



Early View

Research letter

SNPs identified by GWAS affect asthma risk through DNA methylation and expression of *cis*-genes in airway epithelium

Soyeon Kim, Erick Forno, Qi Yan, Yale Jiang, Rong Zhang, Nadia Boutaoui, Edna Acosta-Pérez, Glorisa Canino, Wei Chen, Juan C. Celedón

Please cite this article as: Kim S, Forno E, Yan Q, *et al.* SNPs identified by GWAS affect asthma risk through DNA methylation and expression of *cis*-genes in airway epithelium. *Eur Respir J* 2019; in press (<https://doi.org/10.1183/13993003.02079-2019>).

This manuscript has recently been accepted for publication in the *European Respiratory Journal*. It is published here in its accepted form prior to copyediting and typesetting by our production team. After these production processes are complete and the authors have approved the resulting proofs, the article will move to the latest issue of the ERJ online.

SNPs identified by GWAS affect asthma risk through DNA methylation and expression of *cis*-genes in airway epithelium

Soyeon Kim, PhD^{1,2*}, Erick Forno, MD, MPH^{1,2*}, Qi Yan, PhD^{1,2}, Yale Jiang^{1,3}, Rong Zhang⁴,
Nadia Boutaoui, PhD^{1,2}, Edna Acosta-Pérez, PhD⁵, Glorisa Canino, PhD⁵, Wei Chen, PhD^{1,2^},
Juan C. Celedón^{1,2}, MD, DrPH[^]

¹ Division of Pediatric Pulmonary Medicine, UPMC Children's Hospital of Pittsburgh,

University of Pittsburgh, Pittsburgh, PA, USA

² Department of Pediatrics, School of Medicine, University of Pittsburgh, Pittsburgh, PA, USA

³ School of Medicine, Tsinghua University, Beijing, China

⁴ Department of Statistics, University of Pittsburgh, Pittsburgh, PA, USA

⁵ Behavioral Sciences Research Institute, University of Puerto Rico, San Juan, PR

**Shared first and ^shared senior authors.*

Corresponding Author: Juan C. Celedón, MD, DrPH

Division of Pediatric Pulmonary Medicine

UPMC Children's Hospital of Pittsburgh

4401 Penn Avenue, Pittsburgh, PA 15224

Phone: 412.692.8429; Fax: 412.692.7636

Email: juan.celedon@chp.edu

Author contributions: W.C. and J.C.C. conceived and designed the study. S.K. and E.F. conducted the primary analysis and interpreted data. Q.Y., Y.J., R.Z., N.B., E.A-P., and G.C. participated in data collection and data analysis. S.K., E.F., W.C. and J.C.C. prepared the first draft of the manuscript. All authors reviewed the draft for intellectual content, and approved submission of the final version of the manuscript.

Funding/Support: This study was supported by grants HL079966, HL117191, and MD011764 (PI: Celedón JC) from the U.S. National Institutes of Health (NIH). Dr. Kim is supported by a T32 training grant (HL129949) from the U.S. NIH. Dr. Forno's contribution was supported by NIH grant HL125666.

Running head: GWAS SNPs, methylation and expression in nasal epithelium, and childhood asthma

Subject Category Descriptor Number: 1.18 Asthma (Genetics), 2.09 Racial, Ethnic, or Social Disparities in Lung Disease and Treatment

