Data S1: More details about evaluating the integration of GWAS in biobanks, investigating the pleiotropic effects of loci, and prioritizing functional genes

Using genetic association results to evaluate the integration of association results across biobanks

Here we provide a more extensive description for the section “Integration of association results across biobanks” in the main text. We used the genetic association results to evaluate the integration of different biobanks in the meta-analyses for genetic discovery. First, we compared the effect sizes of top variants with p-value < $1x10^{-10}$ by all-biobank meta-analyses in individual biobanks and the corresponding LOBO meta-analyses. For each biobank and LOBO pair, we fit a Deming regression model [S24], which accounts for standard errors of effect size estimates in both association datasets, with the intercept set to zero. In **Figure S8,** the slope estimates for biobank and LOBO pairs were plotted against the effective sample sizes. Biobanks were annotated by phenotype source (health records (e.g. ICD codes and physician’s diagnosis) only or self-reported data included), sampling strategy, and sample ancestry. Most of the slope estimates were not significantly different from one across biobanks and phenotypes suggesting the genetic association results are robust despite differences among biobanks. However, we observed exceptions to this among biobanks with relatively smaller sample sizes and biobanks containing samples of non-European or multiple ancestries. For example, the multi-ancestry biobanks BioMe, BioVU, and UCLA as well as GNH, have different effects compared to others for multiple phenotypes, including gout, HF, VTE, and POAG. Note that POAG tends to have more phenotypic heterogeneity due to glaucoma types not being well defined by self-reported data, leading to the inclusion of other types of glaucoma, such as the angle-closure glaucoma. As expected, three biobanks using self-reported data for phenotype curation, Lifelines, TWB, and BBJ, showed effect size differences for POAG compared to other biobanks (**Figure S8**). We estimated genetic correlation between individual biobanks and LOBO for the three endpoints with highest heritability estimates: asthma, gout, and COPD. Genetic correlation estimates between biobanks and LOBO were close to 1, although genetic correlation was only possible to estimate for biobanks with non-zero heritability estimates (p-value < 0.05) (**Methods**, **Figure S9**).

To investigate effect size differences between 9 population-based biobanks (CKB, DECODE, ESTBB, GNH, GS, HUNT, Lifelines, TWB, and UKBB) and 6 hospital-/healthcare-based biobanks (BBJ, BioMe, BioVU, MGB, MGI, and UCLA), we conducted meta-analyses for the two biobank groups separately and fitted the Deming regression on effect size estimates of loci identified by all-biobank meta-analyses. Loci with association p-values < $1x10^{-6}$ in both meta-analyses were included in the regression. 4 endpoints (gout, ThC, Asthma and POAG) that had more than 5 qualified loci were analyzed. The scatter plots with regression lines and slope estimates were presented in **Figure S10**. Effect size estimates were observed to be consistent between population-based biobanks and hospital-/healthcare-based biobanks for all 4 endpoints as 95% confidence intervals (CI) for the slope estimates included 1.

We then compared all-biobank meta-analyses results with published GWAS studies. For previously reported loci, consistent effect direction was observed between GBMI and the previous largest studies. For example, all 18 loci that were previously identified by the Trans-National Asthma Genetic Consortium (TAGC) [S25] for asthma had consistent effect directions in GBMI and TAGC and more significant association p-values in GBMI (**Figure S11**). Similarly, all 24 previously identified loci for AAA by MVP [S26] and all 40 previously identified loci for gout [S27] showed effect size consistency between GBMI and the previous GWASs (**Table S15**). Note that by cross-comparing the cohort lists in previous studies and in GBMI, no sample overlap was noted for Asthma and AAA, while 3 biobanks in GBMI (BioVU, GS, and UKBB) were also included in the previous meta-analysis for gout [S27], accounting for about 20% of samples included in the meta-analysis for gout in GBMI.

For some phenotypes such as IPF[S28] and asthma[S29], we observed attenuation of effect size estimates in GBMI compared to disease-specific cohorts which generally study highly ascertained patients. For IPF, the attenuation was seen compared to the IPF-specific cohort: the Allen et al. study[S30]. To further investigate the impacts of case ascertainment on effect size estimates, the IPF working group[S28] in GBMI divided FinnGen into three subsets based on diagnosis and original study cohort for IPF: a clinical IPF cohort (FinnishIPF, n cases = 205), other IPF patients (n = 1,366), and non-IPF ILD patients (n = 1,624) and compared effect size estimates from these cohorts to those of the latest IPF meta-analysis by Allen et al.[S30]. We observed that effect size estimates were 0.9, 1.4, and 2.5-times larger in the latest IPF meta-analysis by Allen et al. [S30] compared with the FinnishIPF, other IPF, and non-IPF ILD cohorts, which provides further evidence that effect sizes in highly ascertained IPF patients are substantially higher compared with patients identified from biobanks. Similarly, for asthma, we observed the attenuation of effect size estimates in GBMI compared to TAGC[S29]. However, for other endpoints, such as VTE[S31], the effect size estimates in GBMI are well aligned with those in the disease-specific consortium, INVENT[S32]. The consistent magnitudes of effect size estimates were also observed for AAA in GBMI and in MVP [S26] **(Table S15).**

