Prioritization of candidate causal genes in GWAS signals of asthma in UK Biobank

Kim Valette
Department of Molecular Medicine, Laval University
Zhonglin Li
Laboratory of Cardiovascular Pathobiology, Quebec Heart and Lung Institute/Research Center, Department of Surgery, Laval University, Quebec https://orcid.org/0000-0003-3659-2389
Valentin Bon-Baret
Laboratory of Cardiovascular Pathobiology, Quebec Heart and Lung Institute/Research Center, Department of Surgery, Laval University, Quebec
Arnaud Chignon
Laboratory of Cardiovascular Pathobiology, Quebec Heart and Lung Institute/Research Center, Department of Surgery, Laval University, Quebec
Jean-Christophe Bérubé
Institut universitaire de cardiologie et de pneumologie de Québec – Université Laval
Aida Eslami
Institut universitaire de cardiologie et de pneumologie de Québec – Université Laval
Jennifer Lamothe
Institut universitaire de cardiologie et de pneumologie de Québec – Université Laval
Nathalie Gaudreault
Quebec Heart and Lung Institute https://orcid.org/0000-0002-5351-596X
Philippe Joubert
Institut universitaire de cardiologie et de pneumologie de Québec, Quebec, Canada
Ma'en Obeidat
The University of British Columbia Centre for Heart Lung Innovation, St Paul's Hospital
Maarten van den Berge
University of Groningen
Wim Timens
University of Groningen
Don Sin
The University of British Columbia Centre for Heart Lung Innovation, St Paul's Hospital
David Nickle
Merck Research Laboratories Boston
Ke Hao
Icahn School of Medicine at Mount Sinai Department of Genetics and Genomic Sciences,
Catherine Labbé
Institut universitaire de cardiologie et de pneumologie de Québec – Université Laval

Krystelle Godbout
Institut universitaire de cardiologie et de pneumologie de Québec – Université Laval
Andréanne Côté
Institut universitaire de cardiologie et de pneumologie de Québec – Université Laval
Michel Laviolette
Institut universitaire de cardiologie et de pneumologie de Québec – Université Laval
Louis-Philippe Boulet
Institut universitaire de cardiologie et de pneumologie de Québec – Université Laval
Patrick Mathieu
Laboratory of Cardiovascular Pathobiology, Quebec Heart and Lung Institute/Research Center, Department of Surgery, Laval University, Quebec https://orcid.org/0000-0002-3805-2004
Sébastien Thériault
Institut universitaire de cardiologie et de pneumologie de Québec-Université Laval, Quebec City https://orcid.org/0000-0003-1893-8307
Yohan Bossé (**yohan.bosse@criucpq.ulaval.ca**)
Department of Molecular Medicine, Laval University https://orcid.org/0000-0002-3067-3711

Article

Keywords: asthma, candidate causal genes, genome-wide association study (GWAS)

Posted Date: September 9th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-68257/v1

Version of Record: A version of this preprint was published at Communications Biology on June 8th, 2021. See the published version at https://doi.org/10.1038/s42003-021-02227-6.

Abstract

To identify susceptibility loci and candidate causal genes of asthma, we performed a genome-wide association study (GWAS) in UK Biobank on a broad asthma definition (n = 56,167 asthma cases and 352,255 controls). We then carried out functional mapping through transcriptome-wide association studies (TWAS) and Mendelian randomization in lung (n = 1,038) and blood (n = 31,684) tissues. The GWAS revealed 72 asthma-associated loci from 116 independent significant variants (P_{GWAS}<5.0E-8). As expected, the yield of exonic variants associated with asthma was low, but nine were identified as potentially deleterious (CADD > 20) including a stop-gain mutation in the filaggrin (*FLG*) gene. The top lung TWAS gene on 17q12-q21 was *GSDMB* (P_{TWAS}=1.42E-54). Other TWAS genes of interest include *TSLP* on 5q22, *RERE* on 1p36, *CLEC16A* on 16p13, and *IL4R* on 16p12, which all replicated in GTEx lung (n = 515). A novel risk locus was also revealed by the lung asthma TWAS on 1q23.3 with the putative gene encoding the gamma chain of the high-affinity IgE receptor (*FCER1G*, P_{TWAS}=2.13E-6), which was also replicated in GTEx lung (P_{TWAS}=3.71E-7). By testing a comprehensive set of cells and tissues, we then demonstrated that the largest fold enrichment of regulatory and functional annotations among asthma-associated variants was in the blood. We mapped 485 eQTL-regulated genes associated with asthma in the blood and 50 of them were shown to be causally associated with asthma by Mendelian randomization. Prioritization of druggable genes revealed known (*IL4R*, *TSLP*, *IL6*, *TNFSF4*) and potentially new therapeutic targets for asthma.

Introduction

Asthma is still causing 420,000 deaths per year and afflicts 300 million individuals worldwide[1]. Our understanding of the genetics of asthma has progressed following the completion of large GWAS by international consortia, namely GABRIEL[2], EVE[3], TAGC[4], and CAAPA[5]. More recently, two groups of investigators tapped into the UK Biobank resource to delineate the genetics of childhood- vs. adult-onset asthma[6,7]. Together, approximately 200 genetic loci have been associated with asthma through GWAS. The objective of this study was to map candidate causal genes of asthma in lung and blood tissues. This was achieved in two steps. First, performing a case-control GWAS on a broad asthma definition in UK Biobank in order to physically define chromosome regions associated with asthma. Second, prioritizing candidate genes by mapping the effects of asthma-associated variants on protein-coding genes, gene expression and chromatin interaction sites using multiple approaches such as transcriptome-wide association study, colocalization, and Mendelian randomization.

Results

Asthma GWAS in UK Biobank

In total, 56,167 asthma cases and 352,255 controls of White British ancestry were selected from UK Biobank (see Methods). Demographics and clinical characteristics of cases and controls are in Table 1. The number of cases corresponds to an asthma population prevalence of 13.8%, which is consistent with the UK lifetime prevalence of patient-reported clinician-diagnosed asthma of 15.6%[8]. The granularity of asthma cases defined based on self-reported questionnaires, hospital records (ICD-9 and ICD-10), and primary care records is

provided in **Supplementary Fig. 1**. For GWAS analysis, 35,270,583 SNPs (filtered by MAF > 0.0001 and imputation info score > 0.3) were available for genetic association testing following standard quality controls and imputation. We observed no evidence of inflation in the test statistics with λ = 1.029 (**Supplementary Fig. 2**). The SNP-heritability on the liability scale was estimated at 11.3%. In total, 14,742 SNPs reached genome-wide significance (P_{GWAS}<5.0E-8) at 73 physically defined loci. Figure 1 shows the Manhattan plot and individual loci are listed in **Supplementary Table 1**. Nine of these loci are novel, with no genetic variant associated with asthma in the literature published before January 1st, 2020. Regional plots for these 9 loci are provided in Fig. 2. The locus 7p14 is characterized by only one rare SNP that passed the significance threshold (rs576468798, P_{GWAS}=2.00E-8, imputation info = 0.61). Allele frequencies in asthma cases (0.00033) and controls (0.00013) range within those observed in reference populations (TOPMed = 0.00015, 1000G European = 0.0006). Nevertheless, we discarded this locus as more validation is needed to robustly establish its association with asthma. We also evaluated the number of independent association signals within the 72 loci by conditional analysis. Sixteen loci had more than one independent association signals, ranging from 2 to 9 independent signals by locus, except for the MHC locus where we observed 12 independent signals. In total, 116 independent associations with asthma risk at a P_{GWAS}<5.0E-8 were observed (**Supplementary Table 2**).

Table 1
Demographics and clinical characteristics of asthma cases and controls in the UK Biobank

	Case n = 56,167	Control n = 352,255
Sex (% male)	42.5	46.4
Age (mean and range)	56.5 (40–71)	57.0 (39–73)
BMI (kg/m^2) (mean and range)	28.2 (13.1–69.0) [212]	27.3 (12.1–74.7) [1079]
Smoking status (%)	[247]	[247]
Never smokers	53.2	54.6
Former smokers	36.3	35.0
Current smokers	10.1	10.1
Lung function (mean and range)		
FEV1 (L)	2.71 (2.08–5.89) [8705]	3.08 (2.63–5.99) [32443]
FVC (L)	3.75 (3.04–7.95) [8688]	3.96 (3.23–7.99) [32383]
FEV1 (L)/FVC (L)	0.72 (0.24-1)	0.78 (0.17-1)
PEF (L/min)	431 (380–995) [8068]	440 (361–999) [28123]
Atopy (%)	45 [59]	21 [387]
Eosinophil count (g/L) (mean and range)	0.22 (0-5.4) [1780]	0.17 (0-9.6) [11039]
Number of missing values is shown in square brackets when applicable.		
BMI: Body mass index; FEV1: forced expiratory volume in 1 second; FVC: forced vital capacity; PEF: Peak expiratory flow.		

GWAS sensitivity analysis

GWAS-nominated loci were re-evaluated by changing exclusion criteria to define asthma cases and controls. The rationale was to evaluate the potential confounding effect of other lung diseases, smoking, and allergy. Genetic association analyses were thus performed in three case-control subsets. First, asthma cases and controls with other lung diseases were excluded. Individuals were excluded if they had self-reported or medical records consistent with the presence of chronic obstructive pulmonary disease (COPD), emphysema, chronic bronchitis, interstitial lung disease or alpha-1 antitrypsin deficiency. This results in the exclusion of 20,998 individuals and genetic analysis performed in 47,391 asthma cases and 340,033 controls. Second, we excluded all asthma cases and controls with a positive smoking history (i.e. former and current smokers). This results in the exclusion of 250,739 individuals and genetic analysis performed in 21,097 asthma cases and 136,586 controls. Third, we excluded control individuals with atopy, including hay fever, allergic rhinitis, and eczema/atopic dermatitis. This results in the exclusion of 84,113 individuals and genetic analysis with

268,142 controls (and the same number of asthma cases, n = 56,167). Note that the three lists of exclusion criteria were applied separately (not cumulatively) and specific UK Biobank data fields and codes used for excluding individuals in each case-control subset are provided in **Supplementary Table 14**.

Positional mapping of deleterious coding SNPs

Our first strategy to prioritize target genes within GWAS-nominated asthma loci was to map deleterious coding variants. A total of 354 exonic variants located in 27 loci were associated with asthma (P_{GWAS}<5.0E-8) or in LD ($r^2 \geq 0.6$, 1000G EUR) with asthma-associated variants (**Supplementary Table 3**). Two-thirds (236 out of 354) of these variants were located in the MHC locus. The top deleterious variants at this locus were rs9269958 in *HLA-DRB1* and rs2855430 in *COL11A2* with CADD scores of 57 and 33, respectively. However, association signals for these variants (P_{GWAS}=6.43E-5 and 6.83E-7) were much smaller compared to the sentinel variant (rs9273386, P_{GWAS}=2.11E-48). The extent of LD at this locus precluded firm conclusion. Outside of the MHC locus, we identified 8 nonsynonymous variants and 1 stop-gain variant with CADD score > 20 located at 7 loci (Table 2). Genes of known biological relevance were identified including filaggrin (*FLG*) on 1p21 and toll like receptor 10 (*TLR10*) on 4p14. On 17q12-q21, three potential target genes were identified, namely, *ERBB2*, *STARD3*, and *GSDMA*. Overall, the yield of candidate genes by mapping of deleterious coding variants was relatively low. This is consistent with previous GWAS results on asthma that showed more genetic associations in noncoding regions of the genome and suggests that most of the risk loci are likely to act through gene regulation.

Table 2
Deleterious coding SNPs associated with asthma or in LD with asthma-associated SNPs outside of the MHC locus.

