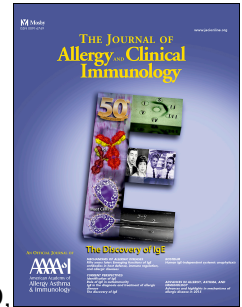


Accepted Manuscript

A Decade of Research on the 17q12-21 Asthma Locus: Piecing Together the Puzzle

Michelle M. Stein, PhD, Emma E. Thompson, PhD, Nathan Schoettler, MD, PhD, Britney A. Helling, PhD, Kevin M. Magnaye, BS, Catherine Stanhope, MS, Catherine Igartua, PhD, Andréanne Morin, PhD, Charles Washington, III, BA, Dan Nicolae, PhD, Klaus Bønnelykke, PhD, Carole Ober, PhD



PII: S0091-6749(17)32943-3

DOI: [10.1016/j.jaci.2017.12.974](https://doi.org/10.1016/j.jaci.2017.12.974)

Reference: YMAI 13186

To appear in: *Journal of Allergy and Clinical Immunology*

Received Date: 15 September 2017

Revised Date: 13 December 2017

Accepted Date: 16 December 2017

Please cite this article as: Stein MM, Thompson EE, Schoettler N, Helling BA, Magnaye KM, Stanhope C, Igartua C, Morin A, Washington III C, Nicolae D, Bønnelykke K, Ober C, A Decade of Research on the 17q12-21 Asthma Locus: Piecing Together the Puzzle, *Journal of Allergy and Clinical Immunology* (2018), doi: 10.1016/j.jaci.2017.12.974.

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

A Decade of Research on the 17q12-21 Asthma Locus: Piecing Together the Puzzle

Michelle M. Stein, PhD^{1,+}, Emma E. Thompson, PhD^{1,+}, Nathan Schoettler, MD, PhD^{1,2,+},
Britney A. Helling, PhD¹, Kevin M. Magnaye, BS¹, Catherine Stanhope, MS¹, Catherine Igartua,
PhD¹, Andréanne Morin, PhD¹, Charles Washington III, BA¹, Dan Nicolae, PhD^{1,3}, Klaus
Bønnelykke, PhD⁴, Carole Ober, PhD^{1,*}

¹Department of Human Genetics, University of Chicago, Chicago, IL 60637

²Section of Pulmonary and Critical Care, Department of Medicine, University of Chicago,
Chicago, IL 60637

³Department of Statistics and Section of Genetic Medicine, Department of Medicine, University
of Chicago, Chicago, IL 60637

⁴COPSAC (Copenhagen Prospective Studies on Asthma in Childhood), Herlev and Gentofte
Hospital, University of Copenhagen, Copenhagen.

⁺The authors contributed equally to this work.

^{*}Corresponding author:

Department of Human Genetics

920 E. 58th Street

University of Chicago

Chicago, IL 60637

TEL: 773-834-0735

FAX: 773-834-0505

e-mail: c-ober@genetics.uchicago.edu

26

27 **Grant Support**

28 This work was supported by R01 HL085197, P01 HL070831, R01 HL129735, U19 AI095230,

29 R01 HL122712, and R01 HL113395. K.M.M. and C.W. are supported by T32 GM007197;

30 M.M.S., N.S. and B.A.H. are supported by T32 HL007605.

Abstract

Chromosome 17q12-21 remains the most highly replicated and significant asthma locus. Genotypes in the core region defined by the first GWAS correlate with expression of two genes, *ORMDL3* and *GSDMB*, making these prime candidate asthma genes, although recent studies have implicated *GSDMA* distal to and *PGAP3* proximal to the core region as independent loci. We review here 10 years of studies on the 17q12-21 locus and suggest that genotype-specific risks for asthma at the proximal and distal loci are not specific to early onset asthma and mediated by *PGAP3*, *ORMDL3*, and/or *GSDMA* expression. We propose that the weak and inconsistent associations of 17q SNPs with asthma in African Americans is due to the high frequency of some 17q alleles, the breakdown of linkage disequilibrium on African-derived chromosomes, and possibly different early life asthma endotypes in these children. Finally, the inconsistent association between asthma and gene expression levels in blood or lung cells from older children and adults suggests that genotype effects may mediate asthma risk or protection during critical developmental windows and/or in response to relevant exposures in early life. Thus, studies of young children and ethnically diverse populations are required to fully understand the relationship between genotype and asthma phenotype and the gene regulatory architecture at this locus.

Key Words

Wheezing, GWAS, immune cells, lung cells, *ORMDL3*, *GSDMB*, *PGAP3*, *GSDMA*, gene expression

Abbreviations

AHR, airway hyperresponsiveness

AR, allergic rhinitis

ASMC, airway smooth muscle cell

BAL, bronchial alveolar lavage

BEC, bronchial epithelial cell

CAAPA, Consortium on Asthma among African Ancestry Populations in the Americans

CBMC, cord blood mononuclear cell

CEU, Utah residents with Northern and Western European ancestry from the CEPH collection

CI, confidence interval

CTCF, CCCTC-binding factor

COAST, Childhood Origins of Asthma birth cohort

COPSAC, Copenhagen Prospective Studies on Asthma in Childhood birth cohort

DHS, DNase hypersensitive site

DNBC, Danish National Birth Cohort

EBV, Epstein-Barr virus

EGEA, Epidemiological Study on Genetics and Environment of Asthma

eQTL, expression quantitative trait loci

ERBB2, erb-b2 receptor tyrosine kinase 2

- 72 ETS, environmental tobacco smoke
- 73 FEV₁, Forced expiratory volume in 1 second
- 74 GPI, glycosylphosphatidylinositol
- 75 GSDMA, gasdermin A
- 76 GSDMB, gasdermin B
- 77 GTEx, Genotype-Tissue Expression project
- 78 GWAS, genome-wide association study
- 79 IgE, immunoglobulin E
- 80 IKZF3, IKAROS family zinc finger 3
- 81 IL, interleukin
- 82 kb, kilobase
- 83 LCL, lymphoblastoid cell line
- 84 LD, linkage disequilibrium
- 85 LPS, lipopolysaccharide
- 86 meQTL, methylation quantitative trait locus
- 87 OR, odds ratio
- 88 ORMDL3, ORM1-like 3
- 89 PASTURE, Protection against Allergy Study in Rural Environments birth cohort
- 90 PBL, peripheral blood leukocyte
- 91 PBMC, peripheral blood mononuclear cell
- 92 PGAP3, post-GPI attachment to proteins 3
- 93 PHA, phytohaemagglutinin
- 94 qPCR, quantitative polymerase chain reaction
- 95 RSV, respiratory syncytial virus

- 96 RV, rhinovirus
- 97 SNP, single nucleotide polymorphism
- 98 TAGC, Transnational Genetics of Asthma Consortium
- 99 TLR3, toll like receptor 3
- 100 TRM, tissue resident memory
- 101 UTR, untranslated region
- 102 WBC, white blood cells
- 103 YRI, Yoruba in Ibadan, Nigeria
- 104 ZPBP2, zona pellucida binding protein 2

The first genome-wide association study (GWAS) of asthma was reported 10 years ago by Moffatt and colleagues¹, who made the seminal discovery of a novel asthma locus on chromosome 17q21, defined by a large 206.5 kilobase (kb) region. This locus has since been extended to include SNPs both proximal and distal¹ to the core locus initially defined by Moffatt *et al.* and is more accurately referred to as the 17q12-21 asthma locus, although in this review, we will simply refer to it as 17q.

Subsequent GWAS and meta-analyses of GWAS have highlighted 17q as the most replicated and most significant asthma locus²⁻¹⁰. Genotypes at SNPs in the core region defined by Moffatt *et al.* have been shown to correlate with the expression of two genes, *ORMDL3* and *GSDMB*^{1, 11-23}, making these the prime candidate asthma genes at this locus. More recent studies have implicated *GSDMA*^{3, 24} distal to and *PGAP3* (previously named *PERLD1*)^{2, 20, 25} proximal to the core region as potentially independent asthma loci (**Fig. 1A-B**). Follow-up studies of 17q have comprised a significant proportion of asthma genetic research over the past decade, however many questions remain. Here, we overview our current understanding of this important locus, focusing on clinical correlations, and the genomic, transcriptomic, and epigenomic studies of this region. Throughout, we highlight 17 SNPs that capture the signature genetic findings (**Fig. 2**). We refer readers to Das *et al.*²⁶ for an excellent review of functional studies of genes at the 17q locus.

Phenotypes and Endotypes Associated with the 17q12-21 Locus

Despite the fact that the 17q locus shows the strongest associations in large GWAS and meta-analyses, the odds ratios (ORs) of these associations are modest. For example, in the Transnational Genetics of Asthma Consortium (TAGC) meta-analysis², which included 75

¹ The terms proximal and distal are used in this review to refer to locations relative to the centromere

independent GWAS comprising ethnically diverse populations totaling 24,925 cases and 95,185 controls, the lead SNP at the 17q locus had a p-value of 2.24×10^{-30} (**Fig. 1A**) but an OR of only 1.15 (95% CIs 1.12 to 1.19). This likely reflects the extensive clinical heterogeneity of asthma, with variation in age of onset, co-occurrence with allergic sensitization, and the manifestation of wheezing illnesses in early life. Moreover, asthma risk is significantly impacted by environmental exposures in the first few years of life, including environmental tobacco smoke (ETS), older siblings, furred pets and large farm animals, that also vary among GWAS subjects. In fact, studies of more clinically homogeneous cases or accounting for relevant exposures have yielded increased ORs for asthma-associated loci, including 17q²⁷.

Age of onset and early life exposures

It is clear that the core 17q locus is associated with early-onset asthma^{2, 6, 14, 28-31}, which was first reported by Bouzigon and colleagues²⁸, who showed that associations with 17q genotypes were restricted to children with onset of asthma symptoms before the age of 4 years (**Fig. 3A**). Furthermore, the association with early-onset asthma was present only in children exposed to ETS in early life. Associations with early age of onset were replicated in subsequent studies that further extended this observation to Asian individuals¹⁴, more severe symptoms in early childhood^{14, 29}, bronchial hyperresponsiveness at one month of age²⁹, exacerbations and hospitalization before the age 6⁶, older sibs³², rhinovirus (RV)-associated wheezing illnesses in the first 3 years of life¹² and wheezing in the first year of life³² (**Fig. 3B**). The observation that 17q-associated asthma risk is modified by early life exposure to ETS²⁸ was replicated in two Dutch birth cohorts³³, in Croatian children³⁴, and in U.K. adults²⁴, but not in the Danish National Birth Cohort (DNBC)³⁵ or the European GABRIEL consortium³⁶.

Surprisingly, the same alleles that are associated with asthma risk in the above studies were associated with *protection* from wheezing among children exposed to animal barns in the first year of life³² (**Fig. 3C**). This protection was dose-dependent: exposure of more than 2 hours a week to animal barns was associated with a >80% reduction in wheezing among children with “high risk” genotypes (rs8076131 AG or GG). In this same study, there was no interaction between genotype and dog ownership on wheezing³², although significant interactions were reported between 17q genotype and furred pets on wheezing in the first 18 months of life in the DNBC study³⁵ and on asthma at school age in Croatian children³⁴ and Danish children in the COPSAC₂₀₀₀ study³⁷. In these studies, the genotypes at 17q SNPs were associated with *protection* among children with a cat or dog. In the COPSAC₂₀₀₀ study, cat (but not dog) allergen levels were associated with asthma among children with the rs7216389 TT genotype.

Associations of the 17q12-21 locus with other phenotypes and diseases

Genotypes at the 17q locus show inconsistent associations with asthma-associated phenotypes. SNPs at this locus were not associated with allergic sensitization^{11, 12, 14, 29, 38, 39} or atopic dermatitis^{29, 40, 41} in GWAS or in 17q-focused studies. Although GWAS of total serum IgE levels did not report associations with 17q SNPs^{3, 42-47}, a 17q-focused study reported an association with IgE in ethnic Chinese subjects from Singapore¹¹. Two 17q-focused studies reported associations with AR^{48, 49}, but a GWAS⁵⁰ and other 17q-focused studies^{11, 29, 32} did not. Neither GWAS⁵¹⁻⁵⁵ nor three 17q-focused studies^{24, 29, 39} reported associations with lung function, whereas one 17q-focused study did³⁴. One GWAS of fractional exhaled nitric oxide reported associations between 17q SNPs⁵⁶ while another did not show evidence of association⁵⁷; and one 17q-focused study reported associations with blood eosinophilia¹¹ but one GWAS did not⁵⁸.

Interestingly, 17q SNPs have been associated with other diseases in GWAS, all of which are autoimmune in etiology: Crohn's disease⁵⁹, ulcerative colitis⁶⁰, inflammatory bowel disease⁶¹, type I diabetes^{62, 63}, rheumatoid arthritis⁶⁴, and primary biliary cirrhosis^{65, 66}. Although the same SNPs are associated with both asthma and autoimmune disease, the risk alleles associated with autoimmune disease are opposite of those associated with asthma risk, as noted^{23, 67, 68}. These findings suggest that genes at the core locus impact immune development broadly²³. Moreover, the overlapping associations of asthma and autoimmune disease with SNPs at the core region, but not at the proximal or distal regions, suggest that the latter two regions may confer lung specificity to asthma whereas SNPs at the core region may be more central to early life immune responses to asthma-promoting or asthma-protective exposures.

Fine Mapping and Ethnic-Specific Associations at the 17q Locus

Fine mapping at the 17q locus has been challenging due to the extensive linkage disequilibrium (LD) in European, Latino, and Asian ancestry populations (**Fig. 1D**) and the weak and inconsistent evidence for association between asthma and 17q SNPs in African ancestry populations^{5, 69-72}. The co-regulation of genes at this locus and the different patterns of gene regulation in blood cells and lung cells impose additional layers of complexity.

Patterns of LD at the 17q Locus

LD is very strong between SNPs at the 17q locus, with an extended haplotype spanning the entire core region in populations of European, Latino, and East Asian ancestry (**Fig. 1D**). This feature makes it virtually impossible to distinguish the effects of specific SNPs with any confidence. In fact, the evidence for association between SNPs across the core region and asthma

is strong and consistent in Europeans^{2, 3, 5}, Latinos^{10, 70}, and ethnic Chinese from Singapore¹¹ and Hong Kong⁷³⁻⁷⁵. Associations are less consistent in smaller studies of Han Chinese from China⁷⁶⁻⁷⁸, and Korean^{14, 79, 80} and Japanese^{47, 81} subjects.

In contrast, LD extends over much shorter distances in African Americans (**Fig. 1D**) and Africans (Supplementary Fig. 1), except for a region in *GSDMB* defined by three SNPs that show near perfect LD in all populations and were the most significant SNPs in GWAS of asthma in Europeans^{3, 6} and ethnically diverse U.S. populations⁵, and in a 17q-focused study in Han Chinese⁷⁷. The three SNPs consist of a common missense variant that is predicted to be “possibly damaging” by PolyPhen^{82, 83} (rs2305480; c.892C>T, p.P298S), an intronic variant that affects alternatively spliced transcript abundances⁸⁴ (rs11078928; c.662T>C), and an intronic variant of unknown function (rs11078927). The rs11078928-T allele is associated with higher expression of *GSDMB* due to aberrant splicing, and the very high frequency of this allele in African Americans (~0.82) could contribute to the high prevalence of asthma among these individuals. In the EVE meta-analysis, the most significant association in African Americans was with rs11078927, making this the most significant association in the combined sample. Although of modest significance in African Americans ($P = 0.0019$), the estimated effect size for rs11078927 was similar to those reported for the other EVE populations. This is likely due to the high frequency of the three *GSDMB* SNPs in African Americans (0.82), which reduces power to detect associations (**Fig. 1C**).

The breakdown of LD in the core region in African Americans should facilitate fine mapping of the 17q locus if the true causal variants are genotyped or tagged by the SNPs typed in those studies. The inconsistency between studies raise the possibility that the 17q-associated risk may involve multiple variants across the extended haplotype that are not in LD and possibly

not tagged by the GWAS SNPs on African-derived chromosomes. Ethnic-specific rare variants at the 17q locus could also influence asthma risk, although a study of rare coding variants in 2,308 African American cases and controls found no significant associations at 17q genes⁸³. However, the array used included only 57% (8 of 14) of the missense, nonsense, or splicing variants discovered in 17q genes in whole genome sequences from 93 of the African American asthmatics in this study, raising the possibility that rare variants may have been missed. These important questions will be addressed by the Consortium on Asthma among African Ancestry Populations in the Americas (CAAPA)⁸⁵ in an asthma GWAS in nearly 18,000 African ancestry individuals, using whole genome sequences from a subset of samples as the imputation reference panel.

Finally, it is also possible that the natural history of early-life asthma among African American children differs from that of European ancestry children with respect to age of onset, viral-associated wheezing illness, or environmental exposures that modify or interact with the 17q-associated risk. In fact, U.S. inner city African American children had lower rates of viral detection overall, fewer RV infections in particular, and higher rates of adenovirus infection and adenovirus-associated illness in the first year of life compared to European ancestry children living in Madison, WI⁸⁶. Thus, the specific type of asthma in African American children may be less influenced by variation at the 17q locus, or by different variants than are associated with asthma in other children. It is likely that one or more of these factors could be masking associations with SNPs at the 17q locus in African ancestry populations.

