nature genetics

ANALYSIS

https://doi.org/10.1038/s41588-020-0594-5

Minimal phenotyping yields genome-wide association signals of low specificity for major depression

Na Cai[1,2,41]✉, Joana A. Revez[3], Mark J. Adams[4], Till F. M. Andlauer[5,6], Gerome Breen[7,8], Enda M. Byrne[3], Toni-Kim Clarke[4], Andreas J. Forstner[9,10,11], Hans J. Grabe[12], Steven P. Hamilton[13], Douglas F. Levinson[14], Cathryn M. Lewis[8,15], Glyn Lewis[16], Nicholas G. Martin[17], Yuri Milaneschi[18], Ole Mors[19,20], Bertram Müller-Myhsok[21,22,23], Brenda W. J. H. Penninx[18], Roy H. Perlis[24,25], Giorgio Pistis[26], James B. Potash[27], Martin Preisig[26], Jianxin Shi[28], Jordan W. Smoller[25,29,30], Fabien Streit[31], Henning Tiemeier[32,33,34], Rudolf Uher[35], Sandra Van der Auwera[12], Alexander Viktorin[36], Myrna M. Weissman[37,38], MDD Working Group of the Psychiatric Genomics Consortium*, Kenneth S. Kendler[39] and Jonathan Flint[40]

Minimal phenotyping refers to the reliance on the use of a small number of self-reported items for disease case identification, increasingly used in genome-wide association studies (GWAS). Here we report differences in genetic architecture between depression defined by minimal phenotyping and strictly defined major depressive disorder (MDD): the former has a lower genotype-derived heritability that cannot be explained by inclusion of milder cases and a higher proportion of the genome contributing to this shared genetic liability with other conditions than for strictly defined MDD. GWAS based on minimal phenotyping definitions preferentially identifies loci that are not specific to MDD, and, although it generates highly predictive polygenic risk scores, the predictive power can be explained entirely by large sample sizes rather than by specificity for MDD. Our results show that reliance on results from minimal phenotyping may bias views of the genetic architecture of MDD and impede the ability to identify pathways specific to MDD.

A key requisite for robust identification of genetic risk loci underlying psychiatric disease is the use of an appropriately large sample. However, the high cost of phenotyping limits sample collection[1]. One solution for reducing the burden of case identification is to use information from hospital registers[2] or individuals' self-reported symptoms, help seeking, diagnoses or medication. We refer to the latter strategy as 'minimal phenotyping', as it minimizes phenotyping costs and reduces data to a single or few self-reported answers.

However, apart from detecting more GWAS) loci[3–5] (Supplementary Table 1), the consequences of sacrificing symptomatic information for genetic analyses have rarely been investigated. The consequences may be particularly important for MDD because of its phenotypic and likely etiological heterogeneity[6], its high degree of comorbidity with other psychiatric diseases[7] and the substantial discrepancies between self-assessment using symptom scales and diagnoses made with full diagnostic criteria[8]. While a majority of the population self-identifies as having one or two depressive symptoms at any one time, only between 9% and 20% of the population has sufficient symptoms to meet criteria for lifetime occurrence of MDD[8–10]. Furthermore, there are high rates of false positives when diagnoses are made without applying diagnostic criteria[11], and antidepressants are prescribed for a wide range of conditions other than MDD[12–14]. As such, a cohort of MDD cases obtained either through the use of either self-reported illness or prescribed treatment may yield a sample that is not representative of the clinical disorder but enriched in those with nonspecific subclinical depressive symptoms and depression secondary to a comorbid disease.

By comparing the genetic architecture of minimal phenotyping definitions of depression with those using full diagnostic criteria for MDD in the UK Biobank[15], a community-based survey of half a million men and women, we assess the implications of a minimal phenotyping strategy for GWAS in MDD. We find that MDD defined by minimal phenotyping has a large nonspecific component, and if GWAS loci from these definitions are chosen for follow-up molecular characterization, they may not be informative about biology specific to MDD.

Results

Definitions of depression in UK Biobank. We identified five ways that MDD could be defined in the UK Biobank. First, self-reports of participants seeking medical attention for depression or related conditions provided 'help-seeking' definitions of MDD (referred to as 'broad depression' in a previous GWAS[3]). Second, participants were diagnosed with 'symptom-based' MDD if, in addition to meeting help-seeking criteria, they reported ever experiencing one or more of the two cardinal features of depression (low mood or anhedonia) for at least 2 weeks[16]. Third, a 'self-report' definition of MDD was based on participants' self-reports of all past and current medical conditions to trained nurses. Fourth, an electronic medical record

A full list of affiliations appears at the end of the paper.

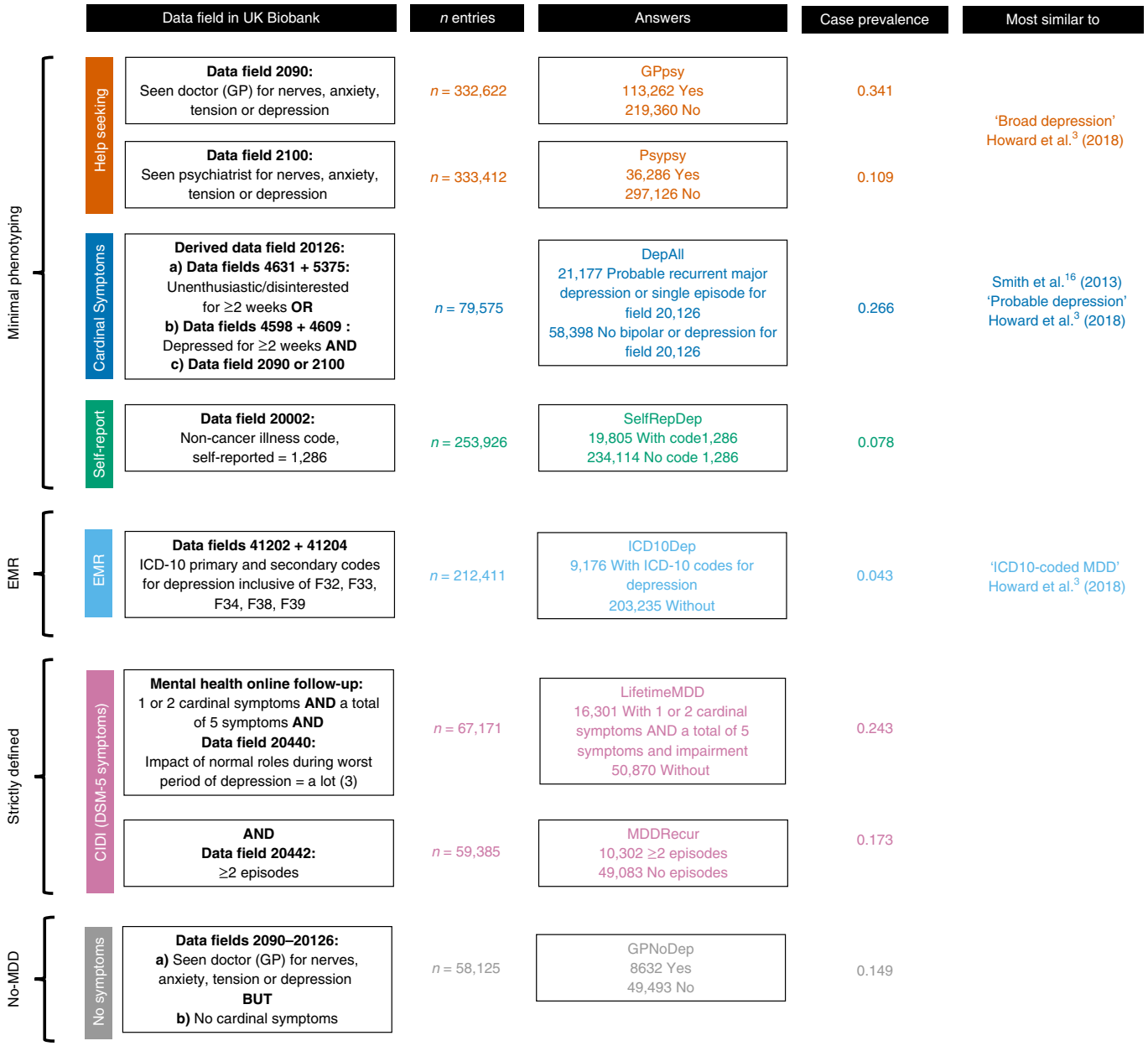


Fig. 1 | Definitions of depression in UK Biobank. This figure shows the different definitions of MDD in the UK Biobank and the color coding used consistently in this paper. The minimal phenotyping definitions of depression are shown in red for help-seeking definitions derived from the Touchscreen Questionnaire; blue for symptom-based definitions derived from the Touchscreen Questionnaire; and green for the self-report-based definition derived from the Verbal Interview. The EMR definition of depression is shown in orange for definitions based on ICD-10 codes. Strictly defined MDD is shown in purple for CIDI-based definitions derived from the Online Mental Health Follow-up. The no-MDD definition is shown in brown for GPNoDep, containing cases in help-seeking definitions that did not have cardinal symptoms for MDD. The data fields in the UK Biobank relevant for defining each phenotype are shown in 'Data field in UK Biobank'; the number of individuals with non-missing entries for each definition are shown in '*n* entries'; the qualifying answers for cases and controls are shown in 'Answers'; the case prevalence in each definition is shown in 'Case prevalence'; and the study and definitions of depression most similar to our definitions are shown in 'Most similar to'. The similarities and differences between help-seeking, EMR and symptom-based definitions in comparison to previously reported definitions of depression can be found in the Supplementary Note.

(EMR) definition was derived from the International Classification of Diseases, Tenth Revision (ICD-10) primary and secondary illness codes in electronic health records. Finally, a 'CIDI-based' diagnosis of lifetime MDD was available from individuals who answered an online 'Mental Health Follow-up' questionnaire (MHQ)[17] based on the Composite International Diagnostic Interview Short Form (CIDI-SF)[18], which included the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5) criteria for MDD (Supplementary Note, Supplementary Fig. 1 and Supplementary Table 2). None of the definitions used trained interviewers applying structured clinical interviews, and only the last applied operationalized criteria, including symptoms, length of episode (more than 2 weeks) and impaired social, occupational or educational function. From here on, we refer to the first three definitions as 'minimal', the fourth as 'EMR-based', and the fifth as 'strictly' defined MDD (Supplementary Note). We also included a category of participants who met the help-seeking definition (part of broad depression in Howard et al.[3]) but failed to meet the symptom-based definition

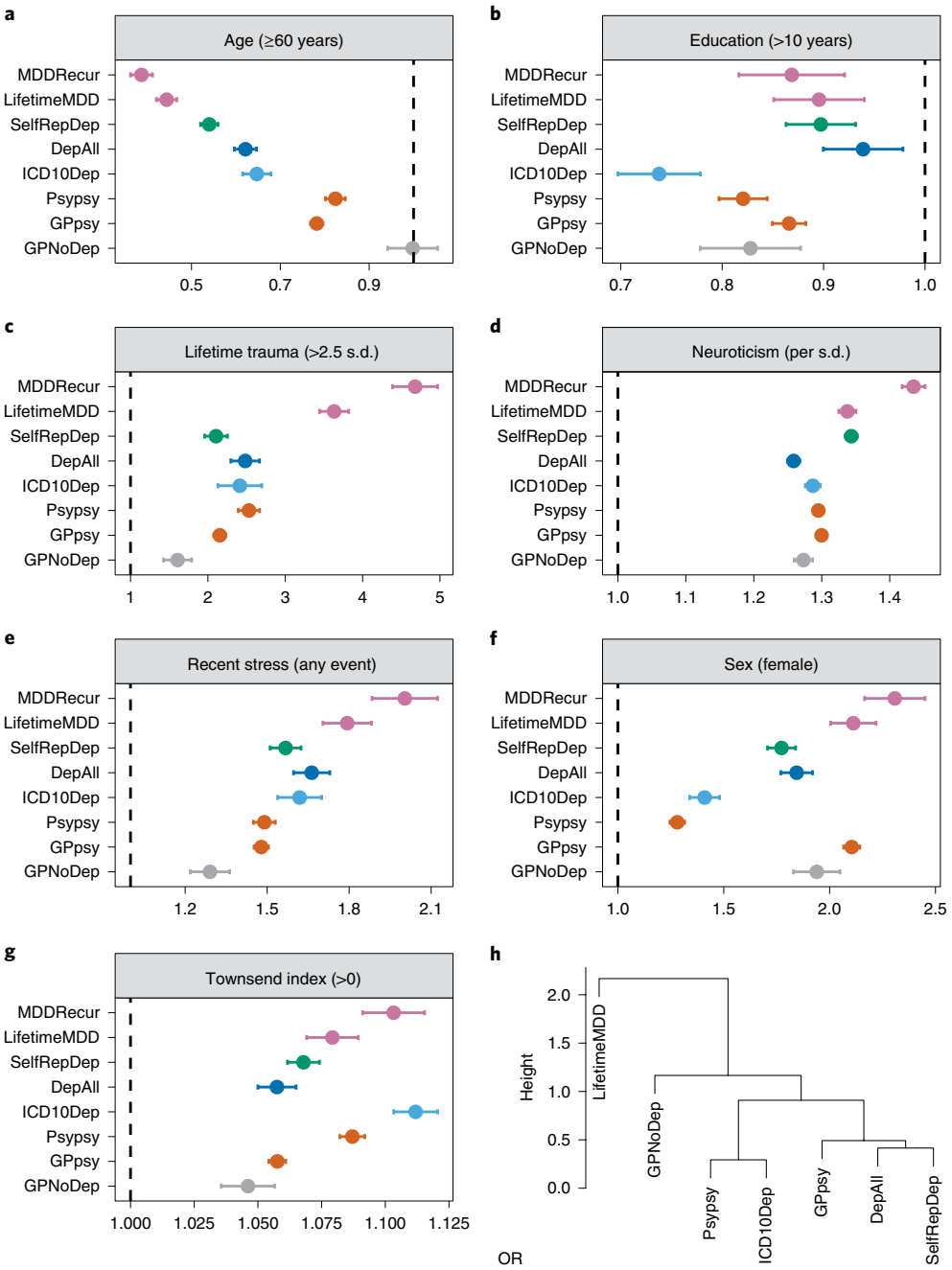


Fig. 2 | Relationship between definitions of depression and environmental risk factors. a–g, Forest plots of ORs of known environmental risk factors and different types (categories) of definitions of depression in the UK Biobank (Definition) from logistic regression, using UK Biobank assessment center, age, sex and years of education as covariates to control for potential geographic and demographic differences between environmental risk factors, except when they were tested. The lifetime trauma measure was derived from the Online Mental Health Follow-up (Supplementary Note and Supplementary Table 7); the Townsend deprivation index, years of education, sex, age, recent stress and neuroticism were derived from Touchscreen Questionnaire (Supplementary Note). **h**, Hierarchical clustering of definitions of depression in the UK Biobank using ORs with environmental risk factors, performed using the hclust function in R; 'height' refers to the Euclidean distance between MDD definitions at the ORs of all six risk factors. MDDRecur was not included in this clustering analysis as it is a subset of the LifetimeMDD definition. The statistics used to generate these plots are presented as source data.