Chr	Chr band	rsID	Position GRCh37	P_{GWAS}	CADD	Gene symbol	Gene name
1	1p36	rs2230624	12,175,658	1.99E-9	22.1	*TNFRSF8*	TNF receptor superfamily member 8
1	1q21	rs61816761	152,285,861	3.95E-22	36	*FLG-AS1/ FLG*	filaggrin
4	4p14	rs11096957	38,776,491	2.49E-10	21.9	*TLR10*	toll like receptor 10
5	5p15	rs16903574	14,610,309	5.3E-12	22.6	*FAM105A/ OTULINL*	OTU deubiquitinase with linear linkage specificity like
11	11q13	rs12146493	65,547,333	7.69E-6	22.2	*AP5B1*	adaptor related protein complex 5 subunit beta 1
12	12q21	rs3763978	71,533,534	2.6E-10	24.5	*TSPAN8*	tetraspanin 8
17	17q12	rs1058808	37,884,037	1.94E-26	23.5	*ERBB2*	erb-b2 receptor tyrosine kinase 2
17	17q12	rs1877031	37,814,080	4.71E-22	23.1	*STARD3*	StAR related lipid transfer domain containing 3
17	17q21	rs3894194	38,121,993	7.95E-33	21.9	*GSDMA*	gasdermin A

All variants are nonsynonymous except rs61816761 in the filaggrin gene that is a stop-gain.

Asthma TWAS in lung tissue

Summary statistics from the UK Biobank GWAS were integrated with our lung eQTL dataset (n = 1,038) to perform a TWAS on asthma. A total of 55 gene-asthma associations (corresponding to 69 probe sets) reached genome-wide significance (P_{TWAS}<2.51E-6) (Fig. 3 and **Supplementary Table 4**). Fifty-three of these genes are located in 21 distinct asthma-associated loci identified in the UK Biobank GWAS (Table 3). **Supplementary Fig. 4** shows the top lung TWAS genes per asthma-associated loci. The top TWAS signal is at the well-known asthma-associated locus on chromosome 17q12-q21. The lead TWAS target gene at this locus is *GSDMB* (P_{TWAS}=1.42E-54). However, nine additional statistically significant TWAS genes are identified including *ORMDL3* (P_{TWAS}=2.12E-44), *GSDMA* (P_{TWAS}=5.52E-23), and *PNMT* (P_{TWAS}=7.87E-23). LocusCompare plots showing the colocalization events for these TWAS genes on 17q12-q21 are provided in **Supplementary Fig. 5** and show that the P value distribution of eQTL for *GSDMB* colocalized better with that of the GWAS. The direction of effects, i.e. whether lower or higher predicted expression of these genes increased asthma risk are presented in Table 3, along with other TWAS genes found at asthma-associated loci. TWAS genes of known

biological interest in asthma include *IL1RL1* on 2q12, *TLR1* on 4p14, *TSLP* on 5q22, *SMAD3* on 15q22-q23, and *IL4R* on 16p12.

Table 3
Lung TWAS genes identified in asthma-associated loci

	Lung eQTL	Replication in GTEx lung*
Chr band	**Genes (direction, P_{TWAS})**	**Genes (direction, P_{TWAS})**
1p36	*RERE* (+, 4.43E-9)	*RERE* (+, 1.69E-9)
1q21	*LINGO4* (+, 6.45E-8)	*LINGO4* (+, 1.67E-6) → *FLG* (-, 3.16E-6)
2q12	*IL1RL1* (-, 4.01E-8) → *SLC9A2* (+, 1.41E-7)	*SLC9A2* (+, 1.72E-8) repl.: *IL1RL1* (-, 0.163)
2q37	*ING5* (-, 7.92E-8)	*RTP5* (-, 4.68E-15) → *D2HGDH* (+, 1.15E-14) → *PDCD1* (-, 9.96E-12) → *ING5* (-, 6.67E-9) → *BOK* (-, 2.52E-7)
3q27-q28	*LPP* (+, 1.33E-6)	*LINC01063* (-, 8.06E-15) repl.: *LPP* (-, 0.006)
4p14	*TLR1* (+, 5.15-17)	repl.: no model for *TLR1*
5q22	*TSLP* (+, 9.43E-14) → *CAMK4* (+, 2.94E-9) → *WDR36* (-, 2.21E-6)	*TSLP* (+, 3.54E-14) → *WDR36* (-, 1.19E-10) repl.: no model for *CAMK4*
5q31	*SEPT8* (-, 3.11E-23) → *PDLIM4* (+, 5.54E-19) → *SLC22A5* (+, 1.03E-18) → *P4HA2* (-, 6.29E-18) → *SLC22A4* (+, 1.50E-17) → *KIF3A* (+, 9.92E-10) → *HSPA4* (+, 1.98E-9) → *RAD50* (+, 5.60E-8) → *CSF2* (-, 1.43E-6)	*SLC22A5* (+, 2.91E-20) → *AFF4* (+, 3.16E-08) → *KIF3A* (-, 3.22E-07) repl.: *HSPA4* (+, 0.182), *RAD50* (+, 2.63E-4), no model for *SEPT8*, *PDLIM4*, *P4HA2*, *SLC22A4*, *CSF2*
6p22-p21	*HLA-DRB6* (+, 1.06E-20) → *HLA-DQB1* (-, 2.03E-14) → *HLA-DQB2* (-, 6.30E-14) → *HLA-DPB1* (-, 2.80E-10) → *TAP2* (-, 4.01E-8) → *PSMB9* (-, 2.18E-7) → *TRIM10* (+, 9.57E-7)	*HLA-DQA1* (-, 2.78E-51) → *HLA-DQB2* (+, 2.64E-40) → *HLA-DQB1* (-, 1.26E-39) → *HLA-DQA2* (+, 1.70E-34) → *HLA-DQB1-AS1* (-, 1.51E-27) → *C6orf47* (-, 8.88E-17) → *HLA-DRB1* (-, 1.77E-12) → *ZNRD1* (-, 1.21E-8) → *COL11A2* (+, 1.00E-7) → *LEMD2* (+, 1.02E-6) → *CFB* (+, 1.37E-6) → *DXO* (-, 2.98E-6) repl.: *HLA-DPB1* (+, 0.001), *TAP2* (-, 0.042), *PSMB9* (+, 0.574), no model for *HLA-DRB6*, *TRIM10*
6q22	*PTPRK* (+, 1.86E-10)	repl.: no model for *PTPRK*
9p24	*C9orf38* (+, 1.19E-12) → *KIAA1432* (+, 1.90E-6)	repl.: no model for *C9orf38*, *KIAA1432*

(+) and (−) indicate predicted gene expression positively or negatively associated with asthma risk. For loci with more than one TWAS genes, the genes are ordered by their level of significance and separated by arrows.

*All Bonferroni-corrected TWAS genes per loci found in GTEx lung are indicated as well as the results of TWAS genes identified in the lung eQTL dataset in order to seek for replication (P_{TWAS}<0.05) in GTEx lung.

	Lung eQTL	Replication in GTEx lung*
10p15	*RBM17* (-, 2.34E-8)	repl.: no model for *RBM17*
10q23	*TSPAN14* (-, 1.20E-6)	repl.: no model for *TSPAN14*
12q13	*CDK2* (+, 2.54E-10) → *FAM62A* (+, 6.03E-8) → *RDH16* (+, 1.81E-6)	*RPS26* (+, 1.29E-13) → *SUOX* (-, 1.21E-8) → *HSD17B6* (+, 3.99E-6) repl.: *RDH16* (+, 5.36E-4), no model for *CDK2*, *FAM62A*
13q32	*UBAC2* (-, 1.10E-12)	repl.: no model for *UBAC2*
15q22-q23	*SMAD3* (+, 5.41E-10) → *MAP2K5* (+, 2.74E-7)	*IQCH* (+, 6.75E-10) → *AAGAB* (-, 3.37E-6) repl.: *SMAD3* (-, 0.010), *MAP2K5* (+, 0.043)
16p13	*CLEC16A* (+, 4.71E-9)	*CLEC16A* (+, 8.47E-07)
16p12	*IL4R* (-, 5.94E-9)	*IL4R* (-, 1.00E-11)
17q12-q21.2	*GSDMB* (+, 1.42E-54) → *ORMDL3* (+, 2.12E-44) → *PERLD1* (+, 2.64E-26) → *GSDMA* (-, 5.52E-23) → *PNMT* (+, 7.87E-23) → *CASC3* (+, 1.53E-9) → *PSMD3* (-, 4.37E-9) → *SMARCE1* (+, 6.39E-9) → *CRKRS* (+, 2.76E-8) → *MED1* (+, 4.89E-7) → *IKZF3* (-, 6.24E-7)	*ORMDL3* (+, 1.05E-54) → *GSDMB* (+, 1.82E-47) → *GSDMA* (-, 8.97E-21) → *PNMT* (+, 7.03E-20) → *PGAP3* (+, 3.15E-19) repl.: *CASC3* (+, 0.005), *SMARCE1* (-, 0.003), no model for *PERLD1*, *PSMD3*, *CRKRS*, *MED1*, *IKZF3*
17q21.32	*KPNB1* (+, 4.88E-7)	repl.: no model for *KPNB1*
22q13	*PHF5A* (+, 2.17E-8) → *MEI1* (+, 1.91E-7)	*MEI1* (+, 4.33E-8) → *ACO2* (+, 1.04E-7) repl.: no model for *PHF5A*

(+) and (−) indicate predicted gene expression positively or negatively associated with asthma risk. For loci with more than one TWAS genes, the genes are ordered by their level of significance and separated by arrows.

*All Bonferroni-corrected TWAS genes per loci found in GTEx lung are indicated as well as the results of TWAS genes identified in the lung eQTL dataset in order to seek for replication (P_{TWAS}<0.05) in GTEx lung.

TWAS can also reveal novel risk loci owing to the resulting power of combining GWAS and eQTL. In this study, two TWAS genes are located in genomic loci that did not reach statistical significance in the GWAS. This includes the gene encoding the gamma chain of the high-affinity IgE receptor (*FCER1G*, z = 4.74, P_{TWAS}=2.13E-6) on chromosome 1q23.3 playing a key role in allergic reactions and DM1 protein kinase (*DMPK*, z = 4.83, P_{TWAS}=1.37E-6) on chromosome 19q13.32 with cellular antioxidant and pro-survival properties[9].

GTEx lung was used to validate the TWAS results. For the two novel asthma risk loci, *FCER1G* was replicated on 1q23.3 (z = 5.08, P_{TWAS}=3.71E-7), but not *DMPK* on 19q13.32 (z = 1.78, P_{TWAS}=0.075). Table 3 shows replication of TWAS results in GTEx lung for the 21 asthma-associated loci. For asthma loci with a single TWAS gene, consistency was observed for *RERE* on 1p36, *CLEC16A* on 16p13, and *IL4R* on 16p12. On 5q22, *TSLP* was the top TWAS gene in both our lung eQTL set and GTEx lung. On 17q12-q21, *GSDMB* and *ORMDL3* were switched as the top TWAS gene. In general, for asthma loci with multiple TWAS genes in our lung eQTL

dataset (the MHC locus for example), some of the genes were replicated in GTEx lung, but the ranking of genes based on level of significance changed, and sometimes different TWAS genes were observed in GTEx lung. Six TWAS genes were replicated, but with a different direction of effect, *SMAD3* on 15q22-q23 is an example. Finally, replication was not feasible for 24 TWAS genes observed in our lung eQTL dataset as they did not yield significant gene expression models in GTEx lung (Table 3).

To further filter lung TWAS genes, we used Bayesian colocalization tests for GWAS and lung eQTL signals in asthma risk loci. A high probability of shared GWAS and lung eQTL signals was observed for *GSDMB* on 17q12-q21 (PP4 = 0.84), *TLR1* on 4p14 (PP4 = 0.75), *TSPAN14* on 10q23 (PP4 = 0.72), *RERE* on 1p36 (PP4 = 0.71), and *UBAC2* on 13q32 (PP4 = 0.65) as well as two genes on 22q13: *PHF5A* (PP4 = 0.87) and *MEI1* (PP4 = 0.63). **Supplementary Table 5** shows the colocalization results for all TWAS genes identified in the lung eQTL dataset.

