THE UNIVERSITY OF CHICAGO

REGULATORY AND FUNCTIONAL GENOMICS STUDIES OF COMPLEX RESPIRATORY DISEASES

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE DIVISION OF THE BIOLOGICAL SCIENCES
AND THE PRITZKER SCHOOL OF MEDICINE
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

COMMITTEE ON GENETICS, GENOMICS, AND SYSTEMS BIOLOGY

BY

MARCUS MAKINA SOLIAI

CHICAGO, ILLINOIS

MARCH 2021

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I would like to extend my gratitude to all of those that have assisted me during my graduate studies.

First, I would like to express my deep appreciation to my advisor, Carole Ober. Carole has had a profound influence on my personal and scientific development during the time that I have been in her lab. I am grateful for all of her support, patience, and mentorship that she has freely given me as I navigated through graduate school and my research projects. Her passion for genetics has been inspiring, and I hope that I can live up to her example.

I am thankful to the members of my committee: Xin He (Chair), Vincent Lynch, and Marcelo Nobrega for generously offering their time, expertise, and guidance throughout my graduate journey. Each of them were invaluable in my own scientific development and in the advancement of my own research. I am grateful to have been able to learn from them.

I am grateful to have been in such a great lab. The members of the Ober Lab, past and present, have each had an impact on my training and I am thankful to have been able to work with each of them. I would especially like to thank Emma Thompson for being so giving of her time and for brainstorming with me about solutions to many problems that I came across in my own research. I am also thankful for Katherine Naughton, Christine Billstrand, and Raluca Nicolae for their critical work on management and preparation of the samples that I used for my dissertation projects. I am also grateful for the work of Catherine Stanhope and William Wentworth-Sheilds, and for their computational and statistical expertise which was invaluable in helping me to complete this research.

I would also like to thank my mother, Kathleen Soliai, and my father, Aiulua Soliai, for instilling in me the importance of education. I am also grateful to my brother, Joshua Soliai, and

my sister, Danielle Uale for their all of their support and encouragement throughout my time in graduate school.

Finally, I would like to thank my wife, Tamsyn Soliai, and my daughter, Yeardley Soliai. I am so grateful for the sacrifice, patience, and support that Tamsyn has given me while in graduate school so that I could earn my PhD. I would also like to thank Yeardley for always brightening my day and for putting a smile on my face. I am so glad that we got to share in this adventure together.

ABSTRACT

Rhinovirus (RV) is the main cause of the common cold and as an exacerbator of chronic airway diseases. Yet, little is known about the genetic and epigenetic responses to this virus, or to the molecular mechanisms that contribute to airway disease after RV infection. This thesis addresses these questions by identifying molecular responses, pathways, and mechanisms of RV infection in a relevant tissue of two chronic airway diseases, chronic rhinosinusitis (CRS) and asthma, and through genome wide association studies (GWAS). In chapter 2, I used an airway epithelial cell (AEC) model of RV infection to identify differential DNA methylation (DNAm) and transcriptional responses to RV infection in CRS cases compared to non-CRS controls. Weighted gene co-expression network analysis (WGCNA) assigned differential molecular responses to co-expression/co-methylation modules enriched in known pathways of CRS pathogenesis and of immune response to viral infection. To study potential genetic causes of CRS, in chapter 3, I used a two-stage genome-wide association study (GWAS) in 2,540 European American adults (483 CRS cases, 2,057 non-CRS controls). Variation at two loci were identified as potentially involved in CRS; the known functions of a gene at one of these loci*, WSB2*, supported the role of the upper airway mucosal immune surface as a key site in disease development. Finally, because of the role of the upper airway epithelium as the primary site of RV infection and in asthma inception, in chapter 4, I leveraged the AEC model of RV infection to identify RV-response mechanisms that contribute to asthma risk. Expression and DNAm quantitative trait loci (eQTLs, meQTLs) were mapped in the vehicle- and RV-treated AECs and integrated with three asthma GWASs. Co-localization analyses of airway epithelial cell molecular QTLs with asthma GWAS variants revealed potential molecular mechanisms of asthma, including QTLs at the *TSLP* locus that were common to both exposure conditions and in all three asthma GWASs, as well as QTLs

at the 17q12-21 asthma locus, the most replicated and most significant asthma susceptibility locus in childhood onset asthma, that were specific to RV exposure and to childhood onset asthma. The results of the combined data at this locus indicated that gene expression of *ERBB2*, a gene associated with asthma severity, may be modulated by RV via an epigenetic mechanism involving long-range chromatin looping. These findings were supported by promoter capture Hi-C data from *ex vivo* bronchial epithelial cells. Together, these results not only highlight the broad regulatory landscape at this important asthma locus, but they also show the importance of context-specificity in identifying molecular mechanisms that contribute to disease risk.

CHAPTER 1

INTRODUCTION

1.1 Complex diseases of the respiratory system: asthma and CRS

Chronic respiratory diseases affect hundreds of millions of people globally and are among the leading causes of death [1, 2]. The most prevalent of these disorders, asthma, affects over 300 million people worldwide and is most common in Western countries where the average prevalence reaches 20.7% and ranges between 8.5% (Italy) and 32.0% (Ireland) [3, 4], and according to the Centers for Disease Control and Prevention (CDC), affects more that 5.5 million children and 19 million adults in the United States [5]. Moreover, asthma is a significant economic burden in the United States, with total annual costs surpassing $80 billion annually [6], though this figure is thought to be higher as researchers only included 'treated asthma' in their study. Chronic rhinosinusitis (CRS) is a debilitating upper airway disease that affects 5% to 15% of the U.S. population [7]. The financial burden associated with CRS is high, costing patients up to $6192 per year and over $5.8 billion in total expenditures in the U.S. [8, 9].

Asthma is a chronic, inflammatory condition of the airways causing symptoms that include wheeze, cough, and shortness of breath. Although asthma is typically diagnosed based on these clinical symptoms, it is comprised of many overlapping sub-phenotypes with underlying distinct endotypes that likely have both shared and unique genetic and environmental risk factors. For example, two subtypes based on age of onset (childhood vs adult onset asthma) is well recognized, and associated with different environmental and genetic risk factors. Overall, asthma has nearly equal contributions from genetic and environmental factors (heritability range: 55-74%) [10-14]. Although environmental factors, such as rhinovirus (RV) infection, are known

to exacerbate asthma symptoms [15-19], less is known about how asthma risk loci interact with environmental factors to increase risk of developing this disease.

Given the high heritability of asthma, there has been deep interest in uncovering the genetic risk factors for this disease. Initially, linkage analyses were applied as attempts to identify asthma risk loci [20, 21] and candidate-gene studies [22, 23], however, these methods either suffer from inadequate power to detect the genetic effects for asthma (i.e. linkage studies [24]), or from methodological limitations (i.e. candidate-gene studies [25]). The development of GWAS in 2004 has made it feasible to test hundreds of thousands of variants in a single analysis [26]. Since the first asthma GWAS in 2007 [27], over 150 asthma susceptibility loci have been identified, with the 17q12-21 locus being the most replicated in childhood onset asthma (reviewed in [28]). However, despite the advances made in asthma genetics due to GWASs, further work is required to move from variant-to-function using robust model systems to determine cell-type and environmental risk specificity.

CRS is a common condition of the paranasal sinuses in which the sinus epithelium becomes inflamed and swollen for more than 12 weeks [29]. CRS is one of the most common chronic diseases of adults in the U.S. [29-31], although recently it has been argued that childhood onset forms have been overlooked [32] CRS is divided into two major subtypes: CRS without nasal polyps (CRSsNP) and a more severe form, CRS with nasal polyps (CRSwNP). Patients with CRS experience thick postnasal drip, nasal obstruction, hyposmia, headache, dry upper respiratory tract syndrome, asthmatic complaints, headaches and other symptoms [33, 34], all of which contribute to a significant loss of quality of life [35].

CRS often occurs in the same patients with asthma and other reactive airway diseases [36, 37], as well as with aspirin-exacerbated respiratory disease, immunodeficiency-related

diseases, and cystic fibrosis (CF), a disease in which polyposis (formation of nasal polyps) also occurs [38]. CF carrier status may even predispose individuals to CRS; over 50% of CF carriers have self-reported CRS, suggesting that heterozygosity for a *CFTR* mutation may contribute to the development of CRS [39, 40]. In a linkage study in the Hutterites, CRS was linked to a region on chromosome 7q31, the region housing the *CFTR* locus [41]. As with asthma, CRS is a clinically challenging disease due to its complex and heterogeneous nature, exemplified by the lack of agreement on the categorization of the disease among clinical experts [42]. Unlike asthma, however, the evidence for genetic susceptibility in CRS is not well established. Most genetic studies of CRS have been limited by small sample size and low statistical power to identify genetic determinants with great confidence [43]. In one of the largest CRS GWAS conducted to date, Kristjansson et al. conducted a large meta-analysis of CRS (n = 5,608 cases) and NP (n = 4,366 cases) from the United Kingdom and Iceland [44], however only two genome-wide significant loci were identified for CRS while ten were identified for NP. This lack of genetic association may be due to substantial clinical heterogeneity that may have masked the effects of other risk loci. And although formal heritability calculations for CRS are not available, a genetic component for this disease has been suspected based on significant familial risk detected in First-degree relatives of CRS patients [45]. Several environmental factors are associated with the risk of developing CRS including tobacco exposure [46, 47], air quality [48, 49], and microbial infection [50].

Asthma and CRS share environmental risk factors, exposure to which can result in chronic inflammation of the upper or lower airways. Among the shared environmental risk factors, viral infection of the upper airways is the most common trigger of asthma exacerbation in children and infants [51, 52], and may also contribute to CRS exacerbations [53, 54]. The

most common viral infection thought to trigger exacerbation in these diseases is from rhinoviruses, the common cold virus.

1.2 Rhinovirus infection and its association with asthma and CRS

RVs belong to the genus *Enterovirus* of the family *Picornaviridae* and are the major cause of the common cold. On average, children experience up to 10 RV infections per year while adults can have between 2 and 5 infection a year [55]. RVs are classified into three groups including the RV-A group (74 serotypes), RV-B group (25 serotypes), and the RV-C group (60 serotypes) [56]. About 90% of RVs belonging to the 'major group' of serotyped strains use intercellular adhesion molecule-1 (ICAM-1) as a cell surface receptor for virus attachment and entry, while the 'minor group' uses a low-density lipoprotein receptor [57]. These receptors are found on the surface of the airway epithelium where the virus attaches itself and infects small clusters of cells. RVs primarily infect the nasal mucosa, though, they have also been detected in the lower respiratory tract [58-61].

RVs are the most common cause of acute exacerbations of asthma, consistently accounting for 60-70% of virus-associated exacerbations in children [62, 63], with similar rates in adults [64]. Additionally, RV-associated wheezing in the first three years of life is a known risk factor for asthma development in childhood [65]. Further evidence to support the role of RVs in asthma exacerbation includes the correlation of hospital admissions with seasonal patterns of upper respiratory infections [66]. Moreover, genetic association with asthma and RV wheezing illness have been observed. Çalışkan et al. reported that genetic variants at one of the most significant and most replicated loci associated with childhood onset asthma (17q12-21) were associated with asthma risk only in children with RV-associated wheezing illness in early life in two birth cohorts of high risk children[12]. This finding was replicated Loss et al. [67] in a

population-based birth cohort, indicating both the robustness of this interaction and the context-specificity of this locus on the development asthma in childhood.

Exposure to RV also has significant effects on the development of CRS. For example, inflammation of the nasal mucosa in CRS and the significantly higher prevalence of RV in CRS patients [68] suggest a role for this virus in the etiology of this disease. Additionally, the seasonal patterns of RV infections closely overlap with seasonal patterns of CRS exacerbations [69]. Abnormal and persistent responses to RV infection in CRS occurs and is often accompanied by severe dysregulation in mucus productions and sinus drainage [70, 71]. Despite the apparent importance of RV to CRS pathogenesis, the genetic and molecular mechanisms by which RV triggers and promotes persistent inflammatory responses in individuals predisposed to CRS is currently unknown.

1.3 In complex disease genetics, context matters

Assigning biologically interpretations to GWAS results is still a major challenge. Not only do we need to consider the where, when and how genes are expressed, but the context by which gene expression occurs is also critical. One of the major difficulties in interpreting GWAS results and prioritizing candidates for further studies is that over 90% of disease-associated variants are located in non-protein-coding regions of the genome [72]. These variants are enriched for chromatin signatures suggestive of enhancers [72, 73] and for eQTLs [72, 74, 75], indicating that the causal single nucleotide polymorphisms (SNPs) at these loci underly disease pathophysiology through their effects on gene regulation. Furthermore, GWAS results alone are not informative of tissue or cell type-specificity that may be central to the pathogenesis of disease.

Because the function of most disease-associated SNPs in GWAS remain unknown, identifying specific causal variants and their target genes at associated loci has been challenging. Databases such as GTEx, ENCODE, and ROADMAP have been useful in annotating GWAS SNPs, predicting molecular mechanisms through which risk variants affect disease phenotypes [75-78], and providing important insights into the interpretation of GWAS results. However, these databases do not include all cell types relevant to all diseases or information on environmental exposures that influence disease outcomes. As a result, for example, annotations of asthma GWAS variants have been largely limited to studies in transformed B cell lines, blood (immune) cells, and whole lung tissue (e.g. [11, 13, 27, 79-81]), even though other tissues are more central to the disease, such as airway smooth muscle [82] and the mucosal epithelium [83].

Characterizing the genetic and molecular responses to environmental exposures in cells from disease-relevant tissues can be hard to test in *ex vivo* samples. For example, variability in exposure to potentially important environmental triggers, such as to microbes and pollution, can confound or even mask inter-individual responses to an environmental exposure of interest. *In vitro* cell models provide an opportunity to address this limitation by allowing for controlled environmental exposures in relevant cell types, and for precise study designs that minimize confounders to an extent that would be impossible to achieve in *ex vivo* studies. To this end, *in vitro* functional studies of the airway epithelium, an asthma- and CRS-relevant tissue, have been used to characterize gene pathways affected by environmental disease modifiers of asthma (e.g., refs. [84-87]).

1.3.1 Dissertation Overview

In this thesis, I describe studies using an *in vitro* cultured AEC model of RV infection for discovery of context-specific molecular disease mechanisms in both asthma and CRS. As part of

the Chronic Rhinosinusitis Integrative Studies Program (CRISP), an AEC model of RV infection was designed to characterize molecular pathways and processes of response. By measuring gene expression and DNAm in these cells, I identified gene pathways and epigenetic regulatory mechanisms in the airway epithelium that are associated with risk of these diseases after RV infection. These data were used to first map context-specific molecular QTLs from the AECs and then integrated with existing asthma GWASs to annotate risk loci with better precision, identify potential RV-specific mechanisms of disease, and to better prioritize genes that may contribute to disease risk.

In the following chapters, I used genomics data col to gain a functional understanding of RV response mechanisms in CRS and asthma. This dissertation is divided into three main analyses. In chapter 2, epigenetic and transcriptional responses to RV that differ between CRS cases and non-CRS controls were identified. Modules of co-regulated gene expression and DNAm profiles were then identified, and potential genes, pathways, and epigenetic regulatory effects that differ between CRS cases and controls were discovered and described. In chapter 3, a two-stage GWAS of CRS was performed with 483 CRS cases and 2,057 non-CRS controls of European ancestry in order to identify any potential risk loci associated with subtypes of this disease. Finally, in Chapter 4, molecular QTLs from the AEC model of RV infection were mapped and integrated with three asthma GWAS, including a GWAS of childhood onset asthma and a GWAS of adult onset asthma. This allowed the identification of potentially functional variants, some of which have context-specific effects on transcriptional and epigenetic responses, and molecular mechanisms of disease.

CHAPTER 2

Epigenetic responses to rhinovirus exposure in airway epithelial cells are correlated with key gene pathways in chronic rhinosinusitis

2.1 ABSTRACT

Rhinovirus (RV) infections of the paranasal sinus epithelium are common in individuals with chronic rhinosinusitis (CRS) and may trigger exacerbation events or result in progression of this disease. Our aim was to identify epigenetic and transcriptional responses to RV that differ between CRS cases and non-CRS controls using a cultured airway epithelial cell model of RV response. Airway epithelial cells were collected from CRS with nasal polyps (CRSwNP) cases and non-CRS controls during endoscopic surgery. Cells were cultured and treated with RV or a vehicle control for 48 hours at which time DNA and RNA were extracted. Gene expression and DNA methylation were measured and response differences to RV infection between cases and controls were determined using linear mixed models. Weighted gene co-expression analysis (WGCNA) was used to identify co-regulated gene expression and DNA methylation (DNAm) profiles and to identify potential genes, pathways, and epigenetic regulatory effects that differ between CRS cases and controls. Our results highlight an important role for epigenetic mechanisms underlying different responses to RV between CRSwNP cases and non-CRS controls, which influence gene expression pathways that reflect dysregulated epithelial cell processes in CRS.

2.2 Introduction

Chronic rhinosinusitis (CRS) is a common disease of the paranasal sinuses in which the sinus epithelium becomes inflamed and swollen for more than 12 weeks [29]. CRS can manifest as one

of two major clinical syndromes: a milder form, CRS sans nasal polyps (CRSsNP), and a more severe form, CRS with nasal polyps (CRSwNP). These subtypes of CRS are associated with different co-morbidity profiles [88-90] and environmental factors [91-93], reflecting heterogeneity with respect to both clinical features and underlying mechanisms, or endotypes. Moreover, there have been relatively few genetic studies of CRS, and the few reported associations have not been replicated [44, 94, 95]. This may reflect the significant effect of environmental exposures on the development of CRS. For example, inflammation of the sinuses in CRS and the significantly higher prevalence of respiratory viruses in CRS patients [54], especially rhinovirus (RV) [68], suggests a role for this common virus in disease etiology and/or pathogenesis. In fact, it has been proposed that RV infection may promote CRS in individuals with constitutively abnormal responses in the sinus epithelium that leads to chronic inflammation [68, 96].

In support of the latter, viral infections of the paranasal sinus epithelium are often associated with triggering events of CRS [97, 98]. In particular, RV infection is detected in CRS patients more frequently than in non-CRS subjects [54], and is thought to induce symptoms of this disease [54, 92, 99]. Moreover, the seasonal patterns of RV infections closely overlap with seasonal patterns of CRS exacerbations [69]. Yet, while exposure to this common pathogen is pervasive throughout the population, only a fraction of people develops CRS. Taken together, these observations suggest that intrinsic features of the sinus epithelium that differ among individuals may mediate responses to RV infection and subsequent risk for CRS. However, the specific molecular response mechanisms to RV infection in the sinus epithelium from CRS patients and non-CRS controls are poorly defined.

In vivo studies of the effects of RV infection on molecular responses in CRS are challenging for two primary reasons. First, significant variability in exposure to potentially important environmental triggers, such as to microbes and pollution, can confound or even mask inter-individual responses to RV infection. Second, although CRS can be grouped into two major phenotypic subgroups (CRSsNP and CRSwNP) based on nasal endoscopy and computerized tomographic (CT) scans [29], additional CRS subtypes exist based on underlying conditions, such as allergic fungal rhinosinusitis (AFRS), aspirin-exacerbated respiratory disease (AERD), cystic fibrosis, primary ciliary dyskinesia, and immune deficiencies. Moreover, CRS comorbidities with common conditions, such as asthma [88] and allergic rhinitis [100], can further mask true differences between CRS cases compared to healthy controls. *In vitro* cell culture models address these limitations by allowing for controlled environmental exposures in relevant cell types, and for precise study designs that minimize confounders to an extent that would be impossible in *ex vivo* studies. Thus, cell culture models would allow for the identification of key genes, biological pathways, and possible epigenetic mechanisms of responses to RV in airway epithelial cells from CRS cases and from non-CRS controls.

Here, we report the results of a multi-omics study to identify molecular regulatory effects of viral infection on CRS using an epithelial cell model of viral response. Because of the importance of both the upper airway epithelium and RV infection in CRS pathogenesis, we used an upper airway (sinonasal) epithelial cell model of gene expression and DNA methylation response to RV exposure to reveal gene pathways and epigenetic regulatory mechanisms that differ between CRSwNP cases and non-CRS controls. Using an integrative systems biology approach, we identified six modules of co-regulated gene expression and DNA methylation. The eigenvectors of these modules were strongly correlated with either CRS or RV exposure. The

modules were enriched for pathways previously implicated in both CRS pathogenesis and immune responses to microbes. Our results highlighted DNA methylation as having an important role in RV response differences between CRS cases and controls, providing an important first step towards understanding the molecular basis for this debilitating disease, and providing clues to the intrinsic, epigenetic differences in airway epithelial cells in CRSwNP and the etiologic role of RV in this disease.

2.3 Methods

2.3.1 Ethics Statement

Sinonasal epithelial cell brushings were collected study participants between March 2012 and August 2015 during endoscopic surgeries at Northwestern University Feinberg School of Medicine. Informed written consent was obtained from each study participant and randomly generated IDs were assigned to all samples, preserving the participants' anonymity and privacy. This study was approved by the institutional review boards at Northwestern University Feinberg School of Medicine and the University of Chicago.

2.3.2 Cohort description and sample collection

This study was a sub-project of the Chronic Rhinosinusitis Integrative Studies Program (CRISP), which had the overall goal of progressing the understanding of chronic rhinosinusitis (CRS) genetics, epidemiology, and immunopathology[101]. The objective of our study was to use a systems biology approach to identify molecular responses to RV that differ between nasal epithelial cells from CRSwNP cases and from non-CRS controls. To this end, we designed a cell culture model of airway epithelial cells and RV infection to characterize molecular pathways and processes of response. The workflow of our study is shown in Fig 2.1.

The cell model employed in this study used nasal epithelial cell brushings of the uncinate process that were collected from 104 study participants during routine endoscopic surgery for sinus disease (cases) or for other unrelated indications (adenoidectomy, dentigerous cysts, septoplasty, and tonsillectomies) in the controls at Northwestern University. The study cohort

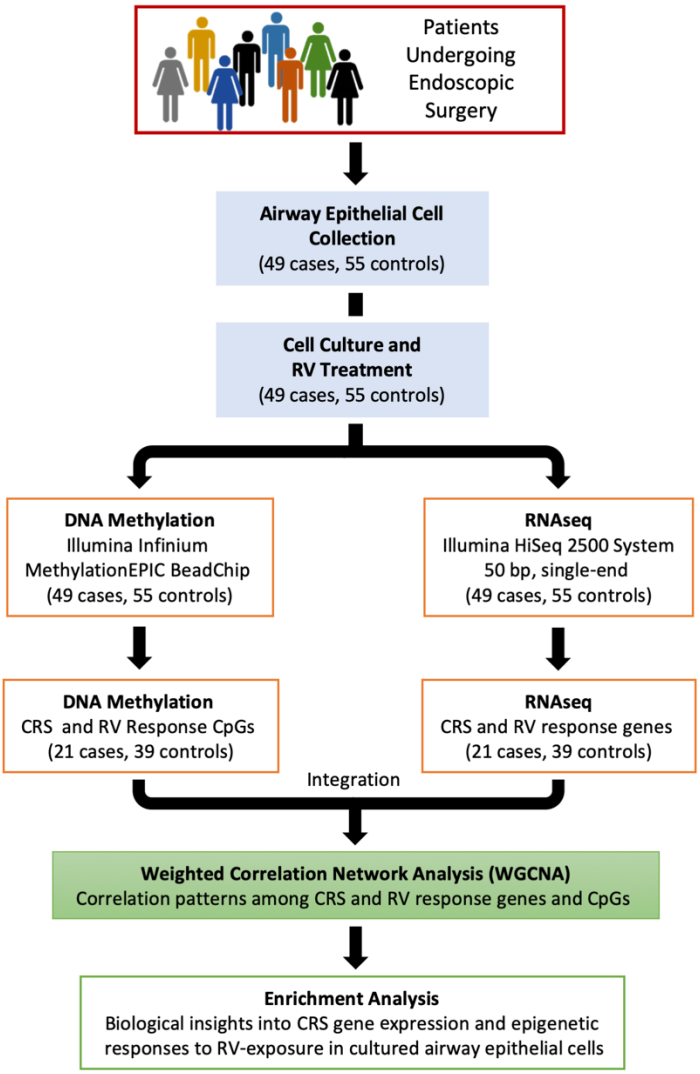


Figure 2.1 Study design and analytical workflow. See Methods for additional details.

included 49 adults with CRSwNP and 55 non-CRS control patients without a history of sinus disease. CRS patients were diagnosed based on the European Position Paper on Rhinosinusitis

and Nasal Polyps (EPOS) criteria [29]. Because asthma is a common co-morbidity with CRS, we excluded subjects with a current or previous physician diagnosis of asthma based on medical chart reviews to reduce potential confounding. The remaining sample included 22 CRS cases and 39 non-CRS controls. The demographic characteristics of the sample with and without the asthmatics excluded, by case/control status are shown in Table 2.1.

Table 2.1 Demographic composition of the subjects in this study.

	Combined	**CRS**	**non-CRS**
Sample Size (N)	61	22	39
Sex (%Female)	23	14	51
Median Age (IQR)	45 (26)	54 (15)	38 (22)
Ethnicity (self-reported)			
White	39	16	23
Black	13	4	9
Hispanic	7	2	5
two or more	2	-	2

2.3.3 Airway epithelial cell culture and RV treatment

The airway epithelial cell culture model and RV treatment protocol has been described in detail previously [102]. Briefly, after isolation, airway epithelial cells were cultured in bronchial epithelial cell growth medium to near confluence, then frozen at -80°C and stored in Liquid Nitrogen for a period ranging from 8 days to 3 years. Airway epithelial cells were thawed, cultured and treated for 48 hours with rhinovirus (HRV-16; RV) or with a vehicle (Bronchial epithelial cell basal medium (BEBM) + Gentamicin/Amphotericin) alone (2-hour treatment followed by a wash; 46 hours culture time subsequent to treatment wash).

2.3.4 Ancestry Principal Components

Genotyping was performed using the Illumina Infinium HumanCore Exome+Custom Array. The 676 ancestry informative markers [103] on the array were used for ancestry principal

components analysis (PCA) of each subject. The first three ancestry principal components (PCs) were included in all subsequent analyses to correct for population structure and ancestry imbalances between the cases and controls.

2.3.5 RNA extraction, sequencing, and QC

RNA was extracted from RV- or vehicle-treated cells. cDNA libraries were constructed using the Illumina TruSeq RNA Library Prep Kit v2 and sequenced on the Illumina HiSeq 2500 System (50 bp, single-end) at the University of Chicago Genomics Core. Sample contamination and swaps were not detected using the VerifyBamID software [104]. RNA sequences were subsequently mapped to the human reference genome (hg19) and reads per gene were quantified using the Spliced Transcripts Alignment to a Reference (STAR) [105] software. Mitochondrial, X and Y chromosome genes, and low count data (genes with <1 count per million [CPM]) were removed while maintaining a greater than 8M mapped read count for each sample. A trimmed mean of M-values method (TMM) of normalization and variance modeling (voom) [106] were then applied to the sample dataset. 11,898 autosomal genes in the 19 CRS cases and 32 non-CRS controls remained after quality control and were included in downstream analyses.

Principal components analysis (PCA) was used to identify biological and technical sources of variation in the normalized RNAseq dataset. Six technical effects contributed to sample variance: technician, days of cell culture, cell lysate batch, RNA concentration, sequencing pool, and percent of mapped RNAseq reads. Thirteen unknown sources of variation (surrogate variables [SV]) were estimated for these samples using the SVA R package [107], after protecting for CRS and treatment. Corrections for technical effects and SVs in the analyses are described below.

2.3.6 DNA extraction, methylation profiling, and QC

DNA was extracted from the cells following cell culture, using the QIAGEN AllPrep DNA/RNA Kit, as described above. DNAm was measured on the Illumina Infinium MethylationEPIC BeadChip at the University of Chicago Functional Genomics Core. Of the 866,836 CpG probes on the array, we removed 74,444 probes that were either located on the X or Y chromosomes, or had detection *P* values greater than 0.01 in more than 10% of samples. Raw probe values or background were corrected using a preprocessing control normalization function, and technical differences between the Infinium type I and type II probes were corrected with the Subset-quantile Within Array Normalization (SWAN) method [108]. Cross-reactive probes and probes within two nucleotides of a SNP with a MAF > 0.05 were removed using the function mSNPandCH() from the R package DMRcate [109]. Measurements of 792,392 autosomal CpG probes in 22 CRS cases and 39 non-CRS controls remained after quality control.

PCA identified biological and technical sources of variation in the normalized methylation dataset after excluding samples with any history of asthma. Variability due to cell harvest date was the only significant technical effect. Age, sex, and ancestral PCs 1-3 were significant biological variables. Unknown sources of variation were predicted with the SVA package [107] in R, which estimated 21 SVs after protecting for CRS and treatment.

2.3.7 Differential gene expression and DNA methylation analysis

To identify genes and DNAm sites associated with RV-response, CRS or both in airway epithelial cells, we used linear mixed effects models to identify differentially expressed genes (DEGs) and differentially methylated CpGs (DMCs; M-values) using the limma R package [110]. The models used to identify DNAm and gene expression differences in CRS cases and

non-CRS controls in vehicle- and RV-treated cells, as well as RV-response genes in CRS and non-CRS samples for each gene or CpG site followed the general form:

$$Y_{(Gx \text{ or } DNAm)} \sim \beta_0 + \beta_1 X_{CRS} + \beta_2 X_{Treatment} + \text{covariates},$$

using an FDR of 10% to control the false positive rate. Biological and technical sources of variation were included as covariates for their respective datasets (described above), as well as age, sex, and ancestry PCs 1-3.To test for interactions between RV-response and CRSwNP on gene expression and on DNAm, we included an interaction term as follows:

$$Y_{(Gx \text{ or } DNAm)} \sim \beta_0 + \beta_1 X_{Treatment} + \beta_2 X_{CRS} + \beta_3 X_{Treatment \times CRS} + \text{covariates}.$$

Because of the general sparsity of interaction effects detected in small samples, only DEGs and DMCs identified between the cases and the controls in vehicle and/or RV treatments (FDR<10%) and RV-responsive genes and DNAm sites in cases and/or controls (FDR<10%) were assessed for interaction effects.

2.3.8 Correlation network construction by WGCNA

Gene expression and DNAm correlation networks were constructed using a supervised weighted gene co-expression network analysis (WGCNA) [111]. For this analysis, we included the 7,474 DEGs and 6,254 DMCs identified in any of the differential gene expression and DNAm analyses at a FDR<10% (discussed above). The residuals for the DEGs and DMCs were merged and quantile normalized before WGNCA analysis.

A weighted adjacency matrix was created for the combined gene expression and DNAm residuals by calculating Pearson correlations between the normalized residuals of gene expression and the M-values of DNAm. The co-expression/methylation similarity matrix was raised to a power $\beta = 5$ ($R^2 = 0.941$) based on the scale-free topology fit index reaching a high value above 0.85 (Fig S2.1A), and with a moderate mean connectivity (Fig S2.1B). The

similarity matrix (β = 5) was used in order to calculate the weighted adjacency matrix [112], and to measure the connection strengths between the nodes, following published methods [111]. A topological overlap matrix (TOM) was then constructed to determine topological similarities (Fig S2.2A), and then used to calculate the corresponding dissimilarities (1-TOM) in order to build clusters (Fig S2.2B). Genes and CpGs with coherent gene expression and DNAm profiles were grouped into modules using the average linkage hierarchical clustering coupled with TOM-based dissimilarity calculations (function = cutreeDynamic(), cut height = 0.25, minimum number of genes/CpGs per module = 20). Colors were randomly assigned to modules, and modules whose gene expression and/or DNAm profiles were considered similar at a threshold height of 0.25, corresponding to a correlation of at least 0.75, were merged using the mergeCloseModules() R function from WGCNA. The Spearman correlation method was then used to correlate module eigenvectors with CRS status and treatment.

2.4 Results

2.4.1 Molecular profiles differ between CRS and controls, and after RV infection

Our first goal was to identify the effects of CRS and RV infection on gene expression (n=19 cases, 32 controls) and DNAm profiles (n=22 cases, 39 controls) in upper airway cells. To this end, we conducted four analyses of each molecular marker comparing CRSwNP cases to non-CRS controls, separately in vehicle- and RV-treated cells, and comparing RV- to vehicle-treated cells, separately in cases and in controls (Fig 2.2). Of the 11,898 expressed genes, 1,501 (12.6%) were up-regulated and 1,517 (12.3%) were down-regulated in cases compared to controls in vehicle-treated cells, and 1,485 (12.5%) were up-regulated and 1,463 (12.3%) were down-regulated in cases in RV treated cells, at a FDR<10% (Fig 2.2A). Of the 3,371 differentially expressed genes (DEG) in either treatment, 2,596 genes (77.0%) were shared between cases and

controls, and 422 and 353 were differentially expressed at FDR < 10% only in the vehicle- or in the RV-treated cells, respectively (Fig 2.2B).

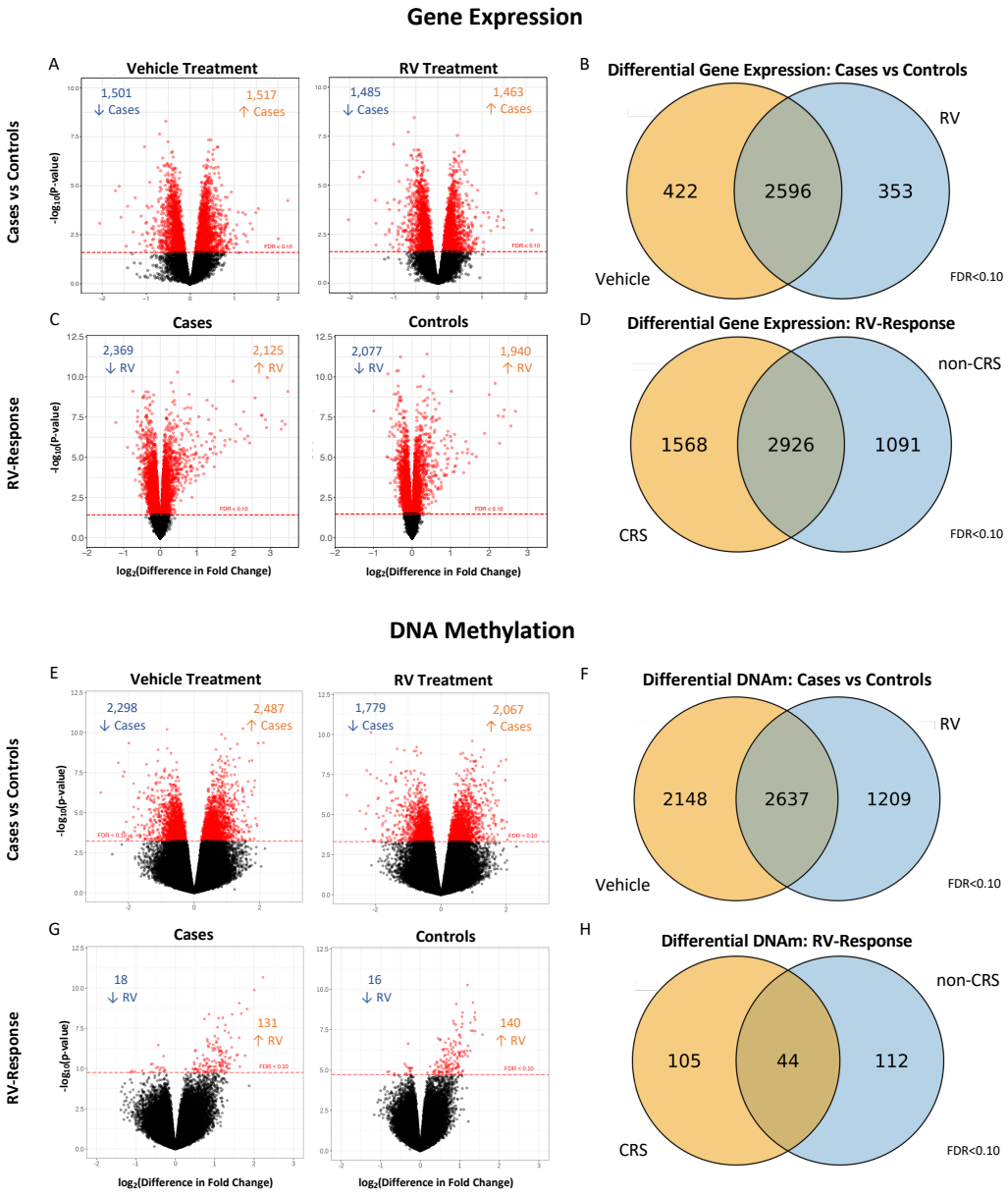


Figure 2.2 Differential gene expression and DNA methylation analysis of cultured airway epithelial cells treated with RV and a vehicle control. A-D Results of the differential gene expression analyses. (A) Volcano plots of differentially expressed genes (DEG) between cases and controls from the vehicle-treated cells (left) and the RV-treated cells (right). The red line in each volcano plot indicates a 10% FDR threshold. (B) Venn diagram of DEGs from the

Figure 2.2 Differential gene expression and DNA methylation analysis of cultured airway epithelial cells treated with RV and a vehicle control. (continued) comparison of cases and controls from the vehicle- and RV-treated cells. C Volcano plots of DEGs of RV-response in the cases (left) and controls (right). (D) Venn diagram of RV-responsive genes in the cases and controls. (E-H) Results of the differential DNA methylation analyses. (E) Volcano plots of the differentially methylated CpGs (DMC) between cases and controls in the vehicle-treated (left) and RV-treated (right) cells. (F) Venn diagram of DMCs between cases and controls from the vehicle- and RV-treated cells. (G) Volcano plot of DMCs in response to RV treatment in cells from the cases (left) and controls (right). (H) Venn diagram of RV-responsive CpGs in cases and controls.

Among the same 11,898 expressed genes, 2,369 (19.9%) were up-regulated and 2,125 (17.8%) were down-regulated in response to RV treatment in cells from CRS cases, and 2,077 (17.5%) were up-regulated and 1,940 (16.3%) were down-regulated in response to RV treatment in cells from the non-CRS controls, at a FDR<10% (Fig 2.2C).