(as they had neither of the two cardinal symptoms of depression: depressed mood or a loss of interest or pleasure in daily activities for more than 2 weeks). We refer to this group as 'no-MDD' (described in detail in the Supplementary Note and Supplementary Table 3). Figure 1 outlines the different diagnostic categories and the number of samples that each group contained.

All definitions were based on recall of episodes or symptoms of depression by participants in the UK Biobank. As priming of recall by current mood affects the reliability of such reports[19–21], we emphasize that each definition is noisy and can be interpreted as being enriched for individuals truly fulfilling its criteria. We explore further characteristics of all definitions and considerations in their

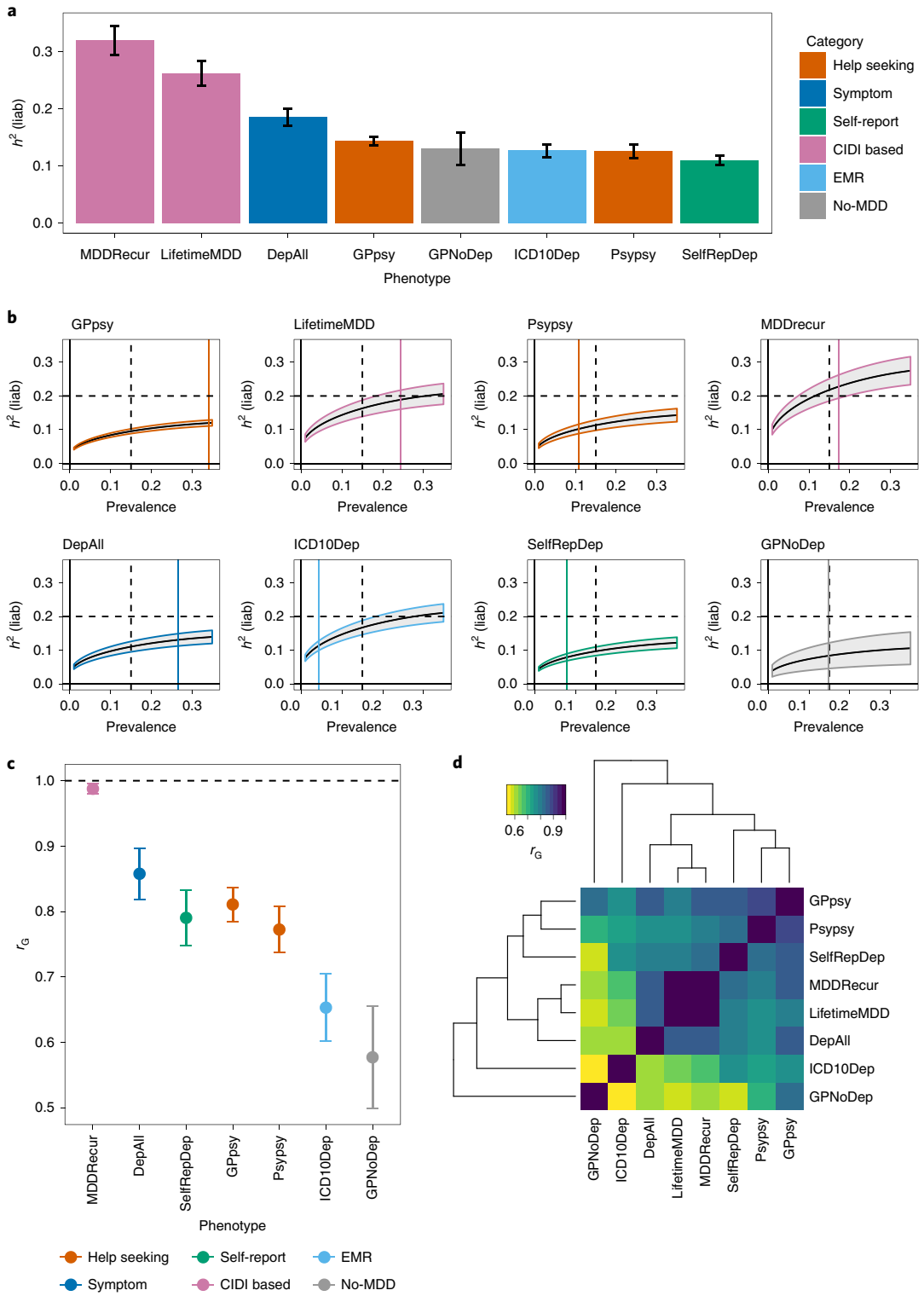


Fig. 3 | SNP heritability and genetic correlation estimates among definitions of MDD in UK Biobank. a, h^2_{SNP} estimates from PCGC[18] on each of the definitions of MDD in the UK Biobank (Methods). h^2_{SNP} (represented as $h2$(liab)) was converted to the liability scale[40,63] using the observed prevalence of each definition of depression in the UK Biobank as both population and sample prevalence (Supplementary Table 4). Error bars show the s.e. of the estimates. **b**, h^2_{SNP} estimates of definitions of MDD in the UK Biobank from LDSC using logistic regression summary statistics on all SNPs with minor allele frequency (MAF) > 5% (Methods), transformed to the liability scale assuming a range of population case prevalence values, from 0 to 0.5. We do not show results for case prevalence from 0.5 to 1, as they would mirror those from 0 to 0.5, with shaded area representing the s.e. of the estimates. We indicate with a black vertical dashed line the population prevalence of 0.15, used in PGC1-MDD; a colored vertical line shows the population prevalence of each definition of depression in the UK Biobank. We also indicate with a black horizontal dashed line the arbitrary liability-scale h^2_{SNP} of 0.2, previously estimated for MDD in PGC1-MDD. Using this, we show that at no prevalence would minimal phenotyping-defined depression such as GPpsy (help-seeking definition) reach this estimate. **c**, Genetic correlation 'r_G' between CIDI-based LifetimeMDD and all other definitions of MDD in the UK Biobank, estimated using PCGC. Error bars show the s.e. of the estimates. **d**, Pairwise r_G between all definitions of depression in the UK Biobank, also detailed in Supplementary Table 15.

GWAS in the Supplementary Note, Supplementary Figs. 2–5 and Supplementary Tables 2–11.

Minimal phenotyping definitions of depression are epidemiologically different from strictly defined MDD. We assessed whether known risk factors for MDD were similar between definitions of depression[22]. Figure 2a–g shows the mean effect (odds ratio, OR) with confidence intervals of each of the following: sex[23,24], age[25], educational attainment[26–28], socioeconomic status[29], neuroticism[24,30], experience of stressful life events in the 2 years leading up to the baseline assessment and cumulative traumatic life events preceding assessment[31,32] (Supplementary Note and Supplementary Table 12). Estimates of the risk factor effect sizes differed substantially, and often highly significantly, as shown by the confidence intervals in Fig. 2. These may reflect differences in methods of ascertainment or underlying pathology between definitions of depression. Next, we asked whether differences in risk factors could be used to classify definitions of depression. We applied a clustering algorithm and found that all minimal phenotyping definitions of depression clustered separately from strictly defined MDD (Fig. 2h).

Minimal definitions of depression are not just milder or noisier versions of strictly defined MDD. Depression defined by minimal phenotyping had lower SNP-based heritabilities (h^2_{SNP}) than more strictly defined versions (Fig. 3a). Self-report (SelfRepDep h^2_{SNP} = 11%, standard error (s.e.) = 0.85%) and help-seeking (Psypsy h^2_{SNP} = 13%, s.e. = 1.18%; GPpsy h^2_{SNP} = 14%, s.e. = 0.81%) definitions had heritabilities of 15% or less. By contrast, strictly defined MDD (LifetimeMDD) had a much higher h^2_{SNP} of 26% (s.e. = 2.15%); imposing the further criterion of recurrence brought the h^2_{SNP} up to 32% (s.e. = 2.56%). Other definitions had intermediate h^2_{SNP}. All h^2_{SNP} values were estimated on the liability scale using phenotype correlation–genotype correlation (PCGC)[33] (Supplementary Note), and the trend held regardless of the method used[33–36] (Supplementary Note and Supplementary Table 13). We further verified that the trend could not be explained by potential case prevalence misestimations (Fig. 3b, Supplementary Note, Supplementary Fig. 3 and Supplementary Table 13) and was not affected by regions of high linkage disequilibrium (LD) or complexity[37] (Supplementary Note and Supplementary Fig. 3). We compared h^2_{SNP} estimates from previous studies of MDD[4,38,39] (Supplementary Fig. 6) with our results and found that they fit squarely into the trend we observed: the less strict the criteria used to diagnose MDD, the lower the h^2_{SNP}.

We examined the roles of a number of additional factors for the lower h^2_{SNP} of minimal phenotyping definitions of MDD. First, minimal phenotyping definitions did not simply have a higher environmental contribution to MDD than the stricter definitions. When we assessed h^2_{SNP} in MDD cases with high and low exposure to environmental risk factors[40], we found that minimal phenotyping definitions of depression (GPpsy and SelfRepDep) showed no significant difference between exposures, which were similar to or lower than those for strictly defined MDD (LifetimeMDD and MDDRecur) (Supplementary Note and Supplementary Table 14). Second, the minimal phenotyping definitions did not merely include milder cases of MDD as previously hypothesized[41]. Inclusion of milder cases is equivalent to lowering the threshold for disease liability in the population above which 'cases' for MDD are defined. Under the liability threshold model[42], this did not reduce the h^2_{SNP} (Supplementary Note and Extended Data Fig. 1). Instead, we showed through simulations that the lower h^2_{SNP} of minimal phenotyping definitions of depression may be due to misdiagnosis of controls as cases of MDD and misclassification of those with other conditions as cases of MDD (Extended Data Figs. 1 and 2).

Genetic correlations between definitions of depression and other diseases. We found that the genetic correlation (r_G) between minimal and strictly defined MDD included a large proportion of nonspecific liability to mental ill health. The r_G between GPpsy (minimally defined MDD) and LifetimeMDD (strictly defined MDD) was 0.81 (s.e. = 0.03), significantly different than unity (Fig. 3c,d, Supplementary Table 15, Supplementary Fig. 6 and Supplementary Note). One interpretation of this finding is that the correlation represents shared genetic liability to MDD[4,5]. However, the majority of the genetic liability of LifetimeMDD due to GPpsy (approximately $r_G^2 = 0.81^2 = 66\%$) was shared with the no-MDD definition, GPNoDep, and the genetic liability of GPNoDep explained approximately 70% of the genetic liability of GPpsy (r_G = 0.84, s.e. = 0.05), and 34% of that of LifetimeMDD (r_G = 0.58, s.e. = 0.08).

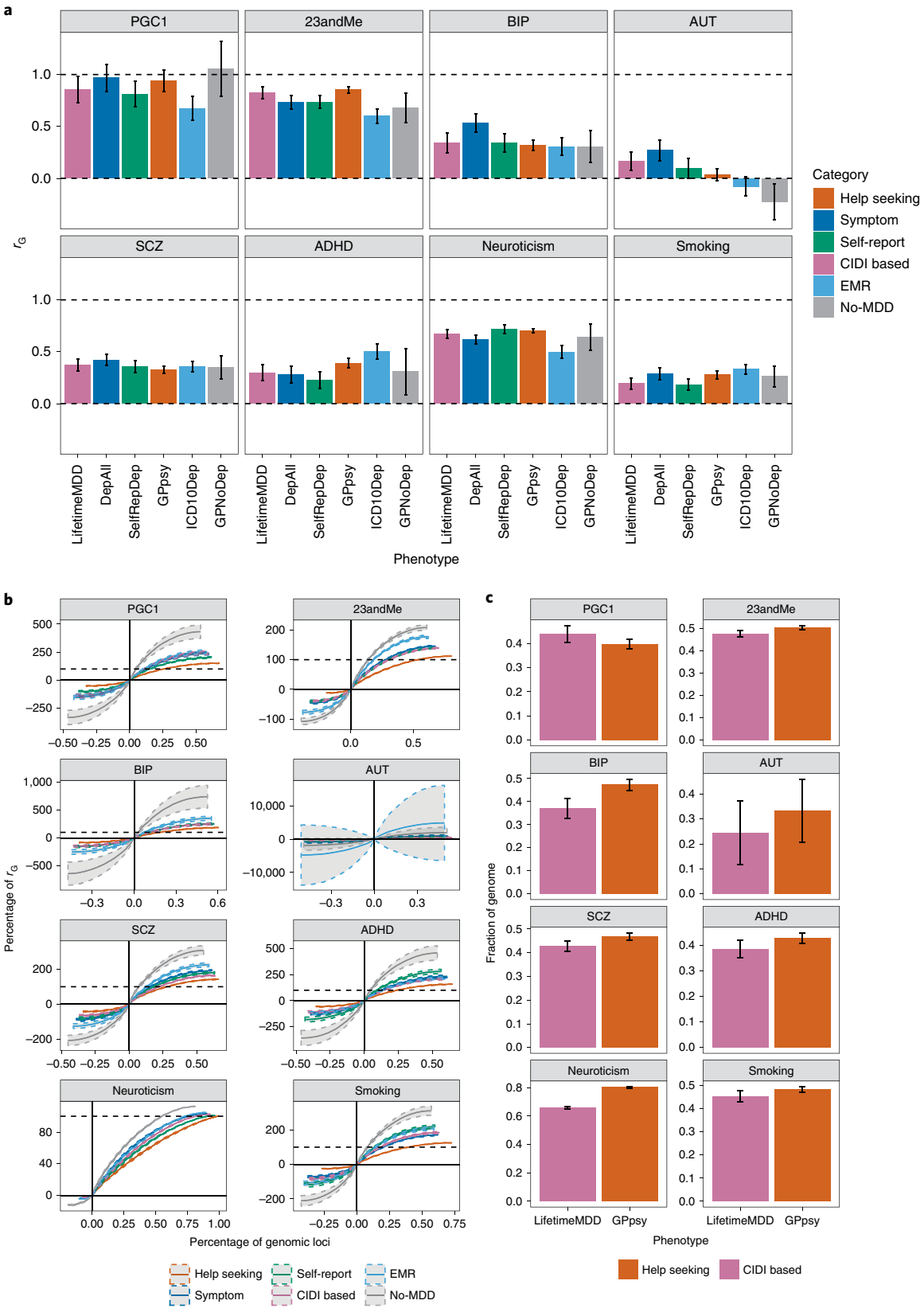
We next examined r_G between different definitions of MDD and comorbid diseases, using cross-trait LD score regression (LDSC)[43] to estimate r_G with neuroticism and smoking (Extended Data Fig. 3 and Supplementary Tables 16 and 17) in the UK Biobank, as well as with all psychiatric conditions in the Psychiatric Genomics Consortium (PGC)[44], including PGC1-MDD[39], and depression defined in 23andMe[4] (Supplementary Table 1). Figure 4a and Supplementary Table 18 show few differences in r_G estimates between other psychiatric disorders and the different definitions of MDD in the UK Biobank, consistent with previous reports[45].