Cell and tissue functional enrichment of asthma-associated SNPs

We used GARFIELD[10] to evaluate the enrichment of asthma-associated loci in regulatory and functional annotations derived from ENCODE and the Roadmap Epigenomics Project. Figure 4 shows functional enrichment within DNase I hypersensitivity site (DHS) hotspots at two GWAS P value cut-offs. The largest fold enrichment values were in the blood. All results are summarized in **Supplementary Table 6**, along with other annotation types.

Functional mapping and annotation in blood

We used the FUMA platform[50] to functionally annotate our GWAS findings. The summary statistics of the asthma GWAS in UK Biobank were uploaded in FUMA. The SNP2GENE function was used to map GWAS SNPs to 1) deleterious coding SNPs (positional mapping), 2) blood eQTL (eQTL mapping), and 3) chromatin contact interactions (chromatin interaction mapping). Positional mapping was performed by selecting exonic variants directly associated with asthma (P_{GWAS}<5E-8) or in LD with asthma-associated variants using a LD r^2 threshold of 0.6 based on the 1000 Genomes EUR reference panel. Protein coding variants (excluding synonymous) with CADD score > 20 were further prioritized. Blood *cis*-eQTL mapping was performed using a publicly available dataset of 31,684 samples[11]. Significant SNP-gene pairs (P_{FDR}<0.05) were identified and then mapped to genetically expressed genes associated with asthma, or eGenes. Chromatin interaction mapping was performed using Hi-C data of a lymphoblastoid B cell line (GM12878). Results of eQTL and chromatin mapping were visualized using circos plots generated by FUMA.

Mendelian randomization in blood with asthma

Two-sample summary-level Mendelian randomization (MR) analyses were performed to infer causal associations between blood eGenes and asthma. The genetic effects on asthma risk were derived from the current GWAS in UK Biobank and the genetic effects on gene expression in blood were derived from a

published eQTL dataset[11]. MR was performed using the inverse-variance weighted (IVW) and Egger methods as implemented in the MendelianRandomization package in R. SNPs were selected within a window of 500 Kb around the transcription start site of each blood eGene. SNPs associated with gene expression ($P < 0.001$ corresponds to ~ F statistics > 10) and independent ($r^2 < 0.1$ based on the 1000 Genomes EUR reference panel) were selected as instrumental variables. We requested at least 3 instrumental variables per gene to perform Mendelian randomization. A P value below the Bonferroni threshold was considered as significant (431 MR with enough instrumental variables: $P_{Bonferroni} < 0.05/431 < 1.16E\text{-}4$). The Cochran's Q test and MRPRESSO (Mendelian randomization pleiotropy residual sum and outlier) global test were used to determine the presence of unmeasured pleiotropy. Heterogeneity ($P_{Qtest}<0.05$) was corrected by applying the MR-PRESSO approach[52].

Drug targets

Target genes of the asthma-associated variants identified in previous sections were then integrated to prioritize druggable genes. In total, we identified 55 lung TWAS genes (**Supplementary Table 4**), 485 blood eGenes (**Supplementary Table 8**), and 563 chromatin contacts mapped genes (**Supplementary Table 10**). Together, 806 unique target genes were identified with overlap across methods shown in Fig. 6. According to the Open Targets Platform[12], 13 of them are the targets of investigational or approved asthma drugs (Table 4). All target genes were also interrogated using the Open Targets Platform[12] for their overall association score with asthma. Results for all target genes are in **Supplementary Table 13**. The 806 target genes were also overlaid with the known druggable genes derived from the drug-gene interaction database (DGIdb)[13] and the druggable genome[14]. Drug-gene interactions were identified for 182 target genes in DGIdb and 201 target genes were part of the druggable genome (**Supplementary Table 13**), which offer numerous opportunities for drug repurposing. We further focused on 29 target genes that were consistently identified by TWAS, eQTL, and chromatin interactions (Fig. 6). Ten of them have known drug targets. Table 5 shows these 10 druggable target genes for asthma and their direction of effect on asthma risk in lung tissue as well as the candidate drugs, interaction types and clinical indications. Target-asthma associations of 1, which is the highest possible score in Open Targets, were observed for two genes including *IL4R* that is the therapeutic target of dupilumab used to treat uncontrolled persistent asthma[15] and *SMAD3* involved in airway remodeling[16] and that may mediate some actions of corticosteroids, which are the cornerstone of asthma treatment. Finally, we filtered the 806 target genes based on three cumulative criteria: 1) those identified by at least two out of three approaches (lung TWAS genes, blood eGenes, and Hi-C genes), 2) those with asthma score of at least 0.5 in Open Targets, and 3) those that are druggable in either the DGIdb or the druggable genome. By excluding the HLA molecules, this strategy revealed 21 prioritized therapeutic targets for asthma (Table 6). In addition to *IL4R* and *SMAD3*, these prioritized genes are known targets of existing asthma drugs including *IL6* (clazakizumab, sirukumab), *TNFSF4* (oxelumab) and *TSLP* (tezepelumab).

Table 4
Investigational or approved asthma drugs acting on identified gene targets

Target genes	Drugs	Action type
CCR4	MOGAMULIZUMAB	Cross-linking agent
CSF2	LENZILUMAB	Inhibitor
IL13	ANRUKINZUMAB	Inhibitor
	LEBRIKIZUMAB	Inhibitor
	DECTREKUMAB	Inhibitor
	TRALOKINUMAB	Inhibitor
IL1R1	ANAKINRA	Antagonist
IL23A	RISANKIZUMAB	Inhibitor
IL2RA	DACLIZUMAB	Inhibitor
IL4	PASCOLIZUMAB	Inhibitor
IL4R	**DUPILUMAB**	Antagonist
IL5	**MEPOLIZUMAB**	Inhibitor
	RESLIZUMAB	Inhibitor
IL6	CLAZAKIZUMAB	Inhibitor
	SIRUKUMAB	Inhibitor
TNF	ADALIMUMAB	Inhibitor
	ETANERCEPT	Inhibitor
	GOLIMUMAB	Inhibitor
TNFSF4	OXELUMAB	Inhibitor
TSLP	TEZEPELUMAB	Inhibitor

In bold are drugs that have demonstrated clinical efficacy in phase 3 clinical trials.

Table 5
Druggable target genes consistently identified across methods

Genes	Asthma score*	Z TWAS	P value	Drug	Interaction	Indication
CAMK4	0.165	5.934	2.94e-09	CHEMBL261720	NA	Experimental
				ESTRADIOL BENZOAT	NA	Oestrogenic hormonal therapy
				GEMCITABINE	NA	Antineoplastic agent (solid cancers)
HSPA4	0.088	5.999	1.98e-09	LITHIUM	NA	Antipsychotic (mania, depression)
				PUROMYCIN	NA	For cell culture (microbiology), no clinical research or application
				ARSENIC TRIOXIDE	NA	Antineoplastic agent (oncohematology)
				BORTEZOMIB	NA	Antineoplastic agent (oncohematology)
				CELECOXIB	NA	Anti-inflammatory and anti-rheumatic drug, non-steroids Investigational as antineoplastic agent
				CHLORPROMAZINE	NA	Neuroleptic antipsychotic, sedative antihistamine
				CYTARABINE	NA	Antineoplastic agent (oncohematology)
				DEFEROXAMINE	NA	Iron chelating agent
				6-DIAZO-5-OXO-L-NORLEUCINE	NA	Experimental as antineopslastic agent
				ENALAPRIL	NA	Angiotensin converting enzyme inhibitor antihypertensive
				EPOIETIN ALFA	NA	Antianemic agent
				FLUTICASONE PROPIONATE	NA	Inhaled corticosteroid used in local treatment of asthma and COPD

*Overall association score for asthma from the Open Targets Platform[12]

Genes	Asthma score*	Z TWAS	P value	Drug	Interaction	Indication
				GOSSYPOL	NA	Experimental as contraceptive and antineoplastic agent
				HALOPERIDOL	NA	Neuroleptic antipsychotic
				HEPARIN	NA	Antithrombotic agent
				HYDRALAZINE	NA	Vasodilatator agent, antihypertensive
				IFOSFAMIDE	NA	Antineoplastic agent
				KETANSERIN	NA	Serotonin antagonist, antihypertensive agent
				NIFEDIPINE	NA	Calcium channel blockers (antihypertensive, vasodilatator)
				NIMESULIDE	NA	Non steroid anti-inflammatory
				PHENYLEPHRINE	NA	Adrenergic agent, vasoconstrictor (hypotension treatment)
				PHOTOPHRIN	NA	Photosensitizer for palliative photodynamic therapy of obstructive cancer
				MIDOSTAURIN	NA	Antineoplastic agent (oncohematology)
				PERILLYL ALCOHOL	NA	Experimental as antineoplastic agent
				RANITIDINE	NA	H2 receptor antagonist drug for ulcer or gastro-oesophageal reflux disease
				SODIUM CHLORIDE	NA	Mineral supplement fluid (hydratation)
				SODIUM SALICYLATE	NA	Analgesic, antipyretic drug
				THIABENDAZOLE	NA	Experimental as antineoplastic agent

*Overall association score for asthma from the Open Targets Platform[12]

Genes	Asthma score*	Z TWAS	P value	Drug	Interaction	Indication
				UREA	NA	Keratolytic agent(onychomycosis)
				VERAPAMIL	NA	Calcium channel blocker (antihypertensive, antiarythmic)
				ASCORBATE	NA	Water-soluble C vitamin (deficiency, infectious disease)
				WORTMANNIN	NA	Experimental as antineoplastic agent
				ISOPROTERENOL	NA	Beta-adrenergic stimulant (cardiac stimulant, bronchodilatator)
				ZALCITABINE	NA	Antiretroviral agent (HIV)
IL4R	1.000	-5.818	5.94e-09	CINTREDEKIN BESUDOTOX	NA	Investigational as anti-neoplastic agent (brain cancer)
				SILYBIN B	agonist	Hepatoprotector herbal drug
				DUPILUMAB	antagonist	Biologic agents for uncontrolled atopic dermatitis, asthma, nasal polyposis
MED1	0.159	5.030	4.89e-07	BECOCALCIDIOL	NA	Investigational drug for psoriasis
PSMB9	0.041	-5.183	2.18e-07	CARFILZOMIB	inhibitor	Antineoplastic agent (oncohematology)
				BORTEZOMIB	inhibitor	Antineoplastic agent (oncohematology)
				IXAZOMIB CITRATE	inhibitor	Antineoplastic agent (oncohematology)
				MARIZOMIB	inhibitor	Investigational as antineoplastic
				OPROZOMIB	inhibitor	Investigational as antineoplastic
PSMD3	0.263	-5.869	4.37e-09	CARFILZOMIB	inhibitor	Antineoplastic agent (oncohematology)

*Overall association score for asthma from the Open Targets Platform[12]

Genes	Asthma score*	Z TWAS	P value	Drug	Interaction	Indication
				BORTEZOMIB	inhibitor	Antineoplastic agent (oncohematology)
				IXAZOMIB CITRATE	inhibitor	Antineoplastic agent (oncohematology)
				OPROZOMIB	inhibitor	Investigational as antineoplastic
RAD50	0.478	5.399	6.68e-08	IRINOTECAN	NA	Antineoplastic agents (solid cancer)
				AZD-7762	NA	Investigational as antineoplastic (checkpoint inhibitor)
				QUINPIROLE	NA	Experimental (psychoactive, neurologic disorders)
SLC22A5	0.189	8.831	1.03e-18	LEVOCARNITINE	NA	Amino acids derivatives (used in metabolic deficiency states)
SMAD3	1.000	6.206	5.41e-10	DEXAMETHASONE	NA	Corticosteroids (anti-inflammatory, immunosuppressive)
				GENISTEIN	NA	Experimental (antineoplastic, menopausal symptoms)
				HALOFUGINONE	NA	Experimental (malaria, cancer, and fibrosis-related and autoimmune diseases)
				LEUPRORELIN ACETATE	NA	Antineoplastic agent (hormone-sensitive tumors) and hormonal treatment
TAP2	0.0415	-5.490	4.01e-08	PRAMLINTIDE	agonist	Antidiabetic agent (amylin analog)
				CLOZAPINE	agonist	Antipsychotic
				ALCOHOL	agonist	Psychotropic substance
				CALCITONIN	agonist	Anti-parathyroid agent, bone antiresorptive agent

