Pleiotropic heritability quantifies the shared genetic variance of common diseases

Yujie Zhao[1,2], Benjamin Strober[2], Kangcheng Hou[2], Gaspard Kerner[2], John Danesh[3,4,5,6,7,8], Steven Gazal[9,10,11], Wei Cheng[1,12,13], Michael Inouye[3,4,5,7,14,15], Alkes L. Price*[2,16,17], Xilin Jiang*[2,3,4,16,18]

Correspondence: Y.Z. (yjzhao22@m.fudan.edu.cn), A.L.P. (aprice@hsph.harvard.edu), X.J. (xj262@medschl.cam.ac.uk)

*: co-senior authors.

[1] Institute of Science and Technology for Brain-Inspired Intelligence, Fudan University, Shanghai, China.
[2] Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, USA
[3] British Heart Foundation Cardiovascular Epidemiology Unit, Department of Public Health and Primary Care, University of Cambridge, Cambridge UK
[4] Victor Phillip Dahdaleh Heart and Lung Research Institute, University of Cambridge, Cambridge UK
[5] British Heart Foundation Centre of Research Excellence, University of Cambridge, Cambridge, UK
[6] National Institute for Health and Care Research Blood and Transplant Research Unit in Donor Health and Behaviour, University of Cambridge, Cambridge, UK
[7] Health Data Research UK Cambridge, Wellcome Genome Campus and University of Cambridge, Cambridge, UK
[8] Department of Human Genetics, Wellcome Sanger Institute, Hinxton, UK
[9] Department of Population and Public Health Sciences, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA
[10] Center for Genetic Epidemiology, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA
[11] Department of Quantitative and Computational Biology, University of Southern California, Los Angeles, CA, USA
[12] State Key Laboratory of Brain Function and Disorders (Fudan University), Ministry of Education; Shanghai, China.
[13] Fudan ISTBI—ZJNU Algorithm Centre for Brain-inspired Intelligence, Zhejiang Normal University, Zhejiang, China.
[14] Cambridge Baker Systems Genomics Initiative, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK
[15] Cambridge Baker Systems Genomics Initiative, Baker Heart and Diabetes Institute, 75 Commercial Rd, Melbourne 3004, Victoria, Australia
[16] Broad Institute of MIT and Harvard, Cambridge, MA, USA
[17] Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA, USA
[18] Cambridge Centre for AI in Medicine, University of Cambridge, Cambridge, UK

Abstract

Common diseases are highly pleiotropic, but the overall contribution of pleiotropy to a target disease's architecture is unknown, as most studies focus on genetic correlations with each auxiliary disease in turn. Here we propose a new method, pleiotropic heritability with bias correction (PHBC), to estimate pleiotropic heritability ($h^2{}_{pleio}$), defined as the liability-scale genetic variance of a target disease that is shared with a specific set of auxiliary diseases. We estimate $h^2{}_{pleio}$ from GWAS summary statistics by estimating the proportion of variance explained from an estimated genetic correlation matrix and employing a Monte-Carlo bias correction procedure to account for sampling noise in genetic correlation estimates. Simulations showed that PHBC produces approximately unbiased estimates of pleiotropic heritability. The average ratio of pleiotropic heritability vs. total SNP-heritability ($h^2{}_{pleio}/h^2$) across 15 diseases from the UK Biobank (spanning 7 disease categories) was 34% (s.e. 7%). Several diseases were dominated by pleiotropic heritability, including depression (71%) and type 2 diabetes (65%). Pleiotropic heritability was broadly distributed across disease categories, with $h^2{}_{pleio}/h^2$ decreasing only slightly when removing all auxiliary diseases in the target disease category (avg = 31% (s.e. 7%)) and only moderately when further removing one other (most informative) category whose removal had the greatest impact (avg = 20% (s.e. 3%)). Average $h^2{}_{pleio}/h^2$ increased to 44% (s.e. 9%) when adding 16 auxiliary quantitative traits in UK Biobank, and 50% (s.e. 5%) when further adding 30 auxiliary diseases from large GWAS meta-analyses. On average, $h^2{}_{pleio}/h^2$ was 2.4x larger than the proportion of liability-scale total phenotypic variance explained by the same set of auxiliary diseases, implying higher pleiotropy for genetic effects than the effects of non-genetic exposures. In conclusion, we have uncovered pervasive sharing of genetic aetiologies, with roughly half of common disease heritability being pleiotropic with diseases from a broad range of disease categories, which strongly motivates the importance of multi-disease approaches to risk prediction and therapeutic development.

Introduction

Common diseases and complex traits are highly pleiotropic and extensive efforts have been made to understand the shared components of disease risk[1–6]. Previous studies have quantified and characterized the shared genetic variance between pairs of diseases[7–13], identified latent genetic components that are shared across many diseases[14–22], and leveraged the sharing of genetic variant associations across diseases to increase association power[16,18,23–26]. However, to our knowledge, we lack a method to estimate the genetic variance of a given target disease that is shared with other diseases and traits; such a method would inform efforts to prioritize therapeutic targets while accounting for disease comorbidities and potential drug side effects[27–31].

To address this gap, we introduce pleiotropic heritability ($h^2{}_{pleio}$), which quantifies the genetic variance of a target disease that is shared with a specific set of auxiliary diseases and traits. We develop a new method, pleiotropic heritability with bias correction (PHBC), to estimate $h^2{}_{pleio}$ from GWAS summary statistics. PHBC employs a Monte-Carlo bias correction procedure to correct bias arising from sampling noise in genetic correlation estimates due to finite GWAS sample size. We validate PHBC using extensive simulations then estimate $h^2{}_{pleio}$ for 15 diseases in UK Biobank and, subsequently, for 30 diseases with GWAS summary statistics available from extensive meta-analyses. Finally, we extend PHBC to estimate the liability-scale total phenotypic variance of a target disease that is shared with a set of auxiliary diseases.

Results

Definition of pleiotropic heritability

We define pleiotropic heritability as the liability-scale genetic variance (estimated from SNPs) of a target disease that is shared with a specific set of auxiliary diseases. We can view the genetic value of the target disease as the sum of a disease-specific genetic value that is not shared with the auxiliary diseases and a pleiotropic genetic value that is shared with auxiliary diseases (**Figure 1**), with the variance of latter denoted the pleiotropic heritability ($h^2{}_{pleio}$). We used the ratio of pleiotropic heritability to total SNP-heritability ($h^2{}_{pleio}/h^2$) to quantify the proportion of genetic variance (from SNPs) that is pleiotropic.

In detail, $h^2{}_{pleio}$ is the liability-scale variance explained by the weighted linear combination of auxiliary disease genetic liabilities that explains the maximum proportion of target disease heritability (in the entire population):

$$h^2_{pleio}/h^2 = \max_{\alpha_j} \; r^2\left[\sum_{j=1}^{n} \alpha_j G_j \, , G_T\right] \qquad (1)$$

where α_j is the weight of the jth auxiliary disease (D_j), G_j is the genetic liability of D_j , and G_T is the genetic liability of the target disease; $h^2{}_{pleio}$ is the corresponding r^2 with the total liability L_T instead of the genetic liability G_T. We note that when an auxiliary disease D_j has zero genetic correlation with the target disease, its contribution to pleiotropic heritability of the target disease

is zero. In this study, we restrict both pleiotropic heritability and total heritability to heritability explained by SNPs.

Estimation of pleiotropic heritability

We estimate ${h^2}_{pleio}$ from estimates of total SNP-heritability and genetic correlations between the target and auxiliary diseases, and employ a Monte-Carlo bias correction procedure to obtain unbiased estimates. We emphasize the need for a bias correction procedure: for example, if the target disease has true genetic correlation of 0 with each auxiliary disease (implying true pleiotropic heritability of 0), genetic correlation *estimates* will be nonzero due to sampling noise, implying a pleiotropic heritability *estimate* greater than 0 in the absence of a bias correction procedure. In this study, we apply cross-trait LD Score regression (LDSC)[32] to summary association statistics and reference LD to estimate genetic correlations. We obtain standard errors via genomic block-jackknife, as in cross-trait LDSC[32]. We have publicly released open-source software implementing our method (Code Availability).

In detail, it can be shown that Eq. 1 is equivalent to (see **Methods**):

$$ {h^2}_{pleio}/h^2 = r_g[\boldsymbol{D}, T]^T r_g[\boldsymbol{D}, \boldsymbol{D}]^{-1} r_g[\boldsymbol{D}, T] \quad (2) $$

where h^2 is the SNP-heritability of the target disease, $r_g[\boldsymbol{D}, T]$ is the $n\times1$ vector of genetic correlations between the target and auxiliary diseases, and $r_g[\boldsymbol{D}, \boldsymbol{D}]$ is the $n\times n$ genetic correlation matrix among the set of auxiliary diseases, where n is the number of auxiliary diseases (**Methods**). We initially estimate ${h^2}_{pleio}/h^2$ by estimating the corresponding quantities in Eq. 2. However, the errors in estimates of $r_g[\boldsymbol{D}, \boldsymbol{D}]$ and $r_g[\boldsymbol{D}, T]$ in finite GWAS sample size introduce an upward bias in estimates of ${h^2}_{pleio}/h^2$. To obtain an unbiased estimate, we first generate Monte-Carlo samples of estimation errors in $r_g[\boldsymbol{D}, \boldsymbol{D}]$ and $r_g[\boldsymbol{D}, T]$ from the sampling covariance matrix computed from genomic block-jackknife[16] (we use the same genomic block-jackknife for the target disease T and all auxiliary diseases $\boldsymbol{D}$; see Code Availability); these samples capture the joint distribution of estimation errors across elements in $r_g[\boldsymbol{D}, \boldsymbol{D}]$ and $r_g[\boldsymbol{D}, T]$. Then, we add the estimation errors to the point estimates of $r_g[\boldsymbol{D}, \boldsymbol{D}]$ and $r_g[\boldsymbol{D}, T]$ to create Monte-Carlo samples of ${h^2}_{pleio}/h^2$, which are larger than the initial estimate. We then use binary search to estimate a scaling coefficient ξ_c multiplying $r_g[\boldsymbol{D}, T]$ so that the average ${h^2}_{pleio}/h^2$ across Monte-Carlo samples matches the initial estimate of ${h^2}_{pleio}/h^2$ (see **Methods**). Finally, we multiply the initial estimate of ${h^2}_{pleio}/h^2$ (and its standard error) by ξ_c^2 to obtain an unbiased estimate of ${h^2}_{pleio}/h^2$.

In the presence of collinearity among auxiliary diseases and uncertainty in genetic correlation estimates, the estimate from Eq. 2 can be numerically unstable, which we define as s.e.$(r_g[\boldsymbol{D}, T]^T r_g[\boldsymbol{D}]^{-1} r_g[\boldsymbol{D}, T]) > 0.5$ and/or $\xi_c < 0.5$ (see **Methods**). Whenever this occurs, we prune auxiliary traits as follows: for each pair of auxiliary traits with $r_g^2 > 0.5$, we remove the trait with lower r_g z-score with target trait; we repeat this procedure for r_g^2 thresholds of 0.5, 0.4, 0.3, 0.2, and 0.1 until the estimate is numerically stable.

We extend our method to estimate the liability-scale total phenotypic variance of a target disease that is shared with a specific set of auxiliary diseases ($V^2{}_{pleio}$) by estimating the liability-scale phenotypic correlation between pairs of diseases using individual-level data[33]. (We note that the term pleiotropy generally refers to genetic effects, but we use the term pleiotropic phenotypic variance for consistency.) Further details are provided in the **Methods** section.

We first estimated $h^2{}_{pleio}$ for 15 highly heritable target diseases (prevalence > 1% and heritability z-score > 6) from the UK Biobank[34] (average N = 157K, spanning 7 PheCode disease categories[35]) (**Supplementary Table 1**); the sample size reflects a restriction to unrelated individuals of British ancestry with diagnostic records from both primary care data and hospital inpatient data, which we imposed to mitigate the impact of missing diagnoses. In our UK Biobank analyses, we used the same 15 diseases as auxiliary diseases; we also performed analyses incorporating 16 quantitative traits from UK Biobank as auxiliary traits. We subsequently incorporated 30 diseases with large meta-analyses available (average N = 483K) as target and/or auxiliary diseases. In all analyses, we restricted to auxiliary diseases with modest genetic correlation to the target disease ($r_g{}^2 < 0.5$). We have publicly released all summary association statistics analyzed in this study, along with estimates of pleiotropic heritability and related quantities (**Data Availability**).

Simulations

We performed simulations to evaluate the unbiasedness of pleiotropic heritability estimates. We specified genetic architectures for 16 diseases using liability-scale heritabilities of 0.25 and prevalences of 0.05. True genetic correlations were set to 0.5 within diseases categories and 0.1 between diseases categories for the first 15 diseases (based on the 7 disease categories from **Supplementary Table 1**) and 0.0 for the last disease (with each other disease), implying 5 different values of true $h^2{}_{pleio}/h^2$ for each target disease (ranging from 0.00 to 0.38). The proportion of causal SNPs was set to 100%. Additional simulation settings were also evaluated. We simulated phenotypes from real UK Biobank genotypes using 417,072 HapMap3[36] SNPs with MAF > 0.05 and r^2 < 0.8 across 157,206 unrelated individuals. We applied PLINK2[37] to compute summary association statistics and applied cross-trait LD score regression[32] to summary association statistics for the 417,072 SNPs and in-sample LD from the 157,206 individuals to estimate genetic correlations. We then estimated $h^2{}_{pleio}$ for each disease using the other 15 diseases as auxiliary diseases. Further details of the simulation framework are provided in the **Methods** section.

Results are reported in **Figure 2** and **Supplementary Table 4**. Before Monte-Carlo bias correction, $h^2{}_{pleio}/h^2$ estimates suffered large upward biases (between 10.97% (s.e. 0.86%) and 18.24% (s.e. 1.09%)), highlighting the need for a bias correction procedure. After Monte-Carlo bias correction, $h^2{}_{pleio}/h^2$ estimates were approximately unbiased across the 16 simulated diseases with varying pleiotropic architectures. Most biases were small (between 0.09% (s.e. 1.40%) and 0.72% (s.e. 1.02%)), but the disease with true $h^2{}_{pleio}/h^2$ equal to 0 had upward bias of 3.66% (s.e. 0.57%), indicating that estimates very close to 0 may be inflated; we note that estimates in empirical data are usually above this range (see below). We compared the estimated standard errors to empirical standard deviations, and determined that estimated standard errors were approximately well-calibrated (**Supplementary Figure 1** and **Supplementary Table 4**).

We performed four secondary analyses. First, we performed simulations with the proportion of causal SNPs set to 5% (instead of 100%), and determined the results were similar to **Figure 2** (**Supplementary Figures 2-3** and **Supplementary Table 12**). Second, we simulated pleiotropic architectures with r_g within disease categories equal to 0.4, 0.3, or 0.2 (instead of 0.5) (and r_g between disease categories still equal to 0.1), and observed modest downward bias for values between 5% and 25% while there was modest upward bias for values of $h^2{}_{pleio}/h^2$ below 5% (**Supplementary Figures 4-5** and **Supplementary Tables 5-7**). Though it should be noted that estimates on empirical data are usually above this range (see below). Third, we estimated $h^2{}_{pleio}/h^2$ without pruning highly correlated auxiliary diseases, and we found that $h^2{}_{pleio}/h^2$ estimates were downward biased when true values were less than 25% but approximately unbiased (similar to **Figure 2**) when true values were above 25% (**Supplementary Figures 6-7** and **Supplementary Table 8**). However, pruning highly correlated auxiliary diseases reduced the bias when true values were less than 25% (**Supplementary Figures 8-9** and **Supplementary Tables 9-11**). Fourth, we performed simulations to evaluate the genomic block-jackknife standard errors on the reduction in $h^2{}_{pleio}/h^2$ in analyses with one auxiliary disease category removed, and determined that standard errors were conservative (**Supplementary Figure 10** and **Supplementary Table 13**).