Key words: SNPs, expression, methylation, airway epithelium, asthma

SUMMARY

The Trans-National Asthma Genetic Consortium (TAGC) identified 878 SNPs associated with asthma. We hypothesized that those SNPs affect asthma risk by regulating gene expression in airway epithelium, and conducted expression quantitative trait loci (eQTL) and mediation analyses to identify direct associations between the SNPs and expression levels of *cis*-genes (within 1 Mb) in nasal (airway) epithelium from Puerto Rican children with (n=228) and without (n=241) asthma. We then tested whether genes whose expression is associated with TAGC SNPs are differentially expressed (DE) in atopic asthma. We identified 1,150 direct associations between 418 TAGC SNPs and the expression of 55 *cis*-genes. Most SNPs regulate distant *cis*-genes (average distance ~200 kb). Our mediation analysis showed that 4,571 (89.2%) of 5,119 (direct and indirect) SNP-gene expression associations are mediated by methylation. Of 114 genes whose expression is associated with TAGC SNPs, 54 are DE in atopic asthma, including novel and previously reported genes. In an independent cohort of 72 African American children, 50 of the 54 DE genes were available, and 21 (42%) were also DE in atopic asthma. Thus, we show that many TAGC SNPs are associated with expression of distant *cis*-genes in airway epithelium, and that this is predominantly mediated by DNA methylation. Moreover, nearly half of the genes whose expression in airway epithelium is associated with TAGC SNPs are also DE in atopic asthma. Our findings support a key role of regulation of airway epithelial gene expression on atopic asthma in children.

Over the last twelve years, many genome-wide association studies (GWAS) have been conducted to identify susceptibility variants for asthma [1-5]. The Trans-National Asthma Genetic Consortium (TAGC) recently conducted a large meta-analysis of GWAS [3], identifying 878 SNPs associated with asthma at 18 loci. The mechanisms underlying such associations are largely unknown but may include effects on gene expression that alter airway epithelial integrity and function.

Studying bronchial airway epithelium, while ideal, requires a bronchoscopy. Both DNA methylation and gene expression in nasal epithelium are well correlated with those in bronchial epithelium, and thus sampling nasal epithelial cells provides a safer and more feasible alternative to sampling the bronchial epithelium for studies of airway epithelial “omics” and asthma [6].

We hypothesized that SNPs identified in the TAGC meta-analysis of asthma (heretofore called “TAGC SNPs”) affect gene expression in airway epithelium, and that DNA methylation would mediate some of the effects of TAGC SNPs on gene expression. We tested these hypotheses using genomic, DNA methylation, and RNA-sequencing data in nasal (airway) epithelium from Puerto Rican children and adolescents – a group disproportionately affected with asthma [7].

The Epigenetic Variation and Childhood Asthma in Puerto Ricans study (EVA-PR) is a case-control study of asthma in Puerto Ricans aged 9 to 20 years [7], in whom we assessed genome-wide genotypes (using the HumanOmni2.5 BeadChip platform, Illumina, San Diego, CA), and genome-wide DNA methylation (using the HumanMethylation450 BeadChip, Illumina) and RNA sequencing (using the Illumina NextSeq 500 platform) in nasal epithelium, as previously

described [7]. SNPs and RNA-seq data were available for our eQTL analysis in 469 subjects with (n=228) and without (n=249) asthma. We analyzed *cis*-eQTL due to limited power for a *trans*-eQTL analysis. After quality control, we tested 869 SNPs and 471 *cis* genes (whose TSS were located <1 Mb from the SNP), comprising 30,855 SNPs-gene pairs. We identified 1,150 SNP-gene expression associations for 418 SNPs and 55 genes located at nine loci (FDR-P <0.05). We denote these SNPs as “eQTL SNPs” and these genes as “eQTL genes”.

Of 46 genes reported in the TAGC meta-analysis of GWAS as “nearby” (defined as the gene in which the SNP is located and the previous and next genes; for intergenic SNPs, as the previous and next genes) their 18 lead SNPs[3], only five had expression levels in nasal epithelium that were significantly associated with the lead SNPs in our analysis. Moreover, of the 188 asthma-susceptibility SNPs ($P < 10^{-8}$) in the GWAS catalog [8], only 9 were associated with expression of nearby genes in our eQTL analysis.