*Overall association score for asthma from the Open Targets Platform[12]

Genes	Asthma score*	Z TWAS	P value	Drug	Interaction	Indication
				MONOETHANOLAMINE	antagonist	Antivaricose therapy (local sclerosing agent)
				OLCEGEPANT	antagonist	Antimigraine

*Overall association score for asthma from the Open Targets Platform[12]

Table 6
Genes prioritized as therapeutic targets for asthma

Genes	Lung TWAS gene	Blood eGene	Hi-C gene	Asthma score*	DGIdb 3.0	Druggable genome
IL4R	yes	yes	yes	1	yes	yes
SMAD3	yes	yes	yes	1	yes	yes
TLR1	yes	yes	yes	1	no	yes
CD247	no	yes	yes	1	yes	yes
IL6	no	yes	yes	1	yes	yes
IL7R	no	yes	yes	1	yes	yes
PTPRC	no	yes	yes	1	yes	yes
TNFSF4	no	yes	yes	1	yes	yes
RORC	no	yes	yes	0.67	yes	yes
NOTCH4	no	yes	yes	0.65	yes	yes
ERBB3	no	yes	yes	0.62	yes	yes
GPR183	no	yes	yes	0.54	yes	yes
ERBB2	no	yes	yes	0.51	yes	yes
GPR18	no	yes	yes	0.50	yes	yes
IL1RL1	yes	yes	no	1	no	yes
TSLP	yes	no	yes	1	no	yes
PLXNC1	no	yes	yes	1	no	yes
TLR10	no	yes	yes	0.94	no	yes
TLR6	no	yes	yes	0.77	no	yes
GLB1	no	yes	yes	0.71	no	yes
CCR7	no	yes	yes	0.55	no	yes

Targets of existing asthma drugs are in bold.

*Overall association score for asthma from the Open Targets Platform[12].

DGIdb, Drug-gene interaction database[13].

Druggable genome[14].

Discussion

An important genetic susceptibility to develop asthma has long been demonstrated by genetic epidemiology studies. However, the predisposing genetic variants have been difficult to identify until the realization of the recent large-scale GWAS. Now, a large number of genetic loci are robustly associated with asthma. The new challenge is to identify the candidate causal genes and best therapeutic targets underpinning GWAS-nominated loci. Here, we leveraged lung and blood transcriptome as well as epigenetic marks to map the most likely causal genes within asthma susceptibility loci derived from UK Biobank. Using a broad asthma definition, we identified 72 physically-defined asthma loci containing 116 independent genetic variants with P_{GWAS} < 5.0E-8. The effect size estimates were robust to more strict asthma definitions excluding other lung diseases, smoking history or allergy within controls. As expected, the yield of deleterious coding variants was low, and we thus focused most analyses on regulatory elements. The UK Biobank asthma GWAS was integrated with the largest lung eQTL study available. Fifty-five significant TWAS genes located in 21 asthma loci were found and 26 of them (in 14 loci) replicated in GTEx lung. As previously reported[17], we demonstrated a strong enrichment of asthma-associated variants in regions of regulatory and functional annotations in blood. We mapped 485 blood eGenes and demonstrated that 50 of them are causally associated with asthma by Mendelian randomization. Chromatin contact mapping in the blood cell line showing the most significant enrichment of DNase I hypersensitive sites (GM12878) revealed 563 Hi-C genes. Prioritization of the 806 candidate causal genes identified in this study based on consistency across methods, druggability, and prior asthma association led to 21 genes prioritized as therapeutic targets for asthma.

Five of the top 21 prioritized genes are the targets of existing asthma drugs including *IL4R* (dupilumab), *SMAD3* (corticosteroids), *IL6* (clazakizumab, sirukumab), *TNFSF4* (oxelumab) and *TSLP* (tezepelumab). This supports the possibility that other genes in that list (Table 6) are credible therapeutic targets for asthma. Among them, there are three members of the toll-like receptor family: *TLR1*, *TLR6*, and *TLR10*. These three TLRs are located at the same 4p14 locus, are phylogenetically related, and require the formation of heterodimers with TLR2 for recognition of invading microbes[18]. We found a strong colocalization signal between the blood eQTL for *TLR10* and the GWAS for asthma (PP4 = 0.84). In MR, the blood expression of *TLR10* was positively associated with the risk of asthma (P_{IVW}=4.34E-6). We also identified two missense mutations in *TLR10*, C-rs4129009 (p.Ile775Val) (P_{GWAS}=1.49E-10) and G-rs11096957 (p.Asn241His) (P_{GWAS}=2.49E-10), which are in moderate LD (r^2 = 0.42 in CEU) and associated with the risk of asthma. The CADD score for G-rs11096957 is 21.9 and is predicted to be “possibly damaging” and “deleterious” by PolyPhen and SIFT, respectively. The alleles Crs4129009 (p.Ile775Val) and G-rs11096957 (p.Asn241His), which decrease the risk of asthma, have been associated with elevated blood cytokine responses to a TLR1/2 agonist, most specifically Pam_3CSK_4-induced interleukin 6 (IL6)[19]. We recently demonstrated that genetically predicted levels of circulating IL6R, a negative regulator of IL6 signaling, are positively associated with the risk of asthma and atopic disorders[20]. These data suggest a complex interaction between TLR10 and IL6 on the risk of asthma and warrant further investigation. One line of enquiry could examine the possibility that TLR10 dampens TLR2 signaling and IL6 production, thereby increasing the risk of asthma. In support of the latter hypothesis, in an ovalbumin-induced asthma mouse model, IL6 lowered Th2 cytokines and decreased bronchial hyperresponsiveness[21].

Other gene targets of interest include cytokine/chemokine receptors (*IL7R*, *IL1RL1*, *CCR7*), two members of the EGFR family of receptor tyrosine kinases (*ERBB2* and *ERBB3*), two G protein-coupled receptors (*GPR18* and

GPR183), an antigen recognition molecule (CD247), a member of the protein tyrosine phosphatase family (*PTPRC*), a transcription factor (*RORC*), a member of the NOTCH family (*NOTCH4*), a member of the plexin family (*PLXNC1*), and galactosidase beta 1 (*GLB1*). All these targets have drug-gene interactions in DGIdb[13] and/or are present in the list of genes encoding druggable human proteins[14]. They are thus representing drug repurposing/development opportunities. Further experimental research will be needed to screen these putative novel therapeutic targets for asthma.

We have been searching for asthma genes for a long time[22]. The last decade of GWAS research has been particularly exciting, witnessing the identification of genetic variants robustly associated with this disease[23,24]. Ten years ago, finding these variants was almost seen as an insurmountable task. Three main factors have allowed us to overcome this challenge: 1) our evolving understanding of the human genome, 2) progress in genotyping technologies, and 3) international collaboration to allow sufficiently large sample size to be analyzed. Now, with advanced bioinformatics skills and resources such as the UK Biobank, we can rediscover the genetic variants associated with asthma. This is remarkable. Obviously, any milestone comes with new challenges. Translating new genetic findings into better management and/or treatment for patients is still falling short. The effect sizes of all genetic variants associated with asthma are relatively small. Although 116 independent variants reached genome-wide significance (P_{GWAS}<5.0E-8) in this study, the ORs range from 1.03 to 3.59 (median = 1.06, note that ORs lower than 1 were converted into their reciprocal (1/OR)). As an example to appreciate the effect size that we are detecting, the top associated asthma variant on chromosome 9 near the *IL33* gene has a p value of 1.25E-56 and an OR of 1.13, which is the result of allele frequencies of 0.73 in cases and 0.75 in controls. One hundred out of the 116 independent variants were common, with minor allele frequency greater than 5% in cases and controls combined. So most risk alleles are common with small effects that we are able to detect owing to the large sample size. Cumulatively, all genetic variants discovered by GWAS explained about 8–9% of the total heritability[17], suggesting much more work is needed to elucidate the full genetic architecture of asthma. More work is also needed to move discovered genetic factors underlying asthma down the clinical translation pipeline. We believe that the current study represents an important step beyond GWAS data. By combining different data sources (eQTL and Hi-C in disease-relevant tissues) and advanced bioinformatics approaches (TWAS, MR, colocalization), we were able to reveal relevant genes and putative therapeutic targets for asthma.

As observed in previous asthma GWAS studies, we had limited success in mapping asthma-associated variants to deleterious coding SNPs. One of our most interesting hit is the stop-gain mutation in the filaggrin (*FLG*) gene. *FLG* is located on chromosome 1q21.3, a locus where we have identified 5 independent significant variants (**Supplementary Table 2**). One of the 5 independent variant is rs61816761 causing a G to A substitution (c.16819G > A) that occurs in exon 3 of the gene, resulting in a stop instead of an arginine in codon 501 (p.Arg501Ter). The allele frequencies in cases and controls are 2.76% and 2.28%, respectively, with P_{GWAS}=3.95E-22. The allele frequency in TOPMED (based on 62,784 individuals) is 1.16%. *FLG* was previously associated with atopic dermatitis[25] where the risk variants are believed to disrupt the skin barrier, allowing allergen sensitization and then promoting the development of asthma[26]. rs61816761 was also found in previous GWAS of asthma in UK Biobank[6,7,17,27,28] and with greater effect on atopic dermatitis than asthma[29], which supports skin barrier dysfunction as a cause of asthma. Another deleterious coding variant (rs2230624, Cys273Tyr) was identified in *TNFRSF8* (also known as *CD30*), which was previously reported[30] and

characterized as a loss of function variant that decreased asthma risk by reducing the trafficking of the CD30 protein on cell surface[29].

As GWAS on asthma in UK Biobank accumulate[6,7,17,28–30], the next milestone will be to identify the function units, most intuitively genes, underlying the GWAS loci. In this study, we have combined the UK Biobank GWAS data with the largest lung eQTL available to perform a TWAS. Plausible causal genes in lung tissues were revealed for 21 asthma loci. On 17q12-q21.2, the first discovered[31] and the most replicated[32] GWAS asthma locus, *GSDMB* was the top TWAS gene in our lung eQTL dataset. Although other TWAS genes were observed in that locus, the p value distributions of GWAS and eQTL colocalized better with *GSDMB*. Using blood as eQTL source, we also identified *GSDMB* as the most likely causal gene on 17q12-q21.2. These results are consistent with eQTL analysis showing that SNPs associated with asthma susceptibility and severity at 17q12-q21.2 are correlated with *GSDMB* expression in cells from human bronchial epithelial biopsy and bronchial alveolar lavage[33,34]. *GSDMB* is thus a gene to focus on in future functional studies. Other lung TWAS genes prioritized by our study for further functional characterization are *RERE* on 1p36, *TLR1* on 4p14, *SLC22A5* or *RAD50* on 5q31, *RBM17* on 10p15, *UBAC2* on 13q32, *SMAD3* on 15q22-q23, *CLEC16A* on 16p13, *IL4R* on 16p12, *KPNB1* on 17q21.32, and *PHF5A* on 22q13.