Evidence for three independent asthma associations at the 17q Locus

Several lines of evidence suggest that SNPs in *GSDMA* and in *PGAP3* may have effects that are independent of SNPs at the core locus. In all populations, the LD that characterizes the core region decays in the distal region encoding *GSDMA* and the proximal region encoding *PGAP3* and *ERBB2* (LD $r^2 \leq 0.5$; **Fig. 1D**), and residual associations with asthma remained at rs3894194 in *GSDMA*³ and at rs2941504 in *PGAP3*¹⁴ after conditioning on SNPs at the core locus. Moreover, because the association with asthma at both regions is reduced in a pediatric sample (see Fig. S8 in Demenais *et al.*²), the effects of variation at the proximal and distal loci may not be specific to childhood-onset asthma. SNP rs2941504 in *PGAP3* was also associated with allergic asthma, but not with atopy among non-asthmatics, in ethnic Chinese adults from Singapore²⁵. SNP rs3894194 in *GSDMA* was also associated with current asthma, forced expiratory volume in 1 second (FEV₁), and airway hyperresponsiveness (AHR), as well as showing interactive effects with cigarette smoking and current asthma in a 17q-focused study of 1,018 U.K. adults²⁴. Another SNP in *GSDMA*, rs3859192, is a strong eQTL for *GSDMA* in whole lung tissue^{15, 87}, although it shows little LD in all populations with the asthma-associated *GSDMA* SNP rs3894194.

Cell-Specific Expression and eQTL Studies of 17q Genes

Patterns of gene expression and eQTLs in asthma relevant tissues (i.e., immune cells and lung cells) can give clues as to which genes contribute to asthma pathogenesis and potentially which SNPs are either themselves causal or tag causal variation. Although nearly all the 17q genes are ubiquitously expressed in asthma-relevant cell types, the relative expression patterns and associations with SNPs (i.e., eQTLs) differ between these cells.

Patterns of gene expression

Six of the seven genes at the 17q locus potentially implicated in asthma risk (*PGAP3*, *ERBB2*, *IKZF3*, *GSDMB*, *ORMDL3*, *GSDMA*) (**Fig. 1B**, shown in red) are expressed at detectable levels in RNA from whole blood⁸⁷, LCLs²³, peripheral blood mononuclear cells (PBMCs)⁸⁸, peripheral blood leukocytes (PBLs) (Stein, unpublished), lung CD4+ tissue resident memory (TRM) cells (Schoettler, unpublished), whole lung tissue⁸⁷, freshly isolated bronchial epithelial cells (BECs)²⁰, and cultured primary airway smooth muscle cells (ASMCs) (Thompson, unpublished). Using quantitative (q)PCR to measure transcript levels, *ZPBP2* expression was also detected in LCLs²³ and cord blood mononuclear cells (CBMCs)⁸⁹. Among immune cell types, transcript levels of *ORMDL3* and *GSDMB* were most highly expressed in T and B cells^{21, 90}, and in one study *ORMDL3* was most highly expressed in eosinophils¹⁷.

ORMDL3 and *GSDMB* transcription increases in response to viral infection. Expression levels of both *ORMDL3* and *GSDMB*, but not *IKZF3*, increased in PBMCs from 96 adults after exposure to RV⁸⁸, and *ORMDL3* expression levels were higher in PBMCs from 10 children with RSV-induced bronchiolitis compared to 15 uninfected children⁹¹. *ORMDL3* expression was increased in primary human lung fibroblasts, but not in primary human BECs or primary human ASMCs, after exposure to poly(I:C), a TLR3 ligand, but not in response to LPS in any of the cells⁴⁷. The response of these genes to important respiratory viruses further supports an important role for this locus in early-life wheezing illness and childhood onset asthma.

As expected, the relative expression of the 17q genes varies by cell type, but the patterns of variation are revealing. In all cells examined here (**Fig. 4**), *GSDMA* transcripts were expressed at levels just above the lower limits of detection in most individuals, so conclusions about this gene should be tempered by this observation. Moreover, this broader view of gene expression

across the 17q locus shows that within individuals, transcript levels of *ORMDL3* were most correlated with those of *PGAP3* and not of *GSDMB* in PBLs, BECs and ASMCs (**Figs. 4-5**). In lung CD4⁺ TRMs, *ORMDL3* expression was not correlated with the expression of any other 17q genes, although the sample size was smallest for these cells.

Cell-type specific eQTLs

Published eQTL studies of 17q SNPs and genes are summarized in **Table 1**. To further evaluate the effects of genotype at the 17 SNPs highlighted in this review on the expression levels of the six candidate 17q genes, we extracted data from the Gene-Tissue Expression (GTEx) Consortium for whole blood and lung tissue⁸⁷ (**Fig. 6**). In whole blood, all 17 SNPs were significant eQTLs for *ORMDL3* and *GSDMB*, but not for any of the other 17q genes. At each SNP, the asthma-associated allele was correlated with increased expression of *ORMDL3* and *GSDMB*. Thus, as in previous studies (**Table 1**), SNPs at the 17q core locus are strong eQTLs for these two genes in peripheral blood cells from primarily non-African ancestry populations.

A different picture emerges in lung tissue cells. SNPs across the region are significant eQTLs for *ORMDL3*, similar to studies in blood but with overall smaller effects (**Fig. 6**). However, only the six most significant eQTLs for *ORMDL3* are also more modest eQTLs for *GSDMB*. Notably, the splice variant in *GSDMB* is not an eQTL for *GSDMB* in lung tissue. In contrast to eQTLs in blood, SNPs in the core and distal region are also eQTLs for *GSDMA*. Additionally, SNPs in the proximal and core regions are strong eQTLs for *PGAP3*. The asthma-associated alleles at each of these SNPs are correlated with increased expression of both *PGAP3* and *ORMDL3*, while in the distal region, the asthma risk alleles at rs3894194 and rs3859192 are correlated with lower expression of *GSDMA*, as previously reported^{15, 92}. These data further

suggest that the expression of *ORMDL3* and *GSDMB* are not co-regulated in lung cells as they are in blood cells, as suggested previously in studies of BECs²⁰.

To examine eQTL patterns in a population with less LD across this region, we extracted eQTL data for 17q SNPs from a study of LCLs from the Nigerian HapMap population (YRI)⁹³ (**Fig. 7**). In contrast to studies in whole blood and in primarily non-African ancestry populations, the 17q SNPs were only significant eQTLs for *ORMDL3* in African-derived LCLs, showing considerably less evidence as eQTLs for *GSDMB*. Moreover, among the SNPs in the core region, the strongest eQTL for *ORMDL3* was rs12936231 in *ZPBP2* ($P=5.35 \times 10^{-10}$); the strength of the associations between other SNPs and *ORMDL3* expression were correlated with the degree of LD with rs12936231, a strong functional candidate for the observed associations with 17q SNPs^{21, 23} (discussed below). Additionally, two SNPs in *PGAP3* are associated with *ORMDL3* expression despite little LD between these SNPs in these regions (**Fig. 1D**), possibly reflecting long-range interactions between a putative enhancer in *PGAP3* and the *ORMDL3* promoter in LCLs²⁰ and further suggesting independent effects of SNPs in the proximal region on *ORMDL3* expression. Although we cannot exclude the six less informative SNPs in the core region from being eQTLs because of the very low power to detect associations in this sample, these data indicate that the effect of rs12936231 on *ORMDL3* expression levels is independent of the other 17q SNPs. Curiously, rs12936231 was not associated with asthma in African Americans in the EVE Consortium⁵, despite having an allele frequency near 0.50 and high statistical power to detect association. These findings suggest that increased expression of *GSDMB* or *ORMDL3* in circulating immune cells may not be causally related to asthma, as has been previously suggested¹⁴.

Gene expression differences between asthmatic and non-asthmatic individuals

The observation that 17q SNPs are associated both with asthma risk and expression levels of 17q genes does not consistently extend to differences in 17q gene expression levels between asthmatics and non-asthmatics. For example, one small study of Swedish asthmatic and non-asthmatic school-aged children⁹⁰ reported differences in 17q gene expression levels in PBMCs: *ORMDL3* transcript abundance was higher in 16 children with controlled asthma compared to 15 healthy controls and 16 children with severe asthma (ANOVA, $P = 0.002$); however, there were no differences between children with severe asthma and controls. In the original GWAS by Moffatt and colleagues¹, transcript levels of *ORMDL3* in LCLs were not significantly different between 112 asthmatic and 266 non-asthmatic children. None of the 17q genes were differentially expressed in BECs from 55 adults with mild-to-severe asthma and 26 healthy controls²⁰ or in 42 adults with mild-to-moderate asthma and healthy controls⁹⁴. In contrast, *ORMDL3* was increased in BECs from 19 adults with stable mild asthma and 16 healthy controls (FDR 8.5×10^{-4})⁹⁵. Although *GSDMB* transcripts were not increased in BECs from asthmatics in any of these studies, Das and colleagues⁹⁶ reported increased *GSDMB* protein by immunohistochemistry in BECs from 14 asthmatics compared to 7 healthy controls ($P < 0.001$), with significantly more *GSDMB*⁺ cells among 7 severe asthmatics compared to controls ($P < 0.01$). The inconsistent results and paucity of studies demonstrating increased expression of 17q genes in asthma is surprising. This could reflect insufficient power to detect gene expression differences^{1, 20}, a focus on cell types or asthma endotypes that are unrelated to the 17q-associated risk, and/or collecting samples at developmental windows or in environments that are irrelevant to the 17q-associated asthma risk.

Chromatin and DNA Methylation Studies at the 17q Locus

Studies of chromatin architecture and DNA methylation patterns in blood and lung cells have provided insights into the specific SNPs that regulate gene expression at this locus.

Allele-specific chromatin states and DNA methylation in immune cells

Verlaan and colleagues²³ were the first to study *cis* regulatory elements associated with gene expression patterns and genotypes at SNPs at the core 17q locus. Using LCLs from 53 European (CEU) HapMap samples, they first showed that the asthma-associated alleles were on an extended haplotype that was associated with increased expression of *ORMDL3* and *GSDMB*. They further showed that the non-risk G allele at rs12936231 in *ZBPB2* was depleted of nucleosomes (a signature of regions with regulatory function) and enriched for inactive chromatin marks and binding of CTCF, an important protein that can act as both an insulator between regulatory domains and a mediator of enhancer-promoter interactions⁹⁷. This study therefore discovered an asthma-associated SNP, rs12936231, that alters the chromatin state of a regulatory domain and is correlated with expression of *ORMDL3* and *GSDMB*.

This observation was extended by Schmiedel *et al.*²¹, who assessed open chromatin state in 62 primary cell types at the core 17q locus (from *IKZF3* to *GSDMA*) as well as allele-specific associations with enhancer activity, CTCF-binding, and gene expression in 10 immune cell types and BECs. Overall, 17q genes were more transcriptionally active in immune cells compared to non-immune cells, as measured by the number of open chromatin sites assessed by DNase hypersensitivity (DHS). Many of the DHS sites in T cells, B cells and NK cells overlapped asthma-associated SNPs, including the rs12936231 SNP in *ZBPB2* identified by Verlaan *et al.*²³, as well as SNPs in the first intron of *ORMDL3*, including rs4065275. These SNPs, among others,

were enriched for H3K27ac chromatin marks of active enhancer activity in T cells; the asthma risk alleles were associated with increased enrichment of H3K27ac. Additionally these two SNPs were predicted to disrupt CTCF-binding motifs, which were experimentally confirmed by ChIP-seq: the asthma-associated allele in *ZBP2* (rs12936231-C) disrupted a CTCF-binding site and the asthma-associated allele in *ORMDL3* (rs4065275-G) introduced a CTCF-binding site (**Fig. 8A**). Finally, chromatin conformation assays (4C-seq) showed that the *ORMDL3* promoter interacts with a long-range enhancer in *IKZF3* that promotes transcription of *ORMDL3* in cells that also express *IKZF3*. The binding of CTCF on chromosomes with the rs12936231-G allele in *ZBP2* blocks this interaction, resulting in reduced transcription of *ORMDL3* on haplotypes with the rs12936231-G allele, presumably independent of rs4065275 (**Fig. 8b**). Because *IKZF3* is a transcription factor in B cells, T cells and selected other immune cells, the chromatin looping and physical interaction between the *IKZF3* enhancer and the *ORMDL3* promoter occurs only in immune cells. This likely accounts for the very high expression of *ORMDL3* in immune cells compared to airway cells, where this interaction does not occur²¹.

Because of the LD structure at the 17q locus, nearly all non-African ancestry haplotypes will carry either both (rs12936231-C and rs4065275-G) or neither (rs12936231-G and rs4065275-A) asthma associated alleles, which correspond to high and low expressers of *ORMDL3*, respectively (**Fig. 8A**). However, a recombinant haplotype with rs12936231-G and rs4065275-G is relatively common on African-derived chromosomes, and is associated with low expression of *ORMDL3* in YRI LCLs, despite carrying the high expressing allele rs4065275-G (**Fig. 8B**). As a result, approximately 30% of African-derived chromosomes carrying the asthma-associated rs4065275-G allele will be low expressers of *ORMDL3*, whereas nearly all European-

or Asian-derived chromosomes with rs4065275-G will be high expressers of *ORMDL3*. This could explain why rs4065275 is a very weak eQTL for *ORMDL3* in YRI LCLs (**Fig. 7**).

Two studies examined allele-specific differences in DNA methylation patterns at the core locus. Combining DNA methylation patterns with nucleosome occupancy and *in vitro* promoter and enhancer activity in LCLs (CEU), Berlivet and colleagues⁹⁸ provided further evidence for an enhancer of *ORMDL3* located in *ZBP2*. Acevedo *et al.*⁹⁰ focused on three SNPs within CpG dinucleotides: the risk allele at rs7216389 in *GSDMB* removes a CpG site and the risk alleles at rs4065275 and rs12603332 in *ORMDL3* create CpG sites. These SNPs affected methylation levels both at the position they are located in and at other CpG sites in the 5' untranslated region (UTR) of *ORMDL3*. The asthma-associated SNPs and the allele-specific methylation patterns were independently associated with expression of *ORMDL3* and *GSDMB* in PBLs, but only methylation patterns at, but not expression of, *ORMDL3* differed between children with and without asthma.

DNA methylation and gene regulation in BECs

Nicodemus-Johnson *et al.*²⁰ performed genome-wide eQTL and DNA methylation (me)QTL mapping in freshly isolated BECs from 74 asthmatic and 41 non-asthmatic adults. The most significant eQTL for *ORMDL3* was rs2517955, located 240 bp from *ORMDL3* in an intron of *PGAP3* (eQTL $P=1.56 \times 10^{-5}$) and showing little LD with genotypes at the core locus (**Fig. 1D**). This SNP was also a significant eQTL for *PGAP3* ($P=2.07 \times 10^{-8}$), but not for *GSDMB* or any other genes at the core locus. SNP rs2517955 was also an meQTL for a nearby CpG, cg05616858 (meQTL $P=8.95 \times 10^{-13}$), which was itself correlated with expression of *ORMDL3* ($P=3.87 \times 10^{-4}$). A conditional analysis that accounted for genotype at the core 17q locus

confirmed that the association between rs2517955 with *ORMDL3* expression is independent of SNPs at the core locus. To elucidate the causal relationship between rs2517955 genotype, cg05616858 methylation, and *ORMDL3* gene expression, the investigators used Mendelian randomization and showed that methylation at this CpG site directly influences *ORMDL3* expression level independent of rs2517955. This SNP was also associated with asthma in TAGC ($P = 7.6 \times 10^{-29}$).

Interestingly, rs2517955 resides within a peak of H3K27ac histone marks in ENCODE data (all cell types pooled), suggestive of an enhancer in this region. Chromatin capture data (Hi-C) in LCLs demonstrated looping and physical interaction between the putative enhancer at rs2517955 and the promoter of *ORMDL3*, ~240 kb away²⁰, consistent with the correlated expression levels of these two genes in PBLs, BECs, and ASMCs (**Fig. 4**), providing a potential mechanism for these observations and supporting an independent asthma risk locus in the proximal 17q region.

Synthesis and Future Directions

Although our understanding of the 17q asthma locus has significantly deepened since its discovery, research over the past few years has revealed appreciable levels of complexity for this locus. Importantly, several lines of evidence support the presence of three independent asthma-associated loci at the extended 17q12-21 region.