Investigating the pleiotropic effects of associated loci

Here we provide a more extensive description for the section “Pleiotropic effects of associated loci ” in the main text. Of 430 loci whose index variants were tested in the UKBB GWAS data, 78 variants identified from 12 GBMI endpoints (except for HCM and UtC) exhibited significant (p-value $< 5 \times 10^{-8}$) pleiotropic associations with at least one other phenotype (**Table S16**). Risk increasing alleles of top variants at two asthma-associated loci, the known asthma locus BACH2 and the novel locus *FGFR1OP*, are both associated with a reduced risk of hypothyroidism. The risk increasing allele of the top variant at the novel locus *GOT1/LINC01475* for acute appendicitis (AcApp) is associated with a decreased risk of ulcerative colitis. A previous study also observed a low risk of ulcerative colitis among people who had undergone an appendectomy for appendicitis and mesenteric lymphadenitis [S33], but the underlying cause remains unclear.

We have done a more extensive investigation on pleiotropic effects of the 52 loci (30 novel) identified for gout by all-biobank meta-analysis, which has been less well studied by previous GWAS studies. A vast majority of these loci (n=40, either same SNP or multiple different SNPs) were associated with serum urate levels [S27, S34–S36] (**Table S17**)**,** including key urate transporter genes *SLC2A9* and *SLC22A12* [S37]. We also found that most of these loci were associated with other relevant traits and diseases **(Table S17).** For example, *RAB24* and *MC4R* were associated with BMI related traits, *MPPED2*, *A1CF*, *BCAS3, LRP2*, *MTX2*, *TRIM46*, *SFMBT1*, and *STC1* were associated with kidney function related traits, *ARID1A*, *BMPR2*, *MLXIPL*, and *HNF4A* were associated with lipid traits, *GCKR* was associated with diabetes, and *PDZK1*, *PDE1A*, and *SLC22A7* were associated with blood pressure traits. Previous studies have already speculated the possible mechanisms for the involvement of these traits or processes in gout etiology. For example, coating of urate crystals with Apolipoprotein B can down-regulate the innate immune system by suppressing neutrophil activation[S38] and neutrophil activation is needed for the endocytosis and lysis of urate crystals and thus the resolution of gout attack[S39]. Similar biological links have also been proposed for other above-mentioned traits and uric acid metabolism and thus can explain the observed association of related genes with gout risk in our study.

Prioritizing functional genes

Here we provide a more extensive description for the section "Prioritization of cell types, tissues, and genes".

To further understand the biology underlying the genetic associations, we first prioritized tissues and cell types in which genes at the associated loci are likely to be highly expressed using the Data-driven Expression-Prioritized Integration for Complex Traits (DEPICT)[S40] (**Table S18).** For example, at FDR < 0.05, the adrenal cortex, which releases the mineralocorticoid aldosterone, was prioritized for AAA. This result agrees with previous functional studies which have shown that the mineralocorticoid aldosterone can induce aortic aneurysm and dissection in the presence of high salt[S41]. Prioritized tissue types for asthma included lymphoid tissue and immune systems (blood cells, antigen presenting cells, and myeloid cells) as well as nasal and respiratory mucosa. Besides muscle cells and connective tissue cells, heart and blood vessels were identified for POAG[S42].