Similar r_G estimates can result from different genetic architectures, indexed by the extent to which genetic liability is spread across the genome. We estimated local r_G and the percentage of the genome contributing to total r_G using rho-HESS[46] (Methods and Fig. 4b). Approximately 65.8% (s.e. = 0.6%), 37.1% (s.e. = 4.5%) and 42.7% (s.e. = 2.3%) of the genome explained 90% of the total rG between strictly defined MDD (LifetimeMDD) and neuroticism, bipolar disorder and schizophrenia, respectively. In comparison, 80.2% (s.e. = 0.6%), 47.3% (s.e. = 2.4%) and 46.8% (s.e. = 0.2%) of the genome was needed to explain the same percentage of total r_G between help-seeking-based GPpsy and the same conditions (Fig. 4c). In other words, minimal phenotyping definitions of depression

Fig. 4 | Genetic correlation between definitions of MDD and other psychiatric conditions. a, The genetic correlation estimated by cross-trait LDSC[43] on the liability scale between definitions of MDD in the UK Biobank and other psychiatric conditions in both the UK Biobank (smoking and neuroticism) and PGC[44] (Supplementary Table 1), including schizophrenia[49] (SCZ) and bipolar disorder[50] (BIP) (Supplementary Table 1). Error bars show the s.e. of the estimates. AUT, autism; ADHD, attention deficit/hyperactivity disorder. **b**, The cumulative fraction of regional genetic correlation (out of the sum of regional genetic correlation across all loci) between definitions of MDD in the UK Biobank and schizophrenia in 1,703 independent loci in the genome[64] estimated using rho-HESS[46], plotted against the percentage of independent loci. CIDI-based LifetimeMDD is shown in purple, while help-seeking-based GPpsy is shown in red. The steeper the curve, the smaller the number of loci explaining the total genetic correlation. The dashed colored curves around each solid line represent the s.e. of the estimate computed using a jackknife approach as described in Shi et al.[36] The dashed black line represents 100% of the sum of genetic correlation between each definition of MDD in the UK Biobank and schizophrenia. The cumulative sums of positive regional genetic correlations (right of *y* axis) go beyond 100%; this is mirrored by the negative regional genetic correlations (left of *y* axis) that go below 0%. **c**, We ranked all 1,703 loci by their magnitude of genetic correlation and asked what fraction of loci summed to 90% of total genetic correlation. This figure shows the percentage of loci summing to 90% of total genetic correlation between either LifetimeMDD (in purple) or GPpsy (in red) and all psychiatric conditions tested, with s.e. estimated using the same jackknife approach. The higher the percentage, the higher the number of genetic loci contributing to 90% of total genetic correlation. Error bars show the s.e. of the estimates.

share more genetic loci with other psychiatric conditions than strictly defined MDD does.

Previous work[4] reported that depression defined through minimal phenotyping shows enrichment of h^2_{SNP} in regions of the genome encoding genes specifically and highly expressed in central nervous system (CNS) tissues represented in Genotype-Tissue Expression (GTEx)[47] project. We assessed this in the definitions of depression in the UK Biobank using LDSC-SEG[48]. As shown in Fig. 5, neither strictly defined MDD (LifetimeMDD) nor MDD defined on the basis of structured clinical assessments



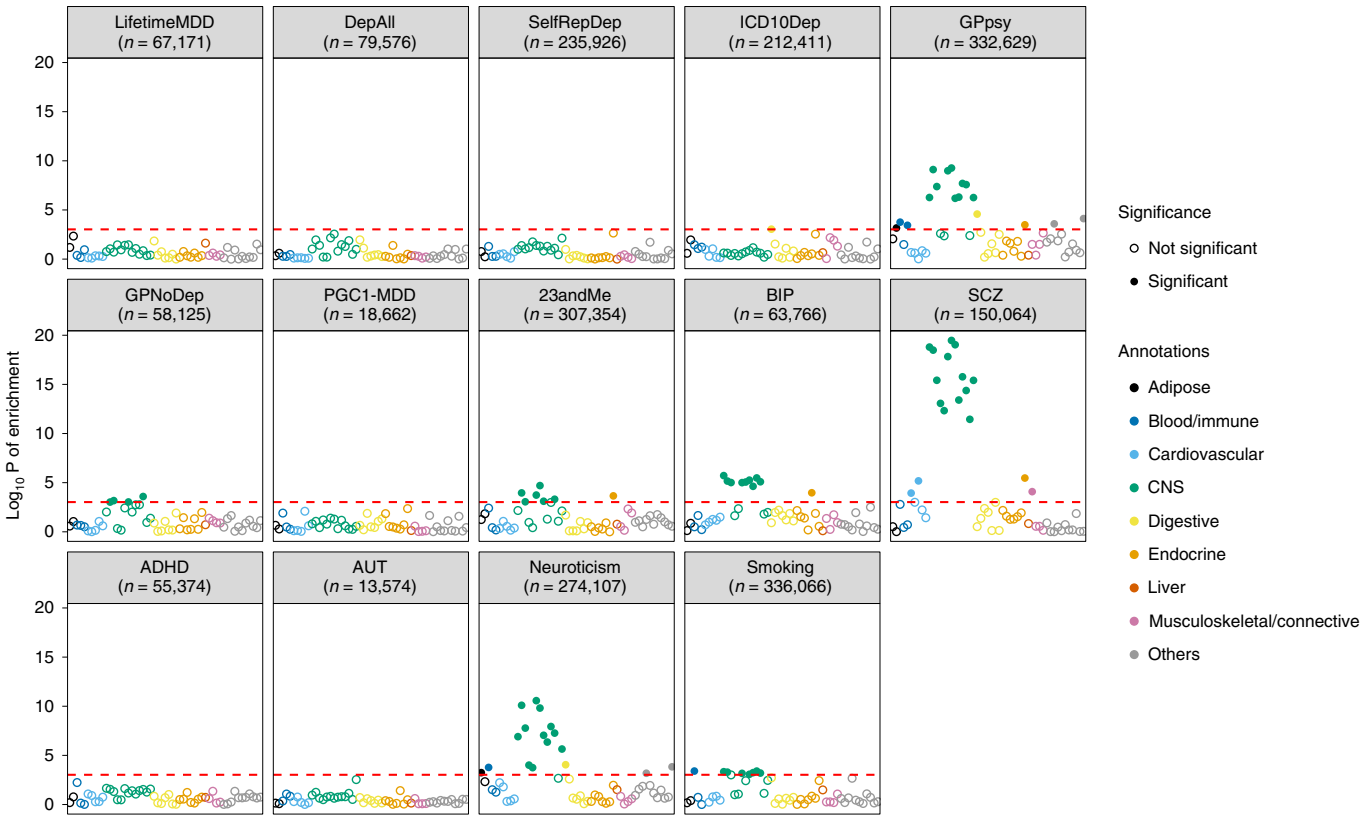


Fig. 5 | Tissue-specific gene expression enrichment in definitions of MDD. The $-\log_{10} P$ value is shown for enrichment in h^2_{SNP} in genes specifically expressed in 44 GTEx tissues, estimated using partitioned h^2_{SNP} in LDSC; the help-seeking based definition of MDD (GPpsy), as well as its constituent no-MDD phenotype (GPNoDep), showed enrichment of h^2_{SNP} in genes specifically expressed in CNS tissues, similarly to an independent cohort of help-seeking-based MDD (23andMe[4]) and other psychiatric conditions such as bipolar disorder[50], schizophrenia[49], autism, personality dimension neuroticism, and the behavioral trait smoking. We indicate the sample size (*n*) for each definition of depression and psychiatric condition.

in PGC1-MDD showed significant CNS enrichments, even though larger and more heterogeneous cohorts did (Methods, Supplementary Note, Supplementary Table 1 and Extended Data Fig. 4). Notably, the minimal phenotyping definition GPpsy showed a significant CNS enrichment, as did the no-MDD help-seeking definition GPNoDep, neuroticism, smoking, and other disorders in the PGC[44], such as schizophrenia[49] and bipolar disorder[50]. Our analysis shows that the degree of CNS enrichment does not relate to the strictness of the definition of MDD and is neither sufficient nor valid evidence that any particular definition of depression better represents MDD or captures the biological mechanisms behind MDD.

GWAS hits from minimal phenotyping are not specific to MDD. We next examined the specificity of the action of individual genetic loci found in GWAS of each definition of MDD. We found that the help-seeking definitions gave the greatest number of genome-wide-significant loci (27 from GPpsy and Psypsy; Supplementary Table 10) in GWAS, consistent with their larger sample sizes and statistical power for finding associations. We examined whether these loci could be detected in strictly defined MDD. Of the 27 loci from minimal phenotyping definitions, 10 showed significant effects (at $P < 0.05$ after multiple-testing correction for 27 loci) on LifetimeMDD, despite the latter's much smaller sample size, consistent with the hypothesis that risk loci for minimal phenotyping MDD also act in strictly defined MDD. However, all ten loci also showed significant effects in neuroticism, smoking, schizophrenia and the no-MDD help-seeking condition (GPNoDep; Supplementary Table 19). Furthermore, all significant SNPs in minimal phenotyping definitions of depression had the same directions of effect on no-MDD phenotypes (Fig. 6).

We found the same pattern of results when we used loci identified from a minimal phenotyping strategy in an independent study by 23andMe that used a minimal phenotyping definition[4]. Of the 17 loci, 10 replicated in GPpsy (at $P < 0.05$, after multiple testing correction for 17 loci) and 3 replicated in LifetimeMDD. All significant SNPs had the same directions of effect on neuroticism, smoking and schizophrenia (Extended Data Fig. 5 and Supplementary Table 20) and are therefore not specific to MDD, consistent with our analysis of minimal phenotyping definitions in the UK Biobank. In summary, GWAS of minimal phenotyping definitions of depression primarily enables the discovery of pathways that are shared with other conditions. It is not currently possible to assess the specificity of GWAS loci from strictly defined MDD in the same way, given that the sample size for strictly defined MDD remains relatively small and GWAS hits relatively few.

Out-of-sample prediction of MDD. Finally, we explored how well the definitions of depression in the UK Biobank predict strictly defined, CIDI-based MDD in independent cohorts, using data from 23 MDD cohorts in the latest data freeze from the MDD Working Group of the Psychiatric Genomics Consortium (PGC29-MDD[5,51]; Supplementary Note, Supplementary Table 21 and Supplementary Fig. 7). We constructed polygenic risk scores (PRSs) on each definition of depression in the UK Biobank (Methods) and examined their prediction in each of the PGC29-MDD cohorts. Of note, PRS from all definitions of depression in the UK Biobank, whether minimally or strictly phenotyped, accounted for a small proportion

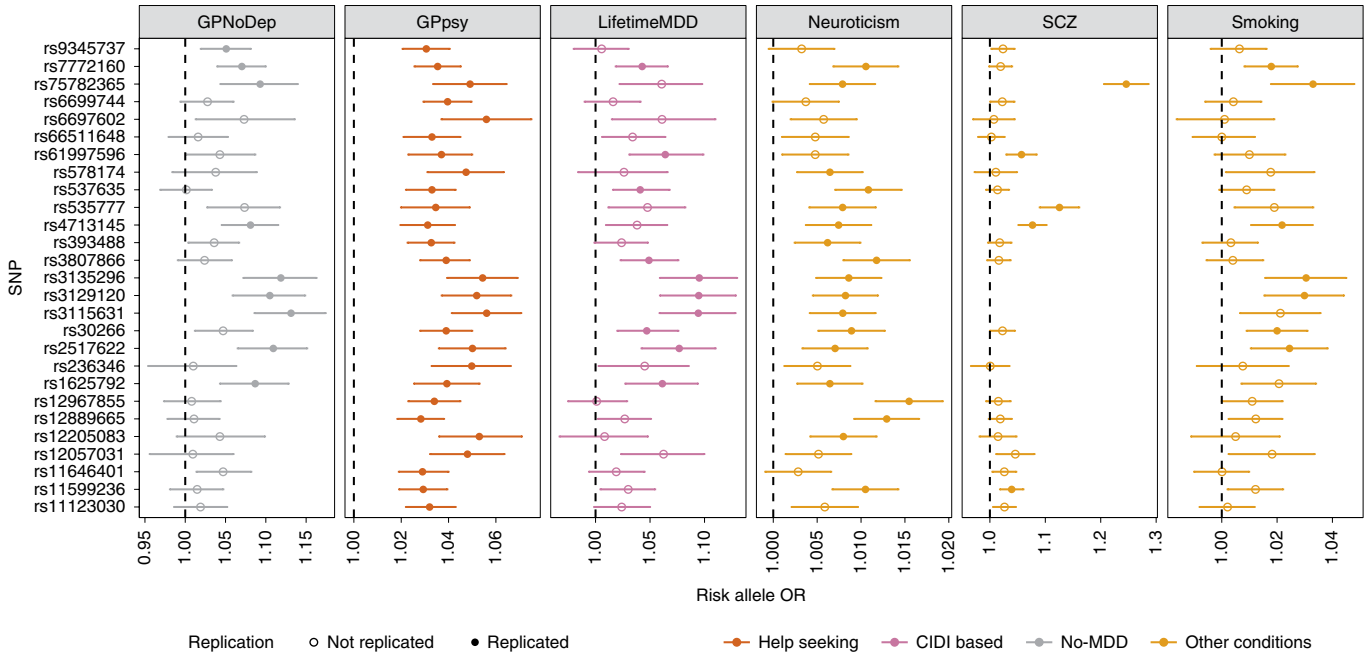


Fig. 6 | GWAS hits from minimal phenotyping definition of MDD in the UK Biobank are not specific to MDD. ORs are shown for the risk alleles at 27 loci significantly associated with help-seeking definitions of MDD in the UK Biobank (GPpsy and Psypsy), in logistic regression GWAS conducted using MDD definitions based on on CIDI (LifetimeMDD, in purple), help seeking (GPpsy, in red) and no-MDD (GPNoDep, in brown) based definitions of MDD. For comparison, we show the same in conditions other than MDD: neuroticism, smoking and schizophrenia (all in pink). SNPs missing in each panel were not tested in the respective GWAS. For clarity of display, scales on different panels vary to accommodate the different magnitudes of ORs of SNPs in different conditions. ORs at all 27 loci were highly consistent across phenotypes, being completely aligned in direction of effect, regardless of whether it was a definition of MDD or a risk factor or condition other than MDD. All results are shown in Supplementary Table 14. Error bars show the s.e. of the estimates.

of variation in disease status in PGC29-MDD (Supplementary Table 22). We observed the following features.

First, the PRS obtained using the full sample of GPpsy performed best at predicting MDD status in independent cohorts from PGC29-MDD (Nargelkerke's $r^2 = 0.018$, area uncer the curve (AUC) = 0.56 at a *P*-value threshold of 0.1; Fig. 7a and Extended Data Fig. 6). However, when equal sample sizes were used (randomly downsampled to 50,000 and case prevalence of 0.15; Methods), GPpsy no longer performed best at predicting MDD status in PGC29-MDD cohorts (Fig. 7b). Rather, the PRS from strictly defined CIDI-based MDD (LifetimeMDD) best predicted MDD disease status (Nargelkerke's $r^2 = 0.0027$, AUC = 0.52 at a *P*-value threshold of 0.1; Extended Data Fig. 6).