At least two SNPs at the core locus directly regulate the expression of *ORMDL3* or *GSDMB* in immune cells. The asthma-associated allele in *GSDMB*, rs11078928-T, is associated with higher expression of this gene due to aberrant splicing associated with the alternate C allele, and the rs12936231-C allele in *ZBP2* that destroys a CTCF motif is associated with increased

expression of *ORMDL3*. Due to LD, two haplotypes corresponding to high and low expressers of *ORMDL3* and *GSDMB* account for 95% of non-African haplotypes. However, these two SNPs have very different allele frequencies and are in low LD on African-derived chromosomes (**Fig. 1C-D**), potentially breaking up the co-regulation of these two genes in African ancestry populations. Moreover, 18.5% of African-derived chromosomes are recombinants carrying the rs12603332-G allele that is associated with low expression of *ORMDL3* and the rs4065275-G allele that is associated with high expression of *ORMDL3* in non-African ancestry individuals (**Fig. 7A**). However, in African individuals this haplotype is associated with low expression of *ORMDL3* (**Fig. 7B**), indicating that rs12936231 in *ZBP2* has a “dominant” effect on expression of *ORMDL3*. It is not surprising, therefore, that rs12936231 is the most significant eQTL for *ORMDL3* in whole blood and lung tissue cells (**Fig. 5**) and in African-derived LCLs (**Fig. 6**). Thus, if high expression of *ORMDL3* by itself was the underlying cause of asthma, this SNP in *ZBP2* should be associated with asthma in GWAS. Yet rs12936231 has never been reported as a lead SNP in GWAS.

The long-range looping and correlated expression patterns between *PGAP3* in the proximal region and *ORMDL3* in the core region, and the correlated expression of their transcripts, are intriguing and raise additional questions. Are the combinations of genotypes at the proximal and core locus SNPs more associated with asthma than each one individually? Are the chromatin interactions and looping between *PGAP3* and *ORMDL3* inhibited by the binding of CTCF at rs12936231, as it is for looping within the core locus? Do *PGAP3* and *ORMDL3* proteins functionally interact to promote asthma? At the present time, little is currently known about the function of *PGAP3*, which is a glycosylphosphatidylinositol (GPI)-specific phospholipase that is ubiquitously expressed and primarily localizes to the Golgi apparatus. It is

predicted to encode a seven-transmembrane protein that removes fatty acids from GPI, which may be important for proper association between GPI-anchored proteins and lipid rafts⁹⁹. A role for PGAP3 in asthma has not yet been explored. In contrast, many potential functions have been attributed to ORMDL3, including sphingolipid metabolism¹⁰⁰⁻¹⁰⁵, ER stress^{17, 106-109}, eosinophil trafficking¹¹⁰, and responses to *Alternaria*¹¹¹, RSV⁹¹, poly(I:C)⁴⁷, and IL-17 secretion¹⁸ (reviewed in Das *et al.*²⁶).

Although less is known about *GSDMA*, a missense variant (rs3894194) has shown independent associations for asthma and specific lung phenotypes, and a SNP (rs3859192) in this gene is an eQTL for *GSDMA* in lung tissue cells. These two SNPs show very little LD with each other in all populations. As a result, a relationship between the eQTL or between *GSDMA* expression and asthma *per se* has not been established. Additionally, the eQTL should be interpreted cautiously because of the very low expression of *GSDMA* in lung cells.

Finally, the combined observations of consistently strong associations between 17q SNPs with both early-onset asthma and the expression of 17q genes but conflicting evidence for increased expression of 17q genes in immune cells or airway cells in older children and adults with asthma is enigmatic, and may suggest that the critical window for regulating the expression of *ORMDL3*, *GSDMB* or other 17q genes occurs in the first few years of life. In that case, studying children at school age or later will miss this relationship. Additionally, specific exposures, such as viruses or other microbes, may enhance genotype-specific differences in transcript abundance between asthmatic and non-asthmatic children, as suggested by a study in CBMCs¹⁸. Prospective birth cohorts will be required to examine early life gene expression and epigenetic patterning and the subsequent development of asthma. Ideally, these studies would be performed in both immune and airway cells, during periods of wellness and during respiratory

494 infections in the first few years of life, and include ethnically diverse children and children
495 exposed to different environments in infancy. More comprehensive and integrative studies of
496 17q genotype with gene expression and epigenetic variation using systems biology approaches in
497 carefully phenotyped children will yield further insights into the genetic risk architecture of this
498 important locus. Given the current pace of research, we expect the next decade to yield answers
499 to these questions and translate findings into asthma prevention strategies for children with the
500 17q high risk genotype.

References

1. Moffatt MF, Kabesch M, Liang L, Dixon AL, Strachan D, Heath S, et al. Genetic variants regulating ORMDL3 expression contribute to the risk of childhood asthma. *Nature* 2007; 448:470-3.
2. Demenais F, Margaritte-Jeannin P, Barnes KC, Cookson WO, Altmueller J, Ang W, et al. Multi-ancestry genome-wide association study identifies new asthma susceptibility loci that co-localize with immune cell enhancer histone marks. *Nat Genet* 2017; in press.
3. Moffatt MF, Gut IG, Demenais F, Strachan DP, Bouzigon E, Heath S, et al. A large-scale, consortium-based genomewide association study of asthma. *N Engl J Med* 2010; 363:1211-21.
4. Pickrell JK. Joint analysis of functional genomic data and genome-wide association studies of 18 human traits. *Am J Hum Genet* 2014; 94:559-73.
5. Torgerson DG, Ampleford EJ, Chiu GY, Gauderman WJ, Gignoux CR, Graves PE, et al. Meta-analysis of genome-wide association studies of asthma in ethnically diverse North American populations. *Nat Genet* 2011; 43:887-92.
6. Bønnelykke K, Sleiman P, Nielsen K, Kreiner-Møller E, Mercader JM, Belgrave D, et al. A genome-wide association study identifies CDHR3 as a susceptibility locus for early childhood asthma with severe exacerbations. *Nat Genet* 2014; 46:51-5.
7. Ferreira MA, McRae AF, Medland SE, Nyholt DR, Gordon SD, Wright MJ, et al. Association between ORMDL3, IL1RL1 and a deletion on chromosome 17q21 with asthma risk in Australia. *Eur J Hum Genet* 2011; 19:458-64.
8. Galanter JM, Gignoux CR, Torgerson DG, Roth LA, Eng C, Oh SS, 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* 2014; 134:295-305.
9. Wan YI, Shrine NR, Soler Artigas M, Wain LV, Blakey JD, Moffatt MF, et al. Genome-wide association study to identify genetic determinants of severe asthma. *Thorax* 2012; 67:762-8.
10. Yan Q, Brehm J, Pino-Yanes M, Forno E, Lin J, Oh SS, et al. A meta-analysis of genome-wide association studies of asthma in Puerto Ricans. *Eur Respir J* 2017; 49.
11. Andiappan AK, Sio YY, Lee B, Suri BK, Matta SA, Lum J, et al. Functional variants of 17q12-21 are associated with allergic asthma but not allergic rhinitis. *J Allergy Clin Immunol* 2016; 137:758-66 e3.

- 534 12. Çalışkan M, Bochkov YA, Kreiner-Moller E, Bønnelykke K, Stein MM, Du G, et al.
535 Rhinovirus wheezing illness and genetic risk of childhood-onset asthma. *N Engl J Med* 2013;
536 368:1398-407.
- 537 13. Dixon AL, Liang L, Moffatt MF, Chen W, Heath S, Wong KC, et al. A genome-wide
538 association study of global gene expression. *Nat Genet* 2007; 39:1202-7.
- 539 14. Halapi E, Gudbjartsson DF, Jonsdottir GM, Bjornsdottir US, Thorleifsson G, Helgadottir H,
540 et al. A sequence variant on 17q21 is associated with age at onset and severity of asthma. *Eur*
541 *J Hum Genet* 2010; 18:902-8.
- 542 15. Hao K, Bosse Y, Nickle DC, Pare PD, Postma DS, Laviolette M, et al. Lung eQTLs to help
543 reveal the molecular underpinnings of asthma. *PLoS Genet* 2012; 8:e1003029.
- 544 16. Li X, Hastie AT, Hawkins GA, Moore WC, Ampleford EJ, Milosevic J, et al. eQTL of
545 bronchial epithelial cells and bronchial alveolar lavage deciphers GWAS-identified asthma
546 genes. *Allergy* 2015; 70:1309-18.
- 547 17. Liu YP, Rajamanikham V, Baron M, Patel S, Mathur SK, Schwantes EA, et al. Association
548 of ORMDL3 with rhinovirus-induced endoplasmic reticulum stress and type I Interferon
549 responses in human leucocytes. *Clin Exp Allergy* 2017; 47:371-82.
- 550 18. Lluís A, Schedel M, Liu J, Illi S, Depner M, von Mutius E, et al. Asthma-associated
551 polymorphisms in 17q21 influence cord blood ORMDL3 and GSDMA gene expression and
552 IL-17 secretion. *J Allergy Clin Immunol* 2011; 127:1587-94 e6.
- 553 19. Murphy A, Chu JH, Xu M, Carey VJ, Lazarus R, Liu A, et al. Mapping of numerous disease-
554 associated expression polymorphisms in primary peripheral blood CD4⁺ lymphocytes. *Hum*
555 *Mol Genet* 2010; 19:4745-57.
- 556 20. Nicodemus-Johnson J, Myers RA, Sakabe NJ, Sobreira DR, Hogarth DK, Naureckas ET, et
557 al. DNA methylation in lung cells is associated with asthma endotypes and genetic risk. *JCI*
558 *Insight* 2016; 1:e90151.
- 559 21. Schmiedel BJ, Seumois G, Samaniego-Castruita D, Cayford J, Schulten V, Chavez L, et al.
560 17q21 asthma-risk variants switch CTCF binding and regulate IL-2 production by T cells.
561 *Nat Commun* 2016; 7:13426.
- 562 22. Sharma S, Zhou X, Thibault DM, Himes BE, Liu A, Szeftler SJ, et al. A genome-wide survey
563 of CD4(+) lymphocyte regulatory genetic variants identifies novel asthma genes. *J Allergy*
564 *Clin Immunol* 2014; 134:1153-62.

23. Verlaan DJ, Berlivet S, Hunninghake GM, Madore AM, Lariviere M, Moussette S, et al. Allele-specific chromatin remodeling in the ZBP2/GSDMB/ORMDL3 locus associated with the risk of asthma and autoimmune disease. *Am J Hum Genet* 2009; 85:377-93.
24. Marinho S, Custovic A, Marsden P, Smith JA, Simpson A. 17q12-21 variants are associated with asthma and interact with active smoking in an adult population from the United Kingdom. *Ann Allergy Asthma Immunol* 2012; 108:402-11 e9.
25. Anantharaman R, Andiappan AK, Nilkanth PP, Suri BK, Wang de Y, Chew FT. Genome-wide association study identifies PERLD1 as asthma candidate gene. *BMC Med Genet* 2011; 12:170.
26. Das S, Miller M, Broide DH. Chromosome 17q21 Genes ORMDL3 and GSDMB in Asthma and Immune Diseases. In: *Advances in Immunology*: Academic Press; 2017.
27. Bønnelykke K, Ober C. Leveraging gene-environment interactions and endotypes for asthma gene discovery. *J Allergy Clin Immunol* 2016; 137:667-79.
28. Bouzigon E, Corda E, Aschard H, Dizier MH, Boland A, Bousquet J, et al. Effect of 17q21 variants and smoking exposure in early-onset asthma. *N Engl J Med* 2008; 359:1985-94.
29. Bisgaard H, Bønnelykke K, Sleiman PM, Brasholt M, Chawes B, Kreiner-Moller E, et al. Chromosome 17q21 gene variants are associated with asthma and exacerbations but not atopy in early childhood. *Am J Respir Crit Care Med* 2009; 179:179-85.
30. Kreiner-Moller E, Strachan DP, Linneberg A, Husemoen LL, Bisgaard H, Bønnelykke K. 17q21 gene variation is not associated with asthma in adulthood. *Allergy* 2015; 70:107-14.
31. Smit LA, Bouzigon E, Pin I, Siroux V, Monier F, Aschard H, et al. 17q21 variants modify the association between early respiratory infections and asthma. *Eur Respir J* 2010; 36:57-64.
32. Loss GJ, Depner M, Hose AJ, Genuneit J, Karvonen AM, Hyvarinen A, et al. The Early Development of Wheeze. Environmental Determinants and Genetic Susceptibility at 17q21. *Am J Respir Crit Care Med* 2016; 193:889-97.
33. van der Valk RJ, Duijts L, Kerkhof M, Willemsen SP, Hofman A, Moll HA, et al. Interaction of a 17q12 variant with both fetal and infant smoke exposure in the development of childhood asthma-like symptoms. *Allergy* 2012; 67:767-74.
34. Blekic M, Kljaic Bukvic B, Aberle N, Marinho S, Hankinson J, Custovic A, et al. 17q12-21 and asthma: interactions with early-life environmental exposures. *Ann Allergy Asthma Immunol* 2013; 110:347-53 e2.

- 596 35. Brauner EV, Loft S, Raaschou-Nielsen O, Vogel U, Andersen PS, Sorensen M. Effects of a
597 17q21 chromosome gene variant, tobacco smoke and furred pets on infant wheeze. *Genes*
598 *Immun* 2012; 13:94-7.
- 599 36. Scholtens S, Postma DS, Moffatt MF, Panasevich S, Granell R, Henderson AJ, et al. Novel
600 childhood asthma genes interact with in utero and early-life tobacco smoke exposure. *J*
601 *Allergy Clin Immunol* 2014; 133:885-8.
- 602 37. Stokholm J, Chawes BL, Vissing N, Bonnelykke K, Bisgaard H. Cat exposure in early life
603 decreases asthma risk from the 17q21 high-risk variant. *J Allergy Clin Immunol* 2017.
- 604 38. Bonnelykke K, Matheson MC, Pers TH, Granell R, Strachan DP, Alves AC, et al. Meta-
605 analysis of genome-wide association studies identifies ten loci influencing allergic
606 sensitization. *Nat Genet* 2013; 45:902-6.
- 607 39. Granell R, Henderson AJ, Timpson N, St Pourcain B, Kemp JP, Ring SM, et al. Examination
608 of the relationship between variation at 17q21 and childhood wheeze phenotypes. *Journal of*
609 *Allergy and Clinical Immunology* 2013; 131:685-94.
- 610 40. Paternoster L, Standl M, Chen CM, Ramasamy A, Bonnelykke K, Duijts L, et al. Meta-
611 analysis of genome-wide association studies identifies three new risk loci for atopic
612 dermatitis. *Nat Genet* 2012; 44:187-92.
- 613 41. Kim KW, Myers RA, Lee JH, Igartua C, Lee KE, Kim YH, et al. Genome-wide association
614 study of recalcitrant atopic dermatitis in Korean children. *J Allergy Clin Immunol* 2015;
615 136:678-84 e4.
- 616 42. Levin AM, Mathias RA, Huang L, Roth LA, Daley D, Myers RA, et al. A meta-analysis of
617 genome-wide association studies for serum total IgE in diverse study populations. *J Allergy*
618 *Clin Immunol* 2013; 131:1176-84.
- 619 43. Granada M, Wilk JB, Tuzova M, Strachan DP, Weidinger S, Albrecht E, et al. A genome-
620 wide association study of plasma total IgE concentrations in the Framingham Heart Study. *J*
621 *Allergy Clin Immunol* 2012; 129:840-5 e21.
- 622 44. Pino-Yanes M, Gignoux CR, Galanter JM, Levin AM, Campbell CD, Eng C, et al. Genome-
623 wide association study and admixture mapping reveal new loci associated with total IgE
624 levels in Latinos. *J Allergy Clin Immunol* 2015; 135:1502-10.
- 625 45. Weidinger S, Gieger C, Rodriguez E, Baurecht H, Mempel M, Klopp N, et al. Genome-wide
626 scan on total serum IgE levels identifies FCER1A as novel susceptibility locus. *PLoS Genet*
627 2008; 4:e1000166.