Application to 15 diseases from the UK Biobank

We estimated $h^2{}_{pleio}/h^2$ for 15 highly heritable target diseases from the UK Biobank (heritability z-score > 6; ${r_g}^2 < 0.5$; average N=157K, spanning 7 PheCode disease categories[35]) (**Supplementary Table 1**). Genetic correlations across these 15 diseases are reported in **Figure 3A** and **Supplementary Table 14**, and corresponding liability-scale phenotypic correlations are reported in **Figure 3B** and **Supplementary Table 15**. We observed moderate genetic correlations within disease categories and between disease categories, and smaller phenotypic correlations.

Estimates of $h^2{}_{pleio}/h^2$ for three representative diseases (type 2 diabetes (T2D), depression (MDD), and hypertension (HTN)), and the average across 15 diseases, are reported in **Figure 4**, **Supplementary Figure 12** and **Supplementary Table 17**. We determined that 34% of SNP-heritability is pleiotropic on average (jackknife s.e.(avg) = 7%, s.d. = 25% across diseases); the Monte-Carlo bias correction procedure had a substantial impact on estimates of $h^2{}_{pleio}/h^2$ (pre-correction average = 45%, jackknife s.e.(avg) = 10%, **Supplementary Figure 12**, **Supplementary Table 17**), consistent with simulations (**Figure 2**) and underscoring the importance of correcting upward bias. 4 of 15 diseases had $h^2{}_{pleio}/h^2$ estimates greater than 50%, including 71% (s.e. 14%) for MDD and 65% (s.e. 21%) for T2D.

We performed three analyses in which we modified the set of auxiliary diseases (**Figure 4**, **Supplementary Figure 12** and **Supplementary Table 17**). First, we included 16 quantitative traits from the UK Biobank (**Supplementary Table 2**) as additional auxiliary traits, causing estimates of $h^2{}_{pleio}/h^2$ to increase (avg = 44%, jackknife s.e.(avg) = 8%, s.d. = 35%). Second, we restricted to auxiliary diseases excluding the target disease category, causing estimates of $h^2{}_{pleio}/h^2$ to decrease only slightly (avg = 31% (s.e. 7%); 69% (s.e. 13%) for MDD (removing mental) and 64% (s.e. 23%) for T2D (removing endocrine)). Third, we further removed one other disease category whose removal had the greatest impact, causing estimates of $h^2{}_{pleio}/h^2$ to

decrease only moderately (avg = 20% (s.e. 3%); 59% (s.e. 11%) for MDD (removing mental and endocrine) and 45% (s.e.13%) for T2D (removing endocrine and symptoms)). These results imply that pleiotropic heritability is broadly distributed across disease categories.

Estimates of ${h^2}_{pleio}/h^2$ with respect to auxiliary diseases in each individual disease category are reported in **Figure 5**, **Supplementary Figure 13** and **Supplementary Table 19**. We determined that auxiliary diseases from a single disease category can often explain the majority of pleiotropic heritability. For example, MDD had 50% (s.e. 11%) and 49% (s.e. 11%) of SNP-heritability shared with endocrine and digestive disease categories, respectively (vs. 71% for all disease categories); T2D had 30% (s.e. 11%) and 28% (s.e. 7%) of SNP-heritability shared with mental disorder and circulatory disease categories, respectively (vs. 65% for all disease categories; these two categories were also implicated in previous studies[21,38]); and the 15 target diseases on average had 21% (s.e. 6%) and 17% (s.e. 5%) of SNP-heritability shared with mental disorder and endocrine categories, respectively (vs. 34% for all disease categories).

However, removing a single disease category from the set of auxiliary diseases generally reduced ${h^2}_{pleio}/h^2$ much less than the ${h^2}_{pleio}/h^2$ shared with that disease category (**Figure 5**, **Supplementary Figure 13** and **Supplementary Table 19**). For example, MDD ${h^2}_{pleio}/h^2$ was reduced by only 10% (s.e. 11%) and 5% (s.e. 7%) when removing endocrine and digestive disease categories, respectively; T2D ${h^2}_{pleio}/h^2$ was reduced by only 7% (s.e. 14%) and 9% (s.e. 8%) when removing mental disorder and circulatory disease categories, respectively; and average ${h^2}_{pleio}/h^2$ across the 15 target diseases was reduced by only 4% (s.e. 8%) and 6% (s.e. 10%) when removing mental disorder and endocrine categories, respectively. These results imply that pleiotropic heritability is broadly shared across disease categories, such that removal of one disease category has a limited impact due to shared pleiotropic heritability with other disease categories.

We performed two secondary analyses. First, we estimated genetic correlations by applying BOLT-REML[33] to individual-data, instead of applying cross-trait LDSC[32] to summary-level data. We determined that estimates of genetic correlation and ${h^2}_{pleio}/h^2$ were broadly consistent (**Supplementary Figures 14-15**, **Supplementary Tables 20-21**). Second, we estimated genetic correlations using three modified versions of cross-trait LDSC: constraining heritability intercept, constraining genetic covariance intercept, and constraining both intercepts. We determined that all three modified versions deviated from BOLT-REML results in estimates of genetic correlation and ${h^2}_{pleio}/h^2$ (**Supplementary Figures 14-15**, **Supplementary Tables 20-21**), supporting our primary use of the default version of cross-trait LDSC.

Application to 30 diseases from publicly available GWAS meta-analyses

We expanded our analyses to 30 highly heritable diseases with publicly available GWAS summary statistics (heritability z-score > 6; ${r_g}^2 < 0.5$; average N=483K, spanning ten PheCode disease categories[35]) (**Supplementary Table 3**). These include many diseases (e.g. autoimmune diseases and cancers) that were not included in the 15 diseases in UK Biobank based on the criteria that we applied. For brevity, we subsequently refer to the above 15 as 'UK Biobank' diseases and the 30 as 'non-UK Biobank' diseases (while duly noting that a subset of the latter includes both non-UK Biobank and UK Biobank data). Genetic correlations across these 30

diseases are reported in **Supplementary Figure 16** and **Supplementary Table 16**. We again observed that most diseases have moderate genetic correlations within disease categories and between disease categories; we note that it is not feasible to estimate corresponding liability-scale phenotypic correlations in the absence of individual-level data.

We first estimated $h^2{}_{pleio}/h^2$ for 15 UK Biobank target diseases with respect to different sets of auxiliary diseases. Results for three representative diseases (T2D, MDD, and HTN), and the average across 15 diseases, are reported in **Figure 6A**, **Supplementary Figure 17** and **Supplementary Table 22**. Average $h^2{}_{pleio}/h^2$ increased to 43% with respect to 30 non-UK Biobank auxiliary diseases (jackknife s.e.(avg) = 3%, s.d. = 30% across diseases), 47% with respect to all 45 UK Biobank + non-UK Biobank auxiliary diseases (jackknife s.e.(avg) = 5%, s.d. = 36% across diseases), and 50% (jackknife s.e. (avg) = 5%, s.d. = 34% across diseases) with respect to all 45 UK Biobank + non-UK Biobank auxiliary diseases and 16 UK Biobank quantitative traits, as compared to 34% with respect to 15 UK Biobank auxiliary diseases.

We next estimated $h^2{}_{pleio}/h^2$ for 30 non-UK Biobank target diseases with respect to different sets of auxiliary diseases. Results for three representative diseases (type 1 diabetes, attention-deficit/hyperactivity disorder, and asthma), and the average across 30 diseases, are reported in **Figure 6B**, **Supplementary Figure 18** and **Supplementary Table 23**. We observed a similar trend but with lower overall estimates: average $h^2{}_{pleio}/h^2$ increased to 31% with respect to 30 non-UK Biobank auxiliary diseases (jackknife s.e.(avg) = 1%, s.d. = 21% across diseases), 36% with respect to all 45 UK Biobank + non-UK Biobank auxiliary diseases (jackknife s.e.(avg) = 4%, s.d. = 25% across diseases), and 35% with respect to all 45 UK Biobank + non-UK Biobank auxiliary diseases and 16 UK Biobank auxiliary quantitative traits (jackknife s.e.(avg) = 1.5%, s.d. = 26% across diseases), as compared to 26% with respect to 15 UK Biobank auxiliary diseases.

Finally, we estimated $h^2{}_{pleio}/h^2$ of 15 UK Biobank target diseases or 30 non-UK Biobank target diseases with respect to auxiliary diseases in each of the 11 PheCode disease categories[35] spanned by all 45 UK Biobank + non-UK Biobank auxiliary diseases (analogous to **Figure 5**). Results are reported in **Supplementary Figure 19** and **Supplementary Table 24**. We again determined that auxiliary diseases from a single disease category often explain the majority of pleiotropic heritability (e.g. MDD had 74% (s.e. 12%) of SNP-heritability shared with other mental disorders), and that removing a single disease category from the set of auxiliary diseases reduced $h^2{}_{pleio}/h^2$ much less than the $h^2{}_{pleio}/h^2$ shared with that disease category (e.g. MDD $h^2{}_{pleio}/h^2$ was reduced by only 29% (s.e. 10%) when removing other mental disorders).

Comparing shared genetic and non-genetic variance

We sought to compare $h^2{}_{pleio}/h^2$ to the proportion of liability-scale total phenotypic variance of a target disease that is shared with a given set of auxiliary diseases (pleiotropic phenotypic variance, $V^2{}_{pleio}$; defined in **Methods**, analogous to Eq. 1, **Supplementary Figure 20**). We extended our method to estimate $V^2{}_{pleio}$ by replacing genetic correlations (r_g) in Eq. 2 with liability-scale phenotypic correlations (r_l), estimated from individual-level data as in ref. 33 (**Methods**). We determined via simulations that $V^2{}_{pleio}$ estimates are unbiased without the need for bias correction, as r_l is accurately estimated at large sample sizes (**Supplementary Figure 21**

and **Supplementary Table 25**). We used the ratio of pleiotropic phenotypic variance vs. total liability-scale phenotypic variance ($V^2{}_{pleio}/V^2$) to quantify the proportion of total liability-scale phenotypic variance that is pleiotropic (noting that V^2 = 1 in standard parametrizations). Because estimating r_l requires individual-level data, we analyzed 15 UK Biobank diseases only (**Supplementary Table 1**).

A comparison of estimates of $h^2{}_{pleio}/h^2$ vs. $V^2{}_{pleio}/V^2$ is reported in **Figure 7** and **Supplementary Table 26**; to ensure a fair comparison, we pruned the same auxiliary diseases in estimates of $V^2{}_{pleio}/V^2$ as in our prior estimates of $h^2{}_{pleio}/h^2$. We determined that estimates of $h^2{}_{pleio}/h^2$ were generally larger than estimates of $V^2{}_{pleio}/V^2$ (ratio of averages = 2.4x (s.e. 0.52)). $h^2{}_{pleio}/h^2$ was significantly larger than $V^2{}_{pleio}/V^2$ ($P < 0.05/15$) for 3 of 15 diseases, including MDD (0.71 (s.e. 0.14) vs. 0.12 (s.e. 0.006), $P = 1.9 \times 10^{-5}$), obesity (0.54 (s.e. 0.11) vs. 0.13 (s.e. 0.007), $P = 2.4 \times 10^{-4}$), and gastroesophageal reflux disease (GERD) (0.50 (s.e. 0.09) vs. 0.15 (s.e. 0.005), $P = 7.5 \times 10^{-5}$), and significantly smaller for only 1 disease, atrial fibrillation (AF) (0.03 (s.e. 0.02) vs. 0.19 (s.e. 0.005), $P < 1 \times 10^{-15}$). Due to relatively small values of SNP-heritability, estimates of the proportion of non-genetic variance that is shared with auxiliary diseases ($E^2{}_{pleio}/E^2$) were only slightly smaller than $V^2{}_{pleio}/V^2$ estimates on average (**Supplementary Figure 23** and **Supplementary Table 26**).

Discussion

We have defined and estimated $h^2{}_{pleio}$, the genetic variance of a target disease that is shared with a specific set of auxiliary diseases. We highlight four findings. First, bias correction had a large impact on estimates of $h^2{}_{pleio}$. Second, roughly half of the total SNP-heritability of common diseases is shared with other disease/traits. Third, the contribution of auxiliary diseases to $h^2{}_{pleio}$ is broadly distributed across disease categories, with $h^2{}_{pleio}$ decreasing only slightly when removing one auxiliary disease category. Finally, we compared $h^2{}_{pleio}/h^2$ to $V^2{}_{pleio}/V^2$ and determined that genetic variance is more strongly shared across diseases than environmental variance.

We emphasize three downstream implications. First, our results can inform the incorporation of genome-wide priors in polygenic risk scores that leverage pleiotropic diseases/traits[24,39–43]. Second, our work motivates the definition and estimation of *partitioned pleiotropic heritability*, in which genetic values of auxiliary diseases/traits are partitioned into tissue/cell-type-specific components[44–46], increasing shared genetic variance with the target disease and elucidating shared biological processes[1–6,14,21,22]. Third, genetic values of auxiliary traits could also be partitioned into gene pathways, elucidating shared biological pathways[47–49]. As such, our method for pleiotropic heritability may be developed in further ways to inform risk prediction, drug repurposing, prioritization of drug targets for multimorbid diseases, or improve the detection of unintended drug side-effects.

We note several limitations of our work. First, jackknife standard errors on estimates of $h^2{}_{pleio}$ do not account for stochasticity that may in principle be introduced by pruning genetically correlated auxiliary diseases to ensure numerical stability; however, we anticipate that the set of

pruned diseases would exhibit minimal variation across genomic jackknife blocks, due to limited variation in genetic correlation estimates across blocks. Second, PHBC suffers small upward biases when the true $h^2{}_{pleio}$ is very low (**Figure 2B**); however, we expect that $h^2{}_{pleio}$ estimates in empirical data are approximately unbiased, as most diseases analyzed have substantial $h^2{}_{pleio}$ estimates. Third, genetic correlation estimates include the effects of assortative mating, which can be substantial for some diseases/traits[50]; our results do not distinguish between genetic correlation due to shared causal effects vs. genetic correlation due to assortative mating, which would require applying new methods to individual-level data[51]. Finally, we restricted our analyses to individuals of European ancestry (represented in large sample size in UK Biobank), but it is important to analyze more diverse cohorts[52,53], particularly for diseases whose prevalence varies across populations. Despite these limitations, our definition and estimation of $h^2{}_{pleio}$ provide a robust quantification of the genetic variance of a target disease that is shared with pleiotropic diseases/traits.

Code Availability

We have released open-source software implementing PHBC at https://github.com/yjzhao1004/pleioh2g. Other publicly available software packages used in this study are listed below: GWAS summary statistics were computed using PLINK2 (https://www.cog-genomics.org/plink/2.0/) and BOLT-LMM v2.4.1 (https://alkesgroup.broadinstitute.org/BOLT-LMM/downloads/), Heritability and genetic correlation were estimated using LDSC (https://github.com/bulik/ldsc), the ldscr R package (https://github.com/mglev1n/ldscr), and BOLT-REML as implemented in BOLT-LMM v 2.4.1.

Data Availability

Summary association statistics for all diseases/traits analyzed in this study have been made publicly available at https://alkesgroup.broadinstitute.org/PHBC/.

