1.19E-08	MHC
Help-seeking	GPpsy	331741	6	rs9345737	66676938	G	A	0.439	0.969	0.005	3.03E-09	
Help-seeking	GPpsy	332629	7	rs3807866	12250378	A	G	0.410	1.039	0.005	5.44E-13	
Help-seeking	GPpsy	327847	9	rs393488	17044971	A	T	0.469	0.967	0.005	2.14E-10	
Help-seeking	GPpsy	331280	9	rs12057031	25235063	T	C	0.108	0.952	0.008	5.29E-09	
Help-seeking	GPpsy	322805	10	rs11599236	106454672	C	T	0.408	0.971	0.005	2.91E-08	
Help-seeking	GPpsy	332122	11	rs537635	88705235	T	C	0.484	1.033	0.005	6.24E-10	SV
Help-seeking	GPpsy	329151	11	rs578174	89959637	G	A	0.102	0.953	0.009	2.10E-08	SV

Help-seeking	GPpsy	332489	14	rs12889665	75234830	T	G	0.462	0.972	0.005	3.16E-08	
Help-seeking	GPpsy	329763	14	rs61997596	104511206	A	G	0.191	1.037	0.007	4.15E-08	
Help-seeking	GPpsy	332629	16	rs11646401	21609978	G	C	0.442	1.029	0.005	3.58E-08	
Help-seeking	GPpsy	327186	18	rs12967855	35138245	A	G	0.329	1.034	0.006	1.57E-09	
Help-seeking	GPpsy	331002	18	rs8097498	53449667	G	A	0.387	1.031	0.005	9.58E-09	
Help-seeking	Psypsy	332603	6	rs66975207	26942146	C	A	0.111	0.925	0.013	1.23E-09	MHC
Help-seeking	Psypsy	333419	6	rs4713145	28106827	C	T	0.244	0.941	0.009	5.32E-11	MHC
Help-seeking	Psypsy	333295	6	rs3129120	29111775	C	T	0.124	0.925	0.012	1.45E-10	MHC
Help-seeking	Psypsy	333387	6	rs2517622	30155149	C	G	0.137	0.926	0.012	4.20E-11	MHC
Help-seeking	Psypsy	319776	6	rs535777	32577633	C	G	0.138	0.930	0.012	9.94E-10	MHC
No-MDD	GPNoDep	57572	6	rs3094146	29970960	C	G	0.130	0.870	0.025	4.74E-08	MHC

Supplemental Table S13: Genome-wide significant loci in GWAS for definitions of MDD

This table shows the genome-wide significant loci (top SNP in 1MB regions) in GWAS for all definitions of MDD in UKBiobank. Only four definitions show genome-wide signficiant hits: DSM-based LifetimeMDD, minimal phenotyping, help-seeking based definitions GPpsy and Psypsy, and minimal phenotyping, help-seeking based no-MDD definition that exclude MDD symptoms GPNoDep. For each of the locus, we show the chromosome (CHR), rsid for the top SNP in 1MB window (SNP), position on the chromosome (BP), test and minor allele (A1), major allele (A0), allele frequency of the test allele (A1FREQ) in all White-British samples in UKBiobank, number of samples included in the linear regression at the locus with no missing genotype, phenotype or covariate data (NMISS), standardized effect size of the minor allele on the phenotype (BETA), standard error of the effect (SE), p value of the association (P), and whether the locus is in the MHC region on chr6:25-35MB or in any SV regions as listed in Price et al 2008[5] (MHC/SV). We note that rs3094146 in GPNoDep lies in the same locus as rs3115631 in GPpsy, and they lie in the same MB region as rs926552 in LifetimeMDD and rs3129120 in Psypsy, with GPNoDep showing the greatest size of effect (OR=0.84, SE=0.025) demonstrating this locus is not specific to MDD but shared and potentially driven by conditions other than MDD captured by GPNoDep.

