HHS Public Access

Author manuscript

J Allergy Clin Immunol. Author manuscript; available in PMC 2022 December 01.

Published in final edited form as:

J Allergy Clin Immunol. 2021 December ; 148(6): 1505–1514. doi:10.1016/j.jaci.2021.04.035.

Inducible expression quantitative trait locus analysis of the *MUC5AC* gene in asthma in urban populations of children

Matthew C Altman, MD[1,2,*], **Kaitlin Flynn, PhD**[2], **Mario G Rosasco, PhD**[2], **Matthew Dapas, PhD**[3], **Meyer Kattan, MD**[4], **Stephanie Lovinsky-Desir, MD**[4], **George T O'Connor, MD**[5], **Michelle A Gill, MD**[6], **Rebecca S Gruchalla, MD**[6], **Andrew H Liu, MD**[7], **Jacqueline A Pongracic, MD**[8], **Gurjit K Khurana Hershey, MD**[9], **Edward M Zoratti, MD**[10], **Stephen J Teach, MD**[11], **Deepa Rastrogi, MBBS, MS**[11], **Robert A. Wood, MD**[12], **Leonard B Bacharier, MD**[13], **Petra LeBeau, PhD**[14], **Peter J Gergen, MD**[15], **Alkis Togias, MD**[15], **William W Busse, MD**[16], **Scott Presnell, PhD**[2], **James E Gern, MD**[16], **Carole Ober, PhD**[3], **Daniel J Jackson, MD**[16], **NIAID Inner City Asthma Consortium**

[1]Department of Medicine, University of Washington, Seattle WA

[2]Benaroya Research Institute, Seattle WA

[3]Department of Human Genetics, University of Chicago, Chicago, IL

[4]Columbia University, New York NY

[5]Boston University School of Medicine, Boston MA

[6]UT Southwestern Medical Center, Dallas TX

[7]Children's Hospital Colorado University of Colorado School of Medicine, Aurora CO

[8]Ann Robert H Lurie Children's Hospital of Chicago, Chicago IL

[9]Cincinnati Children's Hospital Cincinnati OH

[10]Henry Ford Health System Detroit MI

[11]Children's National Hospital, Washington DC

[12]Department of Pediatrics, Johns Hopkins University Medical Center, Baltimore, MD

[13]Washington University, Saint Louis MO

[14]Rho Inc, Chapel Hill NC

[15]NIH/NIAID, Bethesda MD

[*]**Corresponding author**: Matthew C. Altman, MD, Systems Immunology Division, Benaroya Research Institute, 1201 Ninth Avenue, Seattle, WA 98101, USA. Tel. +1 206 287 5648, Fax. 206 287 5682, maltman@benaroyaresearch.org.

Author contributions: M.C.A., M.G.R., J.E.G., C.O., and D.J.J. made contributions to the conception and/or design of the study. M.C.A., K.F., M.G.R, M.D., P.L., S.P., and C.O. oversaw or performed all statistical analyses. M.K., S.L., G.T.O., M.A.G., R.S.G., A.H.L., J.A.P., G.K.H., E.M.Z., D.R., S.J.T., R.A.W., L.B.B., contributed patient samples and/or data. M.C.A. and D.J.J. wrote the manuscript. K.F., M.G.R., M.D., W.W.B., S.P., J.E.G, C.O. read and edited early drafts of the manuscript. All authors reviewed and/or critically revised the manuscript for important intellectual content and provided final approval.

[16]University of Wisconsin School of Medicine and Public Health, Madison WI

Abstract

Background: Mucus plugging can worsen asthma control, lead to reduced lung function and fatal exacerbations. MUC5AC is the secretory mucin implicated in mucus plugging, and *MUC5AC* gene expression has been associated with development of airway obstruction and asthma exacerbations in urban children with asthma. However, the genetic determinants of *MUC5AC* expression are not established.

Objective: To assess single-nucleotide polymorphisms (SNPs) that influence *MUC5AC* expression and relate to pulmonary functions in childhood asthma.

Methods: We used RNA-sequencing data from upper airway samples and performed cis-expression quantitative trait loci (eQTL) and allele specific expression (ASE) analyses in two cohorts of predominantly Black and Hispanic urban children, a high asthma-risk birth cohort and an exacerbation-prone asthma cohort. We further investigated inducible *MUC5AC* eQTLs during incipient asthma exacerbations. We tested significant eQTLs SNPs for associations with lung function measurements and investigated their functional consequences in DNA regulatory databases.

Results: We identified two independent groups of SNPs in the *MUC5AC* gene that were significantly associated with *MUC5AC* expression. Moreover, these SNPs showed stronger eQTL associations with *MUC5AC* expression during asthma exacerbations, consistent with inducible expression. SNPs in one group also showed significant association with decreased pulmonary functions. These SNPs included multiple EGR1 transcription factor binding sites suggesting a mechanism of effect.

Conclusions: These findings demonstrate the applicability of organ specific RNA-sequencing data to determine genetic factors contributing to a key di6sease pathway. Specifically, they suggest important genetic variations that may underlie propensity to mucus plugging in asthma and could be important in targeted asthma phenotyping and disease management strategies.

Capsule summary:

Our results add new information regarding the role of *MUC5AC* genotypes in childhood asthma pathogenesis in populations with high disease burden and suggest genetic predictors for targeted phenotyping and management strategies focused on mucin dysregulation.

Keywords

Asthma; Expression quantitative trait; MUC5AC

Introduction

Estimates of asthma heritability range from 35-90%, but genome-wide association studies (GWAS) only account for a minor fraction of the observed heritability[1–4]. This is likely due in part to the racial, phenotypic, and molecular heterogeneity of asthma, but also to the limitations of GWAS approaches[4]. Moreover, the functional consequences of many asthma

risk variants found by GWAS are incompletely understood[5–9]. Investigation of molecular and phenotypic consequences of genetic variation in or near genes implicated in asthma pathogenesis, and in targeted high-risk populations, can provide important new insights into the genetics of asthma with potential implications for population and personalized disease management.

Urban populations have increased asthma prevalence, suboptimal response to asthma therapies, and are more likely to have reduced lung function as they age[10–13]. Our previous studies highlighted *MUC5AC* overexpression as a specific characteristic in children who develop recurrent wheezing, high levels of allergic sensitization, and progressive loss of lung function during childhood[14], and as a gene upregulated in the airways during incipient asthma exacerbations in children with exacerbation-prone disease[15]. Understanding genetic factors contributing to *MUC5AC* induction is critical given that overproduction can impair mucociliary transport and lead to mucus plugging[16, 17], which is the most common pathologic feature in patients with fatal asthma[18–20] and is a feature of severe disease in adult asthma[21]. Furthermore, genotypes in the *MUC5AC* region of chromosome 11 have recently been associated with asthma in adult European populations[22, 23]

We therefore hypothesized that *MUC5AC* genotypes could differentially regulate its expression in the airway, in particular during acute exacerbations of asthma. To test this hypothesis, we leveraged RNA-sequencing data from airway samples from two urban, predominantly Black and Hispanic cohorts of children to both genotype individuals and then perform combined expression quantitative trait loci (eQTL) and confirmatory allele specific expression (ASE) analyses from samples collected during disease control and during asthma exacerbations.

Methods

Study design and sample composition:

The URECA birth cohort study started enrollment in 2005 in urban neighborhoods with high rates (>20%) of poverty in Baltimore, Boston, New York City, and St. Louis as previously described.[24] Pregnant women, whose children were at high risk for developing asthma due to a history of asthma, allergic rhinitis, or eczema in the mother or father, were recruited prenatally. Written informed consent was obtained from the families, and assent was obtained for children at ages 7 years and older. Asthma diagnosis was determined at age 10 and nasal cells were obtained at age 11 by brushing the anterior turbinate with a cytology brush. RNA was extracted and sequenced to analyze gene expression as previously described[14] in 318 participants used for this analysis (Table 1); in brief total RNA was isolated from nasal epithelial brushes in RLT Plus lysis buffer (Qiagen) by using a Qiashredder column (Qiagen) and RNeasy mini kits (Qiagen). RNA-sequencing libraries were constructed from total RNA using SMART-Seq v4 Ultra Low Input RNA Kit (Takara) and single-read sequencing was carried out on a HiSeq2500 sequencer (illumina), using a HiSeq SBS v4 Kit to generate 58-base reads. This data is publicly available at Gene Expression Omnibus (GEO) accession number GSE145505.

