Article

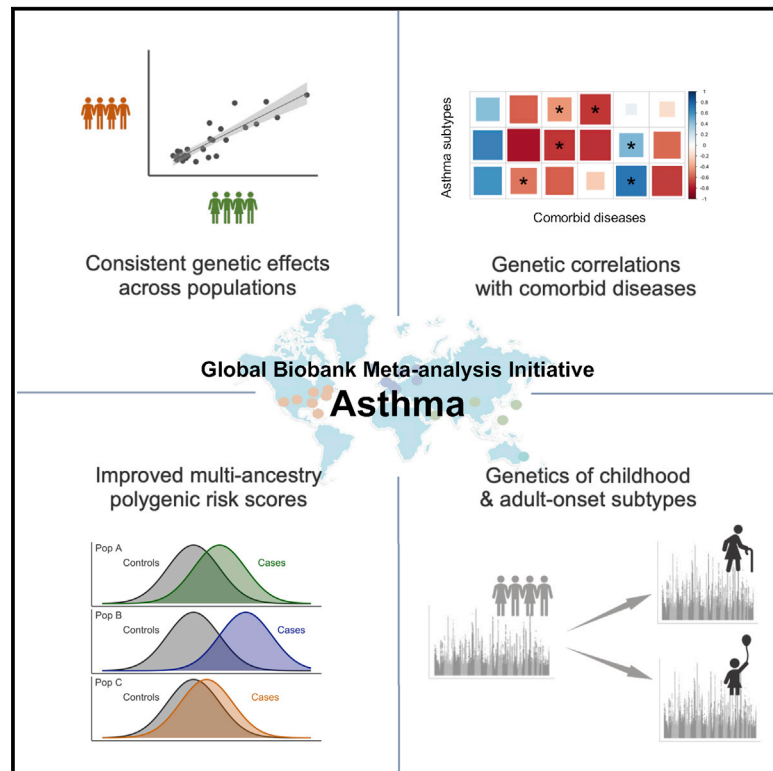# Multi-ancestry meta-analysis of asthma identifies novel associations and highlights the value of increased power and diversity

Graphical abstract

Highlights

- 22 biobank meta-analysis of >150k individuals with asthma discovers 49 novel loci

- Genetic effects are consistent across ancestries and biobanks with varying prevalence

- Greater ancestral diversity improves genetic discovery and risk prediction

- Strong genetic correlations are observed between asthma subtypes and comorbidities

Authors

Kristin Tsuo, Wei Zhou, Ying Wang, ..., Global Biobank Meta-analysis Initiative, Mark J. Daly, Alicia R. Martin

Correspondence

ktsuo@broadinstitute.org (K.T.), mjdaly@broadinstitute.org (M.J.D.), armartin@broadinstitute.org (A.R.M.)

In brief

Tsuo et al. investigated genetic signatures underlying asthma using data from 22 biobanks worldwide. They demonstrate that the increased diversity and sample size of this resource can accelerate gene discovery, improve risk prediction, and advance our understanding of asthma's shared genetic basis across populations and with related diseases.

CellPress

Cell Genomics

Article

Multi-ancestry meta-analysis of asthma identifies novel associations and highlights the value of increased power and diversity

Kristin Tsuo,[1,2,3,14,*] Wei Zhou,[1,2,3] Ying Wang,[1,2,3] Masahiro Kanai,[1,2,3,4,5] Shinichi Namba,[5] Rahul Gupta,[1,2,3,6] Lerato Majara,[7] Lethukuthula L. Nkambule,[1,2,3] Takayuki Morisaki,[8] Yukinori Okada,[5,9,10,11,12] Benjamin M. Neale,[1,2,3] Global Biobank Meta-analysis Initiative, Mark J. Daly,[1,2,3,13,*] and Alicia R. Martin[1,2,3,*]

[1]Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA, USA
[2]Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA, USA
[3]Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA, USA
[4]Department of Biomedical Informatics, Harvard Medical School, Boston, MA, USA
[5]Department of Statistical Genetics, Osaka University Graduate School of Medicine, Suita, Japan
[6]Howard Hughes Medical Institute and Department of Molecular Biology, Massachusetts General Hospital, Boston, MA, USA
[7]Department of Psychiatry and Mental Health, Faculty of Health Sciences, University of Cape Town, Cape Town, South Africa
[8]Division of Molecular Pathology, The Institute of Medical Science, The University of Tokyo, Minatu-ku, Tokyo, Japan
[9]Laboratory for Systems Genetics, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan
[10]Laboratory of Statistical Immunology, Immunology Frontier Research Center (WPI-IFReC), Osaka University, Suita 565-0871, Japan
[11]Department of Genome Informatics, Graduate School of Medicine, The University of Tokyo, Tokyo 113-0033, Japan
[12]Center for Infectious Disease Education and Research (CiDER), Osaka University, Suita 565-0871, Japan
[13]Institute for Molecular Medicine Finland, University of Helsinki, Helsinki, Finland
[14]Lead contact
*Correspondence: ktsuo@broadinstitute.org (K.T.), mjdaly@broadinstitute.org (M.J.D.), armartin@broadinstitute.org (A.R.M.)
https://doi.org/10.1016/j.xgen.2022.100212

SUMMARY

Asthma is a complex disease that varies widely in prevalence across populations. The extent to which genetic variation contributes to these disparities is unclear, as the genetics underlying asthma have been investigated primarily in populations of European descent. As part of the Global Biobank Meta-analysis Initiative, we conducted a large-scale genome-wide association study of asthma (153,763 cases and 1,647,022 controls) via meta-analysis across 22 biobanks spanning multiple ancestries. We discovered 179 asthma-associated loci, 49 of which were not previously reported. Despite the wide range in asthma prevalence among biobanks, we found largely consistent genetic effects across biobanks and ancestries. The meta-analysis also improved polygenic risk prediction in non-European populations compared with previous studies. Additionally, we found considerable genetic overlap between age-of-onset subtypes and between asthma and comorbid diseases. Our work underscores the multi-factorial nature of asthma development and offers insight into its shared genetic architecture.

INTRODUCTION

Asthma is a complex and multi-factorial disease that affects millions of people worldwide, yet much of its genetic architecture has eluded discovery. Genetic factors contribute substantially to asthma risk, with heritability estimates from twin studies ranging between 50% and 90%.[1,2] Early genome-wide association studies (GWASs) provided some evidence for the polygenic architecture of asthma,[3–5] but only in the past few years have genomic studies of asthma collated large enough sample sizes to more definitively articulate its polygenicity.[6] The most recent GWAS of asthma discovered 167 asthma-associated loci across the genome.[7] However, these risk loci only account for a small proportion of the total heritability of asthma. Furthermore, the discovery GWAS, like the majority of previous asthma GWASs,

were primarily conducted in populations of European ancestry. Some major exceptions are the EVE Consortium,[8] one of the first efforts to perform GWASs in populations of African American, African-Caribbean, and Latino ancestries, as well as the Transnational Asthma Genetic Consortium (TAGC),[9] which included modest sample sizes from populations of African, Japanese, and Latino ancestries in their meta-analysis. As these studies noted, efforts to conduct asthma GWASs in diverse populations are particularly important because the prevalence of asthma varies widely around the world. Surveys of asthma worldwide have found that prevalence can vary by as much as 21-fold among countries.[10,11] Within countries, prevalence of asthma ranges considerably as well,[12] and this variation cannot be attributed to any single known risk factor such as air pollution. Rather, the contributing genetic and environmental factors are

complex. Therefore, assessing the genetic architecture of asthma in diverse cohorts is critical to gaining a more comprehensive understanding of asthma risk.

This heterogeneity in prevalence is mirrored by, and may be a consequence of, the heterogeneity of the disease itself. The wide variability in underlying mechanistic pathways and clinical presentations of asthma has led to a shift away from its characterization as a single disease entity.[13–15] Instead, asthma is now commonly viewed as a syndrome encompassing several distinct yet inter-related diseases, each driven by a unique set of genetic and non-genetic risk factors.[13,14] Different subgroups of asthma, for example, share genetic components with various comorbid diseases, including other respiratory diseases such as chronic obstructive pulmonary disease (COPD), allergic diseases, obesity, and neuropsychiatric disorders.[16–22] This complexity in turn complicates standards for defining phenotypes to study; for example, one study found that nearly 60 different definitions of "childhood asthma" were used across more than 100 studies in the literature.[23] The heterogeneity of asthma thus presents many challenges in identifying genetic risk factors for asthma.

A greater understanding of the genetics underlying asthma risk can facilitate the development of more accurate clinical models of asthma that may help inform clinical intervention, prevention, and management strategies.[24] In particular, leveraging GWAS associations for genetic risk prediction models, such as polygenic risk scores (PRS), has shown potential in informing preventive clinical decision-making for several polygenic diseases.[25–27] For asthma, PRS could ultimately play a role in predicting disease severity and development in the clinical setting and serve as a tool for investigating gene-environment interactions in the research setting. So far, some GWASs have been applied to developing PRS for asthma,[28–32] but these models have had limited predictive ability, likely attributable to the insufficient sample sizes and diversity of existing datasets of asthma. This underscores the genetic complexity of asthma and highlights the need for more large-scale genomic studies of asthma.

To more deeply interrogate the genetic architecture of asthma across different populations through genetic discovery and prediction, we analyzed paired phenotypic and genetic data from the Global Biobank Meta-analysis Initiative (GBMI). Participating biobanks shared summary statistics for the meta-analyses of 14 disease endpoints: asthma, COPD, heart failure, stroke, gout, venous thromboembolism, primary open-angle glaucoma, abdominal aortic aneurysm, idiopathic pulmonary fibrosis, thyroid cancer, cardiomyopathy, uterine cancer, acute appendicitis, and appendectomy.[33] More details on the selection of these disease endpoints can be found in Zhou et al.[33] Compared with previous asthma resources and studies, this collaborative effort increased both the sample size and diversity of asthma cases by many-fold, covered more variants with high imputation quality, and harmonized phenotypes using consistent electronic health record definitions for asthma across datasets. Harnessing this resource, we identify 49 loci not previously associated with asthma. Despite prevalence differences of nearly an order of magnitude, we also demonstrate remarkable consistency of genetic effects across the biobanks

and ancestries in the GBMI. Furthermore, we show that the increased sample size and diversity of data from the GBMI improves genetic risk prediction accuracy in multiple populations. Finally, we show that this meta-analysis captures much of the genetic architecture underlying asthma age-of-onset subtypes, and we provide additional evidence for shared genetic architectures between asthma and comorbid diseases such as COPD. Our findings highlight the need for future investigations into how genetic effects shared across different asthma subtypes and with different diseases contribute to the heterogeneity of asthma.

RESULTS

Multi-ancestry meta-analysis for asthma across 18 biobanks in the GBMI

To identify novel loci associated with asthma, we performed fixed-effects inverse-variance weighted meta-analysis using the harmonized GWAS summary statistics for asthma from 18 biobanks participating in the GBMI (Table S1). The combined sample size from all discovery studies was 153,763 cases and 1,647,022 controls, spanning individuals of European (EUR), African (AFR), admixed American (AMR), East Asian (EAS), Middle Eastern (MID), and Central and South Asian (CSA) ancestry (Figures 1 and S1). The meta-analysis of GWASs from four additional biobanks (9,991 cases and 63,605 controls) was used as an independent replication study (Table S1). Despite the standardized phenotype definitions used by each biobank, which included the asthma PheCode and/or self-reported data (Table S3), the prevalence of asthma varies widely across these biobanks, ranging from 3% in the Taiwan Biobank (TWB) to 24% in the Mass General Brigham Biobank (MGB). We applied pre- and post-GWAS quality control filters that resulted in 70.8 million single-nucleotide polymorphisms (SNPs) for meta-analysis; for downstream analyses we analyzed SNPs present in at least two biobanks.[33] The meta-analysis identified 179 loci of genome-wide significance ($p < 5 \times 10^{-8}$), 49 of which have not been previously reported to be associated with asthma (Figures 2A and S2). These potentially novel loci were defined so that the index variants, or the most significant variants in each locus, were at least 1 Mb in distance from a previously discovered genome-wide significant variant associated with asthma (STAR Methods). Additionally, all but one index variant did not have a previously discovered SNP in linkage disequilibrium (LD) at $r^2 > 0.07$, estimated using a reference panel from individuals in 1000 Genomes[34] (Table S2). In the replication meta-analysis, 51 of the 179 loci had index variants with a p value < 0.05, even though the case numbers in the replication data were less than 10% of the case numbers in the discovery data (Table S2). 154 of the 179 index variants had consistent directions of effect in the discovery and replication meta-analyses. We also found that the potentially novel associations had smaller effect sizes on average compared with the previously reported loci across the allele frequency spectrum (Figure 2B). This illustrates that with the increased power and effective sample size of the GBMI, we were able to uncover SNPs with more modest effects on asthma.
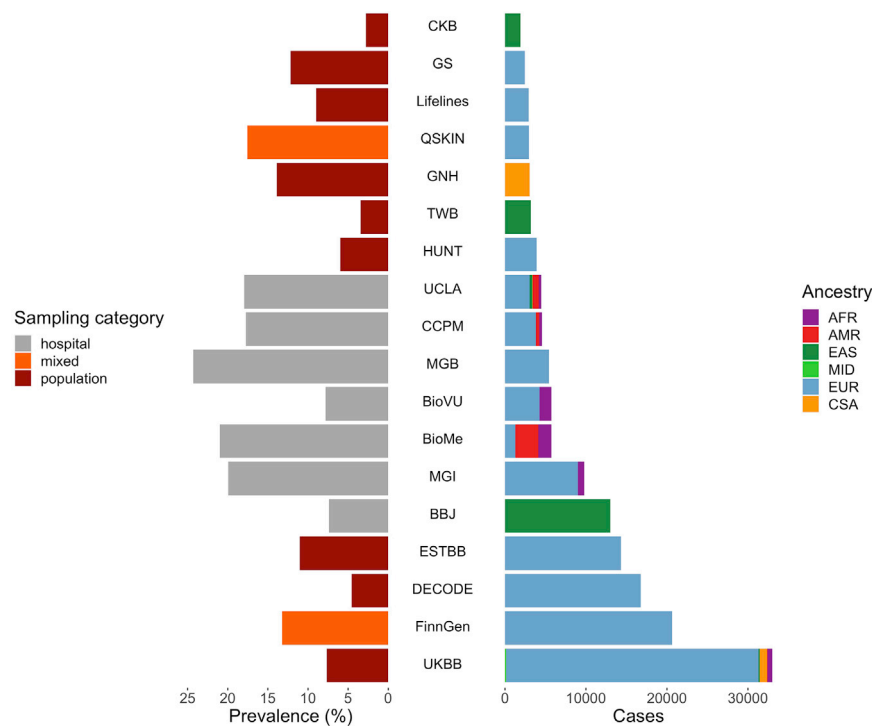
cates *FCGRT*, which regulates immunoglobulin G recycling and is a potential drug target for autoimmune neurological disease therapies.[41] The other previously reported missense variants replicated previous findings; among these, chr4:102267552:C:T (p.Ala391 Thr, $p_{meta-analysis} = 2.5 \times 10^{-12}$) is a highly pleiotropic variant in *SLC39A8* that has been associated with many psychiatric, neurologic, inflammatory, and metabolic diseases[42–48] and has been demonstrated to disrupt manganese homeostasis.[49] Variants implicating well-known asthma-associated genes with large effects, such as *IL1RL1*, *IL2RA*, *STAT6*, *IL33*, *GSDMB*, and *TSLP*, were replicated in the meta-analysis as well.

Because the GBMI meta-analysis includes data from the UK Biobank (UKBB), we compared our results with the TAGC meta-analysis results that did not include the UKBB GWAS to facilitate analyses that require independent samples.[9] Of the index variants within the top 179 loci in the GBMI, 122 were in the TAGC meta-analysis or had a tagging variant in high LD ($r^2 > 0.8$) in the TAGC study; 76 of these had p < 0.05 in the TAGC results. We compared the effect sizes of these 76 SNPs in the GBMI and the TAGC meta-analyses using a previously proposed Deming regression method that considers standard errors in both effect size estimates.[35] We found that all 76 SNPs were directionally consistent and aligned across the studies (Table S4 and Figure S3).