Of the 792,392 CpG sites included in this study, 2,298 (0.30%) were hypermethylated and 2,487 (0.31%) were hypomethylated in cases compared to controls in vehicle-treated cells, and 1,779 (0.22%) were hypermethylated and 2,067 (0.26%) were hypomethylated in cases compared to controls in the RV-treated cells, each at a FDR<10% (Fig 2.2E). Of the 5,994 DMCs in either the cases or controls, 2,637 (43.9%) were shared, and 2,148 and 1,209 DMCs were only present in the vehicle- or RV-treated cells, respectively, at FDR<0.10 (Fig 2.2F). Among the same 792,392 CpG sites, 18 (<0.01%) were hypomethylated and 131 (0.01%) were hypermethylated in response to RV in cases, and 16 (<0.01%) were hypomethylated and 140 (0.01%) were hypermethylated in response to RV in controls (Fig 2.2G). Among the 261 RV-responsive CpGs in either the cases or controls, 44 (16.9%) were shared, whereas 105 and 112 DMCs were only in the cases or controls, respectively, at a FDR<10% (Fig 2.2H).

2.4.2 Interaction effects of CRS and RV exposure suggest that molecular responses to RV differ between CRS and controls

The DEGs and DMCs with FDR<10% in airway epithelial cells only from cases or only from controls in response to RV or DEGs and DMCs between cases and controls in either treatment at FDR <10% reflect different molecular responses of cells from cases and controls to treatments. To more directly assess interactions between case-control status and RV exposure on molecular responses, we included an interaction term (CRS x RV) in a linear mixed effects model (see Methods). To reduce the multiple testing burden and increase power in this small sample, we focused these studies on the subset of 7,474 DEGs and 6,254 DMCs identified in any of the eight analyses at a FDR<10% (Fig 2.2).

No CRS x RV interactions on gene expression were identified, although 117 genes were suggestive of an interaction effect with an uncorrected *P* value <0.01 (P_{adj}≥0.10). Analysis of these 117 genes by gene set enrichment analysis (Enrichr [113]) identified enrichments in immune-related pathways such as the Interleukin-2 signaling pathway, TNF-alpha effects on cytokine activity, and Interleukin-5 regulation of apoptosis (BioPlanet catalogue [114]), pathways previously implicated in CRS pathogenesis (Table S2.1) [115-118]. These results suggest that some components of immune responses to RV differ between CRSwNP cases and controls, by showing either a dampened (66 genes), heightened (42 genes), or no change (9 genes) in response to treatment in the cases.

In contrast to gene expression results, CRS x RV interactions were detected at 61 CpGs at an FDR <10% (Fig 2.3A). These 61 DMCs reflect different DNAm responses to RV in cultured airway epithelial cells from cases compared to controls. Fig 2.3B illustrates examples of DNAm patterns at four CpGs with different types of CRS x RV interactions: two show responses to RV

that are only present in the cases (left panels) and two show responses to RV that go in opposite directions in cases compared to controls (right panels). To investigate the regulatory potential of these 61 CpGs, we overlapped their genomic locations with ENCODE [119] transcription factor binding sites (TFBS) and tested for enrichments of the CpGs involved in interactions at TFBS relative to CpGs without interactions effects. Indeed, CpGs involved in interactions were enriched 1.28-fold in TFBSs ($P = 5.70\text{x}10^{-3}$; hypergeometric test). Overall, 46 of the 61 (75.4%) CpGs overlapped with TFBSs for 135 transcription factors, suggesting their potential for influencing the binding of these transcription factors and influencing the expression of their downstream genes. Taken together, these results suggest that airway epithelial cells from subjects with CRSwNP may have intrinsic differences in their epigenetic responses to RV infection compared to non-CRS controls, with potential to impact the expression of hundreds of genes regulated by transcription factors that target sites overlapping with the CpGs involved in interactions.

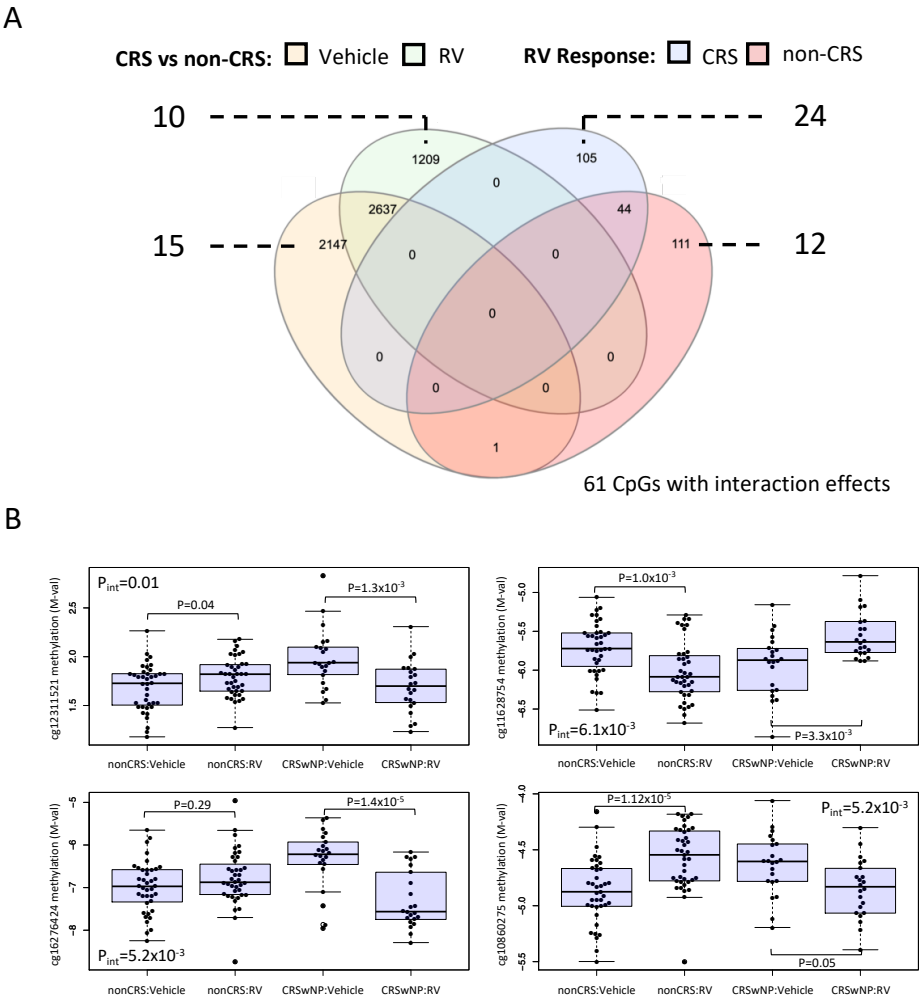


Figure 2.3 DNAm sites tested for interaction effects for CRSwNP status and treatment (CRSwNP x treatment). (**A**) Venn diagram showing the overlap of the 6,254 differentially methylated sites at a FDR<10%. The numbers on the left and right of the Venn diagram indicate the number of CpGs that were identified to have interaction effects (FDR<10%) and the differential analyses from which they were initially identified. (**B**) Box plots showing examples of DNA methylation interaction effects for four of the 61 interactive CpGs. The adjusted *P* values for the interaction effects (P_{int}) are shown in each box plot.

2.4.3 WGCNA identified co-regulated modules of gene expression and DNA methylation

We next used WGCNA [111], a systems biology tool, to evaluate coordinated responses between gene expression and DNAm. For this analysis, we included the 7,474 DEGs and 6,254 DMCs discovered in the analyses described above (Fig 2.2). WGCNA initially assigned all the DEGs

and DMCs to one of eight modules of correlated gene expression and DNAm patterns (Fig 2.4A). Based on the high degrees of correlation between some modules (Spearman $\rho > 0.75$), the black, green, and pink modules (upper bar in Fig 2.4A) were merged into into a single (black) module (lower bar in Fig 2.4A), yielding six modules of co-regulated gene expression and/or DNAm. Three modules included both genes and DNAm sites (black, blue, turquoise), two included only genes (red and yellow), and one included only DNAm sites (brown) (Fig 2.4B). The correlations (and p-values) of each module with CRS and treatment are shown in the first and second columns of Fig 2.4B.

Not surprising, all module eigenvectors were significantly correlated with CRS or treatment after correction for 12 tests (Bonferroni corrected $P < 4.1 \times 10^{-3}$). However, it was unexpected that none of the modules were significantly correlated with both treatment and CRS: three were significantly correlated with CRS (brown, blue, turquoise) and three were significantly associated with treatment (red, black, yellow). The three modules associated with CRS were predominantly comprised of DMCs: 95.6% of all DMCs were in the CRS-associated modules. The three modules that were correlated with treatment include 56.3% of all DEGs, with the remaining 43.6% distributed among two modules (blue and turquoise) associated with CRS. Consistent with the results described above, these data further suggest that DNAm may contribute to intrinsic differences in airway epithelial cells between CRSwNP subjects and non-CRS controls, and further suggests that differences in gene expression responses to RV infection in cases and controls may be secondary to these epigenetic modifications.

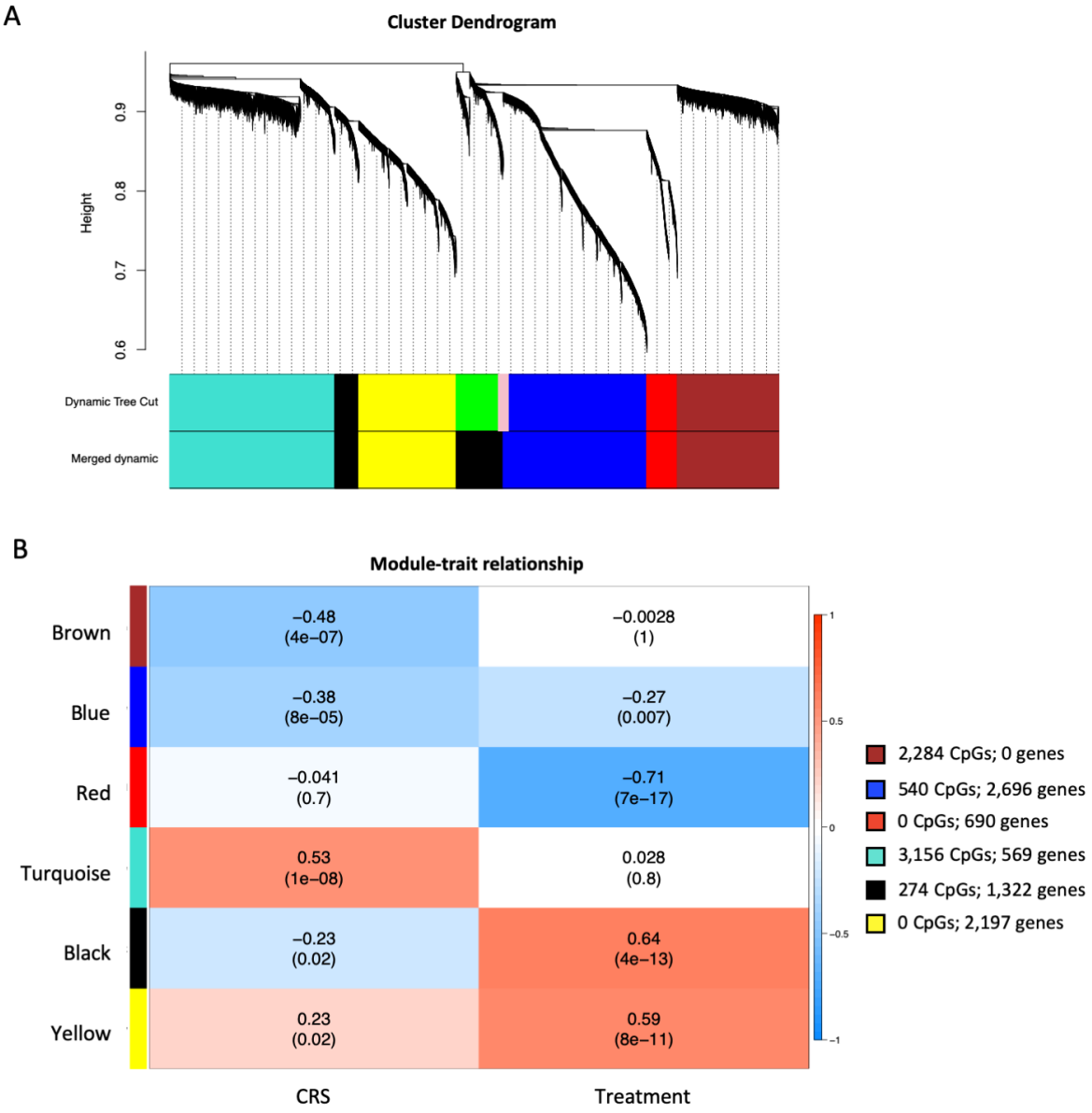


Figure 2.4 Network analysis dendrogram showing co-regulation modules of gene expression and DNA methylation profiles identified by weighted gene co-expression network analysis (WGCNA). (A) Dendrogram showing co-regulated modules of co-expression and DNAm. Colored rows below the dendrogram indicate modules that were identified by WGCNA. The first colored row contain modules of highly correlated gene expression and DNAm (Dynamic Tree Cut); the pink module was enlarged for viewing convenience. The second row is a merger of modules with highly correlated expression or DNAm profiles. (B) The module-trait relationship. Each row corresponds to a module eigenvector while each column corresponds to either CRS status or treatment. Each cell contains corresponding correlations (top value) and *P* value (bottom value). Cells are colored blue for negative correlations while red cells indicate positive correlations from -1 to 1, respectively. The legend on the right indicates the number of co-regulated genes and DNAm sites in each module.

2.4.4 Module genes are enriched in CRS and microbial response-associated gene pathways

To infer biological processes from the genes within the modules, we performed gene enrichment analyses (Enrichr gene enrichment analysis tool [113]) using biological pathways from the WikiPathways database [120] for the five gene-containing modules (Table 2.2). The eigenvectors of the modules correlated with treatment (red, yellow, black) contained genes enriched in pathways representing different components of the innate immune response and molecular signaling pathways involved in response to microbial infection.

Genes from the red module were enriched in innate immune pathways activated after microbial exposure, including immune response to tuberculosis ($P = 6.35\text{x}10^{-19}$; $P\text{adj} = 3\text{x}10^{-16}$) and type II interferon signaling ($P = 3.51\text{x}10^{-17}$; $P_{adj} = 8.27\text{x}10^{-15}$), as well as cytokine-signaling pathways ($P_{adj} < 0.10$). Genes from the yellow module were enriched in five lipid-related pathways including the cholesterol biosynthesis pathway ($P = 9.68\text{x}10^{-8}$; $P_{adj} = 2.28\text{x}10^{-7}$), lipid metabolism pathway ($P = 1.41\text{x}10^{-4}$; $P_{adj} = 0.01$), as well as the sterol regulatory element-binding protein (SREBP) and in cholesterol and lipid homeostasis pathway ($P = 1.17\text{x}10^{-4}$; $P_{adj} = 0.01$). These pathways have been implicated in the regulation of RV replication in bronchial epithelial cells [121], and are upregulated by viruses in order to meet the demand for viral structural elements [122]. The genes in the yellow module were also enriched in ciliary gene pathways, possibly reflecting the response of the airway epithelium to compensate for the loss of ciliated cells during respiratory viral infections [123, 124]. The co-expressed genes from the red and black modules were also enriched in cancer-related pathways associated with cell proliferation and cellular processes that have also been associated with CRSwNP [125, 126]. RV-induced expression of genes that promote cell proliferation may contribute to CRS exacerbations and

progression of CRS symptoms after RV infection, and the development of polyps in more severe CRS. Genes from the black module were enriched in the epidermal growth factor signaling pathway ($P = 1x10^{-3}$; $P_{adj} = 0.09$), a pathway that has been consistently associated with CRS [126-129]. Thus, coordinated gene expression responses to RV that differ between cases and controls are enriched for diverse pathways, some of which are shared between RV-correlated modules and some of which are module-specific.

In contrast, the blue and turquoise modules that were associated with CRSwNP were predominantly enriched in molecular pathways previously linked to CRS. Co-expressed genes in the blue module were primarily enriched for cytoplasmic ribosomal proteins ($P = 1.31x10^{-39}$; $P_{adj} = 6.21x10^{-37}$) and mitochondrial function and processes including the electron transport chain ($P = 1.61x10^{-14}$; $P_{adj} = 3.81x10^{-12}$) and oxidative phosphorylation ($P = 7.66x10^{-10}$; $P_{adj} = 1.21x10^{-7}$). Previous studies have shown that the abundance of both ribosomal and mitochondrial proteins were significantly different in the nasal mucosa of CRS individuals compared to controls [130], and that morphological and functional changes in mitochondria of the airway epithelium was associated with CRSwNP pathogenesis [131]. Co-expressed genes in the turquoise module were significantly enriched in one pathway, the leptin signaling pathway ($P = 9.47x10^{-6}$; $P_{adj} =$

Table 2.2 Module genes enriched in biological pathways. The top five pathways are shown for each module.

	Module (# of Pathways P_{adj}<0.10)	Term	*P* value	Adjusted *P* value	Odds Ratio	Module Genes / Pathway Genes
RV-Response Processes	**Red** (67) Immune Response to Microbial Infection	The human immune response to tuberculosis	$6.35x10^{-19}$	$3.00x10^{-16}$	20.22	16/23
		Type II interferon signaling (IFNG)	$3.5x10^{-17}$	$8.27x10^{-15}$	14.14	18/37
		Retinoblastoma Gene in Cancer	$1.24x10^{-14}$	$1.94x10^{-12}$	7.69	23/87
		DNA IR-damage and cellular response via ATR	$1.99x10^{-11}$	$2.34x10^{-09}$	6.9	19/80
		Apoptosis	$2.34x10^{-08}$	$2.21x10^{-06}$	5.54	16/84
	Yellow (9) Lipid and Cholesterol Biosynthesis	Genes related to primary cilium development	$8.29x10^{-13}$	$3.91x10^{-10}$	3.45	39/103
		Cholesterol Biosynthesis Pathway	$9.68x10^{-10}$	$2.28x10^{-07}$	7.3	12/15
		Sterol Regulatory Element-Binding Proteins (SREBP) signaling	$5.83x10^{-07}$	$9.17x10^{-05}$	3.04	23/69
		Ciliary landscape	$1.57x10^{-04}$	$1.24x10^{-02}$	1.77	42/216
		Lipid Metabolism Pathway	$1.41x10^{-04}$	$1.33x10^{-02}$	3.46	11/29
	Black (4) Cell Proliferation	Retinoblastoma Gene in Cancer	$4.63x10^{-05}$	$2.18x10^{-02}$	2.96	17/87
		DNA Replication	$3.18x10^{-04}$	$7.51x10^{-02}$	3.61	10/42
		Breast cancer pathway	$5.01x10^{-04}$	$7.88x10^{-02}$	2.16	22/154
		EGF/EGFR Signaling Pathway	$1.00x10^{-03}$	$9.45x10^{-02}$	2.06	22/162
		Signaling Pathways in Glioblastoma	$8.88x10^{-04}$	$1.05x10^{-01}$	2.59	14/82
CRS Gene Pathways	**Blue** (10) Oxidative Phosphorylation	Cytoplasmic Ribosomal Proteins	$1.31x10^{-39}$	$6.21x10^{-37}$	12.93	45/89
		Electron Transport Chain (OXPHOS system in mitochondria)	$1.61x10^{-14}$	$3.81x10^{-12}$	6.46	26/103
		Oxidative phosphorylation	$7.66x10^{-10}$	$1.21x10^{-07}$	6.82	16/60
		Mitochondrial complex I assembly model OXPHOS system	$1.80x10^{-07}$	$1.70x10^{-05}$	5.94	13/56
		Nonalcoholic fatty liver disease	$1.68x10^{-07}$	$1.98x10^{-05}$	3.63	22/155
	Turquoise (1) Leptin Signaling	Leptin signaling pathway	$9.47x10^{-06}$	$4.47x10^{-03}$	5.1	11/76
		IL-5 Signaling Pathway	$8.62x10^{-04}$	$1.02x10^{-01}$	5.28	6/40
		MET in type 1 papillary renal cell carcinoma	$1.36x10^{-03}$	$1.07x10^{-01}$	4.18	7/59
		TNF alpha Signaling Pathway	$1.21x10^{-03}$	$1.14x10^{-01}$	3.44	9/92
		Pathways Affected in Adenoid Cystic Carcinoma	$4.90x10^{-04}$	$1.16x10^{-01}$	4.33	8/65

$4.47x10^{-3}$). Expression of leptin receptors was higher in the nasal mucosa of CRSwNP subjects compared to healthy controls in a previous study [132]. Other terms were enriched with genes from the turquoise module, which had the fewest genes compared to the other four modules, but were not significant after multiple testing correction ($P_{adj} \sim 0.10$); these included pathways previously associated with CRS pathogenesis, including the IL-5 signaling pathway [133] and the TNF alpha signaling pathway [118].

2.4.5 DNAm sites with interaction effects are enriched in the black and brown modules

Modules containing correlated DEGs and/or DMCs may reflect potential co-regulated pathways in CRS or RV response, as well as potential sites of interactions between CRS and RV. To further explore this possibility, we first assessed the distribution of the 61 DMCs with CRS x RV interaction effects among the four modules containing CpGs (Fig 2.5A). The majority of the interaction sites (86%) were in two modules: one correlated with CRS (turquoise, 32 CpGs) and one correlated with RV-treatment (black, 21 CpGs). The remaining eight were in two modules correlated with CRS (brown, 6 CpGs; blue, 2 CpGs).

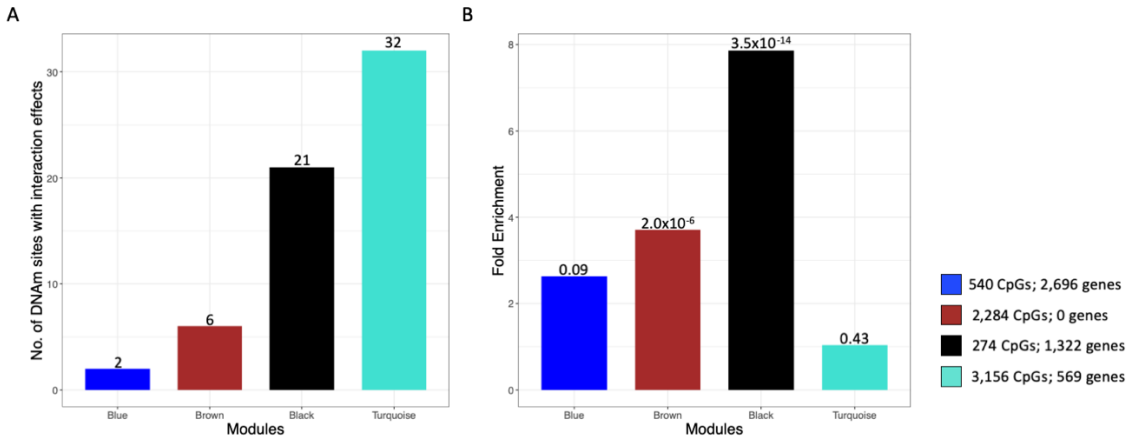


Figure 2.5 Enrichment analysis of DNAm sites with interaction effects in WGCNA modules. (A) Bar plot of the distribution of DNAm sites with interaction effects in modules with co-methylated CpG sites. The x-axis indicates the WGCNA module while the y-axis indicates the number of DNAm sites with interaction effects; the number of DNAm with interaction effects is shown above each bar. (B) Bar plot of enrichments of interaction DNA methylation

Figure 2.5 Enrichment analysis of DNAm sites with interaction effects in WGCNA modules. (continued) sites in WGCNA modules. The x-axis indicates the WGCNA module, and y-axis indicates the fold enrichment of interactive DNA methylation sites. Enrichment *P* values for each module are shown above each bar (hypergeometric test).

Two of the modules were significantly enriched for DMCs with interaction effects compared to DMCs without interaction effects ($P_{adj} \leq 0.05$), with a 3.71-fold enrichment in the CRS-correlated brown module ($P = 2.0x10^{-6}$; $P_{adj} = 8.0x10^{-6}$; hypergeometric test) and 7.86-fold enrichment in the RV-associated black module ($P = 3.5x10^{-14}$; $P_{adj} = 1.4x10^{-13}$; Fig 2.5B). The black module was also enriched for genes involved in proliferative responses, which are important features of both RV response and CRS pathogenesis. These data demonstrate that DMCs involved in CRS x RV interactions are part of correlated networks of molecular responses and underscore the importance of epigenetic mechanisms of RV responses that differ between CRSwNP cases and non-CRS controls.

2.5 Discussion

The airway epithelium serves as the first line of defense against environmental perturbations, and also functions as the initiator of downstream host responses to foreign invaders of the airway. Dysfunction of this barrier can occur through exposure to environmental irritants, including viral infections, which can trigger and promote the progression of chronic inflammatory diseases, such as CRS. By focusing on transcriptional and epigenetic marks of response to RV infection in an airway epithelial cell culture model, we were able to delineate the molecular response mechanisms to RV infection in cells from CRSwNP and non-CRS subjects. Our study highlighted an important role for epigenetic mechanisms of response to RV that differ between cases and control, and which may influence gene pathways reflecting dysregulation of epithelial cell processes in CRS.

Several lines of evidence support these conclusion. First, we observed over 4,000 DMCs between cases and controls in the vehicle treatment (Fig 2E), indicating that significant epigenetic differences are present at baseline, potentially reflecting intrinsic changes to the nasal epithelium of individuals with CRSwNP. Second, the vast majority of DMCs were included in the three WGCNA modules that were primarily associated with CRS, but not in the modules primarily associated with RV treatment. Third, the DMCs with significant CRS x RV interaction effects were enriched in TFBSs, where variation in methylation levels can influence the binding of transcription factors [134] and downstream gene expression. This indicates that these CpGs have the potential to disrupt the binding of over 100 different transcription factors, and perturb the expression their gene targets. Taken together, these findings revealed a central role for epigenetic signatures, or correlated DMCs, in CRS etiology or pathogenesis.

In addition to the important role that DNAm contributes to CRS, our systems biology approach identified co-regulatory modules of genes enriched in pathways that have been previously implicated in CRS or microbial-response mechanisms. Only two of the six modules (black, blue) showed some nominal evidence of correlation with both CRS and RV and both included genes enriched in pathways relevant to both CRS and viral infection. For example, four of the top five enriched pathways of co-expressed genes in the blue module were related to mitochondrial functions. This is consistent with the findings of Yoon et al.[131], who reported increased mitochondrial reactive oxygen species (mtROS) and changes in airway epithelial mitochondrial function and morphology in CRSwNP individuals compared to controls. They demonstrated that mitochondrial changes and mtROS generation is inducible in airway epithelial cells after exposure to *Staphylococcus aureus* enterotoxin B, a potent bacterial antigen that leads to stimulation of cytokine release and inflammation. Enhanced ROS production in the airway

epithelium also occurs during respiratory viral infections, including RV infection [135]. This could explain the link between RV-response and the enriched mitochondrial pathways among co-expressed genes in the blue module, with RV infection leading to increased ROS followed by the induction of mitochondrial gene pathways affecting CRS. Additionally, because the blue module also includes a co-methylation network of 540 CpGs, it is possible that some of these gene pathways are epigenetically regulated. Similar to the blue module, co-expressed genes in the black module were enriched in gene pathways related to RV-response and CRS pathogenesis, although they capture different molecular features. Genes in the black module were enriched in four gene pathways that were all related to cell proliferation. One of these pathways, the EGF/EGFR signaling pathway is responsible for coordinating the repair of epithelial cells by regulating cell proliferation, differentiation, and migration, and has been extensively studied in CRS [127-129]. In particular, genes in this pathway are upregulated in airway cells in several respiratory diseases, including CRS [127]. The remaining three pathways were related to cancer and have not been directly linked to CRS, however, they may be related to the cell proliferative state of the airway epithelium found in individuals with CRS and the overgrowth contributing to the development of polyps in its most severe form. In fact, the expression of apoptotic mediators (*CASP3*, *CASP9*) and a tumor suppressor gene (*TP53*) , both dysregulated in cancer, is significantly lower in the nasal epithelium from patients with CRSwNP, potentially contributing to the greater cell proliferation and the perpetual inflammatory state observed in this disease [136]. Taken together, these results indicate that RV infection may promote cell proliferation in the airway epithelium, which is a feature of CRS. Although we did not observe any differences in gene expression responses to RV infection between cases and controls for these genes, the

DMCs in the black module may differentiate disease outcomes in CRS and non-CRS individuals by promoting a chronic inflammatory and cell proliferative state in CRS-susceptible subjects.

We identified 61 DMCs with CRS x RV interaction effects that were enriched in the black and brown modules. The CRS-correlated brown module contained a co-methylation network of 2,284 DMCs but no DEGs, so directly identifying gene networks for this module was not possible. However, the enrichment of DMCs with interaction effects in the brown module suggests that this co-methylation network may indeed impact gene pathways that affect CRS. This is supported by the enrichment for these DMCs at TFBSs. The lack of co-expressed genes in the brown module may suggest that the DMCs in this module, which differentiate CRS cases and controls, may have temporal-specific effects on gene expression outside of the 48 hour treatment time point of our cell culture model. Alternatively, the DMCs in this module may represent stable, long-term epigenetic states that impact CRS onset (etiology) or progression (pathogenesis) through other mechanisms. The finding that the DMCs with interaction effects were also enriched in the RV-correlated black module further supports the notion of an epigenetic regulatory mechanism for genes enriched in cell proliferative pathways in response to RV exposure.

Using primary airway epithelial cells from CRSwNP and non-CRS individuals to model transcriptional and epigenetic responses to RV infection allowed us to make novel observations of epigenetic response differences between cases and controls. However, our study had limitations. First, our study focused on a CRS-relevant tissue in isolation from the many cell types that contribute to CRS in the sinonasal epithelium. As a result, our model only partially captured RV-responses and case-control differences that are present *in vivo*. Second, our sample sizes were small and our power to detect interaction effects (CRS x RV), especially for gene

expression, was limited. It is possible therefore, and even likely, that many more interactions would be detected in larger samples, including many among those with *P* values < 0.01 but FDRs > 0.10 in our study. Because the genes participating in those interactions were enriched in CRS-relevant pathways, it is likely that many of these represent true interactions. Despite this, however, the use of WGCNA enabled us to extract biological insights from the gene expression and DNAm data associated with CRS and/or RV infection in these samples. Finally, this cross-sectional study at a single time point does not allow us to determine whether the epigenetic patterns associated with differential response to RV in cases compared to controls preceded and contributed to the onset of disease or was a consequence of the disease state itself. Additional investigations in *ex vivo* tissues and in longitudinal studies would be needed to address this important question.

In summary, our study revealed DNAm responses to RV exposure that differed in cultured airway epithelial cells from CRSwNP patients compared to cells from non-CRS controls. Our data suggested for the first time that epigenetic mechanisms are important contributors to the development of this condition, which significantly impacts quality of life and contributes disproportionately to health care costs [137]. Focus on the transcriptional networks potentially perturbed by DNAm effects on transcription factor binding in future studies could identify novel drug targets or therapeutic modalities.

2.6 Supplementary Information

2.6.1 Supplementary Tables

Table S2.1. Enrichment Results of interaction gene expression analysis with the BioPlanet catalogue (*P* value<0.05).

Term	Overlap	P-value	Adjusted P-value	Odds Ratio	Genes
Interleukin-2 signaling pathway	12/847	4.02E-03	1	2.42	DUSP4;TRAP1;CLIC3;UCP2;CREM;PMAIP1; PEMT;SLC39A8;IL7R;SSBP2;LYST;PIGH
Phosphatidylcholine biosynthesis	2/18	4.88E-03	1	18.99	SLC44A3;PEMT
TNF-alpha effects on cytokine activity, cell motility, and apoptosis	4/135	8.18E-03	1	5.06	PMAIP1;ACKR3;IL7R;CX3CL1
Interleukin-5 regulation of apoptosis	4/144	1.02E-02	1	4.75	DUSP4;TRAP1;SDC4;P2RY2
Homologous recombination	2/29	1.24E-02	1	11.79	RAD54L;NBN
Signaling events regulated by Ret tyrosine kinase	2/39	2.18E-02	1	8.77	GAB1;PDLIM7
ATM-mediated phosphorylation of repair proteins	1/5	2.89E-02	1	34.19	NBN
Activation of NOXA and translocation to mitochondria	1/5	2.89E-02	1	34.19	PMAIP1
Ion channel and phorbal esters signaling pathway	1/5	2.89E-02	1	34.19	P2RY2
JNK/MAPK pathway	2/53	3.85E-02	1	6.45	DUSP4;GAB1
Binding of chemokines to chemokine receptors	2/54	3.99E-02	1	6.33	ACKR3;CX3CL1
Acetylcholine biosynthesis	1/7	4.02E-02	1	24.42	PEMT
Inactivation of BCL-2 by BH3-only proteins	1/7	4.02E-02	1	24.42	PMAIP1
Spermatogenesis regulation by CREM	1/7	4.02E-02	1	24.42	CREM
Recruitment of repair and signaling proteins to double-strand breaks	1/8	4.59E-02	1	21.37	NBN
Neurotrophic factor-mediated Trk receptor signaling	2/60	4.82E-02	1	5.70	RIT1;GAB1
p53 signaling pathway	3/139	4.82E-02	1	3.69	PMAIP1;PPM1J;CX3CL1
FGF signaling pathway	2/61	4.97E-02	1	5.60	SDC4;GAB1

2.6.2 Supplementary Figures

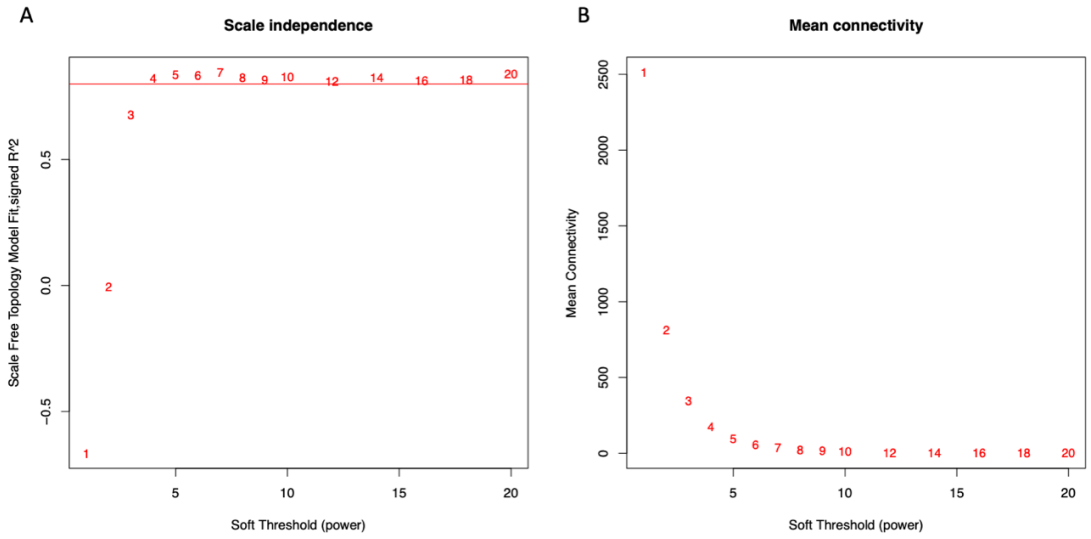


Fig S2.1. Analysis of transcriptomic and DNA methylation topology for various soft-thresholding powers. (A) The scale-free fit index (y-axis) as a function of the soft threshold power (x-axis). (B) Mean connectivity of the degree of mean connectivity (y-axis) as a function of the soft threshold power (x-axis).

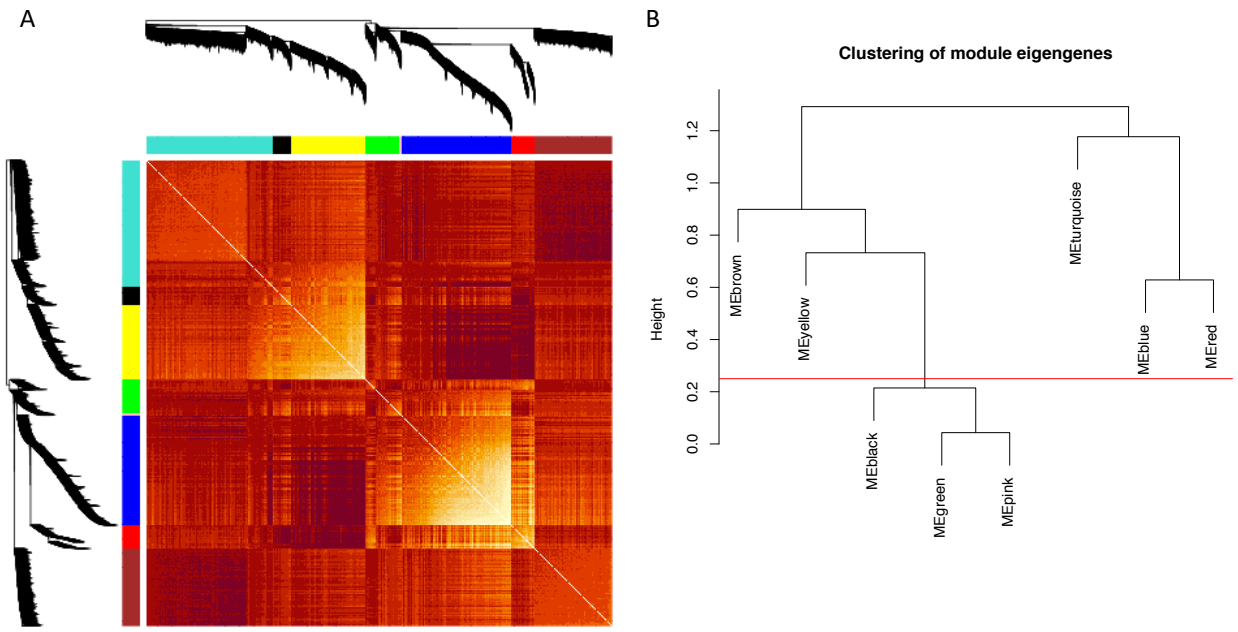


Fig S2.2. Topological overlap matrix and module clustering. (A) Visualization of the gene expression and DNA methylation network using a heatmap plot depicting the topological overlap matrix (TOM) among all the genes and CpGs in the analysis. Dark red colors represent a greater number of topological overlaps between modules and progressively lighter colors represent a lower number of overlaps. The gene expression and DNA methylation dendrogram and module assignments are also shown along the top and left sides of the heatmap. (B) Cluster dendrogram of genes and CpGs with dissimilarity based on topological overlap. The red line indicates a height cutoff of 0.25, corresponding to a correlation of 0.75. Modules below this line were merged into a single module.