The MUPPITS study recruited children with established exacerbation-prone asthma 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, as previously described[15]. In brief, a participant was eligible for enrollment if he or she: was 6 to 17 years of age; was diagnosed with asthma by a clinician greater than 1 year prior to recruitment; had at least 2 asthma exacerbations in the prior year (defined as a requirement for systemic corticosteroids (SCS) and/or hospitalization); was 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 fluticasone-salmeterol 250/50 meg 1 puff twice daily or its equivalent for those aged 12 years and older; had peripheral blood eosinophils ≥ 150 per mm^3; was a non-smoker; 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. Participants were followed prospectively for respiratory illnesses, during which they returned to the study site twice in the 6-day period after the start of symptoms for collection of nasal lavage samples and pulmonary function testing. An exacerbation illness was defined based on whether the participant was treated with SCS within 10 days following the onset of respiratory symptoms or not. RNA was extracted from nasal lavage samples and sequenced to analyze gene expression as previously described[15] in 106 participants used for this analysis (Table 1); in brief, total RNA was isolated from nasal cell pellets in RNAprotect Saliva Reagent (Qiagen) by using a Qiashredder column (Qiagen) and RNeasy mini kits (Qiagen). Sequencing libraries were constructed from total RNA using TruSeq RNA Sample Preparation Kits v2 (illumina) and single-read sequencing was carried out on a HiSeq2500 sequencer (illumina), using a HiSeq SBS v4 Kit to generate 58-base reads. These data are publicly available at GEO accession number GSE115824.

SNP determination and validation:

RNA-sequencing reads were first mapped to the GRCh38.p10 reference genome using the STAR aligner (v2.4.2a). Base calls were made at all SNP positions listed in NCBI dbSNP Build 151 using samtools (v1.7) with the arguments ‘samtools mpileup -f $grch38FastaFile -B -C 50 -d 250 -I $dbSnp151 Bedfile -q 30 -Q 13 -g -e 20 -h 100 -L 250 -o 40 $alignmentFile -o $bcfCallFile’. The call files were converted to VCF format and filtered to exclude sites where fewer than 10 reads aligned using bcftools (v1.9). SNPs were further filtered to select those with at least 2 different genotypes represented among cohort subjects. MUPPITS SNPs and libraries were filtered to require at least 75% concordance in calls among libraries from the same donor and any non-concordant libraries were removed from the analysis. For 149 participants in the URECA cohort who previously had DNA-based microarray genotyping using an illumina Custom 3000 platform (550,224 variants), 8,568 overlapping SNPs were compared between RNA-based and DNA-based calls and found to be 99.2% concordant.

Statistical analyses:

SNPs spanning the *MUC5AC* gene, spanning from the transcriptional start site to the transcriptional end site (chr11:1157953-1201138), were used for eQTL analysis. SNPs were filtered to those genotyped in >50% of the URECA cohort and checked that they satisfied Hardy-Weinberg equilibrium by the chi square goodness-of-fit test with 2 degrees

of freedom (Benjaminin-Hochberg [BH] false discovery rate adjusted p-value [FDR]>0.05) using the HardyWeinberg R package[25]. *MUC5AC* expression was associated to SNP genotypes using a weighted linear model using the limma R package and empirical Bayes method[26, 27] and considering each allele under an additive effect (0, 1, or 2 minor alleles). For multiallelic sites, any minor allele was counted as 1. For the URECA cohort, the model was adjusted for cell percentages in the sample and study site, which contributed to gene expression variability, as previously described[14]. For the MUPPITS cohort, the model was adjusted for cell percentages, study site, viral positivity, age, sequencing depth and a random effect for individual, which contributed to gene expression variability as previously described[15]. In both cohorts further adjusting for self-reported race gave very similar results, so race was not included in the final models. Multiple testing corrections were performed using the BH method and significant eQTL SNPs were defined as those with a FDR<0.05. Pairwise r^2 measures of linkage disequilibrium (LD) were calculated among significant SNPs for each cohort using the LDheatmap R package[28]. A Euclidean distance matrix was constructed from the r^2 values in URECA and used for hierarchical clustering using the average method (hclust function in R), which divided SNPs into 3 cluster groups (called A, B, or other). Group A and B representative SNPs were selected from the eQTL significant SNPs as those with highest percentages of calls to avoid potential bias by smaller sample sizes. These SNP groups were assessed for independent eQTL effects by performing conditional analyses in which we included the representative SNP or SNPs (0, 1, or 2 minor alleles) as covariates in the eQTL linear model and investigated the effect on the remaining SNPs. In the MUPPITS cohort, the interaction effect between a representative SNP and exacerbation status was determined by adding an interaction term to the linear model. ASE was assessed by calculating the number of counts of the minor allele divided by the total counts (minor and major allele) within each individual heterozygous at a given SNP. Significance was determined using a binomial test and the log2 allelic fold change (log2-aFC) was calculated as the log2-ratio between the expression of the minor allele compared to the major allele in heterozygotes and reported on a non-log scale (aFC)[29]. Associations between a group-representative SNP and FEV_1/FVC and FEV_1 % predicted values were tested using a linear model adjusted for sex (URECA) or sex and age (MUPPITS); all URECA participants were the same age (10 years) at the time of pulmonary function assessment.

SNP functional assessment:

Functional consequences of the identified significant SNPs were assessed by investigating amino acid changes using dbSNP[30], transcriptional regulatory consequences in ENCODE and ORegAnno using the UCSC genome browser interface[31–35], and predicted protein functional effects in Meta-SNP[36] and Polyphen-2[37].

Comparison to DNA genotyping data:

LD between the group A or group B representative SNP to previously published non-coding SNPs was performed by examining DNA-sequencing based genotypes of those published SNPs in a subset of 243 (76.4%) individuals in the URECA cohort. LD was computed between these SNPs as described above using the LDheatmap R package[28].

Results

MUC5AC SNPs associated with its expression in a high asthma risk birth cohort

The URECA cohort consists of urban children who are predominantly Black and Hispanic with high rates of poverty. By age 10 years, 29.5% of these children had a diagnosis of asthma (Table 1). 86 SNPs within the *MUC5AC* gene were identified from nasal brush RNA-sequencing data for eQTL analysis. 43 of these showed a significant eQTL association with *MUC5AC* expression (FDRs<0.05) (Figure 1a, Table 2). Clustering of SNPs by LD (r^2), as well as conditional eQTL analyses demonstrated that there were at least two groups of SNPs each independently associated with *MUC5AC* expression (Figure 1b, Supplementary Figure 1, Supplementary Table 1). Group A contained 10 SNPs; at each group A SNP, individuals homozygous for the major allele showed the lowest *MUC5AC* expression levels, with an increase according to number of minor alleles (eQTL β-coefficients=0.12 to 0.21). Representative of group A, rs1132436 (eQTL β-coefficient=0.20; FDR=−1.63E-4), had a minor allele frequency of 35.6% and explained 5.4% of variance of *MUC5AC* expression in this population (Figure 1c). ASE, assessed by RNA copy number in heterozygotes, confirmed significantly higher expression of the minor allele with an allelic fold change (aFC)=2.25 at rs1132436 (p<2E-16) (Figure 1d). Group B contained 29 SNPs and at each SNP the minor allele was associated with decreased *MUC5AC* expression (eQTL β-coefficients = −0.13 to −0.31). Representative of group B, rs1292198170 (β-coefficient=−0.28, FDR=1.9E-3), had a minor allele frequency of 13.0% and accounted for 3.1 % of variance of *MUC5AC* expression (Figure 1e). ASE confirmed lower expression of the minor allele with an aFC=0.42 (p<2E-16) (Figure 1f). rs1132436 in group A and rs1292198170 in group B showed an additive effect and together accounted for 7.9% of variance of *MUC5AC* expression (Figure 1g). The remaining 4 significant SNPs had low LD with the group A or B SNPs and with one another, and all had relatively low minor allele frequencies (<2.8%) so were not investigated further.