The lung TWAS revealed a novel asthma risk locus at 1q23.3 with the putative gene encoding the gamma chain of the high-affinity IgE receptor (*FCER1G*). Note that the *FCER1A* gene at 1q23.2 (approximately 2 Mb away from *FCER1G*) has been associated with total serum IgE levels[35]. Concerning *FCER1G*, hypomethylation at its promoter has been reported in monocytes of patients with atopic dermatitis, resulting in the overexpression of high affinity IgE receptors in these cells[36]. Here, we found that higher expression of *FCER1G* in lung tissue is associated with asthma (z = 4.74, P_{TWAS}=2.13E-6), a finding replicated in GTEx lung (z = 5.08, P_{TWAS}=3.71E-7). This new asthma locus may thus mediate its effect by upregulating *FCER1G*, which may then lead to inflammatory cells with greater surface expression of IgE receptors, that are more prone to allergic reaction.

This study has limitations. We used the best possible bioinformatics approaches to identify causality genes. However, there is no functional validation using experimental models or human studies to confirm the biological role of these genes in the pathogenesis of asthma. Further studies are needed to demonstrate causality. Our analyses are largely based on European-descent individuals. Inference to other ethnic groups is thus a concern and the lack of similarly powerful resources (e.g. UK Biobank) for other ancestries represent a missed opportunity to identify other relevant asthma genes. Environmental risk factors and the specific period of exposures during the lifespan play an important role in the development of asthma. Our genomic datasets (GWAS and eQTL) are retrospective in nature and we have not taken into account likely modifiers of genetic risk. Finally, we used whole lung and blood tissues, which contain heterogeneous cell populations, limiting our ability to identify genes affecting asthma risk through gene regulation. Progress in single-cell transcriptomic is promising for future studies.

In conclusion, this study expands our understanding of the regulatory and functional mechanisms underlying GWAS asthma risk loci in lung and blood tissues. The candidate causal genes identified are key to understand disease etiology, interpret GWAS results and prioritize follow-up functional studies. Our top therapeutic targets represent new opportunities for drug repositioning and testing in pre-clinical models.

Methods

Genome-wide association study on asthma in UK Biobank

UK Biobank is an open access resource of nearly 500,000 participants enrolled at the age of 40–69 and prospectively evaluated for a range of health-related outcomes[37]. The definition of asthma in this study is based on the UK Biobank Outcome Adjudication Group and relies on hospital, death, primary care and self-reported related codes (Phase 2 code list for asthma). Asthma cases include patients with a diagnosis from hospital record (ICD-9 or ICD-10 codes) or primary care medical record as well as those with self-reported asthma (data-field 20002 in UK Biobank). Genotyping data are derived from the Affymetrix UK BiLEVE or UK Biobank Axiom Arrays. Phasing and imputation were performed centrally using the Haplotype Reference Consortium (HRC) and merged UK10K and 1000 Genomes phase 3 reference panels[38,39]. Samples with call rate < 95%, outlier heterozygosity rate, sex mismatch, non-white British ancestry, samples with excess third-degree relatives (> 10), or not used for relatedness calculation were excluded. Variants with an imputation quality score (INFO) ≤ 0.3 or minor allele frequency (MAF) < 0.0001 were removed. Using the aforementioned definition of asthma and quality control filters, 56,167 asthma cases were compared to 352,255 controls. The genetic association analysis was performed using SAIGE (Scalable and Accurate Implementation of GEneralized mixed model, version 0.36.3.1, https://github.com/weizhouUMICH/SAIGE)[40]. SAIGE is a two-step method to perform generalized mixed model GWAS analysis that is robust to unbalanced case-control ratios, sample relatedness and low-frequency variants. In step 1, we fit a null logistic mixed model with 93,511 independent, high-quality genotyped variants, which were used by the UK Biobank data group to estimate the kinship coefficients between samples[39,40]. The following covariates were added: age, sex, and the first 20 ancestry-based principal components. In step 2, we performed association tests between each genetic variant (genotyped and imputed) and asthma. We applied the leave-one-chromosome-out (LOCO) scheme (LOCO = TRUE). The quantile-quantile plot was generated (**Supplementary Fig. 2**). The genomic inflation factor was computed by converting P values into chi-squared values and then dividing the median of the resulting chi-squared statistics by the expected median of the chi-squared distribution. The present analyses were conducted under UK Biobank data application number 25205.

Heritability

LD-score regression was used to estimate SNP-heritability for asthma[41]. To obtain heritability on the liability scale, we provided sample and population prevalence of 13.8% (–samp-prev 0.138) and 15.6%[8] (–pop-prev 0.156), respectively.

GWAS sensitivity analysis

GWAS-nominated loci were re-evaluated by changing exclusion criteria to define asthma cases and controls. The rationale was to evaluate the potential confounding effect of other lung diseases, smoking, and allergy. Genetic association analyses were thus performed in three case-control subsets. First, asthma cases and controls with other lung diseases were excluded. Individuals were excluded if they had self-reported or medical

records consistent with the presence of chronic obstructive pulmonary disease (COPD), emphysema, chronic bronchitis, interstitial lung disease or alpha-1 antitrypsin deficiency. This results in the exclusion of 20,998 individuals and genetic analysis performed in 47,391 asthma cases and 340,033 controls. Second, we excluded all asthma cases and controls with a positive smoking history (i.e. former and current smokers). This results in the exclusion of 250,739 individuals and genetic analysis performed in 21,097 asthma cases and 136,586 controls. Third, we excluded control individuals with atopy, including hay fever, allergic rhinitis, and eczema/atopic dermatitis. This results in the exclusion of 84,113 individuals and genetic analysis with 268,142 controls (and the same number of asthma cases, n=56,167). Note that the three lists of exclusion criteria were applied separately (not cumulatively) and specific UK Biobank data fields and codes used for excluding individuals in each case-control subset are provided in **Supplementary Table 14**.

Number of loci associated with asthma

After the GWAS analysis, we assessed the number of loci that were associated with asthma based on two methods. First, we counted the number of loci based on physical distance only. All SNPs associated with asthma (P < 5.0E-8) were ranked by chromosome order and by position on build 37. Two subsequent SNPs on this list located on the same chromosome and separated by more than 1 Mb were considered distinct loci. The physical boundaries of asthma-associated loci were then defined by adding 500 Kb downstream and upstream of the most 5' and 3' asthma-associated variants (P_{GWAS}<5.0E-8), respectively, within each locus. One exception was the extended MHC region on chromosome 6 that was counted as a single locus and delimited at 25,726,000 to 33,378,000 bp (GRCh37) based on the positions of two genes (*HIST1H2AA* and *KIFC1*). Second, we identified the number of independent variants, as some physically defined loci will contain significant SNPs that are not in LD. This was performed using a stepwise conditional analysis (–cojo-slct)[42]. The procedure consists of a first round of analysis that is conditioned on the top asthma-associated variant at each locus derived from the original GWAS. If significantly associated variants remain, a second round of analysis is conditioned on the top asthma-associated variant from the first round. Subsequent rounds are carried out until no more variants reach P_{GWAS}<5.0E-8.

The lung expression quantitative trait loci

The lung eQTL dataset consists of whole-genome genotyping (Illumina Human1M-Duo BeadChip) and gene expression (Affymetrix) in non-tumor lung tissues from patients who underwent lung surgery at three academic sites, Laval University, University of British Columbia, and University of Groningen, henceforth referred to as Laval, UBC, and Groningen, respectively. All lung specimens from Laval were obtained from patients undergoing lung cancer surgery and were harvested from a site distant from the tumor. At UBC, the majority of samples were from patients undergoing resection of small peripheral lung lesions. Additional samples were from autopsy and at the time of lung transplantation. At Groningen, the lung specimens were obtained at surgery from patients with various lung diseases, including patients undergoing therapeutic resection for lung tumors, harvested from a site distant from the tumor, and lung transplantation. Lung tissue processing and storage, DNA and RNA extraction, genotyping, microarray-based gene expression and lung *cis*-eQTL analyses have been described previously[43,44]. Following standard microarray and genotyping quality controls, data on

1,038 patients were available. At Laval and UBC, written informed consent was obtained from all subjects and the study was approved by their respective ethics committee. At Groningen, lung specimens were provided by the local tissue bank of the Department of Pathology and the study protocol was consistent with the Research Code of the University Medical Center Groningen and Dutch national ethical and professional guidelines ("Code of conduct; Dutch federation of biomedical scientific societies"; http://www.federa.org).

Transcriptome-wide association study

The TWAS was performed using S-PrediXcan[45]. The lung eQTL dataset was used as the training set to derive the expression weights. Gene expression normalized for age, sex and smoking status from Laval, UBC, and Groningen were combined with ComBat[46]. Gene expression traits were then trained with elastic net linear models (alpha = 0.5, n_k_folds = 10, window = 500 Kb). Models with false-discovery rate (FDR) < 0.05 as implemented in S-PrediXcan were obtained for 19,918 probe sets. Predicted expression levels from the lung in the UK Biobank participants were then tested for association with asthma[45]. The Bonferroni correction was used to claim transcriptome-wide significance (S-PrediXcan P_{TWAS}=0.05/19,918 = 2.51E-6).

TWAS replication in GTEx lung

Lung eQTL data from 515 individuals available in the Genotype-Tissue Expression (GTEx) project (version 8)[47] were used for TWAS replication. The TWAS was performed using SPrediXcan as described above. In GTEx lung, models were obtained for 11,518 genes. Bonferroni-corrected TWAS gene was thus set at P_{TWAS}<4.34E-6. We also sought replication of TWAS genes identified in our lung eQTL dataset. Significant replication was considered for genes with the same direction of effect and with P_{TWAS}<0.05 in GTEx lung.

Bayesian colocalization

For specific asthma-associated loci and genes, we evaluated whether the asthma GWAS and lung eQTL signals shared the same causal variants using the COLOC package in R[48]. For the loci of interest, summary statistics from the asthma GWAS in UK Biobank were combined with our lung eQTL results using a window of 1 Mb up- and downstream of the TWAS genes. We considered colocalization events when the posterior probability of shared eQTL and GWAS associations (PP4) was greater than 60%. The colocalization analysis for the 485 blood eGenes were performed using the HyPrColoc package. HyPrColoc analysis was performed in a window of 500 kb up- and downstream of the transcription start site of each eGene. LocusCompare[49] was used to visualize GWAS and eQTL colocalization events.

Cell type and tissue enrichment of asthma-associated loci

We used GARFIELD[10] to overlap our GWAS findings with regulatory and functional annotations derived from ENCODE, GENCODE and Roadmap Epigenomics projects. A total of 1,005 annotation features were considered including chromatin states, histone modifications, genic annotations, transcription factor binding sites and

open chromatin data (FAIRE, DHS hotspots, peaks and footprints), which were evaluated in different cell types and tissues. LD pruning of GWAS SNPs was performed at $r^2 > 0.8$ and fold enrichment was evaluated at two GWAS significance thresholds: 1.0E-5 and 1.0E-8.

Functional mapping and annotation in blood

We used the FUMA platform[50] to functionally annotate our GWAS findings. The summary statistics of the asthma GWAS in UK Biobank were uploaded in FUMA. The SNP2GENE function was used to map GWAS SNPs to 1) deleterious coding SNPs (positional mapping), 2) blood eQTL (eQTL mapping), and 3) chromatin contact interactions (chromatin interaction mapping). Positional mapping was performed by selecting exonic variants directly associated with asthma (P_{GWAS}<5E-8) or in LD with asthma-associated variants using a LD r^2 threshold of 0.6 based on the 1000 Genomes EUR reference panel. Protein coding variants (excluding synonymous) with CADD score > 20 were further prioritized. Blood *cis*-eQTL mapping was performed using a publicly available dataset of 31,684 samples[11]. Significant SNP-gene pairs (P_{FDR}<0.05) were identified and then mapped to genetically expressed genes associated with asthma, or eGenes. Chromatin interaction mapping was performed using Hi-C data of a lymphoblastoid B cell line (GM12878). Results of eQTL and chromatin mapping were visualized using circos plots generated by FUMA.

