An investigation of expression quantitative trait loci in interferon genes in childhood asthma using RNA-sequencing data

Sneha P. Cherukuri

A thesis

Submitted in partial fulfillment of the requirements for the degree of

Master of Science

University of Washington

2021

Committee:

Alison Fohner

Matthew C Altman

Program Authorized to Offer Degree:

Epidemiology

University of Washington

Abstract

An investigation of expression quantitative trait loci in interferon genes in childhood asthma using RNA-sequencing data

Sneha P. Cherukuri

Chair of the Supervisory Committee:
Alison Fohner
Department of Epidemiology

Background: Asthma is the most common chronic disease among children and yet, many of the underlying mechanisms of severe disease remain poorly understood. Expression quantitative trait loci (eQTL) can provide mechanistic insight into the consequences of asthma risk variants and help explain variation in asthma severity. We identified and evaluated eQTLs in interferon molecular networks previously described in a transcriptome network analysis.

Methods: This was a prospective cohort study with a high disease burden cohort consisting primarily of children from racial and ethnic minority populations. Sequenced RNA from nasal lavage samples were collected from 105 children diagnosed with asthma. Single nucleotide polymorphisms (SNPs) from 6 interferon induced genes were assessed for Hardy-Weinberg equilibrium and used for cis-eQTL analysis after exclusions. Functional consequences of these SNPs were evaluated in genomic databases.

Results: 5 SNPs from the *IFI16*, *IFI44L*, and *IFIH1* genes were significantly associated (FDR<.05) with gene expression. 4 out of the 5 were described as missense variants, while 1 was described as a 3 prime untranslated region variant.

Discussion: While 2 of the eQTLs from *IFI16* were previously identified in the mucosa of the esophagus, we identified 3 novel airway epithelium eQTLs in *IFI44L* and *IFIH1*. Investigation of these genetic variants suggests they play a role in recognition of viruses and the prevention of viral infections, which may be crucial in efforts to prevent and treat viral asthma exacerbations.

Background

Around 300 million people worldwide have asthma, and it is ranked 16th among the leading causes of years lived with disability and 28th among the leading causes of burden of disease, as measured by disability-adjusted life years [1]. There is a wide range of disease severity in regard to asthma. It may be mild, barely noticed by the patient, or it may greatly hinder the life of the patient by causing constant symptoms, inability to perform daily activities, poor quality of life, severe and life-threatening attacks, and even death [2]. There is also a significant economic burden due to asthma. It is estimated that, from 2019 to 2038, the total cost of uncontrolled asthma could exceed $963 billion in direct and indirect costs in the United States alone [3]. Therefore, there is great interest in the process of investigating disparities in clinical outcomes and understanding the cellular and molecular basis for severe asthma.

Previous studies have focused on genome wide association studies (GWAS) in an attempt to identify asthma risk variants, but adverse, phenotypic consequences of these risk variants are often unclear. GWAS has demonstrated that the majority of genome variants are found in non-coding regions of the genome and are therefore likely to be involved in gene regulation [4]. Therefore, identifying and evaluating expression quantitative trait loci (eQTLs) can add important mechanistic insight into the consequences of asthma risk variants and help explain variation in asthma severity. An eQTL is a locus that explains the genetic variance associated with a gene expression phenotype [4]. These regulatory variants can be either *cis* or *trans* acting. Cis eQTLs are generally defined as 1 megabase (Mb) upstream or downstream of the gene's

transcriptional start site (TSS), while trans eQTLs are at least 5 Mb from the TSS or on a different chromosome. Cis eQTLs are more easily and more frequently identified in eQTL studies, likely due to the fact that most of the regulatory control takes place locally, in the vicinity of genes [5-7]. Direct association tests between single nucleotide polymorphisms (SNPs) and gene expression levels allow for identification of eQTLs.