CATEGORY	PHENOTYPE	SNP	CHR	BP	A1	OR	L95	U95	P	REPLICATION
Help seeking	GPpsy	rs6699744	1	72825144	A	0.960	0.950	0.971	6.55E-14	Replicated
		rs6697602	1	177039372	G	1.056	1.037	1.075	6.36E-09	Replicated
		rs11123030	2	124976163	T	1.032	1.022	1.043	1.13E-09	Replicated
		rs66511648	3	117515519	C	1.033	1.021	1.045	2.42E-08	Replicated
		rs30266	5	103972357	A	1.039	1.028	1.050	3.50E-12	Replicated
		rs12205083	6	24275483	G	1.053	1.036	1.071	6.24E-10	Replicated
		rs75782365	6	26408551	G	0.951	0.936	0.967	1.36E-09	Replicated
		rs7772160	6	27412386	C	0.965	0.955	0.974	3.27E-12	Replicated
		rs4713145	6	28106827	C	0.969	0.957	0.980	1.51E-07	Replicated
		rs3135296	6	28795856	T	0.946	0.931	0.961	3.75E-12	Replicated
		rs3129120	6	29111775	C	0.948	0.934	0.963	1.48E-11	Replicated
		rs3115631	6	29986324	A	0.944	0.929	0.959	2.86E-13	Replicated
		rs2517622	6	30155149	C	0.950	0.936	0.964	1.26E-11	Replicated
		rs1625792	6	31306420	A	0.961	0.947	0.975	4.59E-08	Replicated
		rs535777	6	32577633	C	0.965	0.951	0.980	4.66E-06	Replicated
		rs236346	6	36832103	C	0.950	0.934	0.967	1.19E-08	Replicated
		rs9345737	6	66676938	G	0.969	0.960	0.979	3.03E-09	Replicated
		rs3807866	7	12250378	A	1.039	1.028	1.049	5.44E-13	Replicated
		rs393488	9	17044971	A	0.967	0.958	0.977	2.14E-10	Replicated
		rs12057031	9	25235063	T	0.952	0.936	0.968	5.29E-09	Replicated
		rs11599236	10	106454672	C	0.971	0.961	0.981	2.91E-08	Replicated
		rs537635	11	88705235	T	1.033	1.022	1.043	6.24E-10	Replicated
		rs578174	11	89959637	G	0.953	0.937	0.969	2.10E-08	Replicated
		rs12889665	14	75234830	T	0.972	0.962	0.982	3.16E-08	Replicated
		rs61997596	14	104511206	A	1.037	1.023	1.050	4.15E-08	Replicated
		rs11646401	16	21609978	G	1.029	1.019	1.040	3.58E-08	Replicated
		rs12967855	18	35138245	A	1.034	1.023	1.045	1.57E-09	Replicated
DSM based	LifetimeMDD	rs6699744	1	72825144	A	0.984	0.959	1.010	2.20E-01	NotReplicated
		rs6697602	1	177039372	G	1.061	1.015	1.110	9.30E-03	NotReplicated
		rs11123030	2	124976163	T	1.024	0.999	1.050	6.35E-02	NotReplicated

		rs66511648	3	117515519	C	1.034	1.006	1.064	1.79E-02	NotReplicated
		rs30266	5	103972357	A	1.047	1.020	1.076	6.55E-04	Replicated
		rs12205083	6	24275483	G	0.992	0.952	1.033	6.89E-01	NotReplicated
		rs75782365	6	26408551	G	0.939	0.902	0.978	2.32E-03	NotReplicated
		rs7772160	6	27412386	C	0.957	0.934	0.982	6.19E-04	Replicated
		rs4713145	6	28106827	C	0.962	0.934	0.990	9.04E-03	NotReplicated
		rs3135296	6	28795856	T	0.905	0.870	0.941	5.48E-07	Replicated
		rs3129120	6	29111775	C	0.905	0.871	0.941	4.08E-07	Replicated
		rs3115631	6	29986324	A	0.906	0.872	0.941	4.49E-07	Replicated
		rs2517622	6	30155149	C	0.923	0.890	0.958	2.10E-05	Replicated
		rs1625792	6	31306420	A	0.939	0.906	0.973	4.91E-04	Replicated
		rs535777	6	32577633	C	0.952	0.918	0.988	9.68E-03	NotReplicated
		rs236346	6	36832103	C	0.955	0.915	0.997	3.65E-02	NotReplicated
		rs9345737	6	66676938	G	0.994	0.970	1.020	6.63E-01	NotReplicated
		rs3807866	7	12250378	A	1.049	1.023	1.076	2.33E-04	Replicated
		rs393488	9	17044971	A	0.976	0.952	1.001	6.00E-02	NotReplicated
		rs12057031	9	25235063	T	0.938	0.900	0.977	1.94E-03	NotReplicated
		rs11599236	10	106454672	C	0.970	0.945	0.996	2.13E-02	NotReplicated
		rs537635	11	88705235	T	1.041	1.016	1.068	1.44E-03	Replicated
		rs578174	11	89959637	G	0.974	0.934	1.016	2.17E-01	NotReplicated
		rs12889665	14	75234830	T	0.973	0.949	0.998	3.31E-02	NotReplicated
		rs61997596	14	104511206	A	1.064	1.031	1.099	1.18E-04	Replicated
		rs11646401	16	21609978	G	1.019	0.994	1.045	1.37E-01	NotReplicated
		rs12967855	18	35138245	A	1.001	0.975	1.029	9.23E-01	NotReplicated
		rs6699744	1	72825144	A	0.972	0.940	1.006	1.02E-01	NotReplicated
		rs6697602	1	177039372	G	1.073	1.013	1.136	1.61E-02	NotReplicated
		rs11123030	2	124976163	T	1.019	0.986	1.052	2.63E-01	NotReplicated
No-MDD	**GPNoDep**	rs66511648	3	117515519	C	1.016	0.979	1.053	4.02E-01	NotReplicated
		rs30266	5	103972357	A	1.047	1.012	1.084	8.94E-03	NotReplicated
		rs12205083	6	24275483	G	1.043	0.989	1.099	1.18E-01	NotReplicated
		rs75782365	6	26408551	G	0.907	0.860	0.956	2.98E-04	Replicated