Pathway analysis

Pathway analysis was performed using the Enrichr web server[51]. Blood eGenes discovered in this study were uploaded in Enrichr and enrichment was assessed using the combined score method for gene sets available in the Kyoto Encyclopedia of Genes and Genomes (KEGG) database.

Mendelian randomization in blood with asthma

Two-sample summary-level Mendelian randomization (MR) analyses were performed to infer causal associations between blood eGenes and asthma. The genetic effects on asthma risk were derived from the current GWAS in UK Biobank and the genetic effects on gene expression in blood were derived from a published eQTL dataset[11]. MR was performed using the inverse-variance weighted (IVW) and Egger methods as implemented in the MendelianRandomization package in R. SNPs were selected within a window of 500 Kb around the transcription start site of each blood eGene. SNPs associated with gene expression (P<0.001 corresponds to ~F statistics>10) and independent (r^2<0.1 based on the 1000 Genomes EUR reference panel) were selected as instrumental variables. We requested at least 3 instrumental variables per gene to perform Mendelian randomization. A P value below the Bonferroni threshold was considered as significant (431 MR with enough instrumental variables: $P_{Bonferroni} < 0.05/431 < 1.16E\text{-}4$). The Cochran's Q test and MR-PRESSO (Mendelian randomization pleiotropy residual sum and outlier) global test were used to determine the presence of unmeasured pleiotropy. Heterogeneity ($P_{Q\text{-}test}$<0.05) was corrected by applying the MR-PRESSO approach[52].

Druggable target genes

A list of druggable genes were obtained from the drug-gene interaction database[13] (DGIdb, www.dgidb.org) and the druggable genome[14]. Target genes of asthma-associated variants identified by TWAS, eQTL and chromatin interactions were integrated and overlaid with the list of druggable genes. Druggable target genes were then queried for candidate drugs, interaction types and clinical indications in DGIdb[13], DrugBank (www.drugbank.ca), ChEMBL (www.ebi.ac.uk/chembl), and PubChem (pubchem.ncbi.nlm.nih.gov). Target genes were also queried on the Open Targets Platform[12] for their association with asthma.

Genome build

GRCh37 (hg19) coordinates were used in this study.

Data availability

The human lung tissue eQTL study is available in dbGaP under accession phs001745.v1.p1.

Declarations

Competing interests

The authors declare no competing interests.

Author contributions

K.V., P.M., S.T., and Y.B. contributed to the conception and study design. N.G., P.J., M.O., M.B., W.T., D.S., D.N., K.H., P.M., S.T., and Y.B. contributed to data collection. K.V., Z.L., V.B.-B., A.C., J.-C.B., A.E., J.L., P.M., S.T., and Y.B. contributed to data analysis. K.V., Z.L., V.B.-B., N.G., C.L., K.G., A.C., M.L., L.-P.B., P.M., S.T., and Y.B. contributed to data interpretation. K.V., P.M., S.T. and Y.B. drafted the manuscript. All authors revised the manuscript.

Acknowledgments

The authors would like to thank the team at the IUCPQ-UL Biobank for their valuable assistance. This research has been conducted using the UK Biobank Resource. This study was supported by grants from the Chaire de pneumologie de la Fondation JD Bégin de l'Université Laval, the Fondation de l'Institut universitaire de cardiologie et de pneumologie de Québec, and the Canadian Institutes of Health Research (MOP – 123369). Patrick Mathieu holds a Fonds de Recherche du Québec-Santé (FRQS) Research Chair. Sébastien Thériault holds a Junior 1 Clinical Research Scholar award from the FRQS. Yohan Bossé holds a Canada Research Chair in Genomics of Heart and Lung Diseases.

References

1. GBD 2016 Causes of Death Collaborators. Global, regional, and national age-sex specific mortality for 264 causes of death, 1980–2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet* **390**, 1151–1210 (2017).
2. Moffatt, M.F. *et al.* A large-scale, consortium-based genomewide association study of asthma. *N Engl J Med* **363**, 1211–21 (2010).
3. Torgerson, D.G. *et al.* Meta-analysis of genome-wide association studies of asthma in ethnically diverse North American populations. *Nat Genet* **43**, 887–92 (2011).
4. 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).
5. Daya, M. *et al.* Association study in African-admixed populations across the Americas recapitulates asthma risk loci in non-African populations. *Nat Commun* **10**, 880 (2019).
6. 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).
7. 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).
8. Mukherjee, M. *et al.* The epidemiology, healthcare and societal burden and costs of asthma in the UK and its member nations: analyses of standalone and linked national databases. *BMC Med* **14**, 113 (2016).
9. Pantic, B. *et al.* Myotonic dystrophy protein kinase (DMPK) prevents ROS-induced cell death by assembling a hexokinase II-Src complex on the mitochondrial surface. *Cell Death Dis* **4**, e858 (2013).
10. Iotchkova, V. *et al.* Discovery and refinement of genetic loci associated with cardiometabolic risk using dense imputation maps. *Nat Genet* **48**, 1303–1312 (2016).
11. Võsa, U. *et al.* Unraveling the polygenic architecture of complex traits using blood eQTL metaanalysis. *bioRxiv*, 447367 (2018).
12. Carvalho-Silva, D. *et al.* Open Targets Platform: new developments and updates two years on. *Nucleic Acids Res* **47**, D1056-D1065 (2019).
13. Cotto, K.C. *et al.* DGIdb 3.0: a redesign and expansion of the drug-gene interaction database. *Nucleic Acids Res* **46**, D1068-D1073 (2018).
14. Finan, C. *et al.* The druggable genome and support for target identification and validation in drug development. *Sci Transl Med* **9**(2017).
15. Wenzel, S. *et al.* Dupilumab efficacy and safety in adults with uncontrolled persistent asthma despite use of medium-to-high-dose inhaled corticosteroids plus a long-acting beta2 agonist: a randomised double-blind placebo-controlled pivotal phase 2b dose-ranging trial. *Lancet* **388**, 31–44 (2016).
16. Le, A.V. *et al.* Inhibition of allergen-induced airway remodeling in Smad 3-deficient mice. *J Immunol* **178**, 7310–6 (2007).
17. Han, Y. *et al.* Genome-wide analysis highlights contribution of immune system pathways to the genetic architecture of asthma. *Nat Commun* **11**, 1776 (2020).

18. Guan, Y. *et al.* Human TLRs 10 and 1 share common mechanisms of innate immune sensing but not signaling. *J Immunol* **184**, 5094–103 (2010).
19. Mikacenic, C., Reiner, A.P., Holden, T.D., Nickerson, D.A. & Wurfel, M.M. Variation in the TLR10/TLR1/TLR6 locus is the major genetic determinant of interindividual difference in TLR1/2-mediated responses. *Genes Immun* **14**, 52–7 (2013).
20. Rosa, M. *et al.* A Mendelian randomization study of IL6 signaling in cardiovascular diseases, immune-related disorders and longevity. *NPJ Genom Med* **4**, 23 (2019).
21. Wang, J., Homer, R.J., Chen, Q. & Elias, J.A. Endogenous and exogenous IL-6 inhibit aeroallergen-induced Th2 inflammation. *J Immunol* **165**, 4051–61 (2000).
22. Bossé, Y. & Hudson, T.J. Toward a comprehensive set of asthma susceptibility genes. *Annu Rev Med* **58**, 171–84 (2007).
23. Kim, K.W. & Ober, C. Lessons Learned From GWAS of Asthma. *Allergy Asthma Immunol Res* **11**, 170–187 (2019).
24. 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 (2017).
25. Palmer, C.N. *et al.* Common loss-of-function variants of the epidermal barrier protein filaggrin are a major predisposing factor for atopic dermatitis. *Nat Genet* **38**, 441–6 (2006).
26. Hudson, T.J. Skin barrier function and allergic risk. *Nat Genet* **38**, 399–400 (2006).
27. Ferreira, M.A. *et al.* Shared genetic origin of asthma, hay fever and eczema elucidates allergic disease biology. *Nat Genet* **49**, 1752–1757 (2017).
28. 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).
29. Olafsdottir, T.A. *et al.* Eighty-eight variants highlight the role of T cell regulation and airway remodeling in asthma pathogenesis. *Nat Commun* **11**, 393 (2020).
30. Johansson, A., Rask-Andersen, M., Karlsson, T. & Ek, W.E. Genome-wide association analysis of 350 000 Caucasians from the UK Biobank identifies novel loci for asthma, hay fever and eczema. *Hum Mol Genet* **28**, 4022–4041 (2019).
31. Moffatt, M.F. *et al.* Genetic variants regulating ORMDL3 expression contribute to the risk of childhood asthma. *Nature* **448**, 470–3 (2007).
32. 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).
33. Li, X. *et al.* eQTL of bronchial epithelial cells and bronchial alveolar lavage deciphers GWAS-identified asthma genes. *Allergy* **70**, 1309–18 (2015).
34. Li, X. *et al.* Genetic analyses identify GSDMB associated with asthma severity, exacerbations, and antiviral pathways. *J Allergy Clin Immunol* (2020).
35. Weidinger, S. *et al.* Genome-wide scan on total serum IgE levels identifies FCER1A as novel susceptibility locus. *PLoS Genet* **4**, e1000166 (2008).
36. Liang, Y. *et al.* Demethylation of the FCER1G promoter leads to FcepsilonRI overexpression on monocytes of patients with atopic dermatitis. *Allergy* **67**, 424 – 30 (2012).

37. Sudlow, C. *et al.* UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med* **12**, e1001779 (2015).
38. McCarthy, S. *et al.* A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet* **48**, 1279–83 (2016).
39. Bycroft, C. *et al.* The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).
40. Zhou, W. *et al.* Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies. *Nat Genet* **50**, 1335–1341 (2018).
41. Bulik-Sullivan, B.K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet* **47**, 291–5 (2015).
42. Yang, J. *et al.* Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat Genet* **44**, 369–75, S1-<background-color:#D279AA;>3</background-color:#D279AA;> (2012).
43. Hao, K. *et al.* Lung eQTLs to Help Reveal the Molecular Underpinnings of Asthma. *PLoS Genet* **8**, e1003029 (2012).
44. Lamontagne, M. *et al.* Leveraging lung tissue transcriptome to uncover candidate causal genes in COPD genetic associations. *Hum Mol Genet* **27**, 1819–1829 (2018).
45. Barbeira, A.N. *et al.* Exploring the phenotypic consequences of tissue specific gene expression variation inferred from GWAS summary statistics. *Nat Commun* **9**, 1825 (2018).
46. Johnson, W.E., Li, C. & Rabinovic, A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* **8**, 118 – 27 (2007).
47. Consortium, G.T. *et al.* Genetic effects on gene expression across human tissues. *Nature* **550**, 204–213 (2017).
48. Giambartolomei, C. *et al.* Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet* **10**, e1004383 (2014).
49. Liu, B., Gloudemans, M.J., Rao, A.S., Ingelsson, E. & Montgomery, S.B. Abundant associations with gene expression complicate GWAS follow-up. *Nat Genet* **51**, 768–769 (2019).
50. Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nat Commun* **8**, 1826 (2017).
51. Kuleshov, M.V. *et al.* Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res* **44**, W90-7 (2016).
52. Verbanck, M., Chen, C.Y., Neale, B. & Do, R. Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. *Nat Genet* **50**, 693–698 (2018).

Supplementary Information

Supplementary Fig. 1. UpSet plot showing the data sources to define asthma status in UK Biobank. A total of 56,167 asthma cases were identified. Affected individuals had one or more diagnosis of asthma based on self-reported questionnaires, hospital records (ICD-9 and ICD-10), and primary care records.

Supplementary Fig. 2. Quantile-quantile plot of test statistics generated by the GWAS in UK Biobank including 56,167 asthma cases and 352,255 controls. Genomic inflation factor λ = 1.029.