MUC5AC SNPs showed an inducible effect on its expression during asthma exacerbations

To validate these results and assess their relevance to asthma exacerbations, we conducted eQTL and ASE analyses in the MUPPITS cohort, a demographically similar population to URECA, but consisting entirely of children with established, exacerbation-prone asthma (Table 1)[15]. Participants had nasal lavage samples collected during respiratory illnesses, which were used for RNA-sequencing. 24 *MUC5AC* SNPs were detected in this dataset, predominantly near the 3’ end of the gene. Among the group A SNPs, 7 of 10 were detected, 5 of which were significantly associated with *MUC5AC* expression (FDR<0.05); the other two showed similar trends (FDRs=0.07 and 0.12) (Figure 2a). eQTL effect sizes were larger in this dataset (rs1132436, β-coefficient=0.57; FDR=8.9E-3) compared to those estimated in URECA (Table 2). Conditional eQTL analysis and LD assessment demonstrated similar co-dependence among group A SNPs in this population (Supplemental Figure 2, Figure 2b, Supplementary Table 2).

Because the MUPPITS samples were collected during respiratory illnesses, we could test the effects of asthma exacerbations on the observed eQTLs. The *MUC5AC* eQTL association at group A SNP rs1132436 was significantly greater during respiratory illnesses that led

to asthma exacerbations relative to uncomplicated upper respiratory illnesses (ΔFC=2.42, p=5.7E-6) (Figure 2c), which accounted for the overall larger effect sizes observed in this population. During exacerbations, rs1132436 explained 20.3% of variance in gene expression as compared to 1.0% of variance during limited upper respiratory illnesses. ASE confirmed significantly higher expression of the minor allele in heterozygotes (p=4.0E-5) (Figure 2d), and the aFC was higher during exacerbations (aFC=2.4) compared to limited upper respiratory illnesses (aFC=1.4).

Identification of group B SNPs, which are closer to the 5' end of the gene, was limited in the MUPPITS cohort because of library preparation method. We were able to identify genotypes at the group B SNP rs1292198170 in only 31.4% of the population. Likely due in part to the small sample size, neither this SNP nor other group B SNPs reached an eQTL FDR cutoff <0.05, but similar to URECA rs1292198170 showed a trend towards lower expression in heterozygotes (β-coefficient=−0.90, FDR=0.24). Despite the small sample size, we did observe a statistically significant inducible eQTL effect at this SNP during respiratory illnesses that led to asthma exacerbations relative to uncomplicated upper respiratory illnesses (ΔFC=0.27, p=3.9E-3) (Figure 2e). During exacerbations, rs1292198170 explained 5.2% of variance as compared to <1% of variance during limited upper respiratory illnesses. Similarly, ASE was significant in heterozygotes (p=8.0E-3) and the overall aFC was 0.32 (Figure 2f).

Together these two SNPs showed an additive inducible eQTL effect (Figure 2g) and explained 42.4% of variance during exacerbations as compared to 9.7% of variance during limited upper respiratory illnesses in the subset of the population with identifiable SNPs at both positions (30.4% of the population).

MUC5AC SNPs associated with pulmonary functions

We investigated the relationship of these SNPs with pulmonary function measures in each cohort. In the URECA cohort, the minor allele at rs1132436 (group A) was significantly associated with lower FEV_1/FVC values (β-coefficient−1.33, p=1.8E-2) (Figure 3a) and showed a non-significant trend towards lower FEV_1 % predicted (β-coefficient=−1.74, p=0.13). The minor allele at rs1292198170 (group B) showed a non-significant trend towards higher values of FEV_1/FVC (β-coefficient=1.45, p=0.10) and FEV_1 % predicted (β-coefficient=2.58, p=0.16). In the MUPPITS cohort, rs1132436 (group A) was also associated with lower lung function as measured by both FEV_1/FVC (β-coefficient=−1.64, p=4.0E-2) and FEV_1 % predicted (β-coefficient=−5.71, p=2.2E-2) (Figure 3b). rs1292198170 (group B) did not show a relationship with pulmonary functions in the MUPPITS cohort (p-values>0.75), perhaps due in part to the small sample size genotyped at this SNP.

Functional inference of *MUC5AC* eQTL SNPs

To investigate potential functional consequences of the identified SNPs, we examined nucleotide and amino acid changes in dbSNP, DNA regulatory elements in ENCODE and ORegAnno through the UCSC genome browser, and predicted mutational consequences in Meta-SNP and Polyphen-2 (Table 2)[30–37]. DNA binding and regulatory elements spanning

these SNPs revealed 4 Early Growth Response 1 (EGR1) transcription factor binding sites, 2 in group A and 2 in group B, as well as 3 CpG islands in group A. 16 of the significant SNPs lead to missense mutations, 4 in group A and 12 in group B. Predictions of functional consequences varied across tools with one missense mutation in block A and 3 in block B predicted as deleterious by at least one of the tools (Table 2).

LD of *MUC5AC* eQTL SNPs with prior disease associated SNPs

Using DNA genotype data from a subset of the URECA population (243 individuals, 76.4% of the population), we examined the LD between the SNPs identified in our study and non-coding *MUC5AC* SNPs previously associated with asthma or cystic fibrosis (CF) severity. rs11602802, which was previously associated with moderate-severe asthma in European ancestry[22], was not in LD with either group of SNPs (rs1132436 $r^2 = 0.08$, rs1292198170 $r^2 = 0.004$) in our sample. rs12788104, which was significantly associated with adult asthma in British white individuals[23], was in weak LD with group A but not group B SNPs (rs1132436 $r^2 = 0.37$, rs1292198170 $r^2 = 0.008$). rs28514396, a *MUC5AC* intronic variant associated with CF severity[38] was in LD with group A but not group B SNPs (rs1132436 $r^2 = 1$, rs1292198170 $r^2 = 0.09$).

Discussion

MUC5AC hypersecretion plays a pathogenic role in asthma through its relation to airway hyperresponsiveness and mucus plugging. In this study, we used RNA-sequencing data from airway samples to derive genotyping information and perform combined eQTL and ASE analyses focused on the *MUC5AC* region in two cohorts of predominantly African-American and Hispanic children using samples when children were well, and during incipient asthma exacerbations. We report two distinct groups of genetic polymorphisms in *MUC5AC* that independently relate to its expression, at least one of which is also significantly associated with decreased lung function. Moreover, we show that the effects of these SNPs on *MUC5AC* expression are increased during respiratory illnesses that result in asthma exacerbation, suggesting that inducible overexpression during exacerbations could contribute to exacerbation pathophysiology in genetically susceptible individuals. Functional assessment of these SNPs suggests potential causal effects, most notably they may alter EGR1 transcription factor binding.

Interestingly, despite long standing recognition of the role of mucus hypersecretion in asthma, GWAS studies have only recently linked specific *MUC5AC* polymorphisms, rs11602802[22] and rs12788104[23], to asthma. These studies were restricted to individuals of European ancestry and the observed associations were with moderate-to-severe and adult-onset asthma. This relatively recent discovery of *MUC5AC* as an asthma risk locus may be due in part to the specificity of this finding to moderate-severe asthma, but may also be due in part to incomplete mapping of the *MUC5AC* gene until 2014[39].

There was previously a large gap in the *MUC5AC* sequence including its large central exon, which spanned nearly all of the SNPs identified in the current study. These SNPs were missing from the data used in each of these prior GWAS studies. Since rs11602802 is not in LD with either group of our identified eQTL SNPs, it presumably has an independent effect

in relation to asthma risk or may be specific to European populations. rs12788104 was in weak LD with group A SNPs in our URECA population, so may represent an overlapping genetic risk to that identified in our study. However, since the URECA cohort includes children with and without asthma, our results suggest that the identified group A and B SNPs are *MUC5AC* eQTLs independent of a diagnosis of asthma. Hence these SNPs may be more relevant to the pathophysiologic endotype of asthma than the risk of developing asthma. Consistent with our results, two prior targeted gene association studies have linked SNPs in LD with group A with pulmonary diseases: the non-synonymous coding SNP, rs1132440 in group A was modestly associated with bronchitis, wheeze, and asthma in a British cohort[40], and the intronic SNP, rs28514396 (r^2=1 with rs1132436 in URECA) was associated with severity of lung disease in CF[38].