		rs7772160	6	27412386	C	0.930	0.900	0.960	1.05E-05	Replicated
		rs4713145	6	28106827	C	0.919	0.885	0.955	1.33E-05	Replicated
		rs3135296	6	28795856	T	0.881	0.837	0.928	1.49E-06	Replicated
		rs3129120	6	29111775	C	0.895	0.851	0.941	1.54E-05	Replicated
		rs3115631	6	29986324	A	0.869	0.826	0.914	5.62E-08	Replicated
		rs2517622	6	30155149	C	0.891	0.849	0.935	2.80E-06	Replicated
		rs1625792	6	31306420	A	0.913	0.872	0.957	1.38E-04	Replicated
		rs535777	6	32577633	C	0.926	0.882	0.973	2.15E-03	NotReplicated
		rs236346	6	36832103	C	0.990	0.937	1.046	7.21E-01	NotReplicated
		rs9345737	6	66676938	G	0.949	0.919	0.981	1.77E-03	Replicated
		rs3807866	7	12250378	A	1.024	0.991	1.058	1.59E-01	NotReplicated
		rs393488	9	17044971	A	0.964	0.933	0.996	2.86E-02	NotReplicated
		rs12057031	9	25235063	T	0.991	0.940	1.044	7.24E-01	NotReplicated
		rs11599236	10	106454672	C	0.985	0.953	1.019	3.82E-01	NotReplicated
		rs537635	11	88705235	T	0.999	0.967	1.031	9.33E-01	NotReplicated
		rs578174	11	89959637	G	0.962	0.911	1.016	1.66E-01	NotReplicated
		rs12889665	14	75234830	T	0.989	0.957	1.022	5.08E-01	NotReplicated
		rs61997596	14	104511206	A	1.043	1.001	1.087	4.57E-02	NotReplicated
		rs11646401	16	21609978	G	1.047	1.014	1.082	5.11E-03	NotReplicated
		rs12967855	18	35138245	A	1.008	0.974	1.044	6.49E-01	NotReplicated
Other Condition	**SCZ**	rs6699744	1	72825144	A	0.978	0.956	1.000	4.34E-02	NotReplicated
		rs6697602	1	177039372	G	0.993	0.955	1.030	7.06E-01	NotReplicated
		rs11123030	2	124976163	T	1.026	1.006	1.047	1.36E-02	NotReplicated
		rs66511648	3	117515519	C	0.997	0.973	1.021	8.26E-01	NotReplicated
		rs30266	5	103972357	A	1.023	1.000	1.045	5.02E-02	NotReplicated
		rs12205083	6	24275483	G	0.985	0.952	1.019	3.92E-01	NotReplicated
		rs75782365	6	26408551	G	0.754	0.714	0.794	8.89E-27	Replicated
		rs7772160	6	27412386	C	0.981	0.960	1.001	6.87E-02	NotReplicated
		rs4713145	6	28106827	C	0.923	0.898	0.948	8.73E-09	Replicated
		rs3135296	NA	NA	T	NA	NA	NA	NA	NotReplicated
		rs3129120	NA	NA	C	NA	NA	NA	NA	NotReplicated