Next, note that prioritizing potentially functional genes based on the genetic variant associations is one of the biggest areas of challenge and research. We applied several methods to prioritize potentially functional genes, including DEPICT (**Table S19)**, the gene-level Polygenic Priority Score (PoPS) [S43] (**Table S20**), transcriptome-wide association studies (TWAS) [S44](**Table S21**), and proteome-wide Mendelian randomization (PWMR) [S45] (**Table S22， Methods**). Using asthma, POAG, and VTE as examples, the gene lists generated by these different methods showed quite little overlap (**Figure S12**). For asthma, 618 genes were prioritized by at least one of the four approaches (FDR < 0.05 by DEPICT, top 1% scores in PoPS, $P < 2.5x10^{-6}$ by TWAS, P < 0.001 by PWMR) (**Figure S12A**). However no genes were prioritized by all four methods and 5 were prioritized by any three methods (*FCER1G, IL18R1, IL4R*, and *SMAD3* by DEPICT, TWAS, and PoPS and *IL2RB* by DEPICT, PoPS, and PWMR) (**Table S22**). All these genes are located at the well-known asthma-associated loci. *FCER1G* encodes the Fc Fragment of IgE Receptor Ig, and *IL18R1, IL4R,* and *IL2RB* encode Interleukin receptors, which are all involved in the immune system. Dupilumab, an anti-interleukin 4 receptor alpha monoclonal antibody, blocks IL-4 and IL-13 and decreases IgE over time, is an FDA approved add-on therapy for asthma [S46, S47]. *SMAD3* encodes a transcription factor whose methylation has been shown to be associated with neonatal production of IL-1β and childhood asthma risk [S48]. We then extracted the nearest genes of the most significant variants (for intergenic variants, the nearest genes on both sides were included if both are located within 50kb from the top hits), which brings the total number of prioritized genes to 729. *FCER1G, IL4R*, and *SMAD3* that were prioritized by DEPICT, TWAS, and PoPS were also the nearest genes of top hits at those loci. 17 more genes were prioritized by any of the two methods and the naive nearest genes approach: *BCL2 ,CD247, CD28, GSDMB, HDAC7, IL13, IL2RA, IL6R, IL7R, ITPKB, JAZF1, NEK6, PTPRC, RUNX3, STAT6, TLR1,* and *TNFSF8*. Similarly, for POAG, 204 genes were prioritized, but no genes were prioritized by all four or any three methods, and five genes were prioritized by two out of the four methods (*CAV1, CDH11, PLCE1*, and *PRSS23* by PoPs and DEPICT, *CDKN2A* by PoPs and TWAS), but none were nearest genes at the loci (**Figure S12B**). The nearest gene approach prioritized 42 more genes that were not prioritized by any of the four methods, 3 nearest genes were also prioritized by DEPICT (*AFAP1,BICC1,* and *COL11A1*) and 2 nearest genes were also prioritized by TWAS (*GAS7* and *PDE7B*).

For VTE, 244 genes were prioritized by the four prioritization methods. One well-known VTE-associated gene, *F2*, that encodes the coagulation factor II, was prioritized by all four methods and is also the nearest gene at the locus. 5 genes were prioritized by three methods, DEPICT, PWMR, and PoPS and all are nearest genes at those loci (*F5, PLCG2, PLEK, PROC*, and *PROS1*). To our best knowledge, no functional study has been done to explore the role of gene *PLEK* in VTE. 16 nearest genes were not prioritized by any of the four methods. (**Figure S12C**). In addition, a gold standard set of 41 VTE genes was curated blindly from the meta-analysis results[S31]. Based on this gene set, the precision and recall of the gene prioritization methods were estimated. PWMR had the highest precision and 6 out of 18 genes (33.3%) prioritized by PWMR were in the gold standard set, followed by DEPICT with precision 26.7% (8 out of 30) and the nearest gene approach with precision 20.6% (7 out of 34). PoPS (with the top one percentile PoPS score cutoff) had the highest recall, which prioritized 17 out of the 41 genes in the gold standard set, followed by DEPICT and the nearest gene approach, both of which prioritized 8 and 7 genes in the gold standard set, respectively[S31]. 13 genes in the gold standard set are located within 1Mb around the VTE top hits. Using these genes as a gold standard, as expected, we observed an increase in the recall of DEPICT and the nearest gene approach from 26.7% to 61.5% and from 20.6% to 53.8%, respectively. This is because both approaches tend to prioritize genes that are located at GWAS loci (**Figure S13C** and **Table S23**). In addition, when a more stringent gene prioritization score cutoff, top 0.1 percentile, was applied for PoPS, there was a decrease in the recall rate and an increase in precision compared to those with the top one percentile cutoff (**Figure S13B and 13C**).

In line with what has previously been discussed [S43], these results showed that the existing gene prioritization methods successfully prioritized relevant genes for diseases but had poor agreement. Note that besides adapting different statistical models and pipelines, these approaches prioritize genes based on different expression data types: DEPICT uses the co-regulation of gene expression data with the pre-annotated gene sets [S40], PoPS leverages the cell-type specific gene expression, biological pathways, and protein-protein interactions [S43], TWAS is based on expression quantitative trait loci (eQTLs) [S44], and PWMR is based on the protein quantitative trait loci (pQTLs)[S45] (**Methods**). Using VTE as an example, the nearest gene approach performs comparable to other methods. Our results highlight the challenges in interpreting genome-wide significant loci and the clear need for robust in silico approaches and pipelines to nominate genes for experimental follow-up.