Prior work has shown EGR1 is an essential transcription factor for the expression of MUC5AC in airway epithelium[41, 42], and it is upregulated in airway epithelium in a model of respiratory virus infection[43]. These lead us to hypothesize that the *MUC5AC* eQTLs in known EGR1 binding sites disrupt its binding as the most likely mechanism for the eQTL effects. Alternatively, given the number of missense mutations among the identified *MUC5AC* SNPs, it is also plausible that these altered codons could affect gene expression, protein function or both[44] and may thereby contribute to the observed eQTLs and/or associations with pulmonary function. Interestingly, the group A SNPs have been shown to be in high LD with an insertion in the central exon encoding tandem repeats of an eight-amino-acid repeat rich in proline, threonine, and serine (PTS-TR domains)[38, 39]. These domains are characteristic of mucin proteins and are heavily glycosylated. The group A SNPs are thus indicative of a higher molecular weight and highly glycosylated protein that may play an important role in mucus plug formation[38, 39, 45], which may functionally contribute to the associations observed with pulmonary function.

An important limitation of our study is that it assessed links between genotype, mRNA expression in the upper airway, and pulmonary physiology. Linking these genetic polymorphisms to lower airway MUC5AC protein expression and/or lung imaging consistent with mucus plugging[16, 21] will help validate the pathogenic importance of these SNPs. Additionally, *in vitro* studies will be necessary to further investigate the functional effects of these SNPs to prove their mechanisms of effect. Our results suggest these as important next studies that can fully define the relevance of *MUC5AC* genotypes in the pathology of asthma, as well as potentially in other “muco-obstructive” lung diseases[46].

An important technical strength of our work is demonstration of the synergistic value of utilizing RNA-sequencing derived genotypes, organ specific gene expression, and sampling during disease exacerbation to gain novel insights into genetic contributions to a disease. These allowed for combined eQTL, ASE, and induction analyses to determine genetic factors contributing to a key pathway in asthma. We anticipate that the insights described in this study can lead to a more personalized understanding of asthma pathogenesis. Ideally, this will lead to strategies targeting mucus hypersecretion in genetically at-risk individuals as a way to reduce risk of disease inception and progression as well as risk of exacerbation[47].

In conclusion, our results add important new and detailed information regarding the role of *MUC5AC* genotypes in childhood asthma pathogenesis, demonstrating their direct effect on gene expression, relationships to pulmonary physiology, and identifying etiologies of their effects.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements:

We are grateful to all participants and their families who took part in these studies. We would like to thank P Woodruff, J Boyce, and S Durham for assistance with methodologic development as well advice and discussion.

We are grateful to all participants and their families who took part in these studies. We would like to thank P Woodruff, J Boyce, and S Durham for assistance with methodologic development as well advice and discussion.

This work in this study was funded by NIAID through the Inner City Asthma Consortium (UM1 AI11427).

This work in this study was funded by NIAID through the Inner City Asthma Consortium (UM1 AI11427).

All authors with the exception of P. Gergen, and A. Togias report grants from NIH/NIAID/DAIT during the conduct of study. M. Altman reports personal fees for consulting from Regeneron. M. Kattan reports consulting fees from Novartis. S. Lovinsky-Desir reports funding from NIH/NHLBI and the Robert Wood Johnson Foundation. G. O'Connor reports consulting fees from AstraZeneca, and reports a grant from Janssen Pharmaceuticals paid to his employing institution. M. Gill reports consulting fees from the American Academy of Allergy, Asthma, and Immunology and the American Academy of Pediatrics. R. Gruchalla reports employment as a special government employee with the Center for Biologies Evaluation and Research, and consulting fees from the Consulting Massachusetts Medical Society. A. Liu is a Consultant for Phadia ThermoFisher, received non-monetary research support from Phadia Thermofisher and Propeller Health/ResMed, and research funding from Avillion; all funds paid to University of Colorado. J. Pongracic reports provisions of study drug for other asthma studies from Glaxo SmithKline, Boehringer-lngelheim, and Genentech/Novartis. G. K. Khurana Hershey reports advisory board fees from Hoth Therapeutics. E. Zoratti reports consulting fees from Wayne State University. S. Teach reports support from NIH/NHLBI and EJF Philanthropies and royalties from UpToDate. L. Bacharier reports grant support from the NIH/NHLBI, Sanofi, and Vectura, as well as personal fees from GlaxoSmithKline, Genentech, Novartis, Merck, DBV Technologies, Teva, Boehringer Ingelheim, AstraZeneca, WebMD/Medscape, Sanofi, Regeneron, Vectura, and Circassia. W. Busse reports grants from NIH/NHLBI, during the conduct of the study; personal fees from Novartis, Regeneron, AstraZeneca, GlaxoSmithKline, Genentech, Elsevier, Med Learning Group, Boston Scientific, Medscape. J. Gern reports grants from NIH, personal fees and stock options from Meissa Vaccines Inc., personal fees from AstraZeneca and Ena Therapeutics and a patent on methods for production of rhinoviruses. C. Ober personal fees from the American Academy of Asthma, Allergy and Immunology. D. Jackson reports grants from NIH/NHLBI and GlaxoSmithKline, personal fees for DSMB from Pfizer and for consulting from Novartis, Sanofi-Regeneron, GlaxoSmithKline, Vifor Pharma and Astra Zeneca. K. Flynn, M.G. Rosasco, M. Dapas, P. LeBeau, P. Gergen, A. Togias, S. Presnell have nothing to disclose.

All authors with the exception of P. Gergen, and A. Togias report grants from NIH/NIAID/DAIT during the conduct of study. M. Altman reports personal fees for consulting from Regeneron. M. Kattan reports consulting fees from Novartis. S. Lovinsky-Desir reports funding from NIH/NHLBI and the Robert Wood Johnson Foundation. G. O'Connor reports consulting fees from AstraZeneca, and reports a grant from Janssen Pharmaceuticals paid to his employing institution. M. Gill reports consulting fees from the American Academy of Allergy, Asthma, and Immunology and the American Academy of Pediatrics. R. Gruchalla reports employment as a special government employee with the Center for Biologies Evaluation and Research, and consulting fees from the Consulting Massachusetts Medical Society. A. Liu is a Consultant for Phadia ThermoFisher, received non-monetary research support from Phadia Thermofisher and Propeller Health/ResMed, and research funding from Avillion; all funds

paid to University of Colorado. J. Pongracic reports provisions of study drug for other asthma studies from Glaxo SmithKline, Boehringer-lngelheim, and Genentech/Novartis. G. K. Khurana Hershey reports advisory board fees from Hoth Therapeutics. E. Zoratti reports consulting fees from Wayne State University. S. Teach reports support from NIH/NHLBI and EJF Philanthropies and royalties from UpToDate. L. Bacharier reports grant support from the NIH/NHLBI, Sanofi, and Vectura, as well as personal fees from GlaxoSmithKline, Genentech, Novartis, Merck, DBV Technologies, Teva, Boehringer Ingelheim, AstraZeneca, WebMD/Medscape, Sanofi, Regeneron, Vectura, and Circassia. W. Busse reports grants from NIH/NHLBI, during the conduct of the study; personal fees from Novartis, Regeneron, AstraZeneca, GlaxoSmithKline, Genentech, Elsevier, Med Learning Group, Boston Scientific, Medscape. J. Gern reports grants from NIH, personal fees and stock options from Meissa Vaccines Inc., personal fees from AstraZeneca and Ena Therapeutics and a patent on methods for production of rhinoviruses. C. Ober personal fees from the American Academy of Asthma, Allergy and Immunology. D. Jackson reports grants from NIH/NHLBI and GlaxoSmithKline, personal fees for DSMB from Pfizer and for consulting from Novartis, Sanofi-Regeneron, GlaxoSmithKline, Vifor Pharma and Astra Zeneca. K. Flynn, M.G. Rosasco, M. Dapas, P. LeBeau, P. Gergen, A. Togias, S. Presnell have nothing to disclose.