Among the 49 novel loci, the index variants of six loci were missense or in high LD ($r^2 > 0.8$) with a missense variant (Table S2). One of these SNPs, chr10:94279840:G:C ($p_{meta-analysis} = 2.5 \times 10^{-9}$), resides in *PLCE1*, an autosomal recessive nephrotic syndrome gene;[36] a high prevalence of atopic disorders, such as asthma, among children with nephrotic syndrome has long been observed in the clinic, suggesting potential shared pathways underlying asthma and nephrotic syndrome.[37] The asthma risk allele has also been previously linked to lower blood pressure.[38] The index SNPs chr14:100883117:G:T ($p_{meta-analysis} = 2.6 \times 10^{-8}$) and chr19:56222056:C:A ($p_{meta-analysis} = 2.4 \times 10^{-8}$) also implicate novel genes, *RTL1* and *ZSCAN5A*, respectively. *RTL1* has been found to play a role in muscle regeneration,[39] and *ZSCAN5A* has been linked to monocyte count.[40] Additionally, three of the novel index SNPs colocalized with fine-mapped *cis* expression quantitative trait loci (eQTL) (Table S2). One of these, chr19:49513502:C:T ($p_{meta-analysis} = 7.98 \times 10^{-9}$), impli-

GWASs from diverse ancestries reveal shared genetic architecture of asthma and improve power for genetic discovery

Given that sample size, disease prevalence, ancestry, and sampling approaches differed across the 18 discovery biobanks, we investigated the consistency of the asthma-associated loci across the biobanks and their attributes. We first implemented an approach that estimates the correlation (r_b) between the effects of the index variants of the 179 top loci in each biobank GWAS and the corresponding meta-analysis excluding that biobank.[50] We observed that most biobanks have highly correlated genetic effects with other biobanks (r_b estimates close to 1) (Table S5). To further interrogate the consistency of the index variants in all biobanks, we computed the ratio of the effect size of these SNPs in the biobank-specific GWAS over that in the corresponding leave-that-biobank-out meta-analysis. We found that the average per-biobank ratios were almost evenly split between those greater than and less than 1 (Figure S4). This indicates that any significant difference in effects likely does not reflect technical artifacts in the meta-analysis or GWAS procedures. We also applied Deming regression[35] to assess the alignment of the SNP effects in each biobank-specific GWAS with the effects in the corresponding leave-that-biobank-out meta-analysis and observed that the effect sizes were comparable across the biobanks (Figures 3A and S5). Furthermore, the genome-wide genetic correlations between the biobanks with non-zero heritability estimates and the respective leave-that-biobank-out meta-analyses were all close to 1.[33]
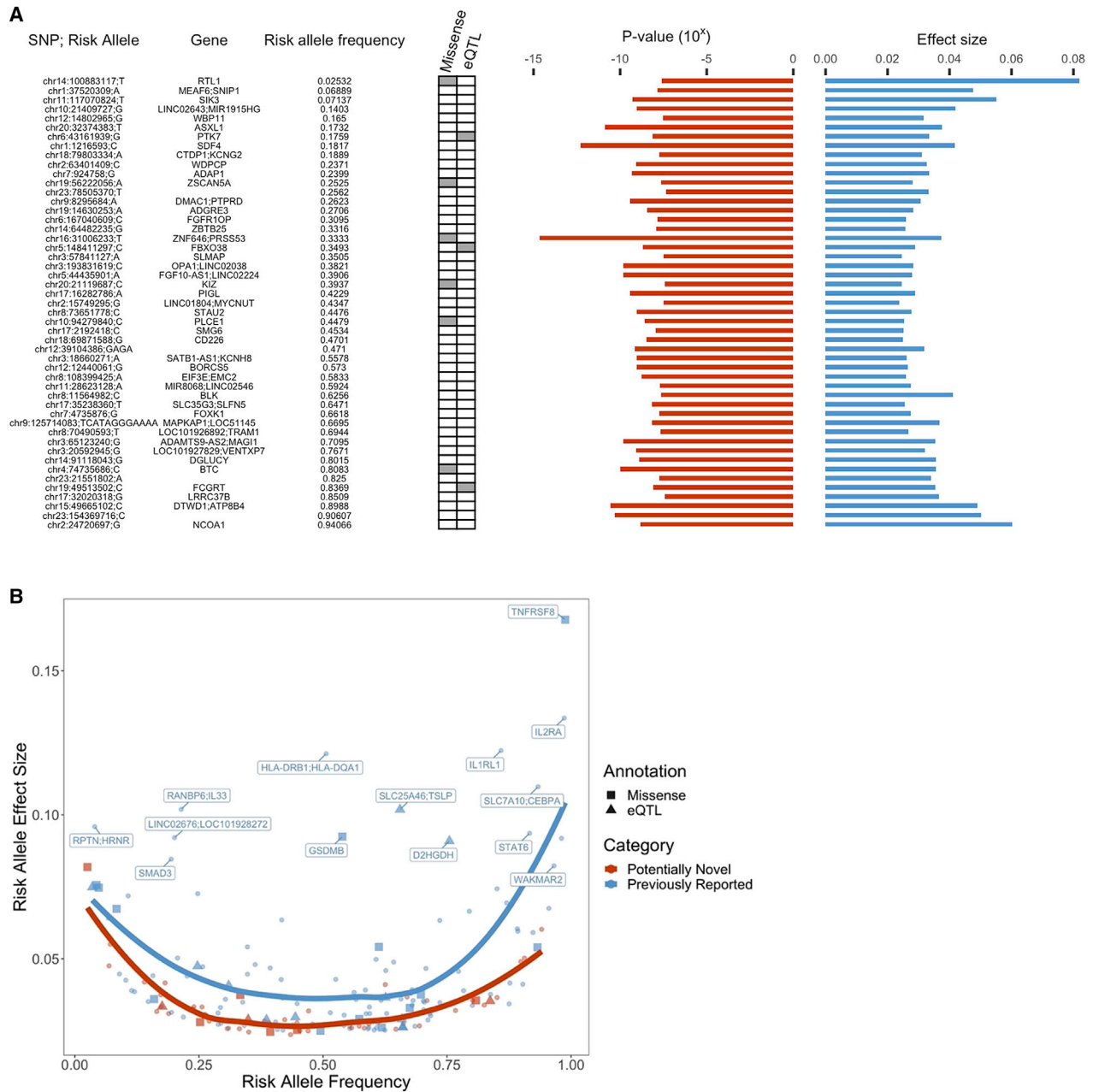
Figure 2. Top loci associated with asthma

(A) Index variants of 49 asthma-associated loci that are potentially novel. Missense variants and *cis*-eQTLs fine-mapped with PIP > 0.9 that overlapped with an index or tagging variant (r² > 0.8) are annotated here. Frequency and meta-analysis effect size estimate of risk-increasing allele , with association p value, are shown on the right.

(B) Frequency and effect size of risk alleles of all 179 index variants. Previously reported genes with large effect sizes are highlighted.

See also Figures S1 and S3; Tables S2 and S4.

To test for potential heterogeneity in effect estimates due to ascertainment, we conducted an additional sensitivity analysis comparing SNP effects in the meta-analyses of the hospital- versus population-based biobanks. We conducted meta-analyses of the nine population-based biobanks (China Kadoorie Biobank [CKB], deCODE Genetics, Estonian Biobank [ESTBB],

East London Genes & Health [GNH], Generation Scotland [GS], Trøndelag Health Study [HUNT], LifeLines, TWB, and UKBB) and six hospital-based biobanks (BioBank Japan [BBJ], Mount Sinai BioMe Biobank [BioMe], BioVU, Mass General Brigham [MGB], Michigan Genomics Initiative [MGI], and ATLAS Community Health Initiative [UCLA]). We then fitted the Deming
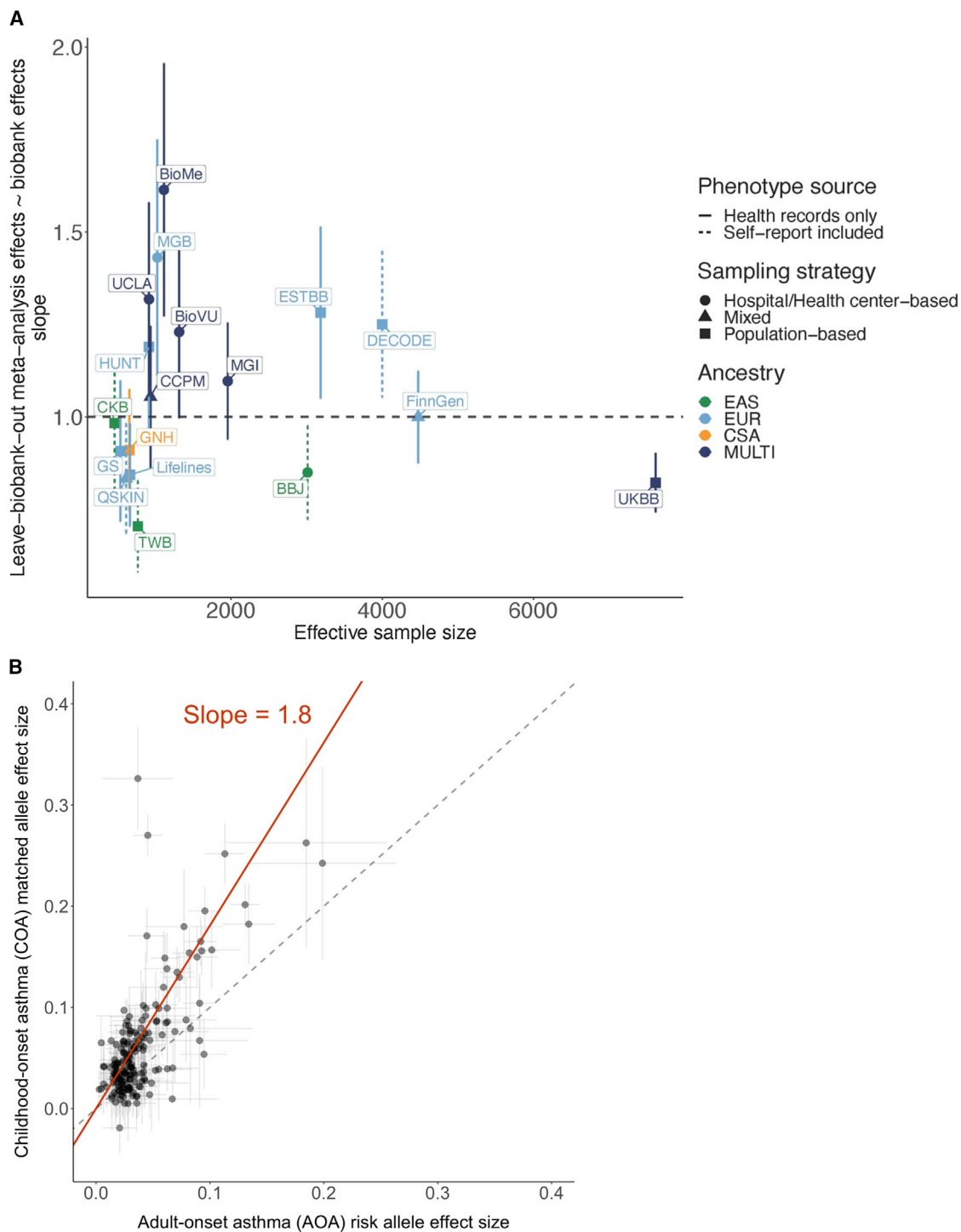
Figure 3. Consistency of loci across biobanks and asthma age-of-onset subtypes

(A) Regression slopes computed using the Deming regression method, which compared effects of index variants in each biobank GWAS against their effects in the corresponding leave-that-biobank-out meta-analysis.[23] The x axis shows the effective sample size of each biobank, computed as 4/(1/cases + 1/controls). Error bars represent 95% confidence intervals of the regression slope estimates.

(B) Effect sizes of the index variants discovered in the all-asthma meta-analysis as estimated in the COA versus AOA meta-analyses compared using the Deming regression method.[35] The intercept was set to be 0; the slope estimated from the regression analysis is reported. Error bars represent 95% confidence intervals of the effect size estimates from the corresponding meta-analysis.

See also Figures S4 and S5; Tables S5 and S11.

regression[35] on the effect size estimates of the loci identified by the all-biobank meta-analysis, using the SNPs with $p < 1 \times 10^{-6}$ in both meta-analyses, and observed high consistency in the effects across the two groups (Figure S6).

Taken together, these analyses indicate that the genetic architecture of asthma is largely shared across cohorts, despite differences in characteristics such as disease prevalence and ascertainment strategy. Furthermore, the consistency of genetic effects across the biobanks suggests that the fixed-effects meta-analysis approach is appropriate for the integration of GWASs from the different datasets. We additionally conducted meta-analysis using the meta-regression approach implemented in MR-MEGA,[51] which accounts for potential effect size heterogeneity across datasets. MR-MEGA identified only two additional loci associated with asthma, one of which is novel (Table S6).

We also found little evidence of heterogeneity in the ancestry-specific effect sizes for the index variants. One SNP, chr10:9010779:G:A, was significantly heterogeneous (p value for Cochran's Q test < 0.0003, the Bonferroni-corrected p value threshold) across the ancestry-specific meta-analyses of AFR, AMR, CSA, EAS, and EUR individuals (Figure 4A and Table S7). One known SNP that nearly reached the Bonferroni-corrected p value threshold for heterogeneity, chr16:27344041:G:A, displayed different effects in the EUR and EAS cohorts. This SNP lies within an intron of *IL4R* (Figure 4B), which has known associations with asthma.[6,52] Previous studies have investigated the association of *IL4R* with asthma in different populations, with inconsistent results, so future studies on the potential population-specific effects of this gene will be needed.[53–55] Our findings demonstrate that despite broad consistency of effect sizes across ancestries among the top loci, the increased power and diversity of the GBMI enabled the detection of SNPs with significantly different effects across ancestries.

Additionally, the greater diversity of the GBMI facilitated the discovery of loci that would not have been identified in association analyses using data from only EUR ancestry cohorts. We found that of the 179 loci identified in the all-biobank meta-analysis, 49 did not reach genome-wide significance in the EUR-only meta-analysis (Table S8). This additional yield of loci may be partially due to the increase in sample size, but the inclusion of GWASs from diverse ancestries also enabled the identification of loci that are more frequent in some non-EUR populations. 19 of these 49 loci were potentially novel, and 13 of these novel loci had an index variant higher in frequency in a non-EUR ancestry group compared with the EUR ancestry group. The consistent effect estimates of the 49 additional variants across populations (45/49 had a p value for Cochran's Q test across ancestries > 0.02) indicate that the additional variants discovered with the incorporation of GWASs from diverse ancestries do not tend to be population-specific loci that only have effects in certain populations. However, owing to differences in frequency across populations, it is essential to conduct asthma GWASs in different populations to uncover the full spectrum of asthma-associated loci.

Meta-analysis across diverse ancestries improves asthma PRS accuracy

We next explored the impact of the increased sample sizes and diversity in the GBMI on genome-wide risk prediction of asthma.

To establish a baseline understanding of PRS performance for asthma as well as other disease endpoints in the GBMI, Wang et al.[56] evaluated and compared the prediction accuracy of PRS derived from the pruning and thresholding (P + T) method and PRS with continuous shrinkage (PRS-CS)[57] in target cohorts of EUR, CSA, EAS, and AFR ancestries, using the leave-one-biobank-out meta-analyses as discovery data. This study observed improvements in prediction accuracy for asthma using PRS-CS across all target cohorts (Figure S7), and additionally the PRS derived from the GBMI leave-one-biobank-out meta-analyses of asthma had higher predictive accuracy, as measured by R^2 on the liability scale $R^2_{liability}$, compared with the PRS constructed from the TAGC meta-analysis[9] (Figure 5).

To expand on these analyses, we tested a recently developed extension of PRS-CS, PRS-CSx,[58] for asthma risk prediction. This method jointly models multiple summary statistics from different ancestries to enable more accurate effect size estimation for prediction. For input to PRS-CSx, we used the AFR, AMR, EAS, CSA, and EUR ancestry-specific meta-analyses from the GBMI; the discovery meta-analysis that matched the ancestry of the target cohort excluded the target cohort (Figure S8). With the posterior SNP effect size estimates from PRS-CSx, we tested the multi-ancestry PRS in the following target populations: AFR ancestry individuals in UKBB, CSA ancestry individuals in UKBB, a holdout set of EAS ancestry individuals in BBJ, and a holdout set of EUR ancestry individuals in UKBB. The final prediction models tested in these target populations were the optimal linear combinations of the population-specific PRS. The average $R^2_{liability}$ in the EAS (0.053) and EUR (0.054) target cohorts approached the SNP-based heritability (h^2_{SNP}), estimated to be 0.085 for asthma using the all-biobank meta-analysis,[56] while the prediction accuracies in the CSA (0.038) and AFR (0.014) target cohorts were lower (Figure 5 and Table S9). When we downsampled the EUR target cohort to 1,000 individuals, to match the sample size of the EAS target cohort, we found a higher average $R^2_{liability}$ (0.063) but, as expected, much larger confidence intervals (Figure S9). Estimates may differ across biobanks and ancestries given differences in disease prevalence, environmental exposures, phenotype definitions, and other factors, and these differences may in turn contribute to the PRS in EAS individuals performing similarly to PRS in EUR individuals in our analyses, despite the smaller sample size of the EAS discovery cohort. The $R^2_{liability}$ across the target populations for the PRS-CSx scores were roughly the same as the $R^2_{liability}$ of the PRS derived from the PRS-CS analyses. It is important to note that the discovery data used in the PRS-CS analyses differed slightly in sample size and composition, since the leave-one-biobank-out approach was used for PRS-CS, but the target cohorts in which the PRS were evaluated were the same (Table S10).

To investigate why improvement in performance using PRS-CSx was only incremental in most of the target cohorts, we examined the performances of each population-specific PRS. We found that across all target cohorts, PRS derived from either the EUR or EAS set of posterior effect size estimates outperformed the linear combination, and the $R^2_{liability}$ values of these PRS were also higher compared to that of the PRS-CS scores (Figure S10 and Table S9). This suggests that the addition of
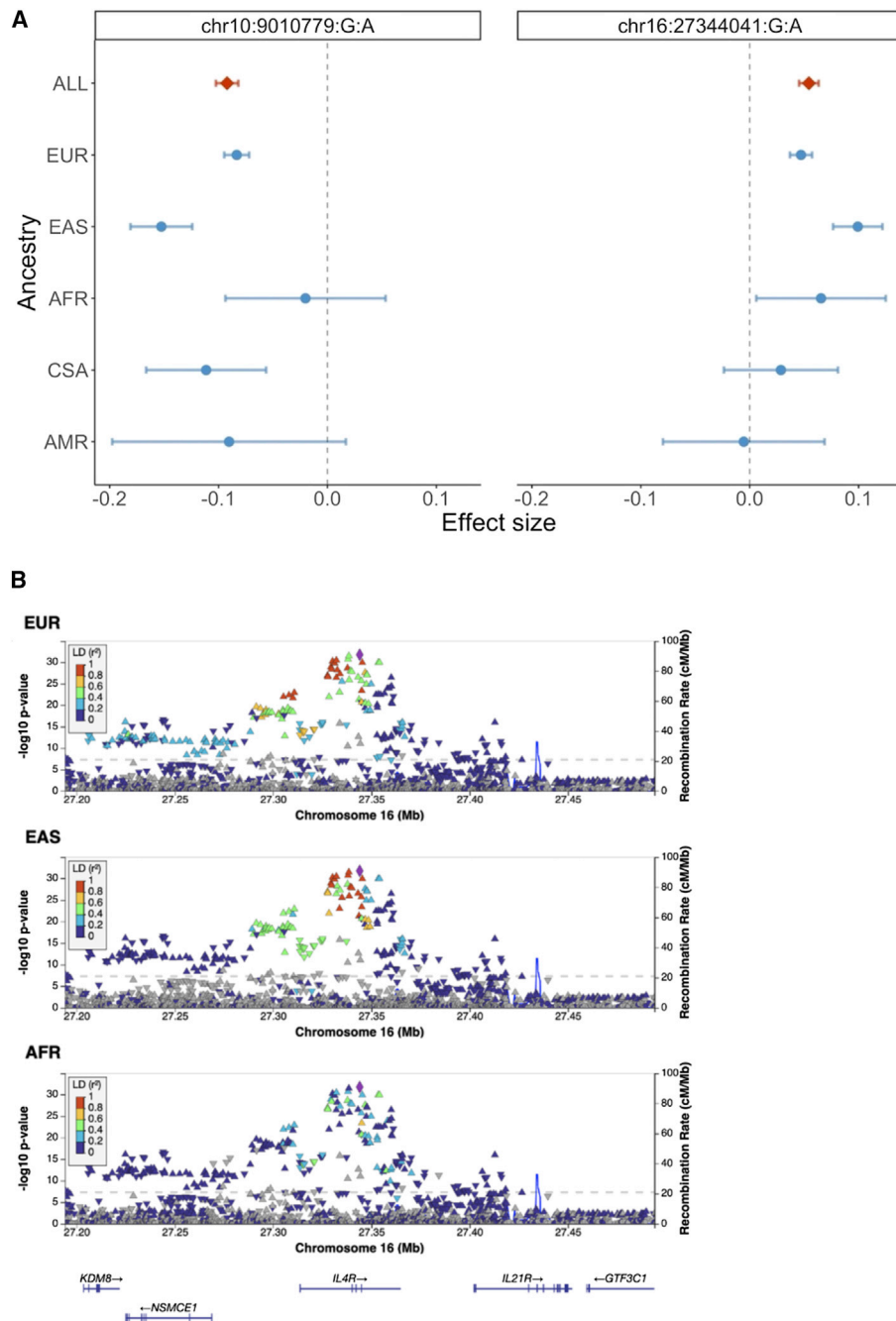
Figure 4. Loci showing heterogeneity in ancestry-specific effect sizes

(A) Index variants with the most significant $p_{Cochran's\ Q}$. Effect sizes of these variants in each ancestry-specific meta-analysis are shown. Error bars represent 95% confidence intervals of effect size estimates.