Second, the higher prediction accuracy of the PRS obtained using the full sample of GPpsy could be entirely explained by the larger sample size[52] (113,260 cases and 219,362 controls; effective sample size = 298,677; Supplementary Note and Extended Data Fig. 7). We calculated the effective sample size needed for other definitions to have the same predictive power: for strictly defined LifetimeMDD, we would need an effective sample size of 129,106 (Supplementary Note and Extended Data Fig. 7), less than half of that of GPpsy.

Third, the PRS from strictly defined LifetimeMDD predicted MDD disease status better in the PGC29-MDD cohorts, which had a higher percentage of cases fulfilling DSM-5 symptom criteria (Supplementary Table 21 and Extended Data Fig. 8; Pearson r^2 between the AUC and percentage of cases in PGC29-MDD cohorts fulfilling DSM-5 symptom criteria = 0.26, $P = 0.025$ at PRS *P* value = 0.1). This is consistent with the interpretation that LifetimeMDD captures signals specific to MDD. We did not observe such a trend for GPpsy (Pearson $r = 0.02$, $P = 0.57$ at PRS *P* value = 0.1) or any other definition of depression (Supplementary Table 23), suggesting their lower specificity for MDD.

Discussion

Our study demonstrates that the genetic architecture of minimal phenotyping definitions of depression is different from that of strictly defined MDD and is enriched for nonspecific effects on MDD. Using a range of definitions of MDD in the UK Biobank, from self-reported help seeking to a full assessment of the DSM-5 criteria for MDD through self-reported symptoms from the MHQ, we made five key observations.

First, the heritabilities of depression defined by minimal phenotyping strategies are lower than those of MDD defined by full DSM-5 criteria using the CIDI questionnaire. Second, although there is substantial genetic correlation between definitions, much of the shared genetic liability is not specific to MDD and significant differences remain, indicating the presence of genetic effects unique to each definition. Third, a larger percentage of the genome contributes to the shared genetic liability between minimal phenotyping definitions of depression and other psychiatric conditions than that between CIDI-based MDD and other conditions, likely driven by misdiagnosis due to nonspecific phenotyping. Fourth, all GWAS hits from the GPpsy minimal definition of depression are shared with genetically correlated conditions such as neuroticism and smoking. Finally, while minimal phenotyping definitions enable greater predictive power for MDD status in independent cohorts, this is due to the large sample size rather than indexing of MDD-specific effects. These results point to the nonspecific nature of genetic factors identified in minimal phenotyping definitions of depression.

A number of factors need to be borne in mind when interpreting the above observations. Importantly, none of the definitions of depression in the UK Biobank were obtained from structured clinical interviews with an experienced rater (the gold standard for diagnosing MDD). The closest to that standard in the UK Biobank is

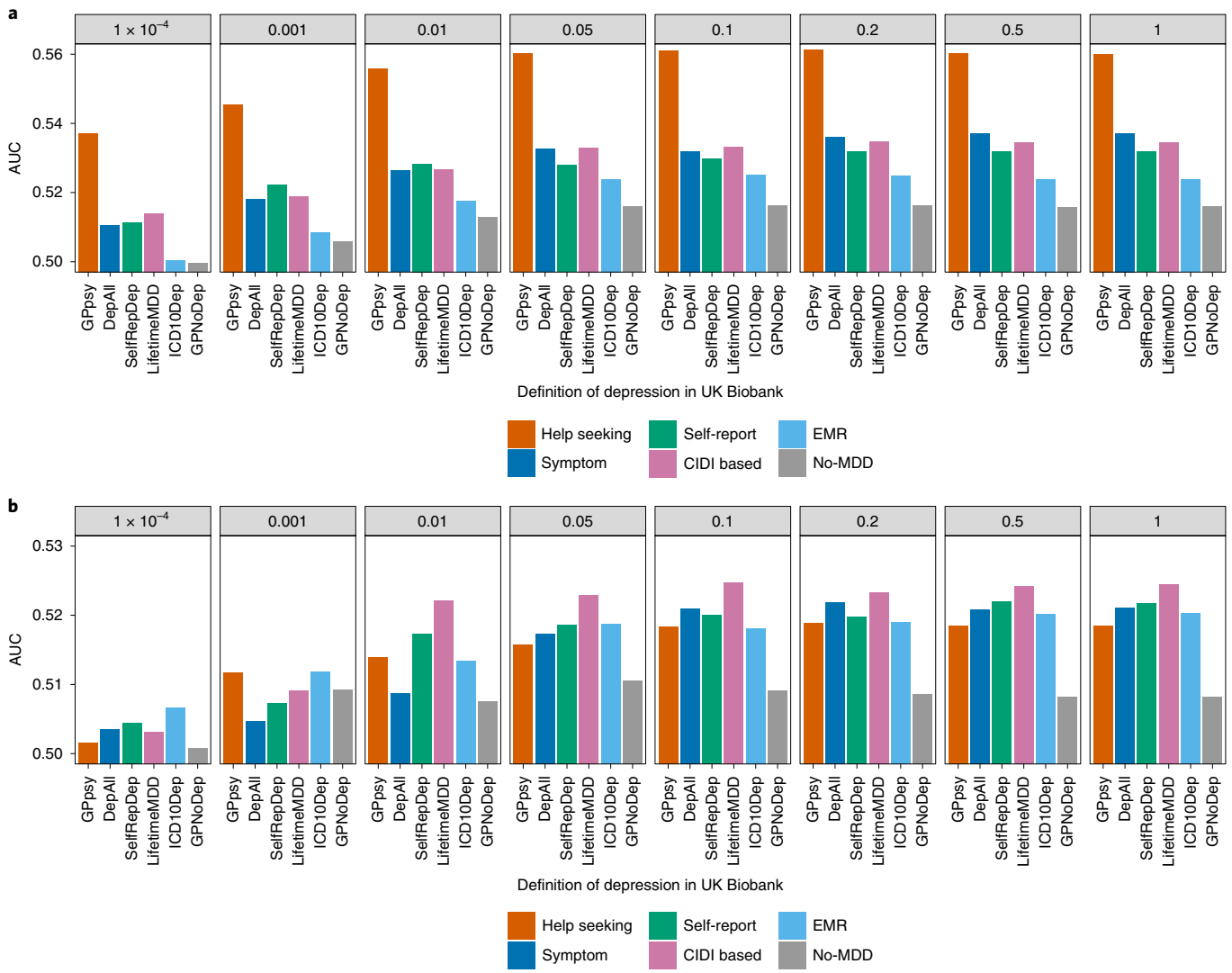


Fig. 7 | Out-of-sample prediction of MDD in PGC cohorts. a, The AUC of PRSs calculated for each definition of depression in the UK Biobank and MDD status indicated in 19 PGC29-MDD cohorts[5], while controlling for cohort-specific effects. PRSs were calculated using effect sizes at independent (LD $r^2 < 0.1$) SNPs passing *P*-value thresholds of 10^{-4}, 0.001, 0.01, 0.05, 0.01, 0.2, 0.5 and 1, in GWAS performed on all definitions of depression in the UK Biobank. **b**, This figure shows the same analysis performed on downsampled data (7,500 cases and 42,500 controls) for each definition of depression.

the online MHQ[17], based on the CIDI-SF[18]. Our results suggest that self-reported diagnoses using CIDI-SF or other diagnostic questionnaire with full DSM-5 criteria lie on the same genetic liability continuum as MDD. This would argue that MDD cases identified through self-report using a full diagnostic questionnaire will be enriched for more strictly defined forms, with the consequence that results from genetic analysis will include loci that contribute to strictly defined MDD disease risk[53,54].

Minimal definitions of MDD do not simply include cases with lower genetic liability to MDD. This is consistent with a recent study of three large twin cohorts, which asked whether a combination of MDD, depressive symptoms and neuroticism could capture all genetic liability of MDD[55] and showed that 65% of the genetic effects contributing to MDD are specific, and minimally defined depression (inclusive of MDD, depressive symptoms and neuroticism) can index only around one-third of the genetic liability to MDD. Similarly, previously reported high degrees of genetic correlation between MDD and depressive symptoms ($r_G = 0.7$, implying that roughly $r_G^2 = 49\%$ of genetic factors contributing to liability of the former is attributable to that of the latter)[22] need to be put in perspective of even higher degrees of sharing between depressive symptoms and other traits such as neuroticism ($r_G = 0.79$–0.94, implying that roughly $r_G^2 = 62$–88% of genetic variance of the former is attributable to that of the latter, especially if both were assayed at a single time point[56]).

Our findings have important implications for downstream investigations. One interpretation is that the nonspecific effects found through using minimal phenotyping approaches will still advance understanding of the biology of psychiatric disorders and their treatment[5,57]. A recent report used the 'quasi-replication' of GWAS loci between depressive symptoms and neuroticism as validation of their functional significance[56]. An alternative view is that these loci reflect the ways in which depressive symptoms can develop as secondary effects, including through susceptibility to adverse life events[58], personality types[24] and use of or exposure to psychoactive agents like cigarettes[59,60]—in which case, while useful for understanding the basis of mental ill health, they are not informative about the genetic etiology of MDD and are not useful for developing disease-specific treatment.

Our findings indicate the need for ways to integrate both strict and minimal phenotyping approaches to determine which loci to prioritize for follow-up functional analyses. They also indicate a

need for means to assess symptoms for diagnosing MDD with specificity at scale, rather than reliance on minimal phenotyping. Fast and accurate diagnostic methods that use a limited number of questionnaire items are becoming available: for example, computerized adaptive diagnostic screening may be as effective for the diagnosis of MDD as an hour-long face-to-face clinician diagnostic interview[61]. There are ongoing attempts to convert behavioral health tracking data from phones or wearable devices into diagnostic information[62]. If successful, these attempts may lead to a dramatic expansion in the ability to collect data appropriate for psychiatric genetics.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41588-020-0594-5.

Received: 2 November 2018; Accepted: 19 February 2020;
Published online: 30 March 2020

References

1. Lu, J. T., Campeau, P. M. & Lee, B. H. Genotype–phenotype correlation: promiscuity in the era of next-generation sequencing. *Obstet. Gynecol. Surv.* **69,** 728–730 (2014).
2. Ripke, S. et al. Genome-wide association analysis identifies 13 new risk loci for schizophrenia. *Nat. Genet.* **45**, 1150–1159 (2013).
3. Howard, D. M. et al. Genome-wide association study of depression phenotypes in UK Biobank identifies variants in excitatory synaptic pathways. *Nat. Commun.* **9**, 1470 (2018).
4. Hyde, C. L. et al. Identification of 15 genetic loci associated with risk of major depression in individuals of European descent. *Nat. Genet.* **48**, 1031–1036 (2016).
5. Wray, N. R. et al. Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nat. Genet.* **50**, 668–681 (2018).
6. Flint, J. & Kendler, K. S. The genetics of major depression. *Neuron* **81**, 484–503 (2014).
7. Kessler, R. C. et al. The epidemiology of major depressive disorder: results from the National Comorbidity Survey Replication (NCS-R). *JAMA* **289**, 3095–3105 (2003).
8. Boyd, J. H., Weissman, M. M., Thompson, W. D. & Myers, J. K. Screening for depression in a community sample. Understanding the discrepancies between depression symptom and diagnostic scales. *Arch. Gen. Psychiatry.* **39**, 1195–1200 (1982).
9. Breslau, N. Depressive symptoms, major depression, and generalized anxiety: a comparison of self-reports on CES-D and results from diagnostic interviews. *Psychiatry Res.* **15**, 219–229 (1985).
10. Weissman, M. M. & Myers, J. K. Rates and risks of depressive symptoms in a United States urban community. *Acta Psychiatr. Scand.* **57**, 219–231 (1978).
11. Mitchell, A. J., Vaze, A. & Rao, S. Clinical diagnosis of depression in primary care: a meta-analysis. *Lancet* **374**, 609–619 (2009).
12. Mojtabai, R. Clinician-identified depression in community settings: concordance with structured-interview diagnoses. *Psychother. Psychosom.* **82**, 161–169 (2013).
13. Druss, B. G. et al. Understanding mental health treatment in persons without mental diagnoses: results from the National Comorbidity Survey Replication. *Arch. Gen. Psychiatry* **64**, 1196–1203 (2007).
14. Marcus, S. C. & Olfson, M. National trends in the treatment for depression from 1998 to 2007. *Arch. Gen. Psychiatry* **67**, 1265–1273 (2010).
15. Sudlow, C. et al. UK Biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med.* **12**, e1001779 (2015).
16. Smith, D. J. et al. Prevalence and characteristics of probable major depression and bipolar disorder within UK Biobank: cross-sectional study of 172,751 participants. *PLoS One* **8**, e75362 (2013).
17. Davis, K. A. S. et al. Mental health in UK Biobank: development, implementation and results from an online questionnaire completed by 157 366 participants. *BJPsych Open* **4**, 83–90 (2018).
18. Kessler, R. C. & Ustun, T. B. The World Mental Health (WMH) Survey initiative version of the World Health Organization (WHO) Composite International Diagnostic Interview (CIDI). *Int. J. Meth. Psych. Res.* **13**, 93–121 (2004).
19. Bromet, E. J., Dunn, L. O., Connell, M. M., Dew, M. A. & Schulberg, H. C. Long-term reliability of diagnosing lifetime major depression in a community sample. *Arch. Gen. Psychiatry* **43**, 435–440 (1986).
20. Kendler, K. S., Neale, M. C., Kessler, R. C., Heath, A. C. & Eaves, L. J. The lifetime history of major depression in women. Reliability of diagnosis and heritability. *Arch. Gen. Psychiatry* **50**, 863–870 (1993).
21. Rice, J. P., Rochberg, N., Endicott, J., Lavori, P. W. & Miller, C. Stability of psychiatric diagnoses. An application to the affective disorders. *Arch. Gen. Psychiatry* **49**, 824–830 (1992).
22. Foley, D. L., Neale, M. C. & Kendler, K. S. Genetic and environmental risk factors for depression assessed by subject-rated symptom check list versus structured clinical interview. *Psychol. Med.* **31**, 1413–1423 (2001).
23. Kendler, K. S., Gardner, C. O., Neale, M. C. & Prescott, C. A. Genetic risk factors for major depression in men and women: similar or different heritabilities and same or partly distinct genes? *Psychol. Med.* **31**, 605–616 (2001).
24. Kendler, K. S., Gatz, M., Gardner, C. O. & Pedersen, N. L. Personality and major depression: a Swedish longitudinal, population-based twin study. *Arch. Gen. Psychiatry* **63**, 1113–1120 (2006).
25. Alexopoulos, G. S. et al. 'Vascular depression' hypothesis. *Arch. Gen. Psychiatry* **54**, 915–922 (1997).
26. Kessler, R. C. et al. Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. *Arch. Gen. Psychiatry* **62**, 593–602 (2005).
27. Kessler, R. C., Foster, C. L., Saunders, W. B. & Stang, P. E. Social consequences of psychiatric disorders. I: Educational attainment. *Am. J. Psychiatry* **152**, 1026–1032 (1995).
28. Lorant, V. et al. Socioeconomic inequalities in depression: a meta-analysis. *Am. J. Epidemiol.* **157**, 98–112 (2003).
29. Kessler, R. C. Epidemiology of women and depression. *J. Affect. Disord.* **74**, 5–13 (2003).
30. Kendler, K. S., Neale, M. C., Kessler, R. C., Heath, A. C. & Eaves, L. J. A longitudinal twin study of personality and major depression in women. *Arch. Gen. Psychiatry* **50**, 853–862 (1993).
31. Kessler, R. C. The effects of stressful life events on depression. *Ann. Rev. Psychol.* **48**, 191–214 (1997).
32. Mazure, C. M. Life stressors as risk factors in depression. *Clinical Psychology: Science and Practice* **5**, 291–313 (1998).
33. Weissbrod, O., Flint, J. & Rosset, S. Estimating SNP-based heritability and genetic correlation in case–control studies directly and with summary statistics. *Am. J. Hum. Genet.* **103**, 89–99 (2018).
34. Bulik-Sullivan, B. K. et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).
35. Loh, P.-R. et al. Contrasting genetic architectures of schizophrenia and other complex diseases using fast variance-components analysis. *Nat. Genet.* **47**, 1385–1392 (2015).
36. Shi, H., Kichaev, G. & Pasaniuc, B. Contrasting the genetic architecture of 30 complex traits from summary association data. *Am. J. Hum. Genet.* **99**, 139–153 (2016).
37. Price, A. L. et al. Long-range LD can confound genome scans in admixed populations. *Am. J. Hum. Genet.* **83**, 132–135 (2008).
38. CONVERGE consortium. Sparse whole-genome sequencing identifies two loci for major depressive disorder. *Nature* **523**, 588–591 (2015).
39. Major Depressive Disorder Working Group of the Psychiatric GWAS Consortiumet al. A mega-analysis of genome-wide association studies for major depressive disorder. *Mol. Psychiatry* **18**, 497–511 (2013).
40. Peterson, R. E. et al. Molecular genetic analysis subdivided by adversity exposure suggests etiologic heterogeneity in major depression. *Am. J. Psychiatry* **175**, 545–554 (2018).
41. Northern Ireland Statistics and Research Agency: 2011 Census aggregate data. UK Data Service https://doi.org/10.5257/census/aggregate-2011-1 (2016).
42. Dempster, E. R. & Lerner, I. M. Heritability of threshold characters. *Genetics* **35**, 212–236 (1950).
43. Bulik-Sullivan, B. et al. An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* **47**, 1236–1241 (2015).
44. Cross-Disorder Group of the Psychiatric Genomics Consortium. Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. *Lancet* **381**, 1371–1379 (2013).
45. Brainstorm Consortium et al. Analysis of shared heritability in common disorders of the brain. *Science* **360**, eaap8757 (2018).
46. Shi, H., Mancuso, N., Spendlove, S. & Pasaniuc, B. Local genetic correlation gives insights into the shared genetic architecture of complex traits. *Am. J. Hum. Genet.* **101**, 737–751 (2017).
47. GTEx Consortium. The Genotype-Tissue Expression (GTEx) project. *Nat. Genet.* **45**, 580–585 (2013).
48. Finucane, H. K. et al. Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. *Nat. Genet.* **50**, 621–629 (2018).

49. Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**, 421–427 (2014).
50. Psychiatric Genomics Consortium Bipolar Disorder Working Group. Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near *ODZ4*. *Nat. Genet.* **43**, 977–998 (2011).
51. Trzaskowski, M. et al. Quantifying between-cohort and between-sex genetic heterogeneity in major depressive disorder. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* **180**, 439–447 (2019).
52. Turley, P. et al. Multi-trait analysis of genome-wide association summary statistics using MTAG. *Nat. Genet.* **50**, 229–237 (2018).
53. Corfield, E. C., Yang, Y., Martin, N. G. & Nyholt, D. R. A continuum of genetic liability for minor and major depression. *Transl. Psychiatry* **7**, e1131 (2017).
54. Direk, N. et al. An analysis of two genome-wide association meta-analyses identifies a new locus for broad depression phenotype. *Biol. Psychiatry* **82**, 322–329 (2017).
55. Kendler, K. S. et al. Shared and specific genetic risk factors for lifetime major depression, depressive symptoms and neuroticism in three population-based twin samples. *Psychol. Med.* **49**, 2745–2753 (2018).
56. Okbay, A. et al. Genetic variants associated with subjective well-being, depressive symptoms, and neuroticism identified through genome-wide analyses. *Nat. Genet.* **48**, 624–633 (2016).
57. McIntosh, A. M., Sullivan, P. F. & Lewis, C. M. Uncovering the genetic architecture of major depression. *Neuron* **102**, 91–103 (2019).
58. Kendler, K. S. & Karkowski-Shuman, L. Stressful life events and genetic liability to major depression: genetic control of exposure to the environment? *Psychol. Med.* **27**, 539–547 (1997).
59. Fluharty, M., Taylor, A. E., Grabski, M. & Munafo, M. R. The association of cigarette smoking with depression and anxiety: a systematic review. *Nicotine Tob. Res.* **19**, 3–13 (2017).
60. Wootton, R. E. et al. Evidence for causal effects of lifetime smoking on risk for depression and schizophrenia: a Mendelian randomisation study. *Psychol. Med.* **6,** 1–9 (2019).
61. Gibbons, R. D. et al. The computerized adaptive diagnostic test for major depressive disorder (CAD-MDD): a screening tool for depression. *J. Clin. Psychiatry* **74**, 669–674 (2013).
62. Freimer, N. B. & Mohr, D. C. Integrating behavioural health tracking in human genetics research. *Nat. Rev. Genet.* **20**, 129–130 (2019).
63. Bycroft, C. et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).
64. Berisa, T. & Pickrell, J. K. Approximately independent linkage disequilibrium blocks in human populations. *Bioinformatics* **32**, 283–285 (2016).

[1]Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, UK. [2]European Bioinformatics Institute (EMBL-EBI), Wellcome Genome Campus, Hinxton, UK. [3]Institute for Molecular Bioscience, The University of Queensland, Brisbane, Queensland, Australia. [4]Division of Psychiatry, University of Edinburgh, Edinburgh, UK. [5]Department of Translational Research in Psychiatry, Max Planck Institute of Psychiatry, Munich, Germany. [6]Department of Neurology, Klinikum rechts der Isar, School of Medicine, Technical University of Munich, Munich, Germany. [7]NIHR Maudsley Biomedical Research Centre, King's College London, London, UK. [8]Social, Genetic and Developmental Psychiatry Centre, King's College London, London, UK. [9]Department of Biomedicine, University of Basel, Basel, Switzerland. [10]Institute of Human Genetics, University of Bonn, School of Medicine & University Hospital Bonn, Bonn, Germany. [11]Centre for Human Genetics, University of Marburg, Marburg, Germany. [12]Department of Psychiatry and Psychotherapy, University Medicine Greifswald, Greifswald, Germany. [13]Department of Psychiatry, Kaiser Permanente Northern California, San Francisco, CA, USA. [14]Department of Psychiatry & Behavioral Sciences, Stanford University, Stanford, CA, USA. [15]Department of Medical & Molecular Genetics, King's College London, London, UK. [16]Division of Psychiatry, University College London, London, UK. [17]Genetics and Computational Biology, QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia. [18]Department of Psychiatry, Amsterdam UMC, Vrije Universiteit and GGZinGeest, Amsterdam, the Netherlands. [19]Psychosis Research Unit, Aarhus University Hospital, Risskov, Denmark. [20]iPSYCH, The Lundbeck Foundation Initiative for Integrative Psychiatric Research, Copenhagen, Denmark. [21]Department of Translational Research in Psychiatry, Max Planck Institute of Psychiatry, Munich, Germany. [22]Munich Cluster for Systems Neurology (SyNergy), Munich, Germany. [23]Institute of Translational Medicine, University of Liverpool, Liverpool, UK. [24]Department of Psychiatry, Harvard Medical School, Boston, MA, USA. [25]Department of Psychiatry, Massachusetts General Hospital, Boston, MA, USA. [26]Department of Psychiatry, Lausanne University Hospital and University of Lausanne, Lausanne, Switzerland. [27]Department of Psychiatry and Behavioral Sciences, Johns Hopkins University School of Medicine, Baltimore, MD, USA. [28]Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD, USA. [29]Psychiatric and Neurodevelopmental Genetics Unit (PNGU), Massachusetts General Hospital, Boston, MA, USA. [30]Stanley Center for Psychiatric Research, Broad Institute, Cambridge, MA, USA. [31]Department of Genetic Epidemiology in Psychiatry, Medical Faculty Mannheim, Central Institute of Mental Health, University of Heidelberg, Mannheim, Germany. [32]Department of Epidemiology, Erasmus University Medical Center Rotterdam, Rotterdam, the Netherlands. [33]Department of Child and Adolescent Psychiatry, Erasmus University Medical Center Rotterdam, Rotterdam, the Netherlands. [34]Department of Social and Behavioral Science, Harvard TH Chan School of Public Health, Boston, MA, USA. [35]Department of Psychiatry, Dalhousie University, Halifax, Nova Scotia, Canada. [36]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden. [37]Department of Psychiatry, Columbia University, Vagelos College of Physicians and Surgeons, New York, NY, USA. [38]Division of Translational Epidemiology, New York State Psychiatric Institute, New York, NY, USA. [39]Virginia Institute for Psychiatric and Behavioral Genetics, Department of Psychiatry, Virginia Commonwealth University, Richmond, VA, USA. [40]Center for Neurobehavioral Genetics, Semel Institute for Neuroscience and Human Behavior, University of California, Los Angeles, Los Angeles, CA, USA. [41]Present address: Helmholtz Pioneer Campus, Helmholtz Zentrum München, Neuherberg, Germany. *A list of members and affiliations appears in the Supplementary Note. ✉e-mail: na.cai@helmholtz-muenchen.de

Methods

Genome-wide associations. To obtain and access the difference between ORs of associations in different definitions of depression in the UK Biobank, as well as for smoking (data field 20160) and neuroticism (data field 20127), we performed logistic regression (or linear regression with –standard-beta for neuroticism) on all 5,276,842 common SNPs (MAF > 5% in all 337,198 White-British unrelated samples) in PLINK[65] (version 1.9) with 20 principal components and genotyping array as covariates.

Estimation of SNP heritability and genetic correlation among definitions of MDD. All estimates of h^2_{SNP} were computed with the PCGC[66] approach implemented with PCGC-ss[33], using 5,276,842 common SNPs (MAF > 5% in all 337,198 White-British unrelated samples). LD scores at SNPs were computed with LDSC[34] in 10,000 random samples drawn from the White-British samples in the UK Biobank as an LD reference, as well as the MAF at all 5,276,842 common SNPs in all 337,198 White-British samples as a MAF reference. Covariates were genotyping array and 20 principal components computed using samples in each definition of MDD with flashPCA[67]. Where we stratified each definition of MDD in the UK Biobank into two strata by risk factors such as sex (Supplementary Note), we computed specific principal components for each definition and stratum (see also the Supplementary Note and Supplementary Table 13).

Estimation of genetic correlation between definitions of MDD and other conditions. Summary statistics for other psychiatric conditions from previous GWAS studies were obtained as described in Supplementary Table 1. Association summary statistics for smoking and neuroticism in the UK Biobank were generated by GWAS (Supplementary Table 15 and 16, and Extended Data Fig. 3). We estimated the genetic correlation between definitions of MDD in the UK Biobank and each of these conditions using LDSC[43], with an LD reference panel generated with European (EUR) individuals from 1000 Genomes[68]. To obtain regional r_G, we partitioned the genome into 1,703 independent loci[64] and estimated regional r_G with rho-HESS[46], using an LD reference panel generated with EUR individuals from 1000 Genomes[68]. We estimated s.e. for each regional r_G and the total r_G across the genome using a jackknife approach implemented in HESS[36]. To assess the percentage of genome contributing to total r_G, we ranked all independent loci by their absolute value of regional r_G, and asked how many loci would contribute 90% of the total r_G.

Enrichment of SNP heritability in genes specifically expressed in tissues. We estimated the enrichment of h^2_{SNP} in genes specifically expressed in 44 tissues in the GTEx[47] project using the partitioned h^2_{SNP} framework in LDSC-SEG[46] and an LD reference panel generated with EUR individuals from 1000 Genomes[68]. We obtained tissue-specific gene expression annotations in GTEx tissues from LDSC-SEG and then estimated the enrichment of h^2_{SNP} in annotations that corresponded to each of the tissues together with 52 annotations in the baseline model[69]. We report the P value of the one-sided test of enrichment of h^2_{SNP} in genes specifically expressed in each tissue against the baseline.

Out-of-sample predictions of MDD. We performed out-of-sample prediction using individual-level genotype and phenotype data from the PGC29-MDD cohorts[5]. We obtained permissions from 20 cohorts with sample sizes greater than 500, among which 17 recorded endorsement of DSM-5 criteria A for MDD (Supplementary Note and Supplementary Table 21). We obtained PRSs from GWAS for each definition of depression in the UK Biobank, using LD-clumped (LD $r^2 < 0.1$) independent SNPs with P values for association below eight thresholds ($P < 10^{-4}$, 0.001, 0.01, 0.05, 0.1, 0.2, 0.5 and 1), and predicted MDD status in the 20 PGC cohorts using the Ricopili pipeline[70–82]. We obtained Nagelkerke's r^2 between the PRSs and MDD status, the AUC of the prediction and the variance of MDD status explained by the PRSs for each cohort. We also obtained the same measures for MDD status pulling data from all cohorts, controlling for cohort differences by including cohort as a covariate.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

Genotype and phenotype data used in this study are from the full release (imputation version 2) of the UK Biobank resource obtained under application no. 28709. We used publicly available summary statistics from other studies downloadable from the website of the Psychiatric Genomics Consortium (https://www.med.unc.edu/pgc/results-and-downloads), the references for which can be found in Supplementary Table 1. We also referenced the 2011 Census aggregate data from the UK Data Service (https://doi.org/10.5257/census/aggregate-2011-2).