CHAPTER 3

Two-stage genome-wide association study of chronic rhinosinusitis and disease sub-phenotypes highlights mucosal immunity contributing to risk

3.1 Abstract[1]

The genetic contributions to chronic rhinosinusitis (CRS) are poorly understood, likely due to low powered studies and extensive clinical heterogeneity. Whether more precise phenotyping of CRS could elucidate relevant genetic associations has not been previously investigated in a genome-wide study. The goal of this study was to identify genetic variants contributing to CRS susceptibility using precise phenotyping definitions. To this end, we performed a two-stage study. In the first stage, a genome-wide association study (GWAS) for CRS was performed using data from the Chronic Rhinosinusitis Integrative Studies Program (CRISP), including adults with CRS (483 cases) identified using survey information and electronic health records, and adults without indications of CRS (2,057 controls) from a health care system biobank. The CRS GWAS identified 82 suggestive significant SNPS ($P<1x10^{-5}$) at six independent loci. Because of the imprecision in defining CRS using surveys and medical records, latent class analysis was applied to 646 individuals with CT scans and identified three CRS subgroups: "no/mild opacification," "localized opacification," and "diffuse opacification". In the second stage, genetic association tests were performed between lead SNP at each of the six loci identified in the GWAS and the latent classes. Two of the six lead GWAS SNPs were also significantly associated with latent classes, with one of the SNPs potentially contributing to pathogen response and barrier function. Although this study is limited in power due to its reduced sample size, we demonstrate that

1. Citation for chapter: Soliai M et al. Two-stage genome-wide association study of chronic rhinosinusitis and disease subphenotypes highlights mucosal immunity contributing to risk. *International Forum of Allergy & Rhinology*. 2020. In Press.

careful phenotyping in CRS may produce insights into this complex disease which might be missed using generalized phenotyping methods.

3.2 Introduction

Chronic rhinosinusitis (CRS) is a prevalent inflammatory disease that results in over $10 billion in health care costs annually in the United States [138]. This disorder (and its most severe form, nasal polyposis [NP]) overlaps clinically with other conditions, including asthma and aspirin-exacerbated respiratory disease (AERD). Despite this enormous public health impact, the underlying molecular mechanisms remain unknown.

Evidence that CRS has a genetic component is suggested by its familial aggregation [45], and co-occurrence with genetic syndromes (e.g., cystic fibrosis, primary ciliary dyskinesia) and disorders with significant genetic components (e.g., asthma, allergic rhinitis) [43]. Recently, Kristjansson et al. reported a large meta-analysis of genome wide association studies (GWAS) of CRS (n = 5,608 cases) and NP (n = 4,366 cases) from the United Kingdom and Iceland [44]. Cases and non-CRS/non-NP controls (n > 700,000) were defined by International Classification of Disease (ICD)-10 diagnosis codes. Ten genome-wide significant loci for NP and two for CRS were identified. A missense variant in the *ALOX15* gene that causes near total loss of arachidonate15-lipoxygenase (15-LO) activity, a pathway previously implicated in AERD [139]. conferred protection against both NP (OR 0.32, 95% CI: 0.26–0.39, $P=8.0\times10^{-27}$) and CRS (OR 0.64, 95% CI: 0.55–0.75, $P=1.1\times10^{-8}$). The finding of only two genome-wide significant loci despite a large sample of CRS cases suggests that substantial clinical heterogeneity may have masked the effects of other risk loci, and that phenotyping using diagnosis codes likely captures a range of conditions with overlapping symptoms. Indeed, a GWAS of chronic sinus infection, which defined cases as subjects who reported having sinus surgery, found no genome-wide

significant associations [140], despite a sample size of 5,291 cases and 79,622 controls. Thus, alternative approaches are needed to delineate the genetic architecture of CRS.

3.3 Methods

3.3.1 Geisinger population and databases

Subjects were selected from 200,769 adult primary care patients from Geisinger, a health system that serves 45 counties in Pennsylvania. The Geisinger primary care population is representative of the general population in the region. Subjects were selected from this patient pool based on diagnostic codes in the electronic health record (EHR) for chronic rhinosinusitis (CRS) and asthma/allergy, and questionnaires were sent to a random sample of 23,700 people, enriched for ethnic minorities and those with certain diagnostic and therapeutic codes [101]. The 7,847 who returned the questionnaire constituted the Chronic Rhinosinusitis Integrative Study Program's (CRISP) study cohort. Rich longitudinal EHR data, four seasonal follow-up questionnaires are available for these subjects in addition to a questionnaire at the time of CT scan. DNA from blood and saliva for a subset of samples in the Geisinger biobank, a MyCode Community Health Initiative, are available for these subjects. CRS was defined by questionnaire responses based on the European Position Paper on Rhinosinusitis (EPOS) epidemiologic criteria [141]: nasal discharge (anterior or posterior) or nasal obstruction, most or all of the time for at least three months, plus at least one other symptom of that frequency and duration (nasal discharge, nasal obstruction, smell loss, facial pain, or facial pressure). Among the CRISP cohort, 1,866 had current CRS (as defined by EPOS in the prior three months) and 2,093 had past CRS (as defined by EPOS in their lifetime prior to three months). Of these individuals with CRS (ever): (1) 1,499 had a MyCode consent and a DNA sample in the Geisinger biobank; opt-out letters were mailed to these individuals prior to using their sample. (2) 372 had MyCode consents but lacked a DNA

sample; these individuals were mailed a consent for the CRISP genetics study with post cards requesting a saliva sample. DNA was collected from 126 of the 372 individuals via saliva collection kits that were mailed to participants (79 samples) or from whole blood (47 samples) that were collected in the interim. (3) Individuals that signed-up for MyCode were sent out mailing in batches, prioritizing those with current CRS, participation in the CT study, or had longitudinal data, until the study sample size was attained. 204 individuals returned consents and 81 were prioritized and mailed DNA kits, of whom 56 returned the kit. Overall, 1,661 of the 1,681 individuals with CRS (ever) had high-quality DNA and were genotyped at the University of Chicago. This subset of the CRISP cohort (CRS by EPOS criteria + genotypes) were the source population for the CRS GWAS cases. The GWAS cases were those with CRS as defined by EPOS epidemiologic criteria of European American ancestry, and who had objective evidence of disease on sinus CT determined by an otorhinolaryngologist (ENT; 127 individuals) or had self-reported sinus surgery for CRS or nasal polyposis (NP) in any one of four CRISP surveys (at baseline, in the six- or 16-month surveys or the CT survey; 3 individuals), had evidence of prior sinus surgery on a sinus CT or had Lund-Mackay (LM) score of ≥ 3 on sinus CT (91 individuals).

Controls for the GWAS were derived from the MyCode Community Health Initiative [142], which includes a Geisinger-system-wide biobank of participants who were genotyped as part of the DiscovEHR cohort, a Geisinger-Regeneron collaboration. Controls were selected from this larger cohort based on the following criteria: European American ancestry and absence of (i) an ICD-9 code for NP, CRS, asthma, or allergic rhinitis, and (ii) a Current Procedural Terminology (CPT) code for sinus CT scan, endoscopy, or surgery in the EHR. Twenty-five of

these controls were also part of the CRISP cohort who never had CRS. These samples are described in Table S3.1.

Among the individuals who completed the baseline CRISP questionnaire (described above), 3,269 subjects were invited for sinus CT scans by stratified random sampling to enrich for those with sinus symptoms [143]. The strata were based on the presence of CRS symptoms as defined by EPOS (CRS) [141] and latent class analysis (LCA) categories on symptoms reported in the questionnaire [144]. Of these subjects, 646 completed the sinus CT scans and LCA was performed again on their sinus CT findings, and subjects were classified into three CT-defined subgroups based on patterns of observed sinus opacification: no/mild opacification LCA class, localized LCA class, and diffuse LCA class. Of the 646 LCA-classified subjects, 341 also had genetic information. Of the 341 with both sinus CT and genetic information, 339 had current or past CRS based on symptom criteria (similar to the case definition in the GWAS). Genetic association studies using the sentinel SNP at each of six suggestive significant loci in the GWAS were performed on the 339 subjects assigned to the latent classes. The LCA cases were those with CRS by epidemiologic definition AND localized LCA class (n = 71) or diffuse LCA class (n = 78), while the LCA controls had CRS by epidemiologic definition but no/mild opacification LCA class (n = 191). In order to differentiate subjects in the two stages of analysis with different sets of cases and controls, the first stage of GWAS is described using the terms GWAS cases and controls, and the second stage of the study using LCA cases and controls; 51% of those in the LCA genetics analysis were also GWAS cases (Fig 3.1).

3.3.2 Genotyping and Imputation

The GWAS cases and LCA cases and controls were genotyped at the University of Chicago using the Human Core Exome Array (550,224 SNPs). The controls were genotyped by

the Geisinger Biobank using either the Illumina Human Core Exome (533,456 SNPs) or the Human Omni Express Arrays (731,306 SNPs). Quality control (QC) was performed for cases and controls separately for each of the genotyping platforms (excluding SNPs with HWE < 0.00001, call rate < 0.95, and MAF = 0). Samples from each of the three platforms were then merged and QC checks were applied in the pooled sample, as was previously performed for each of the genotyping platforms. After this QC, 222,149 overlapping markers for 3,718 individuals were available for analysis. Ancestral principal components analysis (PCA) was performed using 649 ancestral informative markers that overlapped with HapMap release 3 and were included on all of the genotyping arrays.

Phasing and imputation were performed in the merged dataset using the ShapeIt2 [145] and Impute2 [146] software packages, respectively. Variants were imputed in 5 Mb windows across the genome against the 1000 Genomes Phase 3 haplotypes (Build 37; October 2014). Individuals were grouped as European based on how they related to the HapMap reference panel and the k-means clustering of ancestral PCs, using the kmeans [147] function in R, to confirm that all subjects were of European ancestry. After imputation, additional QC steps were performed with gtool [148], including the exclusion of X and Y chromosome-linked SNPs and variants with info score < 0.8, MAF < 0.05, missingness > 0.05 and a probability score < 0.9. Probability scores were converted to dosages for 10,436,943 SNPs used in downstream analyses.

3.3.3 Genome-Wide Association Studies

A GWAS was performed using the Genome-wide Efficient Mixed Model Association (GEMMA) software toolkit [149], fitting a multivariate linear mixed model [150] to test for marker association with CRS while controlling for population stratification. A standardized relatedness matrix was estimated from the genotypes using 0.02 missingness and 0.05 MAF

thresholds. A multivariate linear mixed model was applied using age, sex, ancestry (ancestry PCs 1-3), and relatedness (standardized relatedness matrix) as covariates, and 0.02 missingness and 0.05 MAF thresholds for 10,436,943 SNPs. In total, 4,528,849 SNPs were analyzed in 483 CRS cases and 2,057 non-CRS controls of European descent (Table S3.1). P-values from the likelihood ratio test are reported. The most significant SNP at each of six suggestive significant loci from this GWAS was referred to as the sentinel SNP, and used in the second-stage LCA association study.

3.3.4 Latent class analysis on sinus CT

LCA was used to create unobservable subgroups based on patterns of observed variables in the sinus CT ENT reads. From the LM score on the sinus CT reported by ENT, binary indicators were created (0 and ≥ 1) and included in a series of LCA models fit with an increasing number of classes. Appropriateness of model fit was assessed using multiple fit statistics. The three-class model had the best fit including the "no/mild opacification," "localized opacification," and "diffuse opacification" classes. The LCA analysis was performed using Mplus v.8.1 (Muthén & Muthén, Los Angeles, CA). Traditional LM score is in a linear scale and there are no universally accepted cut-offs for clustering. However, LCA identified subgroups in the sinus CT analysis that likely have clinical significance, making it a good phenotype for this study (Table S3.2).

3.3.5 Statistical analysis for the second-stage

The aim of the LCA genetic analysis was to study the associations of the sentinel SNPs at GWAS loci with surgical CRS and CRS evidence on CT or objective evidence of CRS (CRSo) with the outcome of LCA CT classes. We used chi-square tests to compare the proportion of individuals with different categories of LCA CT by sex, race/ethnicity, smoking, Medical

Assistance, health insurance status (a surrogate for family socioeconomic status), subject reported physician-diagnosed allergic rhinitis and asthma, atopy, migraine headache [101], anxiety sensitivity index (ASI) (below median vs. at or above median), and history of sinus surgery. Analysis of variance was used to examine associations of CRS CT subgroups with continuous variables like age and LM score. Base model-building was performed by adding the above predictors individually to the model and retaining those with significant associations. The final base model included surgical status, sex and physician-diagnosed subject-reported allergic rhinitis. Multinomial logistic regression was used for the associations of genetic data and LCA CT classes, adjusted for base model variables in an additive model with p-values corrected for six tests. All analyses were performed using Stata 15.1 (StataCorp LLC, Texas, USA).

3.4 Results

3.4.1 CRS GWAS

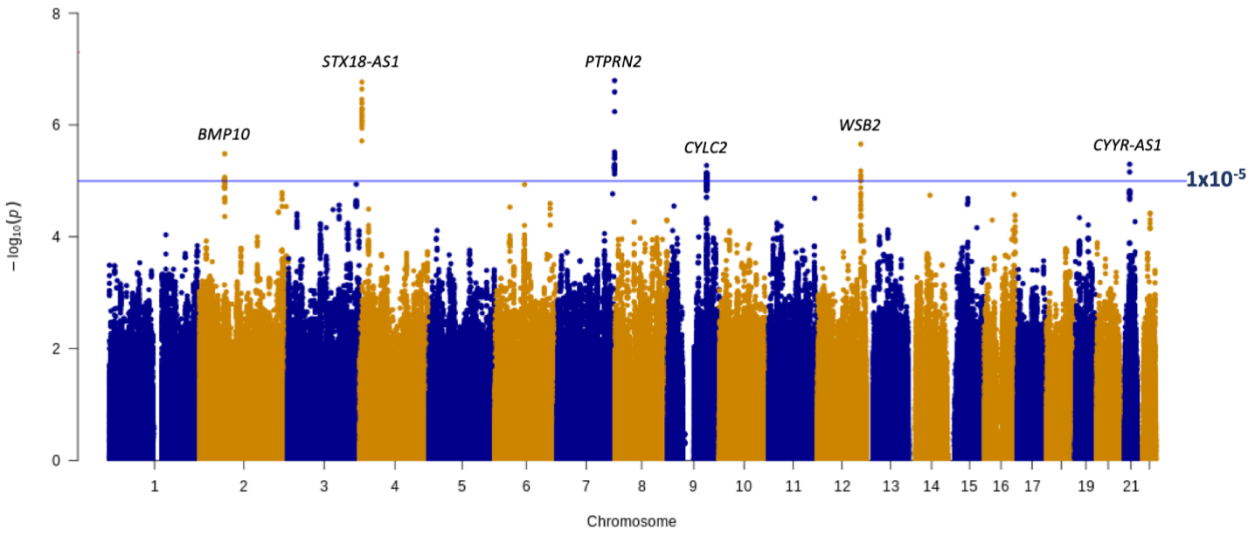
To maximize both the sample size and the detailed phenotyping available, we performed a two-stage analysis. In the first stage, we included all 483 cases and 2,057 controls in a CRS GWAS. This revealed 82 suggestive significant SNPs ($P < 1x10^{-5}$) at six loci (**Fig 3.1A**). The sentinel SNP at each of the six loci is described in **Fig 3.1B**. A SNP in linkage disequilibrium (LD; $r^2 > 0.9$; GRB; 1000 Genomes Project) with one of the two SNPs associated with CRS and NP in the Kristjansson et al. study [44] was not associated with CRS in our GWAS ($P > 0.90$); the second SNP associated with CRS and NP in that study was not imputed or in LD with SNPs in ours.

3.4.2 Association studies of sentinel GWAS SNPs and latent classes

Because defining CRS by medical records may be imprecise, we defined CRS sub-phenotypes in the 646 CRISP subjects with sinus CT scans, including 172 of the cases included in the GWAS,

applying latent class analysis (LCA) to create subgroups based on patterns of sinus opacification using Mplus v.8.1 [151], to identify clinical and immunopathogenic subgroups of CRS [152]. We identified three latent classes, which we referred to as "no/mild opacification," "localized opacification," and "diffuse opacification" (see Table S3.2).

A



B

Chromosome	rsID	Position*	Nearby genes†	Allele‡	AF	OR	P value	eGene	QTL Tissue
2p13.3	rs3833578	2:69088930	LINC01888, BMP10	TA/T	0.061	0.3733	3.22×10^{-06}	-	-
4p16.2	rs6818775	4:4648918	STX18-AS1, STX18, LOC01928279	A/G	0.161	0.4854	1.68×10^{-07}	-	-
7q36.3	rs3935208	7:157843523	PTPRN2, LOC100506585, MIR595	C/A	0.547	1.680	1.58×10^{-07}	-	-
9q31.1	rs10820254	9:105488979	LINC00587, CYLC2	G/A	0.303	0.6084	5.21×10^{-06}	CYCL2	Testis
12q24.23	rs11068795	12:118495743	WSB2, RFC5, VSIG10	C/A	0.238	1.737	2.15×10^{-06}	WSB2	Skeletal Muscle, Skin
21q21.3	rs2830160	21:27651254	MARCKSP1, CYYR1-AS1	G/A	0.140	0.5224	4.96×10^{-06}	AP000230.1	Brain, Adrenal Gland

* SNP position according to Genome Reference Consortium Build 37 (hg19)
† The gene in which the SNP is located is indicated first followed by the genes located upstream and downstream, respectively; for intergenic SNPs, only the upstream and downstream genes are included.
‡ Alleles are shown as risk/non-risk alleles, where the risk allele is associated with increased risk for CRS.

Figure 3.1 CRS GWAS results. (A) Manhattan plot showing chromosome positions along the x axis and -log(*p* value) on the y axis. The horizontal blue line indicates suggestive significance ($p < 1 \times 10^{-5}$). The closest gene to the sentinel SNP at each locus reaching suggestive significance is shown. (B) Description of the 6 sentinel SNPs in regions reaching suggestive significance. CRS = chronic rhinosinusitis; GWAS = genome-wide association studies; SNP = single-nucleotide polymorphism.

In the second stage, we performed association studies between the six sentinel GWAS SNPs (Fig 3.1B) and the latent classes (Fig 3.2). One SNP at the chromosome 12q24.23 locus (rs11068795) was associated with the localized opacification class phenotype ($P_{corrected} = 0.002$) compared to

no/mild opacification, but not with the diffuse opacification class phenotype ($P_{corrected} = 1.0$), after Bonferroni correction for six tests. Another SNP at the chromosome 9q31.1 locus (rs10820254) was associated with the diffuse opacification class phenotype ($P_{corrected} = 0.048$) but not the localized opacification class phenotype ($P_{corrected} = 1.0$). Thus, defining phenotypes based on imaging of inflammatory characteristics in the sinuses highlighted and validated two loci that were associated with CRS in the first-stage GWAS, despite the significantly reduced sample size.

A

	Localized Opacification			Diffuse Opacification		
SNP (Additive Model)	**RRR (95% CI) n=71**	**P-value**	**P-value$_{corr}$**	**RRR (95% CI) n=78**	**P-value**	**P-value$_{corr}$**
rs3833578	1.67 (0.65, 4.26)	0.29	1.00	2.69 (0.88, 8.18)	0.08	0.48
rs6818775	0.88 (0.53, 1.48)	0.64	1.00	1.60 (0.86, 2.98)	0.14	0.84
rs3935208	1.02 (0.68, 1.52)	0.93	1.00	1.28 (0.84, 1.95)	0.25	1.00
rs10820254	0.78 (0.51, 1.19)	0.25	1.00	1.96 (1.19, 3.22)	0.008	0.048
rs11068795	2.14 (1.31, 3.50)	0.002	0.012	1.25 (0.74, 2.12)	0.40	1.00
rs2830160	1.47 (0.80, 2.70)	0.21	1.00	2.09 (1.08, 4.07)	0.03	0.18

B

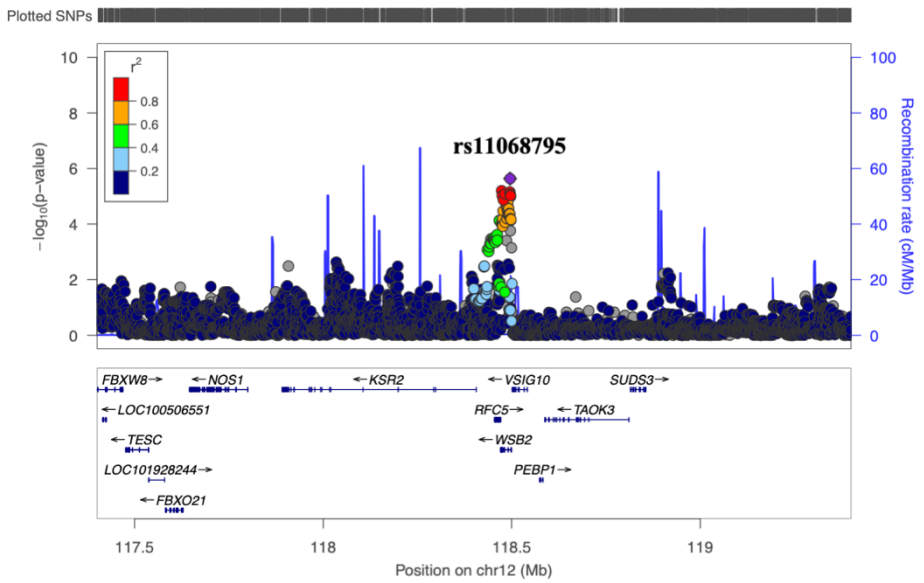


Figure 3.2 Associations of sentinel SNPs with LCA-defined subtypes. (A) Associations between 2 LCA-defined subtypes based on computed tomography scans. Each subtype is compared with the "no/mile opacification" group (n = 191). The *p* values are corrected for 6

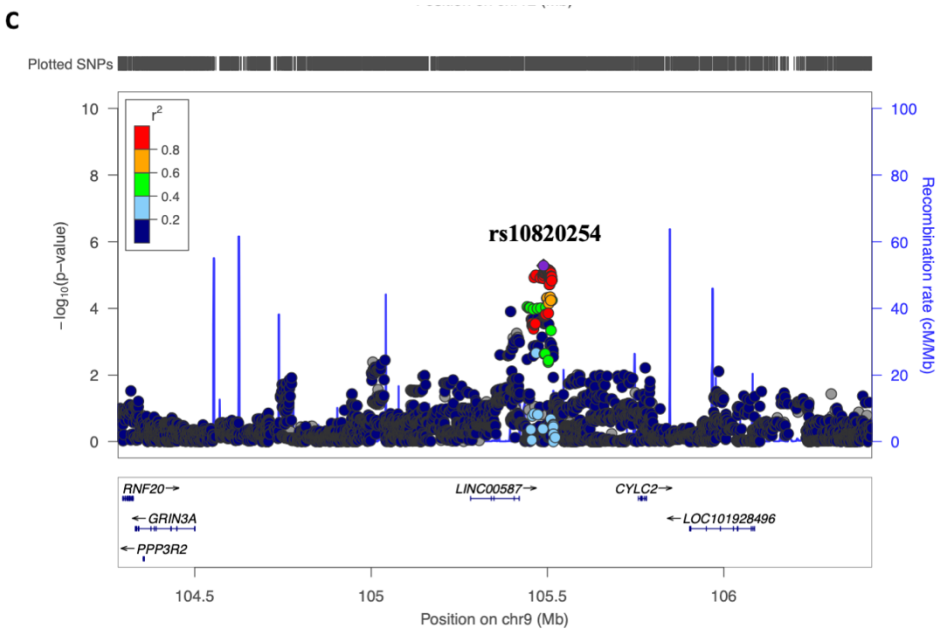


Figure 3.2 Associations of sentinel SNPs with LCA-defined subtypes (Continued). tests. Rows show the sentinel SNPs at each suggestive significant GWAS locus, and its association with each of the subtypes. Results of multinomial logistic regression models, including surgical status, sex, and physician-diagnosed patient-reported allergic rhinitis as covariates, are shown for an additive model. Significant SNPs and results are shown in bold.
The LocusZoom plots for the 2 significant SNPs in (**A**) are showin in (B) (rs11068795) and (C) (rs10820254). The SNPs are shown in a ±1-Mb window with chromosome and gene positions shown on the x axis and the -$\log_{10}$ of the *p* value and the recombination rate (CMMb) are shown on the left and right y axis, respectively. SNPs are colored according to the estimated linkage disequilibrium (r^2) relative to each lead SNP (purple diamond) and calculated using the 1000 Genome European panel. CI = confidence interval; LCA = latent class analysis; RRR = relative risk ratio.

The SNP associated with the localized opacification class phenotype, rs11068795, has been reported as an eQTL for the *WSB2* gene in skin ($P = 2.3\text{x}10^{-7}$) and skeletal muscle ($P = 7.1\text{x}10^{-17}$) [153]. WSB2 functions in the proteasomal degradation of target proteins as a negative regulator of the IL-21 receptor (IL-21R), a molecule associated with microbe-related epithelial conditions (e.g., inflammatory bowel disease [154] and *Helicobacter pylori* infection [155]), which are themselves associated with CRS [156, 157]. IL-21R expression is also upregulated in atopic dermatitis patients [158], who manifest *Staphyloccocus aureus* colonization barrier

dysfunction. This is consistent with previous work showing upregulation of IL21 in CRS with NP [159].

The SNP associated with the diffuse opacification class phenotype, rs10820254, is an intergenic SNP located between the *CYLC2* gene and a gene encoding a long intergenic non-protein coding RNA, *LINC00587*. *CYLC2* is a sperm-specific cyclin 2 and LINC00587 is expressed most highly in testis [153]. Thus, rs10820254 may contribute to CRS pathobiology through long range interactions with the promotors of distal genes, which are yet to be identified.

3.5 Discussion

A possible constraint of this study is that the initial diagnosis of CRS may be prone to provider errors, and although individuals included in this study self-reported a diagnosis of CRS were based on EPOS criteria, these errors may have an impact on our analyses. While this may be a limitation of this study, we note that this method of phenotyping has been used successfully in GWAS of related condition (e.g. asthma).

In summary, using a two-stage population approach, we identified variation at two loci potentially involved in CRS, including a connection to mucosal immunity (pathogen response and barrier function) in the localized disease subgroup, supporting a role for the upper airway mucosal immune surface as a key site in disease development. Lastly, inclusion of subjects with careful phenotyping in genetic studies of CRS may generate insights into the underlying molecular mechanisms of this complex disease and identify potential therapeutic targets.

3.6 Supplementary Information

3.6.1 Supplementary Tables

Table S3.1. Demographic and clinical information for GWAS and LCA subjects.

Characteristic	GWAS source dataset (n = 1661)	GWAS cases (n = 483)	GWAS controls (n = 2057)	No/mild opacification LCA (n = 190)	Localized opacification LCA (n = 71)	Diffuse opacification LCA (n = 78)
Age in years as of 4-2014, mean (SD)	55.40 (14.02)	56.63 (13.51)	57.35 (16.68)	55.93 (12.05)	58.37 (12.91)	58.46 (12.98)
Male, n (column %)	531 (31.97)	204 (42.24)	462 (22.46)	39 (20.53)	26 (36.62)	40 (51.28)
*Self-reported race, n (column %)						
White	1622 (97.65)	477 (98.76)	2054 (99.85)	183 (96.32)	71 (100.00)	76 (97.44)
Hispanic	39 (2.35)	6 (1.24)	3 (0.15)	7 (3.68)	0	2 (2.56)
Smoking, n (column %)						
Current	248 (14.93)	59 (12.22)	313 (15.22)	16 (8.42)	6 (8.45)	8 (10.26)
Former	513 (30.89)	142 (29.40)	662 (32.18)	51 (26.84)	25 (35.21)	31 (39.74)
Never	900 (54.18)	282 (58.39)	1082 (52.60)	123 (64.74)	40 (56.34)	39 (50.00)
BMI kg/m^2, mean (SD)	31.48 (7.60)	31.45 (7.11)	31.07 (7.50)	32.08 (7.95)	31.04 (6.84)	32.18 (6.60)

*All individuals that were included in the GWAS clustered with Europeans according to the k-means clustering of the ancestral principal components and were therefore combined for the analysis.
Abbreviations: BMI = Body Mass Index; GWAS = Genome-Wide Association Study; LCA = Latent Class Analysis; SD = Standard Deviation

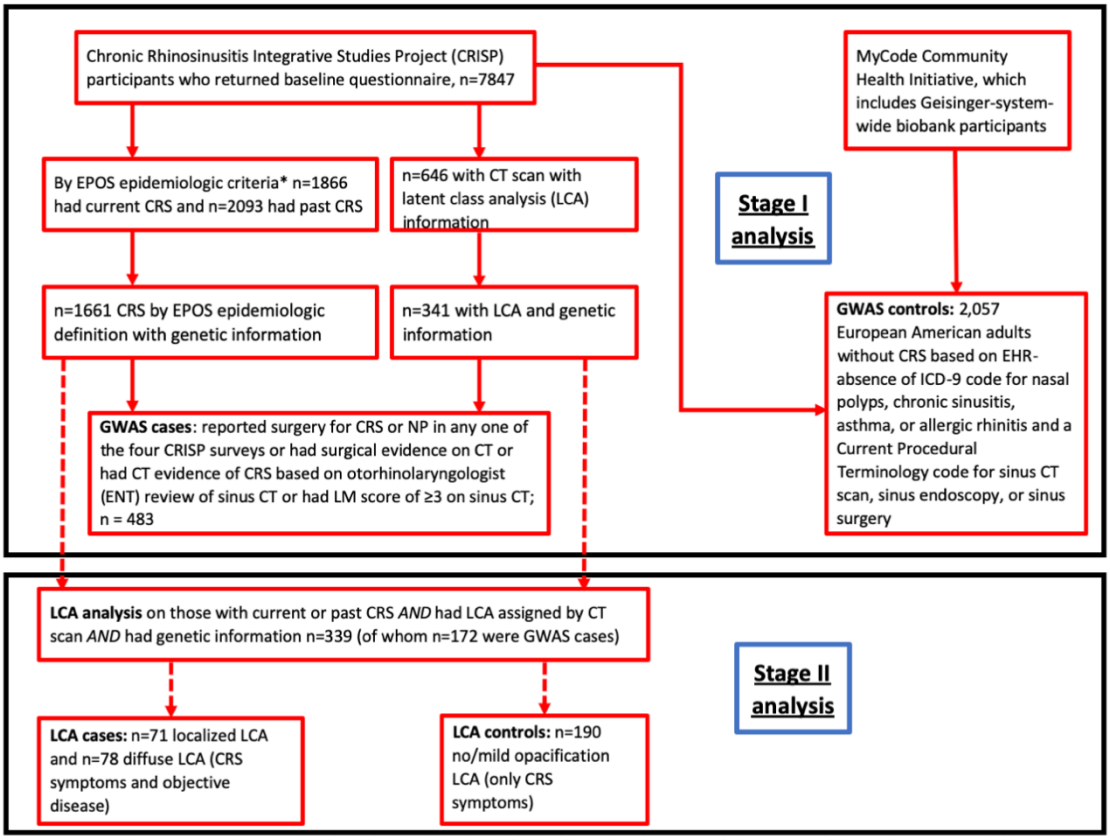
Table S3.2. Clinical profile of the LCA subjects.

Characteristic	No/mild opacification (n =190)	Localized opacification (n= 71)	Diffuse opacification (n= 78)	P-values* for LCA
Age in years, mean (SD) at time of CT	58.20 (12.04)	60.57 (12.91)	60.65 (12.92)	0.21
Males, n (column %)	39 (20.53)	26 (36.62)	40 (51.28)	<0.001
White, n (column %)	183 (96.32)	71 (100.00)	76 (97.44)	0.26
Smoking, n (column %)				
Current	16 (8.42)	6 (8.45)	8 (10.26)	
Former	51 (26.84)	25 (35.21)	31 (39.74)	0.22
Never	123 (64.74)	40 (56.34)	39 (50.00)	
Medical assistance, n (column %)	22 (11.58)	7 (9.86)	9 (11.54)	0.92
Asthma self-reported physician diagnosis, n (column %)	65 (34.21)	20 (28.17)	32 (41.03)	0.26
Allergic rhinitis- self-reported physician diagnosis, n (column %)	98 (51.58)	46 (64.79)	50 (64.10)	0.06
Atopy (self-reported skin prick positivity), n (column %)	82 (43.16)	35 (49.30)	29 (37.18)	0.33
Migraine, n (column %)	78 (41.05)	23 (32.39)	20 (25.64)	0.046
Anxiety sensitivity index ASI score (below median vs. at or above), n (column %)	76 (48.72)	25 (44.64)	36 (55.38)	0.48
Original LM score, mean (SD)	0.22 (0.53)	1.75 (1.09)	7.14 (3.41)	<0.001
Surgery on CT, n (column %)	29 (15.26)	11 (15.49)	34 (43.59)	<0.001

*p-values < 0.05 indicates at least one latent class significantly differs from the others with respect to the selected characteristic.

Abbreviations: CT = Computed Tomography; LCA = Latent Class Analysis; LM = Lund-Mackay; SD = Standard Deviation

3.6.2 Supplementary Figures



*European Position Paper on Rhinosinusitis (EPOS) epidemiological definition for CRS: At least two of six symptoms (obstruction, post-nasal drip, posterior nasal drainage, facial pain, facial pressure, smell loss) most or all of the time for at least three months, where at least one of the symptoms was obstruction, anterior nasal discharge, or posterior discharge

Abbreviations: CRS= chronic rhinosinusitis; CT = computed topography; EHR = electronic health record; GWAS = genome-wide association study; ICD = International Classification of Diseases; LM= Lund-Mackay; NP= Nasal polyposis

Figure S3.1. Flowchart of CRS case and non-CRS control subject inclusion in analyses performed in this study

CHAPTER 4

Multi-omics co-localization with genome-wide association studies reveals a context-specific genetic mechanism at a childhood onset asthma risk locus

4.1 Abstract[1]

Genome-wide association studies (GWASs) have identified thousands of variants associated with asthma and other complex diseases. However, the functional effects of most of these variants are unknown. Moreover, GWASs do not provide context-specific information on cell types or environmental factors that affect specific disease risks and outcomes. To address these limitations, we used an upper airway epithelial cell culture model to assess transcriptional and epigenetic responses to an asthma-promoting pathogen, rhinovirus (RV), and provide context-specific functional annotations to variants discovered in GWASs of asthma. Using genome-wide genetic, gene expression and DNA methylation data in vehicle- and RV-treated upper airway epithelial cells (AECs) from 104 individuals, we mapped *cis* expression and methylation quantitative trait loci (*cis*-eQTLs and *cis*-meQTLs, respectively) in each condition. A Bayesian test for co-localization between AEC molecular QTLs and adult onset and childhood onset GWAS variants was used to assign function to variants associated with asthma. We used Mendelian randomization to demonstrate DNA methylation effects on gene expression at asthma colocalized loci. Co-localization analyses of airway epithelial cell molecular QTLs with asthma GWAS variants revealed potential molecular mechanisms of asthma, including QTLs at the *TSLP* locus that were common to both exposure conditions and to both childhood and adult onset asthma, as well as QTLs at the 17q12-21 asthma locus that were specific to RV exposure and childhood onset asthma, consistent with clinical and epidemiological studies of these loci. This

1. Citation for chapter: Soliai MM et al. Multi-omics co-localization with genome-wide association studies reveals context-specific mechanisms of asthma risk variants. bioRxiv (2020).

study provides information on functional effects of asthma risk variants in airway epithelial cells and insight into a disease-relevant viral exposure that modulates genetic effects on transcriptional and epigenetic responses in cells and on risk for asthma in GWASs.

4.2 Introduction

Since the first asthma genome-wide association study (GWAS) in 2007 [27], over 150 asthma susceptibility loci at genome-wide levels of significance ($p<5x10^{-8}$) have been reported [11, 13, 80, 160], with a locus at 17q12-21 being the most replicated and most significant asthma susceptibility locus in childhood onset asthma (reviewed in [28]). Although asthma is typically diagnosed based on clinical symptoms, such as wheeze, cough, and shortness of breath, it is likely comprised of many overlapping sub-phenotypes with underlying distinct endotypes that have shared as well as unique genetic and environmental risk factors. For example, individuals with asthma differ with respect to age of onset, environmental triggers of exacerbations, response to medications, obesity, and co-occurrence with allergic diseases and other conditions. Recently, Pividori et al. reported 61 independent asthma loci, 23 of which were specific to childhood onset asthma, one that was specific to adult onset asthma, and 37 that were associated with risk for both childhood onset and adult onset asthma [13]. Gene and tissue enrichment patterns at these risk loci suggested epithelial cells (skin) and lung as the primary etiological drivers of childhood onset and adult onset asthma, respectively, while blood (immune) cell gene expression enrichments were shared by both. However, GWASs do not generally consider tissue- or other environment-specific effects, or gene by environment interactions. Moreover, most genome-wide epigenetic studies of exposures (e.g. [86, 161-164]) or of asthma-related traits (e.g. [165-170]) have not integrated their findings with GWAS. Only a few studies have formally integrated asthma GWAS results with epigenetic studies in airway tissues (e.g. [171-173]).

A challenge in interpreting GWAS results and prioritizing candidates for further studies is that over 90% of disease-associated variants are located in non-protein-coding regions of the genome [174]. GWAS-associated variants are enriched for chromatin signatures suggestive of enhancers [73, 174] and for expression quantitative trait loci (eQTLs) [72, 74, 75], suggesting that some of these SNPs are likely to be causal variants, underlying disease pathophysiology through their effects on gene regulation. However, identifying specific causal variants and their target genes at associated loci has been challenging, and the functions of most SNPs associated with diseases in GWASs remain unknown. While databases such as GTEx, ENCODE, and ROADMAP have been used to annotate GWAS SNPs and predict molecular mechanisms through which risk variants affect disease phenotypes [75-78] and provide important insights into the interpretation of GWAS results, they do not include all cell types relevant to all diseases or information on environmental exposures that influence disease outcomes. As a result, annotations of asthma GWAS variants have been largely limited to studies in transformed B cell lines, blood (immune) cells, and whole lung tissue (e.g. [11, 13, 27, 79-81]).

In vitro cell models provide an opportunity to address these limitations by characterizing genetic and molecular responses to environmental exposures in cells from disease-relevant tissues, and identifying genotypes that modify these responses [175, 176]. *In vitro* functional studies of airway epithelium have been used to characterize gene pathways affected by environmental disease modifiers of asthma (e.g., refs. [84-87]), and to identify functional variants that contribute to asthma pathogenesis in *ex vivo* airway epithelial cells, including eQTLs [177-180] and methylation QTLs (meQTLs; e.g., refs. [86, 165, 171]). However, no studies to date have formally integrated molecular QTLs in airway epithelial cells exposed to different conditions with asthma GWAS results. Joint analysis of datasets (e.g. eQTLs/meQTLs

and GWASs) can identify variants associated with both disease risk and molecular traits as candidate causal variants that contribute to mechanisms of disease pathophysiology. To this end, a multi-trait co-localization method (*moloc*) [181] was recently developed to integrate summary data from GWAS and multiple molecular QTL datasets and identify candidate regulatory drivers of complex phenotypes.