Abbreviations:

ASE	Allele Specific Expression
BH	Benjamini-Hochberg
BMI	Body Mass Index
CF	Cystic Fibrosis
eQTL	expression Quantitative Trait Locus
EGR1	Early growth response protein 1
FDR	False Discovery Rate
FEV1	Forced Expiratory Volume in one second
FVC	Forced Vital Capacity
GEO	Gene Expression Omnibus
GWAS	Genome Wide Association Study
ICS	Inhaled Corticosteroid
LD	Linkage Disequilibrium
SCS	Systemic Corticosteroid
SNP	Single Nucleotide Polymorphism

References

1. Duffy DL, Martin NG, Battistutta D, Hopper JL, Mathews JD. Genetics of Asthma and Hay Fever in Australian Twins. American Review of Respiratory Disease 1990; 142:1351–8.
2. van Beijsterveldt CEM, Boomsma DI. Genetics of parentally reported asthma, eczema and rhinitis in 5-yr-old twins. European Respiratory Journal 2007; 29:516–21.
3. Ferreira MA, Vonk JM, Baurecht H, Marenholz I, Tian C, Hoffman JD, et al. Shared genetic origin of asthma, hay fever and eczema elucidates allergic disease biology. Nat Genet 2017; 49:1752–7. [PubMed: 29083406]
4. Ober C Asthma Genetics in the Post-GWAS Era. Ann Am Thorac Soc 2016; 13 Suppl 1:S85–90. [PubMed: 27027959]

5. Moffatt MF, Gut IG, Demenais F, Strachan DP, Bouzigon E, Heath S, et al. A large-scale, consortium-based genomewide association study of asthma. N Engl J Med 2010; 363:1211–21. [PubMed: 20860503]
6. Moffatt MF, Kabesch M, Liang L, Dixon AL, Strachan D, Heath S, et al. Genetic variants regulating ORMDL3 expression contribute to the risk of childhood asthma. Nature 2007; 448:470–3. [PubMed: 17611496]
7. Torgerson DG, Ampleford EJ, Chiu GY, Gauderman WJ, Gignoux CR, Graves PE, et al. Meta-analysis of genome-wide association studies of asthma in ethnically diverse North American populations. Nat Genet 2011; 43:887–92. [PubMed: 21804549]
8. Demenais F, Margaritte-Jeannin P, Barnes KC, Cookson WOC, Altmuller J, Ang W, et al. Multiancestry association study identifies new asthma risk loci that colocalize with immune-cell enhancer marks. Nat Genet 2018; 50:42–53. [PubMed: 29273806]
9. Stein MM, Thompson EE, Schoettler N, Helling BA, Magnaye KM, Stanhope C, et al. A decade of research on the 17q12-21 asthma locus: Piecing together the puzzle. The Journal of allergy and clinical immunology 2018; 142:749–64.e3. [PubMed: 29307657]
10. Naqvi M, Tcheurekdjian H, DeBoard JA, Williams LK, Navarro D, Castro RA, et al. Inhaled corticosteroids and augmented bronchodilator responsiveness in Latino and African American asthmatic patients. Ann Allergy Asthma Immunol 2008; 100:551–7. [PubMed: 18592818]
11. Neophytou AM, White MJ, Oh SS, Thakur N, Galanter JM, Nishimura KK, et al. Air Pollution and Lung Function in Minority Youth with Asthma in the GALA II (Genes-Environments and Admixture in Latino Americans) and SAGE II (Study of African Americans, Asthma, Genes, and Environments) Studies. Am J Respir Crit Care Med 2016; 193:1271–80. [PubMed: 26734713]
12. Urquhart A, Clarke P. US racial/ethnic disparities in childhood asthma emergent health care use: National Health Interview Survey, 2013-2015. J Asthma 2019:1–11.
13. Zahran HS, Bailey CM, Damon SA, Garbe PL, Breysse PN. Vital Signs: Asthma in Children - United States, 2001-2016. MMWR Morb Mortal Wkly Rep 2018; 67:149–55. [PubMed: 29420459]
14. Altman MC, Calatroni A, Ramratnam S, Jackson DJ, Presnell S, Rosasco MG, et al. Endotype of Allergic Asthma with Airway Obstruction in Urban Children. J Allergy Clin Immunol 2021.
15. Altman MC, Gill MA, Whalen E, Babineau DC, Shao B, Liu AH, et al. Transcriptome networks identify mechanisms of viral and nonviral asthma exacerbations in children. Nat Immunol 2019; 20:637–51. [PubMed: 30962590]
16. Bonser LR, Zlock L, Finkbeiner W, Erle DJ. Epithelial tethering of MUC5AC-rich mucus impairs mucociliary transport in asthma. The Journal of clinical investigation 2016; 126:2367–71. [PubMed: 27183390]
17. Evans CM, Raclawska DS, Ttofali F, Liptzin DR, Fletcher AA, Harper DN, et al. The polymeric mucin Muc5ac is required for allergic airway hyperreactivity. Nature communications 2015; 6:6281-.
18. Messer JW, Peters GA, Bennett WA. Causes of Death and Pathologic Findings in 304 Cases of Bronchial Asthma. Diseases of the Chest 1960; 38:616–24. [PubMed: 13769788]
19. Aikawa T, Shimura S, Sasaki H, Ebina M, Takishima T. Marked Goblet Cell Hyperplasia with Mucus Accumulation in the Airways of Patients Who Died of Severe Acute Asthma Attack. Chest 1992; 101:916–21. [PubMed: 1555462]
20. Kuyper LM, Paré PD, Hogg JC, Lambert RK, lonescu D, Woods R, et al. Characterization of airway plugging in fatal asthma. The American Journal of Medicine 2003; 115:6–11. [PubMed: 12867228]
21. Dunican EM, Elicker BM, Gierada DS, Nagle SK, Schiebler ML, Newell JD, et al. Mucus plugs in patients with asthma linked to eosinophilia and airflow obstruction. The Journal of clinical investigation 2018; 128:997–1009. [PubMed: 29400693]
22. Shrine N, Portelli MA, John C, Soler Artigas M, Bennett N, Hall R, et al. Moderate-to-severe asthma in individuals of European ancestry: a genome-wide association study. Lancet Respir Med 2019; 7:20–34. [PubMed: 30552067]

23. 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. The Lancet. Respiratory medicine 2019; 7:509–22. [PubMed: 31036433]
24. Gern JE, Visness CM, Gergen PJ, Wood RA, Bloomberg GR, O'Connor GT, et al. The Urban Environment and Childhood Asthma (URECA) birth cohort study: design, methods, and study population. BMC.Pulm.Med. 2009; 9:17. [PubMed: 19426496]
25. Graffelman J Exploring Diallelic Genetic Markers: TheHardyWeinbergPackage. Journal of Statistical Software 2015; 64.
26. Liu R, Holik AZ, Su S, Jansz N, Chen K, Leong HS, et al. Why weight? Modelling sample and observational level variability improves power in RNA-seq analyses. Nucleic acids research 2015; 43:e97–e. [PubMed: 25925576]
27. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic acids research 2015; 43:e47–e. [PubMed: 25605792]
28. Shin J-H, Blay S, Graham J, McNeney B. LDheatmap: AnRFunction for Graphical Display of Pairwise Linkage Disequilibria Between Single Nucleotide Polymorphisms. Journal of Statistical Software 2006; 16.
29. Mohammadi P, Castel SE, Brown AA, Lappalainen T. Quantifying the regulatory effect size of cis-acting genetic variation using allelic fold change. Genome research 2017; 27:1872–84. [PubMed: 29021289]
30. Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM, et al. dbSNP: the NCBI database of genetic variation. Nucleic acids research 2001; 29:308–11. [PubMed: 11125122]
31. Haeussler M, Zweig AS, Tyner C, Speir ML, Rosenbloom KR, Raney BJ, et al. The UCSC Genome Browser database: 2019 update. Nucleic Acids Res 2019; 47:D853–D8. [PubMed: 30407534]
32. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, et al. The human genome browser at UCSC. Genome Res 2002; 12:996–1006. [PubMed: 12045153]
33. Raney BJ, Dreszer TR, Barber GP, Clawson H, Fujita PA, Wang T, et al. Track data hubs enable visualization of user-defined genome-wide annotations on the UCSC Genome Browser. Bioinformatics (Oxford, England) 2014; 30:1003–5.
34. Rosenbloom KR, Dreszer TR, Pheasant M, Barber GP, Meyer LR, Pohl A, et al. ENCODE whole-genome data in the UCSC Genome Browser. Nucleic acids research 2010; 38:D620–D5. [PubMed: 19920125]
35. Lesurf R, Cotto KC, Wang G, Griffith M, Kasaian K, Jones SJM, et al. ORegAnno 3.0: a community-driven resource for curated regulatory annotation. Nucleic acids research 2016; 44:D126–D32. [PubMed: 26578589]
36. Capriotti E, Altman RB, Bromberg Y. Collective judgment predicts disease-associated single nucleotide variants. BMC genomics 2013; 14 Suppl 3:S2–S.
37. Adzhubei I, Jordan DM, Sunyaev SR. Predicting functional effect of human missense mutations using PolyPhen-2. Current protocols in human genetics 2013; Chapter 7:Unit7.20-Unit7.
38. Guo X, Pace RG, Stonebraker JR, Commander CW, Dang AT, Drumm ML, et al. Mucin variable number tandem repeat polymorphisms and severity of cystic fibrosis lung disease: significant association with MUC5AC. PloS one 2011; 6:e25452–e. [PubMed: 21998660]
39. Guo X, Zheng S, Dang H, Pace RG, Stonebraker JR, Jones CD, et al. Genome reference and sequence variation in the large repetitive central exon of human MUC5AC. Am J Respir Cell Mol Biol 2014; 50:223–32. [PubMed: 24010879]
40. Johnson L, Shah I, Loh AX, Vinall LE, Teixeira AS, Rousseau K, et al. MUC5AC and inflammatory mediators associated with respiratory outcomes in the British 1946 birth cohort. Respirology (Carlton, Vic.) 2013; 18:1003–10.
41. Wang S-B, Zhang C, Xu X-C, Xu F, Zhou J-S, Wu Y-P, et al. Early growth response factor 1 is essential for cigarette smoke-induced MUC5AC expression in human bronchial epithelial cells. Biochemical and Biophysical Research Communications 2017; 490:147–54. [PubMed: 28602698]