		rs3115631	NA	NA	A	NA	NA	NA	NA	NotReplicated
		rs2517622	NA	NA	C	NA	NA	NA	NA	NotReplicated
		rs1625792	NA	NA	A	NA	NA	NA	NA	NotReplicated
		rs535777	6	32577633	C	0.874	0.839	0.910	1.78E-13	Replicated
		rs236346	6	36832103	C	0.999	0.965	1.034	9.72E-01	NotReplicated
		rs9345737	6	66676938	G	0.976	0.955	0.997	2.84E-02	NotReplicated
		rs3807866	7	12250378	A	1.016	0.995	1.037	1.36E-01	NotReplicated
		rs393488	9	17044971	A	0.982	0.961	1.003	9.64E-02	NotReplicated
		rs12057031	9	25235063	T	0.954	0.920	0.989	7.57E-03	NotReplicated
		rs11599236	10	106454672	C	0.961	0.939	0.982	3.63E-04	Replicated
		rs537635	11	88705235	T	1.014	0.993	1.035	1.97E-01	NotReplicated
		rs578174	11	89959637	G	0.989	0.951	1.027	5.83E-01	NotReplicated
		rs12889665	14	75234830	T	0.981	0.960	1.002	7.26E-02	NotReplicated
		rs61997596	14	104511206	A	1.057	1.030	1.085	7.48E-05	Replicated
		rs11646401	16	21609978	G	1.026	1.004	1.048	1.66E-02	NotReplicated
		rs12967855	18	35138245	A	1.015	0.993	1.037	1.77E-01	NotReplicated
Other Condition	**Smoking**	rs6699744	1	72825144	A	0.996	0.986	1.006	4.22E-01	NotReplicated
		rs6697602	1	177039372	G	1.001	0.984	1.019	8.82E-01	NotReplicated
		rs11123030	2	124976163	T	0.998	0.988	1.008	6.74E-01	NotReplicated
		rs66511648	3	117515519	C	1.000	0.990	1.012	9.33E-01	NotReplicated
		rs30266	5	103972357	A	1.020	1.009	1.031	2.01E-04	Replicated
		rs12205083	6	24275483	G	1.005	0.989	1.021	5.48E-01	NotReplicated
		rs75782365	6	26408551	G	0.967	0.952	0.982	2.55E-05	Replicated
		rs7772160	6	27412386	C	0.982	0.973	0.992	3.08E-04	Replicated
		rs4713145	6	28106827	C	0.978	0.967	0.989	1.53E-04	Replicated
		rs3135296	6	28795856	T	0.970	0.955	0.984	5.95E-05	Replicated
		rs3129120	6	29111775	C	0.970	0.956	0.985	6.00E-05	Replicated
		rs3115631	6	29986324	A	0.979	0.964	0.993	4.67E-03	NotReplicated
		rs2517622	6	30155149	C	0.976	0.962	0.990	6.61E-04	Replicated
		rs1625792	6	31306420	A	0.979	0.966	0.993	3.04E-03	NotReplicated
		rs535777	6	32577633	C	0.981	0.967	0.995	9.58E-03	NotReplicated