Here, we report the results of a multi-omics co-localization study to identify condition-specific regulatory effects of asthma risk variants using an epithelial cell model of viral response. Because airway epithelium forms a barrier to inhaled exposures, we used an *in vitro* upper airway epithelial cell model of transcriptional and epigenetic responses to rhinovirus (RV). Primary infection of RV occurs in the nasal epithelium, and RV is a major contributor to asthma inception in young children [182] and asthma exacerbations throughout life [51, 183], underscoring its importance as a contextual promoter of asthma pathophysiology. We first demonstrate a specific enrichment of childhood onset asthma GWAS SNPs among airway epithelial eQTLs, consistent with the important role that the epithelial barrier plays in the inception of asthma in childhood [13, 184, 185]. Then, using an integrative multi-omics approach we describe an environment-specific mechanism of asthma pathogenesis at the 17q12-21 asthma locus in childhood onset asthma, and a molecular mechanism shared between childhood onset and adult onset GWASs in the *TSLP* gene at chromosome 5q22, highlighting central roles of the airway epithelium in the pathogenesis of asthma.

4.3 Methods

4.3.1 Ethics statement

Study participants were recruited between March 2012 and August 2015 [101], and nasal specimens were collected as part of routine endoscopic sinonasal surgeries at Northwestern University Feinberg School of Medicine. Informed written consent was obtained from each study participant and randomly generated ID codes were assigned to all samples thereby preserving the participant's anonymity and privacy. This study was approved by the institutional review boards at Northwestern University Feinberg School of Medicine and the University of Chicago.

4.3.2 Sample collection and composition

Airway epithelial cells were obtained as part of the Chronic Rhinosinusitis Integrative Studies Program (CRISP) [101]. Because of the relevance of the airway epithelium in RV infection and asthma, we leveraged the genomics information collected from these cells to gain a functional understanding of asthma risk loci that are also associated with RV infection. Sinonasal epithelial cells were obtained by brushing the uncinate process collected at elective surgery for CRS or other unrelated indications (adenoidectomy, dentigerous cysts, septoplasty, and tonsillectomies) at Northwestern University from 63 males, 41 females, ages 18 – 73 years old (mean age 44), and self-reported ethnicities as Caucasian (64%), Black (17%), Hispanic (13%), and more than one ethnicity (6%). Samples were collected from 49 CRS and 55 non-CRS subjects, which included 27 and 16 asthmatics (doctor diagnosed current or prior asthma). Study participants were determined to have CRS if they met the European Position Paper on the Primary Care Diagnosis and Management of Rhinosinusitis and Nasal Polyps (EPOS) criteria [141]. DNA from blood was used for genotyping. A summary of the study design and sample composition is shown in Fig S4.1.

4.3.3 Upper airway epithelial cell culture and RV treatment

Nasal epithelial cells were collected from the uncinate tissue with a Rhinoprobe. After isolation, cells were washed with Dulbecco's phosphate-buffered saline (dPBS) and immediately cultured in a 6 well plate in bronchial epithelial cell growth medium (Lonza, BEGM BulletKit, catalog number CC-3170) to near confluence, and then frozen at -80°C for no more than three days and finally stored in Liquid Nitrogen between 8 and 1,075 days. Cells were subsequently thawed and cultured in collagen-coated (PureCol, INAMED BioMaterials, catalog number 5,409, 3 mg/mL, 1:15 dilution) tissue culture plates (6 wells of 2x 12 well plates) using BEGM overnight at 37°C and 5% CO_2. In preparation for rhinovirus (HRV-16; RV) infection/stimulation, plates at 50-60% confluency were incubated overnight in BEGM without hydrocortisone (HC) followed by a two-hour RV infection at a multiplicity of infection (MOI) of 2 and vehicle treatment (Bronchial epithelial cell basal medium (BEBM) + Gentamicin/Amphotericin) at 33°C (low speed rocking, ~15 RPM). RV- and vehicle-treated cells were washed and then were cultured at 33°C for 46 hours (48 hours total) in BEGM without HC. Cell viability was determined based on trypan blue staining or a LDH assay. Paired samples were excluded if the viability of the vehicle-treated cells exceeded 90% or there was over 30% cell death after RV treatment. Prior to our studies, we calculated the MOI as follows. A plaque forming assay was first performed using HeLa cells to generate plaque forming units (PFU). The average number of airway epithelial cells (from 10 donors) was further used to generate the MOI by testing dose-dependent responses by RV treatment at MOI ranges between 0.2-10. A lower range of 2 was within the MOI range for acceptable linearity, and was used for our studies.

4.3.4 Genotyping and imputation

DNA was extracted from whole blood or sinus tissue (if no blood was available) with the Macherey-Nagel NucleoSpin Blood L or NucleoSpin Tissue L Extraction kits, respectively, and quantified with the NanoDrop ND1000. Genotyping of all study participants was performed using the Illumina Infinium HumanCore Exome+Custom Array (550,224 SNPs). After quality control (QC) (excluding SNPs with HWE < 0.0001 by race/ethnicity, call rate < 0.95, and individuals with genotype call rates < 0.05), 529,993 markers with minor allele frequency ≥0.05 were available for analysis in 104 individuals. Ancestry principal component analysis (PCA) was performed using 676 ancestry informative markers [103] that were included on the array and overlapped with HapMap release 3 (Fig S4.2).

Prior to genotype imputation, individuals were categorized into two groups based on the k-means clustering of ancestry PCs, using the kmeans() function in R; individuals were grouped as European or African American based on the mean clustering of their ancestry PCs against the HapMap reference panels (Fig S4.2). Phasing and imputation were performed using the ShapeIt2 [145] and Impute2 [146] software packages, respectively. Variants were imputed in 5 Mb windows across the genome against the 1000 Genomes Phase 3 haplotypes (Build 37; October 2014). After imputation, genotypes from both groups were merged and QC was performed with gtool [148]. X and Y linked SNPs and autosomal SNPs that did not meet the QC criteria (info score < 0.8, MAF < 0.05, missingness > 0.05 and a probability score < 0.9) were excluded from analyses. Probability scores were converted to dosages for 6,665,552 of the remaining sites used in downstream analyses.

4.3.5 RNA extraction and sequencing

Following RV and vehicle treatments, RNA and DNA were extracted from cells using the QIAGEN AllPrep DNA/RNA Kit. RNA quality and quantity were measured using the Agilent RNA 6000 Pico assay and the Agilent 2100 Bioanalyzer. RNA integrity numbers (RIN) were greater than 7.7 for all samples. cDNA libraries were constructed using the Illumina TruSeq RNA Library Prep Kit v2 and sequenced on the Illumina HiSeq 2500 System (50 bp, single-end); RNA sequencing was completed at the University of Chicago Genomics Core. Subsequently, we checked for potential sample contamination and sample swaps using the software VerifyBamID (http://genome.sph.umich.edu/wiki/VerifyBamID) [104] for cells from all 104 individuals in each treatment condition. We did not detect any cross-contamination between samples; one sample swap between two individuals, which was corrected.

Sequences were mapped to the human reference genome (hg19) and reads per gene were quantified using the Spliced Transcripts Alignment to a Reference (STAR) [105] software. X,Y, and mitochondrial chromosome genes, and low count data (genes < 1CPM) were removed prior to normalization via the trimmed mean of M-values method (TMM) and variance modeling (voom) [106]. Nine samples containing < 8M reads were removed from the analysis. Principle components analysis (PCA) identified biological and technical sources of variation in the voom-normalized RNA-seq reads for the remaining 95 samples. We identified contributors to batch and other technical effects (days in liquid nitrogen, experimental culture days, cell culture batches, RNA concentration, RNA fragment length, technician, sequencing pool). Additionally, unknown sources of variation were predicted with the Surrogate Variable Analysis (SVA) [107] package in R where 15 surrogate variables (SVs) were estimated for the samples that were included in the experiment after protecting for treatment in the full model. Voom-normalized RNA-seq data

were then adjusted for technical effects (cell lysate batch, sequencing pool, technician, fragment length, RNA concentration, days frozen in liquid nitrogen, experimental culture days), smoking, SVs, sex, and ancestry PCs (1-3) using the function removeBatchEffect() from the R package limma [110]. The variance in gene expression that was associated with asthma or CRS was correlated with the SVs and was therefore removed by inclusion of SVs in our model . The PCA of gene expression of the vehicle- and RV-treated samples after regression of the covariates are shown in Fig S4.3.

4.3.6 DNA extraction and methylation profiling

Following RV and vehicle treatments, DNA was extracted from cells as described above. DNA methylation profiles for cells from each treatment were measured on the Illumina Infinium MethylationEPIC BeadChip at the University of Chicago Functional Genomics Core. Methylation data for the 104 samples were preprocessed using the minfi package [186]. Probes located on sex chromosomes and with detection p-values greater than 0.01 in more than 10% of samples were removed from the analysis. One sample had > 5% missing probes and was therefore removed from the analysis. A preprocessing control normalization function was applied to the remaining 103 samples to correct for raw probe values or background, and a Subset-quantile Within Array Normalization (SWAN) [108] was used to correct for technical differences between the Infinium type I and type II probes. Additionally, we removed cross-reactive probes and probes within two nucleotides of a SNP with an MAF greater than 0.05 using the function rmSNPandCH() from the R package DMRcate [109].

PCA identified technical and biological sources of variation in the normalized DNA methylation datasets. Technical effects (array and cell harvest date), as well as sex, age, and smoking were significant variables in the PCA. Unknown sources of variation were predicted

with the SVA package. We estimated 37 SVs after protecting for treatment. SWAN and quantile-normalized M-values were then adjusted for batch and technical effects (sample plate and cell harvest date), smoking, SVs, sex, age, and smoking using the function removeBatchEffect() in R. The variance associated with asthma or CRS in the DNA methylation data was correlated with SVs and was therefore removed by inclusion of SVs in the model . The PCA of DNA methylation of the vehicle- and RV-treated samples after regression of the covariates are shown in Fig S4.3.

4.3.7 eQTL and meQTL analyses

Prior to e/meQTL analysis, voom-transformed gene expression values and normalized methylation M-values were adjusted for technical, biological, and surrogate variables as described above. Linear regression between the imputed genotypes (MAF>0.05) and molecular phenotypes (gene expression and methylation residuals) from each treatment condition was performed with the FastQTL [187] software package within *cis*-window sizes of ± 1 Mb (2 Mb total) of the transcription start site [75] and ± 10 kb (20 kb total) from a CpG [188] for eQTL and meQTL analyses, respectively. An FDR threshold of 0.05 was applied to adjust for multiple testing within each experimental dataset with the p.adjust() function in R.

4.3.8 Multivariate adaptive shrinkage analysis (mash)

An Empirical Bayes method of multivariate adaptive shrinkage was applied separately to the eQTL and meQTL data sets as implemented in the R statistical package, mashr (https://github.com/stephenslab/mashr) [189], to produce improved estimates of QTL effects and corresponding significance values in each treatment condition. To do this, mashr estimates patterns of similarity among eQTLs or meQTLs from each treatment condition (vehicle, RV). These patterns are used by mash to improve the accuracy of the molecular QTL effects by an

empirical Bayes approach. Compared to direct comparisons between conditions, *mash* increases power, improves effect-size estimates, and provides better quantitative assessments of effect size heterogeneity of molecular QTLs, thereby allowing for greater confidence in effect sharing and estimates of condition-specificity [189]. As a confidence measurement of the direction of QTL effects, *mash* provides a 'local false sign rate' (lfsr) that is the probability that the estimated effect has the incorrect sign [190], rather than the expected proportion of Type I errors as would be assessed using FDR thresholds. Mashr implements this in two general steps: 1) identification of pattern sharing, sparsity, and correlation among QTL effects, and 2) integration of these learned patterns to produce improved effects estimates and measures of significance for eQTLs or meQTLs in each treatment condition. To fit the mash model, we first estimated the correlation structure in the null test from a random dataset in which 2M gene-SNP or CpG-SNP pairs were randomly chosen for eQTLs and meQTLs, respectively, from the FastQTL nominal pass. The data-driven covariances were then estimated using the most significant e/meQTLs in each gene or CpG from the FastQTL results. Posterior summaries were then computed for the 'top' eQTL and meQTL results (see [189]). The instructions found in the *mashr* eQTL analysis outline vignette were followed to run mash.

4.3.9 Enrichment analysis

The R package, GWAS analysis of regulatory or functional information enrichment with LD correction (GARFIELD) [191], was used to quantify enrichment of GWAS SNPs among eQTLs and meQTLs and assess significance. GARFIELD leverages GWAS results with molecular data to identify features relevant to a phenotype of interest, while accounting for LD and matching for genotyped variants, by applying a logistic regression method to derive statistical significance for enrichment. For this study, molecular QTLs (union of e/meQTLs from each treatment condition)

were tested for GWAS variant enrichment, estimated as odds ratios and enrichment P-values derived at four GWAS P-value thresholds: 10^{-5}, 10^{-6}, 10^{-7}, and 10^{-8}. To evaluate disease-specificity, we selected summary statistics from nine GWASs including three asthma GWASs (adult onset and childhood onset asthma [13], and a meta-analyzed multi-ethnic, all age of asthma onset from the Trans-National Asthma Genetic Consortium [TAGC] [11]), two allergic disease GWASs (hay fever/allergic rhinitis [192] and atopic dermatitis [193]), and four non-allergic GWASs (Alzheimer's disease [194], atrial fibrillation [195], height [196], neuroticism [197]). Summary statistics from these nine GWASs were used for enrichment analyses of the 490,151 molecular QTLs (FDR<0.05) combined from each treatment condition. These non-allergic GWASs were chosen based on similar population backgrounds (European), availability of summary statistics (as of 05/18), and both known to have genetic overlap with asthma (allergic diseases) and not known to have overlapping genetics with asthma (Alzheimer's disease, atrial fibrillation, height, and neuroticism).

To assess tissue-specificity of our results, we examined eQTLs from the adrenal gland, frontal cortex, hypothalamus, ovary, and testis from the GTEx database version 7 (http://gtexportal.org) [75], and tested for enrichment of adult onset, childhood onset asthma, and TAGC asthma GWAS SNPs among the epithelial eQTLs from our study combined across treatment conditions. GTEx data were matched with respect to sample size and number of eQTLs to those of the epithelium, with the exception of testis, which was included to show the consistency of the enrichment results despite it being an outlier in regards to both sample size, which was smaller, and number of eQTLs, which was larger.An OR > 1 and a FDR corrected p-value threshold of < 0.05 was used as the significance threshold for enrichment; FDR adjusted p-

values were calculated using the p.adjust() function in R where 'n' was determined by the number of tests in each respective enrichment analysis.

4.3.10 Co-localization analysis

To estimate the posterior probability association (PPA) that a SNP contributed to the association signal in the GWAS as well as to the eQTL and/or meQTL, we applied a Bayesian statistical framework implemented in the R package multiple-trait-coloc (moloc) [181]. Summary data from adult onset and childhood onset asthma GWASs in the UK Biobank [13], and the TAGC multi-ethnic GWAS [11], along with eQTL and meQTL summary data from upper airway epithelial cells within each treatment condition (described above), were included in the *moloc* analysis. Each co-localization analysis included summary data from a GWAS and epithelial cell eQTLs and meQTLs from each treatment condition. Because a genome-wide co-localization analysis was computationally challenging, genomic regions for co-localization were defined using GARFIELD. First, we analyzed the enrichment pattern of e/meSNPs from each treatment condition in each of the three asthma GWASs using the default package settings. Second, we extracted variants driving the enrichment signals at a GWAS p-value threshold of $1x10^{-4}$. Regions were defined as 2 Mb windows centered around these variants. Only regions with at least 10 SNPs in common between all three datasets or 'traits' (GWAS, eQTL, and meQTL) were assessed by moloc and 15 'configurations' of possible variant sharing was computed across these three traits (see [181] for more details). The PPA was computed by weighting the likelihood of the summary data given the prior probability that a SNP associates with each trait (asthma, gene expression, and DNA methylation) the moloc default prior probabilities were included in our analysis. Prior probabilities of $1x10^{-4}$, $1x10^{-6}$, and $1x10^{-7}$ were chosen for the association of one, two, or three traits, respectively, as recommended by the authors of moloc.

False positive rates remain below 0.05 at a posterior probability thresholds as low as 0.30 [181]. We therefore considered PPAs ≥ 0.50 as evidence for co-localization in this study.

We performed six separate co-localization analyses for each treatment condition with each of the three asthma GWASs. Each analysis provided three possible configurations in which a variant is co-localized between the GWAS and QTLs: eQTL-GWAS pairs, meQTL-GWAS pairs, and eQTL-meQTL-GWAS triplets. Estimates of a posterior probability of association (PPA) is provided, reflecting the evidence for a colocalized SNP being causal for the associations in the GWAS and for the corresponding eQTL and/or meQTL.

4.3.11 Mendelian randomization

To infer causal relationships between DNA methylation and gene expression on asthma risk for colocalized triplets, we performed Mendelian randomization (MR), a method in which genetic variation associated with modifiable exposures (i.e. DNA methylation) can be used as an instrumental variable to estimate the causal influence of an exposure on an outcome (i.e. DNA methylation on gene expression) [198]. We applied a two-stage least squares regression (2SLS) regression using the ivreg2 function [199] in R (https://www.r-bloggers.com/an-ivreg2-function-for-r/) to estimate the effects of DNA methylation (exposure) on gene expression (outcome) in each treatment condition, and used the QTL SNP in the co-localized triplets (eQTL-meQTL-GWAS) to assess the causal effects of DNA methylation on gene expression.

4.4 Results

4.4.1 Genome-wide cis-eQTLs and cis-meQTLs mapping in cultured airway epithelial cells

We performed eQTL and meQTL mapping using gene expression and DNA methylation data from the same cells. The numbers of SNPs associated with gene expression for at least one gene (eQTLs) and genes with at least one eQTL (eGenes), in any treatment, are summarized in Fig

S4.5A; the number of SNPs associated with methylation levels at one or more CpG sites (meQTLs) and CpG sites with at least one meQTL (meCpGs), in any treatment, are shown in Fig S4.5B. Overall, we identified 60,428 eQTLs (40,354 and 37,566 from the vehicle- and RV-treated cells, respectively) associated with 1,710 genes, and 429,725 meQTLs (302,896 and 283,474 from the vehicle- and RV-treated cells, respectively) associated with DNA methylation at 38,942 CpGs.

Because each gene/CpG-variant pair was tested for a linear regression slope that significantly deviated from 0, the estimated effects for the molecular QTLs reflects both the single-SNP effects of each molecular QTL as well as those that are in linkage disequilibrium (LD). Accordingly, these analyses do not differentiate between causal molecular QTLs from those in LD with the QTL.

4.4.2 Estimating shared and condition-specific molecular QTL effects

We first explored the impact of RV exposure on eQTLs and meQTLs by comparing RV-treated to vehicle-treated results to identify condition-specific eQTLs. For this, we analyzed the effect estimates of the most significant eQTL for each of 11,896 genes and assessed sharing of these signals among the RV and vehicle treated cells using mash [189] (see Methods). A pairwise comparison showed that 58.3% of eQTLs were shared between RV and vehicle treatments, representing 1,223 eGenes (Fig 4.1A), and the remaining 41.5% of eQTLs were specific to the vehicle-treated (471 genes) or RV-treated (409 genes) cells. These potentially represent functional genetic variants that modify responses to viral exposure in AECs. Examples of treatment-specific eQTLs are shown in Fig 4.1B.

The effect estimates of the most significant meQTL for each of 792,392 CpG sites were used to identify condition-specific DNA methylation effects, as described above for eQTLs. A

pair-wise analysis of meQTLs revealed that 89.9% of meQTLs were shared between vehicle and RV treatments, representing 48,189 meCpGs, defined here as CpGs with at least one meQTL at a lfsr<0.05 (Fig 4.1C), revealing a much greater proportion shared meQTLs than those observed for eQTLs. Examples of the 21,295 treatment-specific meQTLs are shown in Fig 4.1D. In total, we identified 471 and 409 eGenes (lfsr<0.05) that were specific to vehicle

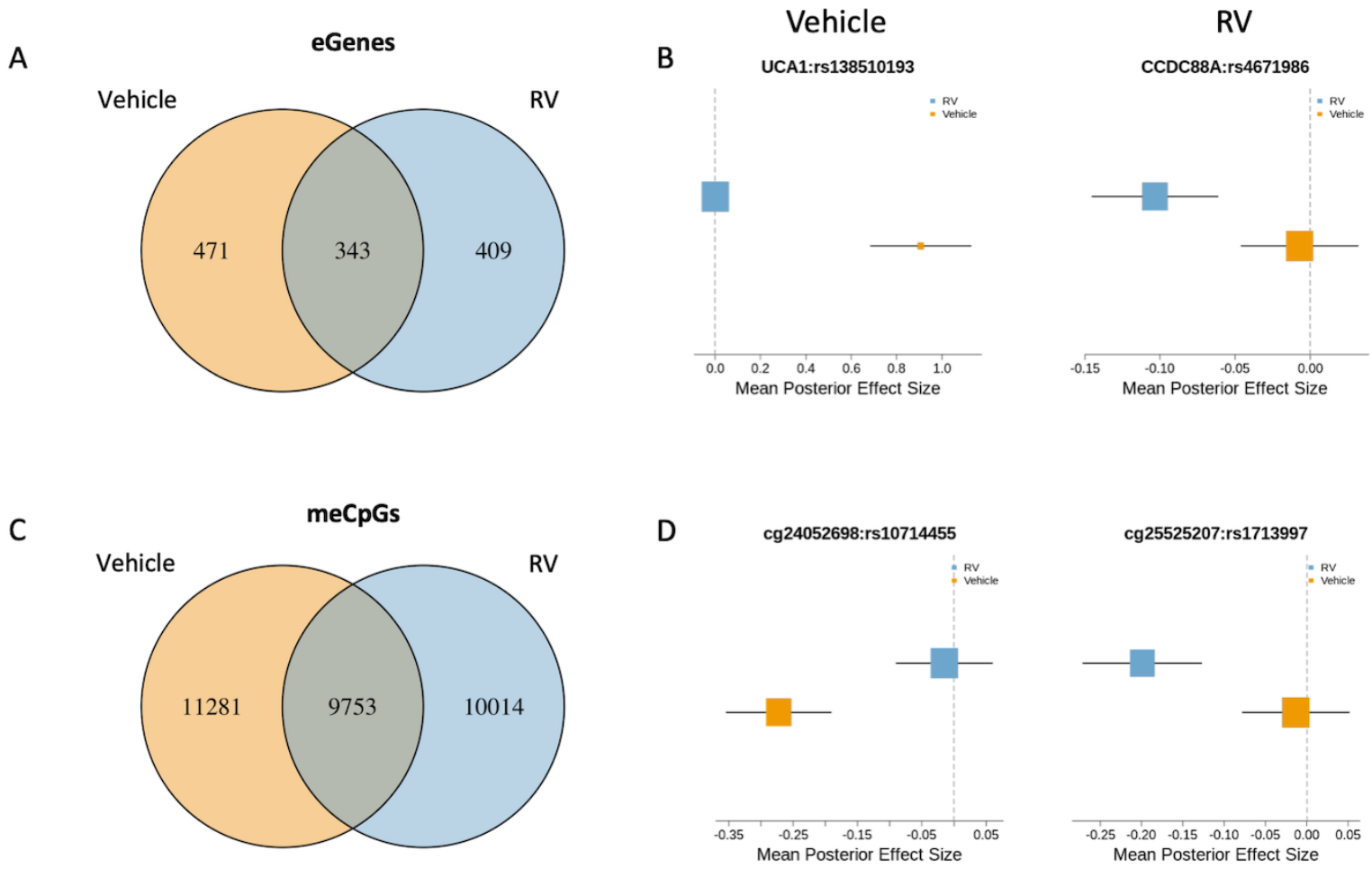


Figure 4.1 Molecular effects sharing across treatment conditions (lfsr<0.05). Venn diagrams of eGenes (A) and meCpGs (C) shared between vehicle- and RV-treated airway epithelial cells. Forest plots showing examples of RV- (left) and Vehicle-specific (right) eQTLs (B) and meQTLs (D).

or RV treatment, respectively, and 11,281 and 10,014 meCpGs that were specific to vehicle-treated or RV-treated cells, respectively.

4.4.3 M*olecular QTLs in the airway epithelium are enriched for asthma GWAS SNPs*

To assess whether the 490,151 molecular QTLs identified in our study (i.e., the union of eQTLs and meQTLs from each treatment condition at FDR<0.05) are enriched for asthma GWAS variants and whether these enrichments show tissue specificity, we first extracted summary statistics from publicly available GWAS data for asthma, including childhood onset and adult onset asthma GWASs in British white individuals [13] and a multi-ethnic asthma GWAS [11], for two allergic diseases (hay fever/allergic rhinitis and eczema/atopic dermatitis [193]), and for four diseases without known allergic or immune etiologies (Alzheimer's disease [194], atrial fibrillation [195], height [196], and neuroticism [197]). These analyses revealed statistically significant enrichments (OR>1 and FDR-adjusted P-value<0.05; see Methods) for SNPs from all three asthma GWAS among the molecular QTLs (eQTLs and meQTLs) at each of four GWAS thresholds (Table 4.1), consistent with the strong epithelial cell involvement in asthma in general and with childhood onset asthma in particular. We also observed significant enrichments among the molecular QTLs for each of the two allergic disease GWASs, also consistent with the central role of epithelial cells in the regulation of allergic diseases [200] and the shared genetic architecture of asthma with these diseases [79]. In contrast, there were no significant enrichments for SNPs from the other GWASs among the epithelial cell molecular QTLs, with exception of the neuroticism GWAS SNPs, which were enriched among the epithelial cell QTLs at three of the four GWAS thresholds. Overall, these results highlight the specific enrichment of asthma and allergic disease GWAS SNPs among airway epithelial molecular QTLs compared to SNPs from GWASs of diseases without known epithelial cell involvement.

Table 4.1 Enrichment estimates of airway epithelial cell eQTLs and meQTLs for GWAS SNPs. *P*-values and diseases that are significant after FDR correction (see Methods) are shown in bolded type (FDR<0.05).

	GWAS	N_{Cases}	$N_{Controls}$	N_{Total}	GWAS Threshold	OR	P	P_{adj}
Non-Allergic Disease GWAS	Alzheimer's Disease [45]	47,793	328,320	376,311	$1x10^{-5}$	0.68	$2.25x10^{-1}$	$2.61x10^{-1}$
					$1x10^{-6}$	0.6	$2.22x10^{-1}$	$2.61x10^{-1}$
					$1x10^{-7}$	0.64	$3.51x10^{-1}$	$3.72x10^{-1}$
					$1x10^{-8}$	0.46	$1.91x10^{-1}$	$2.37x10^{-1}$
	Atrial Fibrillation [46]	60,620	970,216	1,030,836	$1x10^{-5}$	1.19	$3.62x10^{-1}$	$3.72x10^{-1}$
					$1x10^{-6}$	1.17	$5.20x10^{-1}$	$5.20x10^{-1}$
					$1x10^{-7}$	1.41	$2.34x10^{-1}$	$2.63x10^{-1}$
					$1x10^{-8}$	1.7	$1.25x10^{-1}$	$1.73x10^{-1}$
	Height [47]	NA	NA	253,288	$1x10^{-5}$	1.23	$2.49x10^{-1}$	$2.72x10^{-1}$
					$1x10^{-6}$	1.38	$1.43x10^{-1}$	$1.91x10^{-1}$
					$1x10^{-7}$	1.5	$1.22x10^{-1}$	$1.73x10^{-1}$
					$1x10^{-8}$	1.51	$1.56x10^{-1}$	$2.01x10^{-1}$
	Neuroticism [48]	**130,664**	**330,470**	**461,134**	$1x10^{-5}$	1.5	$3.58x10^{-2}$	$5.37x10^{-2}$
					$1x10^{-6}$	**1.98**	**$1.01x10^{-2}$**	**$1.58x10^{-2}$**
					$1x10^{-7}$	**3.53**	**$2.03x10^{-3}$**	**$3.94x10^{-3}$**
					$1x10^{-8}$	**8.64**	**$9.26x10^{-4}$**	**$1.96x10^{-3}$**
Allergic Disease GWAS	**Hay Fever, Allergic Rhinitis** [192] **(multi-ethnic)**	**26,910**	**83,424**	**110,334**	**$1x10^{-5}$**	**4.82**	**$1.55x10^{-5}$**	**$9.30x10^{-5}$**
					$1x10^{-6}$	**7.32**	**$5.50x10^{-5}$**	**$2.48x10^{-4}$**
					$1x10^{-7}$	**5.39**	**$3.20x10^{-3}$**	**$5.49x10^{-3}$**
					$1x10^{-8}$	**9.63**	**$2.36x10^{-3}$**	**$4.25x10^{-3}$**
	Atopic Dermatitis [193]	**11,025**	**40,398**	**51,423**	**$1x10^{-5}$**	**3.86**	**$7.45x10^{-4}$**	**$7.37x10^{-4}$**
					$1x10^{-6}$	**4.54**	**$5.47x10^{-3}$**	**$2.04x10^{-4}$**
					$1x10^{-7}$	**6.35**	**$5.51x10^{-3}$**	**$3.38x10^{-4}$**
					$1x10^{-8}$	**7.96**	**$1.72x10^{-2}$**	**$3.11x10^{-5}$**

Table 4.1 Enrichment estimates of airway epithelial cell eQTLs and meQTLs for GWAS SNPs (Continued). *P*-values and diseases that are significant after FDR correction (see Methods) are shown in bolded type (FDR<0.05; continued).

Asthma GWAS	**Asthma** [10] **(multi-ethnic; all ages of onset)**	**23,948**	**118,538**	**142,486**	**$1x10^{-5}$**	**4.64**	**$1.25x10^{-4}$**	**$3.75x10^{-4}$**
					$1x10^{-6}$	**6.95**	**$1.16x10^{-4}$**	**$3.75x10^{-4}$**
					$1x10^{-7}$	**6.44**	**$2.08x10^{-3}$**	**$3.94x10^{-3}$**
					$1x10^{-8}$	**5.16**	**$8.99x10^{-3}$**	**$1.47x10^{-2}$**
	Adult Onset Asthma [1]	**21,564**	**318,237**	**339,801**	**$1x10^{-5}$**	**7.14**	**$8.94x10^{-7}$**	**$1.61x10^{-5}$**
					$1x10^{-6}$	**8.06**	**$1.44x10^{-4}$**	**$3.99x10^{-4}$**
					$1x10^{-7}$	**11.7**	**$3.87x10^{-4}$**	**$8.71x10^{-4}$**
					$1x10^{-8}$	**31.8**	**$1.83x10^{-4}$**	**$4.71x10^{-4}$**
	Childhood Onset Asthma [1]	**9,433**	**318,237**	**327,670**	**$1x10^{-5}$**	**2.83**	**$3.46x10^{-6}$**	**$3.11x10^{-5}$**
					$1x10^{-6}$	**4.63**	**$1.14x10^{-7}$**	**$4.10x10^{-6}$**
					$1x10^{-7}$	**4.77**	**$4.35x10^{-6}$**	**$3.13x10^{-5}$**
					$1x10^{-8}$	**4.36**	**$9.17x10^{-5}$**	**$3.38x10^{-4}$**

To further assess the specificity of airway epithelial molecular QTLs to asthma, we compared GWAS SNP enrichments among the eQTLs in our study to those from tissues that are not known to be involved in asthma. Comparable cross-tissue data for meQTLs were not available. We tested for enrichment of asthma GWAS SNPs among eQTLs (FDR<0.05) in five different tissues from the GTEx database (adrenal, frontal cortex, hypothalamus, ovary, testis) [75], and compared them to enrichments among the epithelial cell eQTLs from our study. We observed a significant enrichment (OR>1 and FDR-adjusted P≤0.05) of childhood onset asthma (Table 4.2) and TAGC (Table S4.1) GWAS SNPs among the epithelial cell eQTLs at all GWAS P-value thresholds $\leq 1x10^{-8}$, while enrichments for adult onset asthma GWAS SNPs among the epithelial cell eQTLs were not observed at any GWAS threshold (Table S4.2). Except for the hypothalamus, which showed some enrichment at $P \leq 10^{-5}$, no other enrichments of asthma GWAS SNPs were observed among eQTLs in other tissues, further supporting the specificity of our model and previous studies suggesting that epithelial barrier defects underlie risk for childhood onset, but not adult onset, asthma [13, 184, 185].

Table 4.2 Enrichment estimates of eQTLs for childhood onset asthma GWAS SNPs from six tissues. Significant *P*-values after FDR correction are shown in bolded type (FDR≤0.05). Results for TAGC and adult onset asthma are shown in Tables S4.1 and S4.2, respectively.

Tissue	GWAS Threshold	OR	P	P_{adj}	N	N_{eSNP}
Adrenal	$1x10^{-5}$	1.31	$2.15x10^{-1}$	$4.69x10^{-1}$	175	588,348
	$1x10^{-6}$	1.00	$9.98x10^{-1}$	$9.98x10^{-1}$		
	$1x10^{-7}$	0.75	$4.22x10^{-1}$	$5.33x10^{-1}$		
	$1x10^{-8}$	1.04	$9.13x10^{-1}$	$9.53x10^{-1}$		
Brain - Frontal Cortex	$1x10^{-5}$	1.71	$2.41x10^{-2}$	$9.64x10^{-2}$	118	367,312
	$1x10^{-6}$	1.32	$3.60x10^{-1}$	$5.22x10^{-1}$		
	$1x10^{-7}$	1.23	$5.69x10^{-1}$	$6.83x10^{-1}$		
	$1x10^{-8}$	1.46	$3.37x10^{-1}$	$5.22x10^{-1}$		
Brain - Hypothalamus	**$1x10^{-5}$**	**2.25**	**$1.21x10^{-3}$**	**$9.68x10^{-3}$**	**108**	**251,506**
	$1x10^{-6}$	1.70	$1.07x10^{-1}$	$2.85x10^{-1}$		

Table 4.2 Enrichment estimates of eQTLs for childhood onset asthma GWAS SNPs from six tissues. Significant *P*-values after FDR correction are shown in bolded type (FDR≤0.05). Results for TAGC and adult onset asthma are shown in Tables S4.1 and S4.2, respectively (continued).

	1x10-7	1.68	1.78x10-1	4.27x10-1		
	1x10-8	1.62	2.70x10-1	4.98x10-1		
Ovary	1x10-5	1.65	5.51x10-2	1.89x10-1	122	292,461
	1x10-6	1.34	3.70x10-1	5.22x10-1		
	1x10-7	0.94	8.86x10-1	9.53x10-1		
	1x10-8	0.80	6.80x10-1	7.77x10-1		
Testis	1x10-5	0.80	2.40x10-1	4.80x10-1	225	1,358,512
	1x10-6	0.69	1.04x10-1	2.85x10-1		
	1x10-7	0.80	4.10x10-1	5.33x10-1		
	1x10-8	0.75	3.28x10-1	5.22x10-1		
Airway Epithelial Cells	1x10-5	3.68	2.83x10-5	6.79x10-4	104	185,407
	1x10-6	3.28	2.83x10-3	1.69x10-2		
	1x10-7	4.11	8.91x10-4	9.68x10-3		
	1x10-8	3.91	6.16x10-3	2.96x10-2		

4.4.4 Molecular QTL co-localizations with asthma risk loci

Integrating molecular QTLs with GWAS data is a powerful way to identify functional variants that may ultimately influence disease risk [201, 202] and to assign function to known disease-associated variants. We hypothesized that integrating molecular QTLs from RV- and vehicle-exposed epithelial cells with results of GWASs for adult onset and childhood onset asthma would reveal genetic and epigenetic mechanisms that modulate risk for childhood and/or adult onset asthma. Co-localization approaches directly test whether the same genetic variant (or variants in LD) underlie associations between two or more traits (e.g., gene expression and asthma), providing clues to causal disease pathways.

Using this approach, we found evidence for a total of 46 unique multiple trait co-localizations (Table 4.3; Table S4.3). Eleven co-localizations were detected with adult onset asthma GWAS SNPs, of which all were meQTL-GWAS pairs (11 different CpGs). In contrast,

37 co-localizations were detected with childhood onset asthma GWAS SNPs, including 22 eQTL-meQTL-GWAS triplets (13 different genes and 19 different CpGs), five eQTL-GWAS pairs (five different genes), and 10 meQTL-GWAS pairs that were also identified in the adult onset asthma GWAS. Ten co-localizations were detected with the TAGC GWAS SNPs, including two eQTL-meQTL-GWAS triplets (two different genes and two different CpGs from the childhood onset results), and eight me-QTL-GWAS pairs associated with eight CpGs, one of which overlapped with those detected in the other two GWASs (Table 4.3). Overall, only a single co-localization was specific to adult onset asthma. Although the triplets identified in the TAGC GWAS were not the same as those identified in the childhood onset GWAS, a gene (*ERBB2*) and CpG (cg10374813) were associated with triplets in both of these GWASs. Overall, these colocalizations involved 33 SNPs at 17 loci, 14 genes, and 36 CpGs. Moreover, the combined data show that the colocalizations of molecular QTLs in airway epithelial cells with the adult onset asthma GWAS SNPs are subsets of the colocalizations in the childhood onset asthma GWAS SNPs. The largest number of co-localizations for childhood onset asthma GWAS SNPs is consistent with both the observation that genes at the childhood onset asthma loci were most highly expressed in skin, an epithelial cell type [13] and the enrichment of childhood onset asthma GWAS SNPs among epithelial cell eQTLs in our study. Two examples of co-localizations at prominent asthma-associated loci are described in the following sections.

Table 4.3 Number of QTL-GWAS pairs or triplets with evidence of co-localization (PPA≥0.50).