This study contributes to the existing literature by investigating eQTLs in a high disease burden cohort consisting primarily of children from racial and ethnic minority populations. Asthma is the most common chronic disease among children and yet, many of the underlying mechanisms of severe disease remain poorly understood [8]. Significant disparities also exist based on race and ethnicity in that Black, Hispanic, and Indigenous people in the United States are disproportionately burdened by asthma, but research studies involving these groups are limited [9]. From 1993 until 2013, inclusion of members of racial or ethnic minority groups was reported in less than 5% of all NIH-funded published studies of respiratory diseases [10].

In this study, we identified and evaluated eQTLs in interferon molecular networks previously described by Altman, et. all, the first study in asthma to use transcriptome network analysis in a case–control, longitudinal study design [11].

Methods

Study subjects and setting

This study was a prospective cohort study using data from The Mechanisms Underlying Asthma Exacerbations Prevented and Persistent with Immune-Based Therapy (MUPPITS) cohort.

The MUPPITS cohort recruited children who were age 6 to 17 years old; diagnosed with asthma by a clinician over one year before study recruitment; had at least 2 asthma exacerbations in the prior year that required systemic corticosteroids and/or hospitalization; were treated with at least fluticasone 250 mcg 1 puff twice daily or its equivalent for those aged 6 to 11 years or treated with at least Advair 250/50 mcg 1 puff twice daily or its equivalent for those aged 12 years and older; had peripheral blood eosinophils ≥ 150 per mm [15]; was a non-smoker; and lived in a census tract with a density of ≥1000 families per square mile and at least 10% of families with income below the poverty level. An exacerbation was defined as treatment with systemic corticosteroids within 10 days of the first sign of respiratory symptoms. Recruitment occurred between 2015 and 2017 across urban sites in 9 U.S. cities: Boston, New York, Detroit, Denver, Washington D.C., Chicago, Dallas, Cincinnati, and St. Louis [11].

Data collection

Nasal lavage samples from children in the MUPPITS cohort were collected from participants twice in the 6 days following the start of respiratory symptoms. This dataset includes sequenced RNA from those nasal lavage samples in 105 children and is publicly available at the Gene Expression Omnibus (GEO) accession number GSE115824.

Analysis

We identified all known asthma GWAS SNPs using the filter “asthma” from the NHGRI-EBI Catalog of human genome-wide association studies and filtered to only those SNPs that were within genes defined as part of the type 1 interferon response modules (*m11.n* and *neut5.n)* as described by Altman, et. al [11]. After filtering to SNPs with at least one variant identified, 126 genes were identified, each with at least one SNP that fell within the span of that gene’s coding region. For each SNP, minor allele frequencies were calculated using minor allele counts of 0, 1, or 2 from each genotype. A total of 6 interferon induced (IFI) family genes: *IFI16*, *IFI35*, *IFI44*, *IFI44L*, *IFIH1*, and *IFITM3* were selected and used for cis-eQTL analysis.

We calculated p and q values for each SNP and assessed if the SNPs met Hardy-Weinberg (HW) equilibrium using the HardyWeinberg R package. The package uses the HW equilibrium equation: $p^2 + 2pq + q^2 = 1$ and runs the chi-squared test with continuity correction to detect SNPs that do not meet HW equilibrium [12]. Significant deviations from HW equilibrium are often the consequence of genotyping error, and HW equilibrium tests are an efficient way of detecting (gross) genotyping error and assuring data quality [13]. SNPs that did not meet HW equilibrium (p-value <.01) were excluded.

Weighted linear models using SNPs from each gene and gene expression data were run using the limma package in R. Considering each SNP as the independent variable and gene expression as the dependent variable, the model adjusted for percent of each cell type in the sample, the presence or absence of a virus infection, the library

sequencing depth, corticosteroid use, sex, study site, and age, all of which were previously observed to affect global gene expression [11]. Significant eQTL SNPs were defined as those with a false discovery rate (FDR) less than .05. Gene expression levels for each significant SNP were examined and visualized on a log2 scale per genotype.