References

65. Chang, C. C. et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience* **4**, 7 (2015).
66. Golan, D., Lander, E. S. & Rosset, S. Measuring missing heritability: inferring the contribution of common variants. *Proc. Natl Acad. Sci. USA* **111**, E5272–E5281 (2014).
67. Abraham, G. & Inouye, M. Fast principal component analysis of large-scale genome-wide data. *PLoS One* **9**, e93766 (2014).
68. 1000 Genomes Project Consortiumet al. A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
69. Finucane, H. K. et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet.* **47**, 1228–1235 (2015).
70. Lam, M. et al. RICOPILI: Rapid Imputation for COnsortias PIpeLIne. *Bioinformatics* https://doi.org/10.1093/bioinformatics/btz633 (2019).
71. Berardi, D. et al. Increased recognition of depression in primary care. Comparison between primary-care physician and ICD-10 diagnosis of depression. *Psychother. Psychosom.* **74**, 225–230 (2005).
72. Fry, A. et al. Comparison of sociodemographic and health-related characteristics of UK Biobank participants with those of the general population. *Am. J. Epidemiol.* **186**, 1026–1034 (2017).
73. Adams, M. J. et al. Factors associated with sharing email information and mental health survey participation in large population cohorts. *Int. J. Epidemiol.* https://doi.org/10.1101/471433 (2019).
74. Mullins, N. & Lewis, C. M. Genetics of depression: progress at last. *Curr. Psychiatry Rep.* **19**, 43 (2017).
75. Sullivan, P. F. et al. Psychiatric genomics: an update and an agenda. *Am. J. Psychiatry* **175**, 15–27 (2018).
76. Coyne, J. C., Schwenk, T. L. & Smolinski, M. Recognizing depression: a comparison of family physician ratings, self-report, and interview measures. *J. Am. Board Fam. Pract.* **4**, 207–215 (1991).
77. Nevin, R. L. Low validity of self-report in identifying recent mental health diagnosis among U.S. service members completing Pre-Deployment Health Assessment (PreDHA) and deployed to Afghanistan, 2007: a retrospective cohort study. *BMC Public Health* **9**, 376 (2009).
78. Clarke, D. E. et al. DSM-5 field trials in the United States and Canada. Part I: study design, sampling strategy, implementation, and analytic approaches. *Am. J. Psychiatry* **170**, 43–58 (2013).
79. Spitzer, R. L., Forman, J. B. & Nee, J. DSM-III field trials. I. Initial interrater diagnostic reliability. *Am. J. Psychiatry* **136**, 815–817 (1979).
80. Keller, M. B. et al. Results of the DSM-IV mood disorders field trial. *Am. J. Psychiatry* **152**, 843–849 (1995).
81. McCarthy, S. et al. A reference panel of 64,976 haplotypes for genotype imputation. *Nat. Genet.* **48**, 1279–1283 (2016).
82. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).

Acknowledgements

We thank O. Weissbrod, A. Dahl, H. Shi and V. Zuber for insightful discussions. N.C. is supported by the ESPOD Fellowship from European Bioinformatics (EMBL-EBI) and Wellcome Sanger Institute. A.V. is supported by the Swedish Brain Foundation. C.M.L. and G.B. are funded by the National Institute for Health Research (NIHR) Maudsley Biomedical Research Centre at South London Maudsley Foundation Trust and King's College London. In the last 3 years, M.M.W. has received research funds from the US National Institute of Mental Health (NIMH), the Templeton Foundation and the Sackler Foundation and has received royalties for publication of books on interpersonal psychotherapy from Perseus Press and Oxford University Press, on other topics from the American Psychiatric Association Press and royalties on the social adjustment scale from Multihealth Systems. The CoLaus|PsyCoLaus study was and is supported by research grants from GlaxoSmithKline, the Faculty of Biology and Medicine of Lausanne and the Swiss National Science Foundation (grants 3200B0-105993, 3200B0-118308, 33CSCO-122661, 33CS30-139468, 33CS30-148401 and 33CS30-177535/1). The PGC has received major funding from the US NIMH and the US National Institute of Drug Abuse (U01 MH109528 and U01 MH1095320). This research was conducted using the UK Biobank resource under application no. 28709 and with the support and collaboration from all investigators who make up the MDD Working Group of the PGC (full list in the Supplementary Note). We are greatly indebted to the hundreds of thousands of individuals who have shared their life experiences with the UK Biobank and PGC investigators.

Author contributions

N.C. and J.F. designed the study. N.C. and J.A.R. performed the analyses. N.C. and J.F. obtained the data from the UK Biobank resource. M.J.A., T.F.M.A., G.B., E.M.B., T.-K.C., A.J.F., H.J.G., S.P.H., D.F.L., C.M.L., G.L., N.G.M., Y.M., O.M., B.M.-M., B.W.J.H.P., R.H.P., G.P., J.B.P., M.P., J.S., J.W.S., F.S., H.T., R.U., S.V.d.A., A.V., M.M.W. and all investigators from the MDD Working Group of the PGC contributed data from the PGC. N.C., K.S.K. and J.F. interpreted the results and wrote the manuscript.

Competing interests

C.M.L. is on the scientific advisory board of Myriad Neuroscience. H.J.G. has received travel grants and speaker's honoraria from Fresenius Medical Care, Neuraxpharm and

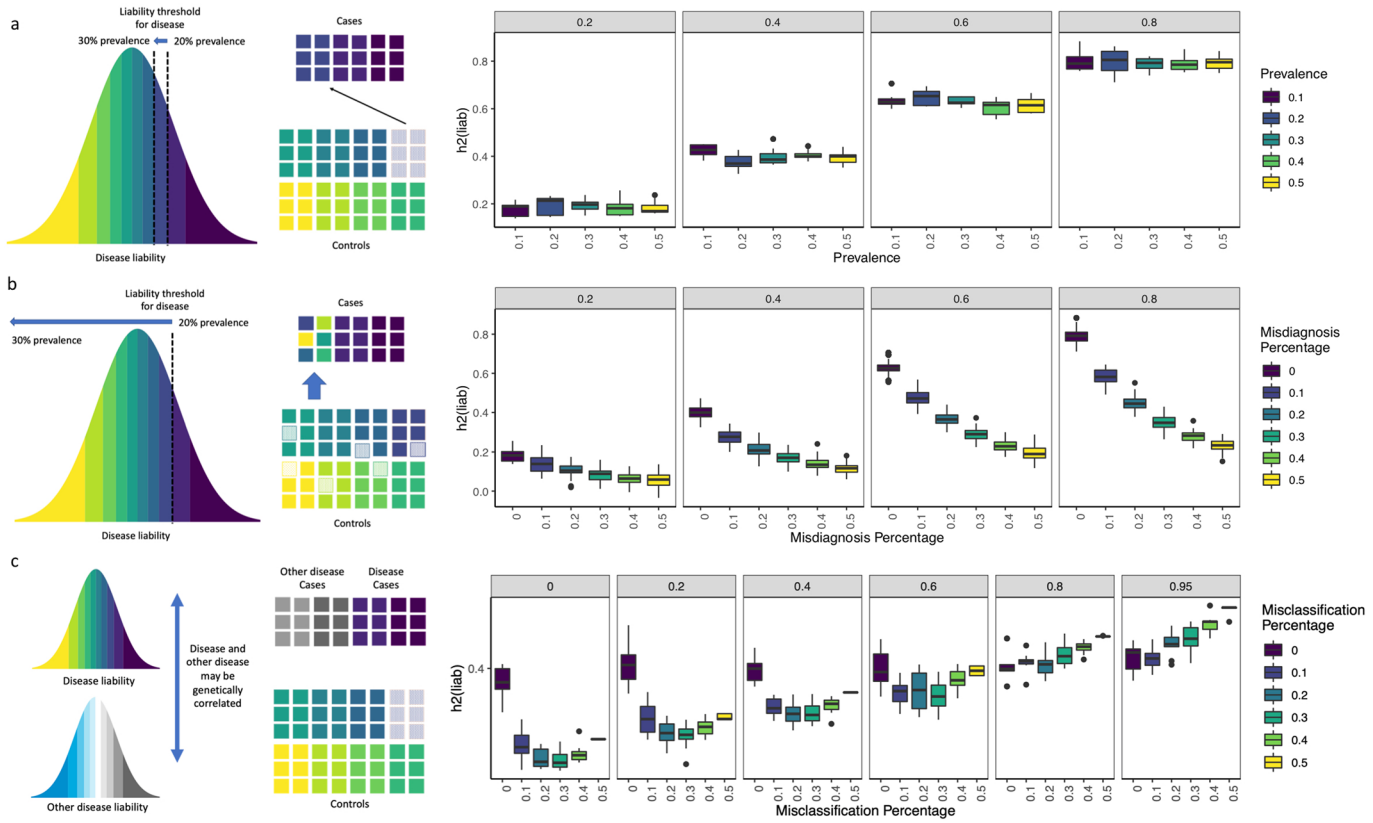
Janssen Cilag as well as research funding from Fresenius Medical Care. B.W.J.H.P. has received (non-related) research grants from Jansen Research and Boehringer Ingelheim.

Additional information

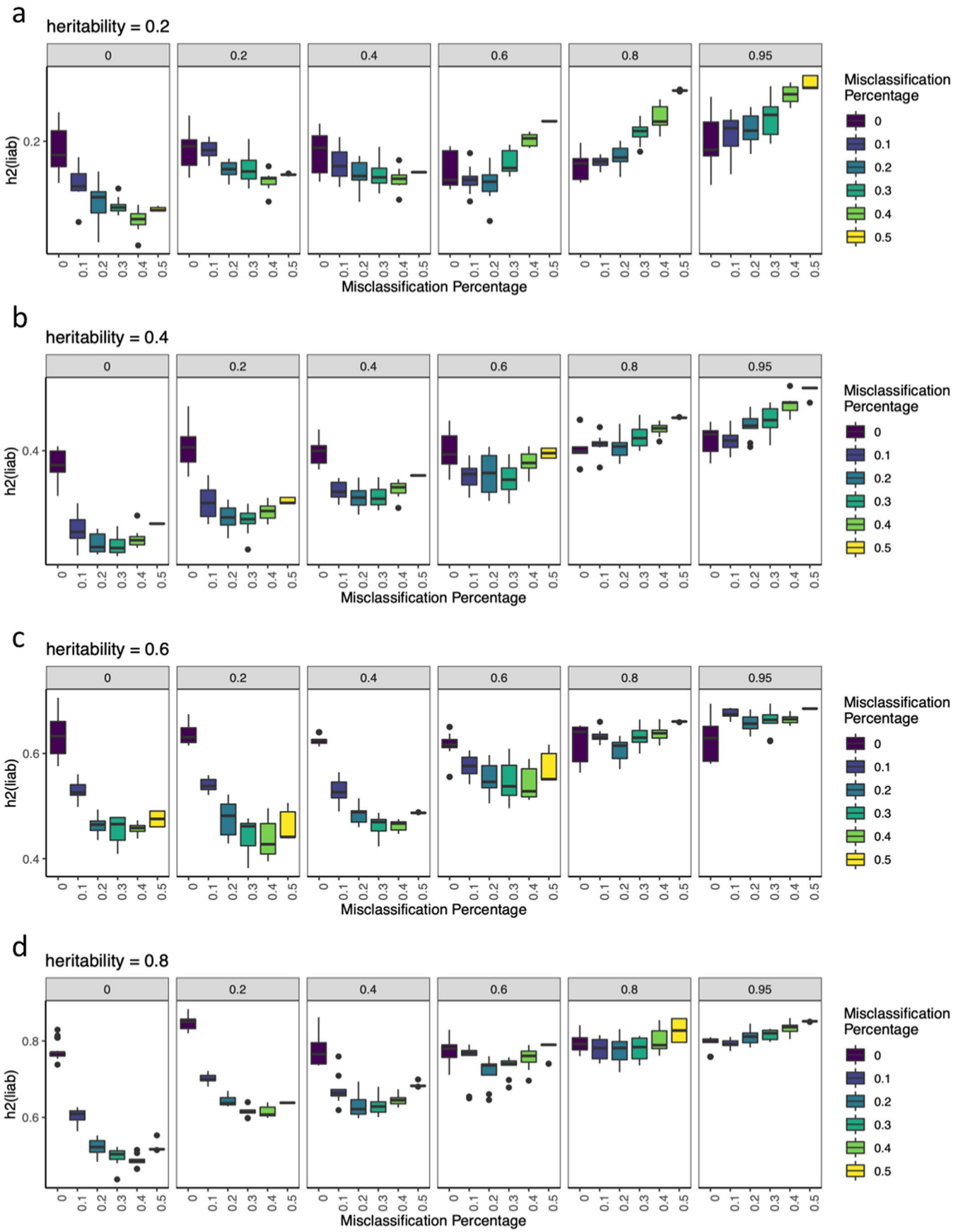
Extended data is available for this paper at https://doi.org/10.1038/s41588-020-0594-5.

Supplementary information is available for this paper at https://doi.org/10.1038/s41588-020-0594-5.

Correspondence and requests for materials should be addressed to N.C.

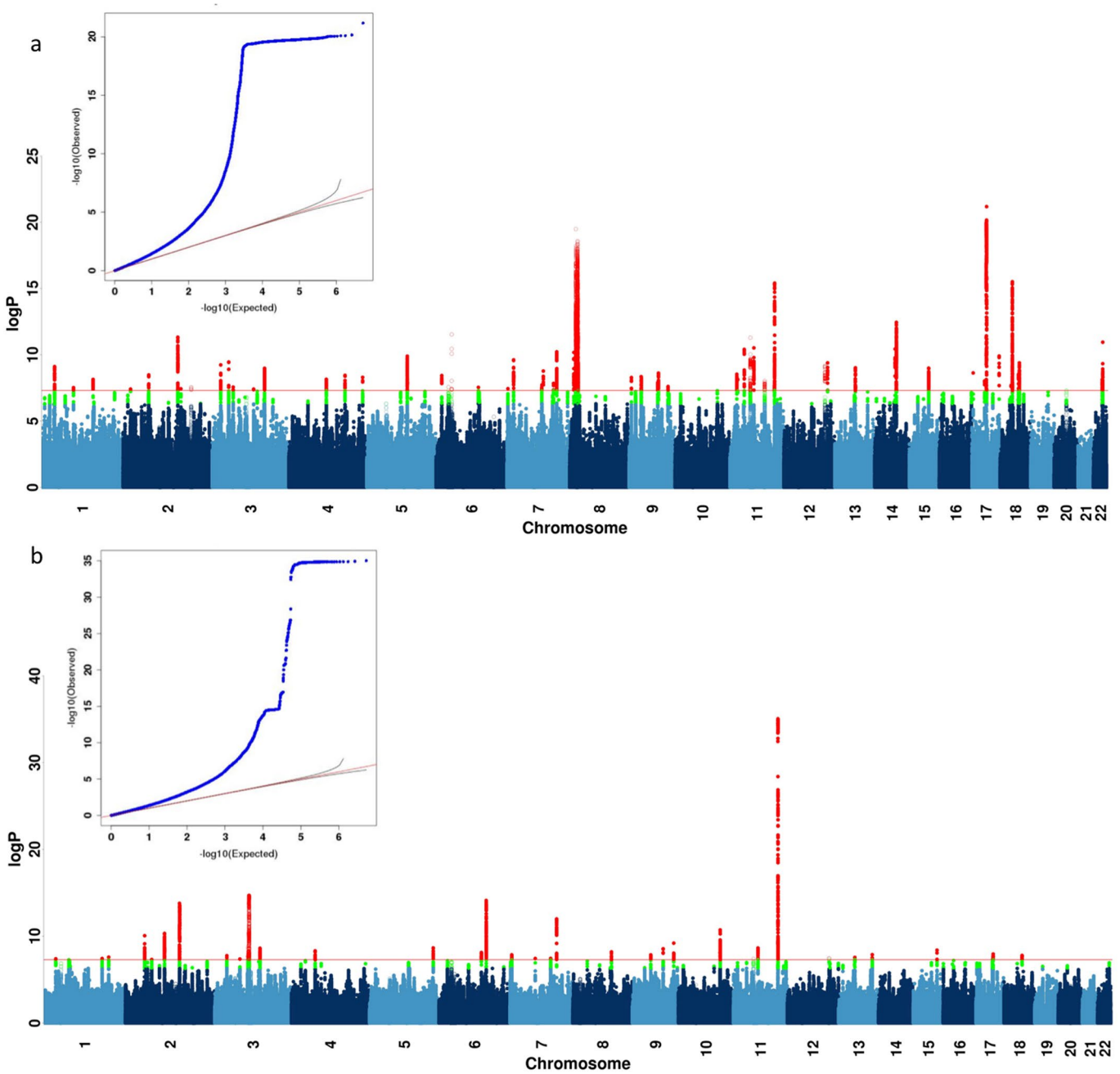


Extended Data Fig. 1 | Simulations of misdiagnosis and misclassification. a-c, Each boxplot show h^2_{SNP} estimates from 10 simulated phenotypes, with upper and lower boundaries of boxes represent the first to third quartiles of all estimates, and the whiskers extends to 1.5 times the interquartile range of the estimates. **a**, This figure shows that liability scale h^2_{SNP} does not change with shifting of liability threshold $K_i \in \{0.1, 0.2, 0.3, 0.4, 0.5\}$ for simulated heritabilities $h_i^2 \in \{0.2, 0.4, 0.6, 0.8\}$. **b**, The figure shows that liability scale h^2_{SNP} is deflated with increasing percentage of controls being misdiagnosed as cases, when prevalence of diagnosed cases is kept constant at K_i=0.2, for simulated heritabilities $h_i^2 \in \{0.2, 0.4, 0.6, 0.8\}$. **c**, This figure shows liability scale h^2_{SNP} is deflated with increasing percentage of misclassification of cases of "other" disease as cases of focal disease, if rG between the two diseases are moderate to low, for simulated $h_{i,1}^2 = 0.4$, for each of which all cases at prevalence $K_{i,1}$=0.2 are correctly identified as cases.