GWAS	eQTL-meQTL-GWAS triplets	eQTL-GWAS	meQTL-GWAS
Adult onset asthma	0	0	11
Childhood onset asthma	22	5	10
Asthma (multi-ethnic; all ages of onset)	2	0	8
Combined (union of co-localizations)	24	5	17

4.4.5 meCpGs at TSLP co-localize with an asthma risk variant

To more deeply characterize the co-localizations, we first focused on shared meQTL-GWAS pairs in all three GWASs. Three of these 11 pairs in the adult and childhood onset GWAS and one of 8 in the TAGC GWAS included an intergenic SNP (rs1837253) located 5.7 kb upstream from the transcriptional start site (TSS) of the *TSLP* gene on chromosome 5q22, encoding an epithelial cell cytokine that plays a key role in the inflammatory response in asthma and other allergic diseases [203]. rs1837253 co-localized with three meQTLs (cg10931190, cg15089387, cg15557878) in both the adult onset asthma ($p_{GWAS} = 2.77x10^{-13}$), childhood onset asthma ($p_{GWAS} = 2.33x10^{-27}$), and TAGC GWASs ($p_{GWAS} = 2.03x10^{-25}$; Fig 4.2; Fig S4.6).

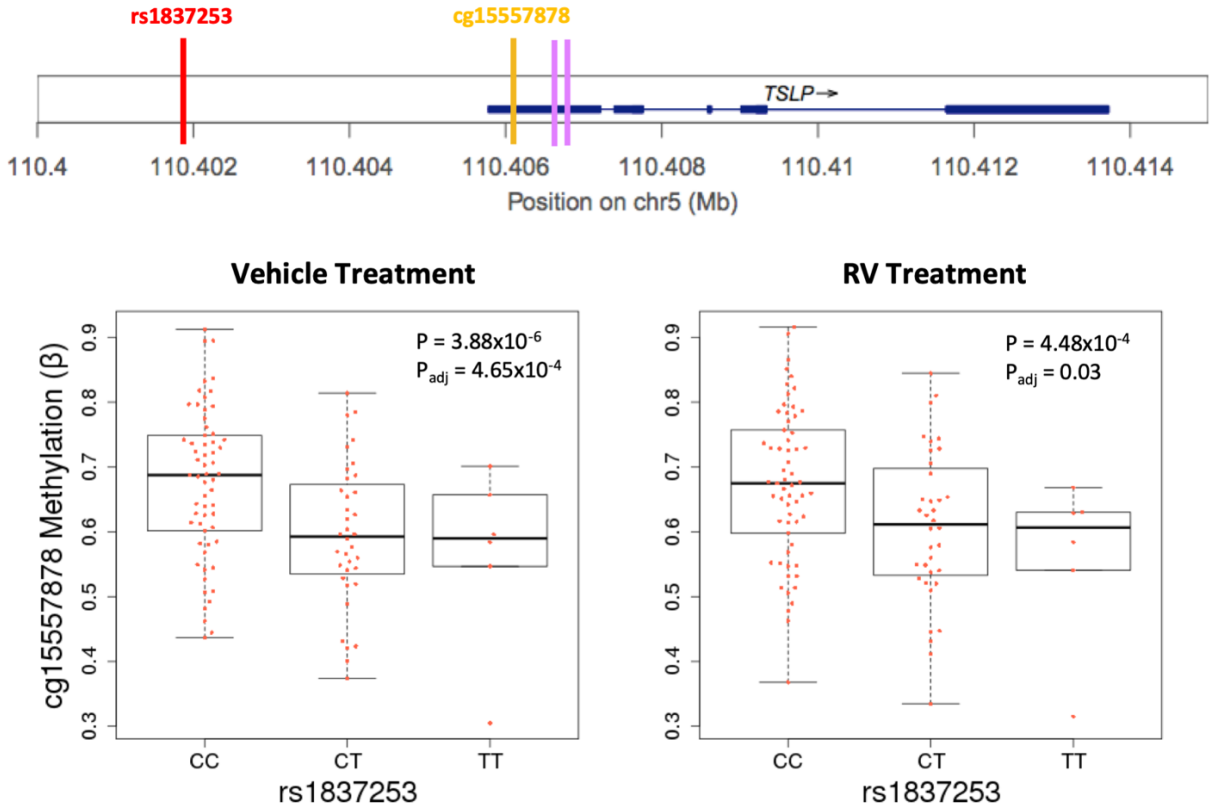


Figure 4.2 Co-localization of rs1837253 with DNA methylation levels for cg15557878 at *TSLP*. rs1837253 (red vertical bar, upper panel) is associated with DNA methylation levels at

Figure 4.2 Co-localization of rs1837253 with DNA methylation levels for cg15557878 at *TSLP*. (continued) cg15557878 (orange vertical bar, upper panel). Box plots show DNA methylation levels (y-axes) for meCpGs by rs1837253 genotype (x-axes) in the vehicle (β = -0.25; 95% CI -0.21, -0.30), and RV (β = -0.26; 95% CI -0.21, -0.30) treatment conditions (lower panel). Purple vertical bars indicate the location of the remaining meCpGs co-localized with rs1837253, which was associated with asthma in all three GWASs used in the co-localization studies ($p_{GWAS} < 10^{-12}$). P-values and FDR adjusted P-values (P_{adj}) are shown in each box plot.

The meCpGs are located in the first (untranslated) exon (5' UTR) of the *TSLP* gene, a region characterized as a promoter in normal human epidermal keratinocyte cells (NHEK; ROADMAP). In fact, rs1837253 was the sentinel SNP at this locus in GWASs of asthma (e.g. [13, 204]) and of moderate-to-severe asthma [205]. In our study, the rs1837253-C asthma risk allele was associated with hypermethylation of cg15557878 in primary cultured AECs (Fig 4.2), but was not associated with the expression of *TSLP* in either treatment condition (not shown). The co-localization of rs1837253 with an meQTL (cg15557878) in three separate asthma GWASs provides robust support for DNA methylation effects on asthma risk at this locus.

4.4.6 Multi-trait co-localizations of molecular QTLs and asthma risk at the 17q12-21 asthma locus

To further explore the possibility that some mechanisms of asthma risk are exposure-specific, we focused on the co-localizations of eQTLs and meQTLs with asthma-associated SNPs at the 17q12-21 (17q) locus, the most significant and most highly replicated locus for childhood onset asthma (reviewed in [28]). This locus is characterized by high LD across a core region of 150 kb, encoding at least 4 genes (including *ORMDL3* and *GSDMB*). SNPs extending both proximal (including *PGAP3* and *ERBB2*) and distal (including *GSDMA*) to the core region show less LD with those in the core region and have been implicated as potentially independent asthma risk loci. Previous studies have shown that SNPs at this extended locus are eQTLs for at least four genes (*ORMDL3*, *GSDMB*, *GSDMA*, *PGAP3*) in blood and/or lung cells [28] and that genotype

effects on risk for childhood onset asthma are modified by early life wheezing illness in general [67, 206], and RV-associated wheezing illness in particular [12].

We identified four co-localizations at this locus (Fig 4.3). One co-localization in the core region was identified only in the TAGC GWAS, and three at the distal end of the 17q locus was identified in both the childhood onset and TAGC GWASs. The TAGC co-localization was an meQTL-GWAS pair that was associated with cg10374813 and the three in both TAGC and the childhood onset asthma GWASs were eQTL-meQTL-GWAS triplets associated with two genes (*ORMDL3* and *ERBB2*) and two CpGs (cg18711369 and cg2270401). The TAGC triplet was co-localized only in RV-treated cells, and included an eQTL for *ORMDL3* (TAGC $p_{GWAS} = 6.20x10^{-45}$; PPA≥0.50; Fig 4.3A-B). The SNP (rs9303281) associated with this co-localization was located within the *GSDMB* gene body and the CpG (cg18711369) was located within the *ORMDL3* gene body (approximately 7.14 kb from rs9303281). The co-localized SNP is in LD (r^2>0.85 in the 1000 Genomes CEU reference panel) with other previously reported asthma-associated GWAS SNPs in this region (see [28]), including some that were reported as eQTLs for *ORMDL3* and *GSDMB*, primarily in resting blood immune cells. Although the *ORMDL3* co-localization was identified only in the RV-treated cells, an eQTL for this gene was also detected in the vehicle treatment. These data indicate that the asthma risk allele (A) is associated with decreased DNA methylation (Fig 4.3D-E) and increased expression of *ORMDL3* in both vehicle and RV-treated cells. However, whereas *ORMDL3* expression significantly decreased after exposure to RV (Fig 4.3C), DNA methylation levels at cg18711369 remains unchanged with RV infection (Fig 4.3F), consistent with our earlier observation of relatively few treatment-specific meQTLs.

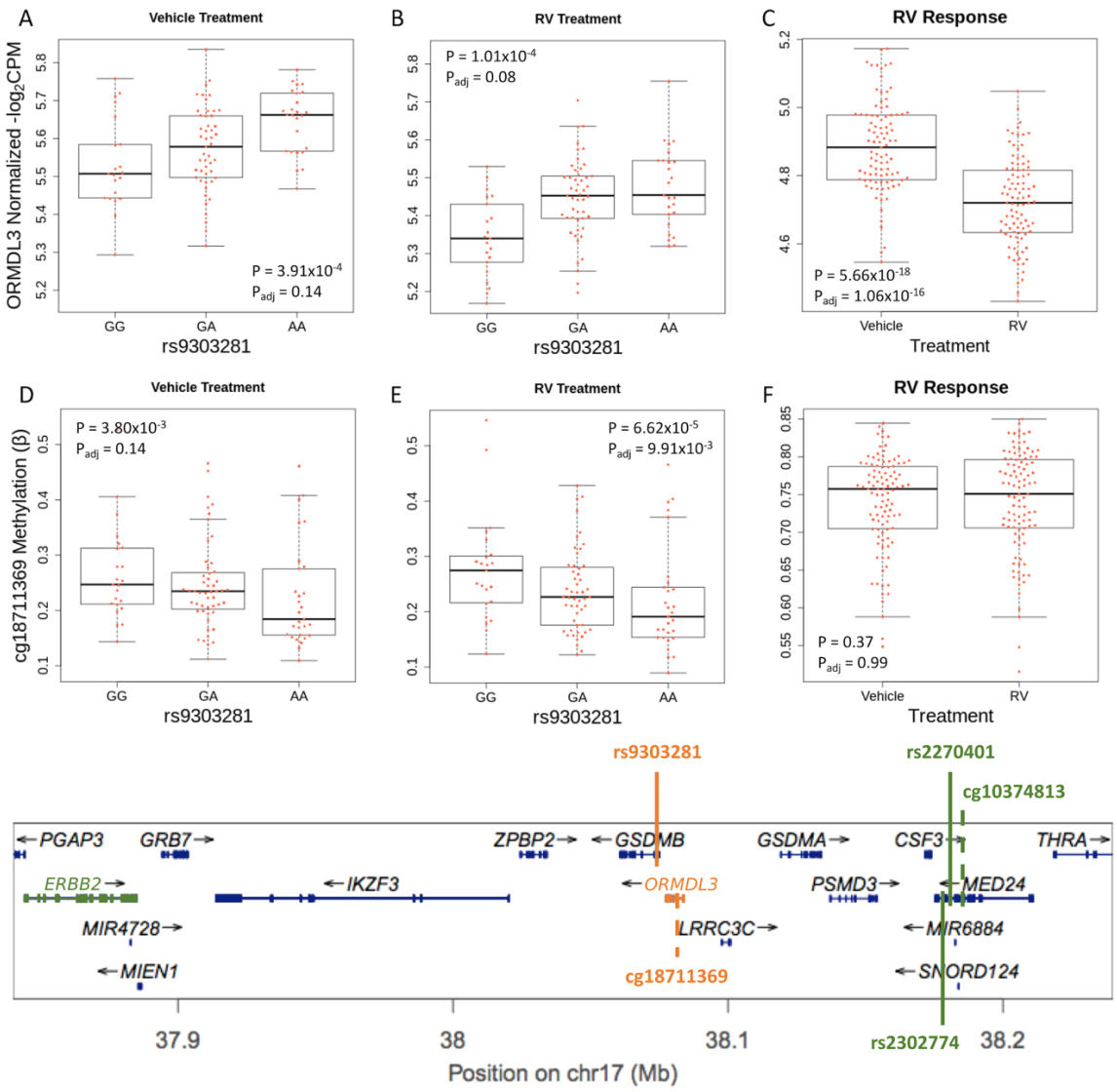


Figure 4.3 Co-localization pairs at the 17q asthma susceptibility locus. Upper panel: Box plots for the co-localized *ORMDL3* eQTL in cultured airway epithelial cells treated with vehicle (A; β = 0.06; 95% CI 0.07, 0.04) and RV (B; β = 0.06; 95% CI 0.08, 0.05). *ORMDL3* gene expression decreases after treatment with RV (C). Box plots of the cg21230266 meQTL treated with vehicle (D) and RV (E), and methylation levels at cg21230266 (F). The meQTLs and overall methylation levels are similar in vehicle and RV treatments. P-values and FDR adjusted P-values (P_{adj}) are shown in each box plot. Lower panel: The extended 17q12-21 locus. Co-localizations are shown by the vertical colored lines. Solid lines indicate the position of the colocalized SNP. Dashed lines indicate the location of meCpG pairs. Traits of the same co-localization are shown in the same color. The eQTL-meQTL-GWAS co-localization for *ORMDL3* is shown in orange, and the eQTL-meQTL-GWAS co-localization for *ERBB2* is shown in green.

The remaining two eQTL-meQTL-GWAS triplets spanned the entire locus (Fig 4.4A, upper panel), included two eQTLs for *ERBB2*, which is located at the proximal end of the locus, more than 290 kb from the two co-localized asthma risk variants rs2270401 (childhood onset) and rs2302774 (TAGC) that are in strong LD (r^2=0.94 in CEU) and one co-localized meCpG (cg10374813) in an intron of *MED24* (Fig 4.4A, middle panel). *MED24* is located beyond the extended 17q12-21 locus as previously defined [28], in a region characterized by ROADMAP as an enhancer and in NHEKs. The eQTLs for *ERBB2* with both SNPs are observed only after exposure to RV (Fig 4.4A middle and lower panels show results for rs2270401), although the meQTLs are present in both vehicle and RV treated cells (Fig 4.4B upper and lower panels, respectively). The asthma risk alleles in the childhood onset asthma (rs2270401-A) and TAGC (rs2302774-G) GWASs were associated with increased DNA methylation of cg10374813 in both conditions and decreased *ERBB2* expression only in RV-treated cells. Overall *ERBB2* expression decreased in response to RV exposure in AECs (Fig 4.4D). The observation that the meQTL for cg10374813 is observed in both conditions suggests an epigenetically poised chromatin state at the distal end of the locus that directly affects transcription of *ERBB2* at the proximal end of the locus after exposure to RV, and possibly to other viruses.

The >290 kb distance between the promoter of *ERBB2* and its eSNPs (rs2270401, rs2302774) suggests a long-range interaction between *ERBB2* and the region harboring both cg10374813 and the asthma-associated SNPs (rs2270401, rs2302774). To examine this possibility, we used promoter capture Hi-C (pcHi-C) data from *ex vivo* primary bronchial epithelial cells (n=8) [87]. These data revealed four chromatin-chromatin interactions (or loops) between the *MED24* gene and the *ERBB2* promoters (Fig 4.5). These data provide empiric support for our co-localization results suggesting a potential epigenetic mechanism at the

MED24 locus that impacts the expression of *ERBB2* and highlights a broad regulatory landscape across the 17q12-21 asthma locus.

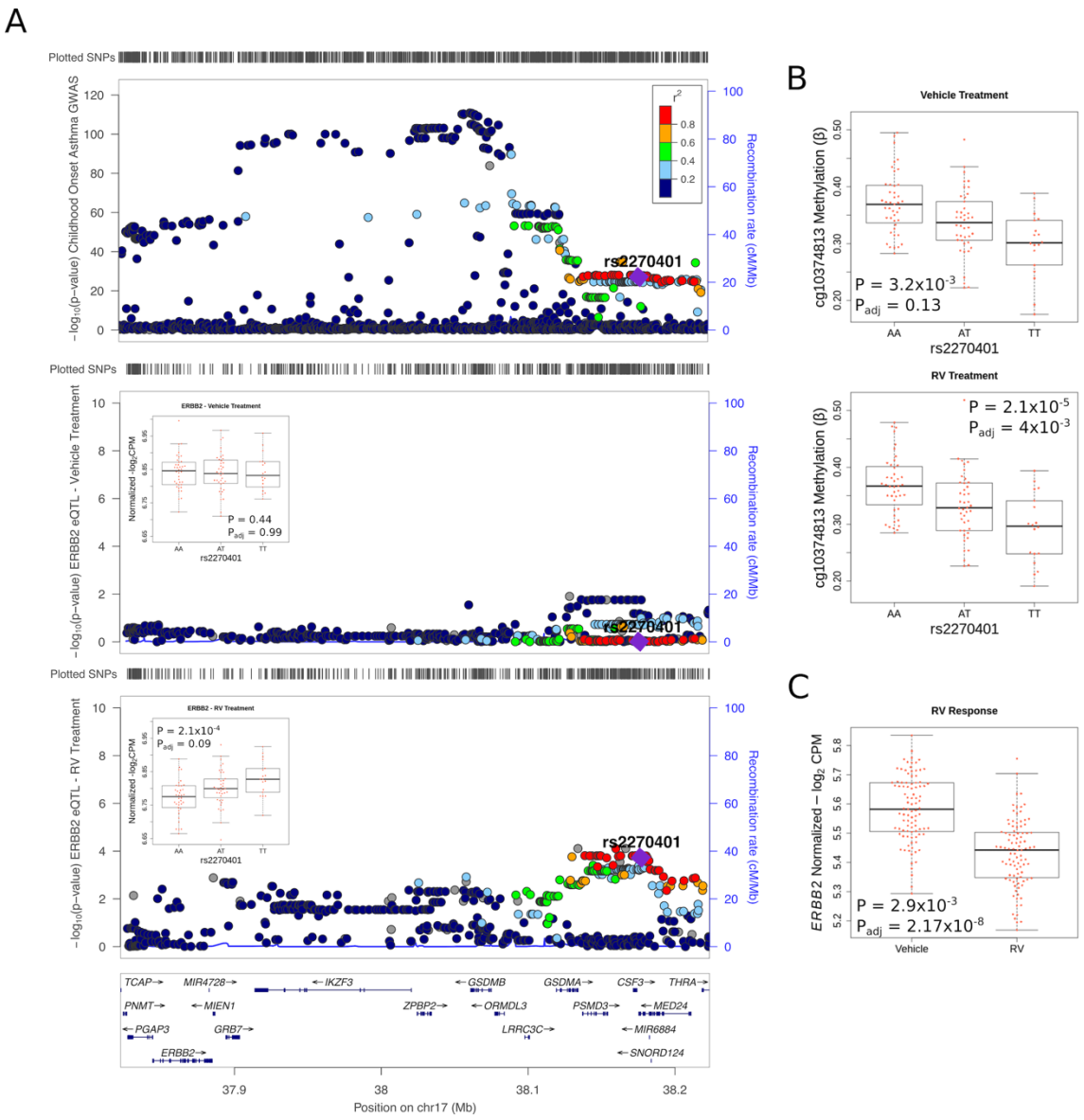


Figure 4.4 Co-localization of rs2270401 with *ERBB2* expression and DNA methylation levels for cg10374813. (A) LocusZoom plots of childhood onset GWAS results at the 17q locus showing the *ERBB2* gene at the proximal (left) end of the locus and the co-localized eQTL (rs2270401) at the distal (right) end of the locus. The SNP (rs2270401), which colocalized with associations for childhood onset asthma, *ERBB2* expression, and DNA methylation at cg10374813, is shown as a purple diamond in each of three LocusZoom plot. Upper panel: childhood onset asthma GWAS (modified from Pividori et al. 2019). Middle panel: *ERBB2* eQTLs for vehicle-treated cultured airway epithelial cells. The association for *ERBB2* gene expression and rs2270401 for the vehicle treatment is shown in the box plot (β = -1.16x10^{-3}; 95% CI -6.40x10^{-3}, -8.73x10^{-3}). Lower panel: *ERBB2* eQTLs for RV-treated airway epithelial

Figure 4.4 Co-localization of rs2270401 with *ERBB2* expression and DNA methylation levels for cg10374813. (continued) cells. Boxplots for *ERBB2* gene expression by rs2270401 genotype is shown within the middle and lower LocusZoom plots. The association of *ERBB2* gene expression and rs2270401 for the RV treatment is shown in the box plot (β = 0.03; 95% CI 0.04, 0.02). (B) Boxplots for cg1037813 meQTLs in vehicle-treated (upper panel; β = -0.07; 95% CI -0.07, -0.09) and RV-treated (lower panel; β = -0.10; 95% CI -0.07, -0.12) cultured airway epithelial cells. (C) *ERBB2* gene expression in vehicle-treated and RV-treated cells. P-values and FDR adjusted P-values (P_{adj}) are shown in each box plot.

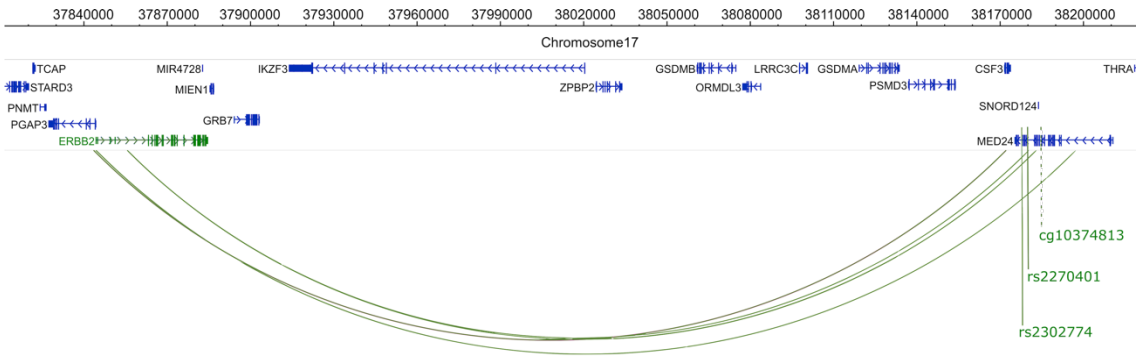


Figure 4.5 pcHi-C Interactions with *ERBB2* and *MED24* from *ex vivo* bronchial epithelial cells. pcHi-C interaction map indicating chromatin interactions (green arcs) between *ERBB2* and *MED24*. The looping is occurring between the *MED24* gene and the *ERBB2* promoters. Solid green vertical lines indicate the SNPs associated with the co-localized triplet from the childhood onset asthma (rs2270401) and TAGC (rs2302774) GWASs. The dashed vertical line shows the location of the CpG associated with this triplet. A Capture Hi-C Analysis of Genomic Organisation (CHiCAGO) score ≥ 5 was considered as evidence for chromatin interactions (range for the four loops: 5.02-7.02).

4.4.7 Mendelian randomization of multi-trait co-localized triplets

Co-localization analyses reveal genetic variants that are associated with both asthma and molecular traits (gene expression and/or DNA methylation) but the question of causality between the molecular traits remains unanswered. To estimate whether the effects of the asthma risk variant on gene expression is mediated by DNA methylation in co-localized triplets, we performed Mendelian randomization (MR) for each of the 24 triplets using a two-stage least squares regression model (2SLS) [199]. In each case, MR suggested a causal relationship between methylation and gene expression, indicating that the genotype effect at each of the co-

localized SNPs on gene expression is mediated by DNA methylation at the co-localized meCpG (Table 4.4). Moreover, four of the 20 co-localizations were significant only in the RV-treated samples. Two of the RV-specific triplets involved *ERBB2* at the 17q12-21 locus, as discussed above Two other RV-specific triplets were located on chromosome 11. Although the SNPs associated with these co-localizations (rs11227318, rs10791827) were in high LD, they were associated with two different genes (*RBM14*, *PACS1*) and two different CpGs (cg15531562, cg15531562).

The results of the MRs provide evidence for the co-localization of the triplets and novel evidence for causal inference with respect to the co-localized molecular traits (the estimated effect of DNA methylation on gene expression). These data also reinforce arguments for epigenetic mechanisms modifying gene expression, and potentially asthma risk, in response to environmental exposures [207, 208].

4.5 Discussion

One of the major challenges of complex disease genetics is to uncover molecular mechanisms of pathogenesis and to understand how genetic and environmental factors interact to influence risks for disease. While GWASs have identified thousands of SNPs associated with disease phenotypes, interpretation and downstream follow-up studies have been challenging [209]. Cell models can advance our understanding of disease pathobiology through experimental testing of disease mechanisms in a controlled environment. In this multi-omics study, we leveraged an AEC model of microbial response to identify potentially functional variants, some of which have context-specific effects on transcriptional and epigenetic responses, and molecular mechanisms of disease. We show that asthma and allergy GWAS SNPs were specifically enriched among molecular QTLs in AECs compared to NPs from non-asthma/allergy GWASs, and among AEC

Table 4.4 Mendelian randomization results of gene expression and DNA methylation identified from co-localization triplets. We considered the relationship between DNA methylation and gene expression to be causal if the adjusted P-value ≤ 0.05.

				Vehicle		RV	
SNP	SNP Position (Chr:Pos)	CpG	Gene	P	P_{adj}	P	P_{adj}
rs34372395	1:152167407	cg15025200	*FLG-AS1*	$< 1x10^{-3}$	$< 1.19x10^{-3}$	$< 1x10^{-3}$	$< 2.78x10^{-3}$
rs34372395	1:152167407	cg23107878	*FLG-AS1*	$1x10^{-3}$	$1.19x10^{-3}$	$< 1x10^{-3}$	$< 2.78x10^{-3}$
rs1552994	1:152171461	cg09127314	*FLG-AS1*	$< 1x10^{-3}$	$< 1.19x10^{-3}$	$< 1x10^{-3}$	$< 2.78x10^{-3}$
rs1552994	1:152171461	cg21280320	*FLG-AS1*	$< 1x10^{-3}$	$< 1.19x10^{-3}$	$< 1x10^{-3}$	$< 2.78x10^{-3}$
rs1552994	1:152171461	cg02754945	*FLG-AS1*	$1x10^{-3}$	$1.79x10^{-3}$	$< 1x10^{-3}$	$< 2.78x10^{-3}$
rs1552994	1:152171461	cg26320663	*FLG-AS1*	$1x10^{-3}$	$1.79x10^{-3}$	$< 1x10^{-3}$	$< 2.78x10^{-3}$
rs1552994	1:152171461	cg13498757	*FLG-AS1*	$2x10^{-3}$	$3.13x10^{-3}$	$< 1x10^{-3}$	$< 2.78x10^{-3}$
rs7545406	1:152193286	cg26879891	*FLG*	$< 1x10^{-3}$	$< 1.19x10^{-3}$	$< 1x10^{-3}$	$< 2.78x10^{-3}$
rs2689273	3:121631451	cg19216788	*SLC15A2*	$1x10^{-3}$	$1.79x10^{-3}$	$2x10^{-3}$	$2.78x10^{-3}$
rs9822474	3:121637966	cg07193051	*SLC15A2*	$< 1x10^{-3}$	$< 1.19x10^{-3}$	$2x10^{-3}$	$2.78x10^{-3}$
rs351855	5:176520243	cg19956155	*FGFR4*	$2.1x10^{-2}$	$2.76x10^{-2}$	$< 1x10^{-3}$	$< 2.78x10^{-3}$
rs3807306	7:128580680	cg26616347	*IRF5*	$< 1x10^{-3}$	$< 1.19x10^{-3}$	$< 1x10^{-3}$	$< 2.78x10^{-3}$
rs11227318	11:65592935	cg15531562	*RBM4B*	$1x10^{-3}$	$1.79x10^{-3}$	$5x10^{-2}$	$5x10^{-2}$
*rs11227318	11:65592935	cg15531562	*RBM14*	$1.15x10^{-1}$	$1.31x10^{-1}$	$< 1x10^{-3}$	$< 2.78x10^{-3}$
rs10791827	11:65596546	cg15531562	*EFEMP2*	$< 1x10^{-3}$	$< 1.19x10^{-3}$	$1.60x10^{-2}$	$1.74x10^{-2}$
*rs10791827	11:65596546	cg15531562	*PACS1*	$8.29x10^{-1}$	$8.29x10^{-1}$	$1x10^{-3}$	$1.56x10^{-3}$
rs9303281	17:38074046	cg18711369	*ORMDL3*	$1x10^{-3}$	$1.79x10^{-3}$	$1x10^{-3}$	$1.56x10^{-3}$
*rs2270401	17:38176256	cg10374813	*ERBB2*	$7.21x10^{-1}$	$7.84x10^{-1}$	$1x10^{-3}$	$1.56x10^{-3}$
*rs2302774	17:38183090	cg10374813	*ERBB2*	$7.70x10^{-1}$	$8.02x10^{-1}$	$5x10^{-3}$	$5.95x10^{-3}$
rs1029792	17:38808941	cg26165421	*SMARCE1*	$2.10x10^{-2}$	$2.76x10^{-2}$	$< 1x10^{-3}$	$< 2.78x10^{-3}$
rs7223354	17:38824672	cg02645492	*SMARCE1*	$3.60x10^{-2}$	$4.29x10^{-2}$	$1x10^{-3}$	$1.56x10^{-3}$
rs132920	22:41810170	cg19274703	*ACO2*	$2x10^{-3}$	$3.13x10^{-3}$	$5x10^{-3}$	$5.95x10^{-3}$
rs9607819	22:41958862	cg07830128	*PMM1*	$4x10^{-3}$	$5.88x10^{-3}$	$5x10^{-3}$	$5.95x10^{-3}$
rs5758461	22:42162189	cg10386501	*PMM1*	$3.10x10^{-2}$	$3.88x10^{-2}$	$2.40x10^{-2}$	$2.50x10^{-2}$

*Treatment-specific causal effects (FDR≤0.05)

eQTLs compared to eQTLs from other tissues. Finally, SNPs that were molecular QTLs in our study co-localized with asthma GWAS SNPs, revealing 46 unique co-localizations that included both known asthma loci (e.g., 17q12-21 and *TSLP*) and loci that did not meet stringent criteria for genome-wide significance in the GWASs (e.g., *IRF5* and *PMM1*) (Table S4.3).

The results of enrichment analyses further highlighted the important role of airway epithelium in asthma pathogenesis. The enrichment of childhood onset asthma and TAGC GWAS SNPs among epithelial eQTLs is particularly noteworthy, as it identified a long-range epigenetic mechanism at the 17q12-21 locus that had not been previously described. These results are consistent with previous studies suggesting that functional variants from disease-relevant tissues are more enriched among GWAS loci for those diseases [72, 210, 211]. The absence of enrichment of adult onset asthma GWAS SNPs among epithelial eQTLs may be due to the overall smaller effect sizes of SNPs at adult onset asthma loci compared to childhood onset asthma loci, to the less important role of epithelial cells in the pathophysiology of adult onset asthma, or to the greater heterogeneity and lesser heritability of adult onset asthma [13]. While we did not observe an enrichment of adult onset asthma GWAS SNPs among AEC eQTLs, SNPs were enriched among AEC molecular QTLs when we considered both eQTLs and meQTLs. This suggests that DNA methylation may be a relatively more important contributor to adult onset compared to childhood onset asthma, consistent with both the greater environmental contributions to adult onset asthma [13] and the more stable nature of DNA methylation across treatments in our study and possibly over time (e.g., [161]). Other differences between the adult onset and childhood onset asthma GWASs were observed. For example, 11 co-localizations were detected with adult onset asthma GWAS SNPs, compared to 39 with childhood onset asthma GWAS SNPs. None of the co-localizations in the adult onset GWAS included an eQTL

compared to 28 childhood onset co-localizations with eQTLs; and the 11 meQTL-GWAS pairs in the adult onset asthma GWAS were also present in the childhood onset asthma GWAS. These differences were additionally surprising because although there were 2.5-times the number of loci associated with childhood onset asthma compared to adult onset asthma in the GWASs [13], there were over 3.5-times more co-localizations in the childhood onset compared to the adult onset GWAS (39 vs. 11, respectively). These observations likely reflect the more important role of gene regulation and dysregulation in airway epithelium in the etiology of childhood onset asthma compared to adult onset asthma [184, 185]. Focusing on other asthma relevant tissues or cells (e.g., immune cells, airway smooth muscle cells) might reveal additional novel molecular mechanisms and differences between childhood onset and adult onset asthma. As with the adult onset asthma GWAS, there were fewer co-localizations using the TAGC GWAS compared to the childhood onset asthma GWAS (10 vs. 39, respectively). This is maybe be due to the inclusion of all ages of asthma onset and multiple ethnicities, fewer reported genome-wide significant loci, and/or smaller effect sizes of the associated alleles in the TAGC GWAS. These combined factors may have reduced the power to identify co-localizations with the TAGC GWAS SNPs. The enrichments of AEC molecular QTLs with neuroticism GWAS SNPs and of asthma GWAS SNPs with hypothalamus eQTLs were unexpected but may reflect the greater than expected co-occurrence of asthma and neuroticism [212, 213] and hypothalamic-pituitary-adrenal axis influence on the development of neuroticism [214, 215]. Whether these observations are due to the co-occurrence of these conditions or a shared genetic architecture cannot be discerned from our studies.

Our study provided mechanistic evidence for associations between GWAS SNPs and asthma at two important loci: the *TSLP* and 17q12-21 loci. SNPs at the *TSLP* locus have been

highly replicated in GWASs, and TSLP is recognized as having an important role in asthma pathogenesis through its broad effects on innate and adaptive immune cells promoting Th2 inflammation [216]. Previous studies have shown *TSLP* to be a methylation-sensitive gene, and hypomethylation at its promoter is associated with atopic dermatitis and prenatal tobacco smoke exposure [217, 218], two asthma-associated features. Another study showed that the rs1837253-CC genotype was associated with increased excretion of TSLP in cultured AECs after exposure to polyI:C (a dsRNA surrogate of viral stimulation) [219]. Through co-localization, we further showed that the effect of rs1837253 genotype on risk for asthma may be mediated through DNA methylation levels at CpG sites in the untranslated first exon of the *TSLP* gene in AECs, suggesting an epigenetic mechanism of disease that is robust to RV and vehicle treatment and contributes to both adult and childhood onset asthma. We were unable to identify any SNPs in LD with rs1837253 (± 50 kb) with $r^2 > 0.12$ in 1000Genomes in European or African ancestry reference panels, suggesting that this SNP may indeed be the causal asthma SNP at this important locus.

Since its discovery over a decade ago, the 17q12-21 locus has been an important focus of asthma research. Several studies have revealed the complex nature of this locus including the differences in LD structure across populations, and different gene expression patterns and eQTLs in different asthma-relevant tissues and cell types (reviewed in [28, 180]). Our AEC model of RV infection revealed additional dimensions of complexity at this locus. For example, we show that one of the candidate genes in the core region, *ORMDL3* [28], is down-regulated by RV in cultured AECs, and that an eQTL for *ORMDL3* co-localizes with a meQTL in a neighboring gene, *GSDMB*, and with a TAGC asthma GWAS SNP. This is in contrast to a recent study [180] in *ex vivo* upper airway epithelial cells, in which 17q12-21 core region SNPs were eQTLs for

GSDMB but not *ORMDL3*, although the co-localized SNP in this study (rs9303281) was not included in that study. These disparate findings could be due to differences between cultured cells, which are comprised of basal epithelial cells, and *ex vivo* cells, which are comprised of fully differentiated epithelial cells. Nonetheless, our study raises the possibility that downregulation of *ORMDL3* during infection with RV, and potentially other viruses, plays a role in the strong association of this locus with early life wheezing [12, 67, 220].

Our study further implicates for the first time a gene at the proximal end of this locus, *ERBB2*, in a genetic study of asthma. Co-localization and Mendelian randomization revealed a novel epigenetic mechanism through which a SNP at the distal boundary of the locus (in *MED24*) was associated with expression of *ERBB2* at the proximal boundary of the locus, only after exposure to RV. We observed this co-localized triplet in both the childhood onset and TAGC GWASs. Although the co-localizations identified different asthma associated alleles in each study, which were in strong LD ($r^2 = 0.94$; 1000 Genomes CEU reference panel), the same epigenetic mechanism was supported by both studies. The eQTL effect on *ERBB2* expression only in RV-treated cells was mediated through a differentially methylated CpG site also in *MED24* at the distal locus, which was present in both treatment conditions and co-localized in both GWASs (Table S4.3). The SNPs that are eQTLs for *ERBB2* in RV infected epithelial cells were associated with asthma in two GWASs (childhood onset asthma $p_{GWAS} = 8.11x10^{-29}$ [13]; TAGC $p_{GWAS} = 2.79x10^{-20}$ [11]), directly connecting the eQTL for *ERBB2* in RV-treated cells to asthma risk. The asthma associated alleles, rs2270401-A and rs2302774-G, respectively, were each associated with decreased expression of *ERBB2* after RV infection (Fig 4.4A), consistent with results of a study in 155 asthma cases and controls reporting an inverse correlation between *ERBB2* expression in *ex vivo* lower AECs and asthma severity [221]. These combined data

suggest that decreased expression of *ERBB2*, which is associated with asthma severity, may be modulated by RV, the most common trigger of asthma exacerbations, via epigenetic mechanisms involving DNA methylation and long-range chromatin looping between the proximal and distal ends of this important locus, which was supported by pcHi-C studies.

The significance threshold ($p<5x10^{-8}$) required to control the false discovery rate in GWASs likely excludes many true associations that do not reach this stringent cutoff. However, distinguishing true from false positive signals for variants among the mid-hanging fruit (e.g., p-values between 10^{-5} and $> 10^{-8}$) can be challenging [222]. We and others have suggested that these SNPs may be environment- or context-specific associations that are missed in GWASs that typically do not control for either [222-224]. Notably, over 52% (24 of 46) of the co-localizations were with a GWAS SNP that did not meet genome-wide significance (five in the adult onset asthma, 16 in the childhood onset asthma, and three in TAGC GWAS). This may be due to the variants having exposure-specific, tissue-specific, or endotype-specific effects, which are heterogeneous among subjects included in GWASs. Therefore, annotating SNPs among the mid-hanging fruit for functionality provides more confidence to these findings, a more complete picture of the genetic architecture of asthma, and a model for prioritizing these loci for further studies.