(B) LocusZoom plots showing the association with asthma of chr16:27344041:G:A (purple symbol) and variants within 150 kb upstream and downstream. Color coding of other SNPs indicates LD with this SNP. EUR, EAS, and AFR indicate the population from which LD information was estimated.

See also Table S7.

more discovery GWASs to PRS-CSx can improve the accuracy of PRS based on a single set of posterior effect size estimates, but the linear combination of PRS from multiple GWASs does not necessarily yield higher accuracy. This may be due to the considerably smaller sample sizes of some of the input discovery meta-analyses in our analyses and, thus, varying signal-to-noise ratios. Collectively, these analyses show that the increase in scale and diversity of discovery GWASs for PRS is the primary driver of increased PRS accuracy in non-EUR populations for asthma, with marginal gains using PRS-CSx over PRS-CS. For
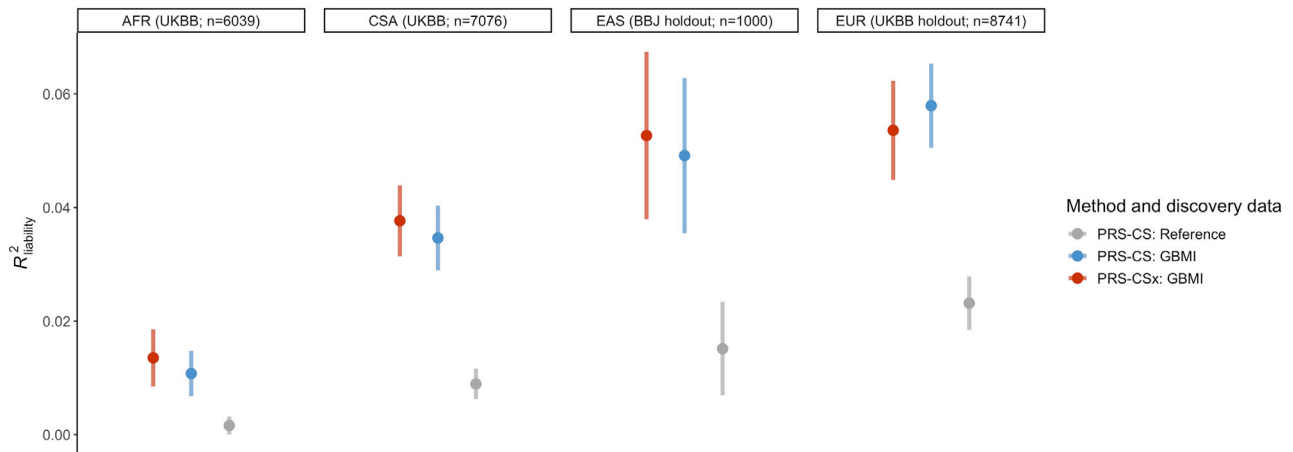
Figure 5. PRS performance across ancestries

Each panel represents a target cohort in which PRS constructed using PRS-CSx and PRS-CS were evaluated. The reference dataset was the TAGC meta-analysis.[5] Sample sizes for the target cohorts are: cases = 849 and controls = 5,190 for AFR; cases = 500 and controls = 500 for EAS; cases = 1,164 and controls = 7,577 for EUR; cases = 1,232 and controls = 6,744 for CSA. Error bars represent standard deviation of R^2 on the liability scale across 100 replicates. See also Figure S10; Tables S9 and S10.

EUR target cohorts, a multi-ancestry PRS construction method such as PRS-CSx does not seem to contribute much improvement in prediction accuracy, likely due to the predominating sample size of EUR discovery GWASs as well as the inclusion of GWASs from smaller, non-EUR discovery cohorts, which may introduce more noise than signal.

Childhood-onset and adult-onset asthma are highly genetically correlated

To increase power for genetic discovery, we used a broad phenotype definition for asthma (STAR Methods) but, given the heterogeneity of the disease, we sought to address the extent to which this meta-analysis captured the genetic architectures of two common subtypes of asthma, childhood-onset asthma (COA) and adult-onset asthma (AOA). We conducted asthma age-of-onset subtype analyses in two of the participating GBMI biobanks for which age at asthma diagnosis information were accessible, UKBB and FinnGen. Using a cutoff age of 19 years at asthma diagnosis to define the subtypes (STAR Methods), we performed GWASs of COA and AOA in FinnGen and the EUR ancestry cohort in UKBB, as well as fixed-effects, inverse-variance weighted meta-analyses of the COA (20,964 cases and 674,014 controls) and AOA (56,744 cases and 674,014 controls) GWASs, respectively. Applying LD score correlation (LDSC),[59] we observed strong genetic correlations between each COA GWAS and the respective leave-that-biobank-out meta-analysis of all other biobanks utilizing the broad phenotype definition (r_g [SE] = 0.73 [0.03], p = 4.70 × 10^{-132} for UKBB and r_g [SE] = 0.80 [0.4], p = 3.19 × 10^{-73} for FinnGen), and even larger genetic correlations between each AOA GWAS and leave-that-biobank-out meta-analysis (r_g [SE] = 0.90 [0.04], p = 1.71 × 10^{-127} for UKBB and r_g [SE] = 0.90 [0.30], p = 1.39 × 10^{-237} for FinnGen). The genetic correlation between the COA and AOA meta-analyses was similarly high (r_g [SE] = 0.78 [0.30], p = 1.32 × 10^{-116}) and similar to the genetic correlation (r_g [SE] = 0.67 [0.02]) reported by a previous study of

asthma age-of-onset subtypes.[60] We also observed substantial overlap between the top loci identified in each subtype meta-analysis and the all-asthma meta-analysis. 75 of the 90 loci (83%) of genome-wide significance (p < 5 × 10^{-8}) and 55 of the 69 loci (80%) identified by the COA and AOA meta-analysis, respectively, overlapped with a locus discovered in the all-asthma meta-analysis (Table S11). Overall, these results suggest that much of the genetic architecture between COA and AOA is shared, as is consistent with previous findings.[60,61] Despite the GBMI meta-analysis drawing from primarily adult cohorts, many of the genetic variants identified contribute to both subtypes.

To investigate whether the genetic effects of the index variants of the asthma-associated loci differ across the subtypes, we compared the estimated effect sizes of the 179 index variants discovered in the all-asthma meta-analysis in the COA and AOA meta-analyses using the Deming regression method. We found that these variants had systematically stronger effects in the COA meta-analysis compared with the AOA meta-analysis (Figure 3B), supporting previous findings that the etiology of COA is likely partially characterized by genes that have smaller (or no) effects on AOA.[60,61]

Asthma and COPD have shared and distinct biological processes

The shared genetic factors between asthma and different diseases that often coexist with asthma, such as COPD, a late-onset respiratory disease, have also been used to investigate and characterize asthma heterogeneity. It has been well documented in the literature that asthma and COPD are frequent co-morbidities of each other,[62] but only a few studies thus far have investigated the extent to which this is driven by a shared genetic basis.[63–65] Utilizing the GBMI meta-analyses of asthma and COPD, we observed a strong genetic correlation between asthma and COPD (r_g [SE] = 0.67 [0.021], p = 1.55 × 10^{-226}). This genetic correlation estimate is higher than estimates from

previous studies, which ranged from 0.38 to 0.42.[64,65] This may be a result of the discovery datasets used by these studies, which were enriched for pediatric asthma cohorts, whereas the GBMI biobanks are primarily composed of adult participants. To more formally test for potential differences in the shared genetic architecture of age-of-onset subtypes and COPD, we computed genetic correlations between the COA and AOA meta-analyses and the GBMI COPD meta-analysis. We found that the AOA meta-analysis had a strong genetic correlation with the COPD meta-analysis (r_g [SE] = 0.60 [0.3], p = 2.65×10^{-94}), while the COA meta-analysis had a more moderate genetic correlation with the COPD meta-analysis (r_g [SE] = 0.33 [0.3], p = 7.60×10^{-31}).

To further evaluate the extent of genetic overlap between asthma and COPD, we applied a gene and gene-set analysis tool, Multi-marker Analysis of GenoMic Annotation (MAGMA),[66] to the GBMI EUR, AFR, EAS, and CSA meta-analyses of asthma as well as the GBMI EUR, AFR, and EAS meta-analyses of COPD. After Bonferroni correction, we found that 442, 149, and 6 genes were significantly associated with asthma in the EUR (p < 2.50×10^{-6}), EAS (p < 2.50×10^{-6}), and CSA (p < 2.52×10^{-6}) populations, respectively, with no significantly associated genes in the AFR cohort (all p > 2.51×10^{-6}) (Table S14). The majority of the genes associated with asthma identified in the EAS meta-analysis overlapped with the genes from the EUR meta-analysis (126 out of 149 genes), and all six genes associated with asthma as identified in the CSA meta-analysis were also significantly associated in the EUR and EAS meta-analyses. We identified 46 and 33 genes significantly associated with COPD in the EUR (p < 2.50×10^{-6}) and EAS (p < 2.50×10^{-6}) cohorts, respectively, and, similarly to asthma, no significantly associated genes from the AFR meta-analysis (all p > 2.51×10^{-6}) (Table S15). Of the 75 genes associated with COPD across the EUR and EAS meta-analyses, 24 overlapped with the asthma-associated genes. We also conducted gene prioritization using Data-driven Expression-Prioritized Integration for Complex Traits (DEPICT)[67] and gene-level Polygenic Priority Score (PoPS).[68] However, only 3 of the 52 genes (6%) prioritized for COPD by DEPICT overlapped with a gene prioritized for asthma using the same method (Table S16), and 17 of the 184 genes (9%) prioritized for COPD by PoPS overlapped with a prioritized gene for asthma (Table S17). Across the shared COPD and asthma genes prioritized by each method, only one gene, *MED24*, was prioritized by more than one method, highlighting that existing gene prioritization methods have poor agreement, an observation that has been previously discussed[68] and is explored in more detail in Zhou et al.[33]

We also adopted MAGMA for gene set enrichment based on the curated and ontology gene sets from the Molecular Signatures Database (MSigDB).[69] We found hundreds of gene sets that were significantly enriched (false discovery rate [FDR] <0.05) by the asthma-associated genes discovered in the EUR and EAS meta-analyses (Table S18). In contrast, only a handful of gene sets were significantly enriched among COPD-associated genes discovered in the AFR meta-analysis, likely reflecting the smaller overall sample size of the COPD meta-analysis (Table S19). The top-ranked asthma pathways from the EUR meta-analysis included cytokine and interleukin signaling and

T cell activation. Consistently biologically, the EAS meta-analysis identified autoimmune thyroid disease and graft-versus-host disease pathways. The top-ranked COPD pathways from the EUR meta-analysis, although not significant, included several pathways related to nicotine receptor activity. These results reinforce the notion that despite the substantial genetic overlap, asthma and COPD are governed by distinct biological processes as well. Future investigations will be required to fully parse out the etiology and comorbidities of asthma, like COPD, that develop later on in adulthood.

Genetic overlap between asthma and other diseases

Non-genetic epidemiological studies have also identified correlations between asthma and many other disease categories beyond COPD.[70–72] More recently, some genome-wide cross-trait studies have found evidence for shared genetic architectures between asthma and other allergic diseases,[21,73] neuropsychiatric disorders,[22] and obesity,[20] suggesting that a comprehensive characterization of the shared genetics among asthma and other complex diseases and traits could provide insights into the variable pathology of asthma.[19] Together, these findings motivated us to assess whether correlations across a broad spectrum of disease endpoints are potentially driven by a shared genetic basis or are purely observational and not driven by a shared biology. Since the GBMI project was limited to 14 disease endpoints, we utilized the wide range of phenotypic data available in UKBB to measure correlations between asthma and additional diseases and traits. Applying LDSC to the UKBB EUR GWAS of 1,008 significantly heritable (heritability Z score > 4) phenotypes and the GBMI leave-UKBB-out meta-analysis of asthma, we obtained pairwise genetic correlation estimates between these phenotypes and asthma. We observed strong correlations ($|r_g| > 0.4$) with 95 of these phenotypes, which spanned prescriptions, PheCodes, and other categories (Table S12). Digestive system disorders, including gastritis and gastroesophageal reflux disease, emerged as a disease category with significant and strong genetic correlations with asthma. Although the association between asthma and digestive disorders has not been as well studied, this does reinforce a previous finding of shared genetics between asthma and diseases of the digestive system,[9] indicating that the commonly observed copresentation of asthma and gastroesophageal disease in the clinic may be partially due to pleiotropic genetic effects. Our results also showed moderate and significant correlations (r_g ranging from 0.2 to 0.3) between asthma and neuropsychiatric diseases, such as anxiety and depression, and obesity-related traits, such as body mass index, which is similarly consistent with previous findings.[20,22]

Leveraging data from another biobank, BBJ, we computed genetic correlation estimates between the GBMI leave-BBJ-out meta-analysis of asthma and 19 significantly heritable disease endpoints in BBJ (Table S13). COPD showed the strongest and most significant correlation with asthma (r_g = 0.29, p = 6.41×10^{-6}), but the notably lower estimate compared with the estimate from the UKBB correlation analyses may be due to differences in phenotype definition and curation. Pollinosis, also known as allergic rhinitis or hay fever, showed moderate correlation with asthma (r_g = 0.28, p = 0.0004), consistent

with the correlation results from UKBB (r_g = 0.39, p = 4.60×10^{-3}). Comparing the phenotypes with significant SNP heritability estimates in both BBJ and UKBB (Figure S11), we found that only COPD had significant genetic correlations with asthma across the biobanks. The rheumatoid arthritis (RA) and type 2 diabetes (T2D) GWASs from UKBB had moderate and significant correlations with asthma, which were partially recapitulated in the BBJ results that showed a moderate but not significant correlation between the BBJ GWASs of RA and asthma, and a small but significant correlation between the BBJ GWAS of T2D and the GBMI leave-BBJ-out meta-analysis of asthma. Several studies in the literature have reported a relationship between risk for RA and asthma[74–79] as well as T2D and asthma,[80–82] but more genetic studies in different populations and biobanks are needed to investigate the potential shared genetic architecture of these diseases. Importantly, causal relationships between asthma and genetically correlated phenotypes are not yet well understood, and other methods such as Mendelian randomization could be applied to identify potential causal associations.[83]

DISCUSSION

Assembling a large and diverse collection of asthma cohorts from around the world, we conducted a GWAS meta-analysis of 18 biobanks as well as a replication meta-analysis of 4 additional biobanks, and identified 49 novel associations among a total of 179. Despite the substantial sample sizes of previous meta-analyses of asthma,[9] our results indicated that the heterogeneity and complexity of asthma, like other common polygenic diseases, will benefit from even larger sample sizes for genomic discovery. We interrogated the overall consistency of genetic effects across the cohorts and found that despite variability in recruitment, continent, sampling strategy, health system design, and disease prevalence, the effects of the loci discovered in the meta-analysis were by and large concordant across the biobanks. Additionally, genetic correlation estimates across ancestries, which ranged from 0.65 to 0.99 for the well-powered ancestry groups, strongly supported the finding that the genetic architecture of asthma is largely shared across the ancestry groups studied.

Importantly, however, the addition of GWASs from more diverse populations aided the discovery of genetic loci with higher frequencies in non-EUR populations that did not reach genome-wide significance in the meta-analysis with only EUR cohorts, highlighting the importance of diversifying genomic studies of asthma. Given the current disproportionate representation of European ancestries, we expect that as the availability of non-EUR GWASs of asthma and other asthma-related diseases and traits continues to increase, it is likely that greater numbers of such variants associated with asthma will be discovered. Previous studies of asthma-related diseases, such as atopic dermatitis, in non-EUR populations have similarly identified additional risk variants that are higher in frequency in other populations but have also found highly shared polygenic architecture between populations, mirroring our findings for asthma.[84,85] This study also provides further evidence for substantial genetic overlap between childhood-onset and

adult-onset asthma, as well as between asthma and well-known immune-related comorbidities such as COPD and allergic diseases. Additionally, we identified genetic correlations between asthma and less well-studied comorbidities such as digestive system disorders while highlighting additional complexity in the etiology and comorbidities of asthma. For example, gene-set enrichment analyses using MAGMA did not yield many shared pathways for asthma and COPD despite the strong genetic correlation.

We also demonstrated that the greater diversity of the GBMI improved polygenic prediction in asthma, particularly for populations of non-European ancestry. Previous studies on asthma PRS in the literature have primarily focused on using PRS to predict asthma in pediatric cohorts, and overall found limited performance of PRS.[28–30,86] Most of these studies used the P + T approach, while a recently published paper by Namjou et al.[32] applied PRS-CS to the TAGC multi-ancestry GWAS and found improved discriminatory power of their PRS (receiver-operating characteristic area under the curve [AUC] of 0.66–0.70 across two pediatric cohorts) compared with the prior studies that used P + T. Sordillo et al.[31] applied another genome-wide approach, lassosum, to the TAGC data, but their PRS evaluated in adult cohorts showed moderate performance (AUC of 0.51–0.57 across cohorts of different ancestries). While we did not assess the lassosum method, we have shown that the greater sample size and diversity of the GBMI compared with TAGC contribute to better-performing PRS (Figure 5). However, we also found that differences in prediction power between the Bayesian PRS construction methods PRS-CSx and PRS-CS were minimal. This may be due to imbalances in the sample sizes of the discovery cohorts, which may need to be taken into careful consideration when using these methods. Previous studies have found that imbalanced sample sizes across ancestries contribute somewhat unpredictably to varying prediction performances, with a low signal-to-noise ratio in ancestry-matched target populations reducing prediction performance.[87] Therefore, further investigation is needed to fully understand the interplay between sample size and ancestry in the context of polygenic prediction. Ultimately, these analyses highlight the pressing need for more well-powered and ancestrally diverse resources that will help reduce these imbalances.

This study and, importantly, the data sharing across biobanks facilitated by this initiative, have laid the groundwork for deeper dives into the shared and distinct genetic signatures of asthma subtypes. We were able to stratify two participating biobanks, UKBB and FinnGen, into COA and AOA based on the participants' ages at first diagnosis. While we found that the GBMI asthma meta-analysis of all biobanks containing both subtypes identified many of the loci contributing to these subtypes, the age-of-onset-stratified meta-analyses uncovered additional subtype-specific loci. Of the top loci associated with COA and AOA, 11 and 12 loci, respectively, (1) did not overlap with a top locus in the other subgroup meta-analysis and (2) were evaluated in the all-asthma GBMI meta-analysis (i.e., in more than three GBMI biobanks) but did not reach genome-wide significance in the meta-analysis (Table S11). Because of the limited availability of information on age at first diagnosis across the biobanks, we were not able to explore age-dependent associations further,

but with sufficient scale it is likely that more of the distinct genetic architectures of COA and AOA will be uncovered.