Potential linkage disequilibrium (LD) was considered and calculated with the LDcorSV package in R [14]. Raw LD calculations typically have a bias that is acknowledged in LDcorSV, which considers and corrects for the structure of the sample or the relatedness of the genotyped individuals [15]. The Pearson coefficient of correlation (r) uses allele frequencies to calculate LD between two loci. Squaring the value eliminates differences in sign. With LDcorSV, LD matrices of all pairwise SNP associations in each gene were created and visualized using calculated r^2 values. SNPs associated with gene expression levels in high LD were further analyzed through existing literature for joint effects.

SNPs associated with gene expression were reviewed with the UCSC genome browser and The Genotype-Tissue Expression (GTEx) portal for clinical and functional effects.

Results

A total of 105 children from the MUPPITS cohort were diagnosed with asthma and included in this study. The female to male ratio was nearly equivalent (52 female, 54 male). The majority (53.8%) of children were Black, non-Hispanic, about a third (33.9%)

were Hispanic, and the remainder consisted of other and White non-Hispanic. The mean age was 10.8 years, with the youngest child being 6 years old and the oldest child being 17 years old.

Among the 6 interferon module genes that were considered for cis-eQTL analysis, 3 genes had at least one statistically significant eQTL (FDR<.05). *IFI35* did not have any SNPs that met HW equilibrium, while *IFI44* and *IFITM3* did not have any SNPs that were associated with gene expression regulation. 2 statistically significant eQTLs in *IFI16* (rs1057027, rs1057028) had a minor allele frequency of 0.43 (table 2). The minor allele (A) for these SNPs was associated with lower expression of *IFI16*. Both rs1057027 and rs1057028 have previously been identified on the GTEx portal as statistically significant eQTLs in skin, spleen, whole blood, artery-tibial, and esophagus-mucosa tissues [15]. These SNPs had an r^2 value of 1, indicating that the two SNPs are in perfect linkage disequilibrium. In *IFIH1*, rs3747517 and rs1990760 were 2 statistically significant eQTLs with minor allele frequencies of 0.32 and 0.24 respectively. The minor allele (C) in rs3747517 was associated with decreased *IFIH1* gene expression and minor allele (T) in rs1990760 was associated with decreased expression. Interestingly, there is only one entry in the GTEx portal of rs3747517 as an eQTL (p-value < .01), and only within cultured fibroblast cells. There was also no record of rs1990760 as a significant eQTL in any tissues in the GTEx portal. These SNPs had an r^2 value of 0.265. In *IFI44L*, rs4650590 was identified as the only eQTL with a minor allele frequency of 0.31. The minor allele (G) was associated with increased *IFI44L* expression. While the other 4 eQTLs in this study were classified as missense variants,

rs4650590 was identified in dbSNP as a 3 prime untranslated region variant and has not been identified as an eQTL in any tissues in GTEx.

Discussion

Interferon gamma inducible protein 16 (*IFI16*) is an intracellular recognition sensor that triggers inflammatory responses against DNA from pathogens [16]. *IFI16* also has an N-terminal region that can bind to DNA and likely plays a role in regulating the activity of transcription factors such as p53 and SP1 in the nucleus that are involved in apoptosis [17, 18]. In this study, rs1057027 and rs1057028 were found to be in high LD, (r^2=1) which is consistent with previous association studies with *IFI16* SNPs [16]. Common mutations in these SNPs are associated with *IFI16* protein damage, although not much else is known about the functional effect of these SNPs. Thus, there may be a greater need for *IFI16* functional experiments in future studies.

The interferon-induced helicase-1 (*IFIH1*) gene encodes the melanoma differentiation associated protein 5 (MDA5), an intracellular pattern recognition receptor (PRR) of importance for the recognition of certain viruses [19]. MDA5 monomers bind to long double stranded RNA molecules formed during the viral replication cycle. Rs3747517 and rs1990760 were identified as statistically significant eQTLs in this study. In a separate study, homozygous and heterozygous asthmatic children for the major allele in rs3747517 were found to have lower exacerbation risk compared to children who were homozygous for the alternate allele [20]. Rs1990760 is associated with an increased risk for development of autoimmune diseases including systemic lupus erythematosus and

multiple sclerosis [21]. In particular, the risk allele (T) at rs1990760 has been associated with type 1 diabetes. Another study found that the homozygous genotype for the minor allele at rs3747517 and rs1990760 was associated with increased viral replication among asthmatic children [20]. The r^2 value for these two SNPs was relatively low (.265), indicating they are not in strong linkage disequilibrium and that the association of one SNP cannot be explained by the other.