Supplementary Fig. 3. Results of sensitivity analysis evaluating the potential confounder effects of other lung diseases, smoking and allergy. The three scatter plots compared the effect size estimates and SE of the main study design (x axis) with the three alternative study designs (y axis). The identity line is shown in blue. The single and most extreme outlier in the upper right corner of scatter plots consists of the sentinel variant on 1q21.3.

Supplementary Fig. 4. The top lung TWAS genes identified per asthma-associated loci. Previously known and new asthma loci are illustrated in blue and green, respectively.

Supplementary Fig. 5. LocusCompare plots for four significant TWAS genes on chromosome 17q12-q21. Association signals for SNPs within 50 Kb up and downstream of target genes are illustrated for *GSDMB*, *ORMDL3*, *GSDMA*, and *PNMT*.

Supplementary Fig. 6. The top causally associated blood eGene identified per asthma-associated loci. Previously known and new asthma loci are illustrated in blue and green, respectively.

Figures

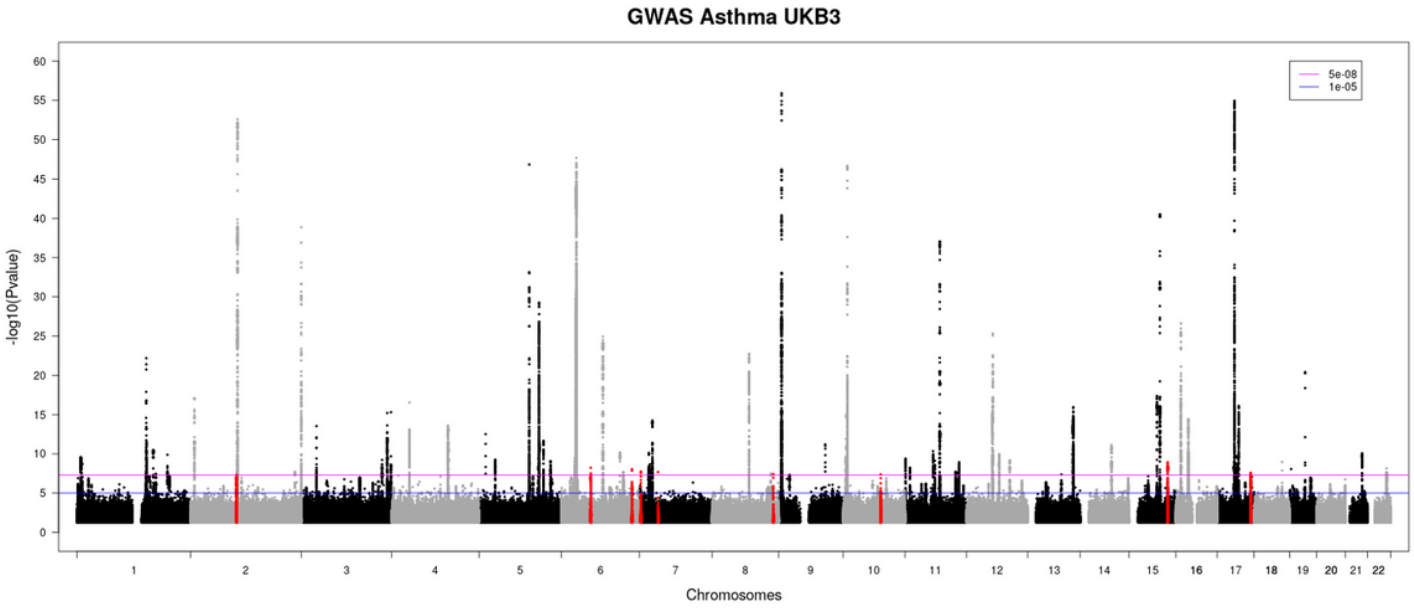

Figure 1

Manhattan plot of the GWAS on asthma in UK Biobank. The GWAS was performed in 56,167 asthma cases and 352,255 controls. The y axis represents P value in -log10 scale. The horizontal blue and magenta lines indicate P value of 1x10-5 and 5x10-8, respectively. Novel asthma loci are in red. Genetic variants with P value > 0.05 were removed to limit the digital information of the Figure.

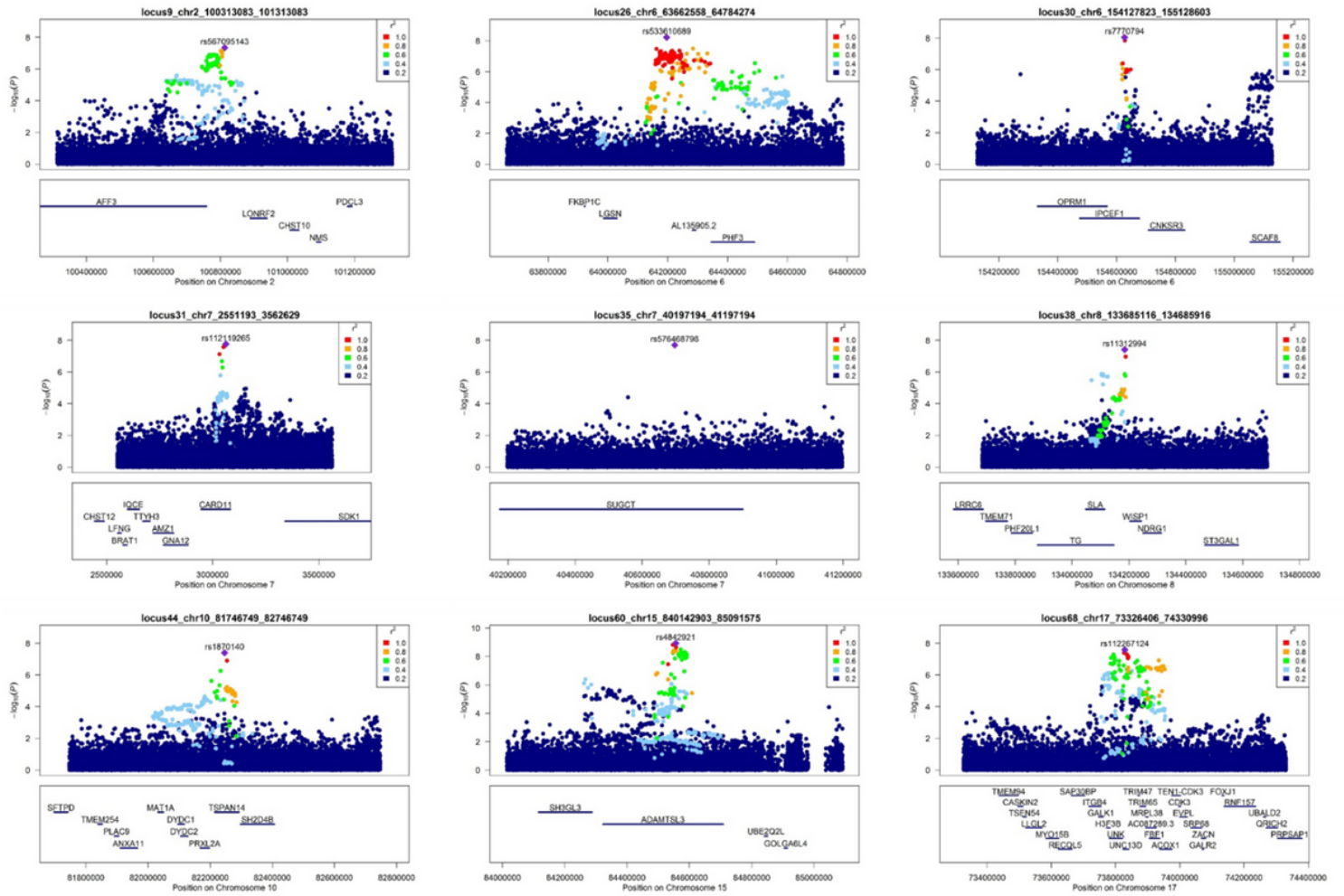

Figure 2

Regional plots showing the nine new asthma-associated loci. The y axis shows the P value in -log10 scale for SNPs up- and downstream of the sentinel SNP (purple dot). The extent of linkage disequilibrium (LD; r2 values) for all SNPs with the sentinel SNP is indicated by colors. The location of genes is shown at the bottom. SNPs are plotted based on their chromosomal position on build 37.

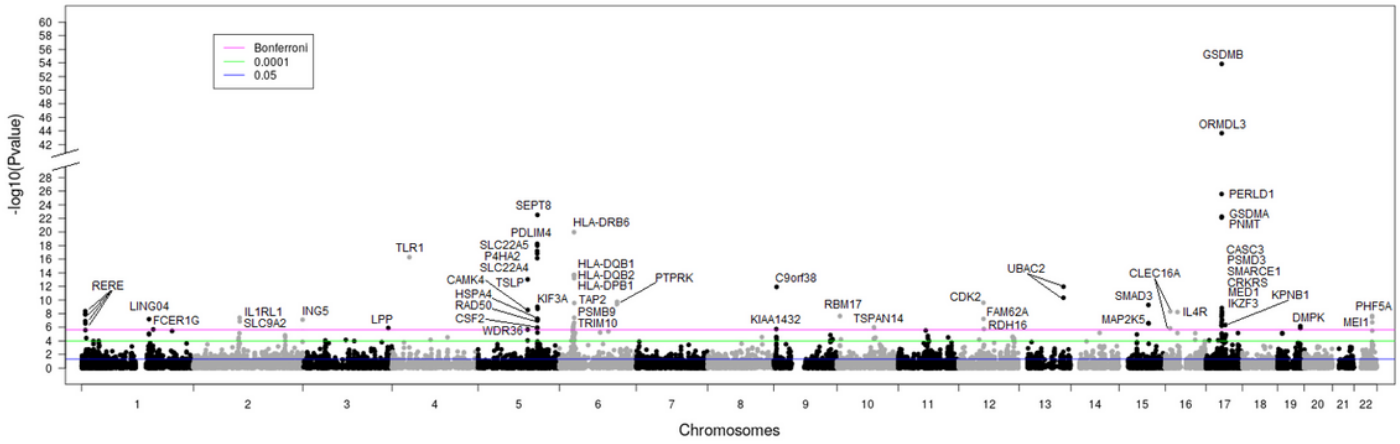

Figure 3

Manhattan plot of the TWAS on asthma integrating the UK Biobank GWAS and the lung eQTL dataset. Each dot represents the association between predicted gene expression and asthma for a specific probe/transcript. P values for gene expression-asthma associations are on the y axis in -log10 scale. The blue, green and magenta horizontal lines represent PTWAS of 0.05, 0.0001 and 2.51E-6 (Bonferroni), respectively. Annotations for genome-wide significant probes/transcripts that passed Bonferroni correction are indicated.