Although our study provides novel observations about epigenetic mechanisms of asthma risk alleles, there are some limitations. First, the sample sizes for the eQTL and meQTL studies were smaller than the sample sizes recommended by moloc ($n_{min}=300$) [181]. In such cases, *moloc* can miss true co-localizations in QTL datasets. For example, an eQTL-GWAS pair may, in reality, be an eQTL-meQTL-GWAS triplet that we were under powered to detect. As a result, the eQTL-GWAS and meQTL-GWAS pairs that we identified could be eQTL-meQTL-GWAS

triplets or we may have missed other co-localizations entirely. This may be evidenced by the fact that only a single meQTL co-localized with a GWAS SNP at the *TSLP* locus in the TAGC GWAS, while the same SNP, rs1837253, was co-localized with two additional meQTLs for two additional CpGs in the childhood and adult onset asthma (Fig S4.6), representing potential contributors to asthma disease mechanisms that were missed in the TAGC GWAS. Additionally, the designation of the co-localized triplet with *ORMDL3* at the 17q12-21 locus as RV-specific is also likely related to the reduced power to detect the same co-localization in the vehicle-treated cells. The fact that the same SNP is an eQTL for *ORMDL3* in both vehicle- and RV-treated cells (Table S4.3), and that the MR results indicated that methylation at cg18711369 mediated *ORMDL3* expression in both the vehicle- and RV-treatments (Table 4.4), suggests that this co-localization may not in fact be RV-specific. Nonetheless, the 46 unique co-localizations detected in our study are likely to be real as we were able to replicate many of our findings in two or all three GWASs, including those for *ERBB2* and *TSLP*, respectively (Table S4.3). Future studies in larger samples will increase confidence in our findings. Second, we focused on one cell type (upper airway sinonasal epithelium), two exposures (vehicle and RV), and one epigenetic mark (DNA methylation). It is possible that other asthma-relevant co-localizations are specific to other tissues or cell types or to other exposures or culture conditions, and that additional epigenetic marks, such as those associated with chromatin accessibility, would be additionally informative. Finally, because of the limited sample size, we did not test for QTLs that differed between subjects with and without asthma or perform ethnic-specific analyses. As a result, we likely only identified QTLs that are robust to disease status, and potentially to ethnicity. However, utilizing the largest sample possible to identify treatment-specific molecular QTLs increased our power to differentiate molecular responses to RV infection.

In summary, we identified *cis*-eQTLs and *cis*-meQTLs in an AEC model of host cell response to RV and integrated those data with three large asthma GWASs to assign potential molecular mechanisms for variants associated with asthma. By combining enrichment studies, co-localization analysis, Mendelian randomization, and pcHi-C, we provide robust statistical, and experimental evidence of epigenetic mechanisms in upper airway cells contributing to childhood onset asthma. We demonstrate that a multi-omics approach using a disease-relevant cell type and disease-relevant exposure allows prioritization of disease-associated variants and provides insight into potential epigenetic mechanisms of asthma pathogenesis.

4.6 Supplementary Information

4.6.1 Supplementary Tables

Table S4.1 Enrichment estimates of eQTLs for TAGC asthma GWAS SNPs from six tissues. P-values that are significant after FDR correction are shown in bolded type.

Tissue	GWAS Threshold	OR	P	P_{adj}	N	N_{eSNP}
Adrenal	1.00E-05	0.68	3.20E-01	5.12E-01	175	588,348
	1.00E-06	0.68	4.00E-01	5.65E-01		
	1.00E-07	0.62	3.80E-01	5.65E-01		
	1.00E-08	0.86	7.90E-01	8.24E-01		
Brain - Frontal Cortex	1.00E-05	0.53	2.30E-01	5.12E-01	118	367,312
	1.00E-06	0.34	1.40E-01	4.00E-01		
	1.00E-07	0.23	1.50E-01	4.00E-01		
	1.00E-08	0.31	2.50E-01	5.12E-01		
Brain - Hypothalamus	1.00E-05	0.83	7.30E-01	8.24E-01	108	251,506
	1.00E-06	0.84	7.70E-01	8.24E-01		
	1.00E-07	0.35	3.10E-01	5.12E-01		
	1.00E-08	0.46	4.60E-01	6.13E-01		
Ovary	1.00E-05	1.02	9.60E-01	9.60E-01	122	292,461
	1.00E-06	1.17	7.40E-01	8.24E-01		
	1.00E-07	1.30	6.40E-01	8.08E-01		
	1.00E-08	1.78	3.20E-01	5.12E-01		
Testis	1.00E-05	0.38	4.21E-03	1.68E-02	225	1,358,512
	1.00E-06	0.18	4.41E-04	3.53E-03		
	1.00E-07	0.14	1.78E-03	1.07E-02		
	1.00E-08	0.19	8.51E-03	2.92E-02		
Airway Epithelial Cells	1.00E-05	6.12	7.14E-07	1.71E-05	104	185,407
	1.00E-06	7.23	1.68E-06	2.02E-05		
	1.00E-07	5.02	2.35E-03	1.13E-02		
	1.00E-08	2.17	3.13E-01	5.12E-01		

Table S4.2 Enrichment estimates of eQTLs for adult onset asthma GWAS SNPs from six tissues. P-values that are significant after FDR correction are shown in bolded type.

Tissue	GWAS Threshold	OR	P	P_{adj}	N	N_{eSNP}
Adrenal	1.00E-05	0.84	6.78E-01	7.75E-01	175	588,348
	1.00E-06	0.73	5.60E-01	7.67E-01		
	1.00E-07	0.75	6.39E-01	7.67E-01		
	1.00E-08	0.68	6.19E-01	7.67E-01		
Brain - Frontal Cortex	1.00E-05	0.58	3.72E-01	6.38E-01	118	367,312
	1.00E-06	0.94	9.17E-01	9.57E-01		
	1.00E-07	136	6.25E-01	7.67E-01		
	1.00E-08	1.24	7.76E-01	8.47E-01		
Brain - Hypothalamus	1.00E-05	1.98	1.27E-01	4.83E-01	108	251,506
	1.00E-06	3.44	8.32E-03	2.00E-01		
	1.00E-07	3	5.14E-02	4.58E-01		
	1.00E-08	3.28	6.98E-02	4.58E-01		
Ovary	1.00E-05	0.22	1.41E-01	4.83E-01	122	292,461
	1.00E-06	0.35	3.04E-01	5.61E-01		
	1.00E-07	0.50	4.96E-01	7.44E-01		
	1.00E-08	0	9.93E-01	9.93E-01		
Testis	1.00E-05	0.55	7.90E-02	4.58E-01	225	1,358,512
	1.00E-06	0.62	2.42E-01	4.84E-01		
	1.00E-07	0.55	2.13E-01	4.84E-01		
	1.00E-08	0.49	2.19E-01	4.84E-01		
Airway Epithelial Cells	1.00E-05	3.18	1.87E-01	4.84E-01	104	185,407
	1.00E-06	1.76	4.54E-01	7.26E-01		
	1.00E-07	2.44	2.40E-01	4.84E-01		
	1.00E-08	3.72	9.55E-02	4.58E-01		

Table S4.3 *moloc* results indicating molecular QTL-GWAS pairs and triplets.

									eQTL p-values		meQTL p-value	
Gene	**CpG**	**rsID**	**SNP Position**	**CpG Position**	**GWAS pval**	**PPA**	**Risk Allele**	**Treatment**	**rv.eqtl npval**	**vehrv.etl npval**	**rv.meqtl npval**	**vehrv.meqtl npval**
Adult onset asthma												
-	cg22356347	rs2949661	1:167424924	1:167427500	2.28E-06	0.5, 0.50	C	rv, vehicle	-	-	3.38E-05	8.37E-06
-	cg10931190	rs1837253	5:110401872	5:110406506	2.77E-13	0.5	C	rv	-	-	1.72E-05	3.87E-05
-	cg15089387	rs1837253	5:110401872	5:110406661	2.77E-13	0.61, 0.50	C	rv, vehicle	-	-	7.91E-06	1.66E-05
-	cg15557878	rs1837253	5:110401872	5:110406049	2.77E-13	0.81, 0.81	C	rv, vehicle	-	-	3.63E-04	1.09E-06
-	cg00786977	rs72925996	6:90930513	6:90926599	4.53E-11	0.58	T	rv	-	-	4.71E-06	1.97E-04
-	cg12606193	rs9408638	9:6096931	9:6097201	1.04E-10	0.53, 0.53	G	rv, vehicle	-	-	6.35E-06	1.57E-07
-	cg17647271	rs10160518	11:76296671	11:76299819	8.88E-09	0.80, 0.80	G	rv, vehicle	-	-	9.16E-08	1.92E-09
-	cg25115154	rs11178644	12:71523541	12:71533408	4.43E-07	0.51, 0.51	T	rv, vehicle	-	-	5.99E-04	8.96E-06
-	cg13355032	rs3851611	12:71524042	12:71524030	4.25E-07	0.68, 0.68	C	rv, vehicle	-	-	1.63E-05	1.51E-08
-	cg01699148	rs9532433	13:40352671	13:40350843	2.52E-05	0.54	A	rv	-	-	4.30E-07	2.15E-05
-	cg05185748	rs2296860	13:99852753	13:99849620	3.19E-05	0.53	G	rv	-	-	9.62E-07	7.51E-05
Childhood onset asthma												
FLG-AS1	-	rs12130219	1:152162106	-	9.28E-14	0.50, 0.50	A	rv, vehicle	4.30E-06	2.05E-05	-	-
FLG-AS1	cg15025200	rs34372395	1:152167407	1:152161237	3.46E-14	0.73	A	rv	4.31E-06	2.04E-05	1.04E-04	3.06E-03
FLG-AS1	cg23107878	rs34372395	1:152167407	1:152161397	3.46E-14	0.90, 0.71	A	rv, vehicle	4.31E-06	2.04E-05	8.90E-06	6.71E-05
FLG-AS1	cg09127314	rs1552994	1:152171461	1:152161683	3.21E-14	0.50	A	rv	4.31E-06	2.01E-05	8.67E-04	2.47E-03
FLG-AS1	cg21280320	rs1552994	1:152171461	1:152162025	3.21E-14	0.50	A	rv	4.31E-06	2.01E-05	7.95E-04	1.12E-05
FLG-AS1	cg13498757	rs1552994	1:152171461	1:152161927	3.21E-14	0.69	A	rv	4.31E-06	2.01E-05	1.61E-04	7.92E-06
FLG-AS1	cg02754945	rs1552994	1:152171461	1:152161885	3.21E-14	0.77	A	rv	4.31E-06	2.01E-05	7.99E-05	1.59E-04
FLG-AS1	cg26320663	rs1552994	1:152171461	1:152161521	3.21E-14	0.79, 0.77	A	rv, vehicle	4.31E-06	2.01E-05	6.11E-05	3.18E-05
FLG	cg26879891	rs7545406	1:152193286	1:152191343	7.24E-13	0.52	T	rv	3.75E-07	3.39E-05	3.85E-04	1.18E-01
FLG	-	rs7545406	1:152193286	-	7.24E-13	0.75	T	rv	3.75E-07	3.39E-05	-	-
-	cg22356347	rs2949661	1:167424924	1:167427500	3.32E-05	0.50, 0.50	C	rv, vehicle	-	-	3.38E-05	8.37E-06
SLC15A2	cg19216788	rs2689273	3:121631451	3:121633802	5.26E-05	0.61, 0.61	T	rv, vehicle	1.23E-03	4.97E-06	8.87E-04	2.27E-05
SLC15A2	cg07193051	rs9822474	3:121637966	3:121633332	7.31E-05	0.54	A	vehicle	1.20E-03	2.91E-06	3.61E-01	6.85E-05

Table S4.3 *moloc* results indicating molecular QTL-GWAS pairs and triplets. (continued)

-	cg10931190	rs1837253	5:110401872	5:110406506	2.33E-27	0.50	C	rv	-	-	1.72E-05	3.87E-05
-	cg15089387	rs1837253	5:110401872	5:110406661	2.33E-27	0.61, 0.50	C	rv, vehicle	-	-	7.91E-06	1.66E-05
-	cg15557878	rs1837253	5:110401872	5:110406049	2.33E-27	0.81, 0.81	C	rv, vehicle	-	-	3.63E-04	1.09E-06
FGFR4	cg19956155	rs351855	5:176520243	5:176511416	1.04E-06	0.61	A	rv	4.69E-07	5.63E-03	2.96E-04	2.64E-03
FGFR4	-	rs351855	5:176520243	-	1.04E-06	0.74	A	rv	4.69E-07	5.63E-03	-	-
-	cg00786977	rs72925996	6:90930513	6:90926599	3.61E-09	0.58	T	rv	-	-	4.71E-06	1.97E-04
IRF5	-	rs3823536	7:128579666	-	1.03E-04	0.52, 0.52	G	rv, vehicle	8.53E-10	5.51E-09	-	-
IRF5	cg26616347	rs3807306	7:128580680	7:128577752	4.51E-05	0.60	G	rv	2.94E-10	1.63E-08	3.28E-04	1.22E-03
-	cg12606193	rs9408638	9:6096931	9:6097201	2.71E-23	0.53, 0.53	G	rv, vehicle	-	-	6.35E-06	1.57E-07
EFEMP2	-	rs7123155	11:65575056	-	1.11E-05	0.50, 0.50	C	rv, vehicle	1.35E-03	2.76E-06	-	-
RBM4B	cg15531562	rs11227318	11:65592935	11:65601754	1.42E-06	0.50	A	vehicle	1.88E-01	1.25E-03	7.51E-07	9.43E-07
RBM14	cg15531562	rs11227318	11:65592935	11:65601754	1.42E-06	0.80	A	rv	1.02E-04	9.86E-02	7.51E-07	9.43E-07
EFEMP2	cg15531562	rs10791827	11:65596546	11:65601754	1.40E-06	0.53, 0.98	A	rv, vehicle	1.21E-03	1.01E-06	6.36E-07	8.07E-07
PACS1	cg15531562	rs10791827	11:65596546	11:65601754	1.40E-06	0.67	A	rv	4.10E-04	6.15E-01	6.36E-07	8.07E-07
-	cg17647271	rs10160518	11:76296671	11:76299819	1.38E-28	0.80, 0.80	A	rv, vehicle	-	-	9.16E-08	1.92E-09
-	cg25115154	rs11178644	12:71523541	12:71533408	1.57E-07	0.51, 0.51	T	rv, vehicle	-	-	5.99E-04	8.96E-06
-	cg13355032	rs3851611	12:71524042	12:71524030	1.53E-07	0.68, 0.68	C	rv, vehicle	-	-	1.63E-05	1.51E-08
-	cg05185748	rs2296860	13:99852753	13:99849620	7.06E-11	0.53	G	rv	-	-	9.62E-07	7.51E-05
ERBB2	cg10374813	rs2270401	17:38176256	17:38183790	8.11E-29	0.56	A	rv	1.55E-04	8.78E-01	2.10E-05	3.20E-03
SMARCE1	cg26165421	rs1029792	17:38808941	17:38805157	1.80E-07	0.56	T	rv	7.71E-05	1.28E-02	3.29E-05	4.82E-04
SMARCE1	cg02645492	rs7223354	17:38824672	17:38827676	1.69E-07	0.60	A	rv	1.06E-04	1.73E-02	8.60E-06	5.62E-06
ACO2	cg19274703	rs132920	22:41810170	22:41806451	1.23E-06	0.57, 0.57	T	rv, vehicle	4.53E-04	2.48E-04	7.29E-04	1.14E-09
PMM1	cg07830128	rs9607819	22:41958862	22:41965204	7.77E-06	0.7, 0.70	C	rv, vehicle	1.60E-03	2.22E-05	1.01E-02	1.06E-06
PMM1	cg10386501	rs5758461	22:42162189	22:42159248	1.48E-05	0.59, 0.59	G	rv, vehicle	2.98E-03	1.06E-04	1.66E-03	3.41E-07
TAGC												
-	cg01613557	rs2071008	2:102615279	2:102616279	2.07E-06	0.56, 0.56	G	rv, vehicle	-	-	2.91E-04	2.49E-06
-	cg15557878	rs1837253	5:110401872	5:110406049	2.03E-25	0.56, 0.56	C	rv, vehicle	-	-	3.63E-04	1.09E-06
-	cg00255919	rs839	5:131819126	5:131827918	1.02E-06	0.76, 0.76	T	rv, vehicle	-	-	1.20E-06	4.23E-08

Table S4.3 *moloc* results indicating molecular QTL-GWAS pairs and triplets. (continued)

-	cg21138405	rs839	5:131819126	5:131827807	1.02E-06	0.71, 0.71	T	rv, vehicle	-	-	5.21E-09	2.22E-09
-	cg12606193	rs437389	9:6099531	9:6097201	8.92E-12	0.67, 0.67	C	rv, vehicle	-	-	3.83E-05	5.17E-07
-	cg02419362	rs668622	12:121198299	12:121203948	4.39E-06	0.61	G	rv	-	-	1.08E-06	2.09E-05
-	cg19468883	rs668622	12:121198299	12:121198447	4.39E-06	0.59, 0.59	G	rv, vehicle	-	-	9.54E-06	1.28E-06
ORMDL3	cg18711369	rs9303281	17:38074046	17:38081186	6.20E-45	0.62	A	rv	1.10E-04	3.91E-04	6.62E-05	3.80E-03
ERBB2	cg10374813	rs2302774	17:38183090	17:38183790	2.79E-20	0.67	G	rv	5.75E-04	9.49E-01	1.66E-06	1.16E-03
-	cg10374813	rs2302774	17:38183090	17:38183790	2.79E-20	0.75	G	rv	-	-	1.66E-06	1.16E-03

Table S4.4 Asthma GWAS risk allele effects on gene expression and DNA methylation.

Gene	CpG	rsID	SNP Position	CpG Position	$Ref_{e/meSNP}$	$Alt_{e/meSNP}$	$Beta_{eQTL}$	$Beta_{meQTL}$	†$Asthma\ Risk\ Allele\ Effect_{eQTL}$	†$Asthma\ Risk\ Allele\ Effect_{meQTL}$	Treatment
Adult onset asthma											
-	cg22356347	rs2949661	1:167424924	1:167427500	T	C*	-	-0.211948		↑	rv
-	cg22356347	rs2949661	1:167424924	1:167427500	T	C*	-	-0.247842		↑	vhiele
-	cg10931190	rs1837253	5:110401872	5:110406506	T	C*	-	-0.231354		↑	rv
-	cg15089387	rs1837253	5:110401872	5:110406661	T	C*	-	-0.194164		↑	rv
-	cg15089387	rs1837253	5:110401872	5:110406661	T	C*	-	-0.207965		↑	vehicle
-	cg15557878	rs1837253	5:110401872	5:110406049	T	C*	-	-0.180076		↑	rv
-	cg15557878	rs1837253	5:110401872	5:110406049	T	C*	-	-0.255426		↑	vehicle
-	cg00786977	rs72925996	6:90930513	6:90926599	C	T*	-	0.39633		↓	rv
-	cg12606193	rs9408638	9:6096931	9:6097201	A	G*	-	-0.35169		↑	rv
-	cg12606193	rs9408638	9:6096931	9:6097201	A	G*	-	-0.424378		↑	vehicle
-	cg17647271	rs10160518	11:76296671	11:76299819	G*	A	-	-0.423118		↓	rv
-	cg17647271	rs10160518	11:76296671	11:76299819	G*	A	-	-0.524326		↓	vehicle
-	cg25115154	rs11178644	12:71523541	12:71533408	C	T*	-	-0.145403		↑	rv
-	cg25115154	rs11178644	12:71523541	12:71533408	C	T*	-	-0.188861		↑	vehicle
-	cg13355032	rs3851611	12:71524042	12:71524030	G	C*	-	0.488457		↓	rv
-	cg13355032	rs3851611	12:71524042	12:71524030	G	C*	-	0.60893		↓	vehicle
-	cg05185748	rs2296860	13:99852753	13:99849620	G*	A	-	0.179284		↑	rv
Childhood onset asthma											
FLG-AS1	-	rs12130219	1:152162106	-	G	A*	0.320388	-	↓		rv
FLG-AS1	-	rs12130219	1:152162106	-	G	A*	0.290545	-	↓		vehicle
FLG-AS1	cg15025200	rs34372395	1:152167407	1:152161237	G	A*	0.320336	-0.206689	↓	↑	rv
FLG-AS1	cg23107878	rs34372395	1:152167407	1:152161397	G	A*	0.320336	-0.241467	↓	↑	rv
FLG-AS1	cg23107878	rs34372395	1:152167407	1:152161397	G	A*	0.290537	-0.229604	↓	↑	vehicle
FLG-AS1	cg09127314	rs1552994	1:152171461	1:152161683	G	A*	0.320507	-0.124623	↓	↑	rv

Table S4.4 Asthma GWAS risk allele effects on gene expression and DNA methylation. (continued)

FLG-AS1	cg21280320	rs1552994	1:152171461	1:152162025	G	A*	0.320507	-0.129427	↓	↑	rv
FLG-AS1	cg13498757	rs1552994	1:152171461	1:152161927	G	A*	0.320507	-0.307961	↓	↑	rv
FLG-AS1	cg02754945	rs1552994	1:152171461	1:152161885	G	A*	0.320507	-0.15468	↓	↑	rv
FLG-AS1	cg26320663	rs1552994	1:152171461	1:152161521	G	A*	0.320507	-0.144778	↓	↑	rv
FLG-AS1	cg26320663	rs1552994	1:152171461	1:152161521	G	A*	0.290898	-0.170867	↓	↑	vehicle
FLG	cg26879891	rs7545406	1:152193286	1:152191343	G	T*	-0.349072	-0.0789825	↑	↑	rv
FLG	-	rs7545406	1:152193286	-	G	T*	-0.349072	-	↑		rv
-	cg22356347	rs2949661	1:167424924	1:167427500	T	C*	-	-0.211948		↑	rv
-	cg22356347	rs2949661	1:167424924	1:167427500	T	C*	-	-0.247842		↑	vehicle
SLC15A2	cg19216788	rs2689273	3:121631451	3:121633802	G	T*	-0.112812	0.212604	↑	↓	rv
SLC15A2	cg19216788	rs2689273	3:121631451	3:121633802	G	T*	-0.146822	0.254206	↑	↓	vehicle
SLC15A2	cg07193051	rs9822474	3:121637966	3:121633332	G	A*	-0.151148	0.192425	↑	↓	vehicle
-	cg10931190	rs1837253	5:110401872	5:110406506	T	C*	-	-0.231354		↑	rv
-	cg15089387	rs1837253	5:110401872	5:110406661	T	C*	-	-0.194164		↑	rv
-	cg15089387	rs1837253	5:110401872	5:110406661	T	C*	-	-0.207965		↑	vehicle
-	cg15557878	rs1837253	5:110401872	5:110406049	T	C*	-	-0.180076		↑	rv
-	cg15557878	rs1837253	5:110401872	5:110406049	T	C*	-	-0.255426		↑	vehicle
FGFR4	cg19956155	rs351855	5:176520243	5:176511416	A*	G	-0.30394	-0.143783	↓	↓	rv
FGFR4	-	rs351855	5:176520243	-	A*	G	-0.30394	-	↓		rv
-	cg00786977	rs72925996	6:90930513	6:90926599	C	T*	-	0.39633		↓	rv
IRF5	-	rs3823536	7:128579666	-	A	G*	0.483761	-	↓		rv
IRF5	cg26616347	rs3807306	7:128580680	7:128577752	T	G*	0.496956	-0.115496	↓	↑	rv
-	cg12606193	rs9408638	9:6096931	9:6097201	A	G*	-	-0.351699		↑	rv
-	cg12606193	rs9408638	9:6096931	9:6097201	A	G*	-	-0.424378		↑	vehicle
EFEMP2	-	rs7123155	11:65575056	-	C*	T	-0.0846154	-	↓		rv
EFEMP2	-	rs7123155	11:65575056	-	C*	T	-0.128349	-	↓		vehicle
RBM4B	cg15531562	rs11227318	11:65592935	11:65601754	A*	C	-0.057114	-0.312265	↓	↓	vehicle

Table S4.4 Asthma GWAS risk allele effects on gene expression and DNA methylation. (continued)

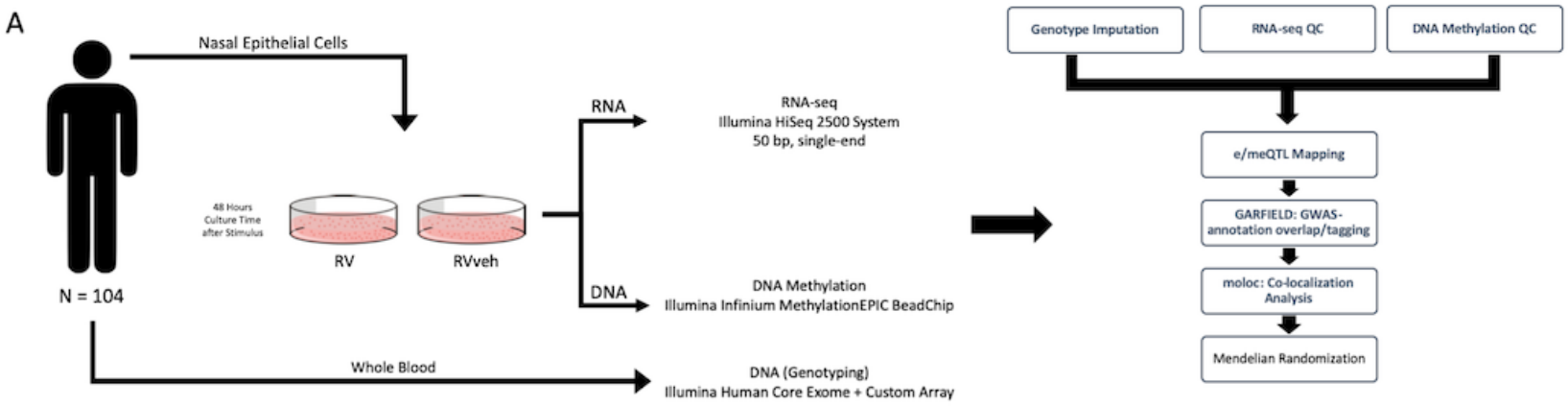
RBM14	cg15531562	rs11227318	11:65592935	11:65601754	A*	C	-0.0772074	-0.336711	↓	↓	rv
EFEMP2	cg15531562	rs10791827	11:65596546	11:65601754	A*	C	-0.0818488	-0.339907	↓	↓	rv
EFEMP2	cg15531562	rs10791827	11:65596546	11:65601754	A*	C	-0.127673	-0.315138	↓	↓	vehicle
PACS1	cg15531562	rs10791827	11:65596546	11:65601754	A*	C	0.0492589	-0.339907	↑	↓	rv
-	cg17647271	rs10160518	11:76296671	11:76299819	G*	A	-	-0.423118		↓	rv
-	cg17647271	rs10160518	11:76296671	11:76299819	G*	A	-	-0.524326		↓	vehicle
-	cg25115154	rs11178644	12:71523541	12:71533408	C	T*	-	-0.145403		↑	rv
-	cg25115154	rs11178644	12:71523541	12:71533408	C	T*	-	-0.188861		↑	vehicle
-	cg13355032	rs3851611	12:71524042	12:71524030	G	C*	-	0.488457		↓	rv
-	cg01699148	rs9532433	13:40352671	13:40350843	A*	C	-	0.361387		↑	rv
-	cg05185748	rs2296860	13:99852753	13:99849620	G*	A	-	0.179284		↑	rv
ERBB2	cg10374813	rs2270401	17:38176256	17:38183790	T	A*	0.0292639	-0.0961591	↓	↑	rv
SMARCE1	cg26165421	rs1029792	17:38808941	17:38805157	T*	G	-0.0535702	-0.221494	↓	↓	rv
SMARCE1	cg02645492	rs7223354	17:38824672	17:38827676	A*	G	-0.0518841	0.172259	↓	↑	rv
ACO2	cg19274703	rs132920	22:41810170	22:41806451	C	T*	0.0569197	-0.092497	↓	↑	rv
ACO2	cg19274703	rs132920	22:41810170	22:41806451	C	T*	0.0552383	-0.168238	↓	↑	vehicle
PMM1	cg07830128	rs9607819	22:41958862	22:41965204	C*	G	-0.0545906	-0.0766181	↓	↓	rv
PMM1	cg07830128	rs9607819	22:41958862	22:41965204	C*	G	-0.070624	-0.137978	↓	↓	vehicle
PMM1	cg10386501	rs5758461	22:42162189	22:42159248	G*	C	-0.0513037	0.209757	↓	↑	rv
PMM1	cg07830128	rs9607819	22:41958862	22:41965204	C*	G	-0.070624	-0.137978	↓	↓	vehicle
TAGC											
-	cg01613557	rs2071008	2:102615279	2:102616279	G*	T	-	0.18786		↑	rv
-	cg01613557	rs2071008	2:102615279	2:102616279	G*	T	-	0.21031		↑	vehicle
-	cg15557878	rs1837253	5:110401872	5:110406049	T	C*	-	-0.180076		↑	rv
-	cg15557878	rs1837253	5:110401872	5:110406049	T	C*	-	-0.255426		↑	vehicle
-	cg00255919	rs839	5:131819126	5:131827918	T*	C	-	0.215952		↑	rv
-	cg00255919	rs839	5:131819126	5:131827918	T*	C	-	0.246712		↑	vehicle

Table S4.4 Asthma GWAS risk allele effects on gene expression and DNA methylation. (continued)

-	cg21138405	rs839	5:131819126	5:131827807	T*	C	-	0.615385		↑	rv
-	cg21138405	rs839	5:131819126	5:131827807	T*	C	-	0.573503		↑	vehicle
-	cg12606193	rs437389	9:6099531	9:6097201	T	C*	-	-0.335459		↑	rv
-	cg02419362	rs668622	12:121198299	12:121203948	A	G*	-	0.265769		↓	rv
-	cg19468883	rs668622	12:121198299	12:121198447	A	G*	-	0.218421		↓	rv
-	cg19468883	rs668622	12:121198299	12:121198447	A	G*	-	0.279779		↓	vehicle
ORMDL3	cg18711369	rs9303281	17:38074046	17:38081186	A*	G	0.063754	-0.142218	↑	↓	rv
ERBB2	cg10374813	rs2302774	17:38183090	17:38183790	T	G*	0.0262859	-0.104907	↓	↑	rv
-	cg10374813	rs2302774	17:38183090	17:38183790	T	G*	-	-0.104907		↑	rv

*Risk allele in adult onset or childhood onset asthma GWASs

†Arrows indicate whether the asthma risk allele is associated with increased or decreased gene expression or DNA methylation relative to the protective allele



B

Sample composition

SQTL Study	Combined	Gene Expression	DNA methylation
Sample Size (N)	104	95	103
Gender (% Female)	39	41	33
Mean Age (Range)	44 (18 – 73)	45 (18 - 73)	45 (18 - 73)
% Asthma (Ever)	43	44	41
%CRSwNP	47	50	47
%Atopic	51	53	50

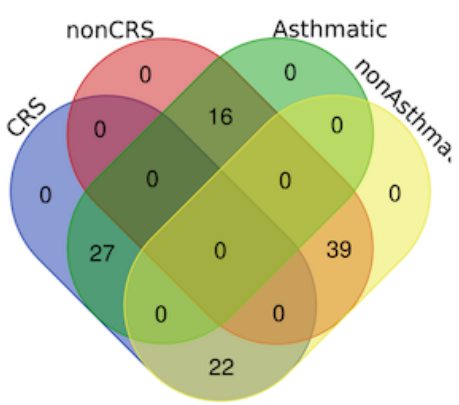


Figure S4.1 Overview of the e/meQTL and co-localization studies in NECs treated with RV. (A) Experimental design used to identify treatment-specific e/meQTLs in upper airway epithelial cells (AECs) from 104 individuals. (B) Sample composition. (C) Venn diagram showing the overlap of asthma and CRS status of the 104 study participants.

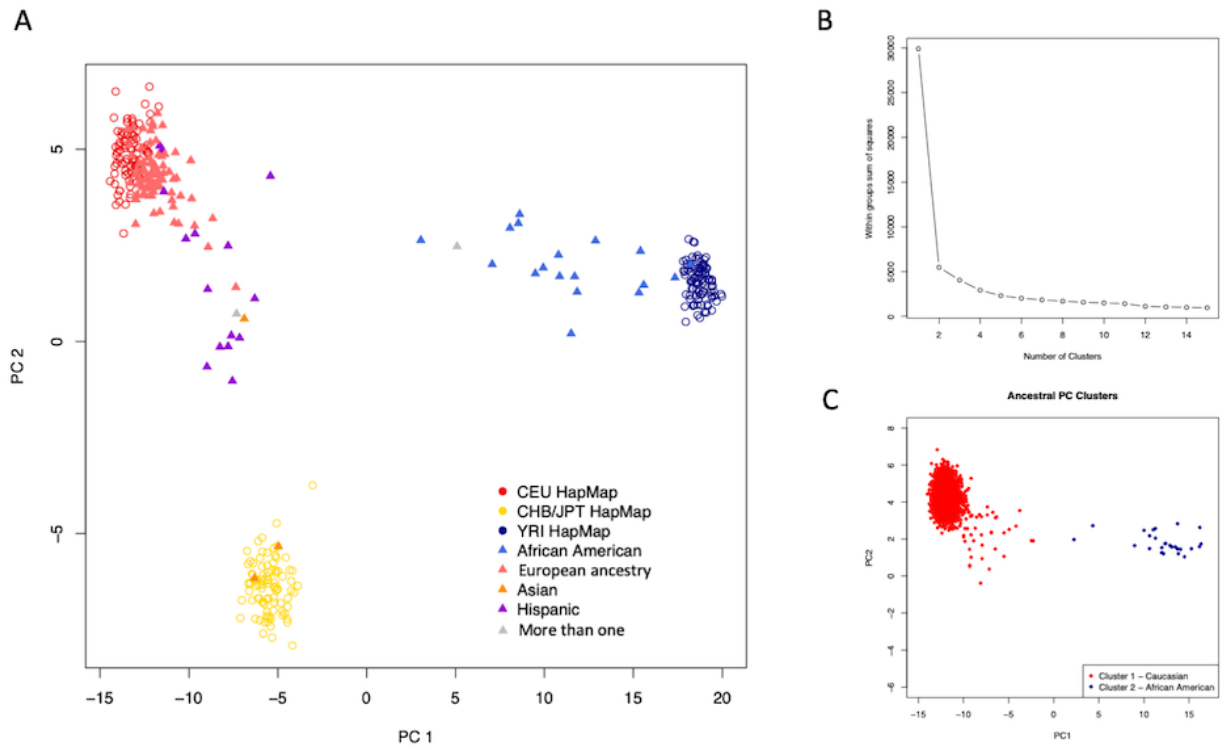


Figure S4.2 PCA and k-means clustering of genotypes. (A) PCA plot of study participant's genotypes (circles) projected on HapMap genotypes (squares). (B) Scree plot of k-means clustering of ancestral PCs in which the within groups sum of squares (y-axis) is plotted against the number of potential group clusters (x-axis); using the 'elbow criterion', it is determined that two clusters are best representative of how many clusters study samples can be grouped into for imputation. (C) PCA plot of study participants grouped into two cluster for genotype imputation, European (red), and African American (Blue), according to the k-means clustering criterion.

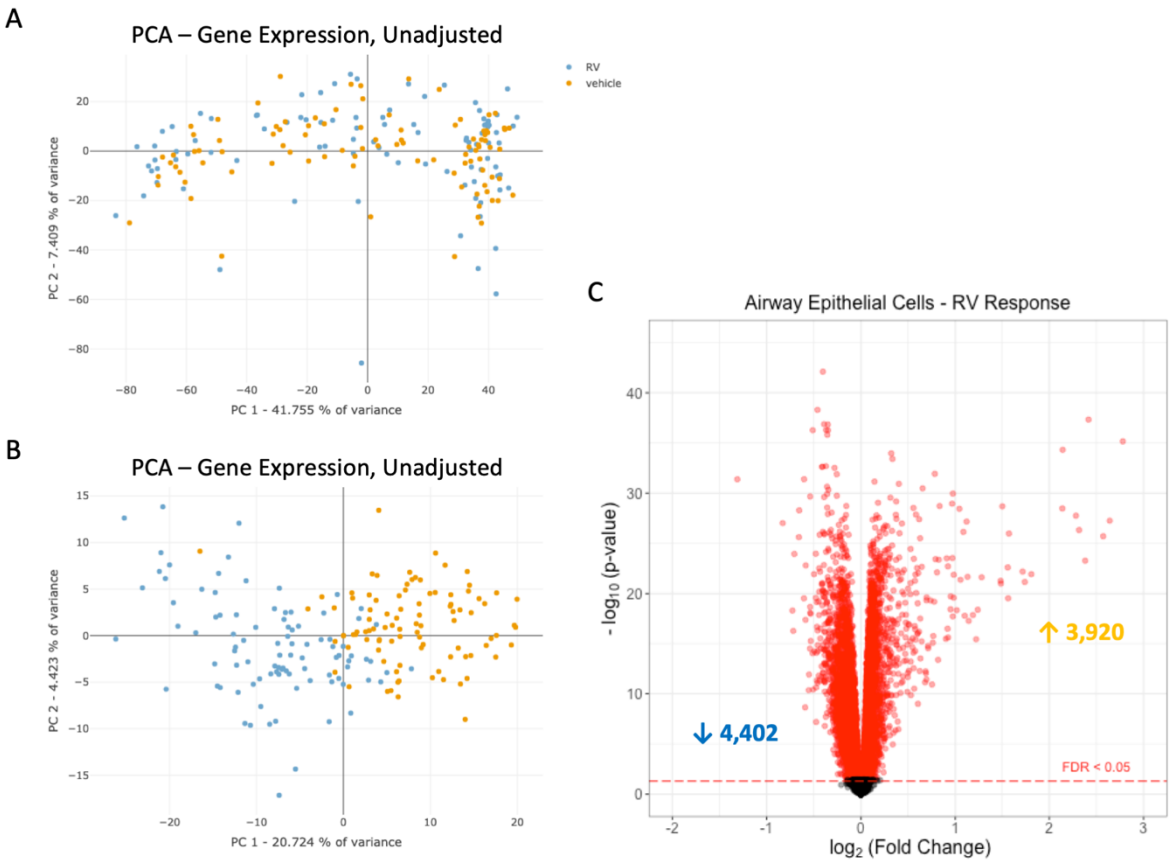


Figure S4.3 PCA of gene expression in vehicle and RV-treated epithelial cells. (A) PCA plot of epithelial cell gene expression from 95 individuals treated with vehicle and RV before regressing out covariates. (B) PCA plot of gene expression in vehicle- and RV-treated cells after regressing out covariates. (C) Volcano plot showing treatment responses in epithelial cells were detected in the combined sample with 8,322 differentially expressed genes identified at a FDR≤0.05.