- 628 46. Yatagai Y, Sakamoto T, Masuko H, Kaneko Y, Yamada H, Iijima H, et al. Genome-wide
629 association study for levels of total serum IgE identifies HLA-C in a Japanese population.
630 PLoS One 2013; 8:e80941.
- 631 47. Hirota T, Harada M, Sakashita M, Doi S, Miyatake A, Fujita K, et al. Genetic polymorphism
632 regulating ORM1-like 3 (*Saccharomyces cerevisiae*) expression is associated with childhood
633 atopic asthma in a Japanese population. *J Allergy Clin Immunol* 2008; 121:769-70.
- 634 48. Fuertes E, Soderhall C, Acevedo N, Becker A, Brauer M, Chan-Yeung M, et al. Associations
635 between the 17q21 region and allergic rhinitis in 5 birth cohorts. *J Allergy Clin Immunol*
636 2015; 135:573-6.
- 637 49. Tomita K, Sakashita M, Hirota T, Tanaka S, Masuyama K, Yamada T, et al. Variants in the
638 17q21 asthma susceptibility locus are associated with allergic rhinitis in the Japanese
639 population. *Allergy* 2013; 68:92-100.
- 640 50. Ramasamy A, Curjuric I, Coin LJ, Kumar A, McArdle WL, Imboden M, et al. A genome-
641 wide meta-analysis of genetic variants associated with allergic rhinitis and grass sensitization
642 and their interaction with birth order. *J Allergy Clin Immunol* 2011; 128:996-1005.
- 643 51. Joost O, Wilk JB, Cupples LA, Harmon M, Shearman AM, Baldwin CT, et al. Genetic loci
644 influencing lung function: a genome-wide scan in the Framingham Study. *Am J Respir Crit*
645 *Care Med* 2002; 165:795-9.
- 646 52. Repapi E, Sayers I, Wain LV, Burton PR, Johnson T, Obeidat M, et al. Genome-wide
647 association study identifies five loci associated with lung function. *Nat Genet* 2010; 42:36-
648 44.
- 649 53. Soler Artigas M, Loth DW, Wain LV, Gharib SA, Obeidat M, Tang W, et al. Genome-wide
650 association and large-scale follow up identifies 16 new loci influencing lung function. *Nat*
651 *Genet* 2011; 43:1082-90.
- 652 54. Tang W, Kowgier M, Loth DW, Soler Artigas M, Joubert BR, Hodge E, et al. Large-scale
653 genome-wide association studies and meta-analyses of longitudinal change in adult lung
654 function. *PLoS One* 2014; 9:e100776.
- 655 55. Yao TC, Du G, Han L, Sun Y, Hu D, Yang JJ, et al. Genome-wide association study of lung
656 function phenotypes in a founder population. *J Allergy Clin Immunol* 2014; 133:248-55 e1-
657 10.
- 658 56. van der Valk RJ, Duijts L, Timpson NJ, Salam MT, Standl M, Curtin JA, et al. Fraction of
659 exhaled nitric oxide values in childhood are associated with 17q11.2-q12 and 17q12-q21
660 variants. *J Allergy Clin Immunol* 2014; 134:46-55.

57. Bouzigon E, Nadif R, Thompson EE, Concas MP, Kuldane S, Du G, et al. A common variant in RAB27A gene is associated with fractional exhaled nitric oxide levels in adults. *Clin Exp Allergy* 2015; 45:797-806.
58. Gudbjartsson DF, Bjornsdottir US, Halapi E, Helgadottir A, Sulem P, Jonsdottir GM, et al. Sequence variants affecting eosinophil numbers associate with asthma and myocardial infarction. *Nat Genet* 2009; 41:342-7.
59. Barrett JC, Hansoul S, Nicolae DL, Cho JH, Duerr RH, Rioux JD, et al. Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease. *Nat Genet* 2008; 40:955-62.
60. Anderson CA, Boucher G, Lees CW, Franke A, D'Amato M, Taylor KD, et al. Meta-analysis identifies 29 additional ulcerative colitis risk loci, increasing the number of confirmed associations to 47. *Nat Genet* 2011; 43:246-52.
61. Huang C, Haritunians T, Okou DT, Cutler DJ, Zwick ME, Taylor KD, et al. Characterization of genetic loci that affect susceptibility to inflammatory bowel diseases in African Americans. *Gastroenterology* 2015; 149:1575-86.
62. Saleh NM, Raj SM, Smyth DJ, Wallace C, Howson JM, Bell L, et al. Genetic association analyses of atopic illness and proinflammatory cytokine genes with type 1 diabetes. *Diabetes Metab Res Rev* 2011; 27:838-43.
63. Barrett JC, Clayton DG, Concannon P, Akolkar B, Cooper JD, Erlich HA, et al. Genome-wide association study and meta-analysis find that over 40 loci affect risk of type 1 diabetes. *Nat Genet* 2009; 41:703-7.
64. Stahl EA, Raychaudhuri S, Remmers EF, Xie G, Eyre S, Thomson BP, et al. Genome-wide association study meta-analysis identifies seven new rheumatoid arthritis risk loci. *Nat Genet* 2010; 42:508-14.
65. Mells GF, Floyd JA, Morley KI, Cordell HJ, Franklin CS, Shin SY, et al. Genome-wide association study identifies 12 new susceptibility loci for primary biliary cirrhosis. *Nat Genet* 2011; 43:329-32.
66. Hirschfield GM, Liu X, Xu C, Lu Y, Xie G, Lu Y, et al. Primary biliary cirrhosis associated with HLA, IL12A, and IL12RB2 variants. *N Engl J Med* 2009; 360:2544-55.
67. Kreiner E, Waage J, Standl M, Brix S, Pers TH, Couto Alves A, et al. Shared genetic variants suggest common pathways in allergy and autoimmune diseases. *J Allergy Clin Immunol* 2017.

68. Li X, Ampleford EJ, Howard TD, Moore WC, Torgerson DG, Li H, et al. Genome-wide association studies of asthma indicate opposite immunopathogenesis direction from autoimmune diseases. *J Allergy Clin Immunol* 2012; 130:861-8 e7.
69. Sleiman PM, Annaiah K, Imielinski M, Bradfield JP, Kim CE, Frackelton EC, et al. ORMDL3 variants associated with asthma susceptibility in North Americans of European ancestry. *J Allergy Clin Immunol* 2008; 122:1225-7.
70. Galanter J, Choudhry S, Eng C, Nazario S, Rodriguez-Santana JR, Casal J, et al. ORMDL3 gene is associated with asthma in three ethnically diverse populations. *Am J Respir Crit Care Med* 2008; 177:1194-200.
71. White MJ, Risse-Adams O, Goddard P, Contreras MG, Adams J, Hu D, et al. Novel genetic risk factors for asthma in African American children: Precision Medicine and the SAGE II Study. *Immunogenetics* 2016; 68:391-400.
72. Almoguera B, Vazquez L, Mentch F, Connolly J, Pacheco JA, Sundaresan AS, et al. Identification of Four Novel Loci in Asthma in European American and African American Populations. *Am J Respir Crit Care Med* 2017; 195:456-63.
73. Tang MF, Sy HY, Kong AP, Ko FW, Wang SS, Liu TC, et al. Genetic effects of multiple asthma loci identified by genomewide association studies on asthma and spirometric indices. *Pediatr Allergy Immunol* 2016; 27:185-94.
74. Fang Q, Zhao H, Wang A, Gong Y, Liu Q. Association of genetic variants in chromosome 17q21 and adult-onset asthma in a Chinese Han population. *BMC Med Genet* 2011; 12:133.
75. Yu X, Yu C, Ren Z, Deng Y, Song J, Zhang H, et al. Genetic variants of 17q21 are associated with childhood-onset asthma and related phenotypes in a northeastern Han Chinese population: a case-control study. *Tissue Antigens* 2014; 83:330-6.
76. Chen J, Zhang J, Hu H, Jin Y, Xue M. Polymorphisms of RAD50, IL33 and IL1RL1 are associated with atopic asthma in Chinese population. *Tissue Antigens* 2015; 86:443-7.
77. Li FX, Tan JY, Yang XX, Wu YS, Wu D, Li M. Genetic variants on 17q21 are associated with asthma in a Han Chinese population. *Genet Mol Res* 2012; 11:340-7.
78. Yang FF, Huang Y, Li QB, Dai JH, Fu Z. Single nucleotide polymorphisms in the ORM1-like 3 gene associated with childhood asthma in a Chinese population. *Genet Mol Res* 2012; 11:4646-53.

- 723 79. Kang MJ, Yu HS, Seo JH, Kim HY, Jung YH, Kim YJ, et al. GSDMB/ORMDL3 variants
724 contribute to asthma susceptibility and eosinophil-mediated bronchial hyperresponsiveness.
725 Hum Immunol 2012; 73:954-9.
- 726 80. Yu J, Kang MJ, Kim BJ, Kwon JW, Song YH, Choi WA, et al. Polymorphisms in GSDMA
727 and GSDMB are associated with asthma susceptibility, atopy and BHR. Pediatr Pulmonol
728 2011; 46:701-8.
- 729 81. Noguchi E, Sakamoto H, Hirota T, Ochiai K, Imoto Y, Sakashita M, et al. Genome-wide
730 association study identifies HLA-DP as a susceptibility gene for pediatric asthma in Asian
731 populations. PLoS Genet 2011; 7:e1002170.
- 732 82. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, et al. A method
733 and server for predicting damaging missense mutations. Nat Methods 2010; 7:248-9.
- 734 83. Igartua C, Myers RA, Mathias RA, Pino-Yanes M, Eng C, Graves PE, et al. Ethnic-specific
735 associations of rare and low-frequency DNA sequence variants with asthma. Nat Commun
736 2015; 6:5965.
- 737 84. Morrison FS, Locke JM, Wood AR, Tuke M, Pasko D, Murray A, et al. The splice site
738 variant rs11078928 may be associated with a genotype-dependent alteration in expression of
739 GSDMB transcripts. BMC Genomics 2013; 14:627.
- 740 85. Mathias RA, Taub MA, Gignoux CR, Fu W, Musharoff S, O'Connor TD, et al. A continuum
741 of admixture in the Western Hemisphere revealed by the African Diaspora genome. Nat
742 Commun 2016; 7:12522.
- 743 86. Gern JE, Pappas T, Visness CM, Jaffee KF, Lemanske RF, Togias A, et al. Comparison of
744 the etiology of viral respiratory illnesses in inner-city and suburban infants. J Infect Dis
745 2012; 206:1342-9.
- 746 87. Consortium G. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis:
747 multitissue gene regulation in humans. Science 2015; 348:648-60.
- 748 88. Çalışkan M, Baker SW, Gilad Y, Ober C. Host genetic variation influences gene expression
749 response to rhinovirus infection. PLoS Genet 2015; 11:e1005111.
- 750 89. Schedel M, Michel S, Gaertner VD, Toncheva AA, Depner M, Binia A, et al. Polymorphisms
751 related to ORMDL3 are associated with asthma susceptibility, alterations in transcriptional
752 regulation of ORMDL3, and changes in TH2 cytokine levels. J Allergy Clin Immunol 2015;
753 136:893-903 e14.

90. Acevedo N, Reinius LE, Greco D, Gref A, Orsmark-Pietras C, Persson H, et al. Risk of childhood asthma is associated with CpG-site polymorphisms, regional DNA methylation and mRNA levels at the GSDMB/ORMDL3 locus. *Hum Mol Genet* 2015; 24:875-90.
91. Wang XH, Shu J, Jiang CM, Zhuang LL, Yang WX, Zhang HW, et al. Mechanisms and roles by which IRF-3 mediates the regulation of ORMDL3 transcription in respiratory syncytial virus infection. *Int J Biochem Cell Biol* 2017; 87:8-17.
92. Nieuwenhuis MA, Siedlinski M, van den Berge M, Granell R, Li X, Niens M, et al. Combining genomewide association study and lung eQTL analysis provides evidence for novel genes associated with asthma. *Allergy* 2016; 71:1712-20.
93. Lappalainen T, Sammeth M, Friedlander MR, t Hoen PA, Monlong J, Rivas MA, et al. Transcriptome and genome sequencing uncovers functional variation in humans. *Nature* 2013; 501:506-11.
94. Woodruff PG, Modrek B, Choy DF, Jia G, Abbas AR, Ellwanger A, et al. T-helper type 2-driven inflammation defines major subphenotypes of asthma. *Am J Respir Crit Care Med* 2009; 180:388-95.
95. Bhakta NR, Christenson SA, Nerella S, Solberg OD, Nguyen CP, Choy DF, et al. Interferon-stimulated Gene Expression, Type-2 Inflammation and Endoplasmic Reticulum Stress in Asthma. *Am J Respir Crit Care Med* 2017.
96. Das S, Miller M, Beppu AK, Mueller J, McGeough MD, Vuong C, et al. GSDMB induces an asthma phenotype characterized by increased airway responsiveness and remodeling without lung inflammation. *Proc Natl Acad Sci U S A* 2016; 113:13132-7.
97. Shukla S, Kavak E, Gregory M, Imashimizu M, Shutinoski B, Kashlev M, et al. CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. *Nature* 2011; 479:74-9.
98. Berlivet S, Moussette S, Ouimet M, Verlaan DJ, Koka V, Al Tuwaijri A, et al. Interaction between genetic and epigenetic variation defines gene expression patterns at the asthma-associated locus 17q12-q21 in lymphoblastoid cell lines. *Hum Genet* 2012; 131:1161-71.
99. <https://www.ncbi.nlm.nih.gov/gene/93210>.
100. Breslow DK, Collins SR, Bodenmiller B, Aebersold R, Simons K, Shevchenko A, et al. Orm family proteins mediate sphingolipid homeostasis. *Nature* 2010; 463:1048-53.

101. Miller M, Rosenthal P, Beppu A, Gordillo R, Broide DH. Oroscomucoid like protein 3 (ORMDL3) transgenic mice have reduced levels of sphingolipids including sphingosine-1-phosphate and ceramide. *J Allergy Clin Immunol* 2017; 139:1373-6 e4.
102. Miller M, Tam AB, Mueller JL, Rosenthal P, Beppu A, Gordillo R, et al. Cutting Edge: Targeting Epithelial ORMDL3 Increases, Rather than Reduces, Airway Responsiveness and Is Associated with Increased Sphingosine-1-Phosphate. *J Immunol* 2017; 198:3017-22.
103. Oyeniran C, Sturgill JL, Hait NC, Huang WC, Avni D, Maceyka M, et al. Aberrant ORM (yeast)-like protein isoform 3 (ORMDL3) expression dysregulates ceramide homeostasis in cells and ceramide exacerbates allergic asthma in mice. *J Allergy Clin Immunol* 2015; 136:1035-46 e6.
104. Yu F, Sun Y, Yu J, Ding Z, Wang J, Zhang L, et al. ORMDL3 is associated with airway remodeling in asthma via the ERK/MMP-9 pathway. *Mol Med Rep* 2017; 15:2969-76.
105. Zhakupova A, Debeuf N, Krols M, Toussaint W, Vanhoutte L, Alecu I, et al. ORMDL3 expression levels have no influence on the activity of serine palmitoyltransferase. *FASEB J* 2016; 30:4289-300.
106. Cantero-Recasens G, Fandos C, Rubio-Moscardo F, Valverde MA, Vicente R. The asthma-associated ORMDL3 gene product regulates endoplasmic reticulum-mediated calcium signaling and cellular stress. *Hum Mol Genet* 2010; 19:111-21.
107. Li M, Zhang S, Qiu Y, He Y, Chen B, Mao R, et al. Upregulation of miR-665 promotes apoptosis and colitis in inflammatory bowel disease by repressing the endoplasmic reticulum stress components XBP1 and ORMDL3. *Cell Death Dis* 2017; 8:e2699.
108. Miller M, Rosenthal P, Beppu A, Mueller JL, Hoffman HM, Tam AB, et al. ORMDL3 transgenic mice have increased airway remodeling and airway responsiveness characteristic of asthma. *J Immunol* 2014; 192:3475-87.
109. Miller M, Tam AB, Cho JY, Doherty TA, Pham A, Khorram N, et al. ORMDL3 is an inducible lung epithelial gene regulating metalloproteases, chemokines, OAS, and ATF6. *Proc Natl Acad Sci U S A* 2012; 109:16648-53.
110. Ha SG, Ge XN, Bahaie NS, Kang BN, Rao A, Rao SP, et al. ORMDL3 promotes eosinophil trafficking and activation via regulation of integrins and CD48. *Nat Commun* 2013; 4:2479.
111. Loser S, Gregory LG, Zhang Y, Schaefer K, Walker SA, Buckley J, et al. Pulmonary ORMDL3 is critical for induction of *Alternaria*-induced allergic airways disease. *J Allergy Clin Immunol* 2017; 139:1496-507 e3.

- 817 112. TGP Consortium, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, et al. A
818 global reference for human genetic variation. *Nature* 2015; 526:68-74.
819

Table 1. Published eQTL studies of 17q12-21 SNPs and genes in blood and lung cells.

Superscripts after reference designates whether gene expression studies were performed by qPCR^a, microarray^b, or RNA-seq³. *P*-values are shown as they are reported in each paper for the genes included in each study; n.s., not significant.

Figure Legends.

Figure 1. The 17q12-21 asthma locus. A) Regional association plot of 17q SNPs with asthma in the TAGC Consortium (children + adults). The lead SNP, rs2952156 (purple diamond), is in *ERBB2* in the proximal 17q region. All other SNPs are colored based on their LD with the lead SNP (see insert). Modified from Demenais *et al.*² with permission. **B)** Location of genes at the 17q locus. The six genes highlighted in this review are shown in blue, orange and green to correspond to the proximal, core, and distal regions, respectively. **C)** Frequencies of the asthma-associated alleles at 17 SNPs in 1000Genome reference panels: African Americans (ASW [American of African Ancestry in SW USA] and ACB [African Caribbeans in Barbados]), Europeans (CEU [Utah Residents with Northern and Western European Ancestry], GBR [British in England and Scotland], TSI [Toscani in Italy]), Latinos (MXL [Mexican Ancestry from Los Angeles USA] and PUR [Puerto Ricans from Puerto Rico]), East Asians (CHB [Han Chinese in Beijing], JPT [Japanese in Tokyo], CHS [Southern Han Chinese], CDX [Chinese Dai in Xishuangbanna], KHV [Kinh in HoChi Minh City]). **D)** LD (r^2) among the 17 SNPs described in Figure 2 in African Americans (upper panel), Europeans (lower left panel), Latinos (lower middle panel), and East Asians (lower right panel), using the same 1000Genome reference panels as in panel C. Asthma-associated allele shown for each SNP; *denotes that the associated allele is ancestral. Data for panels C and D from <ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20130502/>.