Acknowledgements

Use of the UK Biobank data was performed under application number 19542. This research was funded by Wellcome Trust early-career award 227566/Z/23/Z (X.J.) and NIH grants R01 HG006399, U01 HG012009, R01 MH101244, and R37 MH107649 (A.L.P.). M.I. is supported by the Munz Chair of Cardiovascular Prediction and Prevention and the NIHR Cambridge Biomedical Research Centre (NIHR203312) as well as by the UK Economic and Social Research 878 Council (ES/T013192/1). J.D. is supported by a British Heart Foundation Personal Chair. This work was supported by core funding from the British Heart Foundation (RG/18/13/33946: RG/F/23/110103), NIHR Cambridge Biomedical Research Centre (NIHR203312), BHF Chair Award (CH/12/2/29428), Cambridge BHF Centre of Research Excellence (RE/24/130011, RE/18/1/34212) and by Health Data Research UK, which is funded by the UK Medical Research Council, Engineering and Physical Sciences Research Council, Economic and Social Research Council, Department of Health and Social Care (England), Chief

Scientist Office of the Scottish Government Health and Social Care Directorates, Health and Social Care Research and Development Division (Welsh Government), Public Health Agency (Northern Ireland), British Heart Foundation and the Wellcome Trust. The views expressed are those of the authors and not necessarily those of the NIHR or the Department of Health and Social Care. W.C. is supported by grants from the National Key R&D Program of China (No. 2023YFC3605400) and the National Natural Science Foundation of China (No. 82472055, No. 82071997).

Methods

Ethics statement

This study utilized publicly available datasets and therefore did not require ethical approval.

Definition of $h^2{}_{pleio}$

We assume that the total liability of the target disease consists of genetic value (G) and environmental value (E). G of the target disease is divided into two components: disease-specific genetic value ($G_{disease-specific}$) and pleiotropic genetic value (G_{pleio}), with the pleiotropic genetic value defined as a weighted linear combination of auxiliary disease genetic values (**Figure 1**):

$$G_T = G_{disease-specific} + \sum_{j=1}^{n} \alpha_j G_j \ , \qquad (3)$$

where α_j is the weight of the jth auxiliary disease (D_j), G_j is the genetic value of D_j, and G_T is the genetic value of target disease.

We define pleiotropic heritability as the maximum proportion of the total target disease heritability explained by the auxiliary diseases in the entire population. We use the ratio of pleiotropic heritability vs. total heritability ($\frac{h^2{}_{pleio}}{h^2}$) to quantify the proportion of genetic variance that is pleiotropic:

$$\frac{h^2_{pleio}}{h^2} = \max_{\alpha_j} \quad r^2\left[\sum_{j=1}^{n} \alpha_j G_j, \ \ G_T\right] \qquad (4)$$

We emphasize that $h^2{}_{pleio}$ is a function of both the target disease and the selected set of auxiliary diseases/traits.

Initial estimate of $h^2{}_{pleio}$

We employ pleiotropic heritability with bias correction (PHBC) to obtain an unbiased estimate of $\frac{h^2_{pleio}}{h^2}$ with finite sample sizes.

Under the random effect model, we use $\boldsymbol{G}$ to denote the vector of genetic values for target diseases and auxiliary diseases: $\boldsymbol{G} = (G_T, G_1, G_2, \ldots, G_n)$. $\boldsymbol{G}$ follows a multivariate normal distribution $N(0, \Sigma_G)$, with the mean equal to 0 and the covariance equal to genetic covariance matrix Σ_G.

According to the definition in equation (4), pleiotropic heritability is not changed when multiplying $\boldsymbol{\alpha}$ by a non-zero scaling factor. We note that the quantity in equation (4) is equivalent

to regression $\boldsymbol{R^2}$, therefore we change the optimization to:

$$\frac{h^2_{pleio}}{h^2} = \max_{\alpha_j} \quad R^2_{regression}\left[G_T \sim \sum_{j=1}^{n} \alpha_j G_j\ \right], \tag{5}$$

so that the α_j are identifiable. We derive the expression of h^2_{pleio} as follows:

Let $\boldsymbol{\alpha}^*$ be the estimated $\boldsymbol{\alpha}$ that maximizes the right-hand side of equation (5).

Let $\boldsymbol{C}_{DD}$ be the covariance matrix of $\{G_j\}_{j=1}^{n}$.

Let $\boldsymbol{c}_{TD}$ be the covariance vector between G_T and $\{G_j\}_{j=1}^{n}$.

The variance of pleiotropic genetic value $\sum_{j=1}^{n} \alpha_j G_j$ is:

$$Var\left[\sum_{j=1}^{n} \alpha_j G_j\right] = \sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i \alpha_j Cov(G_i, G_j) = \boldsymbol{\alpha}^T \boldsymbol{C}_{DD} \boldsymbol{\alpha} \tag{6}$$

To estimate $h^2{}_{pleio}$, we first estimate the $\boldsymbol{\alpha}^*$ by minimizing the *expectation* of the squared error between G_T and G_{pleio} of $E_G\left[\left(G_T - \sum_{j=1}^{n} \alpha_j G_j\right)^2\right]$ (which is equivalent to maximizing regression $\boldsymbol{R^2}$). We use expectation so that the parameter estimated corresponds to regression $\boldsymbol{R^2}$ in the entire population. By taking the partial derivative of $E_G\left[\left(G_T - \sum_{j=1}^{n} \alpha_j G_j\right)^2\right]$ with respect to $\boldsymbol{\alpha}$:

$$\frac{\partial}{\partial \boldsymbol{\alpha}} E_G\left[\left(G_T - \sum_{j=1}^{n} \alpha_j G_j\right)^2\right]$$

$$= \frac{\partial}{\partial \boldsymbol{\alpha}} \left(E_G[G_T^2] - 2\sum_{j=1}^{n} \alpha_j E_G[G_T G_j] + E_G[(\sum_{j=1}^{n} \alpha_j G_j)^2]\right.$$

$$= \frac{\partial}{\partial \boldsymbol{\alpha}} \left(E_G[G_T^2] - 2\boldsymbol{\alpha}^T \boldsymbol{c}_{TD} + \boldsymbol{\alpha}^T \boldsymbol{C}_{DD} \boldsymbol{\alpha}\right)$$

$$= -2\boldsymbol{c}_{TD} + (\boldsymbol{C}_{DD} + \boldsymbol{C}_{DD}^T)\boldsymbol{\alpha} \tag{7}$$

Where we used $E(XY) = Cov(X,Y) + E(X)E(Y)$ and the expectation of each genetic value is 0. By setting the partial derivative to zero, we can obtain:

$$\boldsymbol{C}_{DD} \boldsymbol{\alpha}^* = \boldsymbol{c}_{TD} \tag{8}$$

If $\boldsymbol{C}_{DD}$ is invertible, the optimal $\boldsymbol{\alpha}^*$ is:

$$\boldsymbol{\alpha}^* = \boldsymbol{C}_{DD}^{-1} \boldsymbol{c}_{TD} \tag{9}$$

After substituting the $\boldsymbol{\alpha}^*$ into equation (6):

$$h^2{}_{pleio} = \boldsymbol{c}_{TD}^T \boldsymbol{C}_{DD}^{-1} \boldsymbol{c}_{TD} \tag{10}$$

Elements in $\boldsymbol{c}_{TD}$ and $\boldsymbol{C}_{DD}$ denote disease heritability and genetic correlation; in this study we focus on SNP-heritability and genetic correlation between genetic values from SNPs. Let h_T^2 be the heritability of the target disease. $\boldsymbol{h_D^2} = \{h_j^2\}_{j=1}^{n}$ is the vector consisting of the heritability of all auxiliary diseases. $r_g[\boldsymbol{D}, T]$ is the genetic correlation vector between the target disease and all auxiliary diseases, $r_g[\boldsymbol{D}, \boldsymbol{D}]^{-1}$ is the inverse of genetic correlation matrix of all auxiliary diseases. diag(.) is diagonal matrix with all non-diagonal elements equal to zero.

$$\boldsymbol{c}_{TD} = diag(\sqrt{h_T^2}\left(\sqrt{\boldsymbol{h_D^2}}\right)^T)r_g[\boldsymbol{D}, T] \tag{11}$$

$$\boldsymbol{C_{DD}^{-1}} = diag\left(1/\sqrt{\boldsymbol{h_D^2}}\right) r_g[\boldsymbol{D}, \boldsymbol{D}]^{-1} \tag{12}$$

After substituting equation (11) and (12) into equation (10), ${h^2}_{pleio}$ can be expressed as:

$${h^2}_{pleio} = h_T^2 \times r_g[\boldsymbol{D}, T]^T r_g[\boldsymbol{D}, \boldsymbol{D}]^{-1} r_g[\boldsymbol{D}, T] \tag{13}$$

$$\frac{{h^2}_{pleio}}{h^2} = r_g[\boldsymbol{D}, T]^T r_g[\boldsymbol{D}, \boldsymbol{D}]^{-1} r_g[\boldsymbol{D}, T] \tag{14}$$

Equation (14) provides the initial pleiotropic heritability estimates, prior to bias correction described in the next section. In this study, we used GWAS summary statistics as input and reimplemented cross-trait LDSC[32] so that the standard errors of all quantities on the right-hand side of equation (14) are estimated through the same genomic jackknife blocks. We used 1000 Genomes Project[54] Europeans as a reference LD panel to estimate genetic correlation. Standard error estimates were obtained via genomic block-jackknife with 200 blocks applied across all traits, which were created by partitioning 1,217,311 HapMap 3 SNPs[36]. We primarily focus on ${h^2}_{pleio}/h^2$ estimates (instead of ${h^2}_{pleio}$ estimates) throughout the manuscript.

Monte Carlo bias correction to initial estimate of ${h^2}_{pleio}$

The initial estimates $\left(\frac{h_{pleio}^2}{h^2}\right)_{initial}$ are based on estimates of genetic correlation in finite GWAS sample size, which introduces an upward bias in estimating $\frac{{h^2}_{pleio}}{h^2}$ due to inverting the estimated genetic correlation matrix and sampling correlation between $r_g[\boldsymbol{D}, T]^T$, $r_g[\boldsymbol{D}, \boldsymbol{D}]^{-1}$, and $r_g[\boldsymbol{D}, T]$ in equation (14):

$$\left(\frac{h_{pleio}^2}{h^2}\right)_{initial} > \left(\frac{h_{pleio}^2}{h^2}\right)_{true} \tag{15}$$

To obtain an unbiased estimate, we employ a Monte-Carlo bias correction procedure to estimate the proportion of bias in the point estimate of $\frac{{h^2}_{pleio}}{h^2}$ and then correct the bias. First, we generate Monte-Carlo samples of the estimation error ϵ_i in the genetic correlation matrix from the

sampling covariance matrix[16] generated from genomic block-jackknife; if the dimension of the genetic correlation matrix is $n \times n$, the sampling covariance matrix is a symmetric matrix with dimension $n(n\text{-}1)/2 \times n(n\text{-}1)/2$, which captures the joint distribution across estimation errors of different elements in $r_g[\boldsymbol{D}, T]^T$ and $r_g[\boldsymbol{D}, \boldsymbol{D}]$. Second, we add the estimation noise to the point estimate of genetic correlation matrix to create Monte-Carlo samples of estimated $r_g[\boldsymbol{D}, T]^T$ and $r_g[\boldsymbol{D}, \boldsymbol{D}]^{-1}$. We filter these Monte-Carlo samples to only keep those whose $r_g[\boldsymbol{D}, \boldsymbol{D}]$ are non-singular (i.e. the smallest eigenvalue of $r_g[\boldsymbol{D}, \boldsymbol{D}]$ is larger than 0). We repeat this Monte-Carlo procedure until we have generated 1,000 samples of genetic correlation estimates.

Using these Monte-Carlo samples, we next estimate a scaling coefficient ξ_c so that:

$$Avg_i[(\xi_c r_g[\boldsymbol{D}, T]^T + \epsilon_i[\boldsymbol{D}, T])^T (r_g[\boldsymbol{D}, \boldsymbol{D}] + \epsilon_i[\boldsymbol{D}, \boldsymbol{D}])^{-1} (\xi_c r_g[\boldsymbol{D}, T]^T + \epsilon_i[\boldsymbol{D}, T])] = \left(\frac{h^2_{pleio}}{h^2}\right)_{initial} \quad (16)$$

The left-hand side is the average across Monte-Carlo samples, ϵ_i are the respective Monte-Carlo samples of estimation errors for $r_g[\boldsymbol{D}, T]$ and $r_g[\boldsymbol{D}, \boldsymbol{D}]$, where i indexes Monte-Carlo samples. The above equation is based on the heuristic that the magnitude of pleiotropic heritability is proportional to the average genetic correlation between target disease and auxiliary diseases $r_g[\boldsymbol{D}, T]$. Downward scaling $r_g[\boldsymbol{D}, T]$ will downward scale $\frac{h^2{}_{pleio}}{h^2}$, which controls the upward bias in equation (15). On the other hand, $r_g[\boldsymbol{D}, \boldsymbol{D}]$ are between pairs of auxiliary diseases which is not directly connected to the scale of $\frac{h^2{}_{pleio}}{h^2}$. The error term ϵ_i is not multiplied by ξ_c as estimation error for genetic correlation is a function of sample size and the heritability, which is not proportional to the scale of the point estimate of genetic correlation.

We estimate scaling coefficient ξ_c using binary search on the interval [0,1], in order to satisfy equation (16). We multiply the initial estimate $\left(\frac{h^2_{pleio}}{h^2}\right)_{initial}$ and its standard error by ξ_c^2 to obtain the bias-corrected estimate of $h^2{}_{pleio}/h^2$ and its standard error.

Pruning auxiliary diseases with high genetic correlation

In the presence of collinearity among auxiliary diseases and large uncertainty in genetic correlation estimates, the estimator in equation (14) can be numerically unstable, which we define as (1) The pre-correction jackknife standard error of $\frac{h^2_{pleio}}{h^2}$ with respect to the pre-pruning auxiliary disease sets is greater than 0.5, and/or (2) the scaling coefficient ξ_c is less than 0.5 (implying that >75% of the initial $\frac{h^2_{pleio}}{h^2}$ estimate is due to uncertainty in genetic correlation matrix estimates). We also apply the pruning procedure if (3) it takes more than 50 iterations to sample a non-singular $r_g[\boldsymbol{D}, \boldsymbol{D}]$ in the Monte-Carlo sampling of bias correction, which indicates that the collinearity of auxiliary diseases causes singularity in samples of $r_g[\boldsymbol{D}, \boldsymbol{D}]$.

To apply the pruning procedure, we prune the auxiliary diseases by progressively applying stricter r_g^2 thresholds between the auxiliary diseases: 0.5, 0.4, 0.3, 0.2, and 0.1. For each r_g^2 threshold, when a pair of auxiliary diseases has r_g^2 above the threshold, we keep the disease with larger r_g z-score with the target disease. We repeat this pruning step until all three criteria no longer hold, and we report the $\frac{h^2{}_{pleio}}{h^2}$ estimate from the pruned set of auxiliary diseases.

Pleiotropic phenotypic variance

To compare $h^2{}_{pleio}/h^2$ with the non-genetic variance that can be explained by the same set of auxiliary diseases, we extend our method to estimate pleiotropic phenotypic variance ($V^2{}_{pleio}$), defined as the proportion of phenotypic variance of target disease liability (σ_{T}) that can be explained by any linear combination of liabilities of auxiliary disease (σ_D) (**Supplementary Figure 20**).