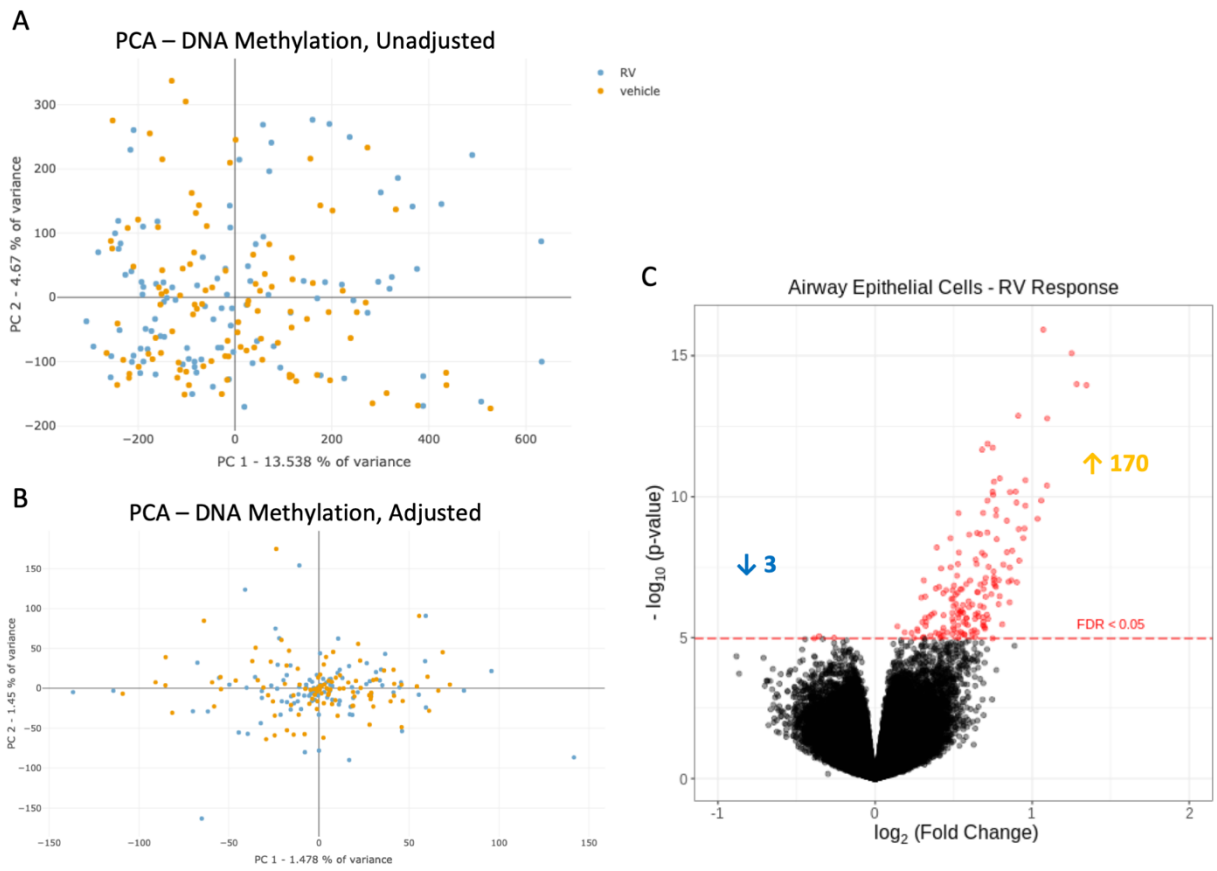
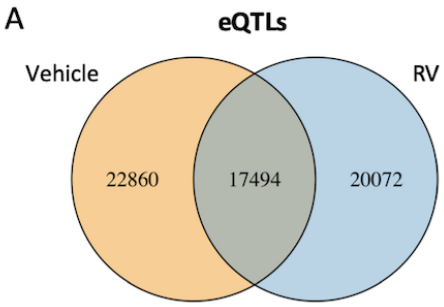
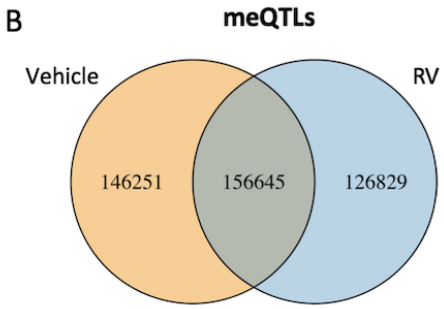


Figure S4.4 PCA of DNA methylation in vehicle- and RV-treated in cultured airway epithelial cells. (A) PCA plot of cultured airway epithelial DNA methylation from 103 individuals treated with vehicle and RV before regressing out covariates. (B) PCA plot of DNA methylation in vehicle- and RV-treated cells after regressing out covariates. Tables showing p-values of correlation with PCs and covariates before (C) Volcano plot showing treatment responses were detected in the combined sample with 173 differentially methylated CpGs at a FDR<0.05





C

eQTL and meQTL mapping results (FDR<0.05)

	Vehicle	RV
eQTLs	40.354	37,566
SNPs	38,461	35,989
Genes	1,174	1,043
meQTLs	302,896	283,474
SNPs	219,905	206,583
CpGs	26,644	25,418

Figure S4.5 Summary results for molecular QTL mappings. Venn diagrams of eQTLs (A) and meQTLs (B) in each condition (FDR<0.05). (C) Summary of eQTL and meQTL mapping results for each treatment condition. The number of SNPs associated with the gene expression of at least one gene or CpG and the number of genes or CpGs whose expression or DNA methylation levels was associated with at least one SNP.

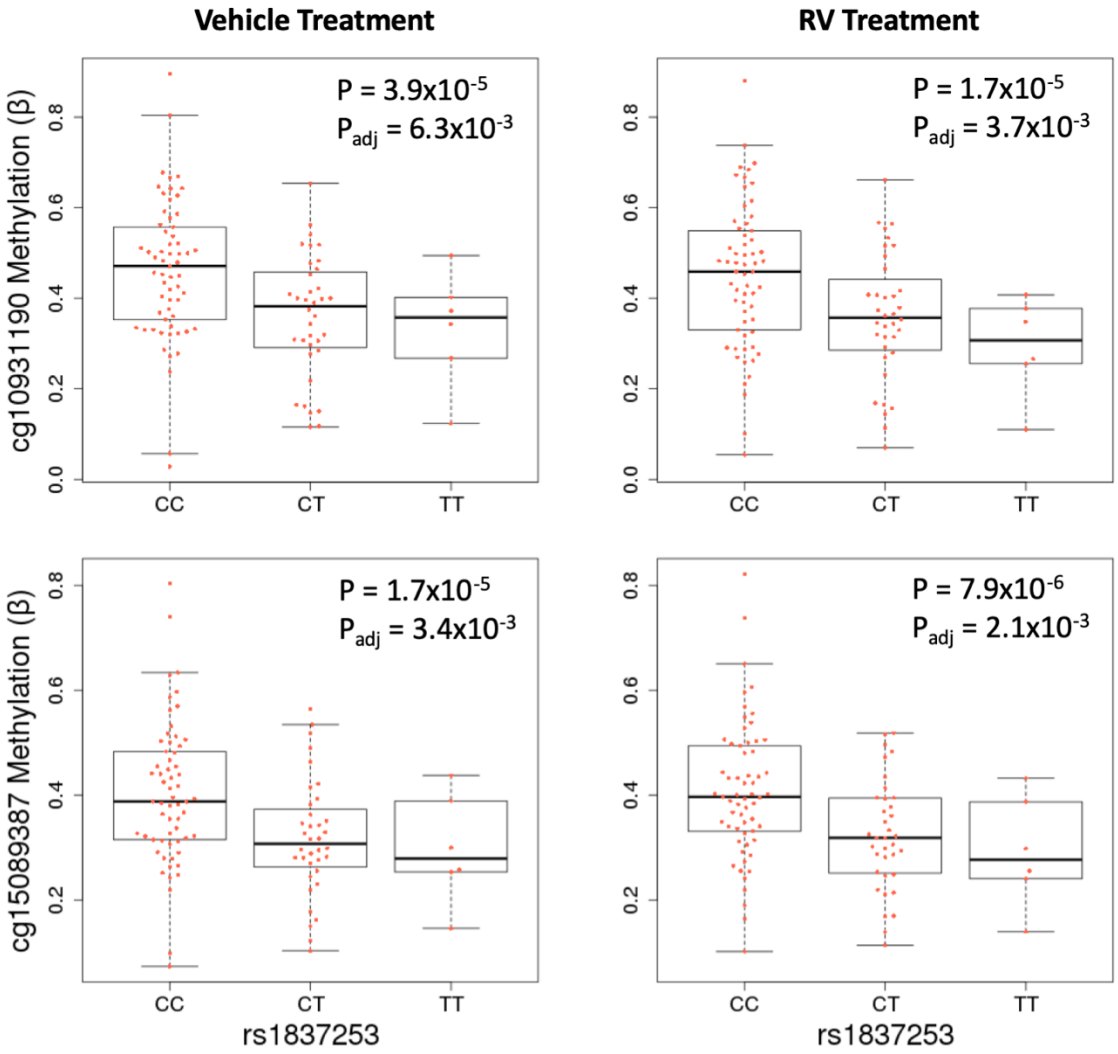


Figure S4.6 meQTLs at rs1837253 located in the first untranslated exon of the *TSLP* gene. Box plots of two meQTLs that were identified in both the vehicle- (left) and RV-treated (right) AECs were co-localized with the asthma risk variant (rs1837253) in adult onset and childhood onset asthma GWASs.

CHAPTER 5

CONCLUSION

As we continue to develop a mechanistic understanding of disease risk loci identified by GWASs, evidence has accumulated showing that GWAS causal variants impact disease risk by altering context-specific (cell type, environment) functional elements and the expression of their target genes as demonstrated in Chapter 4 of this thesis and in published studies [225]. By integrating multi-omics data (gene expression, DNAm) from an *in vitro* epithelial cell model of RV response cell-type (airway epithelial cells) and environmental risk factor (RV infection) with asthma GWAS, we were able to ask questions related to the context-specificity of distinct molecular mechanisms that contribute to asthma risk. Moreover, these combined omics data also enabled the identification of gene and epigenetic pathways associated with CRS, showing that this disease is primarily driven by epigenetic factors.

5.1 Chronic rhinosinusitis is primarily an epigenetic disease

Recently, the largest CRS GWAS to date (5,291 cases, 79,662 controls) only identified a single locus associated with CRS that reached genome-wide significance levels [44]. Given the relatively large sample size of this GWAS compared to previous genetic studies of CRS, two major inferences could explain the paucity of genetic evidence for this disease. First, because CRS is a complex disease, and because the genetics, if any, are likely to have small effect sizes, the variability of classifying CRS sub-phenotypes [43] may contribute to the reduced power in identifying disease variants associated with CRS. Second, several environmental factors are associated with CRS [47, 48, 88, 89]. Exposure to any of these factors may contribute to important epigenetic modifications in the developmental origins and pathogenesis of CRS.

While there is evidence to suggest a possible genetic component in CRS risk [45, 226-229], genetic studies of this disease are widely known to be severely limited by sample size and low statistical power (reviewed in [43]). Moreover, most of these genetic studies lack replication and functional validation. The limited sample size of CRS genetic studies is thought to largely contribute to the lack of robust genetics evidence in CRS.

An important role for epigenetic mechanisms underlying different responses to RV between CRSwNP cases and non-CRS controls was highlighted in Chapter 2. These epigenetic mechanisms influence gene expression pathways that reflect dysregulated epithelial cell processes in CRS and indicate the strong influence that relevant environmental perturbations have in the development of this disease. Additionally, in order to understand the influence of genetics on CRS risk, a CRS GWAS was conducted in Chapter 3. Although this GWAS was underpowered (483 cases, 2,057 controls), 82 suggestive significant SNPs ($P < 1x10^{-5}$) were identified at six loci, two of which were each associated with one of two sub-phenotypes of CRS (localized or diffuse opacification).

Taken together, the results from Chapters 2 and 3, as well as the inconsistent and sparse genetics support for CRS in the literature, strongly suggest that CRS is epigenetic in its origins. The strongest evidence for this comes from the WGCNA analysis (Chapter 2) in which most (95%) of the DMCs identified in airway epithelium from CRS cases and non-CRS controls were in the CRS-correlated modules, whereas the DEGs were more equally distributed between modules correlated with CRS and RV. Moreover, of the three CRS-correlated modules containing correlated DMCs, two of these modules were also correlated with DEGs that were enriched in gene pathways previously implicated in CRS, indicating epigenetic regulatory mechanisms associated with these pathways. Although the results presented in Chapter 2

represent the most comprehensive DNAm analysis of CRS to date, other smaller studies of DNAm in CRS also support a major role of epigenetic variation in CRS pathogenesis [230-232]. Finally, the two-stage GWAS presented in Chapter 3 identified sub-phenotype-specific associations with CRS, but these associations were modest, and we were unable to replicate the finding due to the requirement for CT scans for assigning sub-phenotypes. The sparsity of genetic associations with CRS in Chapter 3 is consistent with other genetics studies, specifically with the largest CRS GWAS [44], in which the evidence for genetics in CRS risk is underwhelming and inconsistent between studies. In sum, these findings point to epigenetic modifications underlying CRS risk and pathogenesis and should be a focus on future therapeutic development for this disease.

5.2 Context is key to interpreting genome-wide association studies

Genome-wide association studies alone are not informative of specific tissues or environmental factors that are involved with or contribute to disease outcomes. Integration of cell type-specific molecular QTLs with GWAS data is not only a powerful way of prioritizing tissues and cell types that contribute to phenotype, but also allows us to identify mechanisms underlying complex diseases. However, currently, these analyses are largely limited to public databases [75-78]. which are based on resting cells as either bulk tissues or cell lines and little to no information on environmental exposures. *In vitro* cell models are a powerful way of overcoming some of these limitations and will help fine-tune GWAS findings into clinically actionable gene sets.

In Chapter 3, we show the advantage of integrating asthma GWASs with molecular QTLs from an *in vitro* cell model of a disease-relevant cell type and environmental exposure. The application of a co-localization approach allowed us to directly test whether the same genetic

variant underlie associations between two or more traits (e.g., gene expression, DNAm, and asthma), providing clues to causal disease pathways, and Mendelian randomization analyses allowed us to test the causal effects of DNAm of gene expression.

We highlighted two important sets of results showing the importance of contextualizing GWAS. First, we show that molecular QTLs in the airway epithelium are enriched for asthma GWAS variants but not for non-allergic disease GWAS variants, highlighting the importance of the airway epithelium in asthma pathogenesis. The specificity of the airway epithelium in childhood onset asthma was further emphasized by showing that eQTLs from tissues that are not known to be involved in asthma were not enriched with childhood asthma GWAS variants. Second, by integrating multi-omics data sets from an *in vitro* cell model of an asthma-relevant cell type and environmental exposure, we were able to identify potential molecular mechanisms underlying asthma risk that are specific to RV infection in airway epithelial cells. It is important to stress the importance of these findings as they likely would have been missed if not discovered in the context of cell-type specificity, environmental exposure, and multi-omics integration.

Most genome-wide epigenetic studies of exposure (e.g. [86, 161-164, 233]) or asthma-related traits (e.g. [165-170]) have not integrated their results with genetic analyses. Only few studies have formally integrated external asthma GWAS results with epigenetic studies in airway tissues (e.g. [171-173]). The work presented in Chapter 3 stresses the importance of contextualizing GWAS results by integrating multi-omics datasets from relevant cell types and environmental exposures. Not only does this work emphasize the importance of context-specific analyses in asthma, but these concepts and analyses can be extended to other complex phenotypes with significant environmental components.

5.3 Future Directions

Using cultured primary epithelial cells to model transcriptional and epigenetic responses to RV infection allowed us to make several observations including epigenetic response differences between CRSwNP cases and non-CRS controls, and context-specific mechanisms in childhood onset asthma. However, the studies presented in this thesis had several limitations. A limitation common to each chapter was the small sample size. This limited our power to detect interaction effects (CRS x RV) in Chapter 2, especially for gene expression, to find additional genetic risk factors for CRS in Chapter 3, and to discovery of additional co-localizations of gene expression, DNAm, and childhood onset asthma in Chapter 4. It is likely, therefore, that many more interactions and co-localizations will be detected in larger samples.

The CRS GWAS presented in Chapter 3 was particularly limited by sample size, which reduced our power to detect SNP associations let alone any functional variants that may also be associated with the disease. Even achieving sample sizes like those in the Kristjansson et al. study [44] may not have aided in identifying any genetic association with CRS considering that they only identified a single locus association. However, sub-phenotyping methods in addition to a larger sample may facilitate the discovery of sub-type-specific variants. Future genetics studies in CRS should move away from combining samples with any CRS diagnosis and focus on careful phenotyping of this disease.

Finally, the studies presented in Chapters 2 and 4 focused on one cell type (upper airway sinonasal epithelium) in isolation from the many cell types that contribute to both CRS and asthma, only included two exposures (vehicle and RV), and one epigenetic mark (DNA methylation). As a result, our model likely only partially captured RV-responses, cases-control differences (CRS vs non-CRS; in Chapter 2), and co-localizations with asthma loci (Chapter 4)

that are present *in vivo*. It is possible that other disease-relevant molecular pathways or co-localizations are specific to other tissues, cell types or to other exposure or culture conditions, and that integrating additional epigenetic marks into these analyses, such as those associated with chromatin accessibility, might be additionally informative. Extended studies addressing these concerns is necessary to confirm the specificity of our findings and provide a more complete picture of disease-relevant gene pathways and co-localizations.

5.4 Concluding Remarks

The work presented in this thesis highlights key advances in our molecular mechanistic understanding of CRS and asthma. Our mechanistic understanding of disease risk loci severely lags behind the discovery of new SNP-disease associations [209]. Likewise, the added dimensions of cell and environment specificity is lacking in disease research and is essential to our understanding if effective drug targets are to be found and therapeutics developed to confront these diseases plaguing millions of people worldwide. The research presented in this study is one step of many needed to achieve these goals by not only developing an understanding of these diseases out of intellectual interest, but to also contribute to advancement of therapies in their treatment.

Bibliography

1. Ferkol T, Schraufnagel D: **The global burden of respiratory disease**. *Ann Am Thorac Soc* 2014, **11**(3):404-406.

2. Organization GWH: **Global surveillance, prevention and control of chronic respiratory diseases: a comprehensive approach.** 2020.

3. Beasley R, Crane J, Lai CK, Pearce N: **Prevalence and etiology of asthma**. *J Allergy Clin Immunol* 2000, **105**(2 Pt 2):S466-472.

4. Collaborators GBDCRD: **Global, regional, and national deaths, prevalence, disability-adjusted life years, and years lived with disability for chronic obstructive pulmonary disease and asthma, 1990-2015: a systematic analysis for the Global Burden of Disease Study 2015**. *Lancet Respir Med* 2017, **5**(9):691-706.

5. **Most Recent National Asthma Data** [https://www.cdc.gov/asthma/most_recent_national_asthma_data.htm]

6. Nurmagambetov T, Kuwahara R, Garbe P: **The Economic Burden of Asthma in the United States, 2008-2013**. *Annals of the American Thoracic Society* 2018, **15**(3):348-356.

7. Rudmik L: **Chronic rhinosinusitis: an under-researched epidemic**. *Journal of Otolaryngology-Head & ...* 2015.

8. Caulley L, Thavorn K, Rudmik L, Cameron C: **Direct costs of adult chronic rhinosinusitis by using 4 methods of estimation: Results of the US Medical Expenditure Panel Survey**. *Journal of Allergy and ...* 2015.

9. Ray NF, Baraniuk JN, Thamer M, Rinehart CS: **Healthcare expenditures for sinusitis in 1996: contributions of asthma, rhinitis, and other airway disorders**. *Journal of Allergy and ...* 1999.

10. Moffatt MF, Gut IG, Demenais F, Strachan DP, Bouzigon E, Heath S, von Mutius E, Farrall M, Lathrop M, Cookson W *et al*: **A large-scale, consortium-based genomewide association study of asthma**. *N Engl J Med* 2010, **363**(13):1211-1221.

11. Demenais F, Margaritte-Jeannin P, Barnes KC, Cookson WOC, Altmuller J, Ang W, Barr RG, Beaty TH, Becker AB, Beilby J *et al*: **Multiancestry association study identifies new asthma risk loci that colocalize with immune-cell enhancer marks**. *Nat Genet* 2018, **50**(1):42-53.

12. Caliskan M, Bochkov YA, Kreiner-Moller E, Bonnelykke K, Stein MM, Du G, Bisgaard H, Jackson DJ, Gern JE, Lemanske RF, Jr. *et al*: **Rhinovirus wheezing illness and genetic risk of childhood-onset asthma**. *N Engl J Med* 2013, **368**(15):1398-1407.

13. Pividori M, Schoettler N, Nicolae DL, Ober C, Im HK: **Shared and distinct genetic risk factors for childhood-onset and adult-onset asthma: genome-wide and transcriptome-wide studies**. *Lancet Respir Med* 2019, **7**(6):509-522.

14. Vicente CT, Revez JA, Ferreira MAR: **Lessons from ten years of genome-wide association studies of asthma**. *Clin Transl Immunology* 2017, **6**(12):e165.

15. Edwards MR, Regamey N, Vareille M, Kieninger E, Gupta A, Shoemark A, Saglani S, Sykes A, Macintyre J, Davies J *et al*: **Impaired innate interferon induction in severe therapy resistant atopic asthmatic children**. *Mucosal Immunol* 2013, **6**(4):797-806.

16. Johnston NW, Johnston SL, Duncan JM, Greene JM, Kebadze T, Keith PK, Roy M, Waserman S, Sears MR: **The September epidemic of asthma exacerbations in children: a search for etiology**. *J Allergy Clin Immunol* 2005, **115**(1):132-138.

17. Orellano P, Quaranta N, Reynoso J, Balbi B, Vasquez J: **Effect of outdoor air pollution on asthma exacerbations in children and adults: Systematic review and multilevel meta-analysis**. *PLoS One* 2017, **12**(3):e0174050.

18. Quinto KB, Kit BK, Lukacs SL, Akinbami LJ: **Environmental tobacco smoke exposure in children aged 3-19 years with and without asthma in the United States, 1999-2010**. *NCHS Data Brief* 2013(126):1-8.

19. Kanchongkittiphon W, Mendell MJ, Gaffin JM, Wang G, Phipatanakul W: **Indoor environmental exposures and exacerbation of asthma: an update to the 2000 review by the Institute of Medicine**. *Environ Health Perspect* 2015, **123**(1):6-20.

20. Cookson WOCM, Faux JA, Sharp PA, Hopkin JM: **Linkage between Immunoglobulin-E Responses Underlying Asthma and Rhinitis and Chromosome-11q**. *Lancet* 1989, **1**(8650):1292-1295.

21. Bouzigon E, Forabosco P, Koppelman GH, Cookson WOCM, Dizier MH, Duffy DL, Evans DM, Ferreira MAR, Kere J, Laitinen T *et al*: **Meta-analysis of 20 genome-wide linkage studies evidenced new regions linked to asthma and atopy**. *Eur J Hum Genet* 2010, **18**(6):700-706.

22. Ohe M, Munakata M, Hizawa N, Itoh A, Doi I, Yamaguchi E, Homma Y, Kawakami Y: **Beta(2) Adrenergic-Receptor Gene Restriction-Fragment-Length-Polymorphism and Bronchial-Asthma**. *Thorax* 1995, **50**(4):353-359.

23. Tizaoui K, Hamzaoui K, Hamzaoui A: **Update on Asthma Genetics: Results From Meta-Analyses of Candidate Gene Association Studies**. *Curr Mol Med* 2017, **17**(10):647-667.

24. Ferreira MAR: **Linkage analysis: Principles and methods for the analysis of human quantitative traits**. *Twin Res* 2004, **7**(5):513-530.

25. Hoffjan S, Nicolae D, Ober C: **Association studies for asthma and atopic diseases: a comprehensive review of the literature**. *Resp Res* 2003, **4**(14).

26. Matsuzaki H, Dong SL, Loi H, Di XJ, Liu GY, Hubbell E, Law J, Berntsen T, Chadha M, Hui H *et al*: **Genotyping over 100,000 SNPs on a pair of oligonucleotide arrays**. *Nature Methods* 2004, **1**(2):109-111.

27. Moffatt MF, Kabesch M, Liang L, Dixon AL, Strachan D, Heath S, Depner M, von Berg A, Bufe A, Rietschel E *et al*: **Genetic variants regulating ORMDL3 expression contribute to the risk of childhood asthma**. *Nature* 2007, **448**(7152):470-473.

28. Stein MM, Thompson EE, Schoettler N, Helling BA, Magnaye KM, Stanhope C, Igartua C, Morin A, Washington C, 3rd, Nicolae D *et al*: **A decade of research on the 17q12-21 asthma locus: Piecing together the puzzle**. *J Allergy Clin Immunol* 2018.

29. Fokkens WJ, Lund VJ, Hopkins C, Hellings PW, Kern R, Reitsma S, Toppila-Salmi S, Bernal-Sprekelsen M, Mullol J, Alobid I *et al*: **European Position Paper on Rhinosinusitis and Nasal Polyps 2020**. *Rhinology* 2020, **58**(Suppl S29):1-464.

30. Kern RC, Conley DB, Walsh W, Chandra R, Kato A, Tripathi-Peters A, Grammer LC, Schleimer RP: **Perspectives on the etiology of chronic rhinosinusitis: an immune barrier hypothesis**. *Am J Rhinol* 2008, **22**(6):549-559.

31. Meltzer EO, Hamilos DL, Hadley JA, Lanza DC, Marple BF, Nicklas RA, Bachert C, Baraniuk J, Baroody FM, Benninger MS *et al*: **Rhinosinusitis: Establishing definitions for clinical research and patient care**. *Otolaryngol Head Neck Surg* 2004, **131**(6 Suppl):S1-62.

32. Chang EH, Willis AL, McCrary HC, Noutsios GT, Le CH, Chiu AG, Mansfield CJ, Reed DR, Brooks SG, Adappa ND *et al*: **Association between the CDHR3 rs6967330 risk allele and chronic rhinosinusitis**. *J Allergy Clin Immunol* 2017, **139**(6):1990-1992 e1992.

33. Deal RT, Kountakis SE: **Significance of nasal polyps in chronic rhinosinusitis: symptoms and surgical outcomes**. *Laryngoscope* 2004, **114**(11):1932-1935.

34. Damm M, Quante G, Jungehuelsing M, Stennert E: **Impact of functional endoscopic sinus surgery on symptoms and quality of life in chronic rhinosinusitis**. *Laryngoscope* 2002, **112**(2):310-315.

35. Hopkins C, Gillett S, Slack R, Lund VJ: **Psychometric validity of the 22-item Sinonasal Outcome Test**. *Clinical ...* 2009.

36. RachelefSky GS, Katz RM, Siegel SC: **Chronic sinus disease with associated reactive airway disease in children**. *Pediatrics* 1984.

37. Jarvis D, Newson R, Lotvall J, Hastan D, Tomassen P: **Asthma in adults and its association with chronic rhinosinusitis: the GA2LEN survey in Europe**. *Allergy* 2012.

38. Ryan MW, Brooks EG: **Rhinosinusitis and comorbidities**. *Current allergy and asthma reports* 2010.

39. Wang XJ, Moylan B, Leopold DA, Kim J: **Mutation in the gene responsible for cystic fibrosis and predisposition to chronic rhinosinusitis in the general population**. *Jama* 2000.

40. Wang XJ, Kim J, McWilliams R: **Increased prevalence of chronic rhinosinusitis in carriers of a cystic fibrosis mutation**. *... –Head & Neck ...* 2005.

41. Pinto JM, Hayes MG, Schneider D: **A genomewide screen for chronic rhinosinusitis genes identifies a locus on chromosome 7q**. *The ...* 2008.

42. Meltzer EO, Hamilos DL: **Rhinosinusitis diagnosis and management for the clinician: a synopsis of recent consensus guidelines**. *Mayo Clin Proc* 2011, **86**(5):427-443.

43. Hsu J, Avila PC, Kern RC, Hayes MG, Schleimer RP, Pinto JM: **Genetics of chronic rhinosinusitis: state of the field and directions forward**. *J Allergy Clin Immunol* 2013, **131**(4):977-993, 993 e971-975.

44. Kristjansson RP, Benonisdottir S, Davidsson OB, Oddsson A, Tragante V, Sigurdsson JK, Stefansdottir L, Jonsson S, Jensson BO, Arthur JG *et al*: **A loss-of-function variant in ALOX15 protects against nasal polyps and chronic rhinosinusitis**. *Nat Genet* 2019, **51**(2):267-276.

45. Oakley GM, Curtin K, Orb Q, Schaefer C, Orlandi RR, Alt JA: **Familial risk of chronic rhinosinusitis with and without nasal polyposis: genetics or environment**. *Int Forum Allergy Rhinol* 2015, **5**(4):276-282.

46. Hastan D, Fokkens WJ, Bachert C, Newson RB, Bislimovska J, Bockelbrink A, Bousquet PJ, Brozek G, Bruno A, Dahlen SE *et al*: **Chronic rhinosinusitis in Europe--an underestimated disease. A GA(2)LEN study**. *Allergy* 2011, **66**(9):1216-1223.

47. Thilsing T, Rasmussen J, Lange B, Kjeldsen AD, Al-Kalemji A, Baelum J: **Chronic rhinosinusitis and occupational risk factors among 20- to 75-year-old Danes-A GA(2) LEN-based study**. *Am J Ind Med* 2012, **55**(11):1037-1043.

48. Wolf C: **Urban air pollution and health: an ecological study of chronic rhinosinusitis in Cologne, Germany**. *Health Place* 2002, **8**(2):129-139.

49. Bhattacharyya N: **Air quality influences the prevalence of hay fever and sinusitis**. *Laryngoscope* 2009, **119**(3):429-433.

50. Bose S, Grammer LC, Peters AT: **Infectious Chronic Rhinosinusitis**. *J Allergy Clin Immunol Pract* 2016, **4**(4):584-589.

51. Johnston SL, Pattemore PK, Sanderson G, Smith S, Lampe F, Josephs L, Symington P, O'Toole S, Myint SH, Tyrrell DA *et al*: **Community study of role of viral infections in exacerbations of asthma in 9-11 year old children**. *Bmj* 1995, **310**(6989):1225-1229.

52. Sykes A, Johnston SL: **Etiology of asthma exacerbations**. *J Allergy Clin Immunol* 2008, **122**(4):685-688.

53. Rowan NR, Lee S, Sahu N, Kanaan A, Cox S, Phillips CD, Wang EW: **The role of viruses in the clinical presentation of chronic rhinosinusitis**. *Am J Rhinol Allergy* 2015, **29**(6):e197-200.

54. Cho GS, Moon BJ, Lee BJ, Gong CH, Kim NH, Kim YS, Kim HS, Jang YJ: **High rates of detection of respiratory viruses in the nasal washes and mucosae of patients with chronic rhinosinusitis**. *J Clin Microbiol* 2013, **51**(3):979-984.

55. McLean GR: **Developing a vaccine for human rhinoviruses**. *J Vaccines Immun* 2014, **2**(3):16-20.

56. Stepanova E, Isakova-Sivak I, Rudenko L: **Overview of human rhinovirus immunogenic epitopes for rational vaccine design**. *Expert Rev Vaccines* 2019, **18**(9):877-880.

57. Makris S, Johnston S: **Recent advances in understanding rhinovirus immunity**. *F1000Res* 2018, **7**.

58. Halperin SA, Eggleston PA, Hendley JO, Suratt PM, Groschel DH, Gwaltney JM, Jr.: **Pathogenesis of lower respiratory tract symptoms in experimental rhinovirus infection**. *Am Rev Respir Dis* 1983, **128**(5):806-810.

59. Gern JE, Galagan DM, Jarjour NN, Dick EC, Busse WW: **Detection of rhinovirus RNA in lower airway cells during experimentally induced infection**. *Am J Respir Crit Care Med* 1997, **155**(3):1159-1161.

60. Seymour ML, Gilby N, Bardin PG, Fraenkel DJ, Sanderson G, Penrose JF, Holgate ST, Johnston SL, Sampson AP: **Rhinovirus infection increases 5-lipoxygenase and**

cyclooxygenase-2 in bronchial biopsy specimens from nonatopic subjects. *J Infect Dis* 2002, **185**(4):540-544.

61. Aponte FE, Taboada B, Espinoza MA, Arias-Ortiz MA, Monge-Martinez J, Rodriguez-Vazquez R, Diaz-Hernandez F, Zarate-Vidal F, Wong-Chew RM, Firo-Reyes V *et al*: **Rhinovirus is an important pathogen in upper and lower respiratory tract infections in Mexican children**. *Virol J* 2015, **12**:31.

62. Heymann PW, Carper HT, Murphy DD, Platts-Mills TA, Patrie J, McLaughlin AP, Erwin EA, Shaker MS, Hellems M, Peerzada J *et al*: **Viral infections in relation to age, atopy, and season of admission among children hospitalized for wheezing**. *J Allergy Clin Immunol* 2004, **114**(2):239-247.

63. Soto-Quiros M, Avila L, Platts-Mills TA, Hunt JF, Erdman DD, Carper H, Murphy DD, Odio S, James HR, Patrie JT *et al*: **High titers of IgE antibody to dust mite allergen and risk for wheezing among asthmatic children infected with rhinovirus**. *J Allergy Clin Immunol* 2012, **129**(6):1499-1505 e1495.

64. Ko FW, Chan PK, Chan RWY, Chan KP, Ip A, Kwok A, Ngai JC, Ng SS, On CT, Hui DS: **Molecular detection of respiratory pathogens and typing of human rhinovirus of adults hospitalized for exacerbation of asthma and chronic obstructive pulmonary disease**. *Respir Res* 2019, **20**(1):210.

65. Jackson DJ, Gern JE, Lemanske RF: **Lessons learned from birth cohort studies conducted in diverse environments**. *J Allergy Clin Immun* 2017, **139**(2):379-387.

66. Johnston SL, Pattemore PK, Sanderson G, Smith S, Campbell MJ, Josephs LK, Cunningham A, Robinson BS, Myint SH, Ward ME *et al*: **The relationship between upper respiratory infections and hospital admissions for asthma: A time-trend analysis**. *American Journal of Respiratory and Critical Care Medicine* 1996, **154**(3):654-660.

67. Loss GJ, Depner M, Hose AJ, Genuneit J, Karvonen AM, Hyvarinen A, Roduit C, Kabesch M, Lauener R, Pfefferle PI *et al*: **The Early Development of Wheeze. Environmental Determinants and Genetic Susceptibility at 17q21**. *Am J Respir Crit Care Med* 2016, **193**(8):889-897.

68. Basharat U, Aiche MM, Kim MM, Sohal M, Chang EH: **Are rhinoviruses implicated in the pathogenesis of sinusitis and chronic rhinosinusitis exacerbations? A comprehensive review**. *Int Forum Allergy Rhinol* 2019, **9**(10):1159-1188.

69. Rank MA, Wollan P, Kita H, Yawn BP: **Acute exacerbations of chronic rhinosinusitis occur in a distinct seasonal pattern**. *J Allergy Clin Immunol* 2010, **126**(1):168-169.

70. Wald ER, Milmoe GJ, Bowen AD, Ledesmamedina J, Salamon N, Bluestone CD: **Acute Maxillary Sinusitis in Children**. *New Engl J Med* 1981, **304**(13):749-754.

71. Slavin RG, Spector SL, Bernstein IL, Kaliner MA, Kennedy DW, Virant FS, Wald ER, Khan DA, Blessing-Moore J, Lang DM *et al*: **The diagnosis and management of sinusitis: a practice parameter update**. *J Allergy Clin Immunol* 2005, **116**(6 Suppl):S13-47.

72. Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, Reynolds AP, Sandstrom R, Qu H, Brody J: **Systematic localization of common disease-associated variation in regulatory DNA**. *Science* 2012:1222794.

73. Calderon D, Nguyen MLT, Mezger A, Kathiria A, Muller F, Nguyen V, Lescano N, Wu B, Trombetta J, Ribado JV *et al*: **Landscape of stimulation-responsive chromatin across diverse human immune cells**. *Nat Genet* 2019, **51**(10):1494-1505.

74. Nicolae DL, Gamazon E, Zhang W, Duan S, Dolan ME, Cox NJ: **Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS**. *PLoS Genet* 2010, **6**(4):e1000888.

75. Consortium GT: **The GTEx Consortium atlas of genetic regulatory effects across human tissues**. *Science* 2020, **369**(6509):1318-1330.

76. Consortium EP: **An integrated encyclopedia of DNA elements in the human genome**. *Nature* 2012, **489**(7414):57-74.

77. Gerasimova A, Chavez L, Li B, Seumois G, Greenbaum J, Rao A, Vijayanand P, Peters B: **Predicting cell types and genetic variations contributing to disease by combining GWAS and epigenetic data**. *PLoS One* 2013, **8**(1):e54359.

78. Roadmap Epigenomics C, Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, Heravi-Moussavi A, Kheradpour P, Zhang Z, Wang J *et al*: **Integrative analysis of 111 reference human epigenomes**. *Nature* 2015, **518**(7539):317-330.

79. Ferreira MA, Vonk JM, Baurecht H, Marenholz I, Tian C, Hoffman JD, Helmer Q, Tillander A, Ullemar V, van Dongen J *et al*: **Shared genetic origin of asthma, hay fever and eczema elucidates allergic disease biology**. *Nat Genet* 2017, **49**(12):1752-1757.

80. Olafsdottir TA, Theodors F, Bjarnadottir K, Bjornsdottir US, Agustsdottir AB, Stefansson OA, Ivarsdottir EV, Sigurdsson JK, Benonisdottir S, Eyjolfsson GI *et al*: **Eighty-eight variants highlight the role of T cell regulation and airway remodeling in asthma pathogenesis**. *Nat Commun* 2020, **11**(1):393.

81. Hao K, Bosse Y, Nickle DC, Pare PD, Postma DS, Laviolette M, Sandford A, Hackett TL, Daley D, Hogg JC *et al*: **Lung eQTLs to help reveal the molecular underpinnings of asthma**. *PLoS Genet* 2012, **8**(11):e1003029.

82. Zuyderduyn S, Sukkar MB, Fust A, Dhaliwal S, Burgess JK: **Treating asthma means treating airway smooth muscle cells**. *Eur Respir J* 2008, **32**(2):265-274.

83. Wang Y, Bai C, Li K, Adler KB, Wang X: **Role of airway epithelial cells in development of asthma and allergic rhinitis**. *Respir Med* 2008, **102**(7):949-955.

84. Heijink IH, van Oosterhout A, Kapus A: **Epidermal growth factor receptor signalling contributes to house dust mite-induced epithelial barrier dysfunction**. *Eur Respir J* 2010, **36**(5):1016-1026.

85. Wark PA, Johnston SL, Bucchieri F, Powell R, Puddicombe S, Laza-Stanca V, Holgate ST, Davies DE: **Asthmatic bronchial epithelial cells have a deficient innate immune response to infection with rhinovirus**. *J Exp Med* 2005, **201**(6):937-947.