42. Xu F, Cao J, Luo M, Che L, Li W, Ying S, et al. Early growth response gene 1 is essential for urban particulate matter-induced inflammation and mucus hyperproduction in airway epithelium. Toxicology Letters 2018; 294:145–55. [PubMed: 29787794]

43. Ramana CV, Cheng GS, Kumar A, Kwon HJ, Enelow RI. Role of alveolar epithelial early growth response-1 (Egr-1) in CD8+ T cell-mediated lung injury. Mol Immunol 2009; 47:623–31. [PubMed: 19786304]

44. Stergachis AB, Haugen E, Shafer A, Fu W, Vernot B, Reynolds A, et al. Exonic transcription factor binding directs codon choice and affects protein evolution. Science 2013; 342:1367–72. [PubMed: 24337295]

45. Lira-Navarrete E, de Las Rivas M, Compañón I, Pallarés MC, Kong Y, Iglesias-Fernández J, et al. Dynamic interplay between catalytic and lectin domains of GalNAc-transferases modulates protein O-glycosylation. Nature communications 2015; 6:6937-.

46. Boucher RC. Muco-Obstructive Lung Diseases. N Engl J Med 2019; 380:1941–53. [PubMed: 31091375]

47. Fahy JV, Dickey BF. Airway mucus function and dysfunction. The New England journal of medicine 2010; 363:2233–47. [PubMed: 21121836]

Key Messages:

- Multiple *MUC5AC* SNPs associated with its expression and two independent groups of SNPs showed additive eQTL effects.
- By examining *MUC5AC* expression during respiratory illnesses, we showed an inducible increase in expression during asthma exacerbations according to genotype.
- By investigating DNA regulatory elements, we identified possible mechanisms for the observed associations of these SNPs.

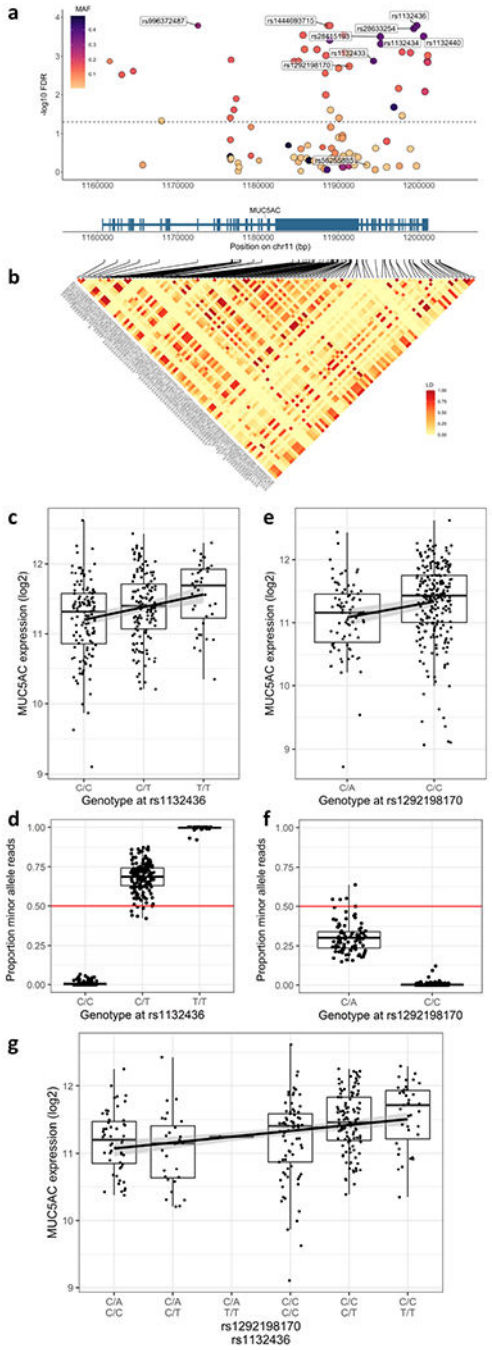

Figure 1.

MUC5AC SNPs showed an eQTL effect in the URECA cohort

a, The locus zoom plot for the URECA cohort showing eQTLs. Among the 86 cis SNPs detected, 43 showed significant association with *MUC5AC* expression (FDR< 0.05). Points are colored according to the minor allele frequency of the SNP and the point size reflects the percent of study subjects with a call at that SNP. **b**, An LD matrix shows the pairwise r^2 values among these 86 SNPs. **c**, For the block A representative SNP, rs1132436, *MUC5AC* gene expression was increased according to the number of minor alleles. **d**, ASE of

heterozygotes at rs1132436 showed an increased proportion of minor allele reads. **e**, For the block B representative SNP, rs1292198170, *MUC5AC* gene expression was decreased in heterozygotes. **f**, ASE of heterozygotes at rs1292198170 showed a decreased proportion of minor allele reads. **g**, Genotypes at rs1132436 and rs1292198170 showed an additive effect on *MUC5AC* expression. Boxplots show mean values, interquartile ranges, and 1.5 interquartile ranges.

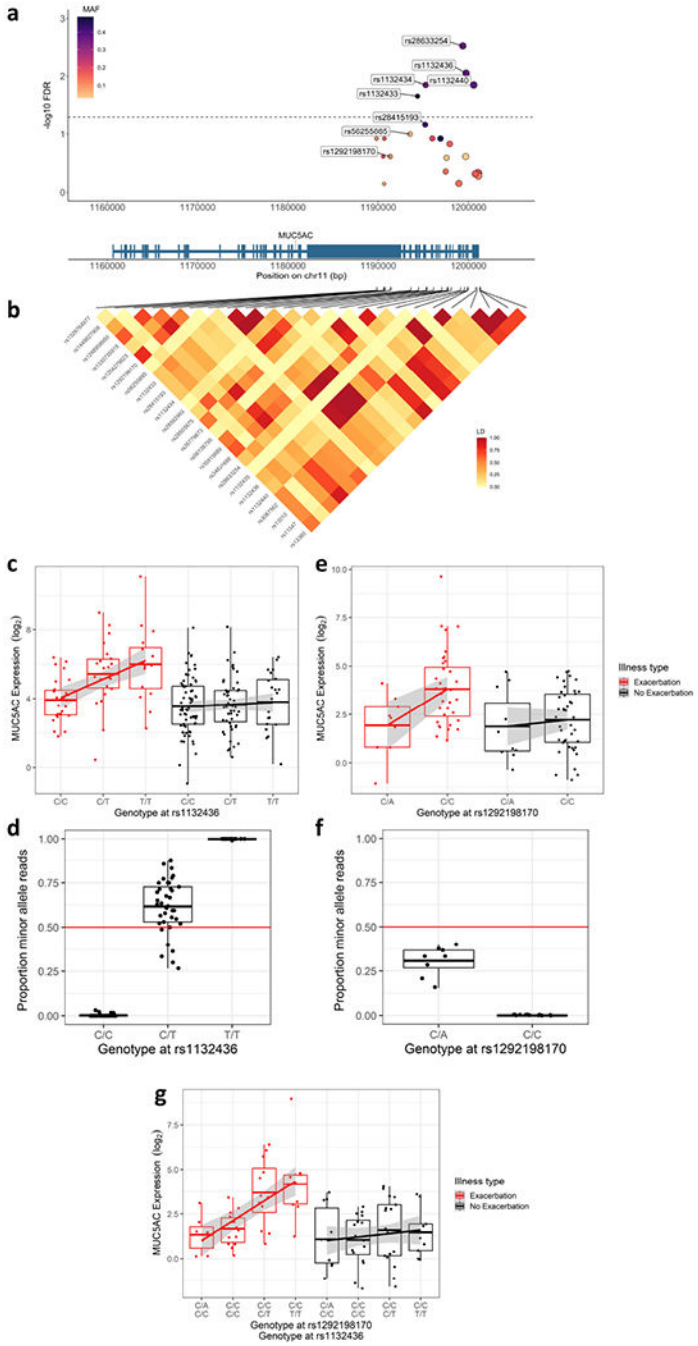

Figure 2.