		rs236346	6	36832103	C	0.992	0.976	1.009	3.73E-01	NotReplicated
		rs9345737	6	66676938	G	0.994	0.984	1.004	2.07E-01	NotReplicated
		rs3807866	7	12250378	A	1.004	0.995	1.015	3.78E-01	NotReplicated
		rs393488	9	17044971	A	0.997	0.987	1.007	5.18E-01	NotReplicated
		rs12057031	9	25235063	T	0.982	0.966	0.998	2.29E-02	NotReplicated
		rs11599236	10	106454672	C	0.988	0.978	0.998	1.79E-02	NotReplicated
		rs537635	11	88705235	T	1.009	0.999	1.019	7.48E-02	NotReplicated
		rs578174	11	89959637	G	0.982	0.967	0.999	3.19E-02	NotReplicated
		rs12889665	14	75234830	T	0.988	0.978	0.998	1.38E-02	NotReplicated
		rs61997596	14	104511206	A	1.010	0.997	1.023	1.19E-01	NotReplicated
		rs11646401	16	21609978	G	1.000	0.990	1.010	9.81E-01	NotReplicated
		rs12967855	18	35138245	A	1.011	1.000	1.022	4.10E-02	NotReplicated
Other Condition	**Neuroticism**	rs6699744	1	72825144	A	0.996	0.993	1.000	5.50E-02	NotReplicated
		rs6697602	1	177039372	G	1.006	1.002	1.010	2.75E-03	NotReplicated
		rs11123030	2	124976163	T	1.006	1.002	1.010	2.23E-03	NotReplicated
		rs66511648	3	117515519	C	1.005	1.001	1.009	1.33E-02	NotReplicated
		rs30266	5	103972357	A	1.009	1.005	1.013	3.75E-06	Replicated
		rs12205083	6	24275483	G	1.008	1.004	1.012	3.17E-05	Replicated
		rs75782365	6	26408551	G	0.992	0.988	0.996	3.35E-05	Replicated
		rs7772160	6	27412386	C	0.989	0.986	0.993	2.72E-08	Replicated
		rs4713145	6	28106827	C	0.993	0.989	0.996	9.14E-05	Replicated
		rs3135296	6	28795856	T	0.991	0.988	0.995	6.03E-06	Replicated
		rs3129120	6	29111775	C	0.992	0.988	0.995	1.53E-05	Replicated
		rs3115631	6	29986324	A	0.992	0.988	0.996	3.67E-05	Replicated
		rs2517622	6	30155149	C	0.993	0.989	0.997	2.15E-04	Replicated
		rs1625792	6	31306420	A	0.994	0.990	0.997	7.07E-04	Replicated
		rs535777	6	32577633	C	0.992	0.988	0.996	4.71E-05	Replicated
		rs236346	6	36832103	C	0.995	0.991	0.999	8.75E-03	NotReplicated
		rs9345737	6	66676938	G	0.997	0.993	1.001	9.08E-02	NotReplicated
		rs3807866	7	12250378	A	1.012	1.008	1.016	9.08E-10	Replicated
		rs393488	9	17044971	A	0.994	0.990	0.998	1.25E-03	Replicated

	rs12057031	9	25235063	T	0.995	0.991	0.999	6.92E-03	NotReplicated
	rs11599236	10	106454672	C	0.989	0.986	0.993	5.16E-08	Replicated
	rs537635	11	88705235	T	1.011	1.007	1.015	1.74E-08	Replicated
	rs578174	11	89959637	G	0.994	0.990	0.997	7.41E-04	Replicated
	rs12889665	14	75234830	T	0.987	0.983	0.991	9.88E-12	Replicated
	rs61997596	14	104511206	A	1.005	1.001	1.009	1.29E-02	NotReplicated
	rs11646401	16	21609978	G	1.003	0.999	1.007	1.38E-01	NotReplicated
	rs12967855	18	35138245	A	1.015	1.012	1.019	1.47E-15	Replicated

Supplemental Table S14: Effects of GWAS loci from help-seeking definitions on other definitions of MDD in UKBiobank and psychiatric conditions

This table shows the genome-wide significant loci (top SNP in 1MB regions) in GWAS minimal phenotyping, help-seeking definitions GPpsy and Psypsy and their effects in the following phenotypes: help-seeking definition GPpsy, DSM-based definition LifetimeMDD, help-seeking no-MDD definition that specifically exclude MDD symptoms GPNoDep, and other psychiatric conditions SCZ, neuroticism and smoking. For each SNP we show the chromosome (CHR), rsid (SNP), position on the chromosome (BP), and test and minor allele (A1). For each SNP-phenotype association we show the odds ratio (OR, in the case of neuroticism which is a quantitative trait, we show exp(BETA) as OR), the lower and upper bounds of the 95% confidence interval of OR (L95 and U95), and the p value of association (P). An association is considered "Replicated' if its p value is below $3{\cdot}09 \times 10^{-4}$ ($p < 0{\cdot}05$ after multiple testing correction for 162 tests in total) and its direction of effect is the same as that in the GPpsy and Psypsy, where the association is discovered.