Figure 2. Seventeen SNPs in the extended 17q12-21 region reported to be associated with asthma, gene expression, or epigenetic modification. Base pair (bp) position from build hg19; SNP type shown as ancestral>derived alleles.

Figure 3. SNPs in the 17q core region are associated with early life onset and risk or protection from wheezing illness. A) Ordered-subset regression analysis of asthma was performed for 11 SNPs that were associated with asthma in the Epidemiological Study on Genetics and Environment of Asthma (EGEA) sample (n=1,511). The maximal difference in likelihood-ratio test statistics for association between the ordered age-specific subsets and the total sample was for an onset of asthma of 4 years or younger. SNP rs9303277 showed the largest difference in the likelihood-ratio test statistics. The EGEA study included children from 388 nuclear families with at least one asthmatic member living in one of five French cities. Modified from Bouzigon *et al.*²⁸ with permission. **B)** Risk for asthma based on rs7216389 genotype is confined to children who wheezed in the first 3 years of life (COAST, n=200 and COPSAC, n=297)¹² or in the first year of life (PASTURE, n=696)³². Dashed horizontal lines show the overall prevalence of asthma at age 6 in each sample. COAST, COPSAC and PASTURE are birth cohort studies in Madison, WI, Copenhagen, DK, and rural areas of Europe, respectively. The children in COAST had at least one parent with asthma or allergies, the children in COPSAC had a mother with asthma, and the children in PASTURE were unselected with respect to asthma or allergy. Modified from Bønnelykke and Ober³⁸. **C)** In the PASTURE study, children with the asthma associated allele rs8076131-A allele and are exposed to animal barns have significantly fewer wheezing illnesses in the first year of life. From Loss *et al.*³² with permission.

Figure 4. Relative expression of 17q genes in different tissues and cells. A) Peripheral blood leukocytes (PBLs, n=112), **B)** lung CD4+ tissue resident memory cells (TRMs, n=18), **C)**

bronchial epithelial cells (BECs, $n=85$)²⁰, **D**) airway smooth muscle cells (ASMCs, $n=67$). The expression levels of *ORMDL3* and *GSDMB* were most correlated in PBLs (*ORMDL3* vs. *PGAP3* $r=0.410$, $P=7.3 \times 10^{-6}$ cf. *ORMDL3* vs. *GSDMB* $r=0.369$, $P=6.2 \times 10^{-5}$), BECs (*ORMDL3* vs. *PGAP3* $r=0.693$, $P=1.9 \times 10^{-13}$ cf. *ORMDL3* vs. *GSDMB* $r=0.320$, $P=2.8 \times 10^{-3}$), and ASMCs (*ORMDL3* vs. *PGAP3* $r=0.450$, $P=1.3 \times 10^{-4}$ cf. *ORMDL3* vs. *GSDMB* $r=0.221$, $P=0.07$). *ORMDL3* expression was not correlated with the expression of any 17q genes in lung CD4+ TRMs (*ORMDL3* v. *PGAP3* $r=0.077$, $P=0.76$ cf. *ORMDL3* vs. *GSDMB* $r=0.400$, $P=0.10$). Gene expression levels in panels A-C are from RNA-seq; counts are shown as relative expression normalized to all genes detected as expressed within each cell type. Gene expression in panel D is based on microarrays; expression levels were normalized to have mean of zero and scaled between -5 and 10. Methods for PBLs, CD4+ TRMs, and ASMCs are described in the Supplement; methods for BECs are reported in Nicodemus-Johnson *et al.*²⁰.

Figure 5. Expression of *ORMDL3* and *PGAP3* transcripts in BECs from 85 adults.

Correlation between *ORMDL3* (x-axis) and *PGAP3* (y-axis) in 85 individuals ($r=0.693$, $P=1.9 \times 10^{-13}$, as in Fig. 4). Expression levels are shown as read counts adjusted for unwanted variation, as described in Nicodemus-Johnson *et al.*²⁰.

Figure 6. GTEx eQTLs for 17q genes in whole blood and lung tissue at an FDR of 5%. 17q

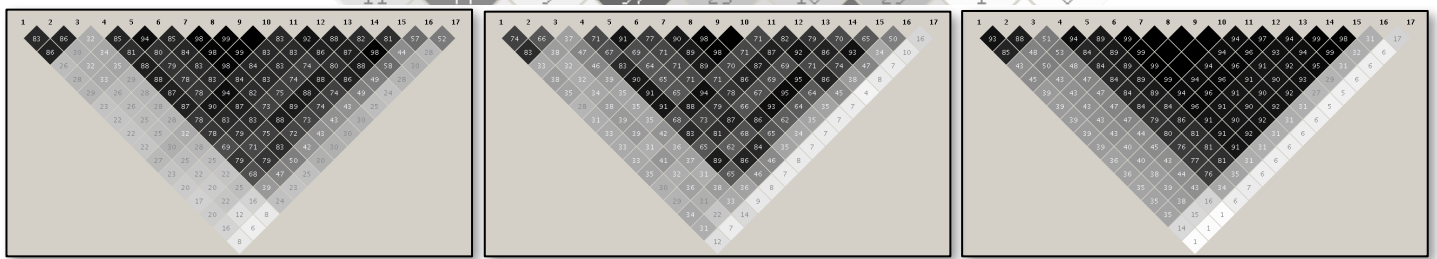
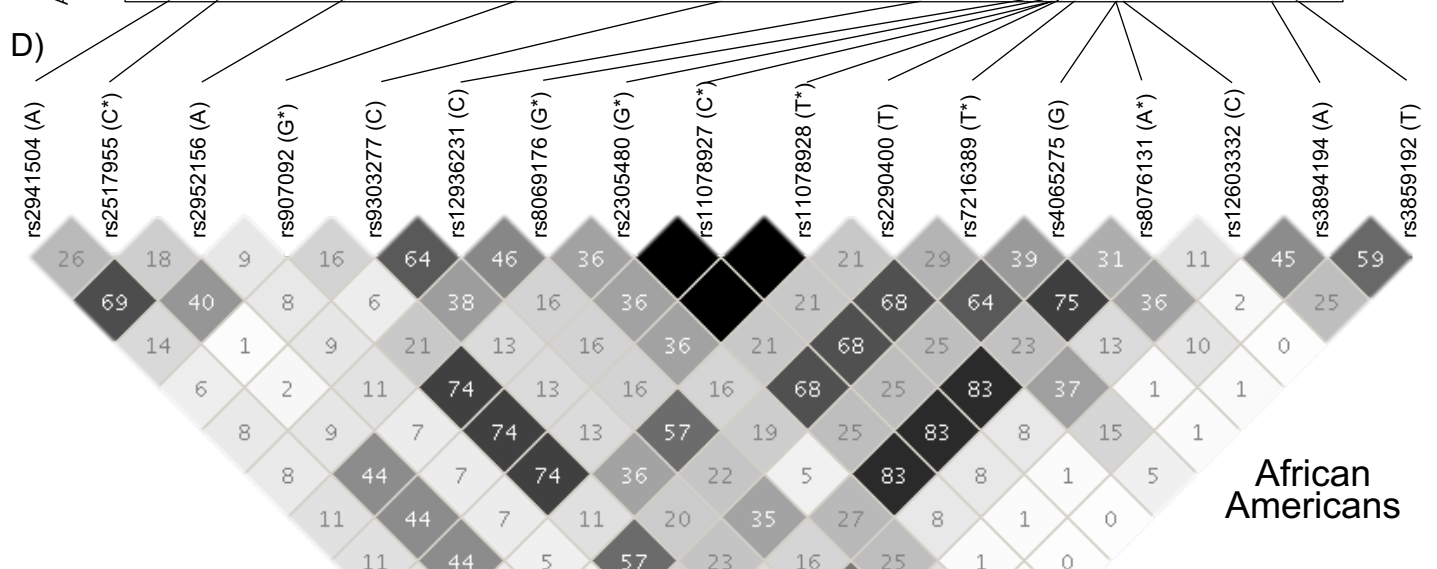
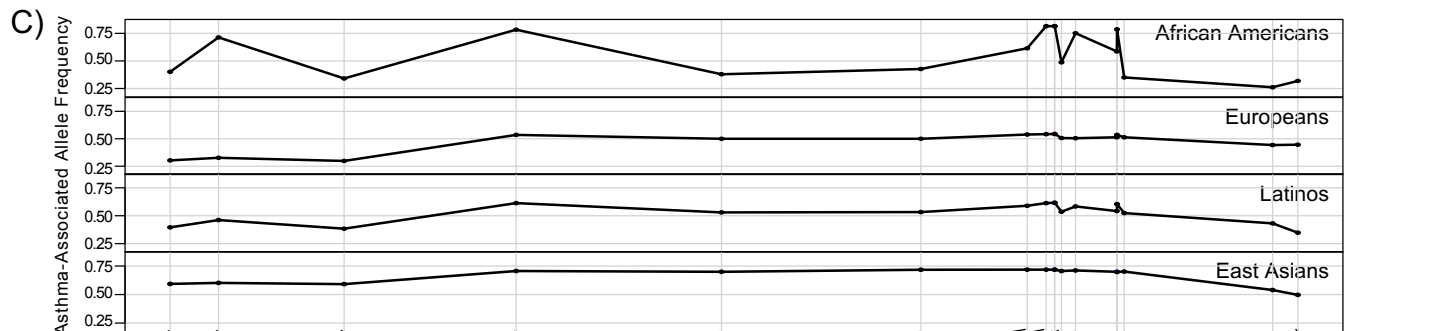
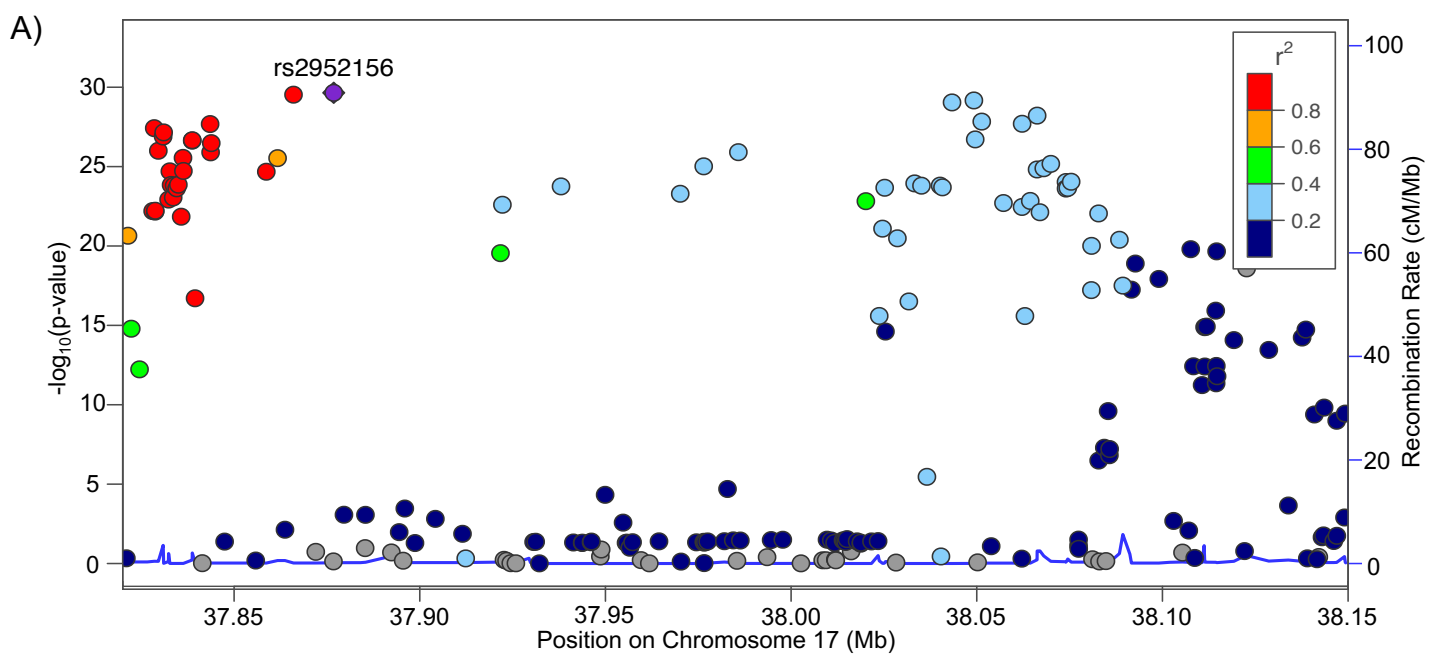
SNPs were eQTLs for only two of the six genes in whole blood and four of the six genes in lung tissue; the remaining genes (without significant eQTLs) are not shown. SNPs in the proximal region are shown on a blue background, SNPs in the core region are shown on an orange background, and SNPs in the distal region are shown on a green background (see **Fig. 1**). The

GTEx sample composition is approximately 84% white, 14% African American, 1% other, 1% unknown (<https://gtexportal.org/home/tissueSummaryPage>). n.s., not significant

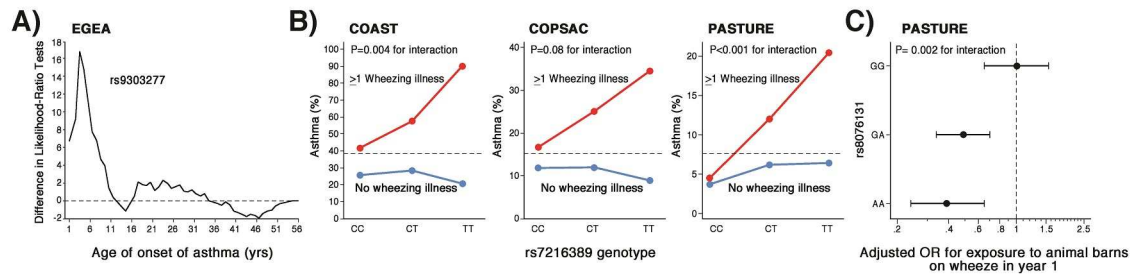
Figure 7. eQTLs for *ORMDL3* and *GSDMB* in Nigerian (YRI) LCLs⁹³ and associations with asthma in African Americans⁵. LD between each SNP and the lead eQTL (rs12936231) are shown in the third column. eQTLs and GWAS *P*-values ≤ 0.01 are shown in bold font. The SNPs in the proximal region are shown on a blue background, SNPs in the core region are shown on a orange background, and SNPs in the distal region are shown on a green background (see **Fig. 1**). n.i., no information available (SNP not imputed).

Figure 8. Allele-specific chromatin modification and gene expression at the 17q locus. A) Schematic representation of the CTCF-binding motifs that overlap with the asthma associated SNPs rs12936231 in *ZBP2* and rs4065275 in *ORMDL3* (modified from Schmiedel *et al.*²¹). Relative expression of *ORMDL3* from the C-A haplotype was inferred from on YRI individual who was heterozygous C-A/G-A (*ORMDL3* FPKM = 26; see panel B). Predicted haplotype frequencies are based on allele frequencies and LD estimates (see Fig. 1). **B)** *ORMDL3* and *GSDMB* gene expression in LCLs and phased genotype data for 29 YRI¹¹² who are homozygous for three of the rs12936231-rs4065274 haplotypes; no individuals were homozygous for the C-A haplotype. The y-axis shows FPKM (fragments per kb million), a measure of gene expression after normalizing for sequence depth and gene length.