MUC5AC SNPs showed an inducible eQTL effect during asthma exacerbations in the MUPPITS cohort

a, The locus zoom plot for the MUPPITS cohort showing eQTLs. Among the 24 cis SNPs detected, 5 showed significant association with *MUC5AC* expression (FDR<0.05). Points are colored according to the minor allele frequency of the SNP and the point size reflects the percent of study subjects with a call at that SNP. **b**, An LD matrix shows the pairwise r^2 values among these 24 SNPs. **c**, For the block A representative SNP rs1132436, *MUC5AC*

gene expression was increased according to the number of minor alleles and showed an inducible eQTL effect during exacerbations compared to non-exacerbation samples. **d**, ASE of heterozygotes at rs1132436 showed an increased proportion of minor allele reads. **e**, For the block B representative SNP, rs1292198170, *MUC5AC* gene expression showed a non-significant (FDR=0.24) decrease in heterozygotes relative to homozygotes and showed an inducible eQTL effect during exacerbations compared to non-exacerbation samples (p=3.9E-3). **f**, ASE of heterozygotes at rs1292198170 showed a decreased proportion of minor allele reads. **g**, Genotypes at rs1132436 and rs1292198170 showed an additive effect on *MUC5AC* expression. Boxplots show mean values, interquartile ranges, and 1.5 interquartile ranges.

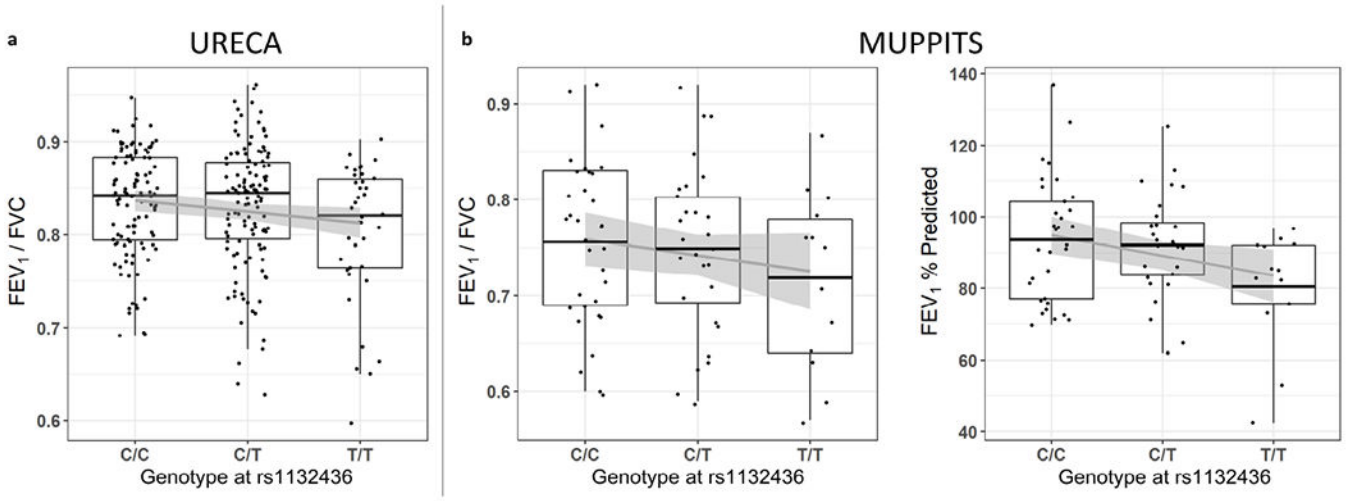

Figure 3.

MUC5AC SNP rs1132436 showed a significant association with pulmonary function in both URECA and MUPPITS.

a, The FEV_1/FVC ratio was significantly associated with the genotype at the block A representative SNP rs1132426 in the URECA cohort. **b**, Both the FEV_1/FVC ratio and FEV_1 % predicted were significantly associated with the genotype at the block A representative SNP rs1132426 in the MUPPITS cohort. Boxplots show mean values, interquartile ranges, and 1.5 interquartile ranges.

Table 1.

Demographics of URECA and MUPPITS cohorts.

Cohort population characteristics		
	URECA	**MUPPITS**
Sex		
Female	151 (47.5%)	52 (49%)
Male	167 (52.5%)	54 (51%)
Race		
African-American, non-Hispanic	233 (73.3%)	57 (53.8%)
Hispanic	59 (18.5%)	36 (33.9%)
Other, non-Hispanic	26 (8.2%)	13 (12.3%)
Age in years, mean	10 *	10.8 [6-17]
Diagnosis		
Asthma	94 (29.5%)	106 (100%)
No asthma	224 (70.5%)	-
ICS use	64 (20.1%)	106 (100%)
Pulmonary function		
FEV_1/FVC, mean	82.7% [59.7% - 98.2%]	76.0% [57.2-96.0%]
FEV_1 percent predicted	99.2% [62.2% - 139.1%]	91.7% [42.3-136.9%]
Gestational age in weeks	38.8 [34 - 42]	Not collected
BMI percentage	69.2% [0.8% - 99.8%]	78.2% [0.9% - 99.9%]
Number of allergen sensitizations	3.2 [0 - 13]	4.7 [0 - 15]
City of residence		
Baltimore	92 (28.9%)	-
Boston	74 (23.2%)	14 (13.2%)
Chicago	-	7 (6.6%)
Cincinnati	-	7 (6.6%)
Dallas	-	8 (7.5%)
Denver	-	13 (12.3%)
Detroit	-	14 (13.2%)
New York	57 (17.9%)	23 (21.7%)
St. Louis	95 (32.6%)	13 (12.3%)
Washington DC	-	6 (5.7%)

Shown are the population characteristics of the URECA and MUPPITS cohorts. Shown are counts of mean values. () indicate percentage of the population. [ ] indicate ranges of values in the population.

* URECA children were all age 10 at the time of pulmonary function assessment and asthma diagnosis, and were age 11 at time of nasal brush collection. BMI is body mass index; ICS is inhaled corticosteroid.

ble 2.