CATEGORY	PHENOTYPE	SNP	CHR	BP	A1	OR	L95	U95	P	REPLICATION
Help seeking	GPpsy	rs301806	1	8482078	C	1.021	1.010	1.031	1.00E-04	Replicated
		rs11209948	1	72811904	G	0.963	0.953	0.973	1.62E-12	Replicated
		rs2422321	1	73293393	G	1.021	1.010	1.032	8.96E-05	Replicated
		rs12065553	1	80793118	G	1.014	1.003	1.025	1.58E-02	NotReplicated
		rs1518395	2	58208074	A	0.989	0.978	0.999	3.21E-02	NotReplicated
		rs1656369	3	158280085	A	0.979	0.969	0.990	1.23E-04	Replicated
		rs10514299	5	87663610	T	1.017	1.005	1.029	4.77E-03	NotReplicated
		rs454214	5	88003403	C	1.018	1.008	1.029	6.71E-04	Replicated
		rs4543289	5	164484948	T	0.977	0.967	0.987	6.52E-06	Replicated
		rs1475120	6	105389953	G	0.981	0.971	0.991	2.71E-04	Replicated
		rs7044150	9	2982931	T	0.987	0.977	0.998	1.90E-02	NotReplicated
		rs6476606	9	37005561	A	1.023	1.012	1.034	2.54E-05	Replicated
		rs10786831	10	106614571	A	1.026	1.015	1.036	1.51E-06	Replicated
		rs2125716	12	84941429	A	1.018	1.006	1.030	4.03E-03	NotReplicated
		rs12552	13	53625781	A	1.011	1.001	1.022	3.25E-02	NotReplicated
		rs8025231	15	37648402	C	1.018	1.008	1.029	5.68E-04	Replicated
		rs2179744	22	41621714	A	1.015	1.004	1.027	7.78E-03	NotReplicated
DSM based	LifetimeMDD	rs301806	1	8482078	C	0.998	0.973	1.023	8.44E-01	NotReplicated
		rs11209948	1	72811904	G	0.977	0.952	1.002	6.88E-02	NotReplicated
		rs2422321	1	73293393	G	1.038	1.012	1.065	4.52E-03	NotReplicated
		rs12065553	1	80793118	G	1.032	1.004	1.060	2.37E-02	NotReplicated
		rs1518395	2	58208074	A	0.990	0.965	1.016	4.63E-01	NotReplicated
		rs1656369	3	158280085	A	0.978	0.952	1.004	8.95E-02	NotReplicated
		rs10514299	5	87663610	T	1.008	0.979	1.037	6.03E-01	NotReplicated
		rs454214	5	88003403	C	1.029	1.003	1.056	2.77E-02	NotReplicated
		rs4543289	5	164484948	T	0.949	0.926	0.974	5.19E-05	Replicated
		rs1475120	6	105389953	G	0.987	0.962	1.012	2.92E-01	NotReplicated
		rs7044150	9	2982931	T	1.006	0.980	1.032	6.77E-01	NotReplicated
		rs6476606	9	37005561	A	1.021	0.995	1.048	1.22E-01	NotReplicated
		rs10786831	10	106614571	A	1.024	0.999	1.051	6.34E-02	NotReplicated

		rs2125716	12	84941429	A	1.013	0.983	1.044	3.97E-01	NotReplicated
		rs12552	13	53625781	A	1.042	1.016	1.068	1.62E-03	Replicated
		rs8025231	15	37648402	C	1.027	1.002	1.053	3.58E-02	NotReplicated
		rs2179744	22	41621714	A	1.044	1.015	1.073	2.47E-03	Replicated
No-MDD	**GPNoDep**	rs301806	1	8482078	C	0.985	0.953	1.018	3.64E-01	NotReplicated
		rs11209948	1	72811904	G	0.984	0.952	1.017	3.32E-01	NotReplicated
		rs2422321	1	73293393	G	1.014	0.981	1.048	4.16E-01	NotReplicated
		rs12065553	1	80793118	G	0.980	0.946	1.015	2.65E-01	NotReplicated
		rs1518395	2	58208074	A	0.999	0.966	1.032	9.38E-01	NotReplicated
		rs1656369	3	158280085	A	0.984	0.951	1.018	3.61E-01	NotReplicated
		rs10514299	5	87663610	T	1.040	1.002	1.079	4.05E-02	NotReplicated
		rs454214	5	88003403	C	1.020	0.986	1.054	2.49E-01	NotReplicated
		rs4543289	5	164484948	T	0.958	0.928	0.990	9.89E-03	NotReplicated
		rs1475120	6	105389953	G	1.022	0.989	1.056	1.89E-01	NotReplicated
		rs7044150	9	2982931	T	1.005	0.972	1.040	7.61E-01	NotReplicated
		rs6476606	9	37005561	A	0.999	0.966	1.034	9.65E-01	NotReplicated
		rs10786831	10	106614571	A	1.050	1.016	1.085	4.02E-03	NotReplicated
		rs2125716	12	84941429	A	1.045	1.005	1.086	2.57E-02	NotReplicated
		rs12552	13	53625781	A	1.006	0.974	1.040	7.20E-01	NotReplicated
		rs8025231	15	37648402	C	1.023	0.990	1.057	1.74E-01	NotReplicated
		rs2179744	22	41621714	A	1.012	0.976	1.049	5.18E-01	NotReplicated
Other condition	**SCZ**	rs301806	1	8482078	C	0.946	0.925	0.967	1.38E-06	NotReplicated
		rs11209948	1	72811904	G	0.980	0.958	1.002	7.77E-02	NotReplicated
		rs2422321	1	73293393	G	1.059	1.037	1.080	2.18E-08	Replicated
		rs12065553	1	80793118	G	1.016	0.993	1.039	1.77E-01	NotReplicated
		rs1518395	2	58208074	A	0.942	0.921	0.963	3.43E-08	Replicated
		rs1656369	3	158280085	A	1.000	0.977	1.022	9.72E-01	NotReplicated
		rs10514299	5	87663610	T	0.964	0.940	0.988	2.75E-03	NotReplicated
		rs454214	5	88003403	C	0.973	0.952	0.993	1.09E-02	NotReplicated
		rs4543289	5	164484948	T	0.995	0.974	1.016	6.67E-01	NotReplicated
		rs1475120	6	105389953	G	1.049	1.029	1.070	1.87E-06	NotReplicated