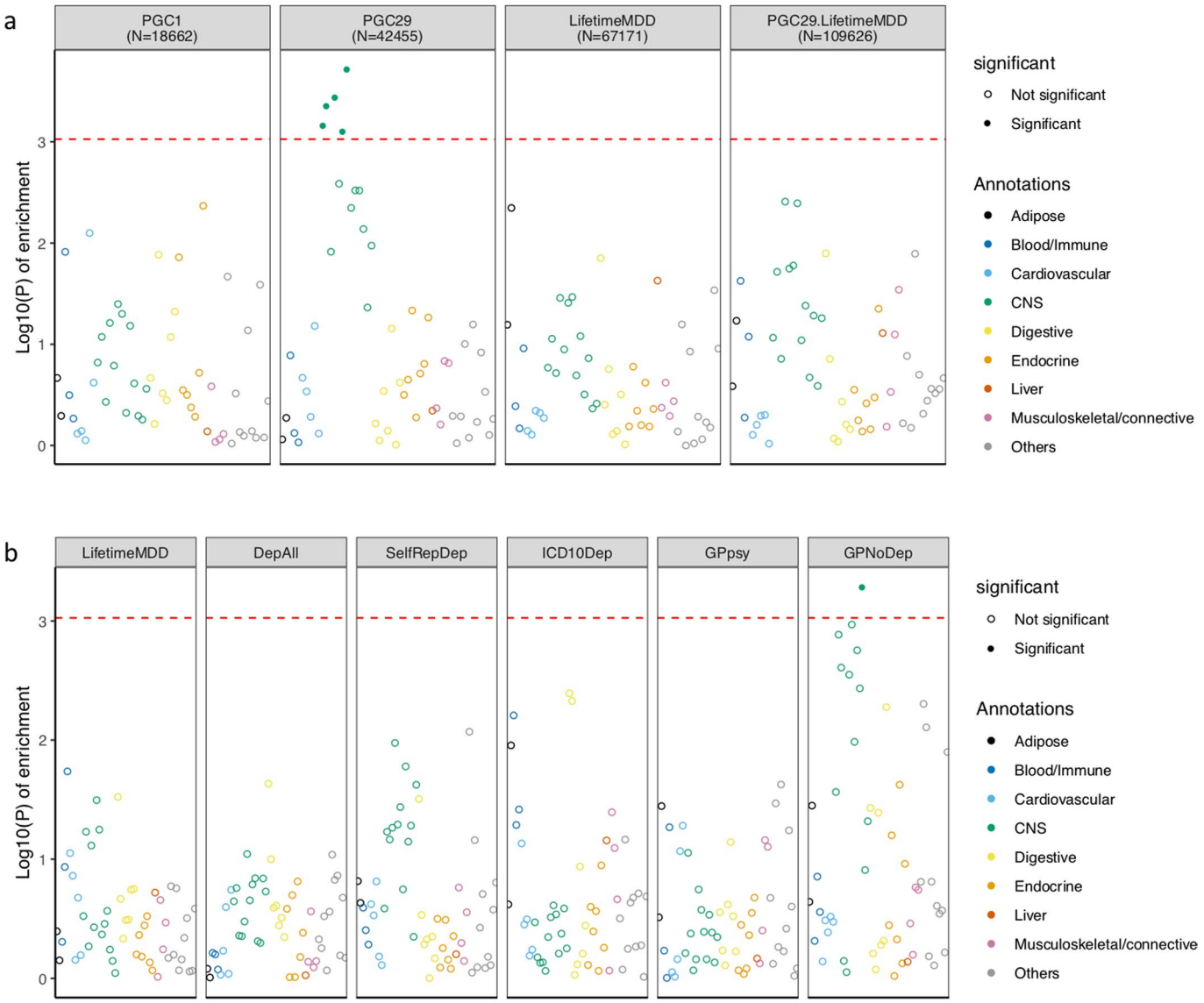


Extended Data Fig. 2 | See next page for caption.

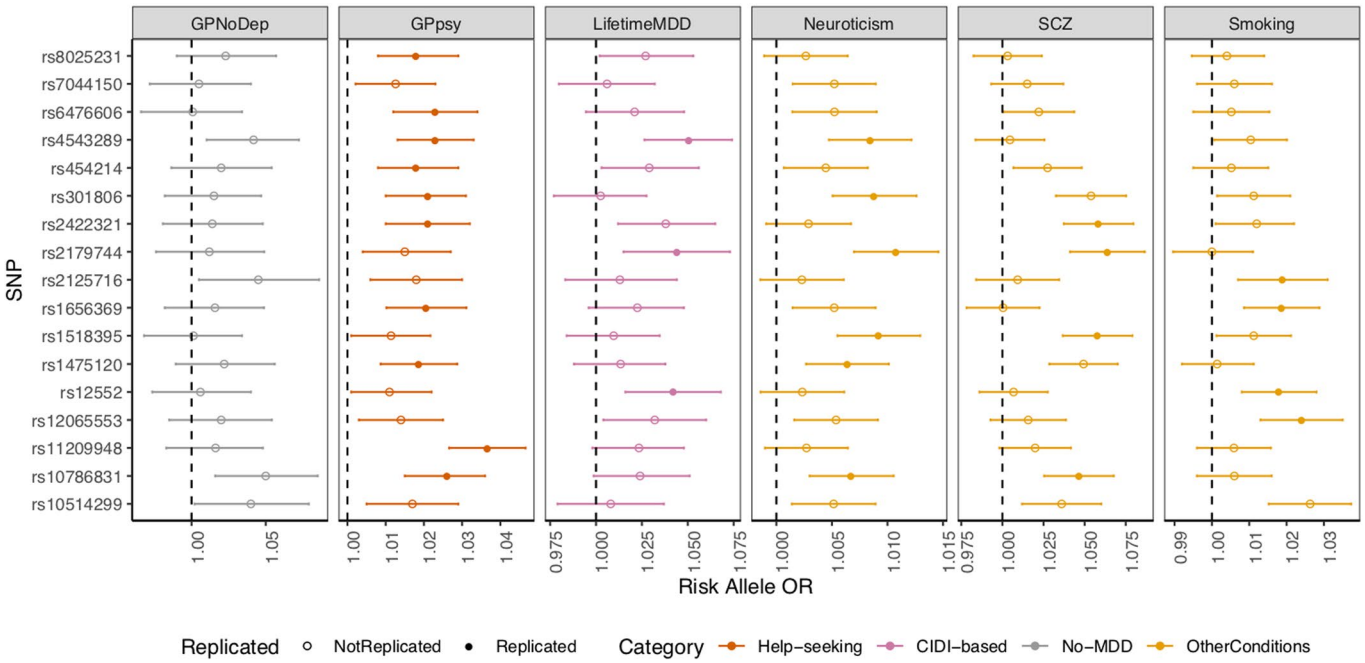
Extended Data Fig. 2 | Simulations of misclassification at different heritabilities. a-d, These figures shows the estimated h^2_{SNP} using –pcgc option with–prevalence K in LDAK, plotted on the *y*-axis) of binary traits ($y_{i,1}$, where i ∈{1..10}) with simulated $h^2_{i,1}$0.2, 0.4, 0.6, and 0.8, 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∈{1..10}) of equal $h^2_{i,1}$ and prevalence as cases of $y_{i,1}$. Genetic correlations between $y_{i,1}$ and $y_{i,2}$ (rG_i∈{0, 0.2, 0.4, 0.6, 0.8, 0.95}) are shown in the grey bars above each panel. Each boxplot show h^2_{SNP} estimates from 10 simulated phenotypes, with upper and lower boundaries of boxes represent the first to third quartiles of all estimates, and the whiskers extends to 1.5 times the interquartile range of the estimates.



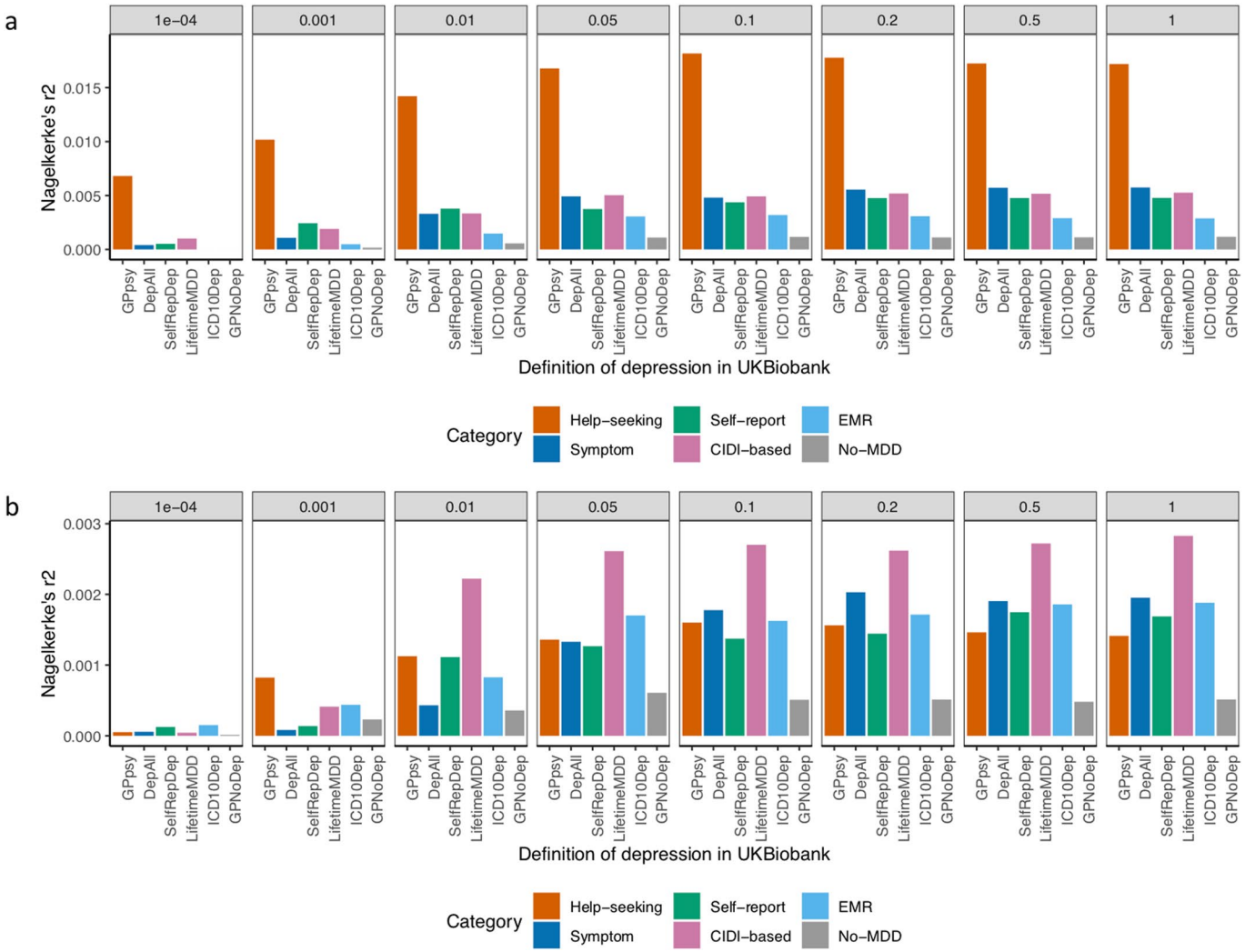
Extended Data Fig. 3 | GWAS on neuroticism and smoking in UK Biobank. a, b, This figure shows the Manhattan plot of neuroticism score (data field 20127, quantitative trait from 0 to 12) in 274,107 individuals and ever smoked status (data field 20160, binary trait of 0 for "No", and 1 for "Yes") in 336,066 individuals in UK Biobank 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)[32] 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 (1-Mb regions) in Supplementary Tables 10 and 11.