We then assessed the position of each eQTL SNP relative to their paired gene. If a SNP was associated with expression of multiple genes, it was considered for each of the associated genes. Only 31 of 1,150 SNPs (2.7%) were located within a gene (**Figure 1A**). The average distance between eQTL SNPs and the TSS of their paired genes was 232,450 bp (**Figure 1B**), suggesting that most of the SNPs of interest regulate expression of genes that are in *cis* but not “nearby”.

Next, we evaluated whether eQTL SNPs are located in enhancer regions, known to regulate expression of distant genes. We used enhancers and their target genes in lung tissue from <http://www.enhanceratlas.org>, finding that five of our 418 eQTL SNPs were located in enhancer

regions of their target genes. Interestingly, all of those target genes were identified as asthma-susceptibility genes in prior GWAS, such as *HLA* genes [9], *GSDMB* [4] *MED24* [10], *STARD3* [10], and *SLC22A5* [5]. Our findings, together with those from prior studies, suggest that SNPs can affect asthma by regulating expression of distant disease-susceptibility genes, possibly by regulating enhancer activity.

We analyzed RNA-sequencing data from 157 subjects with atopic asthma (the most common type of childhood asthma) and 101 non-atopic non-asthmatic controls, finding that 24 of the 55 eQTL genes in nasal epithelium were differentially expressed (DE) in atopic asthma (genome-wide FDR-P < 0.05). Such DE genes include ten genes previously reported in GWAS of asthma (*HLA-DRB1*, *HLA-DQA1*, *ERBB2*, *PGAP3*, *IL33*, *NDFIP1*, *MICB*, *DEXI* [2], *HLA-DQ*, and *SLC22A5* [5]) and 14 genes not previously reported, likely due to the distance between each SNP and the reported asthma-susceptibility gene.

Given the degree of association between TAGC SNPs and DNA methylation in nasal epithelium, we tested whether SNPs regulate gene expression through methylation. After conducting the Sobel test¹⁶ of mediation effects, we identified 12,035 trios with significant mediation effects (FDR-P < 0.05), comprising 4,571 SNP-gene expression pairs. Of the 418 eQTL SNPs, 366 (88%) were associated with gene expression through methylation (**Figure 1C**). Moreover, we identified 91 additional TAGC SNPs that were associated with gene expression solely through indirect effects of methylation. Thus, 509 (59%) of the 869 TAGC SNPs regulated *cis*-gene expression in nasal epithelium (directly or indirectly), and most (89.8%) of these SNPs regulated expression through DNA methylation (**Figure 1D**).

Of 114 genes whose expression was associated with SNPs through methylation, 54 (47.4%) were also associated with atopic asthma (genome-wide FDR $P < 0.05$). As an example, data for *HLA-DQA2* are shown in **Figure 1E**. Through the mediation analysis, we identified 30 additional DE genes in atopic asthma, some of which have been previously reported in GWAS (*IKZF3* [1, 11], *PSORS1C3* [11], *ORMDL3* [12], and *IL1RL1* [3]). We also identified novel genes: while *C3* and *C5* have been associated with allergic asthma [13], *C2* has not; and *SHROOM1* has been associated with sarcoidosis in children [14], but not with asthma.

Through eQTL and mediation analysis, we identified 64 DE genes that were directly and/or indirectly associated with TAGC SNPs. We tested whether genes that were DE in atopic asthma in EVA-PR were also differentially expressed in a cohort of 72 African American children with and without atopic asthma, using publicly available data (GSE65205)[15]. Data were available for 50 of the 64 DE genes of interest in GSE65205: despite small sample size, 21 (42%) of the 50 genes were also DE in atopic asthma in African Americans at $P < 0.05$, mostly in the same direction of association as in EVA-PR.