86. Nicodemus-Johnson J, Naughton KA, Sudi J, Hogarth K, Naurekas ET, Nicolae DL, Sperling AI, Solway J, White SR, Ober C: **Genome-wide methylation study identifies an IL-13-induced epigenetic signature in asthmatic airways**. *Am J Respir Crit Care Med* 2016, **193**(4):376-385.

87. Helling BA, Sobreira DR, Hansen GT, Sakabe NJ, Luo K, Billstrand C, Laxman B, Nicolae RI, Nicolae DL, Bochkov YA *et al*: **Altered transcriptional and chromatin responses to rhinovirus in bronchial epithelial cells from adults with asthma**. *Communications Biology* 2020, **3**(1):678.

88. Jarvis D, Newson R, Lotvall J, Hastan D, Tomassen P, Keil T, Gjomarkaj M, Forsberg B, Gunnbjornsdottir M, Minov J *et al*: **Asthma in adults and its association with chronic rhinosinusitis: the GA2LEN survey in Europe**. *Allergy* 2012, **67**(1):91-98.

89. Tan BK, Chandra RK, Pollak J, Kato A, Conley DB, Peters AT, Grammer LC, Avila PC, Kern RC, Stewart WF *et al*: **Incidence and associated premorbid diagnoses of patients with chronic rhinosinusitis**. *J Allergy Clin Immunol* 2013, **131**(5):1350-1360.

90. Philpott CM, McKiernan DC: **Bronchiectasis and sino-nasal disease: a review**. *J Laryngol Otol* 2008, **122**(1):11-15.

91. Gao WX, Ou CQ, Fang SB, Sun YQ, Zhang H, Cheng L, Wang YJ, Zhu DD, Lv W, Liu SX *et al*: **Occupational and environmental risk factors for chronic rhinosinusitis in China: a multicentre cross-sectional study**. *Respir Res* 2016, **17**(1):54.

92. Jang YJ, Kwon HJ, Park HW, Lee BJ: **Detection of rhinovirus in turbinate epithelial cells of chronic sinusitis**. *Am J Rhinol* 2006, **20**(6):634-636.

93. Van Zele T, Gevaert P, Watelet JB, Claeys G, Holtappels G, Claeys C, van Cauwenberge P, Bachert C: **Staphylococcus aureus colonization and IgE antibody formation to enterotoxins is increased in nasal polyposis**. *J Allergy Clin Immunol* 2004, **114**(4):981-983.

94. Mitts KB, Chang EH: **Genetics of chronic rhinosinusitis**. *J Allergy Clin Immunol* 2020, **145**(3):777-779.

95. Soliai MS, Agnes S.; Morin, Andréanne; Hirsch, Annemarie G.; Stanhope, Catherine; Kuiper, Jordan, Schwartz, Brian S.; Ober, Carole; Pinto, Jayant M.: **Two-stage genome-wide association study of chronic rhinosinusitis and disease sub-phenotypes highlights mucosal immunity contributing to risk**. *International Forum of Allergy & Rhinology* 2020. In Press.

96. Van Crombruggen K, Zhang N, Gevaert P, Tomassen P, Bachert C: **Pathogenesis of chronic rhinosinusitis: inflammation**. *J Allergy Clin Immunol* 2011, **128**(4):728-732.

97. Gwaltney JM, Jr.: **Acute community-acquired sinusitis**. *Clin Infect Dis* 1996, **23**(6):1209-1223; quiz 1224-1205.

98. Osguthorpe JD, Hadley JA: **Rhinosinusitis. Current concepts in evaluation and management**. *Med Clin North Am* 1999, **83**(1):27-41, vii-viii.

99. Lee SB, Yi JS, Lee BJ, Gong CH, Kim NH, Joo CH, Jang YJ: **Human rhinovirus serotypes in the nasal washes and mucosa of patients with chronic rhinosinusitis**. *Int Forum Allergy Rhinol* 2015, **5**(3):197-203.

100. Han JK: **Subclassification of chronic rhinosinusitis**. *Laryngoscope* 2013, **123 Suppl 2**:S15-27.

101. Hirsch AG, Stewart WF, Sundaresan AS, Young AJ, Kennedy TL, Scott Greene J, Feng W, Tan BK, Schleimer RP, Kern RC *et al*: **Nasal and sinus symptoms and chronic rhinosinusitis in a population-based sample**. *Allergy* 2017, **72**(2):274-281.

102. Soliai MM, Kato A, Stanhope CT, Norton JE, Naughton KA, Klinger AI, Kern RC, Tan BK, Schleimer RP, Nicolae DL *et al*: **Multi-omics co-localization with genome-wide association studies reveals context-specific mechanisms of asthma risk variants**. *bioRxiv* 2019.

103. Tandon A, Patterson N, Reich D: **Ancestry informative marker panels for African Americans based on subsets of commercially available SNP arrays**. *Genet Epidemiol* 2011, **35**(1):80-83.

104. Jun G, Flickinger M, Hetrick KN, Romm JM, Doheny KF, Abecasis GR, Boehnke M, Kang HM: **Detecting and estimating contamination of human DNA samples in sequencing and array-based genotype data**. *Am J Hum Genet* 2012, **91**(5):839-848.

105. Dobin A, Gingeras TR: **Mapping RNA-seq Reads with STAR**. *Curr Protoc Bioinformatics* 2015, **51**:11 14 11-19.

106. Law CW, Chen Y, Genome … SW: **Voom: precision weights unlock linear model analysis tools for RNA-seq read counts**. *Voom: precision weights unlock linear model analysis tools for RNA-seq read counts* 2014.

107. Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD: **The sva package for removing batch effects and other unwanted variation in high-throughput experiments**. *Bioinformatics* 2012, **28**(6):882-883.

108. Maksimovic J, Gordon L: **SWAN: Subset-quantile within array normalization for illumina infinium HumanMethylation450 BeadChips**. *SWAN: Subset-quantile within array normalization for illumina infinium HumanMethylation450 BeadChips* 2012.

109. Peters TJ, Buckley MJ: **De novo identification of differentially methylated regions in the human genome**. *De novo identification of differentially methylated regions in the human genome* 2015.

110. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK: **limma powers differential expression analyses for RNA-sequencing and microarray studies**. *Nucleic Acids Res* 2015, **43**(7):e47.

111. Langfelder P, Horvath S: **WGCNA: an R package for weighted correlation network analysis**. *BMC Bioinformatics* 2008, **9**:559.

112. Zhang B, Horvath S: **A general framework for weighted gene co-expression network analysis**. *Stat Appl Genet Mol Biol* 2005, **4**:Article17.

113. Kuleshov MV, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, Koplev S, Jenkins SL, Jagodnik KM, Lachmann A *et al*: **Enrichr: a comprehensive gene set enrichment analysis web server 2016 update**. *Nucleic Acids Res* 2016, **44**(W1):W90-97.

114. Huang R, Grishagin I, Wang Y, Zhao T, Greene J, Obenauer JC, Ngan D, Nguyen DT, Guha R, Jadhav A *et al*: **The NCATS BioPlanet - An Integrated Platform for Exploring the Universe of Cellular Signaling Pathways for Toxicology, Systems Biology, and Chemical Genomics**. *Front Pharmacol* 2019, **10**:445.

115. Lennard CM, Mann EA, Sun LL, Chang AS, Bolger WE: **Interleukin-1 beta, interleukin-5, interleukin-6, interleukin-8, and tumor necrosis factor-alpha in chronic sinusitis: response to systemic corticosteroids**. *Am J Rhinol* 2000, **14**(6):367-373.

116. Van Zele T, Claeys S, Gevaert P, Van Maele G, Holtappels G, Van Cauwenberge P, Bachert C: **Differentiation of chronic sinus diseases by measurement of inflammatory mediators**. *Allergy* 2006, **61**(11):1280-1289.

117. Van Bruaene N, C PN, Van Crombruggen K, De Ruyck N, Holtappels G, Van Cauwenberge P, Gevaert P, Bachert C: **Inflammation and remodelling patterns in early stage chronic rhinosinusitis**. *Clin Exp Allergy* 2012, **42**(6):883-890.

118. Plewka D, Grzanka A, Drzewiecka E, Plewka A, Misiolek M, Lisowska G, Rostkowska-Nadolska B, Gawlik R: **Differential expression of tumor necrosis factor alpha, interleukin 1beta, nuclear factor kappaB in nasal mucosa among chronic rhinosinusitis patients with and without polyps**. *Postepy Dermatol Alergol* 2017, **34**(3):199-206.

119. Consortium EP: **The ENCODE (ENCyclopedia Of DNA Elements) Project**. *Science* 2004, **306**(5696):636-640.

120. Slenter DN, Kutmon M, Hanspers K, Riutta A, Windsor J, Nunes N, Melius J, Cirillo E, Coort SL, Digles D *et al*: **WikiPathways: a multifaceted pathway database bridging metabolomics to other omics research**. *Nucleic Acids Res* 2018, **46**(D1):D661-D667.

121. McCrae C, Dzgoev A, Stahlman M, Horndahl J, Svard R, Grosse A, Grosskopf T, Skujat MA, Williams N, Schubert S *et al*: **Lanosterol Synthase Regulates Human Rhinovirus Replication in Human Bronchial Epithelial Cells**. *Am J Respir Cell Mol Biol* 2018, **59**(6):713-722.

122. Heaton NS, Randall G: **Multifaceted roles for lipids in viral infection**. *Trends Microbiol* 2011, **19**(7):368-375.

123. Wu NH, Yang W, Beineke A, Dijkman R, Matrosovich M, Baumgartner W, Thiel V, Valentin-Weigand P, Meng F, Herrler G: **The differentiated airway epithelium infected by influenza viruses maintains the barrier function despite a dramatic loss of ciliated cells**. *Sci Rep* 2016, **6**:39668.

124. Gelardi M, Ciprandi G: **Ciliocytophthoria of nasal epithelial cells after viral infection: a sign of suffering cell**. *Acta Biomed* 2019, **90**(2-S).

125. Bequignon E, Mangin D, Becaud J, Pasquier J, Angely C, Bottier M, Escudier E, Isabey D, Filoche M, Louis B *et al*: **Pathogenesis of chronic rhinosinusitis with nasal polyps: role of IL-6 in airway epithelial cell dysfunction**. *J Transl Med* 2020, **18**(1):136.

126. Li Y, Li L, Wang T, Zang H, An Y, Li L, Zhang J, Wang F, Zheng Y: **Analysis of epidermal growth factor signaling in nasal mucosa epithelial cell proliferation involved in chronic rhinosinusitis**. *Chin Med J (Engl)* 2014, **127**(19):3449-3453.

127. Ding GQ, Zheng CQ, Bagga SS: **Up-regulation of the mucosal epidermal growth factor receptor gene in chronic rhinosinusitis and nasal polyposis**. *Arch Otolaryngol Head Neck Surg* 2007, **133**(11):1097-1103.

128. Homma T, Kato A, Sakashita M, Takabayashi T, Norton JE, Suh LA, Carter RG, Harris KE, Peters AT, Grammer LC *et al*: **Potential Involvement of the Epidermal Growth Factor Receptor Ligand Epiregulin and Matrix Metalloproteinase-1 in Pathogenesis of Chronic Rhinosinusitis**. *Am J Respir Cell Mol Biol* 2017, **57**(3):334-345.

129. Jiao J, Zhang T, Zhang Y, Li J, Wang M, Wang M, Li Y, Wang X, Zhang L: **Epidermal growth factor upregulates expression of MUC5AC via TMEM16A, in chronic rhinosinusitis with nasal polyps**. *Allergy Asthma Clin Immunol* 2020, **16**:40.

130. Biswas K, Wagner Mackenzie B, Waldvogel-Thurlow S, Middleditch M, Jullig M, Zoing M, Taylor MW, Douglas RG: **Differentially Regulated Host Proteins Associated with Chronic Rhinosinusitis Are Correlated with the Sinonasal Microbiome**. *Front Cell Infect Microbiol* 2017, **7**:504.

131. Yoon YH, Yeon SH, Choi MR, Jang YS, Kim JA, Oh HW, Jun X, Park SK, Heo JY, Rha KS *et al*: **Altered Mitochondrial Functions and Morphologies in Epithelial Cells Are Associated With Pathogenesis of Chronic Rhinosinusitis With Nasal Polyps**. *Allergy Asthma Immunol Res* 2020, **12**(4):653-668.

132. Song SY, Woo HJ, Bae CH, Kim YW, Kim YD: **Expression of leptin receptor in nasal polyps: leptin as a mucosecretagogue**. *Laryngoscope* 2010, **120**(5):1046-1050.

133. Kato A: **Immunopathology of chronic rhinosinusitis**. *Allergol Int* 2015, **64**(2):121-130.

134. Yin Y, Morgunova E, Jolma A, Kaasinen E, Sahu B, Khund-Sayeed S, Das PK, Kivioja T, Dave K, Zhong F *et al*: **Impact of cytosine methylation on DNA binding specificities of human transcription factors**. *Science* 2017, **356**(6337).

135. Khomich OA, Kochetkov SN, Bartosch B, Ivanov AV: **Redox Biology of Respiratory Viral Infections**. *Viruses* 2018, **10**(8).

136. Kupper DS, Valera FC, Malinsky R, Milanezi CM, Silva JS, Tamashiro E, Anselmo-Lima WT: **Expression of apoptosis mediators p53 and caspase 3, 7, and 9 in chronic rhinosinusitis with nasal polyposis**. *Am J Rhinol Allergy* 2014, **28**(3):187-191.

137. Rudmik L: **Chronic rhinosinusitis: an under-researched epidemic**. *J Otolaryngol Head Neck Surg* 2015, **44**:11.

138. Rudmik L: **Economics of Chronic Rhinosinusitis**. *Curr Allergy Asthma Rep* 2017, **17**(4):20.

139. Laidlaw TM: **Clinical updates in aspirin-exacerbated respiratory disease**. *Allergy Asthma Proc* 2019, **40**(1):4-6.

140. Tian C, Hromatka BS, Kiefer AK, Eriksson N, Noble SM, Tung JY, Hinds DA: **Genome-wide association and HLA region fine-mapping studies identify susceptibility loci for multiple common infections**. *Nat Commun* 2017, **8**(1):599.

141. Fokkens WJ, Lund VJ, Mullol J, Bachert C, Alobid I, Baroody F, Cohen N, Cervin A, Douglas R, Gevaert P *et al*: **EPOS 2012: European position paper on rhinosinusitis and nasal polyps 2012. A summary for otorhinolaryngologists**. *Rhinology* 2012, **50**(1):1-12.

142. Carey DJ, Fetterolf SN, Davis FD, Faucett WA, Kirchner HL, Mirshahi U, Murray MF, Smelser DT, Gerhard GS, Ledbetter DH: **The Geisinger MyCode community health initiative: an electronic health record-linked biobank for precision medicine research**. *Genet Med* 2016, **18**(9):906-913.

143. Hirsch AG, Nordberg C, Bandeen-Roche K, Tan BK, Schleimer RP, Kern RC, Sundaresan A, Pinto JM, Kennedy TL, Greene JS *et al*: **Radiologic sinus inflammation and symptoms of chronic rhinosinusitis in a population-based sample**. *Allergy* 2019.

144. Sundaresan AS, Hirsch AG, Young AJ, Pollak J, Tan BK, Schleimer RP, Kern RC, Kennedy TL, Greene JS, Stewart WF *et al*: **Longitudinal Evaluation of Chronic Rhinosinusitis Symptoms in a Population-Based Sample**. *J Allergy Clin Immunol Pract* 2018, **6**(4):1327-1335 e1323.

145. Delaneau O, Zagury JF, Marchini J: **Improved whole-chromosome phasing for disease and population genetic studies**. *Nat Methods* 2013, **10**(1):5-6.

146. Howie B, Marchini J, Stephens M: **Genotype imputation with thousands of genomes**. *G3 (Bethesda)* 2011, **1**(6):457-470.

147. Hartigan JAaW, M. A.: **Algorithm AS 136: A K-means clustering algorithm**. *Applied Statistics* 1979(28):100-108.

148. Freeman C, Marchini J: **GTOOL: A program for transforming sets of genotype data for use with the programs SNPTEST and IMPUTE, Oxford, UK**. *GTOOL: A program for transforming sets of genotype data for use with the programs SNPTEST and IMPUTE, Oxford, UK*.

149. Zhou X, Stephens M: **Genome-wide efficient mixed-model analysis for association studies**. *Nat Genet* 2012, **44**(7):821-824.

150. Zhou X, Stephens M: **Efficient multivariate linear mixed model algorithms for genome-wide association studies**. *Nat Methods* 2014, **11**(4):407-409.

151. Muthén LK, Muthén BO: **Mplus User's Guide: Statisticcal Analysis with Latent Variables**, 7th edn. Los Angeles, CA: Muthén & Muthén; 1998-2012.

152. Meng Y, Lou H, Wang C, Zhang L: **Predictive significance of computed tomography in eosinophilic chronic rhinosinusitis with nasal polyps**. *Int Forum Allergy Rhinol* 2016, **6**(8):812-819.

153. Consortium GT: **Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans**. *Science* 2015, **348**(6235):648-660.

154. Caruso R, Fina D, Peluso I, Stolfi C, Fantini MC, Gioia V, Caprioli F, Del Vecchio Blanco G, Paoluzi OA, Macdonald TT *et al*: **A functional role for interleukin-21 in promoting the synthesis of the T-cell chemoattractant, MIP-3alpha, by gut epithelial cells**. *Gastroenterology* 2007, **132**(1):166-175.

155. Caruso R, Fina D, Peluso I, Fantini MC, Tosti C, Del Vecchio Blanco G, Paoluzi OA, Caprioli F, Andrei F, Stolfi C *et al*: **IL-21 is highly produced in Helicobacter pylori-infected gastric mucosa and promotes gelatinases synthesis**. *J Immunol* 2007, **178**(9):5957-5965.

156. Book DT, Smith TL, McNamar JP, Saeian K, Binion DG, Toohill RJ: **Chronic sinonasal disease in patients with inflammatory bowel disease**. *Am J Rhinol* 2003, **17**(2):87-90.

157. Leason SR, Barham HP, Oakley G, Rimmer J, DelGaudio JM, Christensen JM, Sacks R, Harvey RJ: **Association of gastro-oesophageal reflux and chronic rhinosinusitis: systematic review and meta-analysis**. *Rhinology* 2017, **55**(1):3-16.

158. Jin H, Oyoshi MK, Le Y, Bianchi T, Koduru S, Mathias CB, Kumar L, Le Bras S, Young D, Collins M *et al*: **IL-21R is essential for epicutaneous sensitization and allergic skin inflammation in humans and mice**. *J Clin Invest* 2009, **119**(1):47-60.

159. Xiao L, Wei Y, Zhang YN, Luo X, Yang BY, Yu SF, Wu XM, Wu CY, Li HB: **Increased IL-21 expression in chronic rhinosinusitis with nasalpolyps**. *Clin Exp Allergy* 2015, **45**(2):404-413.

160. Ferreira MAR, Mathur R, Vonk JM, Szwajda A, Brumpton B, Granell R, Brew BK, Ullemar V, Lu Y, Jiang Y *et al*: **Genetic Architectures of Childhood- and Adult-Onset Asthma Are Partly Distinct**. *Am J Hum Genet* 2019, **104**(4):665-684.

161. Morin A, McKennan CG, Pedersen CT, Stokholm J, Chawes BL, Malby Schoos AM, Naughton KA, Thorsen J, Mortensen MS, Vercelli D *et al*: **Epigenetic landscape links upper airway microbiota in infancy with allergic rhinitis at 6 years of age**. *J Allergy Clin Immunol* 2020.

162. Barcelona V, Huang Y, Brown K, Liu J, Zhao W, Yu M, Kardia SLR, Smith JA, Taylor JY, Sun YV: **Novel DNA methylation sites associated with cigarette smoking among African Americans**. *Epigenetics* 2019, **14**(4):383-391.

163. Koo HK, Morrow J, Kachroo P, Tantisira K, Weiss ST, Hersh CP, Silverman EK, DeMeo DL: **Sex-specific associations with DNA methylation in lung tissue demonstrate smoking interactions**. *Epigenetics* 2020:1-12.

164. Provencal N, Arloth J, Cattaneo A, Anacker C, Cattane N, Wiechmann T, Roh S, Kodel M, Klengel T, Czamara D *et al*: **Glucocorticoid exposure during hippocampal neurogenesis primes future stress response by inducing changes in DNA methylation**. *Proc Natl Acad Sci U S A* 2020, **117**(38):23280-23285.

165. Kim S, Forno E, Zhang R, Park HJ, Xu Z, Yan Q, Boutaoui N, Acosta-Perez E, Canino G, Chen W *et al*: **Expression Quantitative Trait Methylation Analysis Reveals Methylomic Associations With Gene Expression in Childhood Asthma**. *Chest* 2020.

166. Cardenas A, Sordillo JE, Rifas-Shiman SL, Chung W, Liang L, Coull BA, Hivert MF, Lai PS, Forno E, Celedon JC *et al*: **The nasal methylome as a biomarker of asthma and airway inflammation in children**. *Nat Commun* 2019, **10**(1):3095.

167. Liang L, Willis-Owen SAG, Laprise C, Wong KCC, Davies GA, Hudson TJ, Binia A, Hopkin JM, Yang IV, Grundberg E *et al*: **An epigenome-wide association study of total serum immunoglobulin E concentration**. *Nature* 2015, **520**(7549):670-674.

168. Qi C, Jiang Y, Yang IV, Forno E, Wang T, Vonk JM, Gehring U, Smit HA, Milanzi EB, Carpaij OA *et al*: **Nasal DNA methylation profiling of asthma and rhinitis**. *J Allergy Clin Immunol* 2020, **145**(6):1655-1663.

169. Yang IV, Pedersen BS, Liu AH, O'Connor GT, Pillai D, Kattan M, Misiak RT, Gruchalla R, Szefler SJ, Khurana Hershey GK *et al*: **The nasal methylome and childhood atopic asthma**. *J Allergy Clin Immunol* 2017, **139**(5):1478-1488.

170. Forno E, Wang T, Qi C, Yan Q, Xu CJ, Boutaoui N, Han YY, Weeks DE, Jiang Y, Rosser F *et al*: **DNA methylation in nasal epithelium, atopy, and atopic asthma in children: a genome-wide study**. *Lancet Respir Med* 2019, **7**(4):336-346.

171. Nicodemus-Johnson J, Myers RA, Sakabe NJ, Sobreira DR, Hogarth DK, Naureckas ET, Sperling AI, Solway J, White SR, Nobrega MA *et al*: **DNA methylation in lung cells is associated with asthma endotypes and genetic risk**. *JCI Insight* 2016, **1**(20):e90151.

172. Thompson EE, Dang Q, Mitchell-Handley B, Rajendran K, Ram-Mohan S, Solway J, Ober C, Krishnan R: **Cytokine-induced molecular responses in airway smooth muscle cells inform genome-wide association studies of asthma**. *Genome Med* 2020, **12**(1):64.

173. Yan Q, Forno E, Herrera-Luis E, Pino-Yanes M, Qi C, Rios R, Han YY, Kim S, Oh S, Acosta-Perez E *et al*: **A genome-wide association study of severe asthma exacerbations in Latino children and adolescents**. *Eur Respir J* 2020.

174. Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, Reynolds AP, Sandstrom R, Qu H, Brody J *et al*: **Systematic localization of common disease-associated variation in regulatory DNA**. *Science* 2012, **337**(6099):1190-1195.

175. Barreiro LB, Tailleux L, Pai AA, Gicquel B, Marioni JC, Gilad Y: **Deciphering the genetic architecture of variation in the immune response to Mycobacterium tuberculosis infection**. *Proc Natl Acad Sci U S A* 2012, **109**(4):1204-1209.

176. Kim-Hellmuth S, Bechheim M, Putz B, Mohammadi P, Nedelec Y, Giangreco N, Becker J, Kaiser V, Fricker N, Beier E *et al*: **Genetic regulatory effects modified by immune activation contribute to autoimmune disease associations**. *Nat Commun* 2017, **8**(1):266.

177. Li X, Hastie AT, Hawkins GA, Moore WC, Ampleford EJ, Milosevic J, Li H, Busse WW, Erzurum SC, Kaminski N *et al*: **eQTL of bronchial epithelial cells and bronchial alveolar lavage deciphers GWAS-identified asthma genes**. *Allergy* 2015, **70**(10):1309-1318.

178. Luo W, Obeidat M, Di Narzo AF, Chen R, Sin DD, Pare PD, Hao K: **Airway Epithelial Expression Quantitative Trait Loci Reveal Genes Underlying Asthma and Other Airway Diseases**. *Am J Respir Cell Mol Biol* 2016, **54**(2):177-187.

179. Imkamp K, Berg M, Vermeulen CJ, Heijink IH, Guryev V, Kerstjens HAM, Koppelman GH, van den Berge M, Faiz A: **Nasal epithelium as a proxy for bronchial epithelium for smoking-induced gene expression and expression Quantitative Trait Loci**. *J Allergy Clin Immunol* 2018, **142**(1):314-317 e315.

180. Ober C, McKennan CG, Magnaye KM, Altman MC, Washington C, 3rd, Stanhope C, Naughton KA, Rosasco MG, Bacharier LB, Billheimer D *et al*: **Expression quantitative trait locus fine mapping of the 17q12-21 asthma locus in African American children: a genetic association and gene expression study**. *Lancet Respir Med* 2020, **8**(5):482-492.

181. Giambartolomei C, Zhenli Liu J, Zhang W, Hauberg M, Shi H, Boocock J, Pickrell J, Jaffe AE, CommonMind C, Pasaniuc B *et al*: **A Bayesian framework for multiple trait colocalization from summary association statistics**. *Bioinformatics* 2018, **34**(15):2538-2545.

182. Jackson DJ, Gangnon RE, Evans MD, Roberg KA, Anderson EL, Pappas TE, Printz MC, Lee WM, Shult PA, Reisdorf E *et al*: **Wheezing rhinovirus illnesses in early life predict asthma development in high-risk children**. *Am J Respir Crit Care Med* 2008, **178**(7):667-672.

183. Busse WW, Lemanske RF, Jr., Gern JE: **Role of viral respiratory infections in asthma and asthma exacerbations**. *Lancet* 2010, **376**(9743):826-834.

184. Busse WW: **The atopic march: Fact or folklore?** *Ann Allergy Asthma Immunol* 2018, **120**(2):116-118.

185. Han H, Roan F, Ziegler SF: **The atopic march: current insights into skin barrier dysfunction and epithelial cell-derived cytokines**. *Immunol Rev* 2017, **278**(1):116-130.

186. Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD, Irizarry RA: **Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays**. *Bioinformatics* 2014, **30**(10):1363-1369.

187. Ongen H, Buil A, Brown AA, Dermitzakis ET, Delaneau O: **Fast and efficient QTL mapper for thousands of molecular phenotypes**. *Bioinformatics* 2016, **32**(10):1479-1485.

188. Banovich NE, Lan X, McVicker G, van de Geijn B, Degner JF, Blischak JD, Roux J, Pritchard JK, Gilad Y: **Methylation QTLs are associated with coordinated changes in transcription factor binding, histone modifications, and gene expression levels**. *PLoS Genet* 2014, **10**(9):e1004663.

189. Urbut SM, Wang G, Carbonetto P, Stephens M: **Flexible statistical methods for estimating and testing effects in genomic studies with multiple conditions**. *Nat Genet* 2019, **51**(1):187-195.

190. Stephens M: **False discovery rates: a new deal**. *Biostatistics* 2017, **18**(2):275-294.

191. Iotchkova V, Ritchie GRS, Geihs M, Morganella S, Min JL, Walter K, Timpson NJ, Consortium UK, Dunham I, Birney E *et al*: **GARFIELD classifies disease-relevant genomic features through integration of functional annotations with association signals**. *Nat Genet* 2019, **51**(2):343-353.

192. Institute B: **Pan-UK Biobank, Pan-ancestry genetic analysis of the UK Biobank**. 2020.

193. Paternoster L, Standl M, Chen CM, Ramasamy A, Bonnelykke K, Duijts L, Ferreira MA, Alves AC, Thyssen JP, Albrecht E *et al*: **Meta-analysis of genome-wide association studies identifies three new risk loci for atopic dermatitis**. *Nat Genet* 2011, **44**(2):187-192.

194. Jansen IE, Savage JE, Watanabe K, Bryois J, Williams DM, Steinberg S, Sealock J, Karlsson IK, Hagg S, Athanasiu L *et al*: **Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer's disease risk**. *Nat Genet* 2019, **51**(3):404-413.

195. Nielsen JB, Thorolfsdottir RB, Fritsche LG, Zhou W, Skov MW, Graham SE, Herron TJ, McCarthy S, Schmidt EM, Sveinbjornsson G *et al*: **Biobank-driven genomic discovery yields new insight into atrial fibrillation biology**. *Nat Genet* 2018, **50**(9):1234-1239.

196. Wood AR, Esko T, Yang J, Vedantam S, Pers TH, Gustafsson S, Chu AY, Estrada K, Luan J, Kutalik Z *et al*: **Defining the role of common variation in the genomic and biological architecture of adult human height**. *Nat Genet* 2014, **46**(11):1173-1186.

197. Luciano M, Hagenaars SP, Davies G, Hill WD, Clarke TK, Shirali M, Harris SE, Marioni RE, Liewald DC, Fawns-Ritchie C *et al*: **Association analysis in over 329,000 individuals identifies 116 independent variants influencing neuroticism**. *Nat Genet* 2018, **50**(1):6-11.

198. Smith GD: **Mendelian Randomization for Strengthening Causal Inference in Observational Studies: Application to Gene x Environment Interactions**. *Perspect Psychol Sci* 2010, **5**(5):527-545.

199. Christopher FB, Mark ES, Steven S: **IVREG2: Stata module for extended instrumental variables/2SLS and GMM estimation**. In., S425401 edn: Boston College Department of Economics; 2002.

200. Mattila P, Joenvaara S, Renkonen J, Toppila-Salmi S, Renkonen R: **Allergy as an epithelial barrier disease**. *Clin Transl Allergy* 2011, **1**(1):5.

201. Giambartolomei C, Vukcevic D, Schadt EE, Franke L, Hingorani AD, Wallace C, Plagnol V: **Bayesian test for colocalisation between pairs of genetic association studies using summary statistics**. *PLoS Genet* 2014, **10**(5):e1004383.

202. Wen X, Pique-Regi R, Luca F: **Integrating molecular QTL data into genome-wide genetic association analysis: Probabilistic assessment of enrichment and colocalization**. *PLoS Genet* 2017, **13**(3):e1006646.

203. Ying S, O'Connor B, Ratoff J, Meng Q, Mallett K, Cousins D, Robinson D, Zhang G, Zhao J, Lee TH *et al*: **Thymic stromal lymphopoietin expression is increased in asthmatic airways and correlates with expression of Th2-attracting chemokines and disease severity**. *J Immunol* 2005, **174**(12):8183-8190.

204. Torgerson DG, Ampleford EJ, Chiu GY, Gauderman WJ, Gignoux CR, Graves PE, Himes BE, Levin AM, Mathias RA, Hancock DB *et al*: **Meta-analysis of genome-wide association studies of asthma in ethnically diverse North American populations**. *Nat Genet* 2011, **43**(9):887-892.

205. Shrine N, Portelli MA, John C, Soler Artigas M, Bennett N, Hall R, Lewis J, Henry AP, Billington CK, Ahmad A *et al*: **Moderate-to-severe asthma in individuals of European ancestry: a genome-wide association study**. *Lancet Respir Med* 2019, **7**(1):20-34.

206. Smit LA, Bouzigon E, Pin I, Siroux V, Monier F, Aschard H, Bousquet J, Gormand F, Just J, Le Moual N *et al*: **17q21 variants modify the association between early respiratory infections and asthma**. *Eur Respir J* 2010, **36**(1):57-64.

207. Liu J, Ballaney M, Al-alem U, Quan C, Jin X, Perera F, Chen LC, Miller RL: **Combined inhaled diesel exhaust particles and allergen exposure alter methylation of T helper genes and IgE production in vivo**. *Toxicol Sci* 2008, **102**(1):76-81.

208. Clifford RL, Jones MJ, MacIsaac JL, McEwen LM, Goodman SJ, Mostafavi S, Kobor MS, Carlsten C: **Inhalation of diesel exhaust and allergen alters human bronchial epithelium DNA methylation**. *J Allergy Clin Immunol* 2017, **139**(1):112-121.

209. Gallagher MD, Chen-Plotkin AS: **The Post-GWAS Era: From Association to Function**. *Am J Hum Genet* 2018, **102**(5):717-730.

210. Consortium GT, Laboratory DA, Coordinating Center -Analysis Working G, Statistical Methods groups-Analysis Working G, Enhancing Gg, Fund NIHC, Nih/Nci, Nih/Nhgri, Nih/Nimh, Nih/Nida *et al*: **Genetic effects on gene expression across human tissues**. *Nature* 2017, **550**(7675):204-213.

211. Finucane HK, Bulik-Sullivan B, Gusev A, Trynka G, Reshef Y, Loh PR, Anttila V, Xu H, Zang C, Farh K *et al*: **Partitioning heritability by functional annotation using genome-wide association summary statistics**. *Nat Genet* 2015, **47**(11):1228-1235.

212. Loerbroks A, Apfelbacher CJ, Thayer JF, Debling D, Sturmer T: **Neuroticism, extraversion, stressful life events and asthma: a cohort study of middle-aged adults**. *Allergy* 2009, **64**(10):1444-1450.

213. Huovinen E, Kaprio J, Koskenvuo M: **Asthma in relation to personality traits, life satisfaction, and stress: a prospective study among 11,000 adults**. *Allergy* 2001, **56**(10):971-977.

214. Ormel J, Bastiaansen A, Riese H, Bos EH, Servaas M, Ellenbogen M, Rosmalen JG, Aleman A: **The biological and psychological basis of neuroticism: current status and future directions**. *Neurosci Biobehav Rev* 2013, **37**(1):59-72.

215. Depue RA, Fu Y: **Neurogenetic and experiential processes underlying major personality traits: implications for modelling personality disorders**. *Int Rev Psychiatry* 2011, **23**(3):258-281.

216. West EE, Kashyap M, Leonard WJ: **TSLP: A Key Regulator of Asthma Pathogenesis**. *Drug Discov Today Dis Mech* 2012, **9**(3-4).

217. Wang IJ, Chen SL, Lu TP, Chuang EY, Chen PC: **Prenatal smoke exposure, DNA methylation, and childhood atopic dermatitis**. *Clin Exp Allergy* 2013, **43**(5):535-543.

218. Luo Y, Zhou B, Zhao M, Tang J, Lu Q: **Promoter demethylation contributes to TSLP overexpression in skin lesions of patients with atopic dermatitis**. *Clin Exp Dermatol* 2014, **39**(1):48-53.

219. Hui CC, Yu A, Heroux D, Akhabir L, Sandford AJ, Neighbour H, Denburg JA: **Thymic stromal lymphopoietin (TSLP) secretion from human nasal epithelium is a function of TSLP genotype**. *Mucosal Immunol* 2015, **8**(5):993-999.

220. Bouzigon E, Corda E, journal … A-H: **Effect of 17q21 variants and smoking exposure in early-onset asthma**. *New England journal ...* 2008.

221. Modena BD, Bleecker ER, Busse WW, Erzurum SC, Gaston BM, Jarjour NN, Meyers DA, Milosevic J, Tedrow JR, Wu W: **Gene expression correlated with severe asthma characteristics reveals heterogeneous mechanisms of severe disease**. *American journal of respiratory and critical care medicine* 2017, **195**(11):1449-1463.

222. Ober C: **Asthma Genetics in the Post-GWAS Era**. *Annals of the American Thoracic Society* 2016, **13 Suppl 1**(Supplement 1):90.

223. McCarthy MI, Hirschhorn JN: **Genome-wide association studies: potential next steps on a genetic journey**. *Human molecular genetics* 2008, **17**(R2):65.

224. Bonnelykke K, Ober C: **Leveraging gene-environment interactions and endotypes for asthma gene discovery**. *J Allergy Clin Immunol* 2016, **137**(3):667-679.

225. Calderon D, Bhaskar A, Knowles DA, Golan D, Raj T, Fu AQ, Pritchard JK: **Inferring Relevant Cell Types for Complex Traits by Using Single-Cell Gene Expression**. *American Journal of Human Genetics* 2017, **101**(5):686-699.

226. Cohen NA, Widelitz JS, Chiu AG, Palmer JN, Kennedy DW: **Familial aggregation of sinonasal polyps correlates with severity of disease**. *Otolaryngol Head Neck Surg* 2006, **134**(4):601-604.

227. Delagrand A, Gilbert-Dussardier B, Burg S, Allano G, Gohler-Desmonts C, Lebreton JP, Dufour X, Klossek JM: **Nasal polyposis: is there an inheritance pattern? A single family study**. *Rhinology* 2008, **46**(2):125-130.

228. Lockey RF, Rucknagel DL, Vanselow NA: **Familial occurrence of asthma, nasal polyps and aspirin intolerance**. *Ann Intern Med* 1973, **78**(1):57-63.

229. Greisner WA, 3rd, Settipane GA: **Hereditary factor for nasal polyps**. *Allergy Asthma Proc* 1996, **17**(5):283-286.

230. Kim JY, Kim DK, Yu MS, Cha MJ, Yu SL, Kang J: **Role of epigenetics in the pathogenesis of chronic rhinosinusitis with nasal polyps**. *Mol Med Rep* 2018, **17**(1):1219-1227.

231. Kidoguchi M, Noguchi E, Nakamura T, Ninomiya T, Morii W, Yoshida K, Morikawa T, Kato Y, Imoto Y, Sakashita M *et al*: **DNA Methylation of Proximal PLAT Promoter in Chronic Rhinosinusitis With Nasal Polyps**. *Am J Rhinol Allergy* 2018, **32**(5):374-379.

232. Li J, Jiao J, Wang M, Gao Y, Li Y, Wang Y, Zhang Y, Wang X, Zhang L: **Hypomethylation of the IL8 promoter in nasal epithelial cells of patients with chronic rhinosinusitis with nasal polyps**. *J Allergy Clin Immunol* 2019, **144**(4):993-1003 e1012.

233. Demanelis K, Argos M, Tong L, Shinkle J, Sabarinathan M, Rakibuz-Zaman M, Sarwar G, Shahriar H, Islam T, Rahman M *et al*: **Association of Arsenic Exposure with Whole Blood DNA Methylation: An Epigenome-Wide Study of Bangladeshi Adults**. *Environ Health Perspect* 2019, **127**(5):57011.