Limitations of the study

There are several limitations of this study that should be taken into consideration. We have highlighted the harmonization of the phenotype definitions across biobanks, but it is important to acknowledge that the criteria used, which allowed for both self-reported and PheCode information, are vulnerable to imprecision and variability in the data collected. Self-reported data for asthma is particularly susceptible to imprecision because it relies on personal recollection of asthma diagnoses that are often given in childhood. On the other hand, PheCodes, which are based on ICD codes, may fail to capture diagnoses made earlier in the lifetime of individuals in hospital-based cohorts. Therefore, including both self-reported and PheCode data—an approach adopted by some but not all biobanks—may be optimal for association analyses for asthma. We were limited in our ability to evaluate the effects of phenotype definition on effect-size estimation, since only three biobanks used self-reported data and two of these three biobanks (TWB and BBJ) only have participants of EAS ancestry. However, we compared the asthma GWAS derived from self-reported versus PheCode data in UKBB and found high genetic correlation (r_g [SE] = 0.95 [0.01]) between the GWASs. This provides some evidence that variation in phenotype definition may not significantly influence genetic discovery, but we cannot confirm the same pattern for all biobanks in the GBMI and especially for other diseases. However, given the relative alignment of genetic effects across the biobanks, we would expect that minor differences in phenotype definition would not substantially change the association results for asthma.

Additionally, we acknowledge that since the definitions used here for asthma and COPD do not exclude individuals with concurrent diagnoses, we are not able to fully distinguish the distinct biological pathways affecting asthma and COPD. Comorbidity rates of asthma and COPD reported in the literature range across studies but population-based estimates generally are low, around 2%–3%,[88,89] while hospital-based prevalence estimates tend to be higher, around 13%.[90] Among biobanks participating in the GBMI, for example, 15.5% of all individuals with asthma in UKBB have a concurrent COPD diagnosis, 21% in BioVU, and 7.4% in BBJ. A previous study found that using stricter definitions of asthma, such as excluding subjects with COPD, resulted in stronger association signals for some of the asthma-associated loci.[7] However, it is important to note that if we excluded participants with a COPD diagnosis, we would not have a fully representative sample of the participants with asthma in the GBMI . As has been documented in other studies,[91,92] this could induce selection bias or collider bias, which could lead to biased genetic associations. Most of the previous genetic studies of asthma in the literature did not exclude individuals with COPD from analyses. However, in the COA and AOA analyses, we did exclude participants with a COPD diagnosis to avoid confounding from potential misclassifications of AOA and COPD. We also note that estimates of genetic correlation by LDSC are not biased by sample overlap.[59] In fact, this has been explored in the context of asthma and allergic diseases, where r_g estimates from LDSC were shown to be robust to overlapping cases and/or controls.[21]

We also recognize the importance of analyzing environmental factors in conjunction with genetic factors for a disease that is heavily influenced by the environment. Our genetic analyses offer insight into the potential shared biological pathways that may be differentially affected by non-genetic factors, but we were not able to explicitly investigate environmental effects given the lack of available environmental exposure data among the biobanks. The high degree of alignment among genetic associations, coupled with the large variability in asthma prevalence, points to a particularly important role of the environment for asthma risk across populations. Gaining a greater understanding of the specific non-genetic factors that contribute to asthma development in different environments may help guide more accurate disease prediction across populations.

Despite these limitations, examples from this study demonstrate that with broader sharing of more extensive phenotype data, biobanks are well positioned to not only facilitate general locus discovery but also advance the study of disease subtypes and comorbidities. The inclusion of individuals of diverse ancestries at a continuously increasing scale will accelerate novel variant and gene discovery. This will more quickly expand the set of genetic findings from which biological inference can be drawn, as well as ensure that predictive models derived from genetic risk factors will be as accurate and informative for individuals of all ancestries and geographical locations as possible.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - ○ Lead contact
  - ○ Materials availability
  - ○ Data and code availability
- METHODS DETAILS
  - ○ Asthma phenotype definitions
  - ○ Principal components (PC) projection
  - ○ Meta-analysis
  - ○ Index variant and locus definitions
  - ○ Index SNP effects across biobanks
  - ○ Ancestry-specific heterogeneity
  - ○ Polygenic risk scores
  - ○ Age-of-onset subtype GWAS and meta-analyses
  - ○ Genetic correlation in UKBB and BBJ
  - ○ Gene- and pathway-based enrichment
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.xgen.2022.100212.

AUTHOR CONTRIBUTIONS

Study design, K.T., A.R.M., M.J.D., and B.M.N.; data collection/contribution, W.Z., Y.O., and T.M.; data analysis, K.T., W.Z., Y.W., M.K., S.N., R.G., L.M., and L.N.N.; writing, K.T., A.R.M., S.N., and R.G.; revision, K.T., A.R.M., Y.W., R.G., and M.J.D.

DECLARATION OF INTERESTS

M.J.D. is a founder of Maze Therapeutics. B.M.N. is a member of the scientific advisory board at Deep Genomics and consultant for Camp4 Therapeutics, Takeda Pharmaceutical, and Biogen.

REFERENCES

1. Thomsen, S.F., van der Sluis, S., Kyvik, K.O., Skytthe, A., and Backer, V. (2010). Estimates of asthma heritability in a large twin sample. Clin. Exp. Allergy *40*, 1054–1061. https://doi.org/10.1111/j.1365-2222.2010.03525.x.

2. Hernandez-Pacheco, N., Pino-Yanes, M., and Flores, C. (2019). Genomic predictors of asthma phenotypes and treatment response. Front. Pediatr. *7*, 6. https://doi.org/10.3389/fped.2019.00006.

3. Moffatt, M.F., Kabesch, M., Liang, L., Dixon, A.L., Strachan, D., Heath, S., Depner, M., von Berg, A., Bufe, A., Rietschel, E., et al. (2007). Genetic variants regulating ORMDL3 expression contribute to the risk of childhood asthma. Nature *448*, 470–473. https://doi.org/10.1038/nature06014.

4. Moffatt, M.F., Gut, I.G., Demenais, F., Strachan, D.P., Bouzigon, E., Heath, S., von Mutius, E., Farrall, M., Lathrop, M., and Cookson, W.O.C.M.; GABRIEL Consortium (2010). A large-scale, consortium-based genomewide association study of asthma. N. Engl. J. Med. *363*, 1211–1221. https://doi.org/10.1056/NEJMoa0906312.

5. Ferreira, M.A., Matheson, M.C., Duffy, D.L., Marks, G.B., Hui, J., Le Souëf, P., Danoy, P., Baltic, S., Nyholt, D.R., Jenkins, M., et al. (2011). Identification of IL6R and chromosome 11q13.5 as risk loci for asthma. Lancet *378*, 1006–1014. https://doi.org/10.1016/S0140-6736(11)60874-X.

6. El-Husseini, Z.W., Gosens, R., Dekker, F., and Koppelman, G.H. (2020). The genetics of asthma and the promise of genomics-guided drug target discovery. Lancet Respir. Med. *8*, 1045–1056. https://doi.org/10.1016/S2213-2600(20)30363-5.

7. Han, Y., Jia, Q., Jahani, P.S., Hurrell, B.P., Pan, C., Huang, P., Gukasyan, J., Woodward, N.C., Eskin, E., Gilliland, F.D., et al. (2020). Genome-wide analysis highlights contribution of immune system pathways to the genetic architecture of asthma. Nat. Commun. *11*, 1776. https://doi.org/10.1038/s41467-020-15649-3.

8. Torgerson, D.G., Chiu, G.Y., Gauderman, W.J., Gignoux, C.R., Graves, P.E., Himes, B.E., Levin, A.M., Mathias, R.A., Hancock, D.B., Baurley, J.W., et al. (2011). Meta-analysis of genome-wide association studies of asthma in ethnically diverse North American populations. Nat. Genet. *43*, 887–892. https://doi.org/10.1038/ng.888.

9. Demenais, F., Margaritte-Jeannin, P., Barnes, K.C., Cookson, W.O.C., Altmüller, J., Ang, W., Barr, R.G., Beaty, T.H., Becker, A.B., Beilby, J., et al. (2018). Multiancestry association study identifies new asthma risk loci that colocalize with immune-cell enhancer marks. Nat. Genet. *50*, 42–53. https://doi.org/10.1038/s41588-017-0014-7.

10. Sembajwe, G., Cifuentes, M., Tak, S.W., Kriebel, D., Gore, R., and Punnett, L. (2010). National income, self-reported wheezing and asthma diagnosis from the World Health Survey. Eur. Respir. J. *35*, 279–286. https://doi.org/10.1183/09031936.00027509.

11. Asher, M.I., García-Marcos, L., Pearce, N.E., and Strachan, D.P. (2020). Trends in worldwide asthma prevalence. Eur. Respir. J. *56*, 2002094. https://doi.org/10.1183/13993003.02094-2020.

12. Akinbami, L.J., Moorman, J.E., Bailey, C., Zahran, H.S., King, M., Johnson, C.A., and Liu, X. (2012). Trends in asthma prevalence, health care use, and mortality in the United States, 2001-2010. NCHS Data Brief, 1–8.

13. Dharmage, S.C., Perret, J.L., and Custovic, A. (2019). Epidemiology of asthma in children and adults. Front. Pediatr. *7*, 246. https://doi.org/10.3389/fped.2019.00246.

14. Borish, L., and Culp, J.A. (2008). Asthma: a syndrome composed of heterogeneous diseases. Ann. Allergy Asthma Immunol. *101*, 1–8, quiz 8–11. https://doi.org/10.1016/S1081-1206(10)60826-5.

15. Kuruvilla, M.E., Lee, F.E.-H., and Lee, G.B. (2019). Understanding asthma phenotypes, endotypes, and mechanisms of disease. Clin. Rev. Allergy Immunol. *56*, 219–233. https://doi.org/10.1007/s12016-018-8712-1.

16. Maselli, D.J., and Hanania, N.A. (2018). Asthma COPD overlap: impact of associated comorbidities. Pulm. Pharmacol. Ther. *52*, 27–31. https://doi.org/10.1016/j.pupt.2018.08.006.

17. Postma, D.S., and Rabe, K.F. (2015). The asthma–COPD overlap syndrome. N. Engl. J. Med. *373*, 1241–1249. https://doi.org/10.1056/NEJMra1411863.

18. Ferreira, M.A.R., Matheson, M.C., Tang, C.S., Granell, R., Ang, W., Hui, J., Kiefer, A.K., Duffy, D.L., Baltic, S., Danoy, P., et al. (2014). Genome-wide association analysis identifies 11 risk variants associated with the asthma with hay fever phenotype. J. Allergy Clin. Immunol. *133*, 1564–1571. https://doi.org/10.1016/j.jaci.2013.10.030.

19. Zhu, Z., Hasegawa, K., Camargo, C.A., and Liang, L. (2021). Investigating asthma heterogeneity through shared and distinct genetics: insights from genome-wide cross-trait analysis. J. Allergy Clin. Immunol. *147*, 796–807. https://doi.org/10.1016/j.jaci.2020.07.004.

20. Zhu, Z., Guo, Y., Shi, H., Liu, C.-L., Panganiban, R.A., Chung, W., O'Connor, L.J., Himes, B.E., Gazal, S., Hasegawa, K., et al. (2020). Shared genetic and experimental links between obesity-related traits and asthma subtypes in UK Biobank. J. Allergy Clin. Immunol. *145*, 537–549. https://doi.org/10.1016/j.jaci.2019.09.035.

21. Zhu, Z., Lee, P.H., Chaffin, M.D., Chung, W., Loh, P.-R., Lu, Q., Christiani, D.C., and Liang, L. (2018). A genome-wide cross-trait analysis from UK Biobank highlights the shared genetic architecture of asthma and allergic diseases. Nat. Genet. *50*, 857–864. https://doi.org/10.1038/s41588-018-0121-0.

22. Zhu, Z., Zhu, X., Liu, C.-L., Shi, H., Shen, S., Yang, Y., Hasegawa, K., Camargo, C.A., Jr., and Liang, L. (2019). Shared genetics of asthma and mental health disorders: a large-scale genome-wide cross-trait analysis. Eur. Respir. J. *54*, 1901507. https://doi.org/10.1183/13993003.01507-2019.

23. Van Wonderen, K.E., Van Der Mark, L.B., Mohrs, J., Bindels, P.J.E., Van Aalderen, W.M.C., and Ter Riet, G. (2010). Different definitions in childhood asthma: how dependable is the dependent variable? Eur. Respir. J. *36*, 48–56. https://doi.org/10.1183/09031936.00154409.

24. Colicino, S., Munblit, D., Minelli, C., Custovic, A., and Cullinan, P. (2019). Validation of childhood asthma predictive tools: a systematic review. Clin. Exp. Allergy *49*, 410–418. https://doi.org/10.1111/cea.13336.

25. Lambert, S.A., Abraham, G., and Inouye, M. (2019). Towards clinical utility of polygenic risk scores. Hum. Mol. Genet. 28, R133–R142. https://doi.org/10.1093/hmg/ddz187.

26. Lewis, C.M., and Vassos, E. (2020). Polygenic risk scores: from research tools to clinical instruments. Genome Med. 12, 44. https://doi.org/10.1186/s13073-020-00742-5.

27. Chatterjee, N., Shi, J., and García-Closas, M. (2016). Developing and evaluating polygenic risk prediction models for stratified disease prevention. Nat. Rev. Genet. 17, 392–406. https://doi.org/10.1038/nrg.2016.27.

28. Dijk, F.N., Folkersma, C., Gruzieva, O., Kumar, A., Wijga, A.H., Gehring, U., Kull, I., Postma, D.S., Vonk, J.M., Melén, E., and Koppelman, G.H. (2019). Genetic risk scores do not improve asthma prediction in childhood. J. Allergy Clin. Immunol. 144, 857–860.e7. https://doi.org/10.1016/j.jaci.2019.05.017.

29. Kothalawala, D.M., Kadalayil, L., Curtin, J.A., Murray, C.S., Simpson, A., Custovic, A., Tapper, W.J., Arshad, S.H., Rezwan, F.I., Holloway, J.W., and On Behalf Of Stelar/Unicorn Investigators. (2022). Integration of genomic risk scores to improve the prediction of childhood asthma diagnosis. J. Pers. Med. 12, 75. https://doi.org/10.3390/jpm12010075.

30. Belsky, D.W., Sears, M.R., Hancox, R.J., Harrington, H., Houts, R., Moffitt, T.E., Sugden, K., Williams, B., Poulton, R., and Caspi, A. (2013). Polygenic risk and the development and course of asthma: an analysis of data from a four-decade longitudinal study. Lancet Respir. Med. 1, 453–461. https://doi.org/10.1016/S2213-2600(13)70101-2.

31. Sordillo, J.E., Lutz, S.M., Jorgenson, E., Iribarren, C., McGeachie, M., Dahlin, A., Tantisira, K., Kelly, R., Lasky-Su, J., Sakornsakolpat, P., et al. (2021). A polygenic risk score for asthma in a large racially diverse population. Clin. Exp. Allergy 51, 1410–1420. https://doi.org/10.1111/cea.14007.

32. Namjou, B., Lape, M., Malolepsza, E., DeVore, S.B., Weirauch, M.T., Dikilitas, O., Jarvik, G.P., Kiryluk, K., Kullo, I.J., Liu, C., et al. (2022). Multi-ancestral polygenic risk score for pediatric asthma. J. Allergy Clin. Immunol. Pre-press. https://doi.org/10.1016/j.jaci.2022.03.035.

33. Zhou, W., Kanai, M., Wu, K.-H.H., Humaira, R., Tsuo, K., Hirbo, J.B., Wang, Y., Bhattacharya, A., Zhao, H., Namba, S., et al. (2021). Global Biobank Meta-analysis Initiative: powering genetic discovery across human diseases. Preprint at medRxiv. https://doi.org/10.1101/2021.11.19.21266436.

34. 1000 Genomes Project Consortium; Auton, A., Brooks, L.D., Durbin, R.M., Garrison, E.P., Kang, H.M., Korbel, J.O., Marchini, J.L., McCarthy, S., McVean, G.A., and Abecasis, G.R. (2015). A global reference for human genetic variation. Nature 526, 68–74. https://doi.org/10.1038/nature15393.

35. Deming, W.E. (1943). Statistical Adjustment of Data (Wiley), p. 261.

36. Jefferson, J.A., and Shankland, S.J. (2007). Familial nephrotic syndrome: PLCE1 enters the fray. Nephrol. Dial. Transplant. 22, 1849–1852. https://doi.org/10.1093/ndt/gfm098.

37. Riar, S.S., Banh, T.H.M., Borges, K., Subbarao, P., Patel, V., Vasilevska-Ristovska, J., Chanchlani, R., Hussain-Shamsy, N., Noone, D., Hebert, D., et al. (2019). Prevalence of asthma and allergies and risk of relapse in childhood nephrotic syndrome: insight into nephrotic syndrome cohort. J. Pediatr. 208, 251–257.e1. https://doi.org/10.1016/j.jpeds.2018.12.048.

38. UK Biobank — Neale lab. http://www.nealelab.is/uk-biobank/.

39. Loo, T.H., Ye, X., Chai, R.J., Ito, M., Bonne, G., Ferguson-Smith, A.C., and Stewart, C.L. (2019). The mammalian LINC complex component SUN1 regulates muscle regeneration by modulating drosha activity. Elife 8, e49485. https://doi.org/10.7554/eLife.49485.

40. Chen, M.-H., Raffield, L.M., Mousas, A., Sakaue, S., Huffman, J.E., Moscati, A., Trivedi, B., Jiang, T., Akbari, P., Vuckovic, D., et al. (2020). Trans-ethnic and ancestry-specific blood-cell genetics in 746, 667 individuals from 5 global populations. Cell 182, 1198–1213.e14. https://doi.org/10.1016/j.cell.2020.06.045.

41. Dalakas, M.C., and Spaeth, P.J. (2021). The importance of FcRn in neuro-immunotherapies: from IgG catabolism, FCGRT gene polymorphisms, IVIg dosing and efficiency to specific FcRn inhibitors. Ther. Adv. Neurol. Disord. 14, 1756286421997381. https://doi.org/10.1177/1756286421997381.

42. Nebert, D.W., and Liu, Z. (2019). SLC39A8 gene encoding a metal ion transporter: discovery and bench to bedside. Hum. Genomics 13, 51. https://doi.org/10.1186/s40246-019-0233-3.

43. Schizophrenia Working Group of the Psychiatric Genomics Consortium (2014). Biological insights from 108 schizophrenia-associated genetic loci. Nature 511, 421–427. https://doi.org/10.1038/nature13595.