We convert the observed-scale phenotypic correlation to liability-scale phenotypic correlation r_l using Monte-Carlo simulations[33] under the liability threshold model[55]. To model the relationship between liability-scale correlation r_l and observed-scale correlations r_o for two binary traits with different prevalences, we simulate liabilities using r_l on a grid between -1 to 1. For each r_l, we simulate pairs of disease liabilities from a multivariate normal distribution with covariance equal to specified r_l, and then use liability threshold model to create the binary disease status by thresholding the liabilities based on their respective prevalences. We then calculate the r_o between the simulated binary traits. This process generates a map between r_l and r_o for two disease traits of given prevalences. Using this mapping, we can map the observed r_o for the 15 UK Biobank diseases to obtain the corresponding r_l. We use the r_l matrix to estimate $V^2{}_{pleio}$ in the same way as pleiotropic heritability:

$$V^2{}_{pleio} = 1 \times r_l[\boldsymbol{\sigma_D}, \sigma_T]^T r_l[\boldsymbol{\sigma_D}, \boldsymbol{\sigma_D}]^{-1} r_l[\boldsymbol{\sigma_D}, \sigma_T] \quad (17)$$

where 1 is the total liability variance for the target disease, $r_l[\boldsymbol{\sigma_D}, \sigma_T]$ is the liability-scale phenotypic correlation matrix between target disease and auxiliary diseases, $r_l[\boldsymbol{\sigma_D}, \boldsymbol{\sigma_D}]^{-1}$ is the inverse of liability-scale phenotypic correlation among auxiliary diseases. We estimate the standard error of r_l and $V^2{}_{pleio}$ by jackknifing over 1000 blocks of individuals across all diseases. We also compare the difference between r_g and r_l with Bonferroni-corrected significant threshold set at $P < 4.76 \times 10^{-4}$ (0.05/105, where 105 is the number of pairs of the 15 diseases) (**Figure 3** and **Supplementary Table 15**).

We compare three quantities: (1) $\frac{h^2{}_{pleio}}{h^2}$, the proportion of genetic variance that is pleiotropic; (2) $\frac{V^2{}_{pleio}}{V^2}$, the proportion of liability variance that is shared with the set of auxiliary diseases, and (3) $\frac{E^2{}_{pleio}}{E^2}$, the proportion of environmental variance that is shared with the set of auxiliary diseases, calculated as:

$$\frac{E^2{}_{pleio}}{E^2} = \frac{V^2{}_{pleio} - h^2{}_{pleio}}{V^2 - h^2} \quad (18)$$

We computed the ratios between $\frac{h^2{}_{pleio}}{h^2}$ vs. $\frac{V^2{}_{pleio}}{V^2}$, $\frac{E^2{}_{pleio}}{E^2}$ vs. $\frac{V^2{}_{pleio}}{V^2}$, $h^2{}_{pleio}$ vs. $V^2{}_{pleio}$, and $\frac{h^2{}_{pleio}}{h^2}$ vs. $\frac{E^2{}_{pleio}}{E^2}$. Full results are listed in **Supplementary Figures 22-25** and **Supplementary Table 26**.

Simulations

We performed simulation using real UK Biobank genotypes from 417,072 independent HapMap3[36] SNPs with MAF > 0.05 and r^2 < 0.8 across 157,206 unrelated individuals (see below). We simulated 16 diseases with liability-scale heritabilities equal to 0.25 and prevalences equal to 0.05. True genetic correlations were set to 0.5 within diseases categories and 0.1 between diseases categories for the first 15 diseases (based on the 7 PheCode disease categories from **Supplementary Table 1**) and 0.0 for the 16th disease (with all other 15 diseases), which results in 5 different values of true $\frac{h^2{}_{pleio}}{h^2}$. The proportion of causal SNPs was set to 100% in the primary simulations (and 5% in a secondary simulation). For each simulation, we first simulated the causal effect sizes using infinitesimal model based on the specified heritability and genetic correlation. Then, we used PLINK2[37] to compute genetic values (which sum over causal_effect_size × causal_allele_dosage) for 157,206 unrelated individuals and added environmental noise to generate the individual liabilities for the 16 diseases. We used the liability threshold model to generate binary phenotype based on the specified prevalence and computed GWAS summary statistics using PLINK2[37]. We applied cross-trait LDSC[32] to the GWAS summary statistics and used in-sample LD scores to estimate genetic correlations. The in-sample LD was generated from the 470,851 HapMap3 SNPs across 157,206 samples. We applied PHBC to estimate $\frac{h^2{}_{pleio}}{h^2}$ and compared it with the true value. We also compared the squared genomic-jackknife standard error of $\frac{h^2{}_{pleio}}{h^2}$ with the squared deviation (defined as the squared difference between the estimated post-corrected $\frac{h^2{}_{pleio}}{h^2}$ and true value) to investigate whether the reported standard errors were well-calibrated.

We also performed a simulation to validate $V^2{}_{pleio}$ estimation. True r_o was set to 0.5 within diseases categories and 0.1 between diseases categories for the 15 diseases (based on the 7 PheCode disease categories). We simulated liabilities for 230,081 individuals (see below). We computed true r_l and true $V^2{}_{pleio}$ based on the simulated liabilities. Then, we used the liability threshold model to generate binary phenotypes based on the empirical prevalences of the 15 UK Biobank diseases. We applied our method to the simulated binary phenotypes to estimate $V^2{}_{pleio}$ and compared it to the true value.

UK Biobank data

For disease phenotype pre-processing, we collected diagnoses from both inpatient data and primary care data. An individual is defined as a case if there is a diagnosis from either inpatient data or primary care data. We selected 230,081 samples with both primary care and hospital inpatient records. Diagnoses from hospital inpatient records were obtained from Hospital Episode Statistics (HES) for England, which are recorded as the International Classification of Diseases (ICD-10) system codes. Diagnoses from primary care data were retrieved from a subset of UK Biobank samples (see below), which are recorded in Read Codes v.2 (Read v.2) and Read Codes Clinical Terms v.3 (Read CTv.3). We mapped primary care data from Read v.2 to ICD-10 codes, and combine with the ICD-10codes from HES data. We kept ICD-10 codes starting with the letters A to N which are disease codes. Then, we mapped the ICD-10 records to PheCode system to obtain the phenotype definition[35,56] and selected the PheCode phenotypes with > 1% prevalence in the 230,081 samples.

We selected 15 relatively independent heritable diseases with heritability *z*-scores larger than 6 and ${r_g}^2 < 0.5$ (both heritability and r_g were computed using cross-trait LDSC[32]), which are distributed across 7 PheCode disease categories. We note that we did not include any UK Biobank autoimmune diseases or cancers, because they did not meet our criteria of prevalence > 1% and heritability z-score > 6. We used 230,081 samples to compute pleiotropic phenotypic variance across these 15 UK Biobank disease phenotypes. Detailed information on phenotype codes, prevalences, and categories are listed in **Supplementary Table 1**.

We selected 157,206 unrelated (defined as having less than 11 putative third-degree relatives in the kinship table) British ancestry individuals from the 230,081 samples to compute GWAS summary statistics for 15 UK Biobank diseases, where we used 470,851 independent HapMap 3 SNPs (see below)[36].We used BOLT-LMM[57,58] to compute summary statistics. Since BOLT-LMM test statistics are well-calibrated for traits with a case fraction of at least 10%[58], for diseases with prevalence lower than 10% we matched each case with a subset of nine randomly sampled controls, to avoid miscalibration in unbalanced case-control samples.

Detailed genotyping and quality control procedures in the UK Biobank are described previously[34]. We excluded SNPs with call rates of < 95%, minor allele frequency of < 0.1%, deviation from the Hardy–Weinberg equilibrium with $P < 1 \times 10^{-10}$. We then generated the PLINK-format data for BOLT-LMM using two steps: 1) subset to the 157,206 unrelated British individuals with genotype data; 2) subset the SNPs to HapMap3 SNPs and using PLINK2 to LD prune at r^2<0.8 to obtain the set of relatively independent SNPs (PLINK2 command "-indep-pairwise 50 5 0.8"). 470,851 SNPs from 157,206 samples remained after the pruning step, and were carried forward to GWAS.

We selected 16 quantitative traits from ref. 59 for the same set of individuals (Average N = 230K) and corrected for medication use, as follows. We corrected blood biochemistry measurements for cholesterol, hypertensive, and diabetic medications: total cholesterol, LDL, and triglyceride levels were corrected for lipid lowering medications[60]; systolic blood pressure and diastolic blood pressure were corrected for hypertension medications[61]; and HbA1c level was corrected for non-insulin diabetic medications and insulin drugs[62]. We verified that the correlation between the corrected quantitative traits and corresponding binary traits was higher compared to uncorrected quantitative traits. Detailed information on trait names, categories, heritability

estimates and genetic correlation estimates for these 16 quantitative traits is provided in **Supplementary Tables 2** and **16**.

GWAS summary statistics for 30 diseases from publicly available GWAS meta-analyses

We collected publicly available GWAS summary statistics for 30 relatively independent heritable diseases under the same criteria (heritability z-score > 6 and ${r_g}^2 < 0.5$) that are primarily from European ancestry. We note that some of these 30 publicly available disease summary statistics are meta-analyses that include UK Biobank samples. The 30 diseases are assigned to 10 PheCode categories[35]. Details of phenotype information, heritability, and genetic correlation estimates from cross-trait LDSC for these 30 diseases are reported in **Supplementary Tables 3** and **16**.

Analyses with one or more disease categories excluded from the set of auxiliary diseases

We performed leave-category-out analyses to investigate the change in $\frac{{h^2}_{pleio}}{h^2}$ when removing PheCode Categories. First, we estimated the $\frac{{h^2}_{pleio}}{h^2}$ with respect to three auxiliary disease sets: 1) all auxiliary diseases; 2) all diseases excluding the auxiliary diseases in the target disease category; 3) all diseases excluding the auxiliary diseases in target disease category and one other category. We implemented leave-category-out analyses in the PHBC software.

To evaluate the contribution of a single PheCode category to $\frac{{h^2}_{pleio}}{h^2}$, we computed the reduction of $\frac{{h^2}_{pleio}}{h^2}$when removing the auxiliary PheCode category. To remove the impact of randomness in Monte-Carlo bias correction when computing the reduction, we first computed the difference of $\frac{{h^2}_{pleio}}{h^2}$ with respect to all auxiliary diseases vs. $\frac{{h^2}_{pleio}}{h^2}$ with respect to auxiliary diseases excluding a PheCode category before bias correction. We multiplied the difference by the scaling coefficient ξ_c^2 of $\frac{{h^2}_{pleio}}{h^2}$ with respect to all auxiliary diseases, which represented the contribution of an auxiliary PheCode category to the $\frac{{h^2}_{pleio}}{h^2}$ with respect to all auxiliary diseases. We reported the jackknife standard error of the before-bias-correction difference of $\frac{{h^2}_{pleio}}{h^2}$ as the standard error of the reduction, which is a conservative standard error shown by simulations (**Supplementary Figure 10** and **Supplementary Table 13**).

Alternative approaches for estimating r_g

To validate the r_g estimated from cross-trait LDSC without constrained intercept, we used BOLT-REML and cross-trait LDSC using constrained intercept to estimate r_g. We applied BOLT-REML to individual-data using 470,851 SNPs across 157,206 unrelated individuals of British ancestry

to estimate r_g among these 15 UK Biobank diseases. We constrained heritability intercept in cross-trait LDSC to be 1. We also computed the analytical intercept for genetic covariance considering sample overlap for each disease pair using equation 16 in the Supplementary Note of ref. 32. The constrained intercepts for genetic covariance for each disease pair are provided in **Supplementary Table 27**.

References

1. Cotsapas, C. *et al*. Pervasive Sharing of Genetic Effects in Autoimmune Disease. *PLoS Genet* **7**, e1002254 (2011).
2. Solovieff, N., Cotsapas, C., Lee, P. H., Purcell, S. M. & Smoller, J. W. Pleiotropy in complex traits: challenges and strategies. *Nat Rev Genet* **14**, 483–495 (2013).
3. Watanabe, K. *et al.* A global overview of pleiotropy and genetic architecture in complex traits. *Nat Genet* **51**, 1339–1348 (2019).
4. Lee, P. H. *et al.* Genomic Relationships, Novel Loci, and Pleiotropic Mechanisms across Eight Psychiatric Disorders. *Cell* **179**, 1469-1482.e11 (2019).
5. Zhao, B. *et al.* Heart-brain connections: Phenotypic and genetic insights from magnetic resonance images. *Science* **380**, abn6598 (2023).
6. Mackay, T. F. C. & Anholt, R. R. H. Pleiotropy, epistasis and the genetic architecture of quantitative traits. *Nat Rev Genet* **25**, 639–657 (2024).
7. Cross-Disorder Group of the Psychiatric Genomics Consortium. Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nat Genet* **45**, 984–994 (2013).
8. ReproGen Consortium *et al.* An atlas of genetic correlations across human diseases and traits. *Nat Genet* **47**, 1236–1241 (2015).
9. Pickrell, J. K. *et al*. Detection and interpretation of shared genetic influences on 42 human traits. *Nat Genet* **48**, 709–717 (2016).
10. Anttila, V. *et al.* Analysis of shared heritability in common disorders of the brain. *Science* **360**, (2018).
11. O'Connor, L. J. & Price, A. L. Distinguishing genetic correlation from causation across 52 diseases and complex traits. *Nat Genet* **50**, 1728–1734 (2018).
12. Morrison, J., Knoblauch, N., Marcus, J. H., Stephens, M. & He, X. Mendelian randomization accounting for correlated and uncorrelated pleiotropic effects using genome-wide summary statistics. *Nat Genet* **52**, 740–747 (2020).
13. Wu, Y. *et al*. Fast estimation of genetic correlation for biobank-scale data. *The American Journal of Human Genetics* **109**, 24–32 (2022).
14. Udler, M. S. *et al*. Type 2 diabetes genetic loci informed by multi-trait associations point to disease mechanisms and subtypes: A soft clustering analysis. *PLoS Med* **15**, e1002654 (2018).
15. Tanigawa, Y. *et al.* Components of genetic associations across 2,138 phenotypes in the UK Biobank highlight adipocyte biology. *Nat Commun* **10**, (2019).
16. Grotzinger, A. D. *et al.* Genomic structural equation modelling provides insights into the multivariate genetic architecture of complex traits. *Nat Hum Behav* **3**, 513–525 (2019).
17. Sakaue, S. *et al.* A cross-population atlas of genetic associations for 220 human phenotypes. *Nat Genet* **53**, 1415–1424 (2021).
18. Grotzinger, A. D. *et al.* Genetic architecture of 11 major psychiatric disorders at biobehavioral, functional genomic and molecular genetic levels of analysis. *Nat Genet* **54**, 548–559 (2022).
19. Ballard, J. L. & O'Connor, L. J. Shared components of heritability across genetically correlated traits. *The American Journal of Human Genetics* **109**, 989–1006 (2022).
20. Zhang, Z. *et al.* A scalable approach to characterize pleiotropy across thousands of human diseases and complex traits using GWAS summary statistics. *The American Journal of Human Genetics* **110**, 1863–1874 (2023).