				MUPPITS cohort results						SNP details						
LD R2 vs rs1292198170	**FDR, rs1132436 conditioned**	**LD R2 vs rs1132436**	**FDR, dual conditioned rs1132436 + rs1292198170**	**% subjs with call**	**Minor Allele Freq**	**FDR**	**Effect size**	**FDR, rs1132436 conditioned**	**LD R2 vs rs1132436**	**Base chng**	**Codon change**	**Mutational consequence**	**Amino aced change**	**Meta-SNP prediction**	**Polyphen-2**	**DNA elements in region**
0.086	NA	NA	NA	71.7%	36.8%	8.89E-03	0.574	NA	NA	C>T	[CCG]>[CTG]	Missense	P5521L	Neutral	Neutral	none
0.082	2.26E-01	0.986	8.10E-01	70.8%	38.2%	1.42E-02	0.549	NA	1.000	G>C	[CGG]>[CGC]	Synonymous				CpG island CpG database
0.089	5.66E-01	0.972	8.99E-01	69.8%	36.5%	3.04E-03	0.665	4.95E-01	0.977	C>T	[CCC]>[CCT]	Synonymous				EGR1 TF bindingsite, ORegAnno
0.098	9.08E-01	0.972	8.99E-01	50.0%	34.9%	1.42E-02	0.636	1.00E+00	1.000	G>C	[CAG]>[CAC]	Missense	Q5148H	Neutral	Neutral	none
0.096	8.28E-01	0.964	9.65E-01	-	-	-	-	-	-	G>A	[CGC]>[CAC]	Missense	R3590H	Neutral	Deleterious	none
0.081	3.81E-01	0.950	8.10E-01	43.4%	41.3%	6.77E-02	0.509	1.00E+00	1.000	G>A	[ACG]>[ACA]	Synonymous				CpG island, CpG database
0.083	2.64E-01	0.938	8.10E-01	31.1%	48.5%	2.21E-02	0.680	4.95E-01	0.958	G>A	[ACG]>[ACA]	Synonymous				CpG island, EGR1 TF binding site
0.082	7.45E-01	0.894	8.99E-01	-	-	-	-	-	-	C>G, T	[CCC]>[GCC], [CCC]>[TCC]	Missense	P1480A, P1480S	Neutral, Neutral	No prediction	none
0.112	7.96E-01	0.690	9.65E-01	51.9%	46.4%	1.20E-01	0.351	5.50E-01	0.623	A>G	[CCA]>[CCG]	Synonymous				none
0.058	1.39E-01	0.494	8.10E-01	-	-	-	-	-	-	C>A	[GGC]>[GGA]	Synonymous				none
NA	1.30E-02	0.086	NA	31.1%	9.1%	2.45E-01	−0.905	4.95E-01	0.186	C>A	[GCC]>[GCA]	Synonymous				none
0.888	8.96E-03	0.109	9.55E-01	-	-	-	-	-	-	T>A	[TCT]>[TCA]	Synonymous				none
0.879	6.46E-02	0.115	8.10E-01	-	-	-	-	-	-	C>T	[ACC]>[ACT]	Synonymous				none
0.843	8.44E-03	0.095	8.10E-01	-	-	-	-	-	-	C>A	[ACC]>[AAC]	Missense	T2541N	Neutral	No prediction	none
0.830	8.96E-03	0.111	8.25E-01	18.9%	7.5%	7.16E-01	−0.319	4.95E-01	0.242	C>T	[ACA]>[ATA]	Missense	T4174I	Neutral	No prediction	none

				MUPPITS cohort results						SNP details						
LD R2 vs rs1292198170	**FDR, rs1132436 conditioned**	**LD R2 vs rs1132436**	**FDR, dual conditioned rs1132436 + rs1292198170**	**% subjs with call**	**Minor Allele Freq**	**FDR**	**Effect size**	**FDR, rs1132436 conditioned**	**LD R2 vs rs1132436**	**Base chng**	**Codon change**	**Mutational consequence**	**Amino aced change**	**Meta-SNP prediction**	**Polyphe n-2**	**DNA elements in region**
0.827	8.96E-03	0.114	8.99E-01	-	-	-	-	-	-	A>G	[AGC]>[GGC]	Missense	S3059G	Neutral	No prediction	none
0.823	1.24E-02	0.110	8.25E-01	22.6%	16.7%	1.20E-01	−0.845	4.95E-01	0.369	C>T	[GCC]>[GTC]	Missense	A4180V	Neutral	No prediction	none
0.823	8.96E-03	0.116	8.99E-01	-	-	-	-	-	-	C>T	[CCT]>[CTT]	Missense	P3434L	Neutral	No prediction	none
0.820	8.96E-03	0.118	8.25E-01	-	-	-	-	-	-	A>G	[AAC]>[GAC]	Missense	N2115D	Neutral	Neutral	none
0.819	7.03E-03	0.116	8.10E-01	-	-	-	-	-	-	G>A	[CGG]>[CAG]	Missense	R3597Q	Neutral	Deleterious	none
0.816	7.03E-03	0.126	8.25E-01	-	-	-	-	-	-	C>T	[GCT]>[GTT]	Synonymous				none
0.809	6.16E-02	0.115	8.44E-01	-	-	-	-	-	-	G>C	[CGT]>[CCT]	Missense	R3426P	Neutral	No prediction	none
0.750	9.57E-02	0.096	8.10E-01	45.3%	13.5%	4.41E-01	−0.351	4.95E-01	0.221	C>T	[GTC]>[GTT]	Synonymous				none
0.744	7.03E-03	0.144	8.99E-01	-	-	-	-	-	-	G>A	[AGC]>[AAC]	Missense	S3503N	Neutral	No prediction	none
0.732	1.30E-02	0.105	8.99E-01	-	-	-	-	-	-	C>A	[ACC]>[ACA]	Synonymous				none
0.728	1.80E-02	0.133	9.92E-01	52.8%	16.1%	1.49E-01	-0.517	4.95E-01	0.247	C>G	[CCC]>[CCG]	Synonymous				none
0.702	4.14E-02	0.119	8.31E-01	63.2%	14.2%	7.05E-01	−0.113	4.95E-01	0.170	C>T	[GAC]>[GAT]	Synonymous				EGR1 TF binding site, ORegAnno
0.691	6.16E-02	0.094	8.99E-01	-	-	-	-	-	-	C>G	[AGC]>[AGG]	Missense	S221R	Disease	No prediction	EGR1 TF binding site, ORegAnno
0.688	8.91E-02	0.119	8.99E-01	-	-	-	-	-	-	C>T	[TGC]>[TGT]	Synonymous				none
0.672	1.20E-02	0.108	8.44E-01	73.6%	11.5%	5.26E-01	−0.189	4.95E-01	0.177	A>C		3 Prime UTR Variant				none
0.628	4.09E-02	0.151	9.87E-01	-	-	-	-	-	-	C>T	[GCC]>[GCT]	Synonymous				none
0.627	2.26E-01	0.109	8.99E-01	-	-	-	-	-	-	C>T	[TGC]>[TGT]	Synonymous				none
0.627	1.39E-01	0.126	8.99E-01	-	-	-	-	-	-	T>C	[TTT]>[TTC]	Synonymous				none
0.618	2.21E-01	0.161	8.25E-01	-	-	-	-	-	-	C>T	[TGC]>[TGT]	Synonymous				none
0.536	1.39E-01	0.150	8.99E-01	84.9%	17.8%	4.67E-01	−0.191	6.08E-01	0.224	A>C		3 Prime UTR Variant				none
0.518	9.57E-02	0.153	8.99E-01	70.8%	16.7%	4.83E-01	−0.200	4.95E-01	0.236	C>T		3 Prime UTR Variant				none

				MUPPITS cohort results						SNP details						
LD R2 vs rs1292198170	**FDR, rs1132436 conditioned**	**LD R2 vs rs1132436**	**FDR, dual conditioned rs1132436 + rs1292198170**	**% subjs with call**	**Minor Allele Freq**	**FDR**	**Effect size**	**FDR, rs1132436 conditioned**	**LD R2 vs rs1132436**	**Base chng**	**Codon change**	**Mutational consequence**	**Amino aced change**	**Meta-SNP prediction**	**Polyphen-2**	**DNA elements in region**
0.487	1.43E-01	0.155	8.99E-01	77.4%	15.2%	4.69E-01	−0.206	4.95E-01	0.228	C>T		3 Prime UTR Variant				none
0.405	3.05E-02	0.064	8.10E-01	-	-	-	-	-	-	G>T	[GTC]>[TTC]	Missense	V52F	Neutral	Deleterious	none
0.317	2.47E-01	0.088	9.65E-01	-	-	-	-	-	-	C>A, T	[CAC]>[CAA], [CAC]>[CAT]	Missense, Synonymous	H3973Q	Neutral	No prediction	none
0.141	3.00E-02	0.019	8.10E-01	19.8%	9.5%	2.45E-01	−1.147	5.50E-01	0.203	C>T	[TCT]>[TTT]	Missense	S4387F	Neutral	Deleterious	none
0.009	1.48E-01	0.027	8.10E-01	-	-	-	-	-	-	C>A	[GCG]>[GAG]	Missense	A5353E	Neutral	Deleterious	none
0.009	1.48E-01	0.027	8.10E-01	-	-	-	-	-	-	G>A	[GCG]>[ACG]	Missense	A5353T	Neutral	Neutral	none
0.009	1.86E-01	0.027	8.10E-01	-	-	-	-	-	-	C>T	[GCG]>[GTG]	Missense	A497V	Neutral	Neutral	none

, A and B, based on LD and conditional analyses. For each SNP we list the chromosomal position,
and FDR values are shown, as well as the FDR after each conditional analysis and the r^2 with the SNP
hort with blanks indicating SNPs that could not be called in the MUPPITS cohort. The base and amino
onal predictions of amino acid changes.