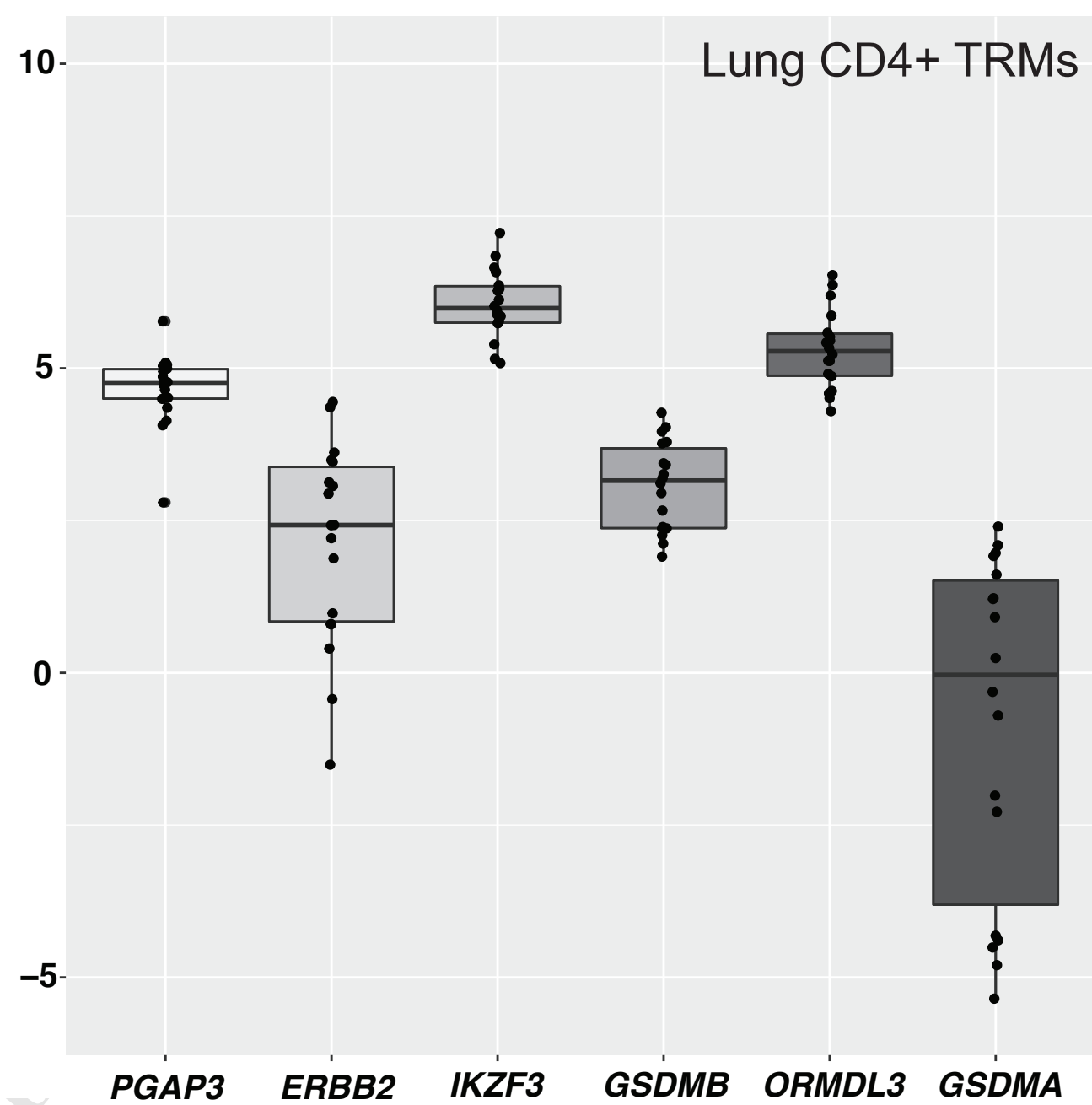
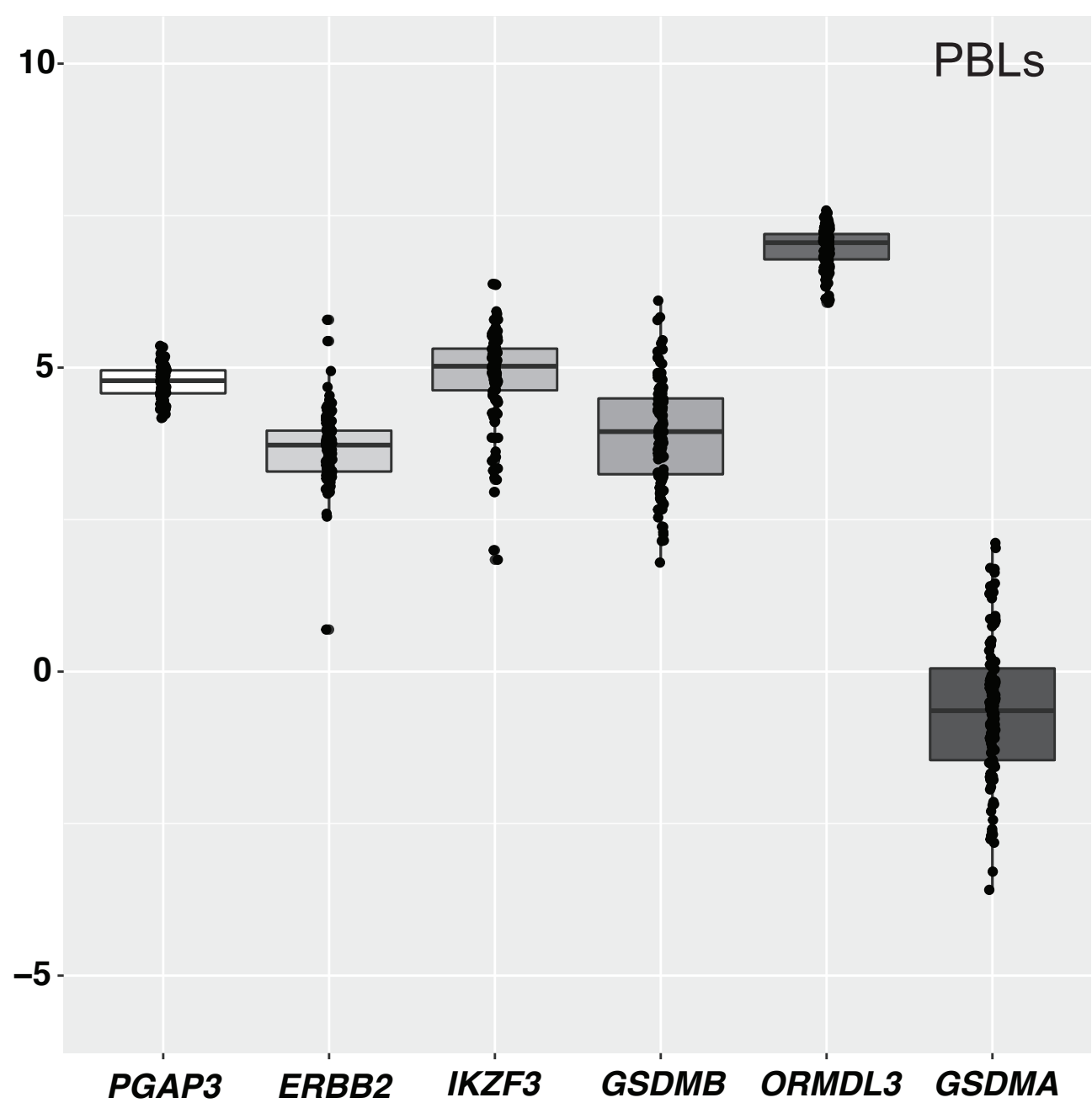
Study	SNP	17q12-21 Genes						
		PGAP3 P-value	ERBB2 P-value	IKZF3 P-value	ZPBP2 P-value	GSDMB P-value	ORMDL3 P-value	GSDMA P-value
Blood Cells								
Schmiedel ²¹ (n=34) Sorted cells ^a	rs7216389					Naïve CD4+ T: <0.05 Tregs: <0.05 Th1: <0.05 B: <0.01 NK: <0.01	Naïve CD4+ T: <0.01 Tregs: <0.01 Th1: <0.001 Th2: <0.001 Th17: <0.001 B: <0.001 NK: <0.01	
Liu ¹⁷ (n=19) Sorted cells ^a	rs7216389						PBMC: ≤0.03 PBMC, RV inf: ≤0.03 CD19+ B: 0.004 CD19+ B, RV inf: <0.001 CD8+ T, RV inf: ≤0.03 CD19-CD8-, RV inf: ≤0.03	
Moffat ¹ (n=378) LCLs ^b	rs7216389 rs93003277 rs2290400 rs3894194 rs3859192						3.9x10 ⁻²³ 1.2x10 ⁻²² 3.9x10 ⁻²³ 1.0x10 ⁻¹¹ 3.1x10 ⁻⁴	
Verlaan ²³ (n=53) LCLs ^a	rs907091/ rs12603332 haplotype				0.0037	2.7x10 ⁻⁵	3.8x10 ⁻³	
Dixon ¹³ (n=308) LCLs ^b	rs2941504 rs2517955 rs907092 rs9303277 rs2305480 rs2290440 rs2716389 rs3894194 rs3859192					3x10 ⁻⁴	2.6x10 ⁻⁸ 7.5x10 ⁻⁸ 2.0x10 ⁻²² 1.2x10 ⁻²² 4.9x10 ⁻²¹ 3.7x10 ⁻²³ 4.2x10 ⁻²³ 1.1x10 ⁻¹¹ 3.0x10 ⁻⁴	
Andiappan ¹¹ (n=71) whole blood ^b	rs8076131			n.s.	0.023	0.026	0.0051	
Çalışkan ¹² (n=100) PBMCs ^a	rs7216389			Untreated: 0.11 RV inf: 0.16		Untreated: <0.001 RV inf: <0.001	Untreated: 0.07 RV inf: 0.03	
Lluis ¹⁸ (n=200) CBMCs ^a	rs7216389				n.s.	n.s.	Der p1 stim: 0.01	PHA stim: 0.0001 Der p1 stim: 0.05
Halapi ¹⁴ (n=473) WBCs ^b	rs2941504 rs907092 rs9303277 rs2305480 rs2290400 rs7216389					1.2x10 ⁻¹¹ 5.1x10 ⁻³⁶ 1.1x10 ⁻⁴¹ 5.5x10 ⁻³⁷ 1.3x10 ⁻⁴¹ 2.3 x10 ⁻³⁸	3.6x10 ⁻²¹ 5.1x10 ⁻⁶³ 8.3x10 ⁻⁶⁸ 9.1x10 ⁻⁶⁵ 1.2x10 ⁻⁶⁰ 8.8x10 ⁻⁵⁸	
Sharma ²² (n=200) CD4+T cells ^b	rs4795405						3.0x10 ⁻¹⁰	
Murphy ¹⁹ (n=200) CD4+T cells ^b	rs2290400 rs7216389	n.s. n.s.	n.s. n.s.	n.s. n.s.			3.1x10 ⁻⁹ 1.6 x10 ⁻⁸	n.s. n.s.
Lung Cells								
Nicodemus ²⁰ (n=81) BECs ^c	rs2517944 rs8076131 rs7216389 rs2305480 rs11078928 rs2941504 rs2941504 rs4065275	2.1x10 ⁻⁸ 8.0x10 ⁻⁵ 7.1x10 ⁻⁵ n.s. n.s. n.s. 4.7x10 ⁻⁷ 2.1x10 ⁻⁵	n.s. n.s. n.s. n.s. n.s. n.s. n.s. n.s.	n.s. n.s. n.s. n.s. n.s. n.s. n.s. n.s.		n.s. n.s. n.s. n.s. n.s. n.s. n.s. n.s.	2.6x10 ⁻⁵ 3.2x10 ⁻⁵ 8.0x10 ⁻⁵ 1.1x10 ⁻⁴ 1.2x10 ⁻⁴ 1.3x10 ⁻⁴ 4.0x10 ⁻⁵ 5.5x10 ⁻⁵	n.s. n.s. n.s. n.s. n.s. n.s. n.s. n.s.
Hao ¹⁵ (n=1,111) Lung tissue ^b	rs2290400 rs7216389 rs2305480 rs3859192 rs3894194	n.s. n.s. n.s. n.s. n.s.	n.s. n.s. n.s. n.s. n.s.	n.s. n.s. n.s. n.s. n.s.		4.0x10 ⁻¹⁵ 4.2x10 ^{-4/14} 2.5x10 ⁻¹² 3.9x10 ⁻⁵ 1.6x10 ⁻⁷	n.s. 1.3x10 ⁻⁶ 2.1x10 ⁻⁵ ns ns	n.s. 8.8x10 ⁻²⁵ 7.4x10 ⁻³² 3.6x10 ⁻¹⁵¹ 2.8x10 ⁻⁷³
Li ¹⁶ (n=107)BEC ^b (n=94) BAL ^b	rs8067378 rs2305480 rs7216389					BEC: 1.9x10 ⁻³ ; BAL: 0.04 BEC: 4.7x10 ⁻³ ; BAL: n.s. BEC: 1.1x10 ⁻³ ; BAL: ns	BEC: n.s.; BAL: n.s. BEC: n.s.; BAL: n.s. BEC: n.s.; BAL: n.s.	