21. Suzuki, K. *et al.* Genetic drivers of heterogeneity in type 2 diabetes pathophysiology. *Nature* **627**, 347–357 (2024).
22. Smith, K. *et al.* Multi-ancestry polygenic mechanisms of type 2 diabetes. *Nat Med* **30**, 1065–1074 (2024).
23. Nieuwboer, H. A., Pool, R., Dolan, C. V., Boomsma, D. I. & Nivard, M. G. GWIS: Genome-Wide Inferred Statistics for Functions of Multiple Phenotypes. *The American Journal of Human Genetics* **99**, 917–927 (2016).
24. Turley, P. *et al.* Multi-trait analysis of genome-wide association summary statistics using MTAG. *Nat Genet* **50**, 229–237 (2018).
25. Baselmans, B. M. L. *et al.* Multivariate genome-wide analyses of the well-being spectrum. *Nat Genet* **51**, 445–451 (2019).
26. Zhou, F., Astle, W. J., Butterworth, A. S. & Asimit, J. L. Improved genetic discovery and fine-mapping resolution through multivariate latent factor analysis of high-dimensional traits. *Cell Genomics* **5**, 100847 (2025).
27. Trajanoska, K. *et al.* From target discovery to clinical drug development with human genetics. *Nature* **620**, 737–745 (2023).
28. Pirmohamed, M. Pharmacogenomics: current status and future perspectives. *Nat Rev Genet* **24**, 350–362 (2023).
29. Carss, K. J. *et al.* Using human genetics to improve safety assessment of therapeutics. *Nat Rev Drug Discov* **22**, 145–162 (2023).
30. Razuvayevskaya, O., Lopez, I., Dunham, I. & Ochoa, D. Genetic factors associated with reasons for clinical trial stoppage. *Nat Genet* **56**, 1862–1867 (2024).
31. Minikel, E. V., Painter, J. L., Dong, C. C. & Nelson, M. R. Refining the impact of genetic evidence on clinical success. *Nature* **629**, 624–629 (2024).
32. Bulik-Sullivan, B. *et al.* An atlas of genetic correlations across human diseases and traits. *Nat Genet* **47**, 1236–1241 (2015).
33. 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).
34. Bycroft, C. *et al.* The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).
35. Bastarache, L. Using Phecodes for Research with the Electronic Health Record: From PheWAS to PheRS. *Annu. Rev. Biomed. Data Sci.* **4**, 1–19 (2021).
36. The International HapMap 3 Consortium. Integrating common and rare genetic variation in diverse human populations. *Nature* **467**, 52–58 (2010).
37. Chang, C. C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, (2015).
38. Eijsbouts, C. *et al.* Genome-wide analysis of 53,400 people with irritable bowel syndrome highlights shared genetic pathways with mood and anxiety disorders. *Nat Genet* **53**, 1543–1552 (2021).
39. Maier, R. *et al.* Joint Analysis of Psychiatric Disorders Increases Accuracy of Risk Prediction for Schizophrenia, Bipolar Disorder, and Major Depressive Disorder. *The American Journal of Human Genetics* **96**, 283–294 (2015).
40. Maier, R. M. *et al.* Improving genetic prediction by leveraging genetic correlations among human diseases and traits. *Nat Commun* **9**, (2018).
41. Chung, W. *et al.* Efficient cross-trait penalized regression increases prediction accuracy in large cohorts using secondary phenotypes. *Nat Commun* **10**, (2019).

42. Patel, A. P. *et al.* A multi-ancestry polygenic risk score improves risk prediction for coronary artery disease. *Nat Med* **29**, 1793–1803 (2023).
43. Truong, B. *et al.* Integrative polygenic risk score improves the prediction accuracy of complex traits and diseases. *Cell Genomics* **4**, 100523 (2024).
44. Finucane, H. K. *et al.* Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat Genet* **47**, 1228–1235 (2015).
45. Finucane, H. K. *et al.* Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. *Nat Genet* **50**, 621–629 (2018).
46. Jagadeesh, K. A. *et al.* Identifying disease-critical cell types and cellular processes by integrating single-cell RNA-sequencing and human genetics. *Nat Genet* **54**, 1479–1492 (2022).
47. Zhu, X. & Stephens, M. Large-scale genome-wide enrichment analyses identify new trait-associated genes and pathways across 31 human phenotypes. *Nat Commun* **9**, (2018).
48. Kim, S. S. *et al.* Genes with High Network Connectivity Are Enriched for Disease Heritability. *The American Journal of Human Genetics* **104**, 896–913 (2019).
49. Siewert-Rocks, K. M., Kim, S. S., Yao, D. W., Shi, H. & Price, A. L. Leveraging gene co-regulation to identify gene sets enriched for disease heritability. *The American Journal of Human Genetics* **109**, 393–404 (2022).
50. Border, R. *et al.* Cross-trait assortative mating is widespread and inflates genetic correlation estimates. *Science* **378**, 754–761 (2022).
51. Zhang, Y. *et al.* The contribution of gametic phase disequilibrium to the heritability of complex traits. *Nat Genet* (2025) doi:10.1038/s41588-025-02192-4.
52. Zhou, W. *et al.* Global Biobank Meta-analysis Initiative: Powering genetic discovery across human disease. *Cell Genomics* **2**, 100192 (2022).
53. The All of Us Research Program Genomics Investigators *et al.* Genomic data in the All of Us Research Program. *Nature* **627**, 340–346 (2024).
54. The 1000 Genomes Project Consortium *et al.* A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
55. Lee, S. H., Wray, N. R., Goddard, M. E. & Visscher, P. M. Estimating Missing Heritability for Disease from Genome-wide Association Studies. *The American Journal of Human Genetics* **88**, 294–305 (2011).
56. Jiang, X. *et al.* Age-dependent topic modeling of comorbidities in UK Biobank identifies disease subtypes with differential genetic risk. *Nat Genet* **55**, 1854–1865 (2023).
57. Loh, P.-R. *et al.* Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nat Genet* **47**, 284–290 (2015).
58. Loh, P.-R., Kichaev, G., Gazal, S., Schoech, A. P. & Price, A. L. Mixed-model association for biobank-scale datasets. *Nat Genet* **50**, 906–908 (2018).
59. Weissbrod, O. *et al.* Leveraging fine-mapping and multipopulation training data to improve cross-population polygenic risk scores. *Nat Genet* **54**, 450–458 (2022).
60. Kofink, D. *et al.* Statin Effects on Metabolic Profiles: Data From the PREVEND IT (Prevention of Renal and Vascular End-stage Disease Intervention Trial). *Circ Cardiovasc Genet* **10**, (2017).
61. Wu, J. *et al.* A Summary of the Effects of Antihypertensive Medications on Measured Blood Pressure. *American Journal of Hypertension* **18**, 935–942 (2005).
62. Esposito, K., Chiodini, P., Colao, A., Lenzi, A. & Giugliano, D. Metabolic Syndrome and Risk of Cancer. *Diabetes Care* **35**, 2402–2411 (2012).

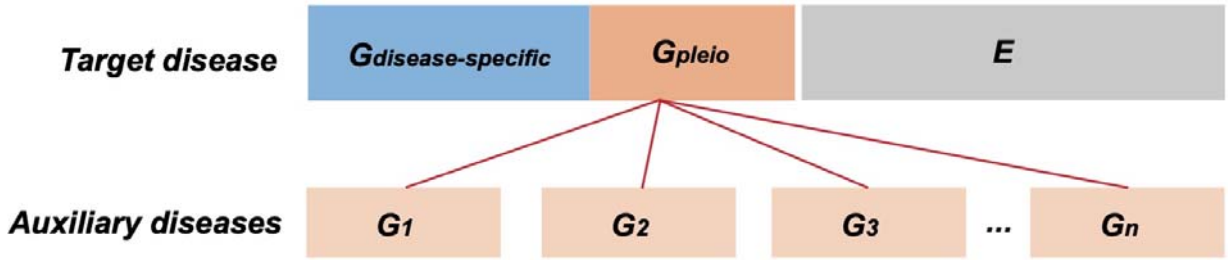


Figure 1. Definition of pleiotropic heritability. The total phenotypic variance of the target disease consists of genetic variance (G) and environmental variance (E). The genetic variance of the target disease is partitioned into a disease-specific component and a pleiotropic component. The private component is not shared with the auxiliary diseases, and the pleiotropic component consists of a linear combination of the genetic values (G_1, G_2, …, G_n) for auxiliary diseases 1 to n.

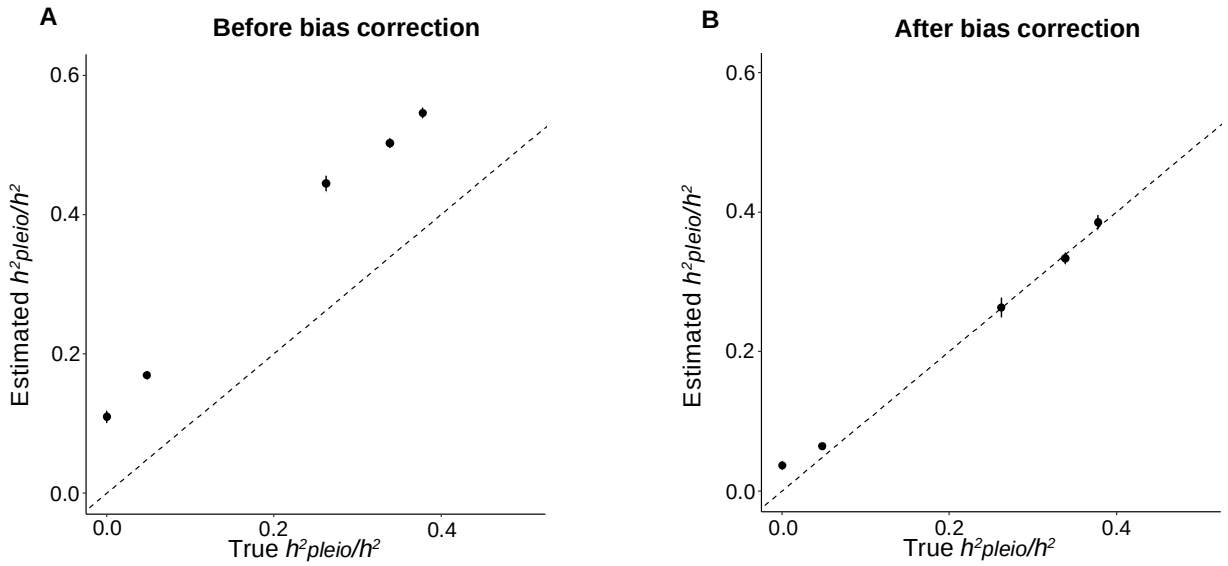


Figure 2. PHBC corrects upwards bias in simulations. (A) Estimates of $h^2{}_{pleio}/h^2$ without the bias correction step suffer upward bias. (B) Estimates of $h^2{}_{pleio}/h^2$ are approximately unbiased after the Monte-Carlo bias correction. Each dot represents the mean of all estimates for diseases with the same true value of $h^2{}_{pleio}/h^2$ across 100 simulations. Error bars denote standard errors. Numerical results are reported in **Supplementary Table 4**.

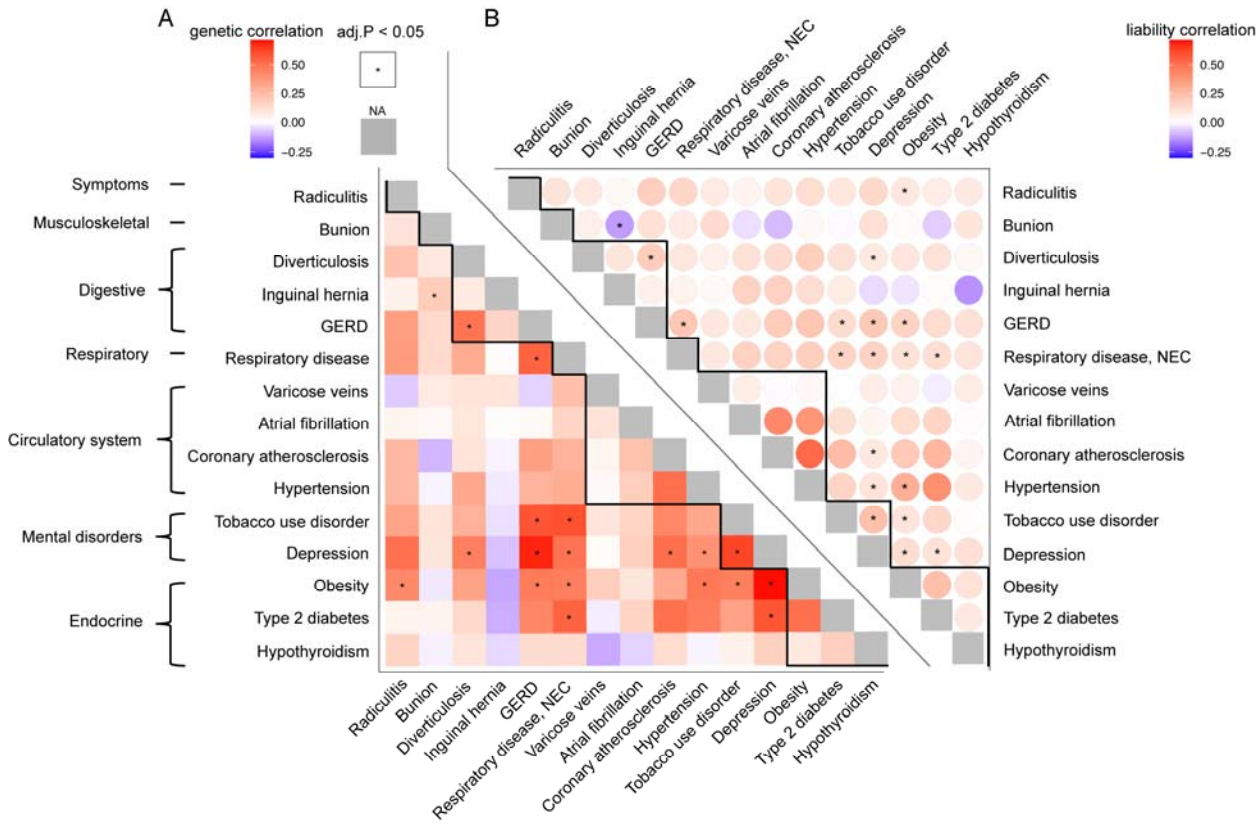


Figure 3. Estimates of genetic correlation and liability-scale phenotypic correlation. (A) Estimates of genetic correlation across 15 UK Biobank diseases. (B) Estimates of liability-scale phenotypic correlation across 15 UK Biobank diseases. Asterisks denote pairs of diseases for which the genetic correlation is significantly larger than the liability-scale phenotypic correlation with Bonferroni-corrected $P < 4.76 \times 10^{-4}$ (= 0.05/105); no pairs have liability-scale phenotypic correlation significantly larger than the genetic correlation. Black boundaries demarcate correlations between diseases within the same disease category. Numerical results are reported in **Supplementary Table 15**.