44. Li, D., Achkar, J.-P., Haritunians, T., Jacobs, J.P., Hui, K.Y., D'Amato, M., Brand, S., Radford-Smith, G., Halfvarson, J., Niess, J.-H., et al. (2016). A pleiotropic missense variant in SLC39A8 is associated with Crohn's disease and human gut microbiome composition. Gastroenterology 151, 724–732. https://doi.org/10.1053/j.gastro.2016.06.051.

45. Huang, H., Fang, M., Jostins, L., Umićević Mirkov, M., Boucher, G., Anderson, C.A., Andersen, V., Cleynen, I., Cortes, A., Crins, F., et al. (2017). Fine-mapping inflammatory bowel disease loci to single-variant resolution. Nature 547, 173–178. https://doi.org/10.1038/nature22969.

46. Speliotes, E.K., Willer, C.J., Berndt, S.I., Monda, K.L., Thorleifsson, G., Jackson, A.U., Lango Allen, H., Lindgren, C.M., Luan, J., Mägi, R., et al. (2010). Association analyses of 249, 796 individuals reveal 18 new loci associated with body mass index. Nat. Genet. 42, 937–948. https://doi.org/10.1038/ng.686.

47. Pickrell, J.K., Berisa, T., Liu, J.Z., Ségurel, L., Tung, J.Y., and Hinds, D.A. (2016). Detection and interpretation of shared genetic influences on 42 human traits. Nat. Genet. 48, 709–717. https://doi.org/10.1038/ng.3570.

48. International Consortium for Blood Pressure Genome-Wide Association Studies; Ehret, G.B., Munroe, P.B., Rice, K.M., Bochud, M., Johnson, A.D., Chasman, D.I., Smith, A.V., Tobin, M.D., Verwoert, G.C., et al. (2011). Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. Nature 478, 103–109. https://doi.org/10.1038/nature10405.

49. Nakata, T., Creasey, E.A., Kadoki, M., Lin, H., Selig, M.K., Yao, J., Lefkovith, A., Daly, M.J., Graham, D.B., and Xavier, R.J. (2020). A missense variant in SLC39A8 confers risk for Crohn's disease by disrupting manganese homeostasis and intestinal barrier integrity. Proc. Natl. Acad. Sci. USA 117, 28930–28938. https://doi.org/10.1073/pnas.2014742117.

50. Qi, T., Wu, Y., Zeng, J., Zhang, F., Xue, A., Jiang, L., Zhu, Z., Kemper, K., Yengo, L., Zheng, Z., et al. (2018). Identifying gene targets for brain-related traits using transcriptomic and methylomic data from blood. Nat. Commun. 9, 2282. https://doi.org/10.1038/s41467-018-04558-1.

51. Mägi, R., Horikoshi, M., Sofer, T., Mahajan, A., Kitajima, H., Franceschini, N., McCarthy, M.I., COGENT-Kidney Consortium T2D-GENES Consortium; Morris, A.P., and Morris, A.P. (2017). Trans-ethnic meta-regression of genome-wide association studies accounting for ancestry increases power for discovery and improves fine-mapping resolution. Hum. Mol. Genet. 26, 3639–3650. https://doi.org/10.1093/hmg/ddx280.

52. Massoud, A.H., Charbonnier, L.-M., Lopez, D., Pellegrini, M., Phipatanakul, W., and Chatila, T.A. (2016). An asthma-associated IL4R variant exacerbates airway inflammation by promoting conversion of regulatory T cells to TH17-like cells. Nat. Med. 22, 1013–1022. https://doi.org/10.1038/nm.4147.

53. Kousha, A., Mahdavi Gorabi, A., Forouzesh, M., Hosseini, M., Alexander, M., Imani, D., Razi, B., Mousavi, M.J., Aslani, S., and Mikaeili, H. (2020). Interleukin 4 gene polymorphism (-589C/T) and the risk of asthma: a meta-analysis and met-regression based on 55 studies. BMC Immunol. 21, 55. https://doi.org/10.1186/s12865-020-00384-7.

54. Nie, W., Zhu, Z., Pan, X., and Xiu, Q. (2013). The interleukin-4 −589C/T polymorphism and the risk of asthma: a meta-analysis including 7345 cases and 7819 controls. Gene 520, 22–29. https://doi.org/10.1016/j.gene.2013.02.027.

55. Battle, N.C., Choudhry, S., Tsai, H.-J., Eng, C., Kumar, G., Beckman, K.B., Naqvi, M., Meade, K., Watson, H.G., LeNoir, M., et al. (2007).

Ethnicity-specific gene–gene interaction between IL-13 and IL-4Rα among african Americans with asthma. Am. J. Respir. Crit. Care Med. *175*, 881–887. https://doi.org/10.1164/rccm.200607-992OC.

56. Wang, Y., Namba, S., Lopera-Maya, E.A., Kerminen, S., Tsuo, K., Lall, K., Kanai, M., Zhou, W., Wu, K.-H.H., Fave, M.-J., et al. (2021). Global biobank analyses provide lessons for computing polygenic risk scores across diverse cohorts. Preprint at medRxiv. https://doi.org/10.1101/2021.11.18.21266545.

57. Ge, T., Chen, C.-Y., Ni, Y., Feng, Y.-C.A., and Smoller, J.W. (2019). Polygenic prediction via Bayesian regression and continuous shrinkage priors. Nat. Commun. *10*, 1776. https://doi.org/10.1038/s41467-019-09718-5.

58. Ruan, Y., Lin, Y.-F., Feng, Y.-C.A., Chen, C.-Y., Lam, M., Guo, Z., Stanley Global Asia Initiatives; He, L., Sawa, A., Martin, A.R., et al. (2022). Improving polygenic prediction in ancestrally diverse populations. Nat. Genet. *54*, 573–580. https://doi.org/10.1038/s41588-022-01054-7.

59. Bulik-Sullivan, B.K., Loh, P.-R., Finucane, H.K., Ripke, S., Yang, J., Schizophrenia Working Group of the Psychiatric Genomics Consortium; Patterson, N., Daly, M.J., Price, A.L., and Neale, B.M. (2015). LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. Nat. Genet. *47*, 291–295. https://doi.org/10.1038/ng.3211.

60. Ferreira, M.A.R., Mathur, R., Vonk, J.M., Szwajda, A., Brumpton, B., Granell, R., Brew, B.K., Ullemar, V., Lu, Y., Jiang, Y., et al. (2019). Genetic architectures of childhood- and adult-onset asthma are partly distinct. Am. J. Hum. Genet. *104*, 665–684. https://doi.org/10.1016/j.ajhg.2019.02.022.

61. Pividori, M., Schoettler, N., Nicolae, D.L., Ober, C., and Im, H.K. (2019). Shared and distinct genetic risk factors for childhood-onset and adult-onset asthma: genome-wide and transcriptome-wide studies. Lancet Respir. Med. *7*, 509–522. https://doi.org/10.1016/S2213-2600(19)30055-4.

62. de Marco, R., Pesce, G., Marcon, A., Accordini, S., Antonicelli, L., Bugiani, M., Casali, L., Ferrari, M., Nicolini, G., Panico, M.G., et al. (2013). The coexistence of asthma and chronic obstructive pulmonary disease (COPD): prevalence and risk factors in young, middle-aged and elderly people from the general population. PLoS One *8*, e62985. https://doi.org/10.1371/journal.pone.0062985.

63. Madore, A.-M., Bossé, Y., Margaritte-Jeannin, P., Vucic, E., Lam, W.L., Bouzigon, E., Bourbeau, J., and Laprise, C. (2022). Analysis of GWAS-nominated loci for lung cancer and COPD revealed a new asthma locus. BMC Pulm. Med. *22*, 155. https://doi.org/10.1186/s12890-022-01890-7.

64. Hobbs, B.D., de Jong, K., Lamontagne, M., Bossé, Y., Shrine, N., Artigas, M.S., Wain, L.V., Hall, I.P., Jackson, V.E., Wyss, A.B., et al. (2017). Genetic loci associated with chronic obstructive pulmonary disease overlap with loci for lung function and pulmonary fibrosis. Nat. Genet. *49*, 426–432. https://doi.org/10.1038/ng.3752.

65. Sakornsakolpat, P., Prokopenko, D., Lamontagne, M., Reeve, N.F., Guyatt, A.L., Jackson, V.E., Shrine, N., Qiao, D., Bartz, T.M., Kim, D.K., et al. (2019). Genetic landscape of chronic obstructive pulmonary disease identifies heterogeneous cell-type and phenotype associations. Nat. Genet. *51*, 494–505. https://doi.org/10.1038/s41588-018-0342-2.

66. de Leeuw, C.A., Mooij, J.M., Heskes, T., and Posthuma, D. (2015). MAGMA: generalized gene-set analysis of GWAS data. PLoS Comput. Biol. *11*, e1004219. https://doi.org/10.1371/journal.pcbi.1004219.

67. Pers, T.H., Karjalainen, J.M., Chan, Y., Westra, H.-J., Wood, A.R., Yang, J., Lui, J.C., Vedantam, S., Gustafsson, S., Esko, T., et al. (2015). Biological interpretation of genome-wide association studies using predicted gene functions. Nat. Commun. *6*, 5890. https://doi.org/10.1038/ncomms6890.

68. Weeks, E.M., Ulirsch, J.C., Cheng, N.Y., Trippe, B.L., Fine, R.S., Miao, J., Patwardhan, T.A., Kanai, M., Nasser, J., Fulco, C.P., et al. (2020). Leveraging polygenic enrichments of gene features to predict genes underlying complex traits and diseases. Preprint at bioRxiv. https://doi.org/10.1101/2020.09.08.20190561.

69. Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., and Mesirov, J.P. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. USA *102*, 15545–15550. https://doi.org/10.1073/pnas.0506580102.

70. Leynaert, B., Neukirch, F., Demoly, P., and Bousquet, J. (2000). Epidemiologic evidence for asthma and rhinitis comorbidity. J. Allergy Clin. Immunol. *106*, S201–S205. https://doi.org/10.1067/mai.2000.110151.

71. Van Lieshout, R.J., and Macqueen, G. (2008). Psychological factors in asthma. Allergy Asthma Clin. Immunol. *4*, 12–28. https://doi.org/10.1186/1710-1492-4-1-12.

72. Miethe, S., Karsonova, A., Karaulov, A., and Renz, H. (2020). Obesity and asthma. J. Allergy Clin. Immunol. *146*, 685–693. https://doi.org/10.1016/j.jaci.2020.08.011.

73. Guo, H., An, J., and Yu, Z. (2020). Identifying shared risk genes for asthma, hay fever, and eczema by multi-trait and multiomic association analyses. Front. Genet. *11*, 270. https://doi.org/10.3389/fgene.2020.00270.

74. Kim, S.Y., Min, C., Oh, D.J., and Choi, H.G. (2019). Increased risk of asthma in patients with rheumatoid arthritis: a longitudinal follow-up study using a national sample cohort. Sci. Rep. *9*, 6957. https://doi.org/10.1038/s41598-019-43481-3.

75. Shen, T.-C., Lin, C.-L., Wei, C.-C., Tu, C.-Y., and Li, Y.-F. (2014). The risk of asthma in rheumatoid arthritis: a population-based cohort study. QJM *107*, 435–442. https://doi.org/10.1093/qjmed/hcu008.

76. Luo, Y., Fan, X., Jiang, C., Arevalo Molina, A.B., Salgado, M., and Xu, J. (2018). Rheumatoid arthritis is associated with increased in-hospital mortality in asthma exacerbations: a nationwide study. Clin. Rheumatol. *37*, 1971–1976. https://doi.org/10.1007/s10067-018-4114-2.

77. Jeong, H.E., Jung, S.-M., and Cho, S.-I. (2018). Association between rheumatoid arthritis and respiratory allergic diseases in Korean adults: a propensity score matched case-control study. Int. J. Rheumatol. *2018*, 3798124. https://doi.org/10.1155/2018/3798124.

78. Rolfes, M.C., Juhn, Y.J., Wi, C.-I., and Sheen, Y.H. (2017). Asthma and the risk of rheumatoid arthritis: an insight into the heterogeneity and phenotypes of asthma. Tuberc. Respir. Dis. *80*, 113–135. https://doi.org/10.4046/trd.2017.80.2.113.

79. Shirai, Y., Nakanishi, Y., Suzuki, A., Konaka, H., Nishikawa, R., Sonehara, K., Namba, S., Tanaka, H., Masuda, T., Yaga, M., et al. (2022). Multi-trait and cross-population genome-wide association studies across autoimmune and allergic diseases identify shared and distinct genetic component. Ann. Rheum. Dis. *81*, 1301–1312. https://doi.org/10.1136/annrheumdis-2022-222460.

80. Ehrlich, S.F., Quesenberry, C.P., Jr., Van Den Eeden, S.K., Shan, J., and Ferrara, A. (2010). Patients diagnosed with diabetes are at increased risk for asthma, chronic obstructive pulmonary disease, pulmonary fibrosis, and pneumonia but not lung cancer. Diabetes Care *33*, 55–60. https://doi.org/10.2337/dc09-0880.

81. Thomsen, S.F., Duffy, D.L., Kyvik, K.O., Skytthe, A., and Backer, V. (2011). Risk of asthma in adult twins with type 2 diabetes and increased body mass index. Allergy *66*, 562–568. https://doi.org/10.1111/j.1398-9995.2010.02504.x.

82. Torres, R.M., Souza, M.D.S., Coelho, A.C.C., de Mello, L.M., and Souza-Machado, C. (2021). Association between asthma and type 2 diabetes mellitus: mechanisms and impact on asthma control—a literature review. Can. Respir. J. *2021*, 8830439. https://doi.org/10.1155/2021/8830439.

83. Sun, Y.-Q., Brumpton, B.M., Langhammer, A., Chen, Y., Kvaløy, K., and Mai, X.-M. (2020). Adiposity and asthma in adults: a bidirectional Mendelian randomisation analysis of the HUNT Study. Thorax *75*, 202–208. https://doi.org/10.1136/thoraxjnl-2019-213678.

84. Zhu, Z., Li, J., Si, J., Ma, B., Shi, H., Lv, J., Cao, W., Guo, Y., Millwood, I.Y., Walters, R.G., et al. (2021). A large-scale genome-wide association analysis of lung function in the Chinese population identifies novel loci and highlights shared genetic aetiology with obesity. Eur. Respir. J. *58*, 2100199. https://doi.org/10.1183/13993003.00199-2021.

85. Tanaka, N., Koido, M., Suzuki, A., Otomo, N., Suetsugu, H., Kochi, Y., Tomizuka, K., Momozawa, Y., Kamatani, Y., Biobank Japan Project, et al.

(2021). Eight novel susceptibility loci and putative causal variants in atopic dermatitis. J. Allergy Clin. Immunol. *148*, 1293–1306. https://doi.org/10.1016/j.jaci.2021.04.019.

86. Park, J., Jang, H., Kim, M., Hong, J.Y., Kim, Y.H., Sohn, M.H., Park, S.-C., Won, S., and Kim, K.W. (2021). Predicting allergic diseases in children using genome-wide association study (GWAS) data and family history. World Allergy Organ. J. *14*, 100539. https://doi.org/10.1016/j.waojou.2021.100539.

87. Majara, L., Kalungi, A., Koen, N., Zar, H., Stein, D.J., Kinyanda, E., Atkinson, E.G., and Martin, A.R. (2021). Low generalizability of polygenic scores in African populations due to genetic and environmental diversity. Preprint at bioRxiv. https://doi.org/10.1101/2021.01.12.426453.

88. Kumbhare, S., Pleasants, R., Ohar, J.A., and Strange, C. (2016). Characteristics and prevalence of asthma/chronic obstructive pulmonary disease overlap in the United States. Ann. Am. Thorac. Soc. *13*, 803–810. https://doi.org/10.1513/AnnalsATS.201508-554OC.

89. Hosseini, M., Almasi-Hashiani, A., Sepidarkish, M., and Maroufizadeh, S. (2019). Global prevalence of asthma-COPD overlap (ACO) in the general population: a systematic review and meta-analysis. Respir. Res. *20*, 229. https://doi.org/10.1186/s12931-019-1198-4.

90. Akmatov, M.K., Ermakova, T., Holstiege, J., Steffen, A., von Stillfried, D., and Bätzing, J. (2020). Comorbidity profile of patients with concurrent diagnoses of asthma and COPD in Germany. Sci. Rep. *10*, 17945. https://doi.org/10.1038/s41598-020-74966-1.

91. Munafò, M.R., Tilling, K., Taylor, A.E., Evans, D.M., and Davey Smith, G. (2018). Collider scope: when selection bias can substantially influence observed associations. Int. J. Epidemiol. *47*, 226–235. https://doi.org/10.1093/ije/dyx206.

92. Griffith, G.J., Morris, T.T., Tudball, M.J., Herbert, A., Mancano, G., Pike, L., Sharp, G.C., Sterne, J., Palmer, T.M., Davey Smith, G., et al. (2020). Collider bias undermines our understanding of COVID-19 disease risk and severity. Nat. Commun. *11*, 5749. https://doi.org/10.1038/s41467-020-19478-2.

93. Karczewski, K.J., Francioli, L.C., Tiao, G., Cummings, B.B., Alföldi, J., Wang, Q., Collins, R.L., Laricchia, K.M., Ganna, A., Birnbaum, D.P., et al. (2021). Author Correction: the mutational constraint spectrum quantified from variation in 141, 456 humans. Nature *590*, E53. https://doi.org/10.1038/s41586-020-03174-8.

94. The GTEx Consortium; Anand, S., Ardlie, K.G., Gabriel, S., Getz, G.A., Graubert, A., Hadley, K., Handsaker, R.E., Huang, K.H., Kashin, S., et al. (2020). The GTEx Consortium atlas of genetic regulatory effects across human tissues. Science *369*, 1318–1330.

95. Cann, H.M., de Toma, C., Cazes, L., Legrand, M.-F., Morel, V., Piouffre, L., Bodmer, J., Bodmer, W.F., Bonne-Tamir, B., Cambon-Thomsen, A., et al. (2002). A human genome diversity cell line panel. Science *296*, 261–262. https://doi.org/10.1126/science.296.5566.261b.

96. Pan-UK Biobank. https://pan.ukbb.broadinstitute.org/.

97. Akiyama, M., Okada, Y., Kanai, M., Takahashi, A., Momozawa, Y., Ikeda, M., Iwata, N., Ikegawa, S., Hirata, M., Matsuda, K., et al. (2017). Genome-wide association study identifies 112 new loci for body mass index in the Japanese population. Nat. Genet. *49*, 1458–1467. https://doi.org/10.1038/ng.3951.

98. Kanai, M., Akiyama, M., Takahashi, A., Matoba, N., Momozawa, Y., Ikeda, M., Iwata, N., Ikegawa, S., Hirata, M., Matsuda, K., et al. (2018). Genetic analysis of quantitative traits in the Japanese population links cell types to complex human diseases. Nat. Genet. *50*, 390–400. https://doi.org/10.1038/s41588-018-0047-6.