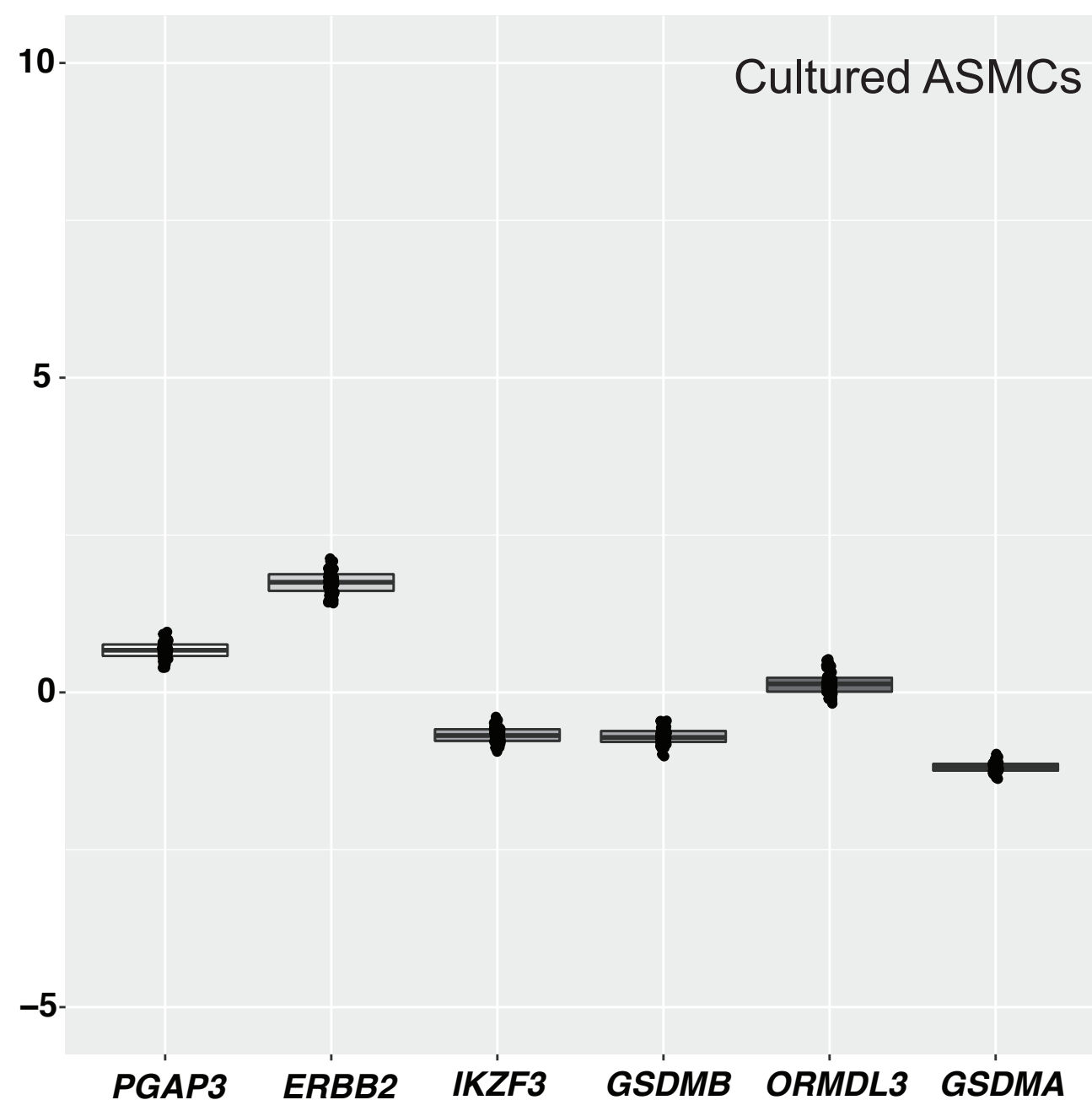
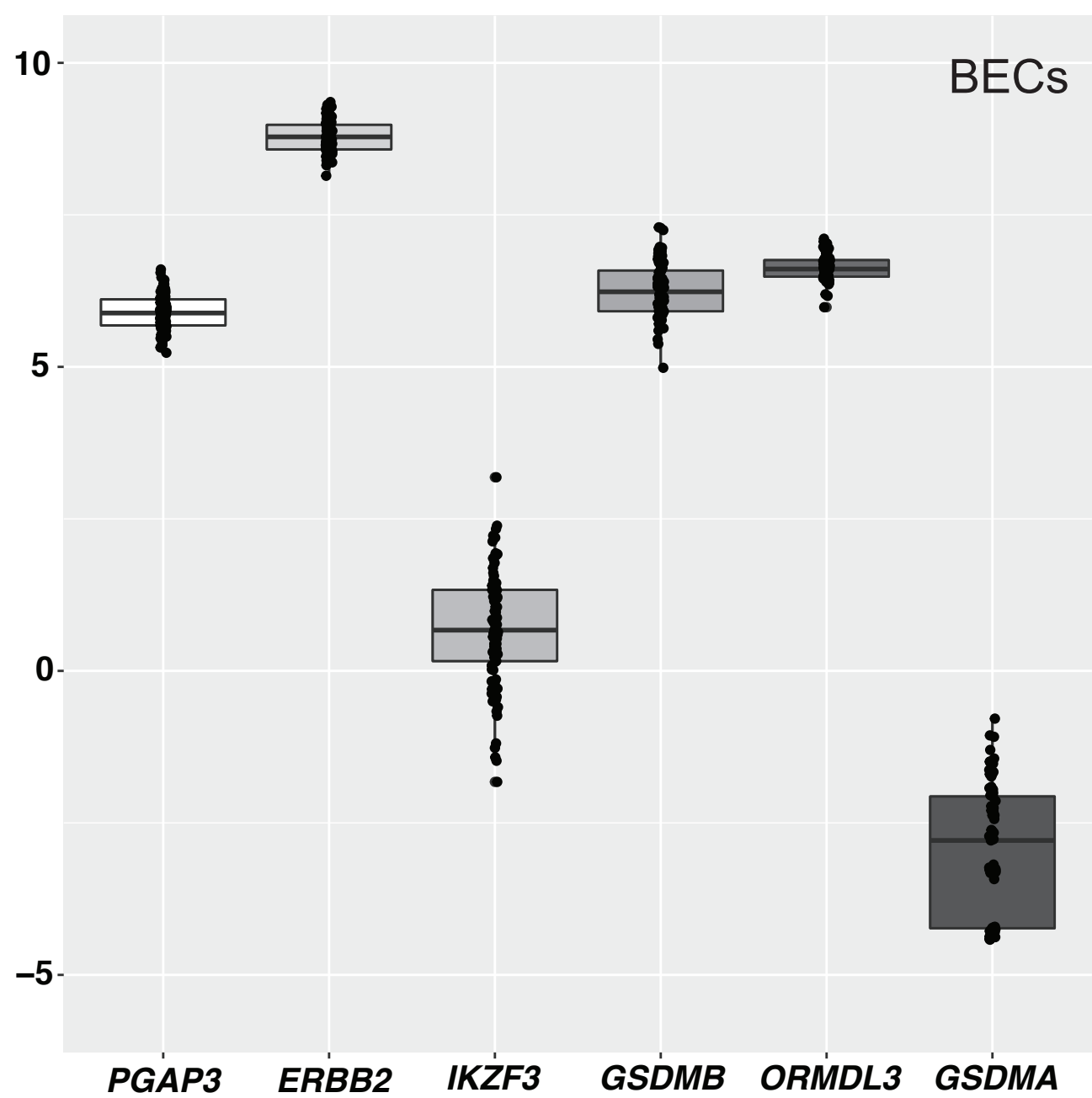
Location	SNP	bp position	SNP Type	Associations in published studies
<i>PGAP3</i>	rs2941504	17:37,830,900	Synonymous (c.465G>A, p.Val104)	Associated with allergic asthma in ethnic Chinese from Singapore ²⁷ and asthma in Icelanders plus four European populations ¹² ; eQTL for <i>PGAP3</i> in white blood cells ¹²
<i>PGAP2</i>	rs2517955	17:37,843,681	Intron (C>T)	eQTL for <i>ORMDL3</i> and meQTL for a nearby CpG site in airway epithelial cells ²⁶
<i>ERBB2</i>	rs2952156	17:37,876,835	Intron (G>A)	Lead SNP in the TAGC GWAS ²
<i>IKZF3</i>	rs907092	17:37,922,259	Synonymous (c.1314G>A; p.Ser438)	Lead SNP in asthma GWAS in Puerto Ricans ¹⁰ and in the EVE study of exome SNPs in Latinos ⁸³
<i>IKZF3</i>	rs9303277	17: 37,976,469	Intron (T>C)	Lead SNP with childhood onset asthma ³⁰
<i>ZBPB2</i>	rs12936231	17: 38,029,120	Intron (T>C)	Lead functional candidate for gene expression and chromatin effects ¹¹ , C disrupts a CTCF-binding motif, which insulates upstream <i>cis</i> -regulatory elements from interacting with the <i>ORMDL3</i> promoter in immune cells ¹⁹
Intergenic	rs8069176	17: 38,057,197	(G>A)	Lead SNP for early onset asthma among children exposed to ETS in early life ³⁰ and asthma in children in the TAGC GWAS ²
<i>GSDMB</i>	rs2305480	17:38,062,196	Missense (c.892G>A, p.Pro298Ser)	Lead SNP in the GABRIEL GWAS ³ and the Exacerbation GWAS ⁶ ; lung eQTL for <i>ORMDL3</i> , <i>GSDMB</i> , <i>GSDMA</i> ⁸⁷
<i>GSDMB</i>	rs11078927	17:38,064,405	Intron (C>T)	Lead SNP in EVE GWAS ⁵
<i>GSDMB</i>	rs11078928	17:38,064,469	Splice variant (c.662T>C)	C allele results in skipping of exon 6 and reduced levels of <i>GSDMB</i> transcript ⁸⁴
<i>GSDMB</i>	rs2290400	17:38,066,240	Intron (T>C)	Interactive with smoking on asthma risk ²⁵
<i>GSDMB</i>	rs7216389	17:38,069,949	Intron (T>C)	Lead SNP in first GWAS and eQTL for <i>ORMDL3</i> and <i>GSDMB</i> in LCLs ¹ ; associated with early onset asthma ¹² ; associated with RV-wheezing interaction in early life and eQTL for <i>ORMDL3</i> and <i>GSDMB</i> in whole blood ⁸⁹ ; lung eQTL for <i>ORMDL3</i> , <i>GSDMB</i> and <i>GSDMA</i> ⁸⁷
<i>ORMDL3</i>	rs4065275	17: 38,080,865	Intron (A>G)	G allele creates a CTCF-binding motif associated with increased expression of <i>ORMDL3</i> ¹⁹
<i>ORMDL3</i>	rs8076131	17:38,080,912	Promoter (A>G)	Strongest association with protective farming effect ³⁴
<i>ORMDL3</i>	rs12603332	17:38,082,807	5'UTR (T>C)	meQTL and eQTL for <i>ORMDL3</i> and <i>GSDMB</i> in blood leukocytes ²⁴ ; lead SNP for asthma in Mexican Americans and African Americans among 7 SNPs tested ⁷⁰
<i>GSDMA</i>	rs3894194	17: 38,121,993	Missense (c.536G>A, p.Ala18Gln)	Independent GWAS signal for asthma ^{3, 25}
<i>GSDMA</i>	rs3859192	17: 38,128,648	Intron (C>T)	eQTL for <i>GSDMA</i> in lung tissue ⁸⁷

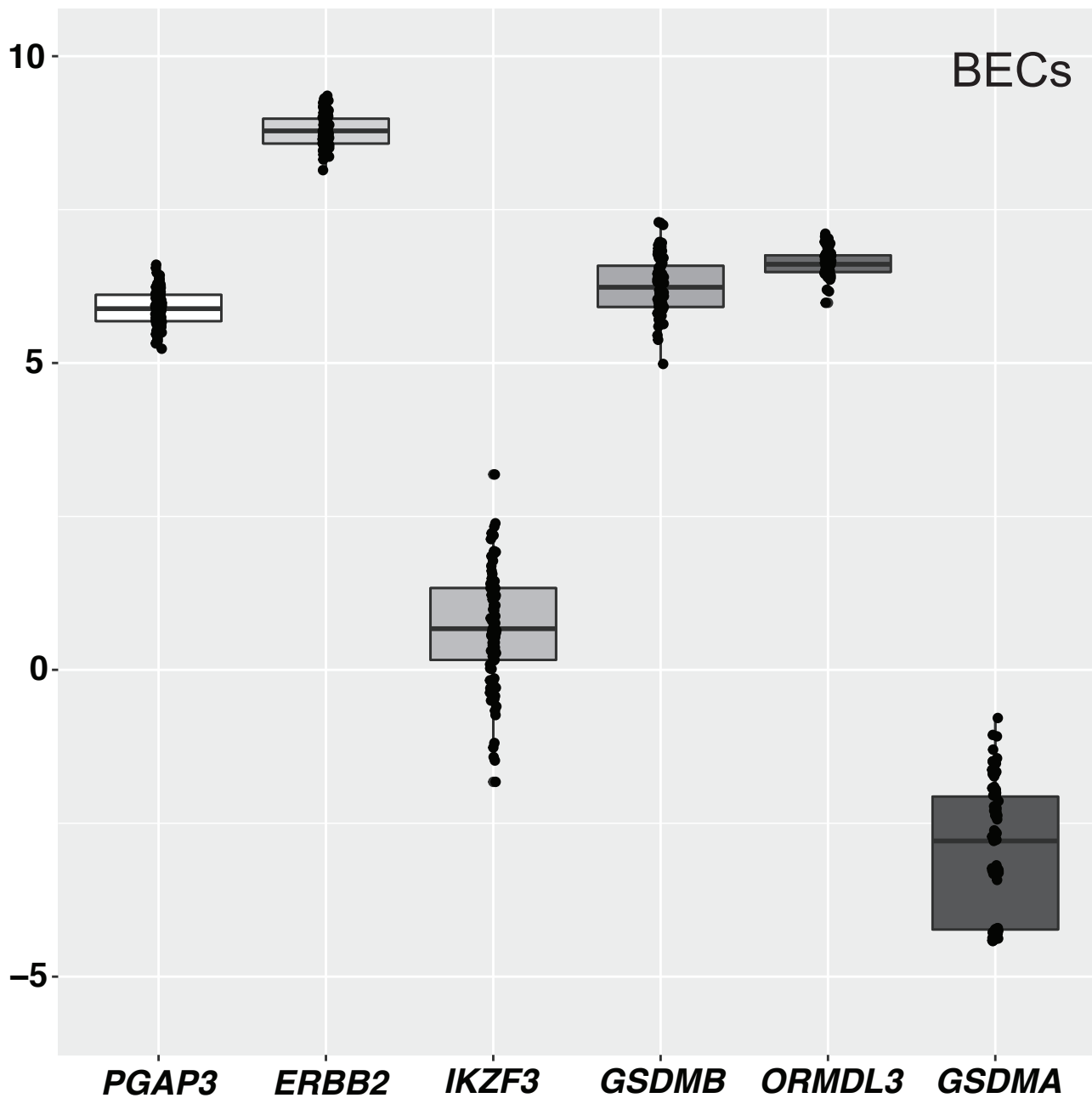


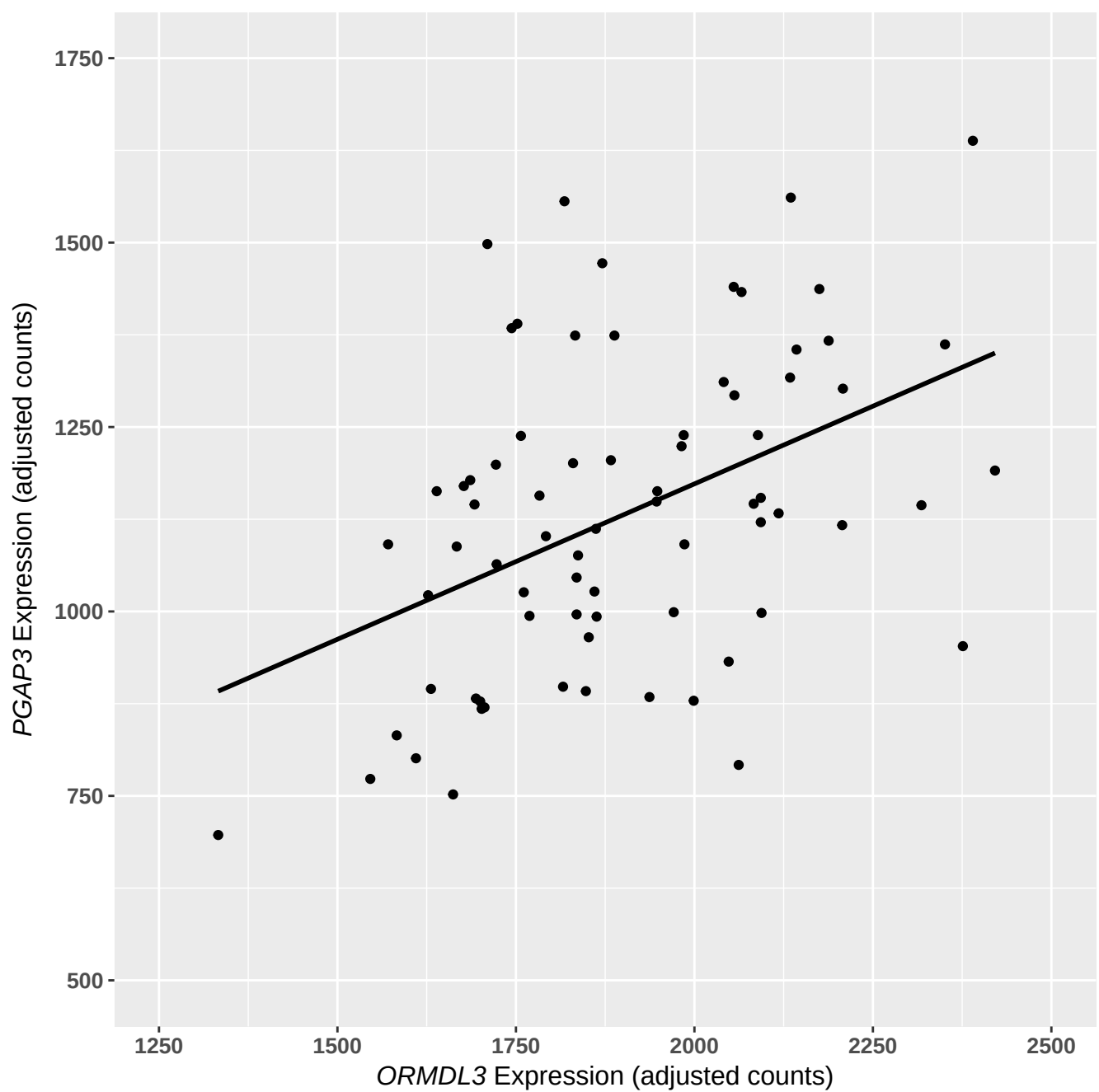
Relative Expression



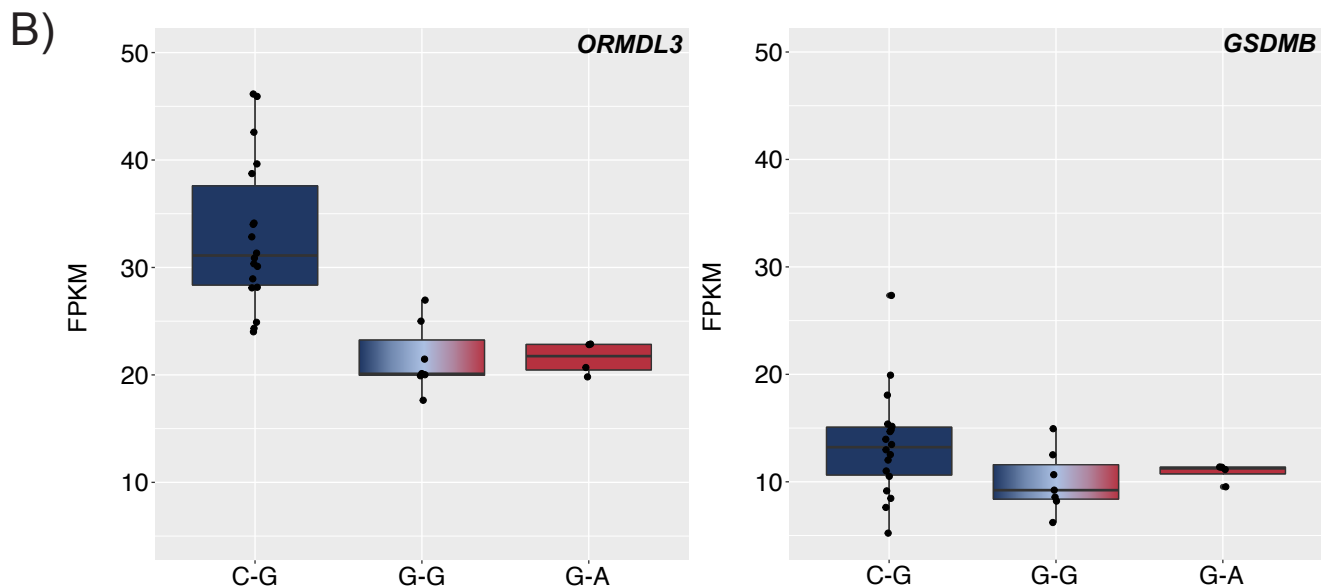
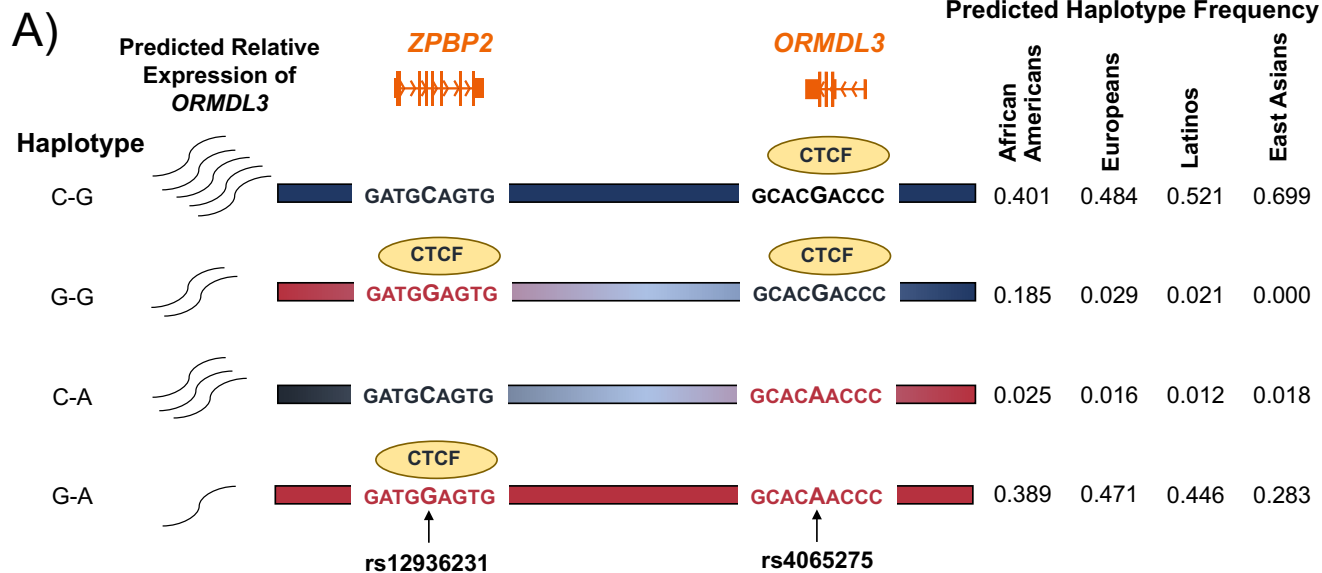
Relative Expression