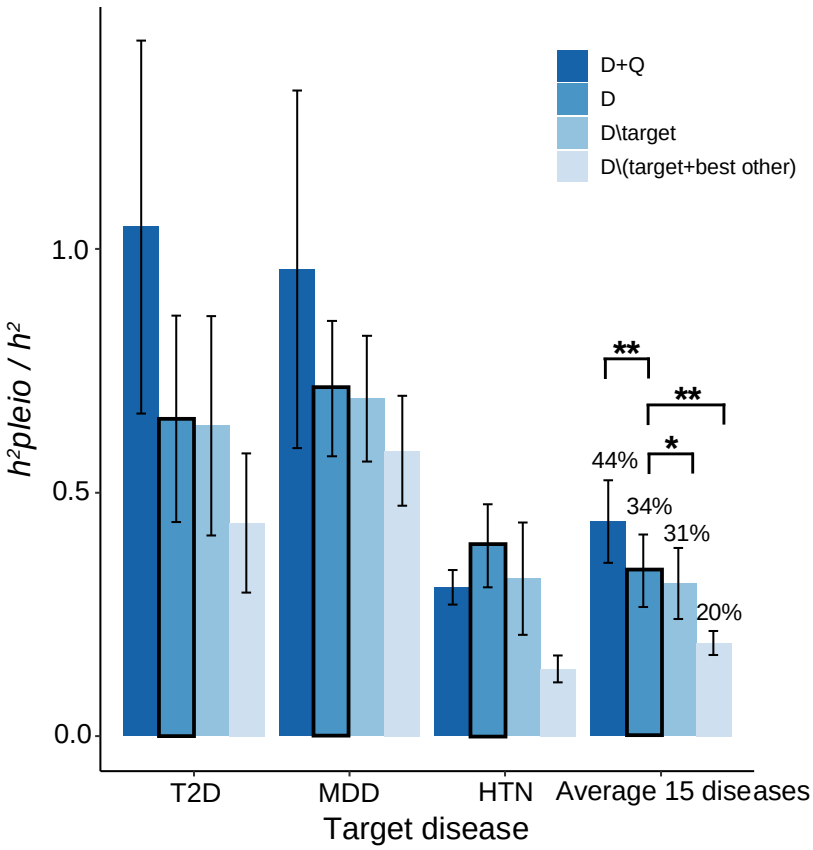


Fig. 4. $h^2{}_{pleio}/h^2$ **estimates for three representative diseases and the average across 15 UK Biobank diseases.** D+Q: $h^2{}_{pleio}/h^2$ estimates with respect to 15 UK Biobank auxiliary diseases and 16 UK Biobank quantitative auxiliary traits. D: $h^2{}_{pleio}/h^2$ estimates with respect to 15 UK Biobank auxiliary diseases. D\target: $h^2{}_{pleio}/h^2$ estimates with respect to 15 UK Biobank auxiliary diseases excluding those from the same disease category as the target disease. D\(target+best other): $h^2{}_{pleio}/h^2$ estimates with respect to 15 UK Biobank auxiliary diseases excluding those from the same disease category as the target disease and those from one other disease category whose removal had the greatest impact. Error bars denote jackknife standard errors. To assess statistical significance, we compare average estimates from D+Q, D\target and D\(target+best other) to estimates from D using inverse-variance weighting across diseases/traits. *: $P < 0.05$; **: $P < 0.001$. T2D: type 2 diabetes; MDD: depression; HTN: hypertension. Numerical results are reported in **Supplementary Tables 17** and **18**.

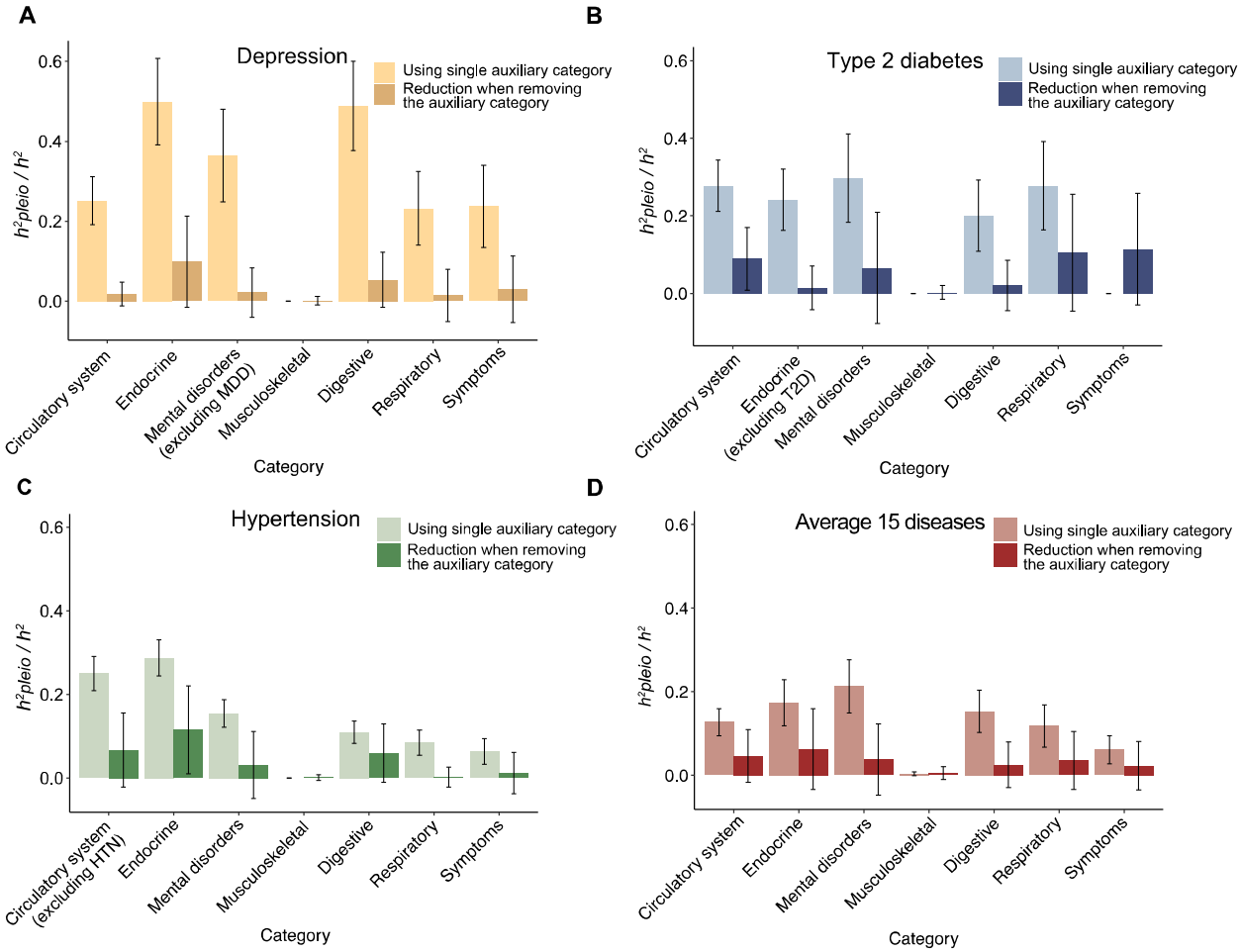


Fig. 5. $h^2{}_{pleio}/h^2$ is broadly shared across PheCode disease categories. (A-C) Comparison between $h^2{}_{pleio}/h^2$ using a single auxiliary disease category (light bars) and the reduction in $h^2{}_{pleio}/h^2$ when removing the auxiliary disease category (dark bars), for three representative UK Biobank diseases (depression, type 2 diabetes, hypertension) and the average across 15 UK Biobank diseases. Error bars denote jackknife standard errors. Numerical results are reported in **Supplementary Table 19**.

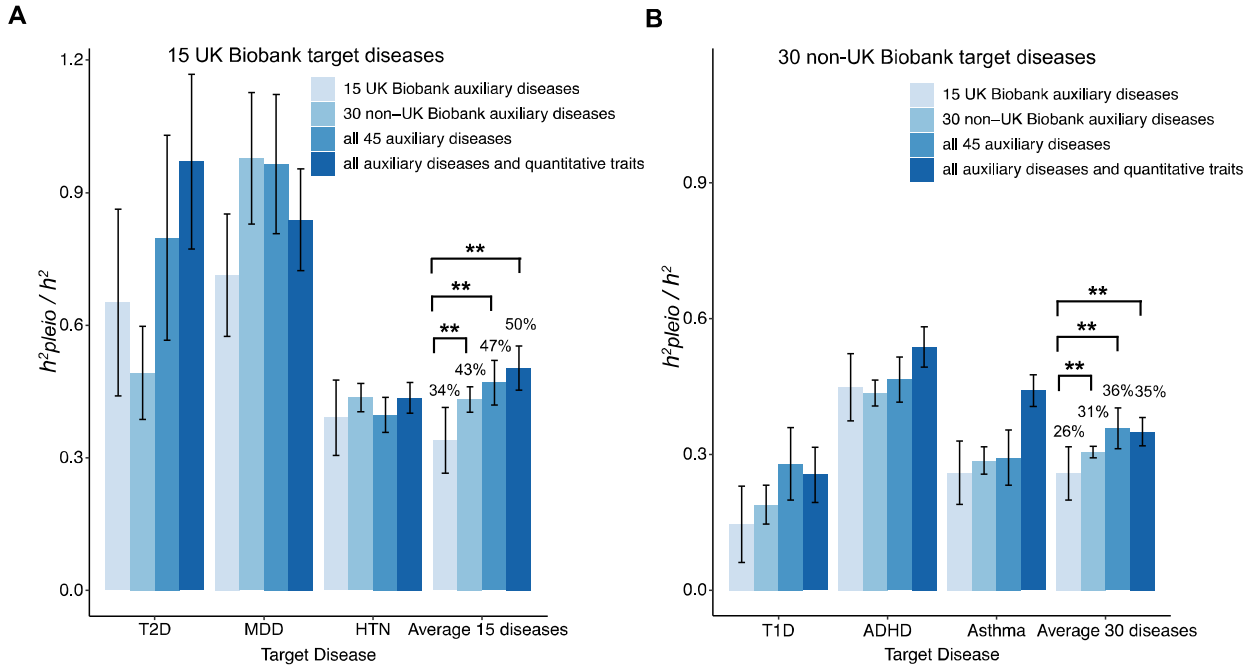


Fig. 6. $\boldsymbol{h^2_{pleio}/h^2}$ **estimates with respect to different sets of auxiliary diseases/traits.** (A) h^2_{pleio}/h^2 estimates for three representative UK Biobank diseases and the average across 15 UK Biobank diseases with respect to four sets of auxiliary diseases/traits. (B) h^2_{pleio}/h^2 estimates for three representative non-UK Biobank diseases and the average across 30 non-UK Biobank diseases with respect to four sets of auxiliary diseases/traits. Error bars denote jackknife standard errors. To assess statistical significance, we compare average estimates from 30 non-UK Biobank auxiliary diseases, all 45 auxiliary diseases, and all 61 auxiliary diseases and quantitative traits to estimates from 15 UK Biobank auxiliary diseases using inverse-variance weighting across diseases/traits: * $P < 0.05$; ** $P < 0.001$. Numerical results are reported in **Supplementary Tables 18, 22** and **23**.