99. Ishigaki, K., Akiyama, M., Kanai, M., Takahashi, A., Kawakami, E., Sugishita, H., Sakaue, S., Matoba, N., Low, S.-K., Okada, Y., et al. (2020). Large-scale genome-wide association study in a Japanese population

identifies novel susceptibility loci across different diseases. Nat. Genet. *52*, 669–679. https://doi.org/10.1038/s41588-020-0640-3.

100. Zhou, W., Nielsen, J.B., Fritsche, L.G., Dey, R., Gabrielsen, M.E., Wolford, B.N., LeFaive, J., VandeHaar, P., Gagliano, S.A., Gifford, A., et al. (2018). Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies. Nat. Genet. *50*, 1335–1341. https://doi.org/10.1038/s41588-018-0184-y.

101. Wang, K., Li, M., and Hakonarson, H. (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. *38*, e164. https://doi.org/10.1093/nar/gkq603.

102. Purcell, S., and Chang, C. Plink 1.9. www.cog-genomics.org/plink/1.9/.

103. Boughton, A.P., Welch, R.P., Flickinger, M., VandeHaar, P., Taliun, D., Abecasis, G.R., and Boehnke, M. (2021). LocusZoom.js: interactive and embeddable visualization of genetic association study results. Bioinformatics *37*, 3017–3018. https://doi.org/10.1093/bioinformatics/btab186.

104. Wei, T., and Simko, V. (2021). R Package "Corrplot": Visualization of a Correlation Matrix.

105. Denny, J.C., Bastarache, L., Ritchie, M.D., Carroll, R.J., Zink, R., Mosley, J.D., Field, J.R., Pulley, J.M., Ramirez, A.H., Bowton, E., et al. (2013). Systematic comparison of phenome-wide association study of electronic medical record data and genome-wide association study data. Nat. Biotechnol. *31*, 1102–1110. https://doi.org/10.1038/nbt.2749.

106. Buniello, A., MacArthur, J.A.L., Cerezo, M., Harris, L.W., Hayhurst, J., Malangone, C., McMahon, A., Morales, J., Mountjoy, E., Sollis, E., et al. (2019). The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. Nucleic Acids Res. *47*, D1005–D1012. https://doi.org/10.1093/nar/gky1120.

107. McLaren, W., Gil, L., Hunt, S.E., Riat, H.S., Ritchie, G.R.S., Thormann, A., Flicek, P., and Cunningham, F. (2016). The ensembl variant effect predictor. Genome Biol. *17*, 122. https://doi.org/10.1186/s13059-016-0974-4.

108. Kerimov, N., Hayhurst, J.D., Peikova, K., Manning, J.R., Walter, P., Kolberg, L., Samoviča, M., Sakthivel, M.P., Kuzmin, I., Trevanion, S.J., et al. (2021). A compendium of uniformly processed human gene expression and splicing quantitative trait loci. Nat. Genet. *53*, 1290–1299. https://doi.org/10.1038/s41588-021-00924-w.

109. Wang, G., Sarkar, A., Carbonetto, P., and Stephens, M. (2020). A simple new approach to variable selection in regression, with application to genetic fine mapping. J. R. Stat. Soc. B *82*, 1273–1300. https://doi.org/10.1111/rssb.12388.

110. Kanai, M., Ulirsch, J.C., Karjalainen, J., Kurki, M., Karczewski, K.J., Fauman, E., Wang, Q.S., Jacobs, H., Aguet, F., Ardlie, K.G., et al. (2021). Insights from complex trait fine-mapping across diverse populations. Preprint at bioRxiv. https://doi.org/10.1101/2021.09.03.21262975.

111. Higgins, J.P.T., and Thompson, S.G. (2002). Quantifying heterogeneity in a meta-analysis. Stat. Med. *21*, 1539–1558. https://doi.org/10.1002/sim.1186.

112. Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M., and Lee, J.J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaScience *4*, 7. https://doi.org/10.1186/s13742-015-0047-8.

113. Finucane, H.K., Bulik-Sullivan, B., Gusev, A., Trynka, G., Reshef, Y., Loh, P.-R., Anttila, V., Xu, H., Zang, C., Farh, K., et al. (2015). Partitioning heritability by functional annotation using genome-wide association summary statistics. Nat. Genet. *47*, 1228–1235. https://doi.org/10.1038/ng.3404.

114. Finucane, H.K., Reshef, Y.A., Anttila, V., Slowikowski, K., Gusev, A., Byrnes, A., Gazal, S., Loh, P.-R., Lareau, C., Shoresh, N., et al. (2018). Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. Nat. Genet. *50*, 621–629. https://doi.org/10.1038/s41588-018-0081-4.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
GBMI summary statistics	Zhou et al.[33]	https://www.globalbiobankmeta.org/resources
PRS-CS weights	Wang et al.[56]	https://www.pgscatalog.org/publication/PGP000262/
gnomAD	Karczewski et al.[93]	https://www.nature.com/articles/s41586-020-2308-7
GTEx v8	Aguet et al.[94]	https://www.science.org/doi/10.1126/science.aaz1776
1000 Genomes	1000 Genomes Project Consortium et al.[34]	https://www.nature.com/articles/nature15393
Human Genome Diversity Project (HGDP)	Cann et al.[95]	https://www.science.org/doi/10.1126/science.296.5566.261b
Pan-UK Biobank summary statistics	Pan-UK Biobank[96]	https://pan.ukbb.broadinstitute.org/
Biobank Japan summary statistics	Akiyama et al.[97]; Kanai et al.[98]; Ishigaki et al.[99]	https://www.nature.com/articles/ng.3951; https://www.nature.com/articles/s41588-018-0047-6; https://www.nature.com/articles/s41588-020-0640-3
Molecular Signatures database v7.4	Sakornsakolpat et al.[69]	https://www.pnas.org/doi/10.1073/pnas.0506580102
Software and algorithms		
Scripts used for quality control, meta-analysis, and summary of results	Zhou et al.[33]; Wang et al.[56]	https://zenodo.org/badge/latestdoi/295461030
Scripts used for analysis of asthma meta-analysis results	This study	https://doi.org/10.5281/zenodo.7130276
PC projection	Zhou et al.[33]	https://zenodo.org/badge/latestdoi/353203447
SAIGE	Zhou et al.[100]	https://www.nature.com/articles/s41588-018-0184-y
MR-MEGA	Mägi et al.[51]	https://academic.oup.com/hmg/article/26/18/3639/3976569
ANNOVAR	Wang et al.[101]	https://academic.oup.com/nar/article/38/16/e164/1749458
LDSC	Bulik-Sullivan et al.[59]	https://www.nature.com/articles/ng.3211
MAGMA	de Leeuw et al.[66]	https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1004219
DEPICT	Pers et al.[67]	https://www.nature.com/articles/ncomms6890
PoPS	Weeks et al.[68]	https://doi.org/10.1101/2020.09.08.20190561
PRS-CS	Ge et al.[57]	https://www.nature.com/articles/s41467-019-09718-5
PRS-CSx	Ruan et al.[58]	https://www.nature.com/articles/s41588-022-01054-7
PLINK v1.9	Purcell and Chang[102]	www.cog-genomics.org/plink/1.9/
LocalZoom	Boughton et al.[103]	https://academic.oup.com/bioinformatics/article/37/18/3017/6178278
Corrplot	Wei and Simko[104]	https://github.com/taiyun/corrplot

RESOURCE AVAILABILITY

Lead contact
Further information and requests for resources should be directed to and will be fulfilled by the lead contact, Kristin Tsuo (ktsuo@broadinstitute.org).

Materials availability
This study did not generate new unique reagents.

Data and code availability
- The all-biobank meta-analysis results and plots for the 14 endpoints (including both ancestry-specific and cross-ancestry meta-analyses) are available for downloading at https://www.globalbiobankmeta.org/resources and browsing at the browser http://results.globalbiobankmeta.org. The PRS-CS weights estimated using the all-biobank multi-ancestry meta-analyses and leave-UKBB-out multi-ancestry meta-analyses have been deposited within the PGS Catalog with study ID PGP000262 (https://www.pgscatalog.org/).
- All original code has been deposited to Zenodo with DOIs as below and is publicly available as of the date of publication. Links are listed in the key resources table.
- Scripts used for quality control, meta-analysis and summary of results are available at https://github.com/globalbiobankmeta and deposited at https://zenodo.org/badge/latestdoi/295461030.
- Scripts for PC projection are deposited at https://zenodo.org/badge/latestdoi/353203447.
- Scripts for analysis of asthma meta-analysis results are available at https://github.com/ktsuo/globalbiobankmeta-Asthma and deposited at https://doi.org/10.5281/zenodo.7130276.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

METHODS DETAILS

Asthma phenotype definitions
The phenotype definition guidelines that were developed by GBMI and shared with all participating biobanks can be found in Zhou et al.[33] Disease endpoints, including asthma, were defined following the PheCode maps, which maps ICD-9 or ICD-10 codes to PheCodes.[105] Asthma cases were all study participants with the asthma PheCode, and controls were all study participants without the asthma PheCode (or asthma-related PheCodes). Biobanks that did not have ICD codes primarily used self-reported data (Table S3).

Principal components (PC) projection
To compare the genetic ancestries represented in different biobanks, we used pre-computed loadings of genetic markers shared across all biobanks and the reference data containing 1000 Genomes (1000G) and the Human Genome Diversity Project (HGDP) to project biobank participants to the same principal components space. 179,195 genetic variants were genotyped/imputed in all biobanks, among which 168,899 are also in the 1000 Genomes[34] and HGDP.[95] The weights corresponding to principal components for those markers were estimated based on the PCA for the reference samples with known ancestry in 1000G and HGDP and shared among biobanks. Biobanks then generated PC loadings based on the pre-estimated weights of those markers. More details are described in Zhou et al.[33]

Meta-analysis
We performed fixed-effects meta-analysis with inverse variance weighting for 18 biobanks in GBMI: China Kadoorie Biobank (CKB), Generation Scotland (GS), Lifelines, QSKIN, East London Genes & Health (GNH), HUNT, UCLA Precision Health Biobank (UCLA), Colorado Center for Personalized Medicine (CCPM), Mass General Brigham (MGB), BioVU, BioMe, Michigan Genomics Initiative (MGI), BioBank Japan (BBJ), Estonian Biobank (ESTBB), deCODE Genetics (DECODE), FinnGen, Taiwan Biobank (TWB), and UK Biobank (UKBB). Basic information on the biobanks are described in Zhou et al.,[33] as well as details on the genotyping, imputation, GWAS, post-GWAS quality control, and meta-analysis procedures.[33] In brief, genetic variants with minor allele count (MAC) < 20 and imputation score <0.3 were excluded from the analyses. Genetic variants with different allele frequencies (AF) compared to gnomAD[93] (Mahalanobis distance between AF-GWAS and AF-gnomAD > 3 standard deviations away from the mean) were also excluded. Altogether, these cohorts had a total sample size of 153,763 cases and 1,647,022 controls (Table S1). GWAS meta-analyses were first conducted within continental ancestry groups to control for population stratification. 5,051 cases and 27,607 controls were of African (AFR) ancestry; 4,069 cases and 14,104 controls were of Admixed American (AMR) ancestry; 18,549 cases and 322,655 controls were of East Asian (EAS) ancestry; 121,940 cases and 1,254,131 were of European (EUR) ancestry; 139 cases and 1,434 controls were of Middle Eastern (MID) ancestry; and 4,015 cases and 27,091 controls were of Central and South Asian (CSA) ancestry.

We also performed fixed-effects meta-analysis with inverse variance weighting for 4 additional biobanks that served as independent replication studies: Canadian Partnership for Tomorrow's Health (CanPath), Qatar Biobank (QBB), Biobank of the Americas (BBofA), and Penn Medicine Biobank (PMBB). Collectively, these cohorts had a total sample size of 9,991 cases and 63,605 controls (Table S1). More information on these biobanks are also described in Zhou et al.[33]

Index variant and locus definitions

We used a threshold of $p < 5 \times 10^{-8}$ to identify SNPs with a genome-wide significant effect. To identify loci, we used a window size of 500 kb upstream and downstream of the SNPs with the strongest evidence of association in the meta-analysis, and merged overlapping regions until no genome-wide significant variants were detected within the ±500 kb region. To designate loci as previously discovered or potentially novel, we compiled a list of known asthma-associated SNPs ($p < 5 \times 10^{-8}$) from the associations collected by El-Husseini et al.[6] and listed in the GWAS catalog (as of 11/14/2021).[106] We extended 500 kb upstream and downstream of each of these variants to define a locus, and intersected these regions with the loci defined by the index variants in our meta-analysis to identify any overlaps. We annotated genetic variants with the nearest genes using ANNOVAR[101] and putative loss-of-function using VEP[107] with the LOFTEE plug[93] as implemented in Hail.[33] We also annotated whether the index or tagging variants ($r^2 > 0.8$) of asthma were fine-mapped in any of the *cis*-eQTL fine-mapping resources. We retrieved *cis*-eQTL fine-mapped variants with posterior inclusion probability (PIP) > 0.9 in any tissues and cell types from the GTEx v8[99] and eQTL catalog release 4.[108] Fine-mapping was conducted using SuSiE[109] with summary statistics and covariate-adjusted in-sample LD matrix as described previously.[110] Genome positions are reported in build hg38 for index variants.

Index SNP effects across biobanks

To estimate the correlation of SNP effects for the 179 top loci between one specific biobank and the leave-that-biobank-out meta-analysis, we used the method proposed by Qi et al.[50] using GWAS summary statistics (Table S5). Specifically, the method directly calculates SNP effect correlation as:

$$\hat{r}_b = \frac{\widehat{cov}(\hat{b}_{biobank}, \hat{b}_{leave-biobank})}{\sqrt{[\widehat{var}(\hat{b}_{biobank}) - \widehat{var}(e_{biobank})][\widehat{var}(\hat{b}_{leave-biobank}) - \widehat{var}(e_{leave-biobank})]}}$$

where $\hat{b}_{biobank}$ and $\hat{b}_{leave-biobank}$ denote the estimated SNP effects from GWAS conducted in one specific biobank and from GWAS performed in the leave-that-biobank-out meta-analysis, respectively. The $\widehat{cov}(\hat{b}_{biobank}, \hat{b}_{leave-biobank})$ is calculated as the sampling covariance between $\hat{b}_{biobank}$ and $\hat{b}_{leave-biobank}$. The $\widehat{var}(\hat{b}_{biobank})$ and $\widehat{var}(\hat{b}_{leave-biobank})$ are the estimated variances of $\hat{b}_{biobank}$ and $\hat{b}_{leave-biobank}$, separately. The $\widehat{var}(e_{biobank})$ and $\widehat{var}(e_{leave-biobank})$ are the estimated variance of the estimation errors of $\hat{b}_{biobank}$ and $\hat{b}_{leave-biobank}$, which are approximated as the mean of the squared standard errors of estimated SNP effect ($\hat{b}_{biobank}$ and $\hat{b}_{leave-biobank}$) across all the top-associated SNPs, respectively. The standard error of r_b is obtained through the jackknife approach by leaving one SNP out each time. SNPs with large standard errors in CKB and HUNT (chr12:123241280:T:C and chr17:7878812:T:C, respectively) were excluded from these analyses.

Then, for the index variants present in each biobank, we computed:

$$\frac{biobank\ meta - analysis\ effect\ size}{leave - that - biobank - out\ meta - analysis\ effect\ size}$$

for the biobank and leave-that-biobank-out pair. We took the average ratio across the index variants for each biobank and leave-that-biobank-out pair. We then used the regression method introduced in Deming et al.,[35] which considers the errors in both the *X*- and *Y*-variables, to compare the effect sizes of these SNPs in each biobank GWAS with their effects in the leave-that-biobank-out meta-analysis. We set the intercept equal to 0 for these analyses.

Ancestry-specific heterogeneity

To assess heterogeneity of per-SNP effect sizes for the 179 top loci across ancestries in the GBMI, we conducted ancestry-specific meta-analyses of the five most well-powered ancestry groups in the GBMI (EUR, AFR, AMR, EAS, and CSA). We applied the Cochran's Q test[111] to the SNP effects in the ancestry-specific meta-analyses and identified SNPs with significant heterogeneity based on a Bonferroni-corrected p value cut-off of 0.05/169 = 0.0003, accounting for the number of SNPs present in all studies (Table S7). Regions displaying heterogeneity in effects across ancestry groups were visualized using the LocalZoom tool.[103]

Polygenic risk scores

A description of the PRS analyses conducted using PRS-CS,[57] as well as the leave-one-biobank-out meta-analysis strategy applied, is provided in Wang et al.[56]

We used PRS-CSx, which jointly models GWAS summary statistics from populations of different ancestries and returns posterior SNP effect size estimates for each input population.[58] We applied this method to the AMR, AFR, CSA, EAS, and EUR ancestry-specific meta-analyses, which served as the discovery data for PRS construction. For the ancestry-specific meta-analysis that matched

the ancestry of the target cohort, we excluded the target cohort. We evaluated the predictive performance of the PRS in 4 target cohorts: 1) AFR ancestry individuals in UKBB (849 cases, 5190 controls), 2) CSA ancestry individuals in UKBB (1232 cases, 6744 controls), 3) EAS ancestry individuals in BBJ that were part of a randomly-selected 1k holdout set (500 cases, 500 controls), and 4) EUR ancestry individuals in UKBB that were part of a randomly-selected holdout set (1164 cases, 7577 controls). We also evaluated the PRS in an additional randomly-selected 1k holdout set (131 cases, 869 controls) of EUR ancestry individuals in UKBB. As an example, for the AFR ancestry individuals, the full set of discovery data for PRS construction consisted of the AMR, CSA, EAS, and EUR ancestry-specific meta-analyses, as well as the AFR ancestry-specific meta-analysis excluding the AFR ancestry individuals in UKBB. The same strategy was applied to the other 3 target cohorts (Table S9). We used ancestry-matched LD reference panels from UKBB data and the default PRS-CSx settings for all input parameters. We evenly and randomly split cases and controls in the target cohorts into validation and testing subsets. Using the posterior SNP effect size estimates from PRS-CSx, we computed one PRS per discovery population for the validation subsets to learn the optimal linear combination of the ancestry-specific PRS using PLINK v1.9.[102,112] Then, with these weights, we evaluated the prediction accuracy of this linear combination of PRS in the testing subset. We reported the average prediction accuracy, measured by variance explained on the liability scale ($R^2_{liability}$), estimated using the prevalence of asthma in the BBJ biobank for the EAS target cohort and in the UKBB biobank for the other target cohorts, across 100 random splits.

