Cell Genomics Transparent Peer Review Record

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

Kristin Tsuo[1,2,3,*], Wei Zhou[1,2], 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,*], Alicia R. Martin[1,2,3,*]

Summary

Initial submission: Received : 12/06/2021

Scientific editor: Orli Bahcall, Emily Marcinkevicius, Laura Zahn

First round of review: Number of reviewers: 3
Revision invited : 3/2/2022
Revision received : 9/1/2022

Second round of review: Number of reviewers: 3
Accepted : 10/12/2022

Data freely available: Yes

Code freely available: Yes

Referees' reports, first round of review

Reviewer #1: Tsuo et al. reported a study on investigating multi-ancestry meta-analysis of asthma GWAS. In general, this is a well-conducted study with many several strengths, including the largest asthma GWAS sample size to date, including multi-ancestry populations, identifying novel loci, ancestry-specific analysis. I have several comments:
Major comments:
1. It's better to conduct conditional analysis to show whether the novel loci are independent from the nearby known loci. The authors can use GCTA-COJO for this.
2. Please provide a Manhattan plot as a main figure to show the asthma GWAS results, and highlighting the novel loci.
3. It is important to know that asthma is a very heterogeneous disease, thus, GWAS of asthma is not always the bigger the better, by contrast, it's important to stratify the analysis by asthma subtypes. I understand this is not easy for the current study, since this study depends on meta-analysis of GWAS summary statistics from many different studies, which provided asthma phenotype, not asthma subtypes. If possible, I would like the authors conduct asthma subtype GWAS analysis. If not possible, the authors should discuss this topic in the discussion section.
4. GWAS of asthma in non-European population is important to know the population specific loci for asthma. From Figure 1, I see a few EAS populations, such as BBJ and CKB, which become more popular in recent years given large sample size and unique genetic background. Please also make a discussion about the future of asthma GWAS in non-European populations, especially Asian populations since large-scale GWAS resources are becoming available. The authors can use asthma-related disease/traits as the examples to discuss, see following papers.
Tanaka N, Koido M, Suzuki A, Otomo N, Suetsugu H, Kochi Y, Tomizuka K, Momozawa Y, Kamatani Y; Biobank Japan Project, Ikegawa S, Yamamoto K, Terao C. Eight novel susceptibility loci and putative causal variants in atopic dermatitis. J Allergy Clin Immunol. 2021;148(5):1293-1306. PMID: 34116867.
Zhu Z, Li J, Si J, Ma B, Shi H, Lv J, Cao W, Guo Y, Millwood IY, Walters RG, Lin K, Yang L, Chen Y, Du H, Yu B, Hasegawa K, Camargo CA Jr, Moffatt MF, Cookson WOC, Chen J, Chen Z, Li L, Yu C, Liang L. 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. 2021;58(4):2100199. PMID: 33766948; PMCID: PMC8513692.
5. Figure 4 results are interesting. In addition to the chr16 locus, it's better to include 17q21. Please check the lead SNP in 17q21 for ancestry heterogeneity.
6. Supplementary Figure 4: Some of the LDSC results do not make too much sense. For example, it is strange for me to see no genetic correlation between asthma and atopic dermatitis in the first 3 columns, and the 4th column UKB results showed significant but not very strong correlation (color bar shows Rg is around 0.25-0.3?) Also, it shows asthma has strong genetic correlation with chronic heart failure, peripheral artery disease, and rheumatoid arthritis? The authors need to double check with this analysis and provide reasonable justification.

Reviewer #2: This manuscript provides important information and reports on the largest and most diverse collection of asthma biobanks to date and includes 18 biobanks with GWAS data. However, the manuscript is a bit ambitious using multiple tools: meta-analysis, PRS, and genetic correlation. Because so many analyses are mentioned, the manuscript lacks details for full understanding of the analyses. The authors could consider deleting the genetic correlation analyses in order to have space to more fully describe the meta-analysis and PRS.

Major comments

1. The investigators could address if the higher predictive power for asthma in non-EUR populations because of the smaller sample sizes.
2. More details of the biobanks are needed. Providing the sample size for cases and controls for each of the races would help understand the reader interpret the results.
3. It's not clear what the major outcome for the asthma PRS is. The manuscript mentions 14 phenotypes analyzed in GBMI, but it's unclear what these phenotypes are.
4. The investigators acknowledge the large variability in asthma prevalence across the biobanks, however the investigators could make clear how they accounted for this variability.
5. What is the average prediction accuracy of the PRS? This is not reported in the Results.
6. The Results state that these analyses suggest genetic architecture of asthma is largely shared across diverse cohorts. Does this mean that diverse populations are not needed for future research? This is probably not the authors conclusion, I think.
7. The genetic correlation analysis with all of the phenotypes in the GBMI biobanks is not needed in this paper. It's unclear what the goal of examining genetic correlations between asthma and all of the heritable diseases is. It appears that chronic heart failure and asthma are correlated—explanations for this could be provided.
8. There is no independent replication population. This is a limitation even though the authors use the leave on out approach.
9. The authors could justify their fixed effects meta-analysis with inverse variance weighting. It seems a random effects meta-analysis may have been more appropriate.
10. In the Discussion, the investigators could provide explanation for why excluding COPD would introduce an additional source of ascertainment bias.
11. Discussion. The manuscript states that Bayesian PRS construction methods can improve prediction in asthma. The authors could explain how prediction is improved and discuss their PRS scores in the context of other asthma PRSs reported on to date.
12. It's not clear if ancestry markers were used at all in these analyses.
13. The authors could mention that they are not able to account for BMI.
14. One of the limitations that the authors appropriately address is the heterogeneity in the phenotype. The authors could consider conducting sensitivity analysis to see if there are differences whether the cases were defined based on self-report (Taiwan Biobank, QSkin) compared to PheCodes or ICD-9 codes.
15. Another limitation is it seems the investigators were unable to distinguish between child-onset and adult-onset asthma. Limiting to adult-onset asthma could help their analysis.
16. The authors conclude that the "genetic effects of associated loci are largely consistent across the biobanks and ancestries." How do the authors put this in the setting of multiple other studies that suggest that genetic effects vary by genetic ancestry?
17. The manuscript does not mention how ancestry groups are defined. This could be added to the supplemental tables. Was self-report used or ancestry markers?
Minor comments:
1. Summary: "Despite the considerable range in prevalence…" Consider adding "of asthma" after "prevalence."
2. Introduction: Need reference
3. Need to write out GWAS the first time in the Introduction
4. Introduction "efforts to diversify asthma" is not clear. Maybe the authors mean "efforts to conduct asthma GWAS in diverse populations."
5. Introduction: Add "initiative" after "Global Biobank Meta-analysis."
6. Consider using the term "allergic rhinitis" in addition to "hay fever."

Reviewer #3: Comments enter in this field will be shared with the author; your identity will remain anonymous.

This is an interesting study representing a huge effort combining multiple biobank based GWASes resulting in the largest asthma GWAS to date.
The study identifies novel susceptibility genes, adding to the understanding of the genetic mechanisms of asthma, and provides novel knowledge on the potential of using large, herterogenous datasets in genetic studies of asthma.
The manuscript is well written and the methodologies seem sound.

I am not sure, calculating gene-based p-values using MAGMA can truly be termed 'gene prioritization". It might be more appropriate to use actual gene prioritization methods such as DEPICT or PoPS which incorporate biological features to search for patterns of shared biology to prioritize genes at GWAS loci.

Comparing gene-based results from MAGMA is not very different from directly comparing summary stats. In order to compare prioritized genes for e.g. asthma and COPD, I think other methods such as DEPICT are more suitable.

The high genetic correlation with COPD might reflect a relatively high age of participants (high age of asthma diagnosis), which is likely to increase the overlap between diagnoses of asthma and COPD compared to asthma studies including younger individuals. How did participant age compare to previous studies reporting lower genetic correlation of asthma and COPD?

Data on participant age should be provided for the individual studies.

Authors' response to the first round of review

We thank the reviewers and editors for their insightful comments and suggestions. Our responses to each comment are below in purple, and quoted text from the manuscript is italicized and in gray. In the manuscript, added and revised text is in purple.

Reviewer 1

Tsuo et al. reported a study on investigating multi-ancestry meta-analysis of asthma GWAS. In general, this is a well-conducted study with many several strengths, including the largest asthma GWAS sample size to date, including multi-ancestry populations, identifying novel loci, ancestry-specific analysis. I have several comments:

We thank the reviewer for these positive comments and respond to each comment below individually.

1. It's better to conduct conditional analysis to show whether the novel loci are independent from the nearby known loci. The authors can use GCTA-COJO for this.

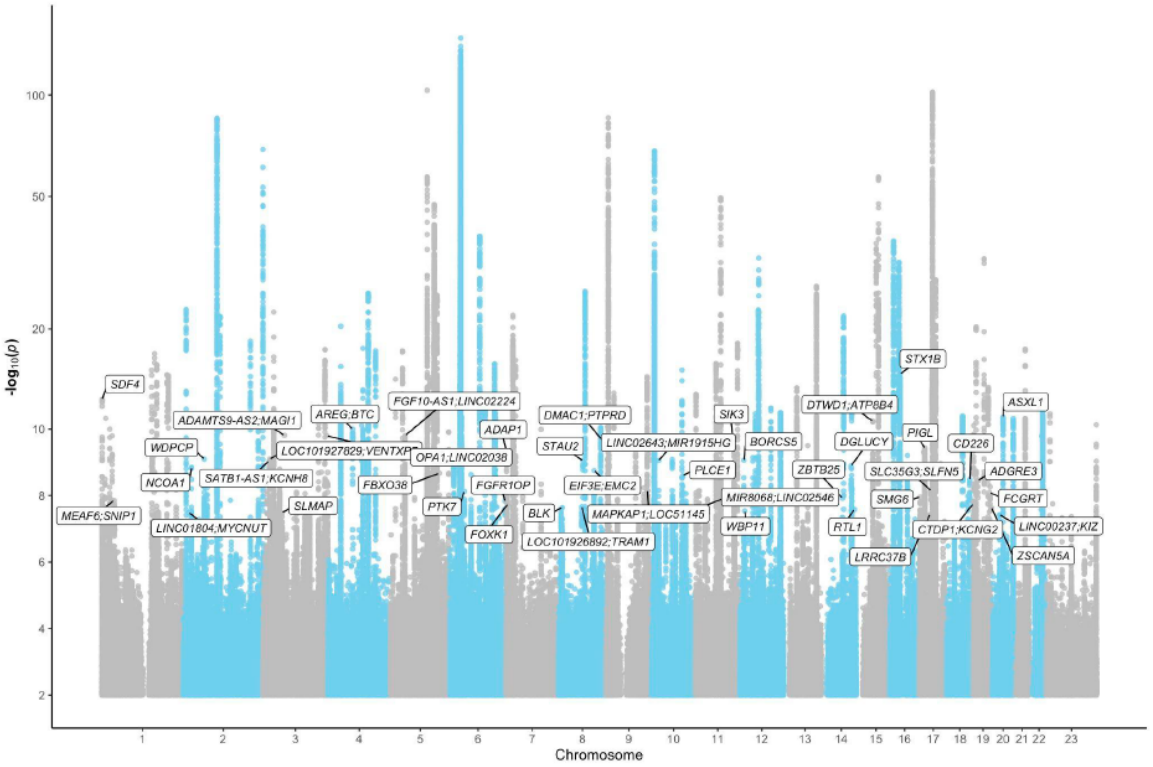
We thank the reviewer for this suggestion. We note that to designate loci as potentially novel, we compiled a list of previously discovered asthma-associated variants ($p < 5x10^{-8}$) from El-Husseini et al. (2020)[1] and as listed in the GWAS catalog (as of 11/14/2021), extended 500kb upstream and downstream each of these variants to define loci, and intersected these with the top loci discovered in the GBMI meta-analysis. Thus, the index variants of the potentially novel loci are at least 1Mb in distance from a previously discovered variant. Additionally, we computed LD using a reference panel from individuals in 1000 Genomes between the potentially novel SNPs and the index variants of each previously discovered loci, defined as the variant with the strongest association in the GBMI meta-analysis. All potentially novel index variants, with the exception of one ($r^2 = 0.16$), had $r^2 < 0.07$ with a previously known SNP (Supplementary Table 2). This provides strong evidence that the novel loci are likely independent from the nearby known loci. Although conditional analysis using GCTA-COJO would account for potential long-range LD, there is currently no gold standard for conducting conditional analysis on meta-analysis data. Given the ancestral diversity of the GBMI meta-analysis, it is unclear which LD reference panel would be appropriate to use for conditional analysis. Lastly, Kanai et al. (2022)[2] shows that applying existing fine-mapping methods to meta-analysis data results in substantial miscalibration due to heterogeneity across biobanks, and thus we expect that applying existing conditional analysis methods to the GBMI data would similarly yield miscalibrated results. We added an explanation of how we defined the potentially novel loci in the "Multi-ancestry meta-analysis for asthma across 18 biobanks in GBMI" Results section:

The meta-analysis identified 179 loci of genome-wide significance ($p < 5x10^{-8}$), 49 of which have not

been previously reported to be associated with asthma (Fig. 2A, Supplementary Fig. 2). These potentially novel loci were defined so that the index variants, or the most significant variants in each locus, were at least 1 Megabase in distance from a previously discovered genome-wide significant variant associated with asthma (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] (Supplementary Table 2).

2. Please provide a Manhattan plot as a main figure to show the asthma GWAS results, and highlighting the novel loci.

We have included a Manhattan plot highlighting the nearest genes to the 49 novel lead loci (Supplementary Figure 2). We included this in the "Multi-ancestry meta-analysis for asthma across 18 biobanks in GBMI" Results section as follows:
The meta-analysis identified 179 loci of genome-wide significance ($p < 5x10\text{-}8$), 49 of which have not been previously reported to be associated with asthma (Fig. 2A, Supplementary Fig. 2).



Supplementary Figure 2. GBMI meta-analysis association results. *Nearest genes to the novel loci are highlighted.*

3. It is important to know that asthma is a very heterogeneous disease, thus, GWAS of asthma is not always the bigger the better, by contrast, it's important to stratify the analysis by asthma subtypes. I understand this is not easy for the current study, since this study depends on meta-analysis of GWAS summary statistics from many different studies, which provided asthma phenotype, not asthma subtypes. If possible, I would like the authors conduct asthma subtype GWAS analysis. If not possible, the authors should discuss this topic in the discussion section.

We thank the reviewer for this suggestion and agree that subtype analysis is particularly important for asthma, given its heterogeneity. We were able to conduct asthma age-of-onset subtype analyses in two of the participating GBMI biobanks, UKBB and FinnGen. We performed GWAS of childhood-onset asthma (COA) and adult-onset asthma (AOA) in Finngen and the EUR ancestry cohort in UKBB, using a cut-off age of 19 years at asthma diagnosis to define the subtypes (Methods). Then, we conducted fixed-effects, inverse-variance weighted meta-analyses of the COA (20,964 cases, 674,014 controls) and AOA (56,744 cases, 674,014 controls) GWAS, respectively. We added the results of these analyses to the "Childhood-onset (COA) and adult-onset (AOA) asthma have high genetic correlations with all-asthma meta-analysis" section as follows:

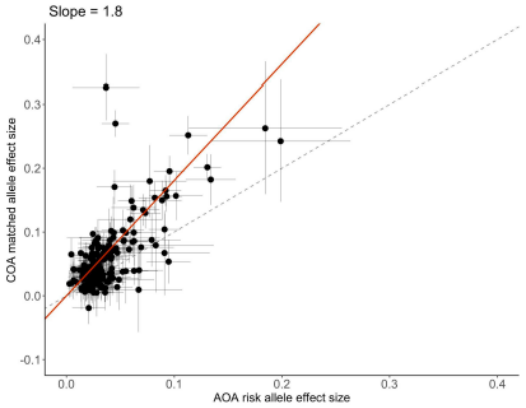
To increase power for genetic discovery, we used a broad phenotype definition for asthma (Methods), but given the heterogeneity of the disease, we sought to address the extent to which this meta-analysis captured the genetic architectures of two common subtypes of asthma, childhood-onset (COA) and adult-onset (AOA) asthma. 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 cut-off age of 19 years at asthma diagnosis to define the subtypes (Methods), we performed GWAS 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, 674,014 controls) and AOA (56,744 cases, 674,014 controls) GWAS, respectively. Applying linkage-disequilibrium score correlation (LDSC), 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[59]. 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 ($p < 5 \times 10^{-8}$) 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 (Supplementary Table 11). Overall, these results suggest that much of the genetic architecture between COA and AOA is shared, as is consistent with previous findings[59,60]. 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 to in the AOA meta-analysis (Supplementary Fig. 11), supporting previous findings that the etiology of COA is likely partially characterized by genes that have smaller (or no) effects on AOA[59,60]

Broadly, the results indicated that the GBMI meta-analysis containing both age-of-onset subtypes captured many of the genetic variants contributing to the subtypes, but additional stratification by subtype revealed that the etiology of COA is characterized by a substantial overlap with AOA, as well as genes that have smaller (or no) effects on AOA. This supports previous findings[3,4]. Due to the limited availability of age of onset information across the biobanks, we did not have as much power to identify

potential subtype-specific associations, which is an important area for future investigation and noted in the Discussion, as follows:

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 3 GBMI biobanks) but did not reach genome-wide significance in the meta-analysis (Supplementary Table 11). Due to the limited availability of age at first diagnosis information 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.



Supplementary Figure 11. Effect size estimates of asthma index variants in COA vs. AOA meta-analyses. *The effect sizes of the 179 index variants discovered in the all-asthma meta-analysis as estimated in the COA vs. AOA 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.*

4. GWAS of asthma in non-European population is important to know the population specific loci for asthma. From Figure 1, I see a few EAS populations, such as BBJ and CKB, which become more popular in recent years given large sample size and unique genetic background. Please also make a discussion about the future of asthma GWAS in non-European populations, especially Asian populations since large-scale GWAS resources are becoming available. The authors can use asthma-related disease/traits as the examples to discuss, see following papers.
Tanaka N, Koido M, Suzuki A, Otomo N, Suetsugu H, Kochi Y, Tomizuka K, Momozawa Y, Kamatani Y; Biobank Japan Project, Ikegawa S, Yamamoto K, Terao C. Eight novel susceptibility loci and putative causal variants in atopic dermatitis. J Allergy Clin Immunol. 2021;148(5):1293-1306. PMID: 34116867.
Zhu Z, Li J, Si J, Ma B, Shi H, Lv J, Cao W, Guo Y, Millwood IY, Walters RG, Lin K, Yang L, Chen Y, Du H, Yu B, Hasegawa K, Camargo CA Jr, Moffatt MF, Cookson WOC, Chen J, Chen Z, Li L, Yu C, Liang L. 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. 2021;58(4):2100199. PMID: 33766948; PMCID: PMC8513692.

We thank the reviewer for this suggestion. In addition to the section in "GWAS from diverse ancestries reveal shared genetic architecture of asthma and improves power for genetic discovery" highlighting the several putative population-specific asthma loci that have heterogeneous effects across ancestry groups, we have added an additional paragraph in this section of the Results explicitly addressing the added benefits of including non-European populations in the meta-analysis:

Additionally, the greater diversity of 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 (Supplementary Table 8). This additional yield of loci may be partially due to the increase in sample size, but the inclusion of GWAS 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 to the EUR ancestry group. The consistent effect estimates of the 49 additional variants across populations (45/49 had p-value for Cochran's Q test across ancestries >0.02) indicate that the additional variants discovered with the incorporation of GWAS from diverse ancestries do not tend to be population-specific loci that only have effects in certain populations. However, due to differences in frequency across populations, it is essential to conduct asthma GWAS in different populations to uncover the full spectrum of asthma-associated loci.

We also added a section to the Discussion:

Importantly, however, the addition of GWAS 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 GWAS 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 also found highly shared polygenic architecture between populations, mirroring our findings for asthma[83,84].

As we discuss in these sections, it is particularly important to include GWAS from diverse cohorts to uncover the full spectrum of genetic variants associated with asthma. We show that with the addition of non-EUR ancestry GWAS, we were able to identify 49 asthma-associated variants that did not reach genome-wide significance in the EUR-only meta-analysis (Supplementary Table 8). 19 of these loci were potentially novel, and 13 of these 19 loci had an index variant higher in frequency in a non-EUR ancestry group compared to the EUR ancestry group. As the availability of non-EUR GWAS of asthma continues to increase, it is likely that additional variants associated with asthma that are higher in frequency in other populations will be discovered.

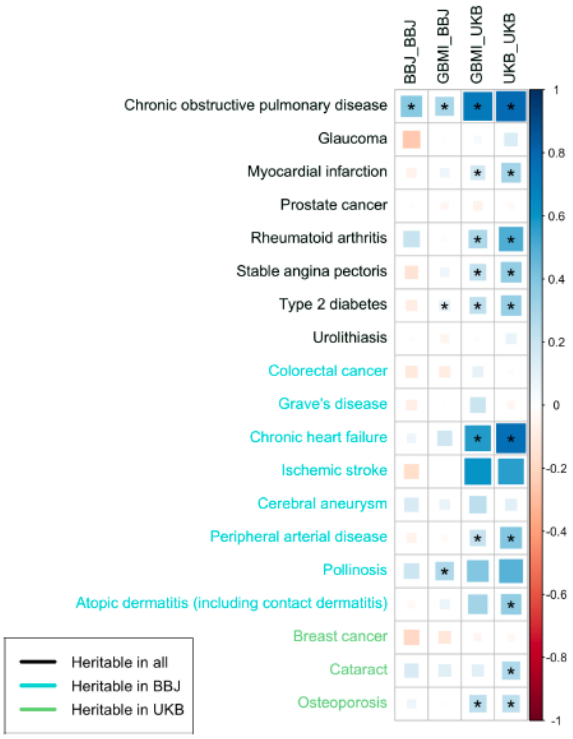
5. Figure 4 results are interesting. In addition to the chr16 locus, it's better to include 17q21. Please check the lead SNP in 17q21 for ancestry heterogeneity.

We thank the reviewer for highlighting this important locus. chr17:39907128:C:T was identified in the meta-analysis as the lead SNP in the 17q21 locus. This SNP lies within GSDMB, which has previously been linked to asthma. However, this SNP did not exhibit heterogeneity in its effects across the

ancestry-specific meta-analysis, with a heterogeneity p-value of 0.17. This SNP did have different allele frequencies across the ancestry groups; for example it is almost 3 times as frequent in EUR than in AFR (in AFR AF = 17.7%; in AMR AF = 36.6%; in CSA AF = 36.5%; in EAS AF = 26.4%; in EUR AF = 52.0%). The overall consistency of effects but difference in allele frequencies across different populations at this locus align with previous observations[8], but as Stein et al. (2018) noted, there may be population-specific rare variants at this locus associated with asthma risk that have been missed. Unfortunately, rare variant analyses are outside of the scope of this study.

6. Supplementary Figure 4: Some of the LDSC results do not make too much sense. For example, it is strange for me to see no genetic correlation between asthma and atopic dermatitis in the first 3 columns, and the 4th column UKB results showed significant but not very strong correlation (color bar shows Rg is around 0.25-0.3?) Also, it shows asthma has strong genetic correlation with chronic heart failure, peripheral artery disease, and rheumatoid arthritis? The authors need to double check with this analysis and provide reasonable justification.

Thank you for pointing out these unexpected results. We note that the atopic dermatitis, chronic heart failure, and peripheral artery disease UKBB GWAS did not have statistically significant SNP heritability estimates, and thus the genetic correlation estimates between asthma and these diseases in UKBB are difficult to interpret. Additionally, the atopic dermatitis PheCode used in BBJ and in UKBB encompasses contact dermatitis, which may contribute to the low genetic correlation observed. We have updated Supplementary Table 13 and Supplementary Figure 12 to more clearly indicate the disease endpoints with significant heritability estimates in both UKBB and BBJ, as shown below:



Supplementary Figure 12. Genetic correlations between asthma and heritable diseases across UKBB and BBJ. *Genetic correlations between asthma and disease endpoints that were significantly heritable in BBJ, UKBB EUR, or both. 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*

The rheumatoid arthritis GWAS in UKBB and BBJ have non-zero heritability estimates, and we also find a relatively high genetic correlation between rheumatoid arthritis and asthma in FinnGen (r_g (se) = 0.375 (0.1), p = 2x10^{-4}). Several studies in the literature have reported a relationship between risk for asthma and rheumatoid arthritis[9–14], but more genetic studies in different populations are needed to investigate the potential shared genetic architecture of these diseases. Significant genetic correlations between asthma and heritable disease endpoints that tend to affect older adults, such as myocardial infarction, could also potentially be driven by the likely predominance of AOA in GBMI. Studies investigating causality are needed here to better understand these relationships. We have revised the description of these genetic correlation results in the Results:

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 (Supplementary Table 13). COPD showed the strongest and most significant correlation with asthma (r_g = 0.29, p = 6.41x10^{-6}), but the notably lower estimate compared to 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 = 4.60x10^{-3}). Comparing the phenotypes with significant SNP heritability estimates in both BBJ and UKBB (Supplementary Fig. 12), we found that only COPD has significant genetic correlations with asthma across the biobanks. The rheumatoid

arthritis (RA) and type 2 diabetes (T2D) GWAS from UKBB have moderate and significant correlations with asthma, which are partially recapitulated in the BBJ results that showed a moderate but not significant correlation between the BBJ GWAS of RA and of 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[73–78], as well as T2D and asthma[79–81], 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[82].

Reviewer 2

This manuscript provides important information and reports on the largest and most diverse collection of asthma biobanks to date and includes 18 biobanks with GWAS data. However, the manuscript is a bit ambitious using multiple tools: meta-analysis, PRS, and genetic correlation. Because so many analyses are mentioned, the manuscript lacks details for full understanding of the analyses. The authors could consider deleting the genetic correlation analyses in order to have space to more fully describe the meta-analysis and PRS.

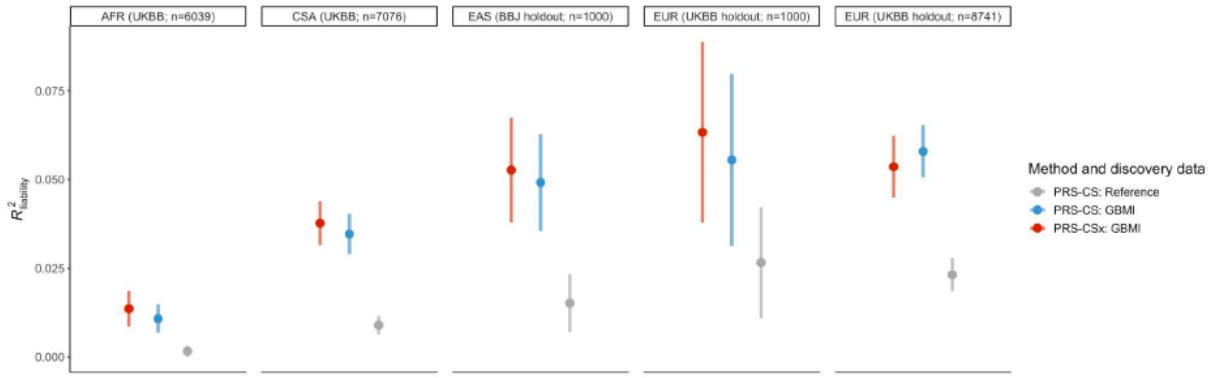
We thank the reviewer for this feedback. We have restructured the genetic correlation sections to cover additional in-depth interpretations of the genetic relationships between asthma subtypes (childhood-onset and adult-onset asthma), asthma and known comorbid diseases (with a focus on COPD), and other comorbid diseases that have been less explored in genetic studies. We have also described the meta-analysis findings and PRS analyses in more detail. We respond to each comment below individually.

Major comments

1. The investigators could address if the higher predictive power for asthma in non-EUR populations because of the smaller sample sizes.

Thank you for pointing this out. To evaluate the effects of the target cohort sample size on the predictive performance of the PRS, we downsampled the EUR target cohort to 1,000 individuals, and found that the average $R^2_{liability}$ was higher than in the EAS target cohort but the confidence intervals as expected were quite large, similar to the confidence intervals of the $R^2_{liability}$ in the BBJ cohort (Supplementary Table 9, Supplementary Fig. 9). Furthermore, we note that phenotype heterogeneity and differences in the precision of phenotype definition used, as well as other factors such as environmental exposures, demographic history, and recruitment strategy of the biobanks, could potentially contribute to differences in predictive power across cohorts. We added these results as below:

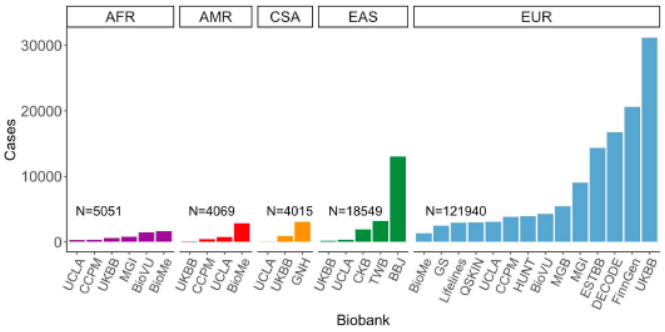
When we downsampled the EUR target cohort to 1,000 individuals, to match the sample size of the EAS target cohort, we found a higher average $R^2_{liability}$ (0.063) but, as expected, much larger confidence intervals (Supplementary Fig. 9).



Supplementary Figure 9. PRS performance in downsampled EUR target cohort. *Fig. 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 Fig. 5.*

2. More details of the biobanks are needed. Providing the sample size for cases and controls for each of the races would help understand the reader interpret the results.

Supplementary Table 1 lists the number of cases and controls for each genetic ancestry group within each biobank. To more clearly present the sample size per ancestry group, we have added an additional supplementary figure (Supplementary Fig. 1) that indicates the number of cases per ancestry



Supplementary Figure 1. Asthma cases in discovery biobanks stratified by ancestry group. *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.*

3. It's not clear what the major outcome for the asthma PRS is. The manuscript mentions 14 phenotypes analyzed in GBMI, but it's unclear what these phenotypes are.

We apologize for the confusion. We have listed in the Introduction which 14 endpoints were analyzed in GBMI, as follows:

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. (2021)[33].

We have also clarified in the Results that we implemented PRS-CS for asthma as well as the other endpoints (see Wang et al. (2021)[19]), and PRS-CSx for asthma specifically, as follows:

To establish a baseline understanding of PRS performance for asthma as well as other disease endpoints in GBMI, Wang et al. (2021)[56] evaluated and compared the prediction accuracy of PRS derived from the pruning and thresholding (P+T) method and PRS-CS[57] in target cohorts of EUR, CSA, EAS, and AFR ancestries, using the leave-one-biobank-out meta-analyses as discovery data.

Our main conclusion from the PRS analyses is that increase in scale and diversity of discovery GWAS 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. We did not observe improvement in PRS accuracy using PRS-CSx vs. PRS-CS in the EUR target cohort, likely because the sample sizes of the EUR discovery GWAS already predominate and including GWAS from smaller, non-EUR discovery cohorts may introduce more noise than signal. We have updated the PRS Results section of the manuscript to clarify this, as follows:

Collectively, these analyses show that the increase in scale and diversity of discovery GWAS 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 EUR target cohorts, a multi-ancestry PRS construction method like PRS-CSx does not seem to contribute much improvement in prediction accuracy, likely due to the predominating sample size of EUR discovery GWAS, as well as the inclusion of GWAS from smaller, non-EUR discovery cohorts which may introduce more noise than signal.

4. The investigators acknowledge the large variability in asthma prevalence across the biobanks, however the investigators could make clear how they accounted for this variability.

We thank the reviewer for noting this important point for clarification. In the comparisons of the lead SNP effects in each biobank GWAS vs. the corresponding leave-that-biobank-out meta-analysis, we observed that the SNP effects were well aligned across the biobanks, despite variability in prevalence and other characteristics. Therefore, in this meta-analysis the large variation in asthma prevalence across the biobanks does not seem to significantly affect the resulting genetic discoveries.

5. What is the average prediction accuracy of the PRS? This is not reported in the Results.

Thank you for pointing this out and we apologize for the oversight. We have updated the PRS section of the Results to include the average prediction accuracy of the PRS in each target population, and added Supplementary Table 9 with these results as well.

6. The Results state that these analyses suggest genetic architecture of asthma is largely shared across diverse cohorts. Does this mean that diverse populations are not needed for future research? This is probably not the authors conclusion, I think.

Thank you for the comment. We have now added a section in “GWAS from diverse ancestries reveals shared genetic architecture of asthma and improves power for genetic discovery” of the Results part of the manuscript explicitly addressing the added benefits of studying diverse populations in genetic studies of asthma, as follows:

Additionally, the greater diversity of 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 (Supplementary Table 8). This additional yield of loci may be partially due to the increase in sample size, but the inclusion of GWAS 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 to the EUR ancestry group. The consistent effect estimates of the 49 additional variants across populations (45/49 had p-value for Cochran's Q test across ancestries >0.02) indicate that the additional variants discovered with the incorporation of GWAS from diverse ancestries do not tend to be population-specific loci that only have effects in certain populations. However, due to differences in frequency across populations, it is essential to conduct asthma GWAS in different populations to uncover the full spectrum of asthma-associated loci.

We also added a section to the Discussion:

Importantly, however, the addition of GWAS 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 GWAS 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 also found highly shared polygenic architecture between populations, mirroring our findings for asthma[83,84].

As we discuss in these sections, it is particularly important to include GWAS from diverse cohorts to uncover the full spectrum of genetic variants associated with asthma. We show that with the addition of non-EUR ancestry GWAS, we were able to identify 49 asthma-associated variants that did not reach genome-wide significance in the EUR-only meta-analysis (Supplementary Table 8). 19 of these loci were potentially novel, and 13 of these 19 loci had an index variant higher in frequency in a non-EUR ancestry group compared to the EUR ancestry group. As the availability of non-EUR GWAS of asthma continues to increase, it is likely that additional variants associated with asthma that are higher in frequency in other populations will be discovered.

7. The genetic correlation analysis with all of the phenotypes in the GBMI biobanks is not needed in this paper. It's unclear what the goal of examining genetic correlations between asthma and all of the heritable diseases is. It appears that chronic heart failure and asthma are correlated—explanations for this could be provided.

Non-genetic epidemiological studies have identified correlations between asthma and many other disease categories[21–23]. More recently, some genome-wide cross-trait studies have found evidence for shared genetic architectures between asthma and other allergic diseases[24,25], neuropsychiatric disorders[26], and obesity[27], 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[28]. 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. Therefore, we leveraged the GBMI meta-analysis of asthma and data from two biobanks in GBMI to estimate genetic correlations between asthma and a wide range of phenotypic endpoints. We have restructured the Results to clarify these motivations:

Genetic overlap between asthma and other diseases
Non-genetic epidemiological studies have also identified correlations between asthma and many other disease categories beyond COPD[69–71]. More recently, some genome-wide cross-trait studies have found evidence for shared genetic architectures between asthma and other allergic diseases[21,72], 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.

The chronic heart failure UKBB GWAS did not have a statistically significant SNP heritability estimate, and thus the genetic correlation estimate between asthma and chronic heart failure in UKBB is difficult to interpret. We have updated Supplementary Table 13 and Supplementary Figure 12 to more clearly indicate the disease endpoints with significant heritability estimates in both UKBB and BBJ. See also related response to Reviewer #1, comment #6.

8. There is no independent replication population. This is a limitation even though the authors use the leave on out approach.

Since the initial submission of this manuscript, 4 more biobanks with asthma data have joined GBMI – Biobank of the Americas (BBofA), Qatar Biobank (QBB), Penn Medicine Biobank (PMBB), and Canadian Partnership for Tomorrow's Health (CanPath) – and thus we were able to use these biobanks as independent replication studies. We have added cohort characteristics of these biobanks to Supplementary Table 1.

We performed a meta-analysis using the GWAS from these replication studies. Although the case numbers in the replication data are less than 10% of the case numbers in the discovery data, 51 of the 179 top loci had index variants with a p-value < 0.05 in the replication meta-analysis, and 154 of the 179 top loci had index variants with consistent directions of effect in the discovery and replication meta-analyses (Supplementary Table 2). We have added this to the "Multi-ancestry meta-analysis for asthma across 18 biobanks in GBMI" Results section:

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 (Supplementary Table 2). 154 of the 179 index variants had consistent directions of effect in the discovery and replication meta-analyses.

9. The authors could justify their fixed effects meta-analysis with inverse variance weighting. It seems a random effects meta-analysis may have been more appropriate.

We thank the reviewer for this point. We chose to report results from the fixed effects meta-analysis because we did not observe substantial heterogeneity in the genetic effects of the genome-wide significant loci across biobanks (only 2 genome-wide significant loci were significantly heterogeneous across ancestries with Bonferroni p-value (Cochran's Q) < 0.05/179). We also conducted meta-analysis using the meta-regression approach in MR-MEGA[29], which accounts for effect size heterogeneity across

the biobanks. We have added these results to the "GWAS from diverse ancestries reveal shared genetic architecture of asthma and improves power for genetic discovery" section of the Results, as follows:

Taken together, these analyses indicate that the genetic architecture of asthma is largely shared across cohorts, despite differences in characteristics like 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 GWAS 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 2 additional loci associated with asthma, 1 of which is novel (Supplementary Table 6).

10. In the Discussion, the investigators could provide explanation for why excluding COPD would introduce an additional source of ascertainment bias.

We thank the reviewer for this suggestion. If we excluded participants with a COPD diagnosis, we would not have a fully representative sample of the participants in GBMI with asthma, potentially introducing selection bias, or collider bias, that could distort genetic associations[30,31]. We would also have decreased power to identify asthma-associated loci. Most of the previous genetic studies of asthma in the literature did not exclude individuals with COPD from analyses. However, in the childhood- and adult-onset asthma analyses, we do exclude participants with a COPD diagnosis to avoid confounding from potential misclassifications of adult-onset asthma and COPD. We have updated the Discussion to clarify this point, as follows:

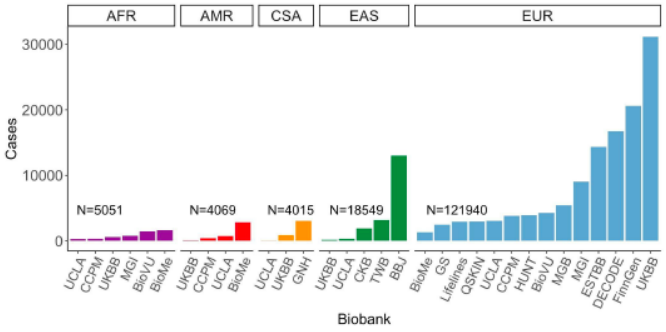
However, it is important to note that if we excluded participants with a COPD diagnosis, we would not have a fully representative sample of the participants in GBMI with asthma. As has been documented in other studies[90,91], this could induce selection bias, or collider bias, which could lead to biased 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 do exclude participants with a COPD diagnosis to avoid confounding from potential misclassifications of adult-onset asthma and COPD.

11. Discussion. The manuscript states that Bayesian PRS construction methods can improve prediction in asthma. The authors could explain how prediction is improved and discuss their PRS scores in the context of other asthma PRSs reported on to date.

Thank you for pointing this out. In a GBMI companion paper focused on PRS, Wang et al. (2021)[19] compared the prediction accuracy of PRS derived from the pruning and thresholding (P+T) method versus PRS-CS for asthma in target cohorts of EUR, CSA, EAS, and AFR ancestries in GBMI, and observed improvements in prediction accuracy using PRS-CS across all target cohorts (Supplementary Fig. 7). Comparisons of the methods are discussed at length in the paper. We thus decided to investigate potential added improvements using PRS-CSx, the multi-ancestry extension of PRS-CS, in this manuscript. Despite its optimization for multi-ancestry GWAS discovery cohorts we observed similar performances across these two Bayesian methods. We have updated the PRS section of the Results to mention the results from Wang et al. (2021)[19], as follows:

We next explored the impact of the increased sample sizes and diversity in 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 GBMI, Wang et al. (2021)[56] evaluated and compared the prediction accuracy

of PRS derived from the pruning and thresholding (P+T) method and 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 (Supplementary Fig. 7), 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 to the PRS constructed from the TAGC meta-*R* analysis[9] (Fig. 5).



Supplementary Figure 1. Asthma cases in discovery biobanks stratified by ancestry group. *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.*

Other studies on asthma PRS in the literature have primarily focused on using PRS to predict asthma in childhood, and overall found limited performance of PRS[33–36]. Most of these studies used the P+T approach, while a recently published paper, Namjou et al. (2022)[37], applied PRS-CS to the TAGC multi-ancestry GWAS and found improved discriminatory power of their PRS (AUC of 0.66-0.70 across two pediatric cohorts) compared to the prior studies that used P+T. Sordillo et al. (2021)[38] 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 GBMI compared to TAGC contribute to better performing PRS (Fig. 5). We have updated the Discussion to include a discussion of prior work on asthma PRS:

We also demonstrated that the greater diversity of 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[28–30,85]. Most of these studies used the P+T approach, while a recently published paper, Namjou et al. (2022)[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, or AUC, of 0.66-0.70 across two pediatric cohorts) compared to the prior studies that used P+T. Sordillo et al. (2021)[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 GBMI compared to TAGC contribute to better performing PRS (Fig. 5).

12. It's not clear if ancestry markers were used at all in these analyses.

We did not use ancestry markers to define ancestry groups, although since we used genome-wide methods, ancestry-informative SNPs are likely included in the analyses. Biobanks defined the ancestry

groups for their participants before conducting GWAS, and we added this information to Supplementary Table 1. 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 and the Human Genome Diversity Project to project biobank participants to the same principal components space[18]. We have added a section in Methods to clarify this:

Principal components (PC) projection for genetic ancestry comparison
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[94]. The weights corresponding to principal components for those markers were estimated based on the PCA analysis 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. (2021)[33].

13. The authors could mention that they are not able to account for BMI.
BMI is not typically included as a covariate in GWAS for asthma, and thus we did not adjust for BMI in our association analyses.

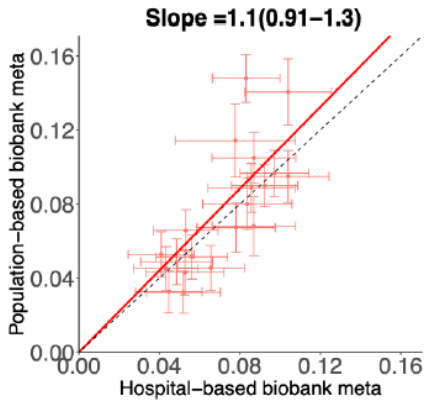
14. One of the limitations that the authors appropriately address is the heterogeneity in the phenotype. The authors could consider conducting sensitivity analysis to see if there are differences whether the cases were defined based on self-report (Taiwan Biobank, QSkin) compared to PheCodes or ICD-9 codes.

We thank the reviewer for this suggestion. We agree that an additional sensitivity analysis comparing SNP effects in the meta-analyses of the biobanks that used the PheCode definition vs. self-reported data would be helpful to evaluate the effects of phenotype heterogeneity on effect size estimation. However, only three biobanks used self-reported data, and 2 of the 3 biobanks (TWB and BBJ) only have participants of EAS ancestry. Unfortunately, this limits our ability to rigorously assess the effects of phenotype definition on differences in effect size estimates without confounding. However, as we mention in the Discussion, we compared the asthma GWAS derived from self-reported vs. PheCode data in UKBB and found high genetic correlation (r_g (se) = 0.95 (0.01)) between the GWAS. This provides some evidence that minor differences in phenotype definition may not substantially change the association results for asthma.

To address another potential source of heterogeneity, we conducted an additional sensitivity analysis comparing SNP effects in the meta-analyses of the biobanks with different ascertainment. 9 biobanks were population-based (CKB, DECODE, ESTBB, GNH, GS, HUNT, Lifelines, TWB, and UKBB) and 6 (BBJ, BioMe, BioVU, MGB, MGI, and UCLA) were hospital-based. We fit the Deming regression on the effect size estimates of loci identified by the all-biobank meta-analysis, using the SNPs with p-value < $1x10^{-6}$ in both meta-analyses, and observed high consistency in the effects across the two groups (Supplementary Fig. 6). We report these results in the "GWAS from diverse ancestries reveal shared genetic architecture of asthma and improves power for genetic discovery" section of the Results:

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- vs. population-

based biobanks. We conducted meta-analyses of the 9 population-based biobanks (CKB, DECODE, ESTBB, GNH, GS, HUNT, Lifelines, TWB, and UKBB) and 6 hospital-based biobanks (BBJ, BioMe, BioVU, MGB, MGI, and UCLA). We then fitted the Deming regression[35] on the effect size estimates of the loci identified by the all-biobank meta-analysis, using the SNPs with p-value < $1x10^{-6}$ in both meta-analyses, and observed high consistency in the effects across the two groups (Supplementary Fig. 6).



Supplementary Figure 6. Consistency of asthma index variants across biobanks with different ascertainment. *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 < $1x10^{-6}$ 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.*

15. Another limitation is it seems the investigators were unable to distinguish between child-onset and adult-onset asthma. Limiting to adult-onset asthma could help their analysis.

We thank the reviewer for this suggestion and agree that stratifying by age of onset is particularly important for studying asthma. We were able to conduct asthma age-of-onset subtype analyses in two of the participating GBMI biobanks, UKBB and FinnGen. We performed GWAS of childhood-onset asthma (COA) and adult-onset asthma (AOA) in Finngen and the EUR ancestry cohort in UKBB, using a cut-off age of 19 years at asthma diagnosis to define the subtypes (Methods). Then, we conducted fixed-effects, inverse-variance weighted meta-analyses of the COA (20,964 cases, 674,014 controls) and AOA (56,744 cases, 674,014 controls) GWAS, respectively. Broadly, the results indicated that the GBMI meta-analysis containing both age-of-onset subtypes captured many of the genetic variants contributing to the subtypes, but additional stratification by subtype revealed that the etiology of COA is characterized by a substantial overlap with AOA, as well as genes that have smaller (or no) effects on AOA. This supports previous findings[3,4]. Due to the limited availability of age of onset information across the biobanks, we did not have as much power to identify potential subtype-specific associations, which is an important area for future investigation. Please see response to Reviewer #1, comment #3 for a full explanation.

16. The authors conclude that the "genetic effects of associated loci are largely consistent across the biobanks and ancestries." How do the authors put this in the setting of multiple other studies that suggest that genetic effects vary by genetic ancestry?

We have highlighted a couple loci that showed significant heterogeneity in SNP effects across genetic ancestry groups, and recognize that other studies of asthma have discovered genetic loci that may have

population-specific effects. Additional studies from diverse populations are needed to more fully investigate and uncover genetic loci that may have differing effects across populations. However, we note that across the other 13 phenotypes studied in GBMI, Zhou et al. (2021)[18] did not observe substantial heterogeneity in genetic effects between the ancestry groups.

17. The manuscript does not mention how ancestry groups are defined. This could be added to the supplemental tables. Was self-report used or ancestry markers?

We thank the reviewer for this suggestion. We have added a column to Supplementary Table 1 with information on how ancestry groups were defined by each biobank. 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 and the Human Genome Diversity Project to project biobank participants to the same principal components space[18]. Additional information on ancestry groups can be found in Zhou et al. (2021)[18]. Please see also related response to Reviewer #1, comment #12.

Minor comments:
1. Summary: "Despite the considerable range in prevalence…" Consider adding "of asthma" after "prevalence."
2. Introduction: Need reference
3. Need to write out GWAS the first time in the Introduction
4. Introduction "efforts to diversify asthma" is not clear. Maybe the authors mean "efforts to conduct asthma GWAS in diverse populations."
5. Introduction: Add "initiative" after "Global Biobank Meta-analysis."
6. Consider using the term "allergic rhinitis" in addition to "hay fever."

We have addressed all above comments in the manuscript.

Reviewer 3

This is an interesting study representing a huge effort combining multiple biobank based GWASes resulting in the largest asthma GWAS to date.
The study identifies novel susceptibility genes, adding to the understanding of the genetic mechanisms of asthma, and provides novel knowledge on the potential of using large, herterogenous datasets in genetic studies of asthma.
The manuscript is well written and the methodologies seem sound.

We thank the reviewer for the summary and positive comments and respond to each comment below individually.

I am not sure, calculating gene-based p-values using MAGMA can truly be termed 'gene prioritization". It might be more appropriate to use actual gene prioritization methods such as DEPICT or PoPS which incorporate biological features to search for patterns of shared biology to prioritize genes at GWAS loci. Comparing gene-based results from MAGMA is not very different from directly comparing summary stats. In order to compare prioritized genes for e.g. asthma and COPD, I think other methods such as DEPICT are more suitable.

We thank the reviewer for this suggestion. We conducted gene prioritization for both asthma and COPD using DEPICT and PoPS. Compared to the 41% of prioritized genes for COPD that overlapped with the

genes prioritized for asthma, only 6% of the genes prioritized for COPD by DEPICT overlapped with asthma genes (Supplementary Table 16), and 9% of the genes prioritized for COPD by PoPS overlapped (Supplementary Table 17). The methods did not prioritize the same genes shared between COPD and asthma, with the exception of one shared gene which was prioritized by both DEPICT and MAGMA. Therefore, it is difficult to draw consistent conclusions about shared biology between asthma and COPD based on these approaches. This challenge is addressed in greater detail in Zhou et al. (2021)[18]. We added these analyses to the Results:

We also conducted gene prioritization using Data-driven Expression-Prioritized Integration for Complex Traits (DEPICT)[66] and gene-level Polygenic Priority Score (PoPS)[67]. However, only 3 of the 52 genes (6%) prioritized for COPD by DEPICT overlapped with a gene prioritized for asthma using the same method (Supplementary Table 16), and 17 of the 184 genes (9%) prioritized for COPD by PoPS overlapped with a prioritized gene for asthma (Supplementary Table 17). Across the shared COPD and asthma genes prioritized by each method, only 1 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[67] and is explored in more detail in Zhou et al. (2021)[33].

The high genetic correlation with COPD might reflect a relatively high age of participants (high age of asthma diagnosis), which is likely to increase the overlap between diagnoses of asthma and COPD compared to asthma studies including younger individuals. How did participant age compare to previous studies reporting lower genetic correlation of asthma and COPD?

Thank you for pointing this out. Sakornsakolpat et al. (2019)[43] estimated genetic correlation between asthma and COPD using summary statistics from the Trans-National Asthma Genetic Consortium (TAGC)[32], which included 66 studies, 27 of which were entirely or partially pediatric asthma cohorts. Hobbs et al. (2017)[44] utilized summary statistics from the GABRIEL Consortium, which also included pediatric asthma cohorts. On the other hand, GBMI biobanks are primarily composed of adult participants (Supplementary Table 1). To further investigate the genetic overlap between asthma and COPD, we stratified two of the participating biobanks with age of onset information, UKBB and FinnGen, into childhood-onset asthma (COA) and adult-onset asthma (AOA) cohorts and performed meta-analyses of COA and AOA GWAS from these two biobanks (Methods). We found that the AOA meta-analysis had similarly strong genetic correlation with the GBMI COPD meta-analysis (r_g (se) = 0.60 (0.3), p = 2.65×10^{-94}), compared to the all-asthma GBMI meta-analysis, while the COA meta-analysis had a more moderate genetic correlation with the GBMI COPD meta-analysis (r_g (se) = 0.33 (0.3), p = 7.60×10^{-31}). Since we excluded participants with concurring asthma and COPD diagnoses from the asthma subtype meta-analyses, this suggests that the higher genetic correlation between the GBMI asthma and COPD meta-analyses is not solely a function of more potential overlaps between asthma and COPD diagnoses in GBMI but potentially also due to greater shared genetic architecture between AOA, which is likely overrepresented in GBMI, and COPD. We report the results of these analyses in the "Asthma and COPD have a shared genetic basis but are also influenced by distinct biological processes" section of the Results, as follows:

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 0.38-0.42[63,64]. This may be a result of the discovery datasets used by these studies, which were enriched for pediatric asthma cohorts, while 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 = $2.65x10^{-94}$), while the COA meta-analysis had a more moderate genetic correlation with the COPD meta-analysis (r_g (se) = 0.33 (0.3), p = $7.60x10^{-31}$).
Data on participant age should be provided for the individual studies.

We have added participant age information to Supplementary Table 1.

References

1. El-Husseini, Z. W., Gosens, R., Dekker, F. & Koppelman, G. H. The genetics of asthma and the promise of genomics-guided drug target discovery. Lancet Respir Med 8, 1045–1056 (2020).
2. Kanai, M. et al. Meta-analysis fine-mapping is often miscalibrated at single-variant resolution. bioRxiv (2022) doi:10.1101/2022.03.16.22272457.
3. Ferreira, M. A. R. et al. Genetic Architectures of Childhood- and Adult-Onset Asthma Are Partly Distinct. Am. J. Hum. Genet. 104, 665–684 (2019).
4. Pividori, M., Schoettler, N., Nicolae, D. L., Ober, C. & Im, H. K. 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 (2019).
5. Deming, W. E. Statistical adjustment of data. 261, (1943).
6. Zhu, Z. et al. 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, (2021).
7. Tanaka, N. et al. Eight novel susceptibility loci and putative causal variants in atopic dermatitis. J. Allergy Clin. Immunol. 148, 1293–1306 (2021).
8. Stein, M. M. et al. A decade of research on the 17q12-21 asthma locus: Piecing together the puzzle. J. Allergy Clin. Immunol. 142, 749–764.e3 (2018).
9. Kim, S. Y., Min, C., Oh, D. J. & Choi, H. G. Increased risk of asthma in patients with rheumatoid arthritis: A longitudinal follow-up study using a national sample cohort. Sci. Rep. 9, 6957 (2019).
10. Shen, T.-C., Lin, C.-L., Wei, C.-C., Tu, C.-Y. & Li, Y.-F. The risk of asthma in rheumatoid arthritis: a population-based cohort study. QJM 107, 435–442 (2014).
11. Luo, Y. et al. Rheumatoid arthritis is associated with increased in-hospital mortality in asthma
Jeong, H. E., Jung, S.-M. & Cho, S.-I. Association between Rheumatoid Arthritis and Respiratory Allergic Diseases in Korean Adults: A Propensity Score Matched Case-Control Study. Int. J. Rheumatol. 2018, 3798124 (2018).
13. Rolfes, M. C., Juhn, Y. J., Wi, C.-I. & Sheen, Y. H. Asthma and the Risk of Rheumatoid Arthritis: An Insight into the Heterogeneity and Phenotypes of Asthma. Tuberc. Respir. Dis. 80, 113–135 (2017).
14. Shirai, Y. et al. Multi-trait and cross-population genome-wide association studies across autoimmune and allergic diseases identify shared and distinct genetic component. Ann. Rheum. Dis. (2022) doi:10.1136/annrheumdis-2022-222460.
15. Ehrlich, S. F., Quesenberry, C. P., Jr, Van Den Eeden, S. K., Shan, J. & Ferrara, A. 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 (2010).
16. Thomsen, S. F., Duffy, D. L., Kyvik, K. O., Skytthe, A. & Backer, V. Risk of asthma in adult twins with type 2 diabetes and increased body mass index. Allergy 66, 562–568 (2011).
17. Torres, R. M., Souza, M. D. S., Coelho, A. C. C., de Mello, L. M. & Souza-Machado, C. Association between Asthma and Type 2 Diabetes Mellitus: Mechanisms and Impact on Asthma Control—A Literature Review. Can. Respir. J. 2021, (2021).
18. Zhou, W. et al. Global Biobank Meta-analysis Initiative: powering genetic discovery across human

diseases. bioRxiv (2021) doi:10.1101/2021.11.19.21266436.
19. Wang, Y. et al. Global biobank analyses provide lessons for computing polygenic risk scores across diverse cohorts. medRxiv (2021).
20. Ge, T., Chen, C.-Y., Ni, Y., Feng, Y.-C. A. & Smoller, J. W. Polygenic prediction via Bayesian regression and continuous shrinkage priors. Nat. Commun. 10, 1776 (2019).
21. Leynaert, B., Neukirch, F., Demoly, P. & Bousquet, J. Epidemiologic evidence for asthma and rhinitis comorbidity. J. Allergy Clin. Immunol. 106, S201–5 (2000).
22. Van Lieshout, R. J. & Macqueen, G. Psychological factors in asthma. Allergy Asthma Clin. Immunol. 4, 12–28 (2008).
23. Miethe, S., Karsonova, A., Karaulov, A. & Renz, H. Obesity and asthma. J. Allergy Clin. Immunol. 146, 685–693 (2020).
24. Zhu, Z. et al. A genome-wide cross-trait analysis from UK Biobank highlights the shared genetic architecture of asthma and allergic diseases. Nat. Genet. 50, 857–864 (2018).
25. Guo, H., An, J. & Yu, Z. Identifying Shared Risk Genes for Asthma, Hay Fever, and Eczema by Multi-Trait and Multiomic Association Analyses. Front. Genet. 11, 270 (2020).
26. Zhu, Z. et al. Shared genetics of asthma and mental health disorders: a large-scale genome-wide cross-trait analysis. Eur. Respir. J. 54, (2019).
27. Zhu, Z. et al. Shared genetic and experimental links between obesity-related traits and asthma subtypes in UK Biobank. J. Allergy Clin. Immunol. 145, 537–549 (2020).
28. Zhu, Z., Hasegawa, K., Camargo, C. A. & Liang, L. Investigating asthma heterogeneity through shared and distinct genetics: Insights from genome-wide cross-trait analysis. J. Allergy Clin. Immunol. 147, 796–807 (2021).
29. Mägi, R. et al. 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 (2017).
30. Munafò, M. R., Tilling, K., Taylor, A. E., Evans, D. M. & Davey Smith, G. Collider scope: when selection bias can substantially influence observed associations. Int. J. Epidemiol. 47, 226–235 (2018).
31. Griffith, G. J. et al. Collider bias undermines our understanding of COVID-19 disease risk and severity. Nat. Commun. 11, 5749 (2020).
32. Demenais, F. et al. Multiancestry association study identifies new asthma risk loci that colocalize with immune-cell enhancer marks. Nat. Genet. 50, 42–53 (2018).
33. Dijk, F. N. et al. Genetic risk scores do not improve asthma prediction in childhood. J. Allergy Clin. Immunol. 144, 857–860.e7 (2019).
34. Kothalawala, D. M. et al. Integration of Genomic Risk Scores to Improve the Prediction of Childhood Asthma Diagnosis. J Pers Med 12, (2022).
35. Belsky, D. W. et al. 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 (2013).
36. Park, J. et al. Predicting allergic diseases in children using genome-wide association study (GWAS) data and family history. World Allergy Organ. J. 14, 100539 (2021).
37. Namjou, B. et al. Multiancestral polygenic risk score for pediatric asthma. J. Allergy Clin. Immunol. (2022) doi:10.1016/j.jaci.2022.03.035.
38. Sordillo, J. E. et al. A polygenic risk score for asthma in a large racially diverse population. Clin. Exp. Allergy 51, 1410–1420 (2021).
39. 1000 Genomes Project Consortium et al. A global reference for human genetic variation. Nature 526, 68–74 (2015).
40. Cann, H. M. et al. A human genome diversity cell line panel. Science 296, 261–262 (2002).
41. Pers, T. H. et al. Biological interpretation of genome-wide association studies using predicted gene functions. Nat. Commun. 6, 5890 (2015).

42. Weeks, E. M. et al. Leveraging polygenic enrichments of gene features to predict genes underlying complex traits and diseases. bioRxiv (2020) doi:10.1101/2020.09.08.20190561.
43. Sakornsakolpat, P. et al. Genetic landscape of chronic obstructive pulmonary disease identifies heterogeneous cell-type and phenotype associations. Nat. Genet. 51, 494–505 (2019).
44. Hobbs, B. D. et al. Genetic loci associated with chronic obstructive pulmonary disease overlap with loci for lung function and pulmonary fibrosis. Nat. Genet. 49, 426–432 (2017).

Referees' report, second round of review

Reviewers' Comments:

Reviewer #1: The authors have addressed all my comments.

Reviewer #2: Comments enter in this field will be shared with the author; your identity will remain anonymous.

The authors have done an excellent job responding to my comments. I am particularly appreciative that the authors conducted additional analyses and were able to add replication analyses. Additionally, the authors were able to do sensitivity analyses looking at child-onset asthma versus adult-onset asthma.

Reviewer #3: The authors have addressed my concerns adequately.

Authors' response to the second round of review

Our responses to each comment are below in purple, and quoted text from the manuscript is italicized and in gray.

1) Please respond to the referee comments and incorporate requested revisions.

Reviewers' Comments:

Reviewer #1: The authors have addressed all my comments.

Reviewer #2: The authors have done an excellent job responding to my comments. I am particularly appreciative that the authors conducted additional analyses and were able to add replication analyses. Additionally, the authors were able to do sensitivity analyses looking at child-onset asthma versus adult-onset asthma.

Reviewer #3: The authors have addressed my concerns adequately.

We thank the reviewers for their positive comments.