To summarize, the vast majority of the 418 eQTL SNPs were located outside genes, and most were associated with expression of distant *cis*-genes. Of the 55 eQTL genes, 24 (~44%) were DE genes in atopic asthma. Of all 5,119 SNP-gene expression associations in EVA-PR, 4,571 (86.8%) were mediated by methylation. We cannot infer temporal relationships in this cross-sectional study, and thus some methylation signals could be disease-induced (caused by atopic asthma). However, our findings suggest that some methylation changes in airway epithelium (due to genetic variation and environmental factors) could affect the development or progression

of asthma by regulating gene expression.

The variance in asthma liability explained by 878 TAGC SNPs was estimated to be 3.5% ². However, the estimated effects of the TAGC SNPs would be much larger if both direct and indirect associations with gene expression in airway epithelium were considered.

In conclusion, a high proportion of asthma-susceptibility SNPs identified in a meta-analysis of GWAS were associated with *cis*-expression and *cis*-methylation of genes in nasal (airway) epithelium, but most genes whose expression was affected were not nearby the corresponding SNPs of interest. Most SNP-gene expression associations were mediated by DNA methylation, supporting a strong regulatory role of methylation on gene expression in airway epithelium. Consistent with a key role of airway epithelium in the pathogenesis of childhood asthma, nearly half of the genes whose expression in airway epithelium was associated with TAGC SNPs were also DE genes in atopic asthma.

FIGURE LEGENDS

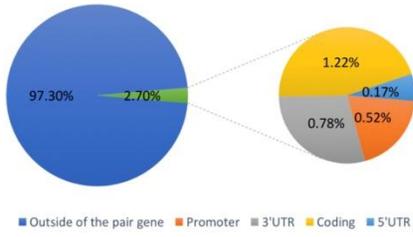
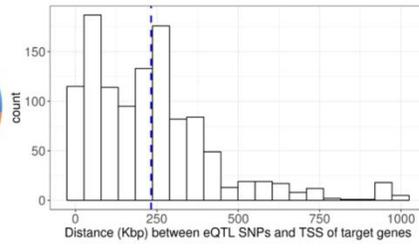
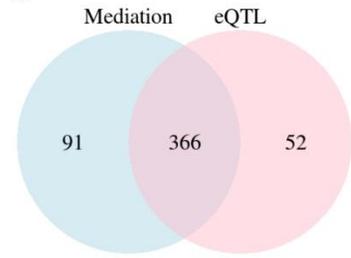
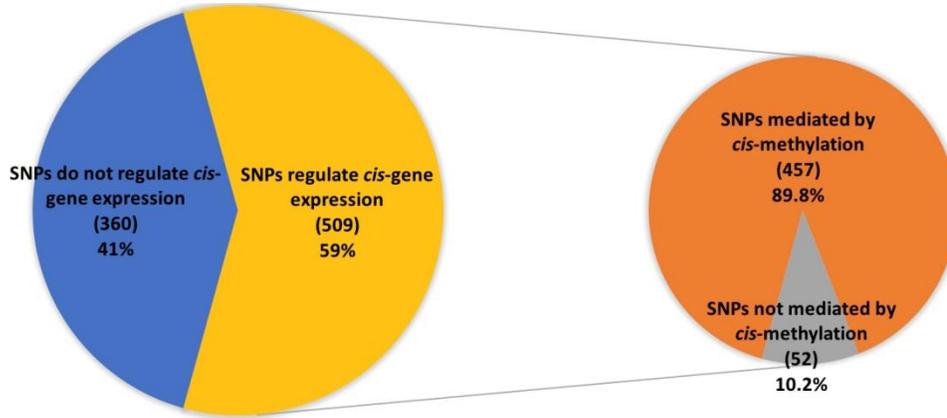
Figure 1. eQTL and mediation analysis of 869 TAGC SNPs