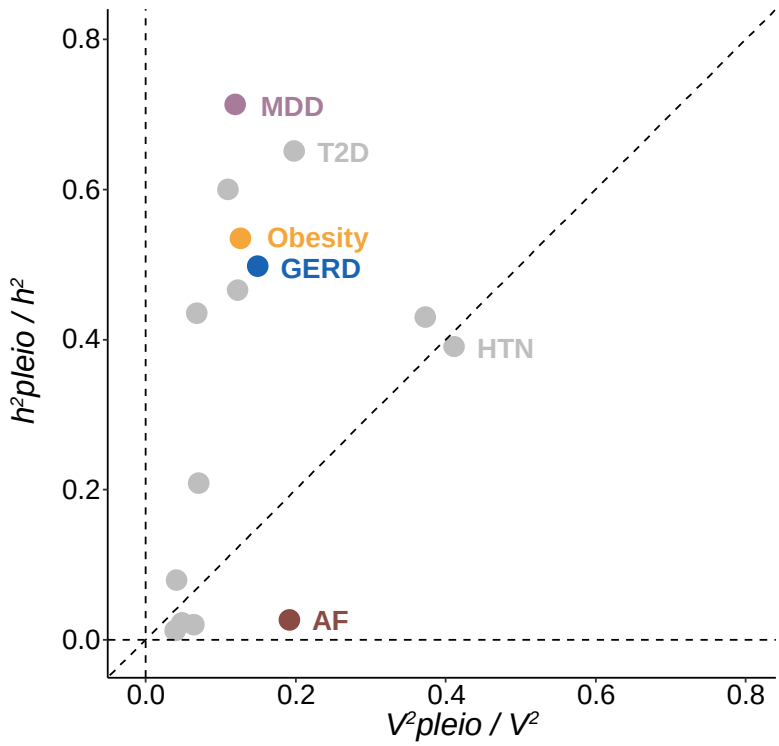
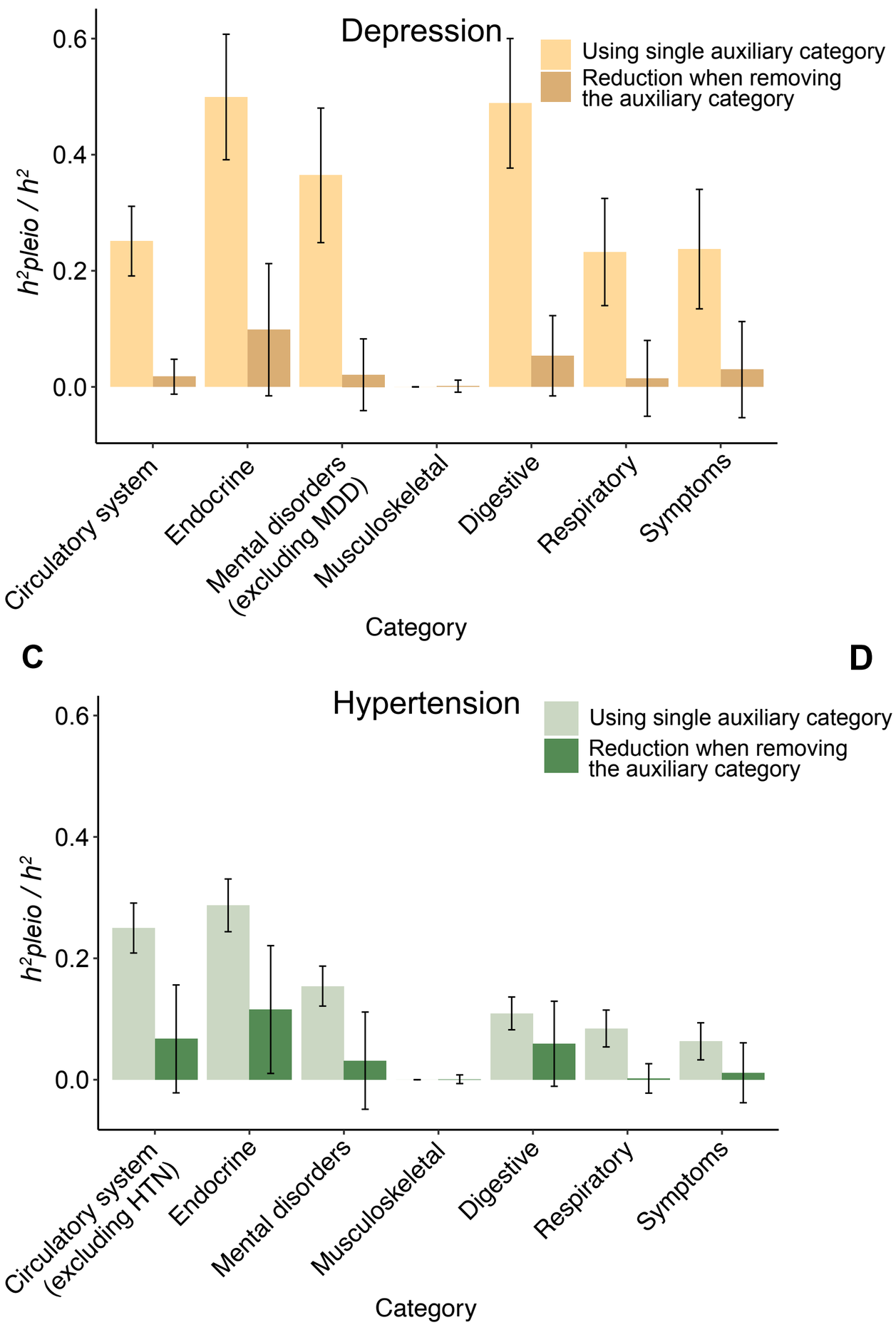
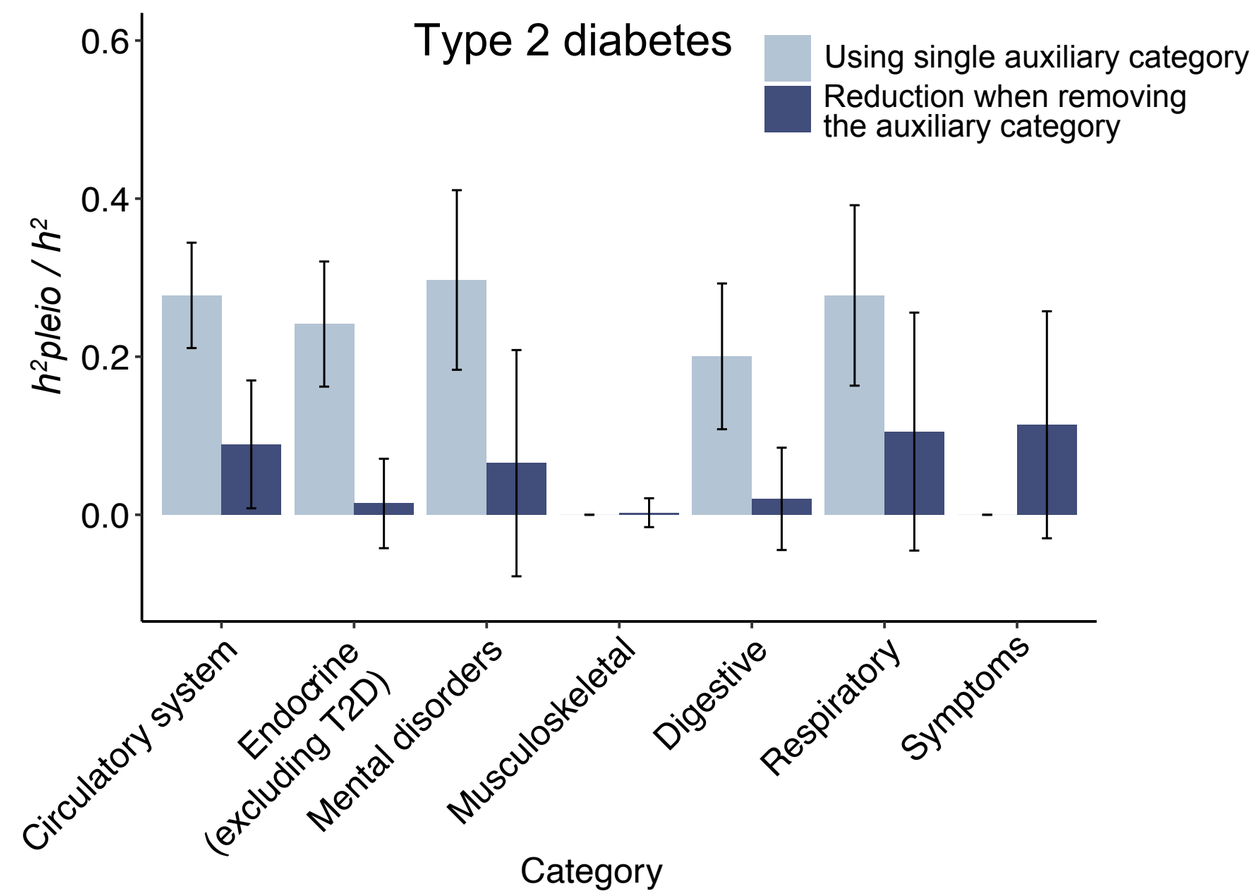
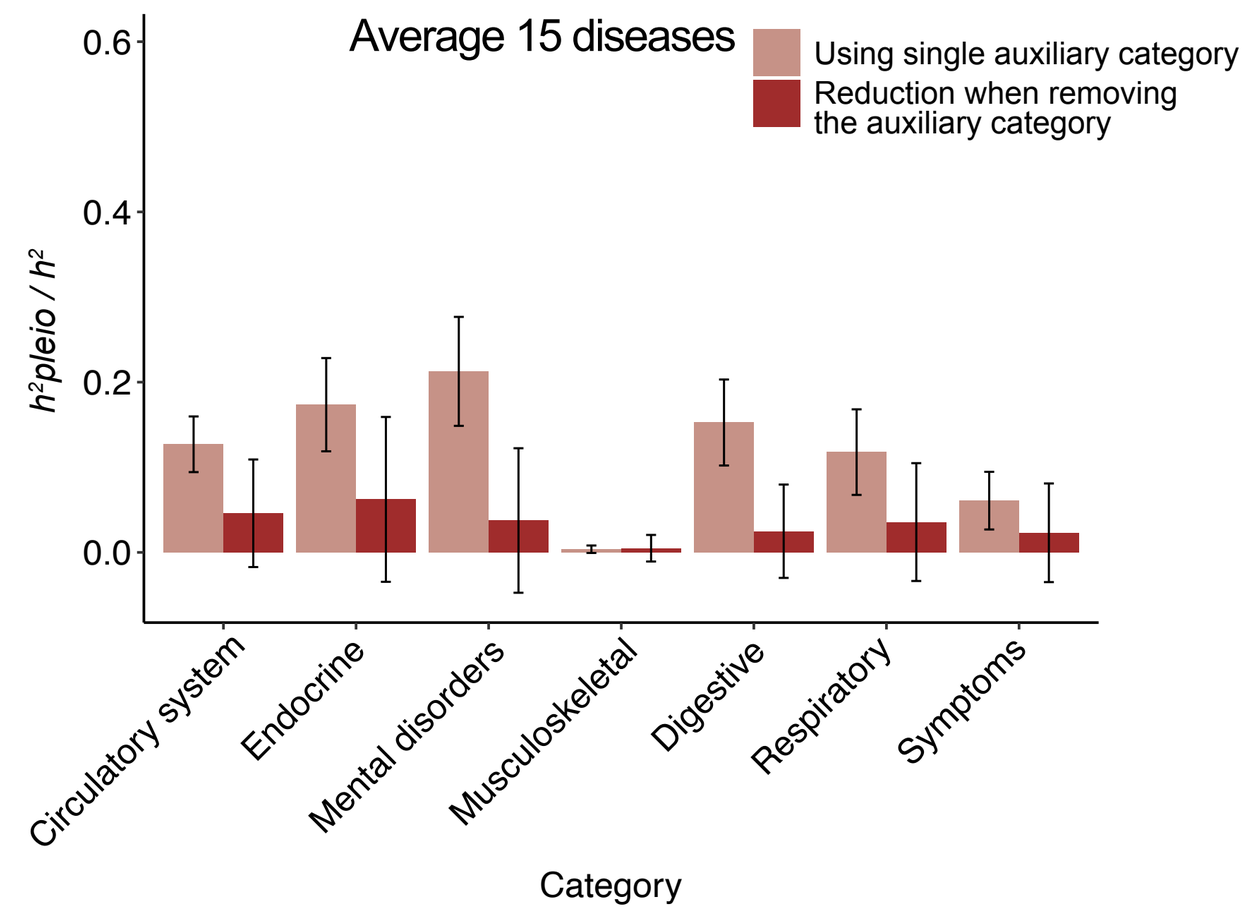


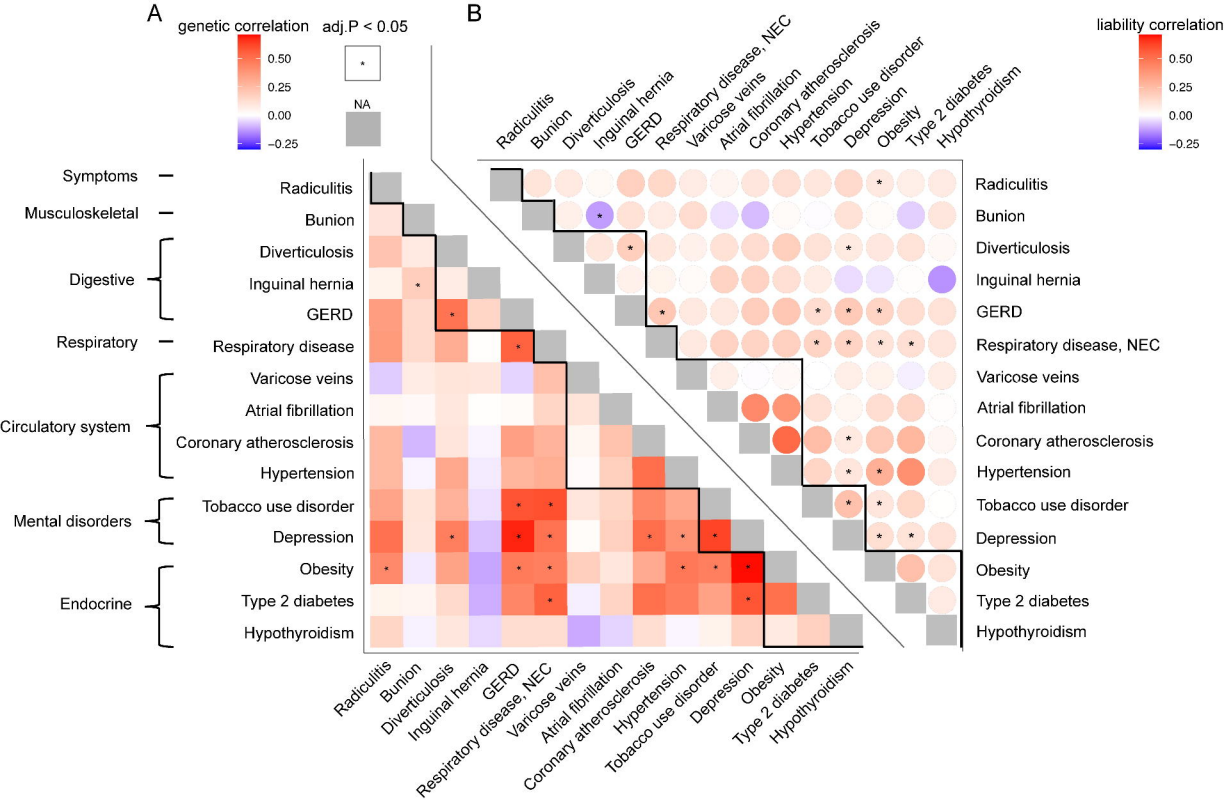
Fig. 7. Comparison of $h^2{}_{pleio}/h^2$ vs. $V^2{}_{pleio}/V^2$ for 15 diseases in UK Biobank. Colored dots denote diseases for which $h^2{}_{pleio}/h^2$ estimates are significantly different from $V^2{}_{pleio}/V^2$ estimates ($P < 0.05/15$). Numerical results are reported in **Supplementary Table 26**. MDD: depression; T2D: type 2 diabetes; HTN: hypertension; GERD: gastroesophageal reflux disease; AF: Atrial Fibrillation.

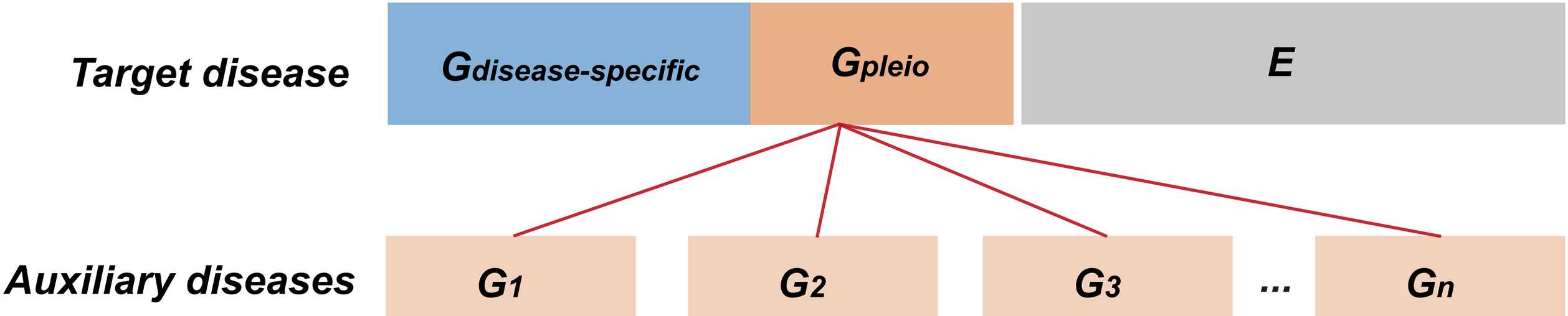

A
Depression
Using single auxiliary category
Reduction when removing the auxiliary category
h^2pleio / h^2
0.6
0.4
0.2
0.0
Circulatory system
Endocrine
Mental disorders (excluding MDD)
Musculoskeletal
Digestive
Respiratory
Symptoms
Category

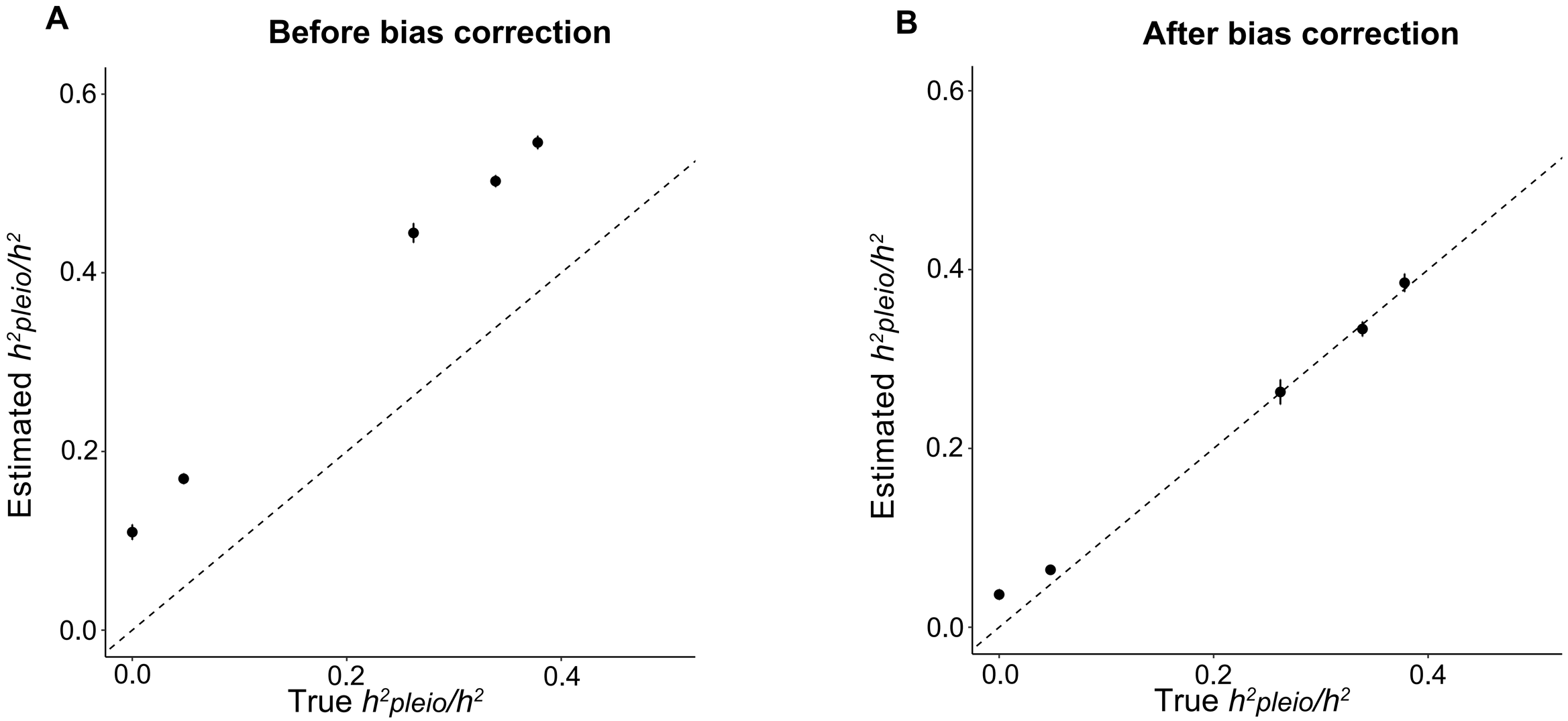

B
Type 2 diabetes
Using single auxiliary category
Reduction when removing the auxiliary category
h^2pleio / h^2
0.6
0.4
0.2
0.0
Circulatory system
Endocrine (excluding T2D)
Mental disorders
Musculoskeletal
Digestive
Respiratory
Symptoms
Category