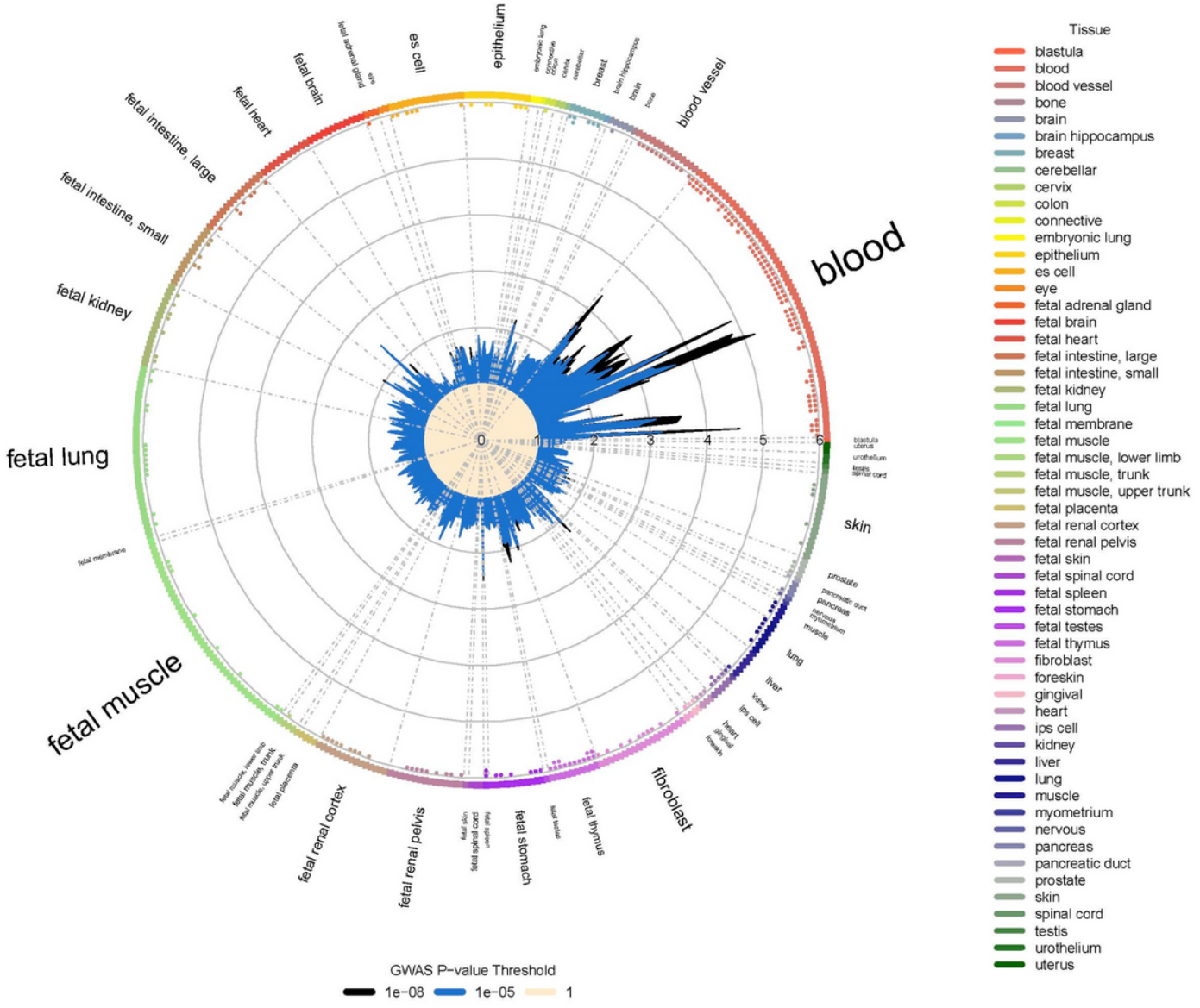

Figure 4

GARFIELD functional enrichment analyses. The wheel plot shows functional enrichment for asthma variants within DHS hotspot regions in ENCODE and Roadmap Epigenomics data. The radial axis represents the enrichment (OR) for each of 424 cell types that are sorted by tissue along the outside edge of the plot. Tissues are labeled with font size proportional to the number of cell types. Boxes forming the edge are colored by tissue. Enrichment is calculated for two GWAS significance thresholds: 1.0E-5 and 1.0E-8, which are plotted in blue and black, respectively, inside the plot. Dots along the inside edge of the plot are colored with respect to tissue and represent significant enrichment for a specific cell type (one dot = P < 1E-5 and two dots = P < 1E-6).

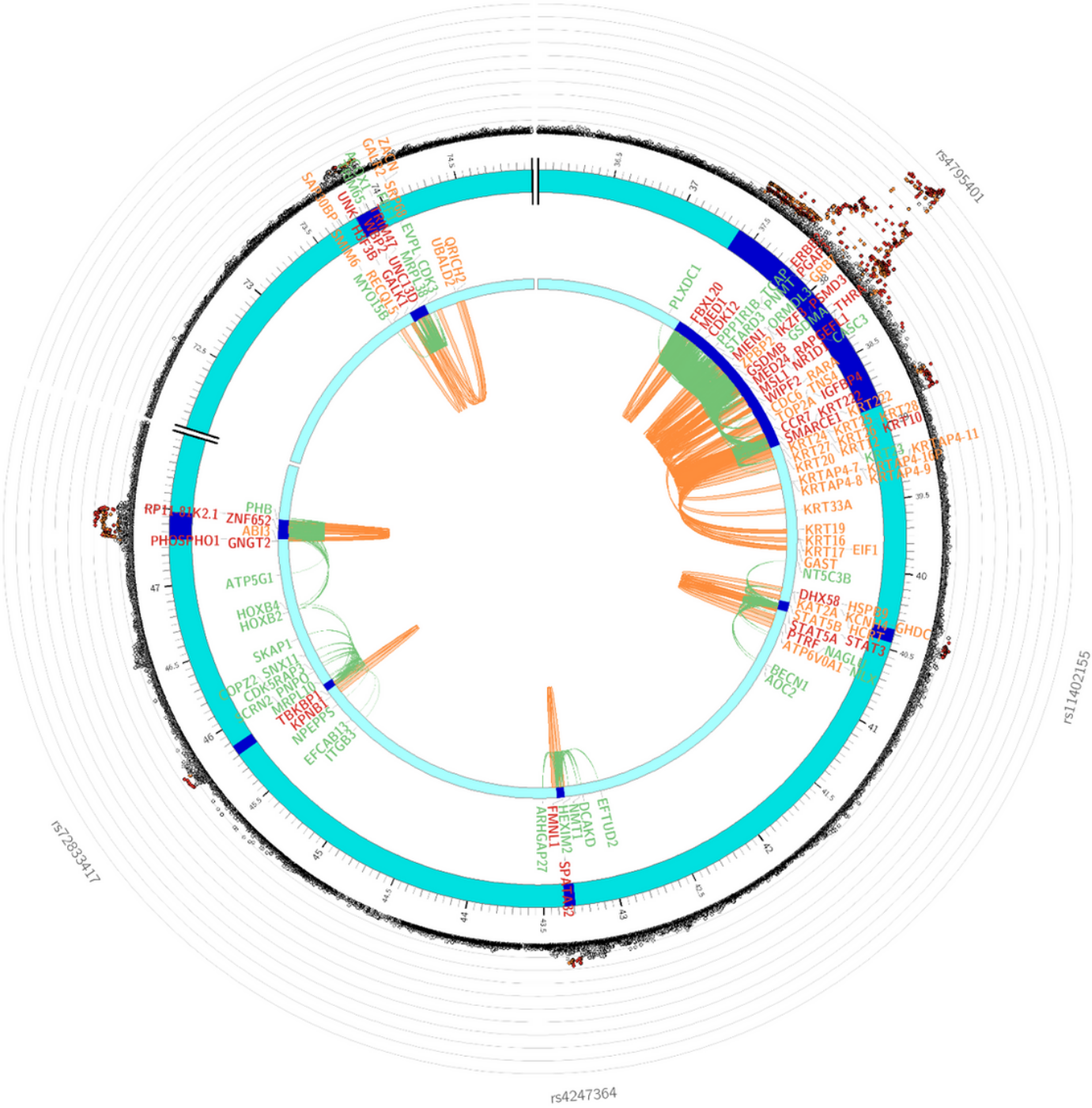

Figure 5

Blood eQTLs and chromatin interactions in GM12878 at asthma risk loci on chromosome 17. Top asthma GWAS SNPs at independent loci are indicated at the most outer border of the circos plot. The subsequent layers show 1) genetic association results from the asthma GWAS in UK Biobank for SNPs with P value < 0.05 with the color of dots reflecting the level of LD with the sentinel variant at each locus (red: r2 > 0.8, orange: r2 > 0.6), 2) the chromosome coordinate and the asthma GWAS loci highlighted in dark blue, 3) blood eQTL genes, genes mapped by Hi-C, and genes mapped by both eQTL and Hi-C labelled in green, orange, and red, respectively, 4) green and orange lines link the position of eQTLs and chromatin interactions, respectively.

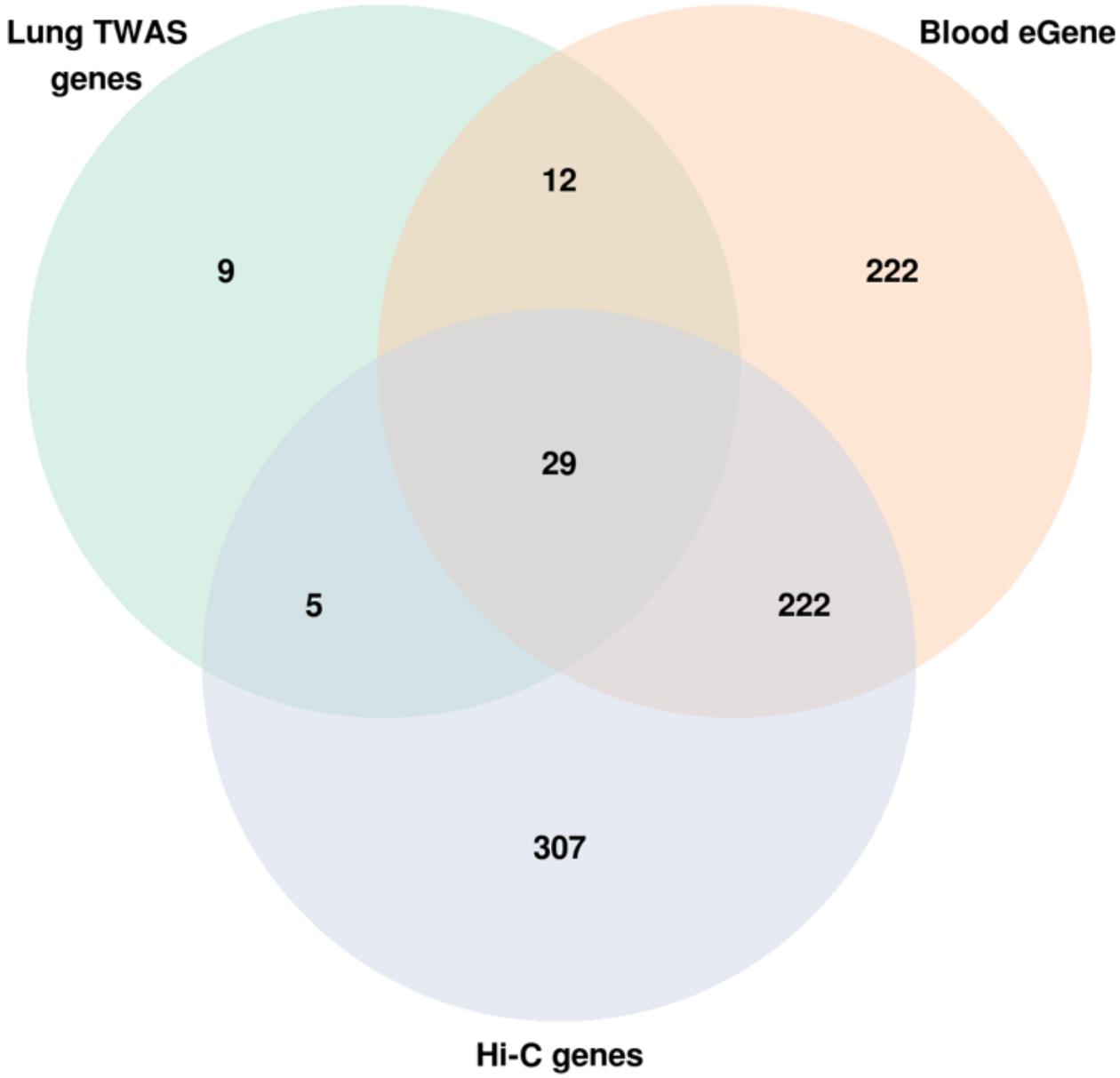

Figure 6

Venn diagram showing the number of target genes that overlapped among 55 lung TWAS genes, 485 blood eGenes, and 563 chromatin contacts mapped genes.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- TWASasthmav6SupplTables.xlsx
- SI.docx