		rs7044150	9	2982931	T	0.985	0.963	1.007	1.74E-01	NotReplicated
		rs6476606	9	37005561	A	0.978	0.956	0.999	4.24E-02	NotReplicated
		rs10786831	10	106614571	A	1.047	1.025	1.068	2.34E-05	Replicated
		rs2125716	12	84941429	A	0.991	0.965	1.016	4.72E-01	NotReplicated
		rs12552	13	53625781	A	0.993	0.972	1.014	5.22E-01	NotReplicated
		rs8025231	15	37648402	C	1.003	0.982	1.024	7.63E-01	NotReplicated
		rs2179744	22	41621714	A	1.064	1.041	1.087	8.31E-08	Replicated
Other condition	**Smoking**	rs301806	1	8482078	C	0.989	0.979	0.999	2.59E-02	NotReplicated
		rs11209948	1	72811904	G	0.994	0.984	1.004	2.47E-01	NotReplicated
		rs2422321	1	73293393	G	1.012	1.001	1.022	2.44E-02	NotReplicated
		rs12065553	1	80793118	G	1.024	1.013	1.035	1.99E-05	Replicated
		rs1518395	2	58208074	A	0.989	0.979	0.999	2.85E-02	NotReplicated
		rs1656369	3	158280085	A	0.981	0.971	0.991	3.07E-04	Replicated
		rs10514299	5	87663610	T	0.974	0.963	0.985	3.95E-06	NotReplicated
		rs454214	5	88003403	C	0.995	0.985	1.005	3.06E-01	NotReplicated
		rs4543289	5	164484948	T	0.990	0.980	0.999	3.74E-02	NotReplicated
		rs1475120	6	105389953	G	0.999	0.989	1.008	7.76E-01	NotReplicated
		rs7044150	9	2982931	T	0.994	0.984	1.004	2.49E-01	NotReplicated
		rs6476606	9	37005561	A	0.995	0.985	1.005	3.16E-01	NotReplicated
		rs10786831	10	106614571	A	1.006	0.996	1.016	2.40E-01	NotReplicated
		rs2125716	12	84941429	A	1.019	1.007	1.031	1.93E-03	Replicated
		rs12552	13	53625781	A	1.018	1.008	1.028	5.80E-04	Replicated
		rs8025231	15	37648402	C	1.004	0.995	1.014	3.78E-01	NotReplicated
		rs2179744	22	41621714	A	1.000	0.990	1.011	9.40E-01	NotReplicated
Other condition	**Neuroticism**	rs301806	1	8482078	C	1.009	1.005	1.013	4.26E-06	Replicated
		rs11209948	1	72811904	G	0.997	0.994	1.001	1.56E-01	NotReplicated
		rs2422321	1	73293393	G	1.003	0.999	1.007	1.37E-01	NotReplicated
		rs12065553	1	80793118	G	1.005	1.002	1.009	5.13E-03	NotReplicated
		rs1518395	2	58208074	A	0.991	0.987	0.995	1.38E-06	Replicated
		rs1656369	3	158280085	A	0.995	0.991	0.999	6.42E-03	NotReplicated
		rs10514299	5	87663610	T	1.005	1.001	1.009	7.18E-03	NotReplicated