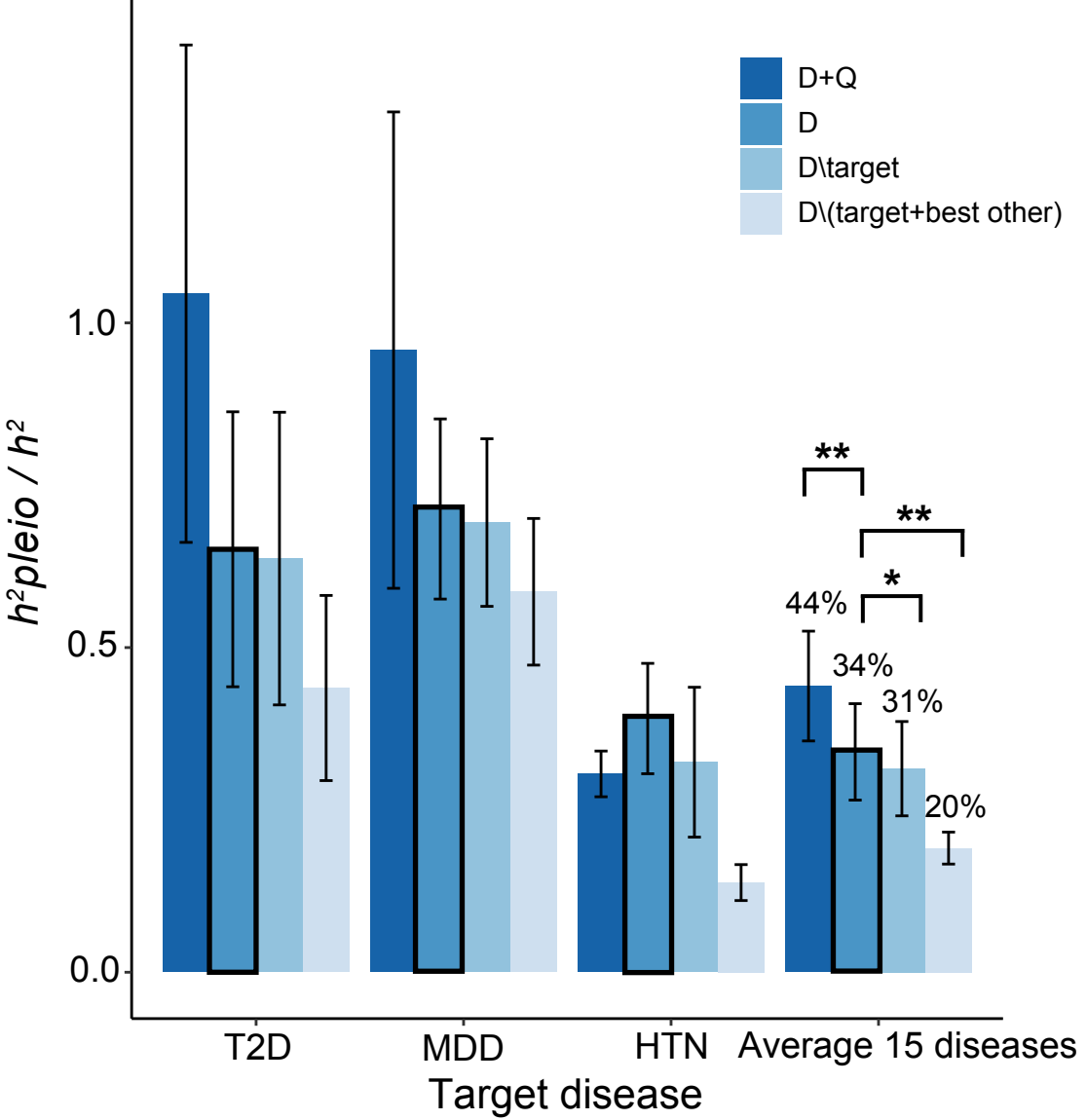
C
Hypertension
Using single auxiliary category
Reduction when removing the auxiliary category
h^2pleio / h^2
0.6
0.4
0.2
0.0
Circulatory system (excluding HTN)
Endocrine
Mental disorders
Musculoskeletal
Digestive
Respiratory
Symptoms
Category

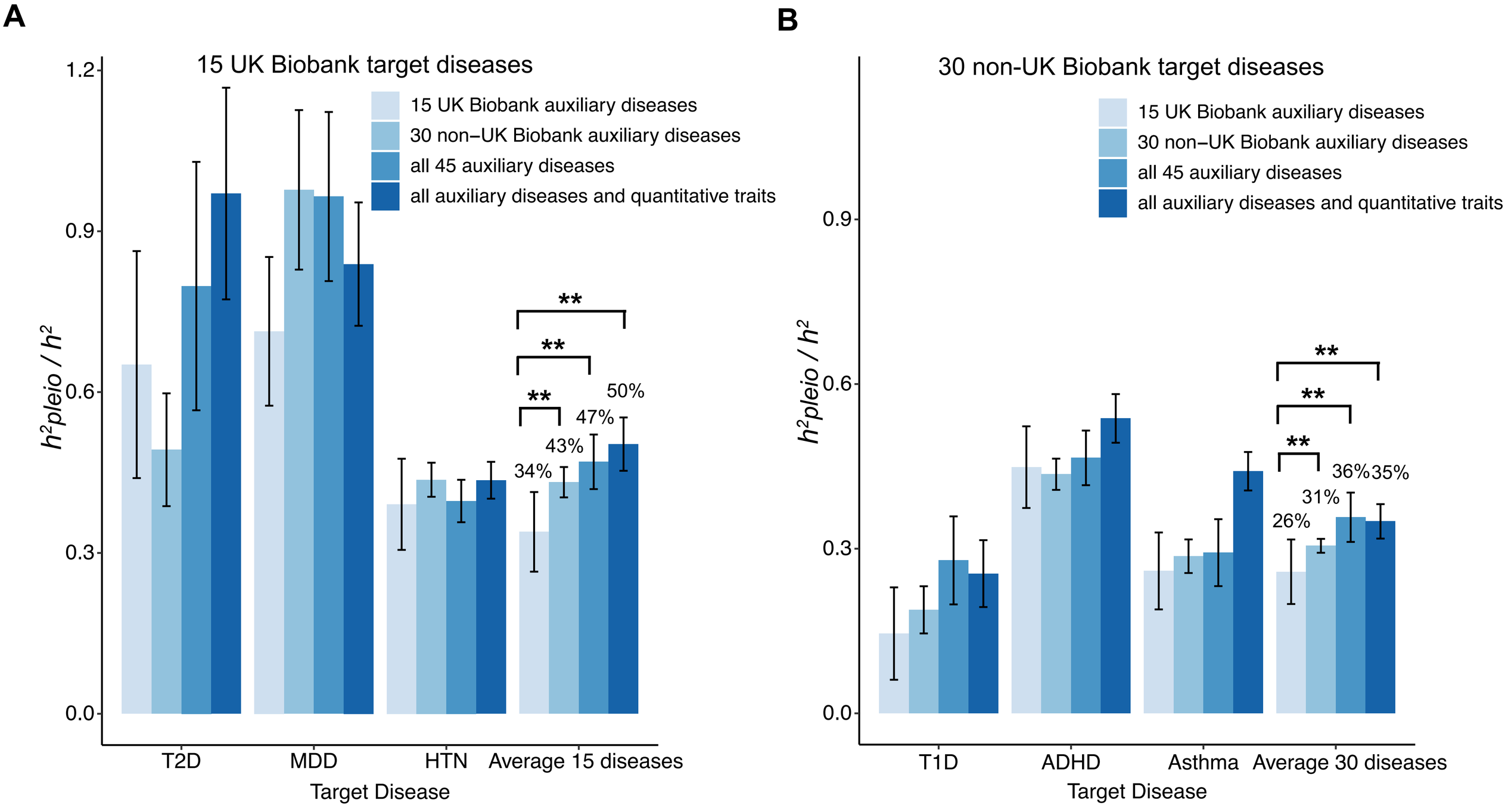

D
Average 15 diseases
Using single auxiliary category
Reduction when removing the auxiliary category
h^2pleio / h^2
0.6
0.4
0.2
0.0
Circulatory system
Endocrine
Mental disorders
Musculoskeletal
Digestive
Respiratory
Symptoms
Category

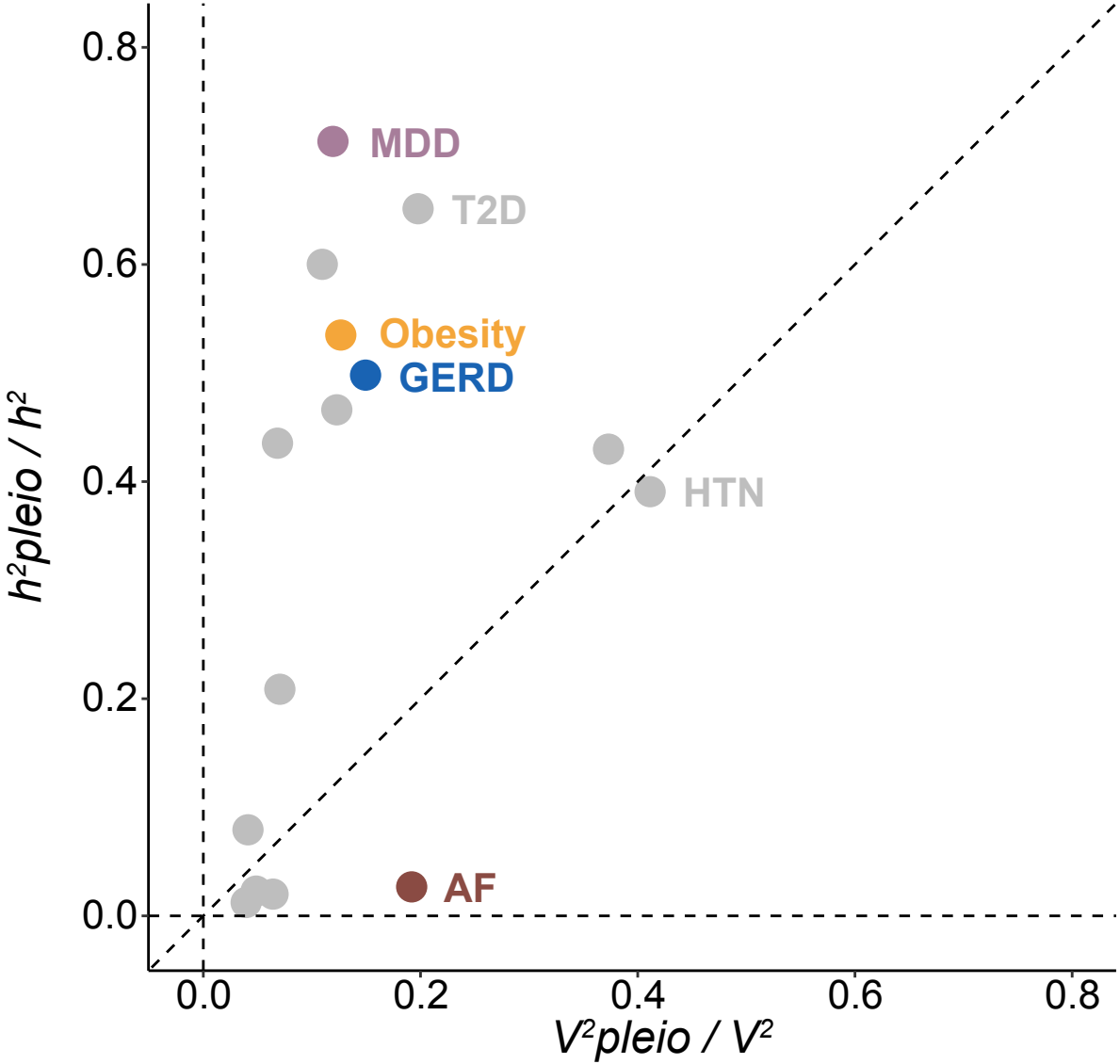

A
genetic correlation
0.50
0.25
0.00
−0.25
adj.P < 0.05
*
NA
B
liability correlation
0.50
0.25
0.00
−0.25
Symptoms
Musculoskeletal
Digestive
Respiratory
Circulatory system
Mental disorders
Endocrine
Radiculitis
Bunion
Diverticulosis
Inguinal hernia
GERD
Respiratory disease
Varicose veins
Atrial fibrillation
Coronary atherosclerosis
Hypertension
Tobacco use disorder
Depression
Obesity
Type 2 diabetes
Hypothyroidism
Respiratory disease, NEC


Target disease
Gdisease-specific
Gpleio
E
Auxiliary diseases
G1
G2
G3
...
Gn


A
Before bias correction
Estimated h^2pleio/h^2
True h^2pleio/h^2
0.0
0.2
0.4
0.6
B
After bias correction
Estimated h^2pleio/h^2
True h^2pleio/h^2
0.0
0.2
0.4
0.6


D+Q
D
D\target
D\(target+best other)
1.0
0.5
0.0
h^2pleio / h^2
**
**
*
44%
34%
31%
20%
T2D
MDD
HTN
Average 15 diseases
Target disease


A
15 UK Biobank target diseases
15 UK Biobank auxiliary diseases
30 non−UK Biobank auxiliary diseases
all 45 auxiliary diseases
all auxiliary diseases and quantitative traits
h^2pleio / h^2
1.2
0.9
0.6
0.3
0.0
**
**
**
34%
43%
47%
50%
T2D
MDD
HTN
Average 15 diseases
Target Disease
B
30 non-UK Biobank target diseases
15 UK Biobank auxiliary diseases
30 non−UK Biobank auxiliary diseases
all 45 auxiliary diseases
all auxiliary diseases and quantitative traits
h^2pleio / h^2
0.9
0.6
0.3
0.0
**
**
**
26%
31%
36%
35%
T1D
ADHD
Asthma
Average 30 diseases
Target Disease


MDD
T2D
Obesity
GERD
HTN
AF
0.0
0.2
0.4
0.6
0.8
h^2pleio / h^2
V^2pleio / V^2