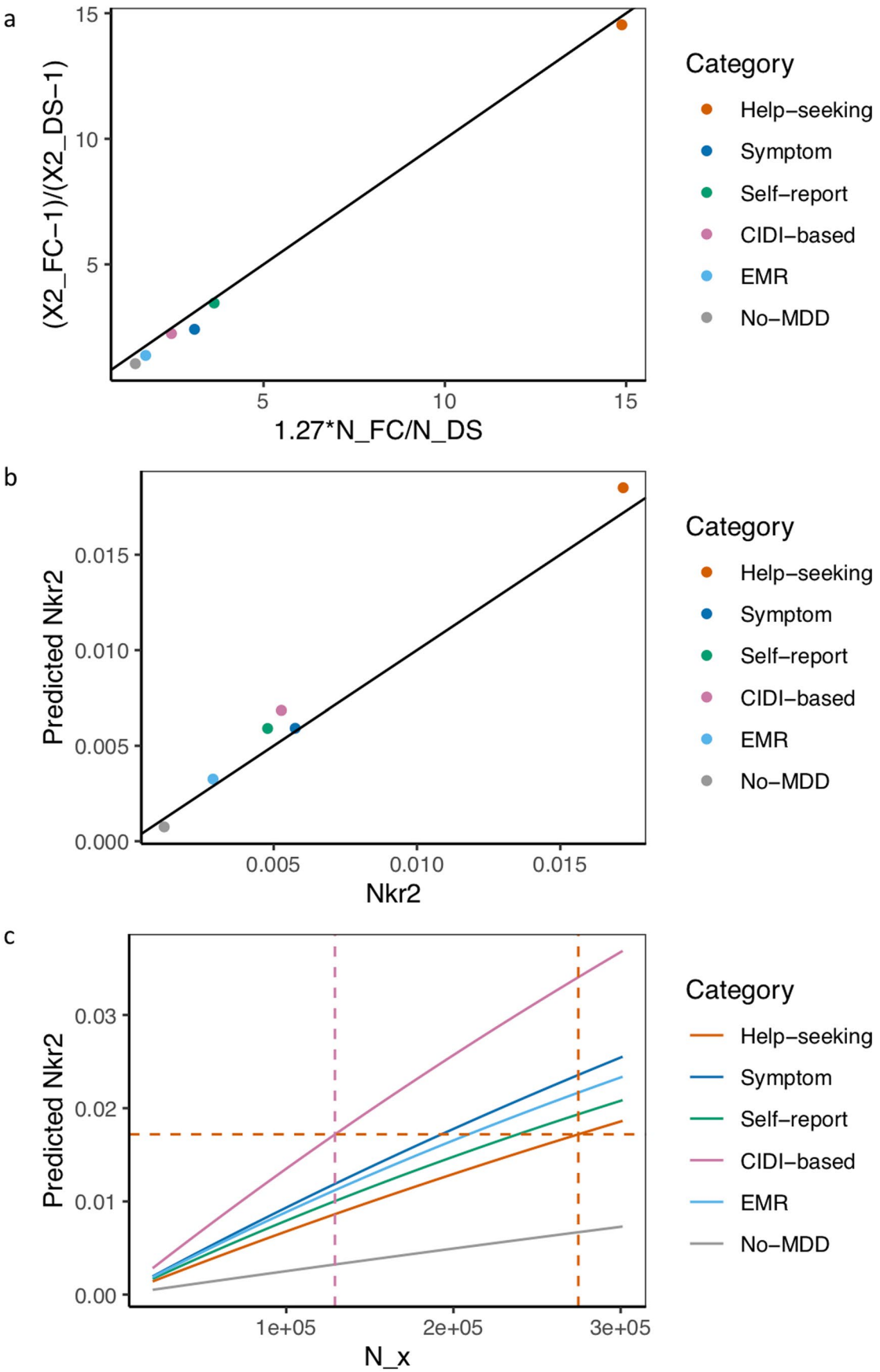
Extended Data Fig. 4 | LDSC-SEG analysis of tissue-specific enrichment of h^2_{SNP}. a, This figure shows -$\log_{10}(P)$ of enrichment in heritability in genes specifically expressed in 44 GTEx tissues, estimated using partitioned heritability in LDSC-SEG, on LifetimeMDD (n = 67,171), PGC1-MDD (n = 18,759), PGC29 (n = 42,455) and a meta-analysis of LifetimeMDD and PGC29 (n = 109,626, PC29.LifetimeMDD, Methods). While PGC29 shows CNS enrichment, neither LifetimeMDD nor the meta-analysis shows the same enrichment. This suggests sample size and differences in genetic architecture and cohort heterogeneity affects results from LDSC-SEG. **b**, This figure shows the same analysis performed on down-sampled data for each definition of depression. Each definition is randomly down-sampled to 7,500 cases and 42,500 controls, a constant prevalence of 0.15, to remove confounding from sample size and difference in statistical power on the enrichment analysis. This figure shows that at equal sample size and prevalence, GPNoDep (no-MDD Help-seeking phenotype) is the only one showing CNS enrichment, suggesting it may be driving the CNS enrichment signal in GPpsy in Fig. 5.



Extended Data Fig. 5 | GWAS hits from 23andMe are not specific to MDD. This figure shows the odds ratios of risk alleles (Risk Allele ORs) at 17 loci significantly associated with help-seeking based definitions of MDD in 23andMe27, in GWAS conducted on CIDI-based (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 Supplementary Table 20. Error bars show the standard errors of the estimates.

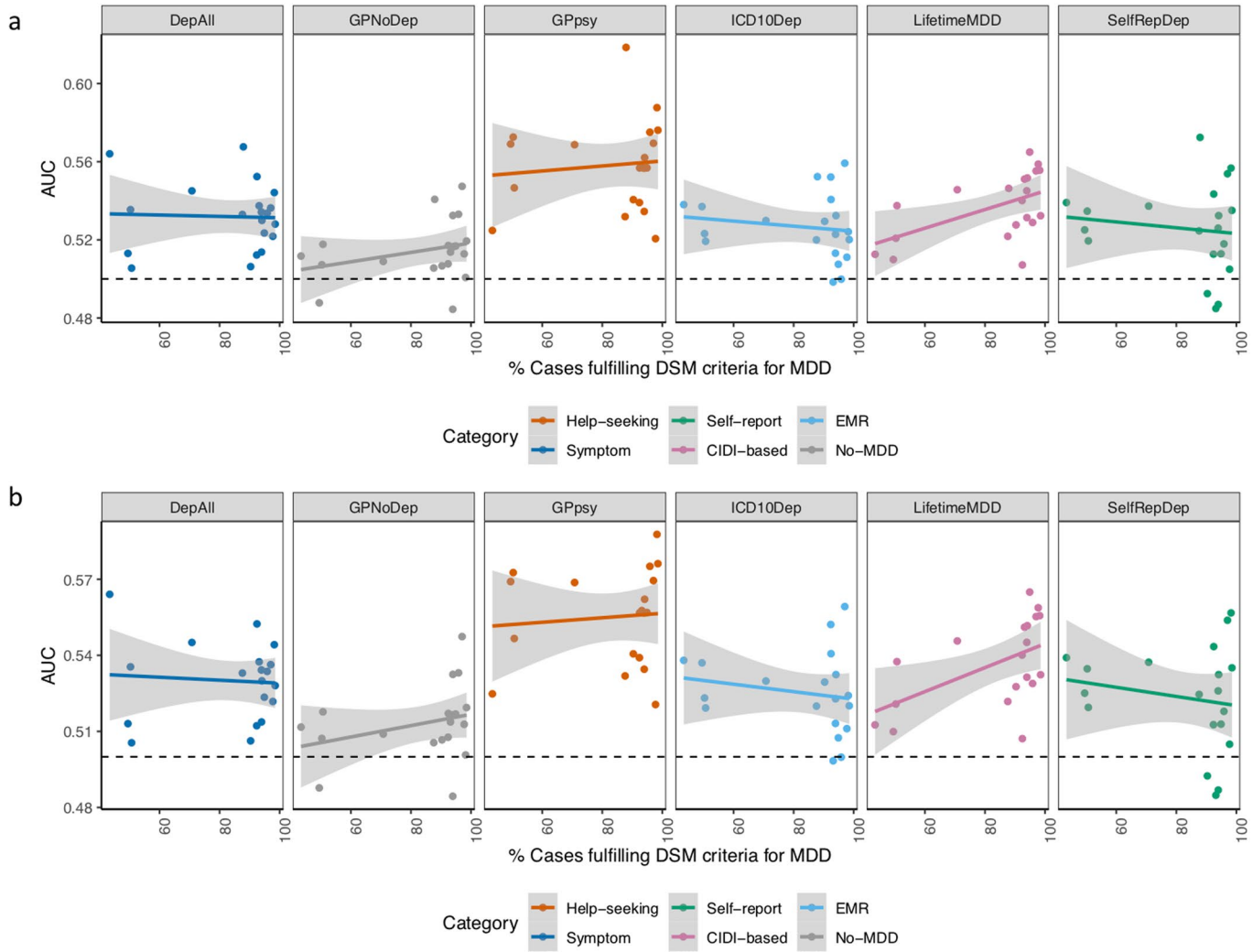


Extended Data Fig. 6 | Out-of-sample prediction in PGC cohorts. a, This figure shows the Nagelkerke's r^2 of polygenic risk scores (PRS) calculated for each definition of depression in UK Biobank and MDD status indicated in 19 PGC29-MDD cohorts, while controlling for cohort specific effects. PRS were calculated using effect sizes at independent (LD $r^2 < 0.1$) SNPs passing *P*-value thresholds 10^{-4}, 0.001, 0.01, 0.05, 0.01, 0.2, 0.5 and 1 respectively, in GWAS performed on all definitions of depression in UK Biobank. **b**, This figure shows the same analysis performed on down-sampled data (7,500 cases, 42,500 controls) for each definition of depression.



Extended Data Fig. 7 | See next page for caption.

Extended Data Fig. 7 | Relationship between effective sample size and prediction accuracy. a, This figure shows the relationship between the ratio of effective sample sizes between the full cohort (N_{FC}) and down-sampled (N_{DS}) data for each definition of depression and the ratio of their mean Chi-square (χ^2) statistic from GWAS, with black line $x = y$ for reference. Across all definitions of depression, $\frac{\overline{\chi^2_{FC}}-1}{\overline{\chi^2_{DS}}-1}$ is highly correlated with $\frac{N_{FC}}{N_{DS}}$ (Pearson $r^2 = 0.999$, $P = 5.50 \times 10^{-7}$), and $\frac{N_{FC}}{N_{DS}}$ has an effect of beta = 1.27 (s.e. = 0.02) on $\frac{\overline{\chi^2_{FC}}-1}{\overline{\chi^2_{DS}}-1}$. **b**, This figure shows the Nagelkerke's r^2 (Nkr2) for MDD status in PGC29 cohorts predicted for PRS of different definitions of depression at N_{FC}, plotted against their respective empirical Nkr2 at N_{FC}, both at P-value threshold = 1. The Pearson correlation r^2 between predicted and actual NKr2 across all definitions were 0.989 ($P = 4.46 \times 10^{-5}$). **c**, This figure shows for each definition of depression the effective sample size N_X required for each predicted Nkr2 in out-of-sample prediction of MDD status in PGC29 cohorts. While N_x= 274,677 (indicated with orange vertical dotted line) for GPpsy to achieve a Nkr2 of 0.0172 (indicated with orange horizontal dotted line), a smaller N_x= 129,106 (indicated with pink vertical dotted line) is needed to achieve the same Nkr2 for LifetimeMDD.



Extended Data Fig. 8 | Prediction accuracy in cohorts with different percentage of DSM MDD cases. a, This figure shows the area under the curve (AUC) of polygenic risk scores (PRS) calculated for each definition of depression in UK Biobank and MDD status indicated in 20 PGC29-MDD cohorts at *P*-value threshold of 0.1 (using all SNPs after LD-clumping, see results at all *P*-value thresholds in Supplementary Table 23), plotting AUC for each cohort against their respective percentage of cases fulfilling DSM-5 criteria A for MDD (see Supplementary Table 21). It shows that strictly defined CIDI-based LifetimeMDD is the only definition of depression in UK Biobank that shows increases in AUC as percentage of cases fulfilling DSM-5 criteria A for MDD in PGC cohorts increases, despite not giving the highest AUC. **b**, This figure shows the same analysis removing the PGC29-MDD cohort rad3, which is the outlier giving AUC > 0.6 in GPpsy in **a**. As this is a UK-based cohort, it is possible it contains relatives of individuals in UK Biobank that upwardly biased prediction accuracy in it. For all analysis shown in Fig. 7, Extended Data Figs. 6 and 7 and Supplementary Table 23, we have removed this cohort.

nature research

Corresponding author(s): Na Cai

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated
☐	☒	Clearly defined error bars *State explicitly what error bars represent (e.g. SD, SE, CI)*

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about availability of computer code

Data collection	No new data was collected for this study. We used the genotype and phenotype data from 502,637 samples in the full release (imputation version 2) of the UK Biobank Resource under application no. 28709. We used publicly available summary statistics from other studies downloadable from the website of Psychiatric Genomics Consortium (https://www.med.unc.edu/pgc/results-and-downloads), and the references for which can be found in Supplemental Table 1. We also referenced the 2011 Census aggregate data from the UK Data Service (http://dx.doi.org/10.5257/census/aggregate-2011-2).
Data analysis	We conducted our analyses using the following published and publicly available software: 1) for computing PCA on all White British sample in UKBiobank: flashPCA (https://github.com/gabraham/flashpca); 2) for GWAS using logistic regression: PLINK version 1.9 (https://www.cog-genomics.org/plink2); 3) for estimation of heritability, genetic correlation and/or enrichment of heritability: LDSC v1.0.0 (https://github.com/bulik/ldsc), PCGCs (https://github.com/omerwe/PCGCs), HESS and rho-HESS (https://huwenboshi.github.io/hess/); 4) simulations of pairs of phenotypes with particular heritability and genetically correlations: LDAK v5 (http://dougspeed.com/downloads2/); 5: sanity check for population structure using per chromosome heritability: BOLT-REML v2.3 (https://data.broadinstitute.org/alkesgroup/BOLT-LMM/)

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Genotype and phenotype data used in this study are from the full release (imputation version 2) of the UKBiobank Resource obtained under application no. 28709. We used publicly available summary statistics from other studies downloadable from the website of Psychiatric Genomics Consortium (https://www.med.unc.edu/pgc/results-and-downloads), and the references for which can be found in Supplemental Table 1. We also referenced the 2011 Census aggregate data from the UK Data Service (http://dx.doi.org/10.5257/census/aggregate-2011-2).

Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/authors/policies/ReportingSummary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	We used the genotype and phenotype data from 337,198 samples, a subset of 502,637 samples in the full release (imputation round 2) of the UK Biobank Resource under application no. 28709 (exclusion criteria see below). We defined 7 definitions of depression in UKBiobank, based on criteria detailed in the results section of the main text "Definitions of depression in UKBiobank", and Supplemental Methods section "DSM-based definition of MDD". The number of samples in each of the definitions of depression are tabulated in Supplemental Tables S4 and S5.
Data exclusions	We excluded all samples with 1) poor genotyping quality, 2) high level of relatedness to other samples, 3) ancestries other than White British as indicated by the QC metrics from UKBiobank (Bycroft et al 2018, https://doi.org/10.1038/s41586-018-0579-z), 4) sex chromosome aneuploidy, 5) withdrawal of consent from being included in research on data from the UKBiobank, 6) a history of substance abuse, and 7) manic or psychotic conditions. This gives us our final sample of 337,198 White-British, unrelated individuals. Details of exclusion criteria can be found in Supplemental Methods section "Sample filtering".
Replication	Not applicable, as our aim is not to identify particular loci of association with depression, but observe patterns across genetic architectures of different definitions of depression. We show that minimal phenotyping definitions of depression in UKBiobank (GPpsy) give similar heritability estimates with a previous study also employing a minimal phenotyping approach that does not contain overlapping samples as UKBiobank: 23andMe (Hyde et al 2016, https://doi.org/10.1038/ng.3623). Similarly, we replicated the trend that minimal phenotyping definitions of depression in UKBiobank (GPpsy) give GWAS hits with high similarity in directions of effect as smoking and neuroticism, in a previous study also employing a minimal phenotyping approach that does not contain overlapping samples as UKBiobank: 23andMe (Hyde et al 2016, https://doi.org/10.1038/ng.3623).
Randomization	Not applicable.
Blinding	Not applicable.

Reporting for specific materials, systems and methods

Materials & experimental systems

n/a	Involved in the study
☒	☐ Unique biological materials
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☒	☐ Human research participants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging