

# Large-scale Genetic Analysis Identifies 66 Novel Loci for Asthma

Yi Han<sup>1,2</sup>, Qiong Jia<sup>1,2</sup>, Pedram Shafiei Jahani<sup>3</sup>, Benjamin P. Hurrell<sup>3</sup>, Calvin Pan<sup>4</sup>, Pin Huang<sup>1,2</sup>, Janet Gukasyan<sup>1,2</sup>, Nicholas C. Woodward<sup>1,2</sup>, Eleazar Eskin<sup>5</sup>, Frank Gilliland<sup>1</sup>, Omid Akbari<sup>3</sup>, Jaana A. Hartiala<sup>1,2\*</sup>, and Hooman Allayee<sup>1,2\*</sup>

Departments of <sup>1</sup>Preventive Medicine, <sup>2</sup>Biochemistry & Molecular Medicine, and <sup>3</sup>Molecular Microbiology and Immunology, Keck School of Medicine, University of Southern California, Los Angeles, CA 90033; <sup>4</sup>Department of Human Genetics, David Geffen School of Medicine at UCLA, Los Angeles, CA 90095; <sup>5</sup>Department of Computer Science and Inter-Departmental Program in Bioinformatics, University of California, Los Angeles, Los Angeles, CA 90095.

\*These authors contributed equally to this work.

## Address correspondence and reprint requests to:

Hooman Allayee, PhD  
Department of Preventive Medicine  
Keck School of Medicine of USC  
2250 Alcazar Street, CSC202  
Los Angeles, CA 90033  
Phone: (323) 442-1736  
Fax: (332) 442-2764  
Email: hallayee@usc.edu

# Abstract

We carried out a genome-wide association study (GWAS) for asthma in UK Biobank, followed by a meta-analysis with results from the Trans-National Asthma Genetic Consortium (TAGC). 66 novel genomic regions were identified, bringing the number of known asthma susceptibility loci to 211. Significant gene-sex interactions were also observed where susceptibility alleles, either individually or as a function of polygenic risk scores, increased asthma risk to a greater extent in men than women. Bioinformatics analyses demonstrated that asthma-associated variants were enriched for colocalizing to regions of open chromatin in immune cells and identified candidate causal genes at 52 of the novel loci, including *CD52*. An anti-CD52 ( $\alpha$ -CD52) antibody mimicked the immune cell-depleting effects of an FDA-approved human  $\alpha$ -CD52 antibody and reduced allergen-induced airway hyperreactivity in mice. These results further elucidate the genetic architecture of asthma, provide evidence that the immune system plays a prominent role in its pathogenesis, and suggest that *CD52* represents a potentially novel therapeutic target for treating asthma.

**Keywords:** asthma; genome-wide association study; meta-analysis, gene-environment interactions; polygenic risk score, immune system; CD52.

# Introduction

Asthma is a global chronic respiratory disease affecting over 300 million people and a significant cause of premature death and reduced quality of life<sup>1</sup>. The disease develops through repeated cycles of inflammation, often in response to environmental triggers, that lead to progressive obstruction of the airways and deterioration of respiratory function<sup>2</sup>. While advances have been made in understanding these complex pathogenic processes, translation of research findings to development of novel therapies has not progressed rapidly enough to address the unmet clinical needs of many patients, particularly those with severe forms of asthma.

It is generally accepted that susceptibility to asthma is a combination of genetic predisposition, exposure to environmental factors, and their interactions. Consistent with this notion, estimates for the heritability of asthma have ranged between 35%-95%<sup>3</sup>, although those based on twin studies have been lower<sup>4</sup>. Genome-wide association studies (GWAS) have confirmed the contribution of genetic factors to risk of asthma with over 140 susceptibility loci having been identified over the last decade<sup>5-24</sup>. However, the risk alleles, most of which are common in the population, still only explain a small fraction of the overall heritability for asthma<sup>25,26</sup>. This observation implies either the existence of additional variants with smaller effect sizes, rare susceptibility alleles, and/or interactions between genes and environmental factors. In this regard, rare variants do not appear to explain a significant proportion of the “missing heritability” for asthma<sup>27</sup>. Gene-environment (GxE) interactions also remain poorly understood due to the inherent difficulties of carrying out such studies in humans, particularly with respect to accurate exposure assessment, adequately powered sample sizes with both genetic and exposure data, and the heterogeneous nature of asthma itself<sup>28,29</sup>. Thus, additional

studies are needed to further elucidate the genetic architecture of asthma and determine the applicability of the findings for translational purposes.

Here, we identified 66 novel risk loci for asthma through a large-scale genetic study with 88,486 cases and 447,859 controls from UK Biobank and the Trans-National Asthma Genetic Consortium (TAGC). Follow up analyses demonstrated interactions between asthma susceptibility alleles and sex, provide further evidence that the immune system plays a prominent role in asthma pathogenesis, and functionally validated *CD52* as one candidate causal gene with the potential of being a novel therapeutic target.

# Results

**GWAS in UK Biobank:** To further elucidate the genetic architecture of asthma, we first carried out a genome-wide association study (GWAS) in 64,538 cases and 329,321 controls from the UK Biobank (**Fig. 1** and **Supplemental Table 1**). This GWAS analysis identified 32,813 significantly associated SNPs distributed among 145 loci (**Supplemental Fig. 1** and **Supplemental Table 2**). Of these regions, 41 loci were novel and not previously known to be associated with asthma or the combined phenotype of asthma and allergic disease. The remaining 104 overlapped with the 145 loci previously identified for either asthma alone in adults, children, or subjects from multiple ancestries, or the combined phenotype of asthma, eczema, and hay fever<sup>5-24</sup> (**Supplemental Table 2**).

**Meta-analyses of GWAS Data for Asthma in UK Biobank and TAGC:** We next meta-analyzed the GWAS results from UK Biobank with publicly available summary data from the TAGC<sup>21,30</sup> (**Fig. 1**). After harmonizing the summary GWAS results from TAGC to match the data in UK Biobank (see Methods for additional details), we performed a Z-score meta-analysis for asthma with 8,365,715 SNPs common to both datasets and a total of 88,486 cases and 447,859 controls (**Supplemental Table 1**). The meta-analysis revealed 33,017 significantly associated variants distributed across 167 loci (**Fig. 2; Supplemental Fig. 2; Supplemental Table 3**). Of these, 58 loci were novel and identified for asthma for the first time herein (**Fig. 2** and **Table 1**) whereas the other 109 overlapped with the 145 previously known loci (**Supplemental Table 3**).

However, despite the P-values not exceeding the genome-wide significance level of  $5.0 \times 10^{-8}$ , our meta-analysis did yield evidence for association of 33 of the remaining 36 known asthma loci at a Bonferroni-corrected threshold for testing 145 regions ( $P=0.05/145=3.4 \times 10^{-4}$ ) (**Supplemental**

**Table 4).** The only exceptions were two childhood asthma loci and one for the combined phenotype of asthma and allergic disease (**Supplemental Table 4**). Eight other novel regions were also significantly associated with asthma in UK Biobank alone, but their association signals fell slightly below the genome-wide significance threshold in the meta-analysis (**Fig. 2** and **Table 1**). Thus, our collective analyses with UK Biobank and TAGC identified 66 novel loci for asthma (**Fig. 1, Table 1, and Supplemental Fig. 3**) and replicated nearly all (142/145) previously known genomic regions. Altogether, the 66 novel loci explained an additional 1.5% of the heritability for asthma and bring the total number of susceptibility loci to 211 at the time of this analysis. Since effect sizes were provided in the publicly available data from TAGC<sup>21</sup>, we also conducted a HapMap-based fixed-effects meta-analysis for asthma with 1,978,494 SNPs common to both datasets and the same number of 536,345 cases and controls. However, no additional novel loci were identified beyond those from the Z-score-based meta-analysis described above.

Sex-stratified and Polygenic Risk Score Analyses with Asthma Susceptibility Loci: We next used primary level data from UK Biobank to explore how association signals at the 211 asthma susceptibility loci differed in males and females and as a function of cumulative genetic burden. Thirty SNPs yielded nominal P-values for interaction with sex (P-int) that were <0.05, of which only four previously known loci on chromosomes 1p34.3, 2q37.3, 5q22.1, and 6p25.3 would be considered significant at the Bonferroni-corrected threshold for testing 211 loci (P-int=0.05/211=2.4x10<sup>-4</sup>) (**Supplemental Table 5**). Interestingly, the effect sizes at these four loci (and most of the others) were stronger in males than in females, although it should be noted that the significance of the interaction with the chromosome 1p34.3 locus could be inflated due to

frequency of the susceptibility allele (**Supplemental Table 5**). We next constructed three weighted genetic risk scores (GRS) in UK Biobank based on the number of risk alleles carried at the 66 novel, 145 known, or all 211 loci. Each GRS was weighted by the overall effect sizes of the included alleles that were derived from the association analysis that included both sexes. As shown in **Fig. 3A**, individuals in the highest decile for the GRS constructed with all 211 risk alleles had nearly a 5-fold increased risk of asthma compared to those in the lowest decile (OR=4.7, 95% CI 4.5-4.9;  $P < 1.0 \times 10^{-305}$ ). Increased risk of asthma was also observed for individuals in the highest versus lowest deciles of the GRS for the 145 known and 66 novel loci (**Fig. 3A**), although the overall effect size for the 66-locus GRS was weaker (OR=2.0, 95% CI 1.9-2.1;  $P = 2.6 \times 10^{-267}$ ). To determine whether cumulative genetic burden differed in men and women, we also used sex-specific effect sizes to construct weighted GRS in men and women separately for the 66 novel, 145 known, or all 211 asthma loci. This analysis revealed a pattern where risk of asthma among subjects in the highest versus lowest decile of the GRS was increased by approximately one order of magnitude greater in men than in women (**Fig. 3B-D**). For example, compared to the lowest decile, the OR for asthma among men in the highest decile for the GRS constructed with all 211 loci was 5.5 (95% CI 5.2-5.9;  $P < 1.0 \times 10^{-305}$ ) whereas the same analysis in women revealed that the increased risk of asthma among subjects in the highest decile was only 4.4 (95% CI 4.2-4.6;  $P < 1.0 \times 10^{-305}$ ). These results were accompanied by significant gene-sex interactions when comparing the highest versus lowest decile for the GRSs constructed from all 211 ( $P\text{-int} = 1.2 \times 10^{-9}$ ), 145 known ( $P\text{-int} = 2.9 \times 10^{-9}$ ), and 66 novel ( $P\text{-int} = 0.029$ ) loci. Significant interactions with sex were also observed when considering the entire spectrum of cumulative genetic burden for the 211 ( $P\text{-int} = 4.0 \times 10^{-18}$ ), 145 known ( $P\text{-int} = 8.0 \times 10^{-18}$ ), and 66 novel ( $P\text{-int} = 0.009$ ) loci. The interactions with sex across all deciles remained

significant even after exclusion of the four risk alleles on chromosomes 1p34.3, 2q37.3, 5q22.1, and 6p25.3 from the 211- ( $P_{\text{int}}=2.3 \times 10^{-13}$ ) and 145-locus ( $P_{\text{int}}=1.6 \times 10^{-13}$ ) GRSs.

Association of Novel Loci with Asthma-related Phenotypes and Other Disease Traits: We next used data from UK Biobank to characterize the novel loci with respect to effects sizes for risk of asthma or allergic disease as well as for association with lung function traits. Nearly all 66 novel loci yielded ORs  $\leq 1.10$  for association with asthma, with the majority having even more modest effect sizes (**Supplemental Table 6**). Furthermore, most of the novel loci were associated with the more broadly defined phenotype of asthma that also included allergic disease, although some yielded stronger associations with a stricter definition of asthma that excluded allergic disease (**Supplemental Table 6**). We next evaluated the association of all 66 novel asthma loci for associations with forced expiratory volume in one second ( $FEV_1$ ) and forced vital capacity (FVC). At the Bonferroni-corrected significance threshold for testing 66 loci ( $P=0.05/66=7.6 \times 10^{-4}$ ), the asthma risk alleles at 17 of the novel loci exhibited directionally consistent associations with decreased  $FEV_1$ , of which 10 were also associated with decreased FVC (**Supplemental Table 6**). To further explore their clinical relevance, we used the PhenoScanner database<sup>31</sup> to determine whether the novel loci identified in our meta-analyses were associated with other disease phenotypes. Interestingly, lead variants (or tightly linked proxies) at ~half of the newly identified loci had been previously associated with other asthma-relevant traits, including blood cell parameters (19 loci), height (7 loci), BMI and waist circumference (4 loci), inflammatory bowel disease (3 loci), and cardiovascular disease (1 locus) (**Supplemental Table 7**).



Enrichment of Asthma-associated Variants in Regulatory Elements or Tissues: To gain biological insight into the newly identified asthma loci, we used two bioinformatics tools to carry out enrichment analyses for regulatory elements and pathways. GWAS Analysis of Regulatory or Functional Information Enrichment with LD correction (GARFIELD) revealed highly significant 3 to 4-fold enrichment of asthma-associated variants colocalizing to DNase I hypersensitive sites in several tissues with biological relevance to asthma, such as, epithelium, lung, and thymus (**Fig. 4A; Supplemental Table 8**). However, enrichment in DNase I hypersensitive elements was particularly evident in blood, and specifically in GM12878 B cells, CD20<sup>+</sup> B cells, and CD3<sup>+</sup>, CD4<sup>+</sup>, or CD8<sup>+</sup> T cells (**Fig. 4B; Supplemental Table 8**). A Data-driven Expression-Prioritized Integration for Complex Trait (DEPICT) analysis also yielded consistent results where the most significantly enriched biological pathways and tissues were leukocyte activation/differentiation/proliferation, cytokine production, lymphoid organ development, blood cells, lymphocytes, lymphoid tissue, and spleen (**Supplemental Tables 9 and 10**). Taken together, these results further highlight the immune system as an important component in the pathophysiology of asthma.

Prioritization of Loci and Positional Candidate Genes: We next used publicly available expression quantitative trait loci (eQTL) data to identify candidate causal genes and prioritize the 66 novel regions for functional validation (**Fig. 1**). The lead SNP or tightly linked ( $r^2 > 0.8$ ) proxy variants at 52 of the novel loci yielded at least one *cis* eQTL for a positional candidate gene in one or more tissues, including those relevant to asthma (**Supplemental Table 11**). Notably, the most significant eQTLs were in blood, some of which yielded  $P < 1.0 \times 10^{-300}$ . These observations are consistent with the GARFIELD enrichment analyses pointing to blood as playing an

important role in asthma. We further prioritized loci by focusing on positional candidates that also yielded eQTLs in immune cell populations highlighted by our enrichment analyses, such B and T cells. *CD52*, *AHII*, and *CLUAP1* on chromosomes 1p36.11, 6q23.3, and 16p13.3 (**Supplemental Fig. 3**), respectively, were three genes that met these criteria and yielded the most significant eQTLs in these specific immune cell types. However, eQTLs were also observed in the lung for *CD52* and *CLUAP1*, making these even stronger candidate causal genes for functional validation (**Supplemental Table 11**). *CLUAP1* was initially identified as a protein that interacts with clusterin and is required for ciliogenesis<sup>32</sup> but has not otherwise been directly implicated in asthma or pulmonary biology. By comparison, *CD52* encodes a membrane glycoprotein present at high levels on the surface of various immune cells, including lymphocytes, monocytes, and dendritic cells. In addition, alemtuzumab is a monoclonal anti-CD52 ( $\alpha$ CD52) antibody that results in preferential and prolonged depletion of circulating T and B cells<sup>33,34</sup> and is FDA approved for the treatment of relapsing remitting multiple sclerosis<sup>35-37</sup> and B-cell chronic lymphocytic leukemia<sup>38-40</sup>. Because of its biological role in immune cells relevant to asthma as well as its potential translational implications, we therefore prioritized *CD52* as a strong causal positional candidate gene for functional validation (**Fig. 1**).

Functional Validation of *CD52*: To functionally validate *CD52*, we used an *in vivo* pharmacological approach. Since alemtuzumab does not target the mouse CD52 protein, we first investigated the immune cell depleting activity of a mouse  $\alpha$ CD52 antibody that was previously reported to have biological effects<sup>41</sup>. We designed a mouse experiment that would mimic the clinical protocol for administering alemtuzumab to humans, in conjunction with the induction of airway hyperreactivity through exposure to house dust mite (HDM) (**Supplemental Fig. 4A**).

Consistent with the clinical effects of alemtuzumab on circulating lymphocytes in humans<sup>33,34</sup>, the mouse  $\alpha$ CD52 antibody efficiently depleted CD4<sup>+</sup> and CD8<sup>+</sup> T cells as well as CD19<sup>+</sup> B cells in both the lung and spleen (**Supplemental Figs. 4B-E and 5**). We next tested whether  $\alpha$ CD52 treatment reduced HDM-induced airway hyperreactivity and lung inflammation (**Fig. 5A**). As expected, mice exposed to HDM had significantly increased lung resistance and decreased dynamic compliance compared to PBS-exposed mice (**Fig. 5B and C**). However, the induction of lung resistance and deterioration of dynamic compliance as a result of HDM exposure was significantly blunted in mice treated with  $\alpha$ CD52 antibody compared to isotype control antibody (**Fig. 5B and C**). Treatment of HDM-exposed mice with  $\alpha$ CD52 antibody also significantly decreased numbers of total cells, eosinophils, T cells, and neutrophils in bronchial alveolar lavage (BAL) (**Fig. 5D and Supplemental Fig. 5**). Histological analysis of the lung further showed reduced infiltration of inflammatory cells and decreased thickness of the airway epithelium in HDM-exposed mice treated with  $\alpha$ CD52 antibody compared to isotype control antibody (**Fig. 5E**). Collectively, these data demonstrate that targeting CD52 with an antagonizing antibody ameliorates cellular and physiological lung function traits in mice exposed to HDM and provide functional *in vivo* evidence that *CD52* is at least one candidate causal gene at the novel asthma locus on chromosome 1p36.11.

# Discussion

In the present study, we identified 66 novel regions not previously known to be associated with asthma and replicated all but three of the 145 previously identified regions. Despite increasing in the number of genomic regions associated with asthma by nearly 50%, the newly identified 66 loci only explained an additional 1.5% of the heritability for asthma, reflecting the modest nature of the susceptibility alleles' effect sizes on disease risk. Taking into consideration the ~7% attributable to previously known loci, the results of our genetic analyses and the 211 susceptibility loci identified to date still only explain 8-9% of the total heritability for asthma. Compared to other common diseases<sup>42</sup>, the relatively low fraction of the heritability explained by genetic factors may be due to the more complex underlying phenotypic and genetic heterogeneity of asthma. For example, most of the novel asthma loci also exhibited associations with allergic disease, consistent with recent studies demonstrating that these phenotypes share genetic determinants<sup>20</sup>. By comparison, only ~25% of the novel regions exhibited associations with FEV<sub>1</sub> and FVC. This is the proportion of asthma susceptibility loci that studies have consistently shown to be associated with FEV<sub>1</sub> and FVC and suggest that most genetic factors predisposing to asthma do not necessarily manifest through associations with lung function traits.

GxE interactions are also recognized as important components of asthma's complex genetic etiology but progress in this area been limited, with most studies focusing on exposures such as allergens, cigarette smoke, and air pollutants<sup>43,44</sup>. Our data suggest that sex is another potentially important endogenous environmental factor that can modulate risk through interactions with asthma susceptibility alleles, both individually or as a function of cumulative genetic burden. This was particularly evident among men in whom increased risk of asthma across all categorized risk alleles was consistently more accentuated than in females, even after

exclusion of four susceptibility alleles that individually yielded statistical evidence for being more strongly associated with asthma in men. Evidence for such gene-sex interactions has been demonstrated for several complex phenotypes<sup>45</sup>, including anthropometric traits<sup>46-48</sup>, coronary artery disease<sup>49</sup>, as well as asthma<sup>50-53</sup>. However, the previously reported sex-specific associations for respiratory outcomes<sup>50-53</sup>, which were identified in much smaller sample sizes relative to our study, were not detected in our present analyses (data not shown). Interestingly, the male-specific locus (rs2549003) identified on chromosome 5<sup>52</sup> was associated with asthma in our meta-analysis (overall  $P=2.0 \times 10^{-14}$ ) but the association signal was derived equally in men and women. Nonetheless, these data collectively suggest that GxE interactions are likely to have important contributions to risk of asthma and represent an unexplored area for future investigation, particularly as new statistical methods are developed to carry out such analyses on a genome-wide level<sup>54</sup>.

Our results also add to the growing body of genetic and bioinformatics evidence that immune system-related processes are important drivers of asthma pathophysiology, consistent with prior studies<sup>21</sup>. For example, enrichment analyses highlighted pathways related to various aspects of leukocyte function and immune system regulation/development. Furthermore, asthma-associated variants were highly enriched in regions of open chromatin in several tissues and especially in B and T lymphocytes. This concept is also supported by the nature of specific candidate causal genes located at some of the novel loci. One such example is the chromosome 1p36.22 locus where the lead SNP (rs2230624; G>A) yielded the strongest effect size for asthma (OR=1.19) out of all 211 risk variants. Rs2230624 leads to a rare (~1% frequency in European populations) but computationally predicted deleterious Cys273Tyr substitution in exon 8 of *TNFRSF8*, which encodes the costimulatory molecule CD30. However, the common Cys273

allele of rs2230624 is associated with increased risk, suggesting that loss of CD30 function may be protective for asthma. This hypothesis would be consistent with prior studies showing that CD30 increases pro-inflammatory cytokine production by CD4<sup>+</sup> T cells and is involved in Th1, Th2, and Th17 immune responses<sup>55</sup>. Furthermore, soluble CD30 levels are elevated in patients with asthma or atopic dermatitis and correlate with disease severity<sup>56,57</sup>. When gene expression data were used to prioritize candidate causal genes, *CD200R1* and *AH11* emerged as two other potential genes of interest related to immune function, given the highly significant *cis* eQTLs ( $P < 1.0 \times 10^{-300}$ ) observed in blood with the lead variants on chromosomes 3q13.2 (rs7622814) and 6q23.3 (rs4526212), respectively. For example, CD200:CD200R1 signaling suppresses expression of proinflammatory molecules<sup>58</sup> and the induction of CD8<sup>+</sup> T cells<sup>59</sup>. By comparison, *AH11* may be involved in JAK-STAT signaling<sup>60</sup> and downstream effects on Th1 and Th2 cells<sup>61</sup>. In addition, a recent GWAS identified a variant (rs9647635) at the chromosome 6q23.3 locus in near perfect LD with our peak SNP ( $r^2 = 0.98$ ) that was associated with selective IgA deficiency<sup>62</sup>, which is correlated with asthma<sup>63</sup>.

Another important aspect of our study are the *in vivo* pharmacological experiments functionally validating *CD52* as at least one candidate causal gene at the chromosome 1p36.11 locus. These findings not only point to a novel pathway for development of asthma but may also have potential translational implications. For example, we demonstrated that treatment of mice with an  $\alpha$ CD52 antibody mimicked the immune cell-depleting activity of alemtuzumab in the circulation in humans<sup>33,34</sup>. We readily observed depletion of immune cells in spleen, which in mice is considered as accurately reflecting the pool of peripheral immune cells<sup>64</sup>. By also depleted leukocytes in lung and BAL, the mouse  $\alpha$ CD52 antibody had biological effects in tissues that, to our knowledge, have not been reported with alemtuzumab in humans. Of most

importance and relevance to asthma, mice receiving the  $\alpha$ CD52 antibody exhibited significantly reduced pulmonary inflammation, airway epithelium thickness, and allergen-induced airway hyperreactivity. Given that alemtuzumab is already FDA-approved<sup>35-40</sup>, it is tempting to consider CD52 as a novel therapeutic target for asthma. However, serious, albeit very rare, side effects have been reported with alemtuzumab, and as an immunosuppressive agent, patients being treated with this monoclonal antibody are at increased risk of infections and malignancies<sup>65</sup>. Nonetheless, there may still be a population of severe asthmatics with limited therapeutic options<sup>66</sup>, such as those with corticosteroid-resistant or eosinophilic asthma, in whom the clinical benefits of alemtuzumab may justify its potential adverse effects. The generally favorable long-term outcomes observed with alemtuzumab in small numbers of patients with hypereosinophilic syndrome<sup>67,68</sup> suggest this treatment strategy may be a potentially viable concept worthy of further consideration. Approaches to limiting or eliminating serious adverse effects of alemtuzumab could be through reduced dosing strategies, such as those used for multiple sclerosis compared to leukemia<sup>69</sup>, and development of other  $\alpha$ CD52 antibodies that may be more efficacious but less immunogenic<sup>70</sup>. Alternatively, the precise mechanisms through which alemtuzumab exerts its therapeutic effects are still not entirely known. Thus, a better understanding of CD52's biological functions may identify alternative targets that could be evaluated for therapeutic development and which may lead to less severe or minimal side effects.

While the present analyses reveal novel genetic determinants of asthma, our study should also be taken in the context of certain limitations. First, our inclusion of subjects in UK Biobank with self-reported asthma could have led to misclassification of cases and biased the results towards the null. However, the large sample size in our study and our ability to replicate all but three of the known 145 asthma loci indicate that our analyses were robust to the presence of such

potential confounding factors. Second, most subjects in UK Biobank and TAGC were of European ancestry and it is possible that the genetic association results may not be generalizable to other populations. This notion has previously been observed with both common and rare variants even among closely related Latino populations and in comparison, to subjects of African ancestry<sup>27,71</sup>. Third, the asthma loci identified in our meta-analysis relied on GWAS results in TAGC that were derived from imputed Z scores and P-values. Although this approach allowed us to include the largest number of SNPs possible in the meta-analysis with UK Biobank, it may not have provided the most precise summary statistics in TAGC. The likelihood of this potential limitation affecting our overall results is low since the Z-score meta-analysis replicated the association signals at nearly all known asthma susceptibility loci. Moreover, even though the Z-score meta-analysis did not provide effect sizes, the ORs based on association tests in UK Biobank alone are still likely to be very similar to those derived from the combined datasets given the large size of UK Biobank. Lastly, our study was primarily focused on discovery of main genetic effects of common susceptibility alleles. However, rare variants or GxE interactions, particularly with respect to other environmental exposures besides sex, may still play important roles in modulating asthma risk and will need to be addressed in appropriately designed future studies.

In summary, large-scale genetics analyses identified 66 novel risk loci for asthma and implicated regulation of immune processes as an important component to disease susceptibility. These results provide opportunities for further exploration of the biological mechanisms underlying the pathogenesis of asthma and evaluating additional validated causal genes as potentially novel therapeutic targets.



# **Materials and Methods**

## **Study Populations**

UK Biobank recruited participants between 40-69 years of age who were registered with a general practitioner of the UK National Health Service (NHS). Between 2006-2010, a total 503,325 individuals were included. All study participants provided informed consent and the study was approved by the North West Multi-centre Research Ethics Committee. Detailed methods used by UK Biobank have been described elsewhere<sup>72</sup>. Participating cohorts in the Trans-National Asthma Genetic Consortium (TAGC) have been previously described in detail<sup>30</sup>. Briefly, TAGC includes 56 studies of European populations, 7 studies of African populations, 2 studies of Japanese populations and one study of a Latino population. All participants gave written consent for participation in genetic studies, and the protocol of each study was approved by the corresponding local research ethics committee or institutional review board. The present study was approved by the Institutional Review Boards of the USC Keck School of Medicine.

## **GWAS Analyses in UK Biobank**

Asthma cases was defined based on UK Biobank field code 6152\_8 (doctor diagnosed asthma), International Classification of Diseases version-10 (ICD10) J45 (asthma)/J46 (severe asthma), and self-reported asthma. Field 6152 is a summary of the distinct main diagnosis codes a participant has had recorded across all their hospital visits. Non-asthmatic controls were defined as individuals free from 6152\_8 (doctor diagnosed asthma), 6152\_9 (doctor diagnosed allergic disease), ICD10 J45/J46/J30 (hay fever)/L20 (dermatitis and eczema), as well as self-reported asthma/hay fever/eczema/allergy/allergy to house dust mite (HDM). This definition strategy resulted in 64,538 asthma cases and 329,321 controls in UK Biobank, and was selected to

increase the likelihood of identifying loci associated specifically with asthma rather than loci for asthma and/or allergic disease. Quality control of samples and DNA variants, and imputation were performed by the Wellcome Trust Centre for Human Genetics, as described in detail elsewhere<sup>72</sup>. Briefly, ~90 million SNPs imputed from the Haplotype Reference Consortium, UK10K, and 1000 Genomes imputation were available in UK Biobank. Of these, 9,572,556 variants were used for GWAS analysis after filtering on autosomal SNPs with INFO scores >0.8 (directly from UK Biobank) and with minor allele frequencies (MAF) >1% in the 487,409 individuals with imputed genotypes. A GWAS analysis was performed with BOLT-LMM v2.3.2 using a standard (infinitesimal) mixed model to correct for structure due to relatedness, ancestral heterogeneity, with adjustment for age, sex, the first 20 principal components, and genotyping array<sup>73</sup>. The genome-wide significance threshold was set at  $P=5.0 \times 10^{-8}$ . Since BOLT-LMM relies on linear models even for qualitative traits, SNP effect size estimates on the quantitative scale were transformed to obtain odds ratios (ORs) and standard errors (SEs) using the following formula:  $\beta$  or  $SE/(\mu * (1 - \mu))$ , where  $\mu$  = case fraction<sup>73</sup>.

### **Imputation in TAGC**

Publicly available GWAS summary statistics for asthma with 2,001,281 SNPs in predominately European ancestry populations from TAGC were downloaded and imputed with summary statistics imputation (SSimp) software<sup>74</sup> using the European population from the 1000 Genomes Project (Phase 1 release, v3) as a reference panel for LD computation. This imputation resulted in Z-scores and P-values for association of 9,415,011 SNPs with asthma in TAGC.

## Meta-analyses for asthma in UK Biobank and TAGC

The imputed Z-score and P-value summary level data in TAGC were combined with the results of our linear mixed model GWAS for asthma in UK Biobank. We first performed a Z-score meta-analysis for asthma with 8,365,359 SNPs common to both datasets assuming an additive model, as implemented in METAL<sup>75</sup>. We also performed a fixed effect meta-analysis for asthma with betas and standard errors obtained in our GWAS analysis of UK Biobank and those provided in TAGC for 1,978,494 SNPs common to both datasets. The genome-wide threshold for significant association in both meta-analyses was set at  $P=5.0 \times 10^{-8}$ . A locus was defined as novel if our sentinel SNP was >1Mb away or in weak or no linkage disequilibrium ( $r^2 \leq 0.1$ ) with the lead variants at the 145 previously reported loci for asthma and/or allergic disease<sup>5-11,13,15-21,23,24</sup>. Replication of the known asthma/ allergic disease loci in our meta-analysis was considered significant at a Bonferroni-corrected threshold of  $P=3.4 \times 10^{-4}$  for testing 145 loci (0.05/145).

## Sex-stratified associations with asthma susceptibility loci

Sex-stratified association with variants at all 211 loci and risk of asthma was carried out with primary level data in UK Biobank using logistic regression in males and females separately, adjusted for age, the first 20 principal components, and genotyping array. Logistic regression with the inclusion of a SNP \* sex interaction term was also used to test for an interaction between SNP and sex on risk of asthma according to the following model:  $\text{Logit}[p] = \beta_0 + \beta_1 \text{SEX} + \beta_2 \text{SNP} + \beta_3 \text{SEX} * \text{SNP}$ . Interaction P-values (P-int) were considered significant at the Bonferroni-corrected threshold for testing 211 loci ( $0.05/211=2.4 \times 10^{-4}$ ). Sex-stratified and gene-sex interaction analyses were performed with STATA (v15.0, StataCorp LP, Texas, USA).

## **Evaluation of asthma loci as a function of cumulative genetic burden**

Primary level data from UK Biobank were also used to generate weighted genetic risk scores (GRSs) for the 66 novel, 145 known, or all 211 loci. For each variant, the number of risk alleles was multiplied by its respective weight (i.e., the Ln of the OR) obtained from the GWAS analysis carried out in UK Biobank and summed together across all variants to generate the three GRSs. Sex-specific weighted GRSs were also constructed for the 66 novel, 145 known, or all 211 loci using effect sizes obtained from the logistic regressions carried out in males and female separately. Association between weighted GRSs and asthma in all subjects and in males and females separately was tested by classifying participants into deciles according to the distribution of weighted GRS and using logistic regression to calculate ORs for subjects in the highest decile compared to the lowest decile, with adjustment for age, sex (if applicable), the first 20 principal components, and genotyping array. Tests for an interaction between sex and the GRS on risk of asthma were carried out by including a sex \* sex-specific GRS interaction term in the logistic regression comparing the highest versus lowest decile or by including the sex-interaction term in a model across all deciles, with inclusion of the same covariates. Sex-stratified and gene-sex interaction analyses were performed with STATA (v15.0, StataCorp LP, Texas, USA).

## **Association of novel loci with asthma-specific phenotypes, allergic-specific phenotypes, lung function, and other disease traits**

Association signals at the novel loci were characterized further by testing the lead SNPs for association with asthma- vs. allergic disease-specific phenotypes as well as lung function traits in UK Biobank. Asthma-specific cases were defined as positive for 6152\_8 (doctor diagnosed asthma) or ICD10 J45 (asthma)/J46 (severe asthma) or self-reported asthma and negative for

6152\_9 (doctor diagnosed allergic disease), ICD10 J30 (hay fever) /L20 (dermatitis and eczema), and self-reported hay fever/eczema/allergy/allergy to house dust mite. Allergic disease-specific cases were defined as positive for 6152\_9 (doctor diagnosed allergic disease), ICD10 J30 (hay fever) /L20 (dermatitis and eczema), or self-reported hay fever/eczema/allergy/allergy to house dust mite and negative for 6152\_8 (doctor diagnosed asthma), ICD10 J45/J46, and self-reported asthma. Controls were defined as individuals free from 6152\_8 (doctor diagnosed asthma), 6152\_9 (doctor diagnosed allergic disease), ICD10 J45/J46/J30/L20, and self-reported hay fever/eczema/allergy/allergy to HDM. These definitions lead to the inclusion of 33,830 cases/329,321 controls for the asthma-specific phenotype and 93,468 cases/329,321 controls for the allergic disease-specific phenotype in UK Biobank. The lead SNPs were also tested for association with forced expiratory volume in 1-second spirometry (FEV<sub>1</sub>; field codes 3063, N=446,697) and forced vital capacity spirometry (FVC, field code 3062, N=446,697). Both lung function traits were log-transformed prior to analysis and a Bonferroni-corrected  $P=7.6 \times 10^{-4}$  (0.05/66) for testing 66 loci was considered as the threshold for significant association with FEV<sub>1</sub> and FVC. Analyses were performed with BOLT-LMM V2.3.2 using a standard (infinitesimal) mixed model to correct structure due to relatedness, ancestral heterogeneity, or other factors, with adjustment for age, sex, the first 20 principal components, and genotyping array<sup>73</sup>. Evaluation of the novel asthma loci for association with other disease traits was carried out using the Phenoscanner database<sup>31</sup>. Only associations with  $P < 5.0 \times 10^{-8}$  and derived from SNPs in high LD ( $r^2 > 0.8$ ) with our lead GWAS variants are reported.

## **Enrichment of asthma loci in epigenetic marks**

Enrichment of asthma-associated variants in DNase I hypersensitive sites (peaks) was determined using the GWAS Analysis of Regulatory or Functional Information Enrichment with LD correction (GARFIELD) method<sup>76</sup>. Briefly, GARFIELD leverages GWAS findings with regulatory or functional annotations (primarily from ENCODE and Roadmap epigenomics data)<sup>77</sup> to find features relevant to a phenotype of interest. It performs greedy pruning of GWAS SNPs ( $r^2 > 0.1$ ) and then annotates them based on functional information overlap. Next, it quantifies ORs at various significance cutoffs and assesses them by employing a generalized linear model framework, while matching for minor allele frequency, distance to nearest transcription start site and number of LD proxies ( $r^2 > 0.8$ ).

## **Data-driven Expression-Prioritized Integration for Complex Traits (DEPICT)**

Additional gene prioritization analysis and tissue enrichment analysis were carried out by DEPICT (version 1.1)<sup>78</sup> using 33,017 genome-wide significant SNPs ( $P < 5.0 \times 10^{-8}$ ) associated with asthma that were identified in the meta-analysis with UK Biobank and TAGC. Both nominal P-values and false discovery rates (FDRs) were calculated for gene set enrichment and tissue enrichment.

## **Heritability**

Heritability was calculated with the INDI-V calculator under multifactorial liability threshold model. The baseline population risk (K) was set at 15%<sup>25</sup> and twin heritability ( $h_L^2$ ) was set at 65% (average of 35% and 95%)<sup>3</sup>. GWAS results from UK Biobank were used to estimate the heritability attributable to asthma-associated variants at the 66 novel loci.

## Expression Quantitative Trait Locus (eQTL) analyses

Functional evaluation of SNPs at the 66 novel asthma loci and prioritization of candidate causal genes was determined using multi-tissue eQTL data from the GTEx Project (version 6)<sup>79</sup> and the PhenoScanner database<sup>31</sup>. Consideration was only given to *cis* eQTLs with  $P < 5.0 \times 10^{-8}$  that were derived from our lead GWAS variants or proxy SNPs in high LD ( $r^2 > 0.8$ ).

## *In vivo* evaluation of airway hyperreactivity after pharmacological perturbation of CD52

Female BALB/cByJ mice between 6-8 weeks of age were immunized on day 1 with 100µg of house dust mite (HDM) in 2mg of aluminum hydroxide (alum) by intraperitoneal (*i.p.*) injection. On day 7, mice were randomly assigned to three groups. Group 1 was intraperitoneally administered 500µg of monoclonal αCD52 IgG2a antibody (clone BTG-2G, MBL International, Woburn, MA) and Group 2 was intraperitoneally administered 500µg IgG2a isotype control antibody (clone 2A3, BioXCell, West Lebanon, NH). Group 3 only received PBS as a control. On days 8, 9 and 10, mice in Groups 1 and 2 were intravenously (*i.v.*) administered 250µg of the αCD52 antibody or isotype control antibody, respectively, and simultaneously challenged intranasally (*i.n.*) with 50µg HDM. Mice in Group 3 were only challenged intranasally with PBS on days 8, 9 and 10. On day 11, airway hyperreactivity was measured by invasive plethysmography. Lung resistance and dynamic compliance at baseline and in response to sequentially increasing doses of methacholine (5-40mg/ml) was measured using a Buxco FinePointe (Data Sciences International, St. Paul, MN) respiratory system in tracheostomized immobilized mice that were mechanically ventilated under general anesthesia, as described elsewhere<sup>80</sup>. After airway hyperreactivity measurements were completed, the trachea was cannulated and 2ml of PBS was used to collect bronchial alveolar lavage (BAL) fluid for

analysis of immune cells by flow cytometry. All animal studies were approved by the USC Keck School of Medicine Institutional Animal Care and Use Committee and conducted in accordance with the Department of Animal Resources' guidelines.

## Flow Cytometry

Single cell suspensions were prepared from lung and spleen tissue in accordance with standard protocols. Immune cells were quantified by means of flow cytometry after staining cells with phycoerythrin (PE)-anti-CD52 (clone BTG-2G, MBL International, Woburn, MA), allophycocyanin (APC)/Cy7-anti-CD45 (clone 30-F11), fluorescein isothiocyanate (FITC)-anti-CD19 (clone MB19-1), peridinin-chlorophyll-protein complex (PerCP)/Cy5.5-anti-CD3 $\epsilon$  (clone 17A2), brilliant violet 421<sup>TM</sup> (BV)-anti-CD4 (clone GK1.5), PE/Cy7-anti-CD8 (clone 53-6.7), (BioLegend, San Diego, CA) in the presence of anti-mouse FC-block (clone 2.4G2, BioXcell, West Lebanon, NH). Leukocytes in BAL fluid were quantified after staining cells with PE-anti-Siglec-F (clone E50-2440, BD Biosciences, San Jose, CA), PE/Cy7-anti-CD45 (clone 30-F11), APC/Cy7-anti-CD11c (clone N418), FITC-anti-CD19 (clone MB19-1), PerCP/Cy5.5-anti-CD3 $\epsilon$  (clone 17A2), APC-anti-Gr-1 (clone RB6-8C5) (BioLegend, San Diego, CA), and eFluor450-anti-CD11b (clone M1/70, eBioscience, San Diego, CA) in the presence of anti-mouse FC-block (clone 2.4G2, BioXcell, West Lebanon, NH). The gating strategy used with these antibodies is shown in **Supplemental Fig. 5**. CountBright absolute count beads (Thermo Fisher Scientific, Waltham, Mass) were used to calculate absolute cell number, according to the manufacturer's instructions. At least 10<sup>5</sup> CD45<sup>+</sup> cells were acquired on a BD FACSCanto II (BD Biosciences). Data were analyzed with FlowJo software (TreeStar, Ashland, OR).



## **Histology**

Lungs were harvested, fixed overnight in 4% paraformaldehyde, and embedded in paraffin. 4 $\mu$ m sections were cut and stained with H&E according to standard protocols. Images of the H&E-stained slides were acquired with a KeyenceBZ-9000 microscope (Keyence, Itasca, Ill) and analyzed for the number of inflammatory cells and thickness of the airway epithelium with the ImageJ Analysis Application (NIH & LOCI, University of Wisconsin).

## **Statistical Analysis**

Differences in measured variables between groups of mice were determined by unpaired Student's t-test (SAS 9.3; SAS Institute Inc). Data are expressed as mean  $\pm$  SE and differences were considered statistically significant at  $P < 0.05$ .

**URLs.** UK Biobank, <http://www.ukbiobank.ac.uk/>; Trans-National Asthma Genetic Consortium, <https://www.ebi.ac.uk/gwas/downloads/summary-statistics>; Genotype-Tissue Expression Portal, <http://gtexportal.org/>; Phenoscanner, <http://www.phenoscaner.medschl.cam.ac.uk/phenoscaner>; GARFIELD, <https://www.ebi.ac.uk/birney-srv/GARFIELD/>; DEPICT, <http://www.broadinstitute.org/mpg/depict>; R statistical software, <http://www.R-project.org/>;

## **Data Availability**

Full summary statistics relating to the GWAS meta-analysis will be deposited with The NHGRI-EBI Catalog of published genome-wide association studies: <https://www.ebi.ac.uk/gwas/>. All other relevant data will be made available by the corresponding author upon reasonable request.

## **Acknowledgments**

This work was supported, in part, by National Institutes of Health Grants R01HL133169, R01ES021801, R01ES025786, R01ES022282, R21ES024707, R21AI109059, P30ES007048, and P01ES022845, and U.S. EPA Grant RD83544101. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## **Author Contributions**

Concept and design: Y.H., J.A.H., and H.A. Acquisition, analysis, and interpretation of data: Y.H., Q.J., P.S.J., B.P.H., C.P., P.H., J.G., N.C.W., E.E., F.G., O.A., J.A.H., and H.A. Drafting of the manuscript: Y.H., J.H., and H.A. Critical revision of the manuscript for important intellectual content: All authors.

## **Disclosures**

None.

## References

- 1 Global Asthma Network Report (<http://www.globalasthmareport.org/>). (2018).
- 2 Holt, P. G. & Sly, P. D. Viral infections and atopy in asthma pathogenesis: new rationales for asthma prevention and treatment. *Nat. Med.* **18**, 726-735, doi:10.1038/nm.2768 (2012).
- 3 Ober, C. & Yao, T. C. The genetics of asthma and allergic disease: a 21st century perspective. *Immunol. Rev.* **242**, 10-30, doi:10.1111/j.1600-065X.2011.01029.x (2011).
- 4 Vinkhuyzen, A. A., Wray, N. R., Yang, J., Goddard, M. E. & Visscher, P. M. Estimation and partition of heritability in human populations using whole-genome analysis methods. *Annu. Rev. Genet.* **47**, 75-95, doi:10.1146/annurev-genet-111212-133258 (2013).
- 5 Moffatt, M. F. *et al.* Genetic variants regulating ORMDL3 expression contribute to the risk of childhood asthma. *Nature* **448**, 470-473, doi:10.1038/nature06014 (2007).
- 6 Moffatt, M. F. *et al.* A large-scale, consortium-based genomewide association study of asthma. *N. Engl. J. Med.* **363**, 1211-1221 (2010).
- 7 Sleiman, P. M. *et al.* Variants of DENND1B associated with asthma in children. *N. Engl. J. Med.* **362**, 36-44, doi:10.1056/NEJMoa0901867 (2010).
- 8 Ferreira, M. A. *et al.* Identification of IL6R and chromosome 11q13.5 as risk loci for asthma. *Lancet* **378**, 1006-1014, doi:10.1016/S0140-6736(11)60874-X (2011).
- 9 Hirota, T. *et al.* Genome-wide association study identifies three new susceptibility loci for adult asthma in the Japanese population. *Nat. Genet.* **43**, 893-896, doi:10.1038/ng.887 (2011).
- 10 Noguchi, E. *et al.* Genome-wide association study identifies HLA-DP as a susceptibility gene for pediatric asthma in Asian populations. *PLoS Genet* **7**, e1002170, doi:10.1371/journal.pgen.1002170 (2011).
- 11 Torgerson, D. G. *et al.* Meta-analysis of genome-wide association studies of asthma in ethnically diverse North American populations. *Nat. Genet.* **43**, 887-892 (2011).
- 12 Forno, E. *et al.* Genome-wide association study of the age of onset of childhood asthma. *J. Allergy Clin. Immunol.* **130**, 83-90 e84, doi:10.1016/j.jaci.2012.03.020 (2012).
- 13 Lasky-Su, J. *et al.* HLA-DQ strikes again: genome-wide association study further confirms HLA-DQ in the diagnosis of asthma among adults. *Clin. Exp. Allergy* **42**, 1724-1733, doi:10.1111/cea.12000 (2012).
- 14 Ding, L. *et al.* Rank-based genome-wide analysis reveals the association of ryanodine receptor-2 gene variants with childhood asthma among human populations. *Human genomics* **7**, 16, doi:10.1186/1479-7364-7-16 (2013).
- 15 Galanter, J. M. *et al.* Genome-wide association study and admixture mapping identify different asthma-associated loci in Latinos: the Genes-environments & Admixture in Latino Americans study. *J. Allergy Clin. Immunol.* **134**, 295-305, doi:10.1016/j.jaci.2013.08.055 (2014).
- 16 Pickrell, J. K. *et al.* Detection and interpretation of shared genetic influences on 42 human traits. *Nat. Genet.* **48**, 709-717, doi:10.1038/ng.3570 (2016).
- 17 White, M. J. *et al.* Novel genetic risk factors for asthma in African American children: Precision Medicine and the SAGE II Study. *Immunogenetics* **68**, 391-400, doi:10.1007/s00251-016-0914-1 (2016).
- 18 Almoguera, B. *et al.* Identification of Four Novel Loci in Asthma in European American and African American Populations. *Am. J. Respir. Crit. Care Med.* **195**, 456-463, doi:10.1164/rccm.201604-0861OC (2017).

- 19 Yan, Q. *et al.* A meta-analysis of genome-wide association studies of asthma in Puerto Ricans. *Eur. Respir. J.* **49**, doi:10.1183/13993003.01505-2016 (2017).
- 20 Ferreira, M. A. *et al.* Shared genetic origin of asthma, hay fever and eczema elucidates allergic disease biology. *Nat. Genet.* **49**, 1752-1757, doi:10.1038/ng.3985 (2017).
- 21 Demenais, F. *et al.* Multiancestry association study identifies new asthma risk loci that colocalize with immune-cell enhancer marks. *Nat. Genet.* **50**, 42-53, doi:10.1038/s41588-017-0014-7 (2018).
- 22 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, doi:10.1038/s41588-018-0121-0 (2018).
- 23 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*, doi:10.1016/S2213-2600(19)30055-4 (2019).
- 24 Shrine, N. *et al.* Moderate-to-severe asthma in individuals of European ancestry: a genome-wide association study. *Lancet Respir Med* **7**, 20-34, doi:10.1016/S2213-2600(18)30389-8 (2019).
- 25 Vicente, C. T., Revez, J. A. & Ferreira, M. A. R. Lessons from ten years of genome-wide association studies of asthma. *Clin Transl Immunology* **6**, e165, doi:10.1038/cti.2017.54 (2017).
- 26 Kim, K. W. & Ober, C. Lessons Learned From GWAS of Asthma. *Allergy Asthma Immunol Res* **11**, 170-187, doi:10.4168/aair.2019.11.2.170 (2019).
- 27 Igartua, C. *et al.* Ethnic-specific associations of rare and low-frequency DNA sequence variants with asthma. *Nat Commun* **6**, 5965, doi:10.1038/ncomms6965 (2015).
- 28 London, S. J. & Romieu, I. Gene by environment interaction in asthma. *Annu. Rev. Public Health* **30**, 55-80, doi:10.1146/annurev.publhealth.031308.100151 (2009).
- 29 Bonnelykke, K. & Ober, C. Leveraging gene-environment interactions and endotypes for asthma gene discovery. *J. Allergy Clin. Immunol.* **137**, 667-679, doi:10.1016/j.jaci.2016.01.006 (2016).
- 30 Demenais, F. *et al.* Multiancestry association study identifies new asthma risk loci that colocalize with immune-cell enhancer marks. *Nat. Genet.* **50**, 42-53, doi:10.1038/s41588-017-0014-7 (2018).
- 31 Staley, J. R. *et al.* PhenoScanner: a database of human genotype-phenotype associations. *Bioinformatics* **32**, 3207-3209, doi:10.1093/bioinformatics/btw373 (2016).
- 32 Pasek, R. C., Berbari, N. F., Lewis, W. R., Kesterson, R. A. & Yoder, B. K. Mammalian Clusterin associated protein 1 is an evolutionarily conserved protein required for ciliogenesis. *Cilia* **1**, 20, doi:10.1186/2046-2530-1-20 (2012).
- 33 Evan, J. R., Bozkurt, S. B., Thomas, N. C. & Bagnato, F. Alemtuzumab for the treatment of multiple sclerosis. *Expert Opin Biol Ther* **18**, 323-334, doi:10.1080/14712598.2018.1425388 (2018).
- 34 Li, Z., Richards, S., Surks, H. K., Jacobs, A. & Panzara, M. A. Clinical pharmacology of alemtuzumab, an anti-CD52 immunomodulator, in multiple sclerosis. *Clin. Exp. Immunol.* **194**, 295-314, doi:10.1111/cei.13208 (2018).
- 35 Camms Trial Investigators *et al.* Alemtuzumab vs. interferon beta-1a in early multiple sclerosis. *N. Engl. J. Med.* **359**, 1786-1801, doi:10.1056/NEJMoa0802670 (2008).

- 36 Cohen, J. A. *et al.* Alemtuzumab versus interferon beta 1a as first-line treatment for patients with relapsing-remitting multiple sclerosis: a randomised controlled phase 3 trial. *Lancet* **380**, 1819-1828, doi:10.1016/S0140-6736(12)61769-3 (2012).
- 37 Coles, A. J. *et al.* Alemtuzumab for patients with relapsing multiple sclerosis after disease-modifying therapy: a randomised controlled phase 3 trial. *Lancet* **380**, 1829-1839, doi:10.1016/S0140-6736(12)61768-1 (2012).
- 38 Keating, M. J. *et al.* Therapeutic role of alemtuzumab (Campath-1H) in patients who have failed fludarabine: results of a large international study. *Blood* **99**, 3554-3561 (2002).
- 39 Rai, K. R. *et al.* Alemtuzumab in previously treated chronic lymphocytic leukemia patients who also had received fludarabine. *J. Clin. Oncol.* **20**, 3891-3897, doi:10.1200/JCO.2002.06.119 (2002).
- 40 Hillmen, P. *et al.* Alemtuzumab compared with chlorambucil as first-line therapy for chronic lymphocytic leukemia. *J. Clin. Oncol.* **25**, 5616-5623, doi:10.1200/JCO.2007.12.9098 (2007).
- 41 Wang, H. *et al.* Anti-mouse CD52 monoclonal antibody ameliorates intestinal epithelial barrier function in interleukin-10 knockout mice with spontaneous chronic colitis. *Immunology* **144**, 254-262, doi:10.1111/imm.12366 (2015).
- 42 Shi, H., Kichaev, G. & Pasaniuc, B. Contrasting the Genetic Architecture of 30 Complex Traits from Summary Association Data. *Am. J. Hum. Genet.* **99**, 139-153, doi:10.1016/j.ajhg.2016.05.013 (2016).
- 43 Kauffmann, F. & Demenais, F. Gene-environment interactions in asthma and allergic diseases: challenges and perspectives. *J. Allergy Clin. Immunol.* **130**, 1229-1240; quiz 1241-1222, doi:10.1016/j.jaci.2012.10.038 (2012).
- 44 Lee, J. U., Kim, J. D. & Park, C. S. Gene-Environment Interactions in Asthma: Genetic and Epigenetic Effects. *Yonsei Med. J.* **56**, 877-886, doi:10.3349/ymj.2015.56.4.877 (2015).
- 45 Khramtsova, E. A., Davis, L. K. & Stranger, B. E. The role of sex in the genomics of human complex traits. *Nat Rev Genet* **20**, 173-190, doi:10.1038/s41576-018-0083-1 (2019).
- 46 Randall, J. C. *et al.* Sex-stratified genome-wide association studies including 270,000 individuals show sexual dimorphism in genetic loci for anthropometric traits. *PLoS Genet* **9**, e1003500, doi:10.1371/journal.pgen.1003500 (2013).
- 47 Winkler, T. W. *et al.* The Influence of Age and Sex on Genetic Associations with Adult Body Size and Shape: A Large-Scale Genome-Wide Interaction Study. *PLoS Genet* **11**, e1005378, doi:10.1371/journal.pgen.1005378 (2015).
- 48 Pulit, S. L. *et al.* Meta-analysis of genome-wide association studies for body fat distribution in 694 649 individuals of European ancestry. *Hum. Mol. Genet.* **28**, 166-174, doi:10.1093/hmg/ddy327 (2019).
- 49 Hartiala, J. A. *et al.* Genome-wide association study and targeted metabolomics identifies sex-specific association of CPS1 with coronary artery disease. *Nat Commun* **7**, 10558, doi:10.1038/ncomms10558 (2016).
- 50 Szczeklik, W., Sanak, M. & Szczeklik, A. Functional effects and gender association of COX-2 gene polymorphism G-765C in bronchial asthma. *J. Allergy Clin. Immunol.* **114**, 248-253, doi:10.1016/j.jaci.2004.05.030 (2004).
- 51 Yang, K. D. *et al.* Polymorphism of the immune-braking gene CTLA-4 (+49) involved in gender discrepancy of serum total IgE levels and allergic diseases. *Clin. Exp. Allergy* **34**, 32-37 (2004).

- 52 Myers, R. A. *et al.* Genome-wide interaction studies reveal sex-specific asthma risk alleles. *Hum. Mol. Genet.* **23**, 5251-5259, doi:10.1093/hmg/ddu222 (2014).
- 53 Mersha, T. B. *et al.* Genomic architecture of asthma differs by sex. *Genomics* **106**, 15-22, doi:10.1016/j.ygeno.2015.03.003 (2015).
- 54 Moss, L. C., Gauderman, W. J., Lewinger, J. P. & Conti, D. V. Using Bayes model averaging to leverage both gene main effects and G x E interactions to identify genomic regions in genome-wide association studies. *Genet. Epidemiol.* **43**, 150-165, doi:10.1002/gepi.22171 (2019).
- 55 Lombardi, V., Singh, A. K. & Akbari, O. The role of costimulatory molecules in allergic disease and asthma. *Int. Arch. Allergy Immunol.* **151**, 179-189, doi:10.1159/000242355 (2010).
- 56 Heshmat, N. M. & El-Hadidi, E. S. Soluble CD30 serum levels in atopic dermatitis and bronchial asthma and its relationship with disease severity in pediatric age. *Pediatr. Allergy Immunol.* **17**, 297-303, doi:10.1111/j.1399-3038.2006.00405.x (2006).
- 57 Oflazoglu, E. *et al.* CD30 expression on CD1a+ and CD8+ cells in atopic dermatitis and correlation with disease severity. *Eur. J. Dermatol.* **18**, 41-49, doi:10.1684/ejd.2008.0309 (2008).
- 58 Vaine, C. A. & Soberman, R. J. The CD200-CD200R1 inhibitory signaling pathway: immune regulation and host-pathogen interactions. *Adv. Immunol.* **121**, 191-211, doi:10.1016/B978-0-12-800100-4.00005-2 (2014).
- 59 Prodeus, A. *et al.* A Synthetic Cross-Species CD200R1 Agonist Suppresses Inflammatory Immune Responses In Vivo. *Mol Ther Nucleic Acids* **12**, 350-358, doi:10.1016/j.omtn.2018.05.023 (2018).
- 60 Zhou, L. L. *et al.* AHI-1 interacts with BCR-ABL and modulates BCR-ABL transforming activity and imatinib response of CML stem/progenitor cells. *J. Exp. Med.* **205**, 2657-2671, doi:10.1084/jem.20072316 (2008).
- 61 Pernis, A. B. & Rothman, P. B. JAK-STAT signaling in asthma. *J. Clin. Invest.* **109**, 1279-1283, doi:10.1172/JCI15786 (2002).
- 62 Bronson, P. G. *et al.* Common variants at PVT1, ATG13-AMBRA1, AHI1 and CLEC16A are associated with selective IgA deficiency. *Nat. Genet.* **48**, 1425-1429, doi:10.1038/ng.3675 (2016).
- 63 Urm, S. H. *et al.* Asthma and risk of selective IgA deficiency or common variable immunodeficiency: a population-based case-control study. *Mayo Clin. Proc.* **88**, 813-821, doi:10.1016/j.mayocp.2013.05.021 (2013).
- 64 Pinchuk, L. M. & Filipov, N. M. Differential effects of age on circulating and splenic leukocyte populations in C57BL/6 and BALB/c male mice. *Immun Ageing* **5**, 1, doi:10.1186/1742-4933-5-1 (2008).
- 65 Ruck, T., Bittner, S., Wiendl, H. & Meuth, S. G. Alemtuzumab in Multiple Sclerosis: Mechanism of Action and Beyond. *Int J Mol Sci* **16**, 16414-16439, doi:10.3390/ijms160716414 (2015).
- 66 Canonica, G. W. *et al.* Therapeutic interventions in severe asthma. *World Allergy Organ J* **9**, 40, doi:10.1186/s40413-016-0130-3 (2016).
- 67 Wechsler, M. E. *et al.* Novel targeted therapies for eosinophilic disorders. *J. Allergy Clin. Immunol.* **130**, 563-571, doi:10.1016/j.jaci.2012.07.027 (2012).
- 68 Strati, P., Cortes, J., Faderl, S., Kantarjian, H. & Verstovsek, S. Long-term follow-up of patients with hypereosinophilic syndrome treated with Alemtuzumab, an anti-CD52



- antibody. *Clin Lymphoma Myeloma Leuk* **13**, 287-291, doi:10.1016/j.clml.2012.09.018 (2013).
- 69 Havrdova, E., Horakova, D. & Kovarova, I. Alemtuzumab in the treatment of multiple sclerosis: key clinical trial results and considerations for use. *Ther Adv Neurol Disord* **8**, 31-45, doi:10.1177/1756285614563522 (2015).
- 70 Holgate, R. G., Weldon, R., Jones, T. D. & Baker, M. P. Characterisation of a Novel Anti-CD52 Antibody with Improved Efficacy and Reduced Immunogenicity. *PLoS One* **10**, e0138123, doi:10.1371/journal.pone.0138123 (2015).
- 71 Galanter, J. M. *et al.* Cosmopolitan and ethnic-specific replication of genetic risk factors for asthma in 2 Latino populations. *J. Allergy Clin. Immunol.* **128**, 37-43 e12, doi:10.1016/j.jaci.2011.03.050 (2011).
- 72 Bycroft, C. *et al.* The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203-209, doi:10.1038/s41586-018-0579-z (2018).
- 73 Loh, P.-R. *et al.* Efficient Bayesian mixed-model analysis increases association power in large cohorts. **47**, 284 (2015).
- 74 Rueger, S., McDaid, A. & Kutalik, Z. Evaluation and application of summary statistic imputation to discover new height-associated loci. *PLoS Genet* **14**, e1007371, doi:10.1371/journal.pgen.1007371 (2018).
- 75 Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190-2191 (2010).
- 76 Iotchkova, V. *et al.* GARFIELD classifies disease-relevant genomic features through integration of functional annotations with association signals. *Nat. Genet.* **51**, 343-353, doi:10.1038/s41588-018-0322-6 (2019).
- 77 Roadmap Epigenomics Consortium. *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317-330, doi:10.1038/nature14248 (2015).
- 78 Pers, T. H. *et al.* Biological interpretation of genome-wide association studies using predicted gene functions. *Nat Commun* **6**, 5890, doi:10.1038/ncomms6890 (2015).
- 79 GTEx Consortium. The Genotype-Tissue Expression (GTEx) project. *Nat. Genet.* **45**, 580-585, doi:10.1038/ng.2653 (2013).
- 80 Maazi, H. *et al.* ICOS:ICOS-ligand interaction is required for type 2 innate lymphoid cell function, homeostasis, and induction of airway hyperreactivity. *Immunity* **42**, 538-551, doi:10.1016/j.immuni.2015.02.007 (2015).



**Table 1. 66 novel loci identified for asthma in meta-analysis of UK Biobank and TAGC.**

SNP	CHR	BP (hg19)	Nearest Gene(s)	EA/NEA	EAF*	UK Biobank	TAGC	Meta-analysis		
						P	P	P	Direction	P-het
rs2230624	1p36.22	12,175,658	<i>TNFRSF8</i>	G/A	0.99	4.3E-12	0.95	2.4E-09	++	4.4E-04
rs11577318	1p36.11	26,601,570	<i>CEP85</i>	G/A	0.18	4.8E-05	5.8E-05	2.8E-08	++	0.18
rs2228552	1p35.2	32,165,495	<i>COL16A1</i>	T/G	0.63	3.4E-08	0.09	2.2E-08	++	0.16
rs4127124	1q32.1	206,624,362	<i>SRGAP2</i>	A/C	0.49	4.6E-09	0.10	4.5E-09	++	0.11
rs72787718	2p23.1	30,474,351	<i>LBH</i>	A/G	0.23	3.2E-08	0.04	7.7E-09	++	0.26
rs68110799	2p16.1	60,945,555	<i>PAPOLG</i>	A/G	0.26	8.9E-08	0.01	5.2E-09	++	0.50
rs35548551	2p13.2	72,039,714	<i>DYSF</i>	T/G	0.12	4.4E-07	1.1E-04	2.7E-10	++	0.48
rs6741949	2q24.2	162,910,223	<i>DPP4</i>	G/C	0.56	1.0E-08	0.03	1.5E-09	++	0.29
rs2595389	2q32.1	187,534,183	<i>ITGAV</i>	T/G	0.54	1.0E-05	8.9E-04	3.8E-08	++	0.57
rs13101202	3p25.2	12,699,361	<i>RAF1</i>	C/A	0.17	2.1E-09	0.81	**1.5E-07	++	3.9E-03
rs2014490	3p24.3	16,972,211	<i>PLCL2</i>	G/A	0.65	9.2E-10	0.08	7.5E-10	++	0.10
rs2036226	3p24.3	23,597,986	<i>UBE2E2</i>	T/C	0.59	7.5E-08	4.7E-04	1.4E-10	++	0.82
rs73072483	3p21.2	50,771,624	<i>DOCK3</i>	G/A	0.87	1.4E-07	0.02	1.2E-08	++	0.47
rs9813229	3p14.3	56,741,585	<i>ARHGEF3</i>	A/G	0.62	1.4E-07	5.1E-03	2.6E-09	++	0.75
rs7372960	3p13	71,254,751	<i>FOXP1</i>	T/C	0.75	1.4E-07	0.05	3.5E-08	++	0.29
rs7622814	3q13.2	112,650,431	<i>CD200R1</i>	T/G	0.46	9.8E-09	9.5E-03	4.1E-10	++	0.46
rs9877891	3q13.33	119,260,866	<i>CD80</i>	C/T	0.18	2.7E-08	0.14	3.4E-08	++	0.11
rs17485347	3q26.2	169,127,519	<i>MECOM</i>	C/A	0.75	2.3E-06	2.0E-04	2.5E-09	++	0.45
rs6842889	4q31.21	145,479,880	<i>LOC105377462</i>	T/C	0.61	3.0E-06	3.1E-04	4.6E-09	++	0.49
rs11950815	5q31.1	130,955,487	<i>RAPGEF6</i>	G/A	0.63	8.6E-12	4.8E-04	2.0E-14	++	0.60
rs62379371	5q31.1	133,439,274	<i>LOC105379185</i>	G/A	0.95	6.5E-09	0.79	**3.2E-07	++	0.01
rs11746314	5q33.3	156,752,957	<i>CYFIP2</i>	G/A	0.06	4.2E-11	7.7E-03	2.1E-12	++	0.26
rs3777755	6p24.1	12,159,699	<i>HIVEP1</i>	C/T	0.70	2.1E-07	7.0E-03	5.3E-09	++	0.72
rs13203384	6p22.3	19,341,301	<i>MBOAT1</i>	A/G	0.21	1.8E-06	5.8E-03	3.5E-08	++	0.92
rs952579	6p22.3	21,884,440	<i>CASC15</i>	A/G	0.14	2.4E-09	0.19	7.1E-09	++	0.05
rs766406	6p22.2	26,319,588	<i>HIST1H4H/BTN3A2</i>	T/G	0.65	7.8E-09	6.4E-07	5.8E-14	++	0.20
rs9394288	6p21.31	35,222,010	<i>LOC101929285</i>	T/C	0.20	2.3E-09	0.38	2.6E-08	++	0.02
rs11759732	6q21	109,370,006	<i>SESNI</i>	A/G	0.85	1.9E-08	3.4E-03	2.5E-10	++	0.70
rs4526212	6q23.3	135,804,631	<i>AH11</i>	A/C	0.37	6.5E-11	7.2E-04	2.1E-13	++	0.64
rs13190880	6q24.2	143,222,770	<i>HIVEP2</i>	T/G	0.09	8.5E-08	5.1E-03	1.6E-09	++	0.72

rs431362	6q25.1	149,787,378	<i>ZC3H12D</i>	A/G	0.36	1.5E-08	2.8E-03	1.6E-10	++	0.72
rs4599658	6q25.2	155,074,505	<i>SCAF8</i>	G/T	0.47	1.8E-08	0.02	1.7E-09	++	0.37
rs73033536	7p22.2	3,149,883	<i>LOC105375130</i>	T/C	0.91	8.1E-11	1.6E-03	6.2E-13	++	0.52
rs7014953	8p23.1	8,168,413	<i>SGK223</i>	A/C	0.60	2.3E-07	9.9E-03	8.3E-09	++	0.65
rs34173062	8q24.3	145,158,607	<i>MAF1/SHARPIN</i>	A/G	0.07	1.3E-12	0.83	6.0E-10	++	5.2E-04
rs1537504	9q22.33	101,829,542	<i>COL15A1</i>	A/G	0.23	1.5E-07	1.1E-05	1.3E-11	++	0.29
rs4978607	9q32	117,508,437	<i>TNFSF15</i>	C/T	0.88	2.3E-08	0.19	4.5E-08	++	0.08
rs10986311	9q33.3	127,071,493	<i>NEK6</i>	C/T	0.37	2.7E-08	1.1E-05	2.1E-12	++	0.37
rs782134971	9q34.2	136,139,907	<i>ABO</i>	G/GAAACTGCC	0.25	1.5E-05	3.7E-04	3.0E-08	++	0.41
rs9284092	10p11.23	30,801,718	<i>MAP3K8</i>	G/T	0.73	8.5E-06	3.3E-04	1.5E-08	++	0.43
rs12769745	10q11.21	43,749,700	<i>RASGEF1A</i>	A/G	0.28	1.5E-06	4.8E-03	2.4E-08	++	0.95
rs1134777	10q22.2	75,538,651	<i>FUT11</i>	C/G	0.73	2.3E-07	0.02	1.4E-08	++	0.54
rs714417	11p11.2	45,247,176	<i>PRDM11</i>	C/T	0.69	7.5E-08	1.9E-05	9.5E-12	++	0.37
rs12303699	12q22	94,582,336	<i>PLXNC1</i>	A/G	0.38	1.9E-09	9.7E-04	7.5E-12	++	0.79
rs61960013	13q14.11	44,490,181	<i>LOC107984576</i>	G/C	0.21	7.0E-09	0.96	**6.1E-07	++	3.3E-03
rs3751289	14q23.1	61,983,943	<i>PRKCH</i>	G/A	0.22	9.8E-08	0.03	1.5E-08	++	0.35
rs10131197	14q32.12	93,015,394	<i>RIN3</i>	G/A	0.33	2.6E-10	0.02	4.6E-11	++	0.19
rs11259930	15q25.2	84,577,350	<i>ADAMTSL3</i>	A/G	0.47	8.5E-09	3.1E-05	1.4E-12	++	0.55
rs11645975	16p13.3	3,749,397	<i>TRAP1</i>	G/C	0.74	3.5E-05	1.7E-04	4.2E-08	++	0.28
rs223819	16q13	57,394,862	<i>CCL22</i>	T/C	0.91	2.3E-06	3.7E-03	3.0E-08	++	0.96
rs72842819	17p13.1	7,328,821	<i>C17orf74/SPEM2</i>	C/A	0.12	3.2E-09	0.06	1.6E-09	++	0.15
rs750065349	17p12	12,193,443	<i>MAP2K4</i>	GC/G	0.46	8.6E-10	0.38	1.1E-08	++	0.02
rs1991401	17q23.3	62,502,435	<i>CEP95/DDX5</i>	A/G	0.69	1.1E-08	0.50	**1.6E-07	++	0.02
rs111365807	17q25.1	73,825,463	<i>UNC13D</i>	C/G	0.12	1.2E-07	0.06	3.5E-08	++	0.27
rs76848919	17q25.3	76,352,554	<i>SOCS3</i>	C/G	0.20	5.6E-09	0.55	**2.7E-06	+-	4.5E-04
rs12956924	18q21.1	46,451,146	<i>SMAD7</i>	A/G	0.32	3.7E-08	0.21	**7.9E-08	++	0.08
rs117552144	19p13.3	3,136,091	<i>GNAI5</i>	T/C	0.06	3.6E-11	0.26	**3.5E-07	+-	1.2E-05
rs10420217	19p13.3	4,355,871	<i>MPND</i>	C/T	0.28	2.0E-08	0.21	**3.1E-05	+-	7.6E-05
rs755023315	19p13.3	6,579,029	<i>CD70</i>	G/GC	0.29	2.7E-09	0.27	1.5E-08	++	0.03
rs34006614	19p13.11	16,442,782	<i>KLF2</i>	T/C	0.33	1.1E-07	0.02	1.0E-08	++	0.43
rs143432496	19q13.32	45,252,714	<i>BCL3</i>	A/ATAT	0.27	2.3E-11	0.10	4.8E-11	++	0.04
rs8103278	19q13.32	46,370,381	<i>FOXA3</i>	G/A	0.65	1.2E-08	5.1E-05	3.1E-12	++	0.59
rs11670020	19q13.41	52,314,161	<i>FPR3/LOC105369197</i>	G/A	0.14	1.7E-06	5.6E-03	3.2E-08	++	0.93
rs8125525	20q13.12	45,681,788	<i>EYA2</i>	C/T	0.74	1.7E-07	0.02	1.5E-08	++	0.46

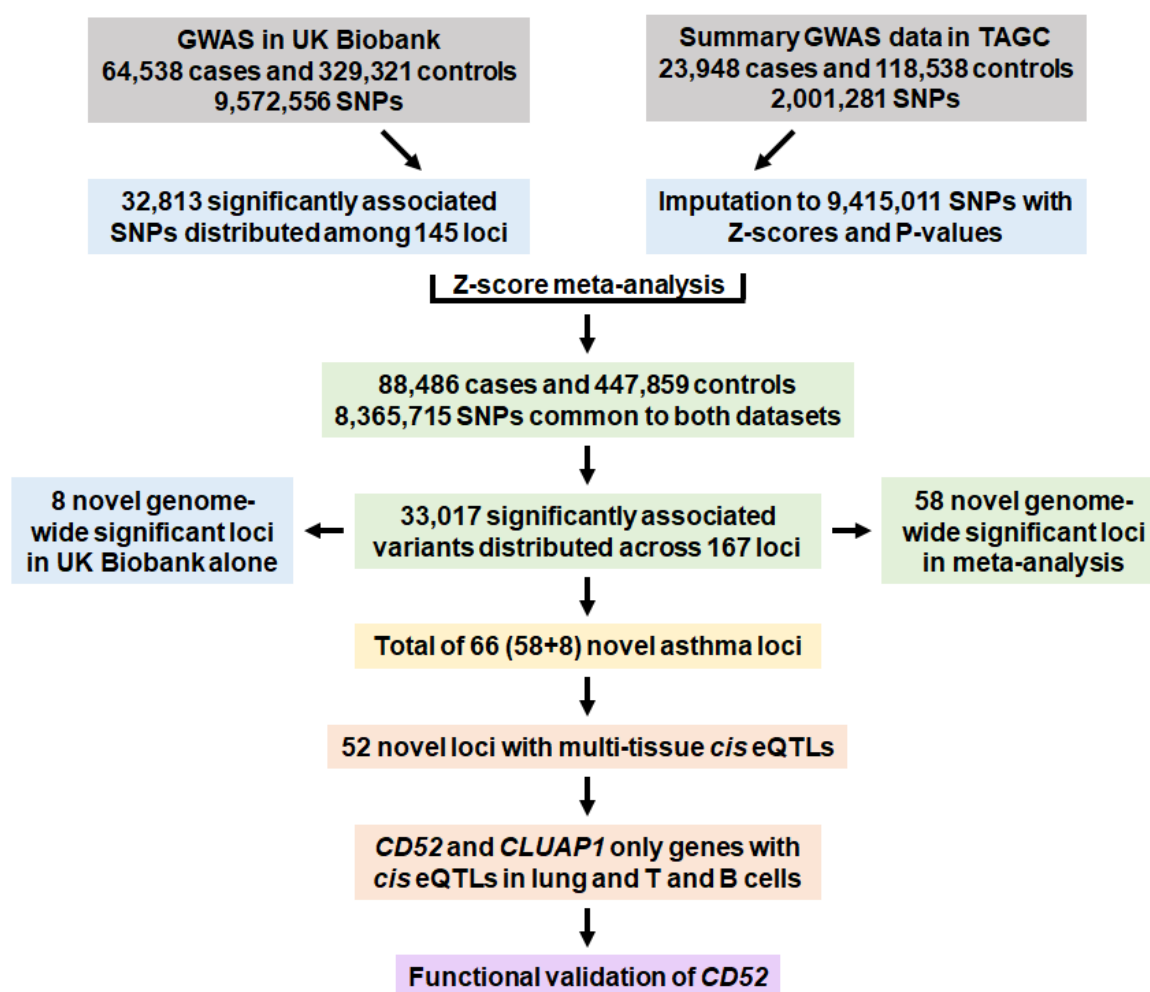
rs1736147	21q21.1	16,813,053	<i>LOC101927745</i>	G/A	0.57	1.1E-07	6.1E-03	2.5E-09	++	0.70
rs34846236	21q22.2	41,049,344	<i>PCP4</i>	G/GC	0.48	5.3E-06	2.2E-03	4.3E-08	++	0.78

EA/NEA, effect allele/non-effect allele.

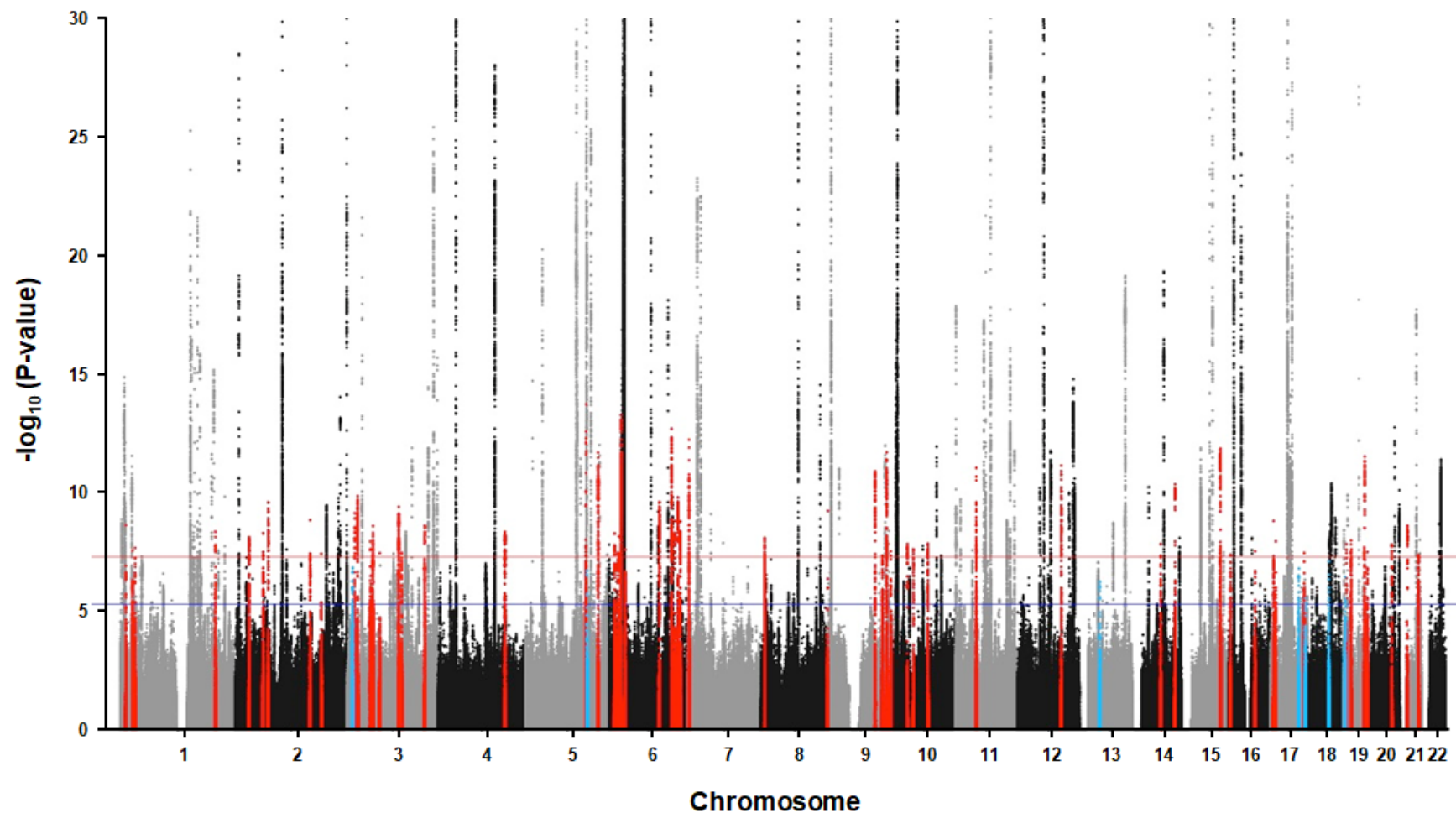
EAF, effect allele frequency (\*only available from UK Biobank).

P-het, p-value for heterogeneity between UK Biobank and TAGC.

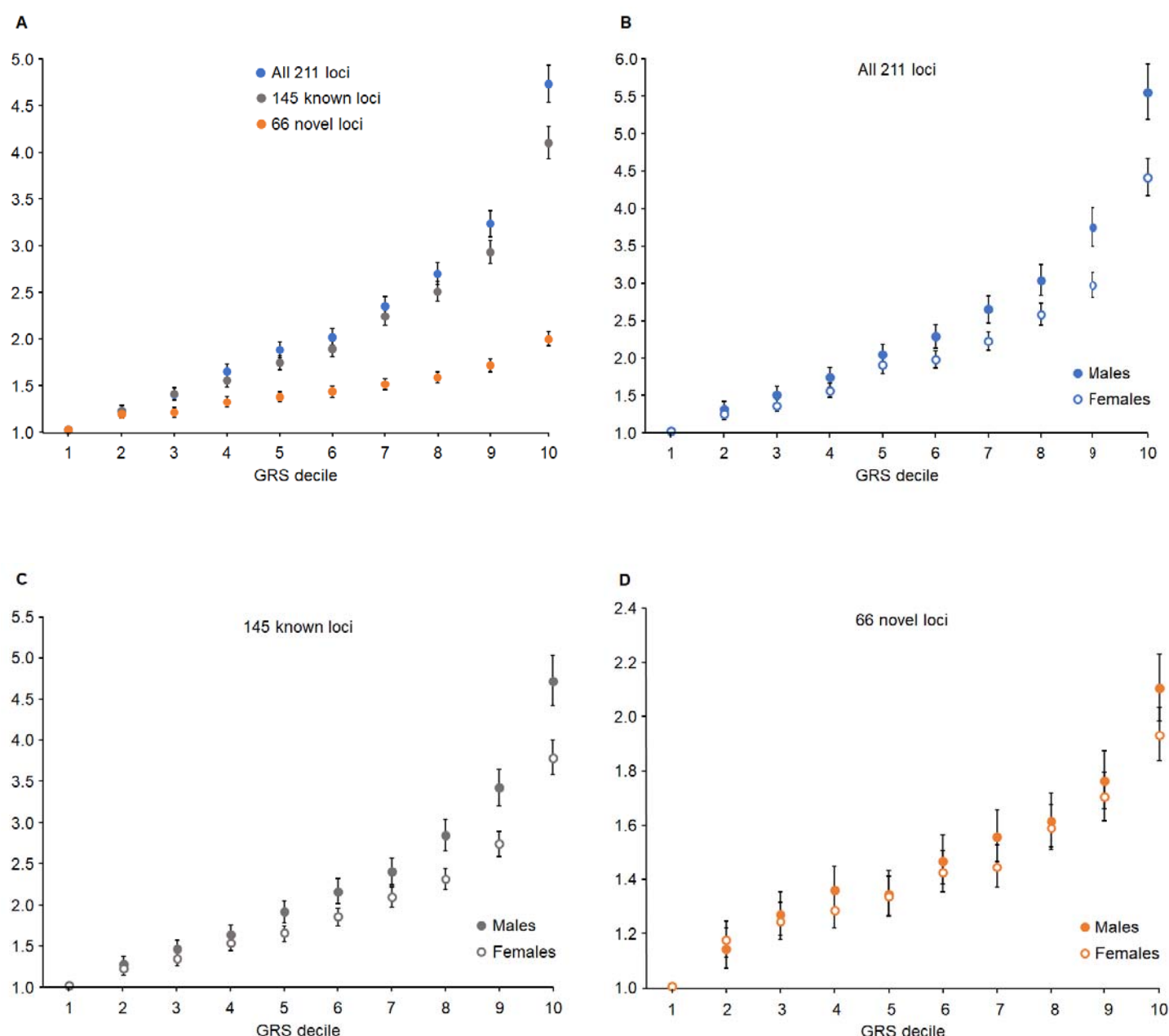
\*\*Locus was genome-wide significant ( $P < 5.0E-08$ ) in UK Biobank alone.



**Figure 1. Overview of genetic and functional analyses.** A GWAS was first carried out using primary level data in UK Biobank with ~9.5 million SNPs. In parallel, summary statistics were imputed based on publicly available GWAS data from TAGC. The results were combined in a Z-score meta-analysis that included 88,484 asthma cases and 447,859 controls with 8,365,715 variants common to both datasets. 58 previously unknown genome-wide significant loci for asthma were identified in the meta-analysis. Combined with the eight novel genome-wide significant loci from the GWAS analysis in UK Biobank alone, a total of 66 novel asthma susceptibility loci were identified. Follow up bioinformatics and eQTL analyses prioritized *CD52* for *in vivo* functional validation studies.

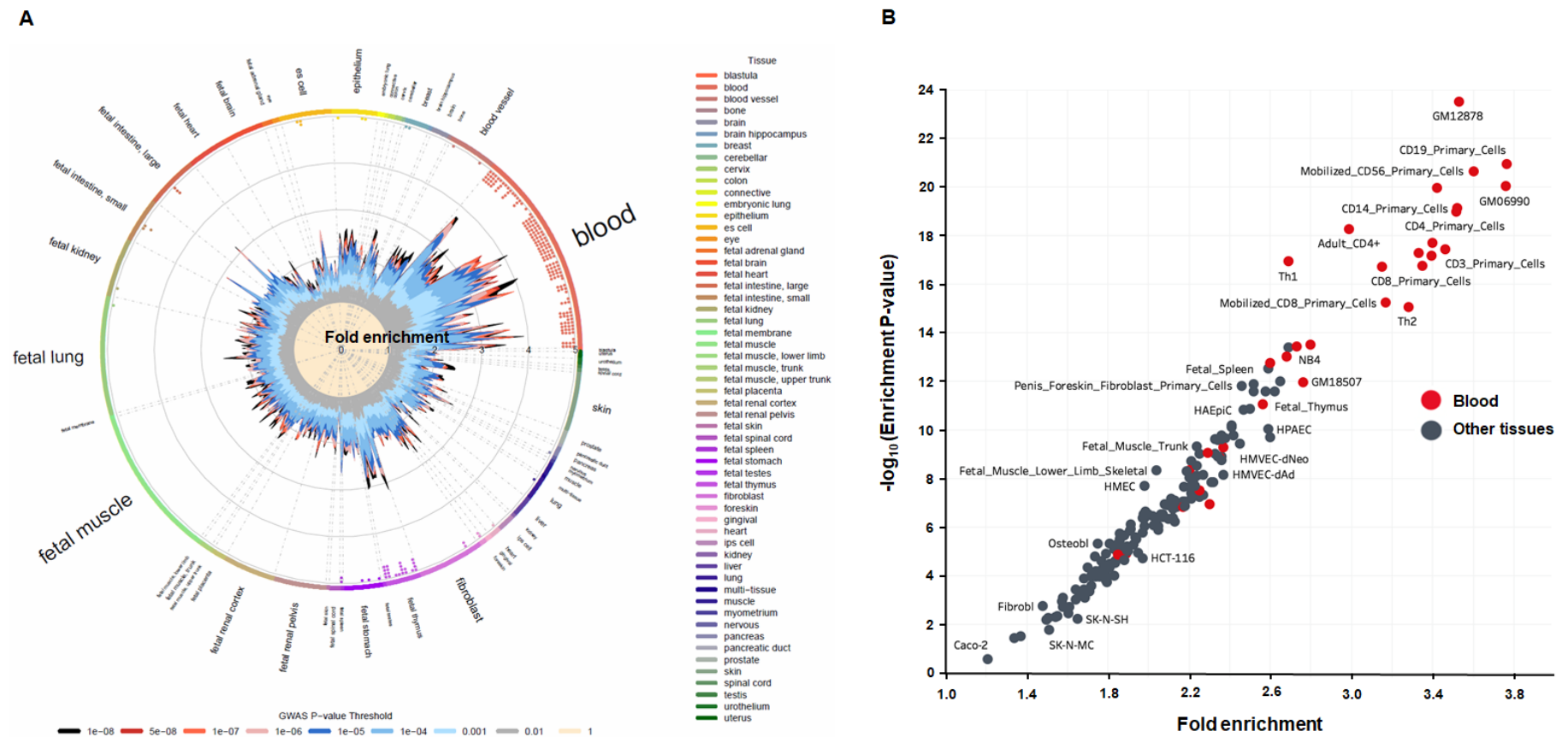


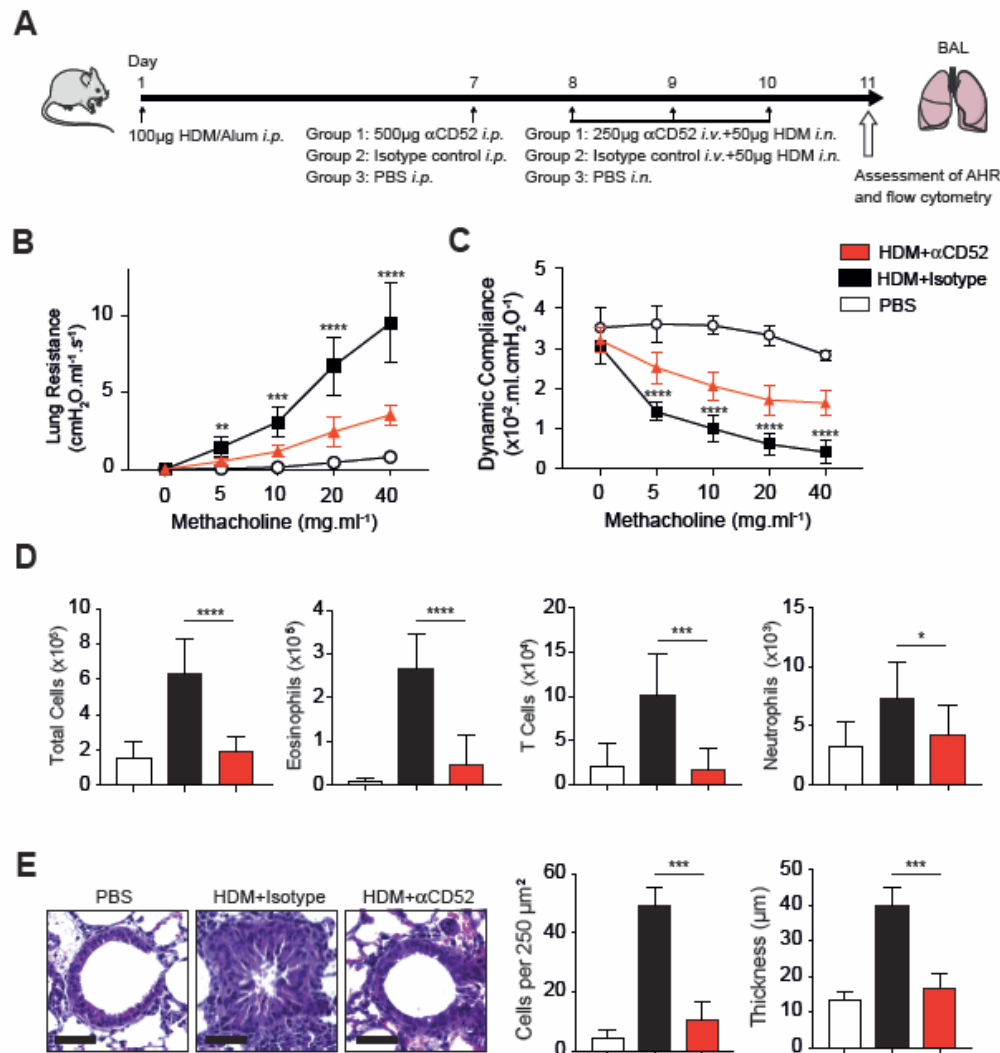
**Figure 2. GWAS analyses in UK Biobank and TAGC identifies 66 novel loci for asthma.** Of the 66 novel loci, 58 were significantly associated with asthma in the meta-analysis with UK Biobank and TAGC (red dots). Eight loci were associated with asthma in the GWAS analysis with UK Biobank alone but fell slightly below the genome-wide significance threshold in the meta-analysis with TAGC (light blue dots). The Z-score meta-analysis included a total of 88,486 cases and 447,859 controls from UK Biobank (64,538 asthma cases and 329,321 controls) and TAGC (23,948 asthma cases and 118,538 controls) and 8,365,715 SNPs common to both datasets. Genome-wide thresholds for significant ( $P=5.0 \times 10^{-8}$ ) and suggestive ( $P=5.0 \times 10^{-6}$ ) association are indicated by the horizontal red and dark blue lines, respectively. P-values are truncated at  $-\log_{10}(P)=30$ .



**Figure 3. Association of cumulative genetic risk with asthma in UK Biobank.** (A) Individuals in the highest decile for three weighted polygenic genetic risk scores (GRSs) based on the known or newly identified susceptibility loci have significantly elevated risk for asthma compared the lowest decile. (B-D) Compared to females, cumulative genetic risk for asthma was more significantly pronounced in males across the entire spectrum of sex-specific weighted GRSs constructed from all 211 ( $P_{\text{int}}=4.0 \times 10^{-18}$ ), 145 known ( $P_{\text{int}}=8.0 \times 10^{-18}$ ), and 66 novel ( $P_{\text{int}}=0.009$ ) loci. Interactions with sex were also significant when comparing the highest versus lowest decile for the GRSs constructed from all 211 ( $P_{\text{int}}=1.2 \times 10^{-9}$ ), 145 known ( $P_{\text{int}}=2.9 \times 10^{-9}$ ), and 66 novel ( $P_{\text{int}}=0.029$ ) loci.







**Figure 5. Treatment of mice with anti-CD52 antibody decreases airway hyperreactivity and pulmonary inflammation.** (A) The experimental protocol for testing the effect of a mouse anti-CD52 ( $\alpha$ CD52) antibody on pulmonary function and inflammation was designed to mimic the clinical protocol used to treat multiple sclerosis patients with a human  $\alpha$ CD52 antibody (alemtuzumab). Female BALB/cByJ mice were immunized on day 1 with 100µg of house dust mite (HDM) in 2mg of aluminum hydroxide (alum) by intraperitoneal (*i.p.*) injection. On day 7, mice were intraperitoneally administered with either 500µg of  $\alpha$ CD52 antibody (Group 1; n=8), 500µg isotype control antibody (Group 2; n=9), or PBS (Group 3; n=4). On days 8, 9 and 10, mice in Groups 1 and 2 were intravenously (*i.v.*) administered 250µg  $\alpha$ CD52 antibody or isotype control antibody, respectively, and simultaneously challenged intranasally (*i.n.*) with 50µg HDM. Mice in Group 3 were only challenged intranasally with PBS on days 8, 9 and 10. On day 11, airway hyperreactivity was measured by invasive plethysmography and leukocyte counts were determined by means of flow cytometry. Treatment of mice with  $\alpha$ CD52 antibody significantly reduced lung resistance (B) and improved dynamic compliance (C) in HDM-exposed mice. (D)  $\alpha$ CD52 antibody significantly decreased numbers of total cells, eosinophils, T cells, and neutrophils in bronchial alveolar lavage (BAL) of mice exposed to HDM compared to the control HDM-exposed isotype antibody group. (E) Inflammation and thickness of the airway epithelium was significantly decreased in lungs of HDM-exposed mice treated with  $\alpha$ CD52 antibody compared HDM-exposed mice treated with isotype control antibody. Data are shown as mean  $\pm$  SE. \* $P < 0.05$ ; \*\* $P < 0.005$ ; \*\*\* $P < 0.0005$ ; \*\*\*\* $P < 0.0001$ .