There is evidence that interferon-induced protein 44-like (*IFI44L*) gene is an important effector of innate antiviral immunity with a plausible role in immune response priming and protection against disease [21]. *IFI44L* is a part of the *IFI44* gene family and is stimulated by type 1 interferon. Previous studies have demonstrated that *IFI44L* may also inhibit Hepatitis C Virus replication and act as a novel tumor suppressor in human hepatocellular carcinoma [22]. While rs4650590 has specifically been associated with measles-specific humoral immunity, there is not much clinical significance data for this SNP [21].

IFI16, *IFIH1*, and *IFI44L* have all been associated with recognition of viruses and the prevention of viral infections. Because the majority of asthma exacerbations are provoked by viral respiratory infections [11], targeted intervention could provide therapeutic benefits. Future studies may also investigate functional effects of the SNPs we identified from this study and their clinical outcomes.

Our study is in some part limited by population substructure: allele frequencies may differ in sub populations with different ancestry simply due to chance, particularly in genetic studies. This limitation may have attenuated or enhanced significant findings from this study. However, excluding SNPs outside of Hardy-Weinberg equilibrium can help mitigate this limitation [23], so we expect the impact of this limitation to be minimal. In conclusion, this study identified and examined novel airway sample eQTLs in *IFI16*, *IFI44L*, and *IFIH1* in order to better understand severe clinical outcomes associated with childhood asthma.

References

1. Dharmage, Shyamali C et al. “Epidemiology of Asthma in Children and Adults.” *Frontiers in pediatrics* vol. 7 246. 18 Jun. 2019, doi:10.3389/fped.2019.00246
2. Bostantzoglou, Clementine et al. “Clinical asthma phenotypes in the real world: opportunities and challenges.” *Breathe (Sheffield, England)* vol. 11,3 (2015): 186-93. doi:10.1183/20734735.008115
3. “AAFA Rolls Out National Roadmap to Address Serious Asthma Disparities Impacting Black and Hispanic Americans.” 16 Sept. 2020, www.aafa.org/media/2752/press-release-aafa-asthma-disparities-in-america-report-16-september-2020.pdf.
4. Nica, Alexandra C, and Emmanouil T Dermitzakis. “Expression quantitative trait loci: present and future.” *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* vol. 368,1620 20120362. 6 May. 2013, doi:10.1098/rstb.2012.0362
5. Göring, Harald H H et al. “Discovery of expression QTLs using large-scale transcriptional profiling in human lymphocytes.” *Nature genetics* vol. 39,10 (2007): 1208-16. doi:10.1038/ng2119
6. Schadt, Eric E et al. “Mapping the genetic architecture of gene expression in human liver.” *PLoS biology* vol. 6,5 (2008): e107. doi:10.1371/journal.pbio.0060107