SNP rsID	Asthma Risk Allele	LD r^2 with rs12936231 in YRI	Asthma Risk Allele Frequency in YRI	eQTLs in YRI LCLs (n=56)		EVE Meta-Analysis of Asthma GWAS in African Americans	
				<i>GSDMB</i> P-value	<i>ORMDL3</i> P-value	Meta P-value	OR
rs2941504	A	0.09	0.49	0.59	1.3E-03	0.196	1.05
rs2517955	C	0.00	0.89	0.40	0.77	0.144	1.06
rs2952156	A	0.09	0.52	0.57	0.01	0.607	1.02
rs907092	G	0.02	0.97	0.10	0.69	0.022	1.14
rs9303277	C	0.73	0.56	0.01	5.48E-07	0.115	1.06
rs12936231	C	1.0	0.54	0.03	5.35E-10	0.348	1.06
rs8069176	G	0.39	0.67	0.11	2.43E-04	0.064	1.11
rs2305480	G	0.07	0.96	0.29	0.31	0.009	1.17
rs11078927	C	0.07	0.96	0.34	0.21	0.001	1.23
rs11078928	T	0.07	0.96	0.26	0.32	n.i.	n.i.
rs2290400	T	0.50	0.61	0.04	1.96E-04	0.174	1.03
rs7216389	T	0.16	0.89	0.29	0.06	0.097	1.06
rs4065275	G	0.28	0.75	0.16	0.01	n.i.	n.i.
rs8076131	A	0.10	0.09	0.47	0.53	0.003	1.17
rs12603332	C	0.43	0.56	0.24	2.52E-03	0.089	1.05
rs3894194	A	0.17	0.30	0.50	0.09	0.241	1.05
rs3859192	T	0.10	0.32	0.77	0.03	0.433	1.03



A Decade of Research on the 17q12-21 Asthma Locus: Piecing Together the Puzzle

Michelle M. Stein, PhD^{1,+}, Emma E. Thompson, PhD^{1,+}, Nathan Schoettler, MD, PhD^{1,2,+},
Britney A. Helling, PhD¹, Kevin M. Magnaye, BS¹, Catherine Stanhope, MS¹, Catherine Igartua,
PhD¹, Andréanne Morin, PhD¹, Charles Washington III, BA¹, Dan Nicolae, PhD^{1,3}, Klaus
Bønnelykke, PhD⁴, Carole Ober, PhD^{1,*}

¹Department of Human Genetics, University of Chicago, Chicago, IL 60637

²Section of Pulmonary and Critical Care, Department of Medicine, University of Chicago,
Chicago, IL 60637

³Department of Statistics and Section of Genetic Medicine, Department of Medicine, University
of Chicago, Chicago, IL 60637

⁴COPSAC (Copenhagen Prospective Studies on Asthma in Childhood), Herlev and Gentofte
Hospital, University of Copenhagen, Copenhagen.

⁺The authors contributed equally to this work.

Supplementary Materials

Contents

1. Supplementary Methods.....Page 3
2. Figure S1. Linkage Disequilibrium Plot in YRI.....Page 7
3. Table S1. Asthma-Associated Allele Frequencies in YRI.....Page 8

Supplementary Methods

Peripheral blood leukocytes were obtained from 112 Hutterite individuals (ages 7-76), a founder population of European ancestry in a study approved by the Institutional Review Board at the University of Chicago. Written consent for these studies was obtained from the adult participants and parents of children under 18; written assent was obtained from all children. One milliliter of whole blood was drawn into a TruCulture (Myriad RBM) tube containing proprietary TruCulture media and incubated upright in a dry heat block at 37C for 30 hours. Samples were washed twice with Buffer EL (Qiagen) and cell pellets were resuspended in 350 µl RLT Buffer (Qiagen) and frozen on dry ice. RNA was extracted from thawed cell pellets using AllPrep DNA/RNA Mini Kits (Qiagen). RNA-seq libraries were made using the TruSeq Library kit (Illumina); quality and concentration of libraries were assessed with an Agilent 2100 Bioanalyzer (Agilent Technologies) and quantitative PCR using the Kapa library quantification kit (Kapa Biosystems). Samples were sequenced in pools of 16-18 samples across three flow cells of an Illumina HiSeq 2500; 119 samples with low read count were re-sequenced on two flow cells on the same machine. Reads were mapped to hg19 and genes were counted using STAR¹. Samples with >7 million uniquely mapped reads underwent Trimmed Means of M-values normalization (TMM) and a voom transformation, used to correct for differences in library sizes³. Confounding technical effects were assessed in the normalized expression data using principal components analysis (PCA), and sequencing pool was adjusted using the function RemoveBatchEffect() from the R package Limma⁴.

Lung CD4 TRMs were sorted from 20 human organ donors whose lungs were not used for transplantation and provided by the Gift of Hope Organ and Tissue Donor Network. Fifty million

lung leukocytes were thawed and centrifuged over Histopaque 1077 (Sigma) gradient as previously described⁵. Cells were then stained with antibodies for FACS sorting of CD4TRMs with the CD45+CD206-CD14-CD3+CD11b-CD4+CD8-CDRO+CD11a+ phenotype. RNA was extracted from cells using AllPrep DNA/RNA Mini Kits (Qiagen). RNA-seq libraries were made using the SMART seq v4 Ultra Low Input RNA kit (Takara); quality and concentration of libraries were assessed with an Agilent 2100 Bioanalyzer (Agilent Technologies) and quantitative PCR using the Kapa library quantification kit (Kapa Biosystems). Samples were sequenced in pools with 84 samples (including other samples) across three flow cells of an Illumina HiSeq 2500; 4 samples with low read count were re-sequenced on two flow cells on the same machine. Reads were mapped to hg19 and genes were counted using TopHat2². Samples with >8 million uniquely mapped reads underwent Trimmed Means of M-values normalization TMM and a voom transformation, used to correct for differences in library sizes³. Confounding technical effects were assessed in the normalized expression data using PCA, and no confounding effects were seen with extraction batch, library batch, pool or lane. Effects for age, sex, race and date of cell sorting were removed using linear regression in the R package Limma⁴.

Primary airway smooth muscle cells (ASMCs) were isolated from 75 human donor lungs that were not suitable for transplantation and provided by the Gift of Hope Organ and Tissue Donor Network. Cells were isolated from trachea and main bronchi using established techniques⁶. Cells were cultured in 75cm² flasks in DMEM/F-12 media (Invitrogen) supplemented with 10% FBS, 5% non-essential amino acids (Invitrogen), and 5% antibiotic/antimycotic (Invitrogen). RNA was isolated using the QIAgen AllPrep Kit (Qiagen) and hybridized to Illumina Human HT-12 v4 arrays at The University of Chicago Genomics Facility. Probe level raw intensity values

across arrays were normalized using quantile normalization and background corrected normalized expression values were obtained for each probe using the R package lumi⁷. A total of 67 samples had adequate array intensities and were further processed. Probes that were indistinguishable from background intensity ($P < 0.01$), contained more than one HapMap single nucleotide polymorphism (SNP), or mapped to multiple locations in the genome⁸ were removed. Median probe intensity was used to represent the transcriptional abundance of each gene. Extraction batch, chip, RNA concentration, cell line age and plate were identified as potential confounders by PCA⁹ of the gene expression data. The effects of culture and extraction batch, chip and plate were removed using ComBat¹⁰, while the effects of the quantitative variables (RNA concentration and cell line age) were removed by linear regression.

References

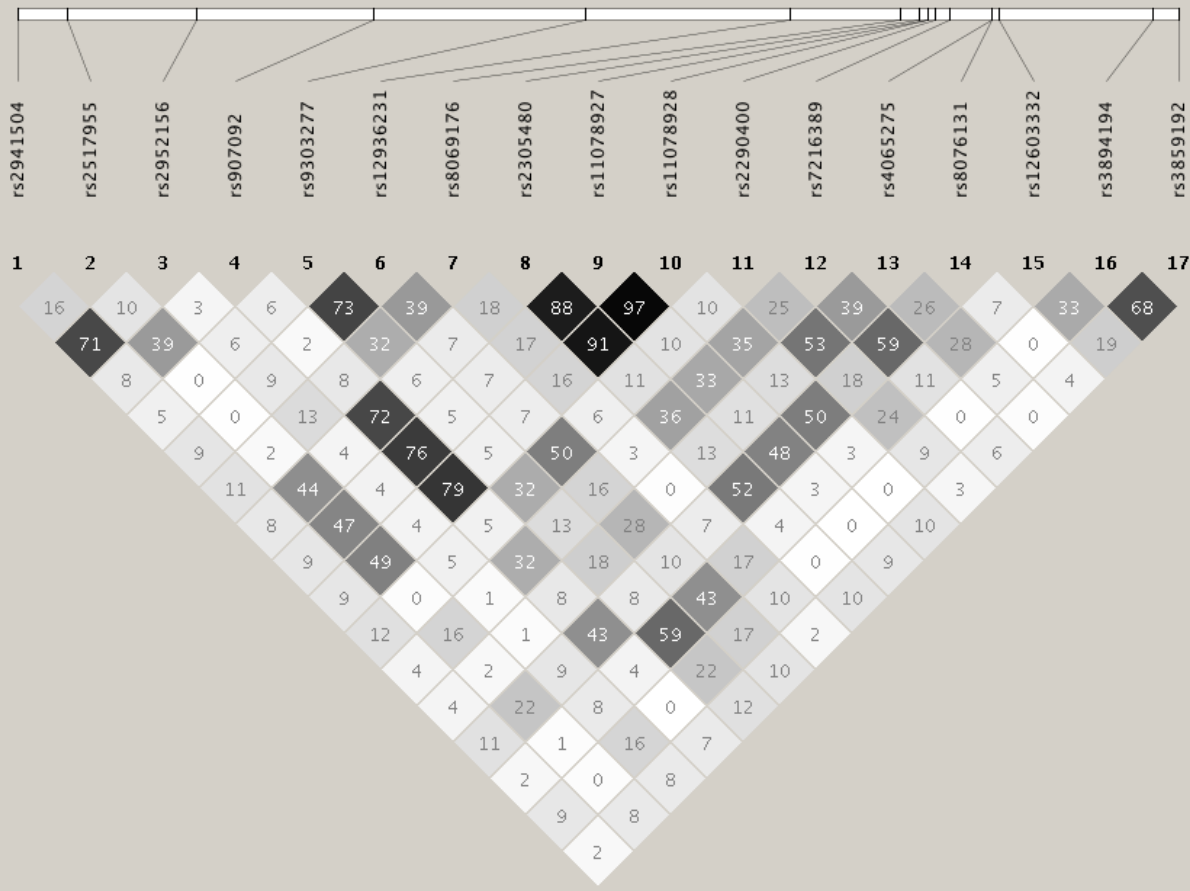
1. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 2013; 29:15-21.
2. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol* 2013; 14:R36.
3. Law CW, Chen Y, Shi W, Smyth GK. voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol* 2014; 15:R29.
4. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* 2015; 43:e47.
5. Adegunsoye A, Hrusch CL, Bonham CA, Jaffery MR, Blaine KM, Sullivan M, et al. Skewed Lung CCR4 to CCR6 CD4+ T Cell Ratio in Idiopathic Pulmonary Fibrosis Is Associated with Pulmonary Function. *Frontiers in Immunology* 2016; 7.
6. Panettieri RA, Jr. Isolation and culture of human airway smooth muscle cells. *Methods Mol Med* 2001; 56:155-60.
7. Du P, Kibbe WA, Lin SM. lumi: a pipeline for processing Illumina microarray. *Bioinformatics* 2008; 24:1547-8.
8. Nicodemus-Johnson J, Naughton KA, Sudi J, Hogarth K, Naurekas ET, Nicolae DL, et al. Genome-Wide Methylation Study Identifies an IL-13-induced Epigenetic Signature in Asthmatic Airways. *Am J Respir Crit Care Med* 2016; 193:376-85.
9. Leek JT, Scharpf RB, Bravo HC, Simcha D, Langmead B, Johnson WE, et al. Tackling the widespread and critical impact of batch effects in high-throughput data. *Nat Rev Genet* 2010; 11:733-9.
10. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* 2007; 8:118-27.

Figure S1. LD (r^2) among the 17 SNPs described in Figure 2 in Nigerians (YRI), using the 1000Genome reference panel.

ACCEPTED MANUSCRIPT

Table S1. Asthma-associated allele frequencies in Nigerians (YRI), from the 1000Genome reference panel.

SNP	Asthma Associated Allele	Freq	N
rs2941504	A	0.5	207
rs2517955	C	0.8575	207
rs2952156	A	0.4469	207
rs907092	G	0.8986	207
rs9303277	C	0.372	207
rs12936231	C	0.4227	207
rs8069176	G	0.6377	207
rs2305480	G	0.9058	207
rs11078927	C	0.91063	207
rs11078928	T	0.91304	207
rs2290400	T	0.5266	207
rs7216389	T	0.814	207
rs4065275	G	0.6667	207
rs8076131	A	0.8599	207
rs12603332	C	0.3623	207
rs3894194	A	0.2681	207
rs3859192	T	0.3213	207



1 **Box 1****What we know?**

- The asthma locus on 17q12-21 is the most replicated and most significant finding in genome-wide association studies (GWAS) of asthma
- Variation at this locus is specifically associated with very early onset asthma, possibly before age 3 years when it manifests as wheezing illness
- The 17q genotype effects on asthma risk and protection are modified by early life exposures
- The extensive linkage disequilibrium (LD) across this region in populations of European and Asian ancestry is greatly reduced in populations of African ancestry
- The asthma-associated 17q SNPs are expression quantitative trait loci (eQTLs) for nearby genes, *GSDMA*, *ORMDL3*, *GSDMB*, and *PGAP3*, in immune cells or lung cells or both
- The regulatory architecture of the 17q genes differs in immune and lung cells
- Variation at this locus is also associated with autoimmune diseases; alleles associated with risk for asthma are often associated with protection from autoimmune disease, and vice versa

What is still unknown?

- Is genetic variation at the extended 17q12-21 locus associated with early life wheezing illness and childhood onset asthma in African American children?
- Which 17q gene or genes (*ORMDL3*, *GSDMB*, *PGAP3*, and/or *GSDMA*) are involved in early life wheezing and asthma pathogenesis?
- What are the functions of and through what mechanisms do the 17q genes impact asthma onset and severity?
- In which cell types (blood cells, airway cells, or both) are the expression of these genes most relevant to the 17q asthma-associated risks or specific endotypes?
- Do the same 17q gene(s) modulate both protection against asthma in children exposed to animals and risk for asthma in children not exposed to animals?
- What is the natural history of asthma in later childhood and post-puberty in individuals with the high risk 17q genotype?
- Which genes and what mechanisms account for the opposite associations of 17q alleles with asthma and autoimmune diseases?