Age-of-onset subtype GWAS and meta-analyses
UKBB
We first identified EUR individuals in UKBB with an asthma diagnosis based on information from the asthma PheCode or field 20002, which has self-reported diagnoses from verbal interviews. We then excluded individuals with either (1) a COPD diagnosis based on the COPD PheCode or field 20002, (2) missing information for field 3786, which has age at first asthma diagnosis information, (3) an asthma diagnosis after age 60 based on field 3786, or (4) greater than 10 years between the age reported in field 3786 and the age reported in field 22147, another age at first asthma diagnosis field that only a subset of participants filled out as part of a follow-up questionnaire. Then, using the age at first diagnosis reported in field 3786, we divided these individuals into asthma age of onset groups: those with diagnoses at or before age 19 were childhood-onset (n = 12,577) and after age 19 were adult-onset (n = 23,533). We then conducted separate COA and AOA GWAS using Scalable and Accurate Implementation of GEneralized mixed model (SAIGE).[100] The same set of controls was used (n = 359,116) for both GWAS, derived based on the PheCode guidelines provided by the GBMI.[33]

FinnGen
We identified individuals with an asthma diagnosis based on the PheCode guidelines provided by the GBMI[33] (Table S3). We excluded individuals with either (1) a COPD diagnosis based on the COPD PheCode definition, or (2) an asthma diagnosis after age 60. Those with diagnoses at or before age 19 were childhood-onset (n = 8,387) and after age 19 were adult-onset (n = 33,191). We conducted separate COA and AOA GWAS using SAIGE.[100] The same set of controls was used (n = 314,898) for both GWAS, derived based on the PheCode guidelines provided by the GBMI.[33]

Meta-analyses
We performed fixed-effects meta-analysis with inverse variance weighting for the COA GWAS from UKBB and FinnGen and the AOA GWAS from both biobanks. We used linkage-disequilibrium score correlation (LDSC)[59] to compute pairwise genetic correlations (r_g) between (1) the subtype meta-analyses, (2) each subtype meta-analysis and the GBMI COPD meta-analysis, (3) each subtype GWAS from UKBB and the GBMI all-asthma leave-UKBB-out meta-analysis, and (4) each subtype GWAS from FinnGen and the GBMI all-asthma leave-FinnGen-out meta-analysis. Finally, using the regression method introduced in Deming et al.,[35] we compared the effect sizes of the 179 index variants discovered in the GBMI all-asthma meta-analysis in each subtype meta-analysis. We set the intercept equal to 0 for this analysis.

Genetic correlation in UKBB and BBJ
Using LDSC, we estimated r_g between all EUR-ancestry UKBB phenotypes with heritability Z score > 4 and (1) the GBMI leave-UKBB-out meta-analysis for asthma and (2) the UKBB EUR-ancestry GWAS of asthma (PheCode ID 495 in UKBB) (Table S12). The heritability Z-scores were obtained from the stratified-LDSC (S-LDSC) computations of heritability reported by the Pan-UK Biobank team.[96,113,114] Summary statistics from the UKBB EUR GWAS were obtained from the Pan-UK Biobank team as well.[96]

We also used LDSC[59] to compute r_g between 48 phenotypes in BioBank Japan (BBJ) and (1) the GBMI leave-BBJ-out meta-analysis for asthma and (2) the BBJ GWAS of asthma (Table S13). We used publicly available GWAS summary statistics for all traits.[97–99] Genetic correlation results were visualized using the R corrplot package.[104]

Gene- and pathway-based enrichment
Fixed-effects meta-analysis with inverse variance weighting was performed for 16 biobanks in the GBMI with COPD data: BBJ, BioMe, BioVU, CCPM, CKB, ESTBB, FinnGen, GNH, GS, HUNT, Lifelines, MGB, MGI, TWB, UCLA, and UKBB. The same processing and methods were applied here as for the asthma meta-analysis. These cohorts had a total sample size of 81,568 cases and 1,310,798 controls. COPD cases were defined based on the COPD PheCode, and controls were all study participants without the COPD or COPD-related PheCodes. Biobanks that did not have ICD codes available used spirometry data (Lifelines) or self-reported

data (TWB). Details can be found in Zhou et al.[33] Meta-analyses were also conducted within continental ancestry groups: 19,044 cases and 310,689 controls of EAS ancestry, 1,978 cases and 27,704 controls of AFR ancestry, and 58,559 cases and 937,358 controls of EUR ancestry.

MAGMA

We used Multi-marker Analysis of GenoMic Annotation (MAGMA)[66] v1.09b for gene and gene-set enrichment analyses, applying this method to the GBMI asthma EUR, AFR, EAS, and CSA ancestry-specific meta-analyses (Table S14) and the GBMI COPD EUR, AFR, and EAS ancestry-specific meta-analyses (Table S15). For the gene-level analyses in MAGMA, we first mapped the SNPs to the provided list of genes using a window size of 20kb, and then performed gene analysis using the ancestry-matched 1000G LD reference panels to account for LD structure. Gene-set enrichment was performed using the default settings to correct for gene length, gene density, and the inverse mean minor allele count. The gene sets used were the c2, "curated gene sets," and c5, "ontology gene sets," obtained from the Molecular Signatures Database v7.4[69] (Tables S18 and S19).

DEPICT

We also used Data-driven Expression-Prioritized Integration for Complex Traits (DEPICT),[67] which performs gene prioritization based on correlation of genes across gene sets. We used a 1000G LD reference panel from individuals of EUR ancestry to calculate LD and identify tag SNPs from GWAS results. We report results from the gene prioritization using a p value threshold of 5×10^{-8} and a minimum of 10 index variants. We defined significant enrichment results by FDR <0.05 (Table S16). Full details can be found in Zhou et al.[33]

PoPS

We used another gene prioritization method, Polygenic Priority Score (PoPS), to identify potential causal genes.[68] PoPS performs gene prioritization based on the integration of GWAS data with gene expression, biological pathway, and predicted protein-protein interaction data. We similarly used a 1000G LD reference panel from individuals of EUR ancestry to obtain gene-level associations. Next, MAGMA was applied to integrate the gene-level associations and gene-gene correlations to perform enrichment analysis for gene features selection. Finally, we computed a PoPS score by fitting a joint model with all the selected features simultaneously. We considered genes with a PoPS score in the top one percentile as the prioritized genes (Table S17). Full details can be found in Zhou et al.[33]

QUANTIFICATION AND STATISTICAL ANALYSIS

All statistical analysis was performed using R 4.0.5 and Hail 0.2.98. All methodological details can be found in the STAR Methods section, and all statistical tests are named as they are used.