7. Dixon, Anna L et al. “A genome-wide association study of global gene expression.” *Nature genetics* vol. 39,10 (2007): 1202-7. doi:10.1038/ng2109
8. Reese, Sarah E et al. “Epigenome-wide meta-analysis of DNA methylation and childhood asthma.” *The Journal of allergy and clinical immunology* vol. 143,6 (2019): 2062-2074. doi:10.1016/j.jaci.2018.11.043
9. Asthma and Allergy Foundation of America, (2020). [Asthma Disparities in America: A Roadmap to Reducing Burden on Racial and Ethnic Minorities]. Retrieved from aafa.org/asthmadisparities.
10. Burchard, Esteban G et al. “Moving toward true inclusion of racial/ethnic minorities in federally funded studies. A key step for achieving respiratory health equality in the United States.” *American journal of respiratory and critical care medicine* vol. 191,5 (2015): 514-21. doi:10.1164/rccm.201410-1944PP
11. Altman, Matthew C et al. “Transcriptome networks identify mechanisms of viral and nonviral asthma exacerbations in children.” *Nature immunology* vol. 20,5 (2019): 637-651. doi:10.1038/s41590-019-0347-8
12. Graffelman, Jan, and Jair Morales Camarena. “Graphical tests for Hardy-Weinberg equilibrium based on the ternary plot.” *Human heredity* vol. 65,2 (2008): 77-84. doi:10.1159/000108939
13. Graffelman, J, and B S Weir. “Testing for Hardy-Weinberg equilibrium at biallelic genetic markers on the X chromosome.” *Heredity* vol. 116,6 (2016): 558-68. doi:10.1038/hdy.2016.20
14. David Desrousseaux, Florian Sandron, Aurélie Siberchicot, Christine Cierco-Ayrolles and Brigitte Mangin (2020). LDcorSV: Linkage Disequilibrium Corrected by the Structure and the Relatedness. R package version 1.3.3. https://CRAN.R-project.org/package=LDcorSV
15. Mangin, B et al. “Novel measures of linkage disequilibrium that correct the bias due to population structure and relatedness.” *Heredity* vol. 108,3 (2012): 285-91. doi:10.1038/hdy.2011.73
16. Marchesan, Julie T et al. “Common Polymorphisms in IFI16 and AIM2 Genes Are Associated With Periodontal Disease.” *Journal of periodontology* vol. 88,7 (2017): 663-672. doi:10.1902/jop.2017.160553
17. Ansari, Mairaj Ahmed et al. “Constitutive interferon-inducible protein 16-inflammasome activation during Epstein-Barr virus latency I, II, and III in B and epithelial cells.” *Journal of virology* vol. 87,15 (2013): 8606-23. doi:10.1128/JVI.00805-13
18. Aglipay, Jason A et al. “A member of the Pyrin family, IFI16, is a novel BRCA1-associated protein involved in the p53-mediated apoptosis pathway.” *Oncogene* vol. 22,55 (2003): 8931-8. doi:10.1038/sj.onc.1207057

19. Domsgen, Erna et al. “An IFIH1 gene polymorphism associated with risk for autoimmunity regulates canonical antiviral defence pathways in Coxsackievirus infected human pancreatic islets.” *Scientific reports* vol. 6 39378. 21 Dec. 2016, doi:10.1038/srep39378
20. Barrow, KA et. al. Deficient IFIH1 Expression and Type-I Interferon Response to Rhinovirus by Nasal Epithelial Cells from Asthmatic Children Homozygous for the IFIH1 Minor Alleles of SNPs rs3747517 and rs1990760 Is Associated with Increased Viral Replication Ex Vivo [abstract]. In: American Thoracic Society International Conference Abstracts; 2018 May 23; San Diego, California (CA): AJRCCM; 2018. Abstract nr D21.
21. Haralambieva, Iana H et al. “Genome-wide associations of CD46 and IFI44L genetic variants with neutralizing antibody response to measles vaccine.” *Human genetics* vol. 136,4 (2017): 421-435. doi:10.1007/s00439-017-1768-9
22. Huang, Wei-Chieh et al. “IFI44L is a novel tumor suppressor in human hepatocellular carcinoma affecting cancer stemness, metastasis, and drug resistance via regulating met/Src signaling pathway.” *BMC cancer* vol. 18,1 609. 30 May. 2018, doi:10.1186/s12885-018-4529-9
23. Tian, Chao et al. “Accounting for ancestry: population substructure and genome-wide association studies.” *Human molecular genetics* vol. 17,R2 (2008): R143-50. doi:10.1093/hmg/ddn268
24. GTEx Consortium. The GTEx Consortium atlas of genetic regulatory effects across human tissues. In press.
25. Sherry,S.T., Ward,M. and Sirotkin,K. (1999) dbSNP—Database for Single Nucleotide Polymorphisms and Other Classes of Minor Genetic Variation. *Genome Res*., 9, 677–679.
26. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D. The human genome browser at UCSC. *Genome Res*. 2002 Jun;12(6):996-1006.