rs454214	5	88003403	C	1.004	1.001	1.008	2.09E-02	NotReplicated
rs4543289	5	164484948	T	0.992	0.988	0.995	8.95E-06	Replicated
rs1475120	6	105389953	G	0.994	0.990	0.997	7.84E-04	Replicated
rs7044150	9	2982931	T	0.995	0.991	0.999	6.36E-03	NotReplicated
rs6476606	9	37005561	A	1.005	1.001	1.009	6.82E-03	NotReplicated
rs10786831	10	106614571	A	1.007	1.003	1.011	4.39E-04	Replicated
rs2125716	12	84941429	A	1.002	0.999	1.006	2.30E-01	NotReplicated
rs12552	13	53625781	A	1.002	0.999	1.006	2.25E-01	NotReplicated
rs8025231	15	37648402	C	1.003	0.999	1.006	1.67E-01	NotReplicated
rs2179744	22	41621714	A	1.011	1.007	1.015	2.24E-08	Replicated

Supplemental Table S15: Effects of GWAS loci from minimal phenotyping definition of MDD in 23andMe on definitions of MDD in UKBiobank and psychiatric conditions

This table shows the genome-wide significant loci (top SNP in 1MB regions) in GWAS minimal phenotyping definitions of MDD in 23andMe and their effects in the following phenotypes: help-seeking definition GPpsy, DSM-based definition LifetimeMDD, help-seeking no-MDD definition that specifically exclude MDD symptoms GPNoDep, and other psychiatric conditions SCZ, neuroticism and smoking. For each SNP we show the chromosome (CHR), rsid (SNP), position on the chromosome (BP), and test and minor allele (A1). For each SNP-phenotype association we show the odds ratio (OR, in the case of neuroticism which is a quantitative trait, we show exp(BETA) as OR), the lower and upper bounds of the 95% confidence interval of OR (L95 and U95), and the p value of association (P). An association is considered "Replicated' if its p value is below $4{\cdot}90x10^{-4}$ ($p < 0{\cdot}05$ after multiple testing correction for 102 tests in total) and its direction of effect is the same as that in 23andMe, where the association is discovered.

Supplemental References

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

2 Wain, L. V. *et al.* Novel insights into the genetics of smoking behaviour, lung function, and chronic obstructive pulmonary disease (UK BiLEVE): a genetic association study in UK Biobank. *The Lancet Respiratory Medicine* **3**, 769-781, doi:10.1016/S2213-2600(15)00283-0 (2015).

3 Association, A. P. *Diagnostic and statistical manual of mental disorders (5th ed.)*. (American Psychiatric Association, 2003).

4 Office for National Statistics, N. R. o. S., Northern Ireland Statistics and Research Agency (2016). 2011 Census aggregate data. UK Data Service (Edition: June 2016). (2011).

5 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).

6 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).

7 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).

8 Speed, D., Hemani, G., Johnson, M. R. & Balding, D. J. Improved heritability estimation from genome-wide SNPs. *Am J Hum Genet* **91**, 1011-1021, doi:10.1016/j.ajhg.2012.10.010 (2012).

9 Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet* **88**, 76-82, doi:10.1016/j.ajhg.2010.11.011 (2011).

10 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).

11 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).

12 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).

13 Evans, L. M. *et al.* Comparison of methods that use whole genome data to estimate the heritability and genetic architecture of complex traits. *Nat Genet* **50**, 737-745, doi:10.1038/s41588-018-0108-x (2018).

14 Speed, D. *et al.* Reevaluation of SNP heritability in complex human traits. *Nat Genet* **49**, 986-992, doi:10.1038/ng.3865 (2017).

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

16 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).

17 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).

18 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).

19 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).

20 Dahl, A., Cai, N., Flint, J. & Zaitlen, N. GxEMM: Extending linear mixed models to general gene-environment interactions. *bioRxiv* (2018).

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

22 Speed, D. & Balding, D. Better estimation of SNP heritability from summary statistics provides a new understanding of the genetic architecture of complex traits. *bioRxiv* (2018).

23 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).

24 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).

25 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).

26 Demontis, D. *et al.* Discovery Of The First Genome-Wide Significant Risk Loci For ADHD. *bioRxiv* (2017).

27 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).

28 Merikangas, K. R. *et al.* Prevalence and correlates of bipolar spectrum disorder in the world mental health survey initiative. *Arch Gen Psychiatry* **68**, 241-251, doi:10.1001/archgenpsychiatry.2011.12 (2011).

29 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).