(A) Percentage of eQTL SNPs that are located outside (blue) or inside (orange) of their paired gene region (Promoter – 5'UTR). When SNPs were associated with expression of multiple genes, they were counted multiple times. (B) Histogram of distance between eQTL SNPs and transcription start site of their associated genes (in Kb). (C) Venn diagram of the number of SNPs associated with gene expression (eQTL) (pink) and the number of SNPs associated with gene expression through mediation by methylation (blue). To test mediation effects of methylation between SNPs and gene expression, the Sobel test was conducted (at FDR-P < 0.05). (D) Composition of TAGC GWAS SNPs according to whether they are associated with *cis*-gene expression (left) and composition of the SNPs that regulate *cis*-gene expression, according to whether they are directly associated with gene expression (eQTL only) or indirectly associated with gene expression through mediation effects of methylation. (E) An example of a SNP that is indirectly associated with *HLA-DQA2* gene expression through DNA methylation, and the gene is differentially expressed in atopic asthma. 1. Violin plots of the relationship between the eQTL SNP and an associated methylation probe (red, homozygous major allele; green, heterozygous; blue, homozygous minor allele) 2. Scatter plots of the relationship between DNA methylation and expression of the gene. 3. Violin plots of data from atopic asthmatic vs. those from non-atopic non-asthmatic controls against differential gene expression.

REFERENCES

1. Shrine, N., et al., *Moderate-to-severe asthma in individuals of European ancestry: a genome-wide association study*. The Lancet Respiratory Medicine, 2019. **7**(1): p. 20-34.
2. Wyss, A.B., et al., *Multiethnic meta-analysis identifies ancestry-specific and cross-ancestry loci for pulmonary function*. Nature Communications, 2018. **9**(1): p. 2976.
3. Demenais, F., et al., *Multiancestry association study identifies new asthma risk loci that colocalize with immune-cell enhancer marks*. Nature Genetics, 2018. **50**(1): p. 42-53.
4. Yan, Q., et al., *A meta-analysis of genome-wide association studies of asthma in Puerto Ricans*. European Respiratory Journal, 2017. **49**(5).
5. Moffatt, M.F., et al., *A large-scale, consortium-based genomewide association study of asthma*. The New England journal of medicine, 2010. **363**(13): p. 1211-1221.
6. Brugha, R., et al., *DNA methylation profiles between airway epithelium and proxy tissues in children*. Acta Paediatrica, 2017. **106**(12): p. 2011-2016.
7. Forno, E., et al., *DNA methylation in nasal epithelium, atopy, and atopic asthma in children: a genome-wide study*. The Lancet Respiratory Medicine, 2019. **7**(4): p. 336-346.
8. McMahon, A., et al., *The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019*. Nucleic Acids Research, 2018. **47**(D1): p. D1005-D1012.
9. Movahedi, M., et al., *Association of HLA class II alleles with childhood asthma and Total IgE levels*. Iranian journal of allergy, asthma, and immunology, 2008. **7**(4): p. 215-220.
10. Ried, J.S., et al., *Integrative genetic and metabolite profiling analysis suggests altered phosphatidylcholine metabolism in asthma*. Allergy, 2013. **68**(5): p. 629-636.

11. Galanter, J.M., et al., *Genome-wide association study and admixture mapping identify different asthma-associated loci in Latinos: The Genes-environments & Admixture in Latino Americans study*. Journal of Allergy and Clinical Immunology, 2014. **134**(2): p. 295-305.
12. Miller, M. and D.H. Broide, *Why Is ORMDL3 on Chromosome 17q21 Highly Linked to Asthma?* American Journal of Respiratory and Critical Care Medicine, 2018. **199**(4): p. 404-406.
13. Zhang, X. and J. Köhl, *A complex role for complement in allergic asthma*. Expert review of clinical immunology, 2010. **6**(2): p. 269-277.
14. Calender, A., et al., *Whole exome sequencing in three families segregating a pediatric case of sarcoidosis*. BMC medical genomics, 2018. **11**(1): p. 23-23.
15. Yang, I.V., et al., *The nasal methylome and childhood atopic asthma*. Journal of Allergy and Clinical Immunology, 2017. **139**(5): p. 1478-1488.

A**B****C****D****E**