Supplemental information

Multi-ancestry meta-analysis of asthma identifies

novel associations and highlights the value

of increased power and diversity

Kristin Tsuo, Wei Zhou, Ying Wang, Masahiro Kanai, Shinichi Namba, Rahul Gupta, Lerato Majara, Lethukuthula L. Nkambule, Takayuki Morisaki, Yukinori Okada, Benjamin M. Neale, Global Biobank Meta-analysis Initiative, Mark J. Daly, and Alicia R. Martin
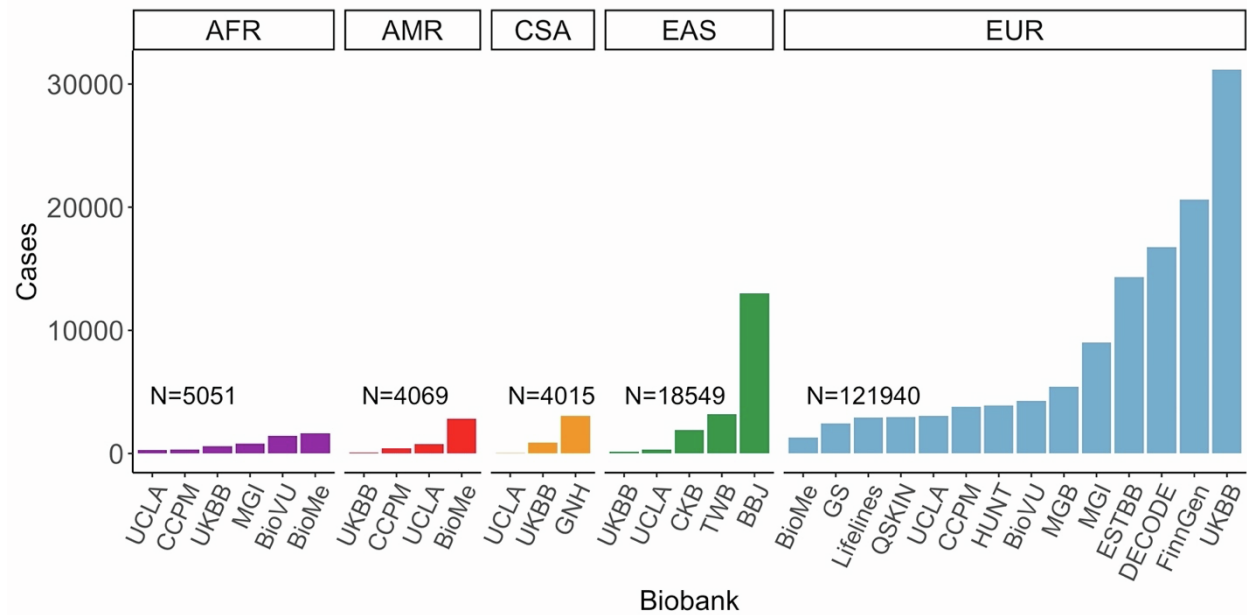
Figure S1. Asthma cases in discovery biobanks stratified by ancestry group, Related to Figure 1 and Table S1. GBMI biobank participants were projected to the same principal components space using pre-computed loadings of genetic markers to compare the genetic ancestries represented in each biobank, indicated on the x-axis. N indicates the total number of cases per ancestry group.
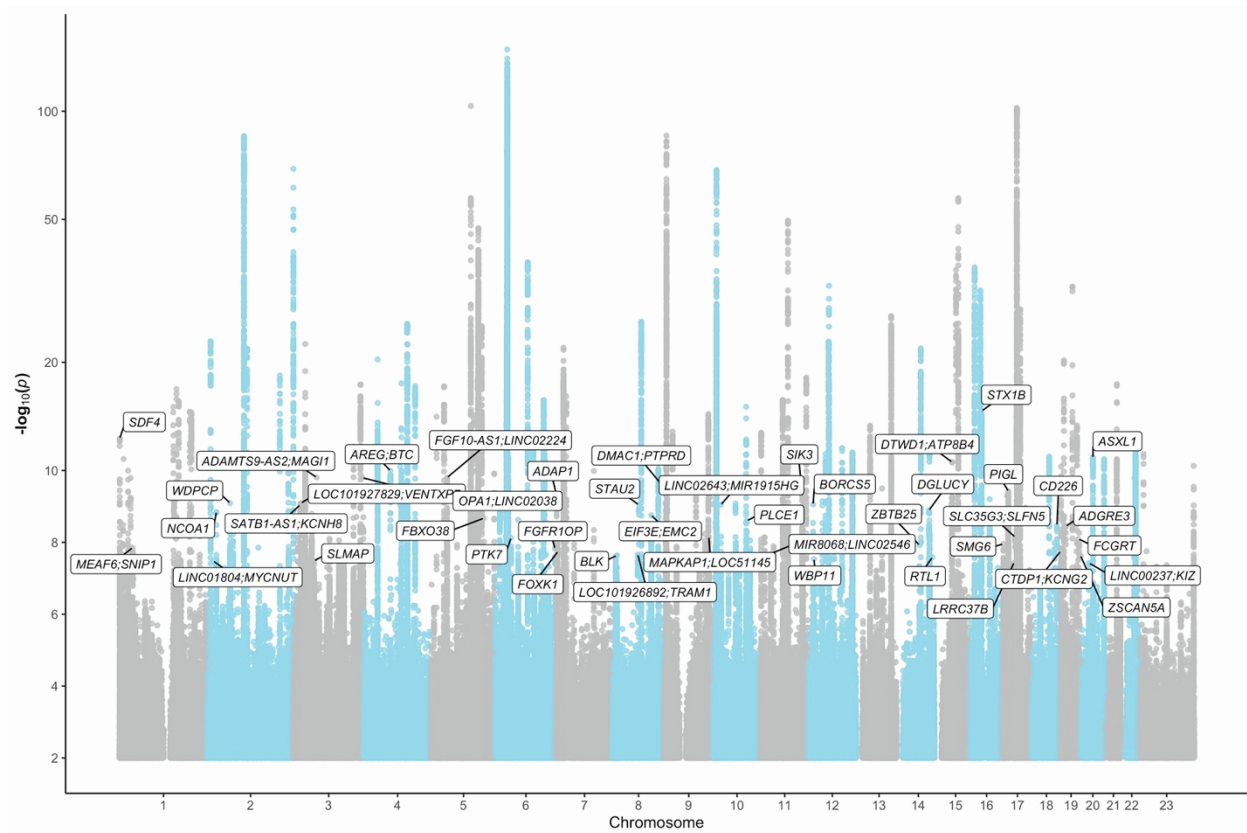
Figure S2. GBMI meta-analysis association results, Related to Figure 2 and Table S2.
Nearest genes to the novel loci are highlighted.
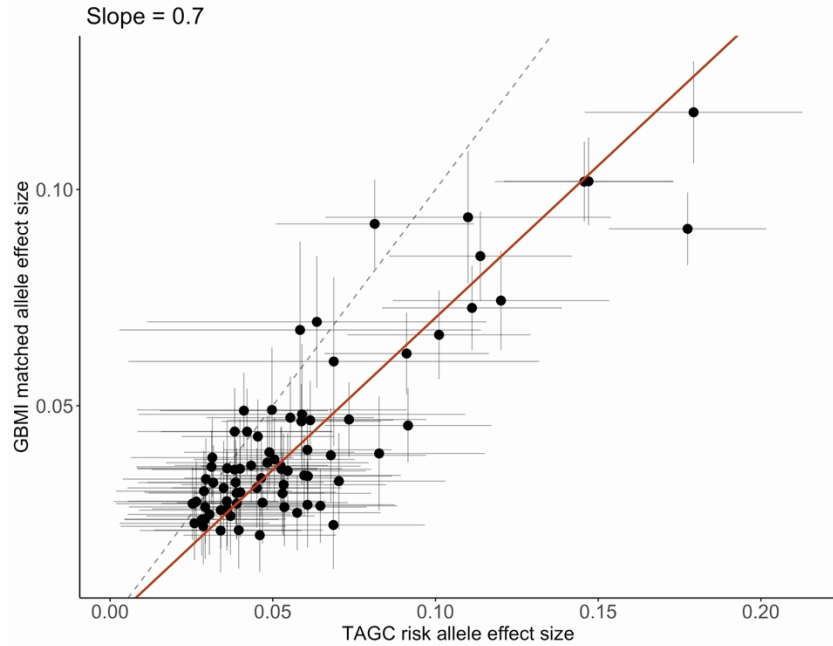
Figure S3. GBMI index variants in TAGC, Related to Figure 2 and Table S4. 76 of the 179 index variants associated with asthma discovered in the GBMI meta-analysis were found in the TAGC meta-analysis of asthma, or had a tagging variant ($r^2 > 0.8$) in the TAGC study, with a p-value < 0.05[9]. The effect sizes of these 76 variants as estimated in the TAGC vs. GBMI meta-analyses were compared using the Deming regression method[35]. The intercept was set to be 0; the slope estimated from the regression analysis is reported here. Error bars represent 95% confidence intervals of the effect size estimates from the corresponding meta-analysis.
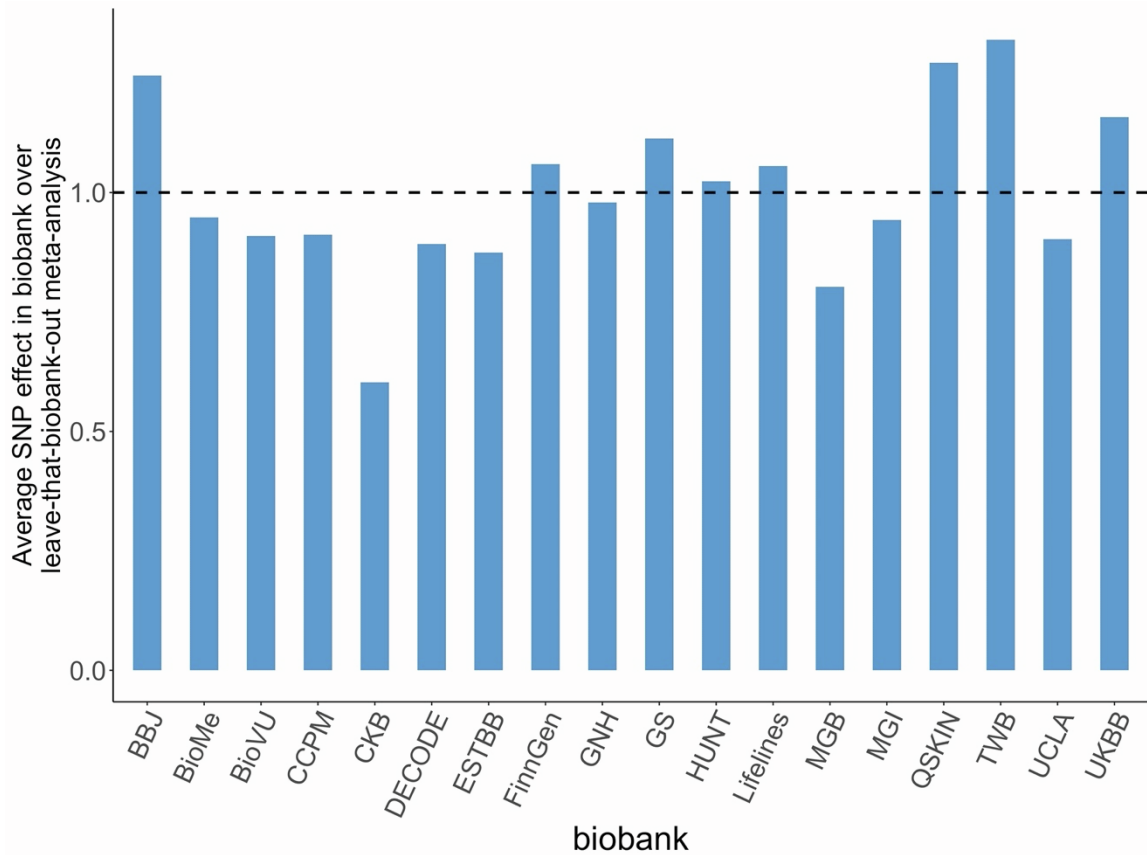
Figure S4. Consistency of asthma index variants across biobanks, Related to Figure 3. For each biobank shown on the x-axis, the average ratio of effect sizes of the index variants in the biobank vs. in the corresponding leave-that-biobank-out meta-analysis is reported here.
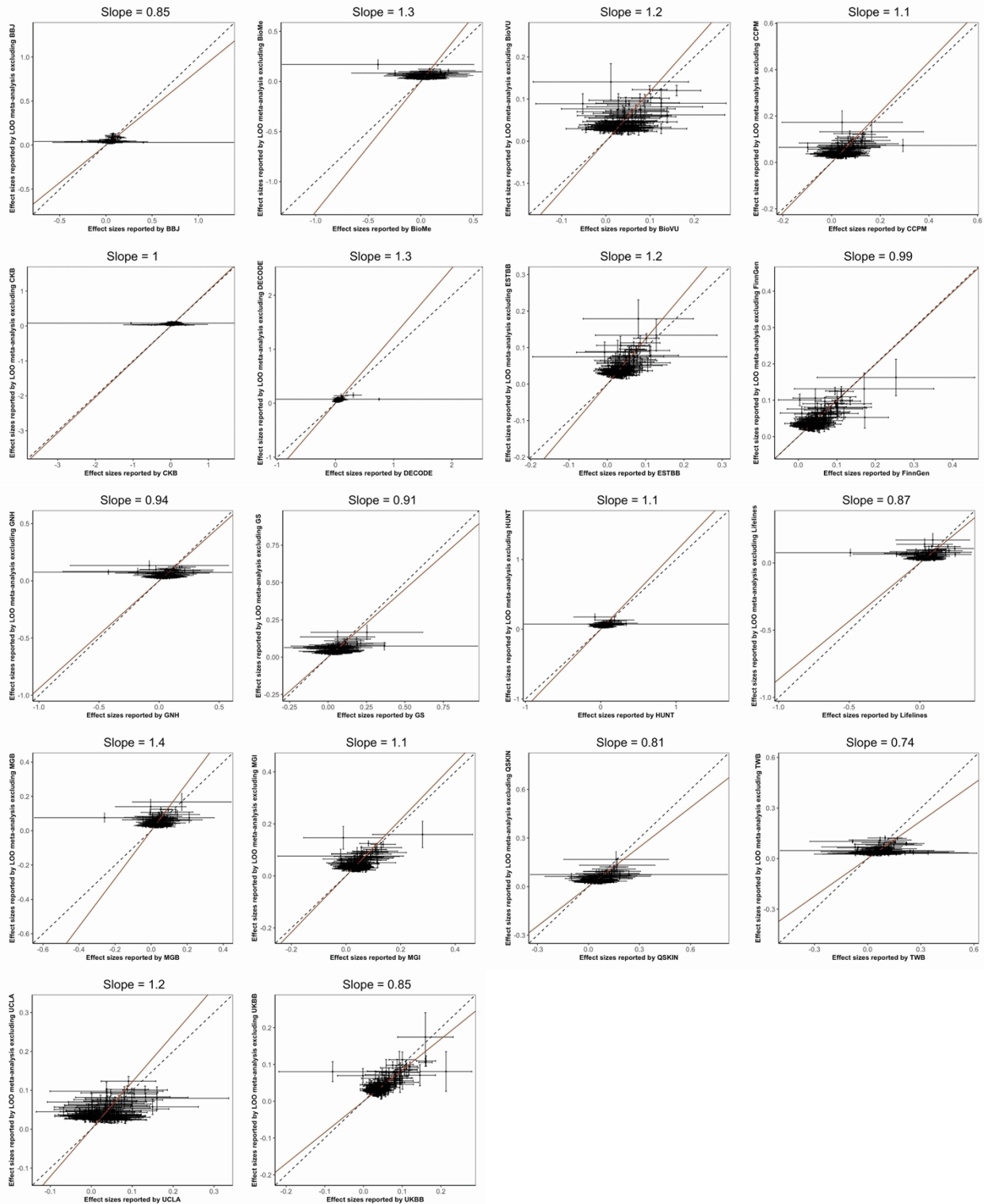
Figure S5. Consistency of asthma index variants across biobanks using Deming regression, Related to Figure 3. The effect sizes of the asthma index variants as estimated in each biobank GWAS vs. in the corresponding leave-that-biobank-out meta-analysis were compared using the Deming regression method[35]. Intercepts were set to be 0; slopes from the regression analyses are reported here. Error bars represent 95% confidence intervals of the effect size estimates from the corresponding meta-analysis or GWAS.
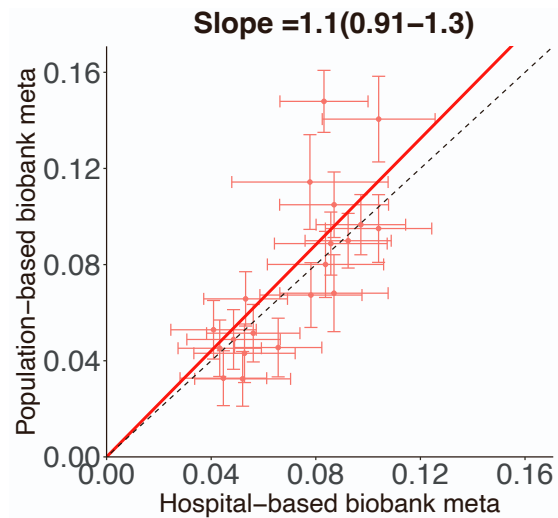
Figure S6. Consistency of asthma index variants across biobanks with different ascertainment, Related to Figure 3. The effect sizes of the asthma index variants as estimated in the meta-analyses of the hospital- vs. population-based biobanks, using the SNPs with p-value < 1×10^{-6} in both meta-analyses, were compared using the Deming regression method[35]. The intercept was set to be 0, and the slope and corresponding 95% confidence interval are reported here. Error bars represent 95% confidence intervals of the effect size estimates from the corresponding meta-analysis.
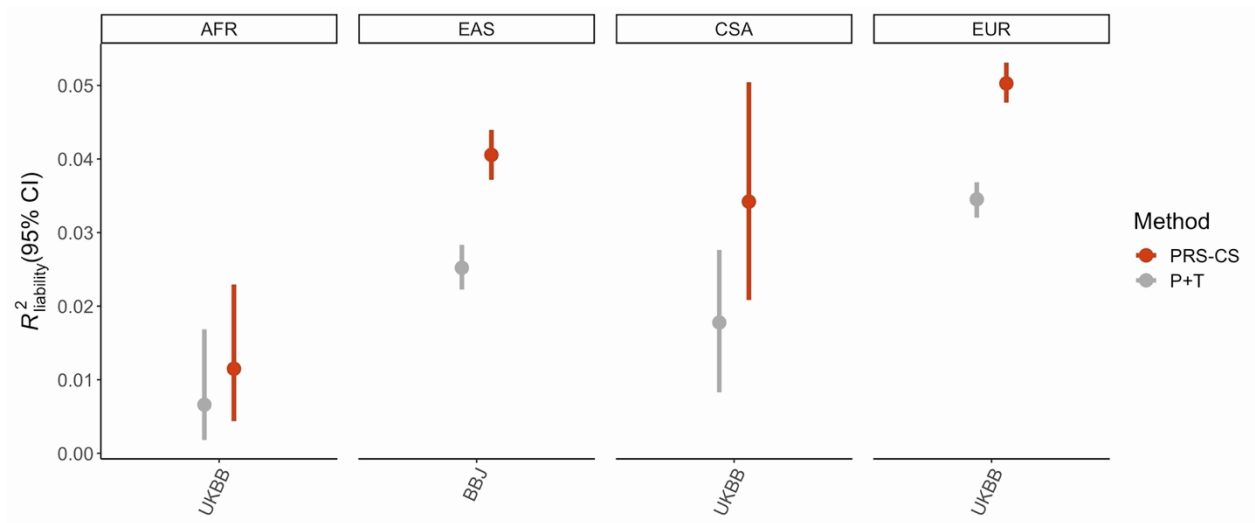
Figure S7. PRS performance of P+T vs. PRS-CS across target cohorts of different ancestries, Related to Figure 5. Each panel represents a target cohort, with the ancestry of the target cohort on the top and the biobank which the target cohort is from on the bottom x-axis. This figure was adapted from Wang et al.[56] Error bars represent the 95% confidence intervals.
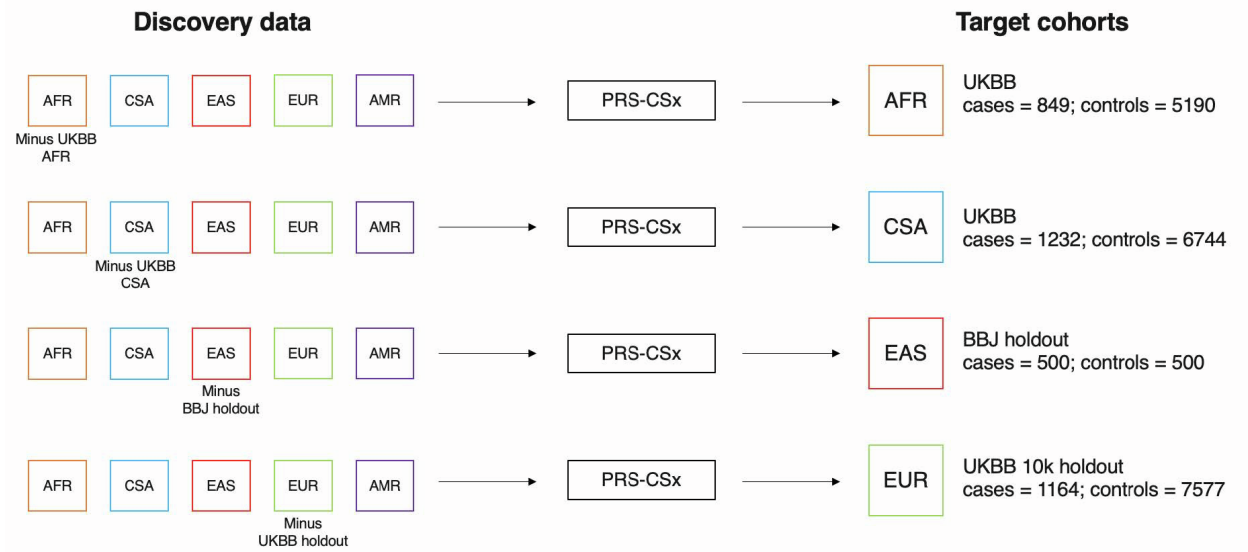
Figure S8. Workflow for PRS-CSx analyses, Related to Figure 5, Figure S9, Figure S10, Table S9, Table S10, and STAR Methods. The discovery data consisted of ancestry-specific meta-analyses, indicated by the squares on the left, that were inputs for PRS-CSx[58]. PRS-CSx returned separate sets of posterior effect size estimates for each input dataset, which were then used to construct PRS. The target cohorts were randomly evenly split; optimal weights for the linear combination of the PRS were selected in one subset and the linear combination of the PRS was evaluated in the other subset.
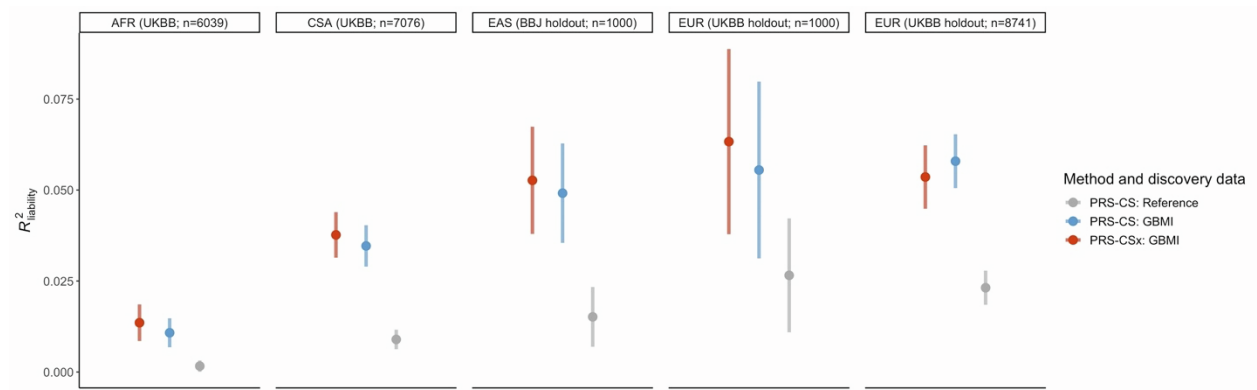
Figure S9. PRS performance in downsampled EUR target cohort, Related to Figure 5, Figure S8, Figure S10, Table S9, and Table S10. Figure 5 is extended here to include results from PRS evaluated in a target cohort of 1,000 randomly selected individuals from the EUR UKBB 10k holdout. Discovery datasets and methods used were the same as described in Figure 5. Error bars represent standard deviation of R² on the liability scale across 100 replicates.
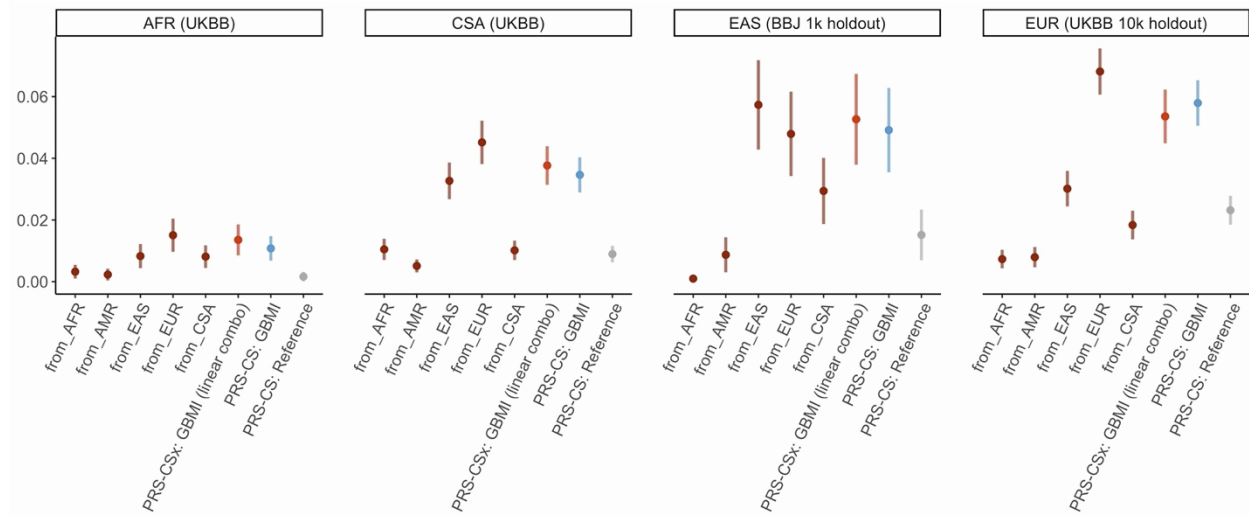
Figure S10. PRS performance of individual PRS vs. linear combination of PRS using PRS-CSx across ancestries, Related to Figure 5, Figure S8, Figure S9, Figure S10, Table S9, and Table S10. Each panel represents a target cohort. The performance of the individual PRS, computed from a single set of posterior effect size estimates corresponding to each input ancestry population from PRS-CSx, is plotted here. The prediction accuracy of the linear combination of the PRS from PRS-CSx, as well as the PRS from the PRS-CS analyses (shown in Figure 5), are also plotted for comparison. PRS-CS results used the GBMI leave-BBJ-out meta-analysis and GBMI leave-UKBB-out meta-analysis as discovery data for the BBJ and all UKBB target cohorts, respectively[56]. The reference dataset was the TAGC meta-analysis[9]. Error bars represent standard deviation of R^2 on the liability scale across 100 replicates.
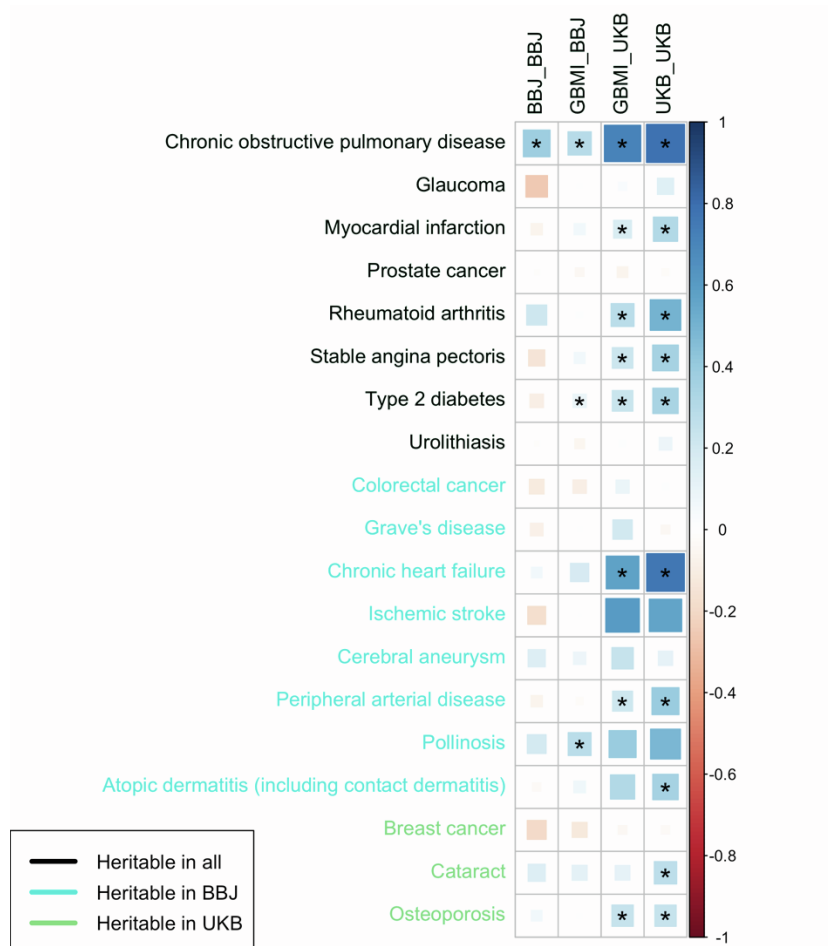
Figure S11. Genetic correlations between asthma and heritable diseases across UKBB and BBJ, Related to STAR Methods, Table S12 and Table S13. Genetic correlations between asthma and disease endpoints that were significantly heritable in BBJ, UKBB EUR, or both. Asterisks indicate genetic correlation estimates with significant p-values at Bonferroni-corrected p-value threshold (p-value < 0.05/20). On x-axis: BBJ_BBJ = BBJ GWAS of asthma vs. BBJ GWAS of diseases on y-axis; GBMI_BBJ = GBMI-excluding-BBJ meta-analysis of asthma vs. BBJ GWAS of diseases on y-axis; GBMI_UKB = GBMI-excluding-UKB meta-analysis of asthma vs. UKB GWAS of diseases (EUR only) on y-axis; UKB_UKB = UKB GWAS of asthma vs. UKB GWAS of diseases (EUR only) on y-axis.

Data S1: GBMI Consortium Biobank and Cohort Acknowledgements

144623

The HUNT Study

Hail is an open-source Python library that simplifies genomic data analysis in the cloud. It provides powerful, easy-to-use data science tools that can be used to interrogate biobank-scale genomic data and was used in the analysis of the data for this paper. We would especially like to thank Daniel King from the Hail team and Sam Bryant from the Stanley Center Data Management team for helping with the Google bucket set up and data sharing.
Website for Hail can be found here: https://hail.is/