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

To the Editor:

Over the past 12 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 single nucleotide polymorphisms (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 hypothesised 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 9–20 years [7], in whom we assessed genome-wide genotypes (using the HumanOmni2.5 BeadChip platform, Illumina, San Diego, CA, USA), 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 expression quantitative trait loci (eQTL) analysis in 477 subjects with (n=228) and without (n=249) asthma. We analysed *cis*-eQTL due to limited power for a *trans*-eQTL analysis. After quality control, we tested 869 SNPs and 471 *cis* genes (whose transcription start sites (TSS) were located <1 Mb from the SNP), comprising 30 855 SNPs-gene pairs. We identified 1150 SNP-gene expression associations for 418 SNPs and 55 genes located at nine loci (false discovery rate (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 catalogue [8], only nine 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 1150 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".
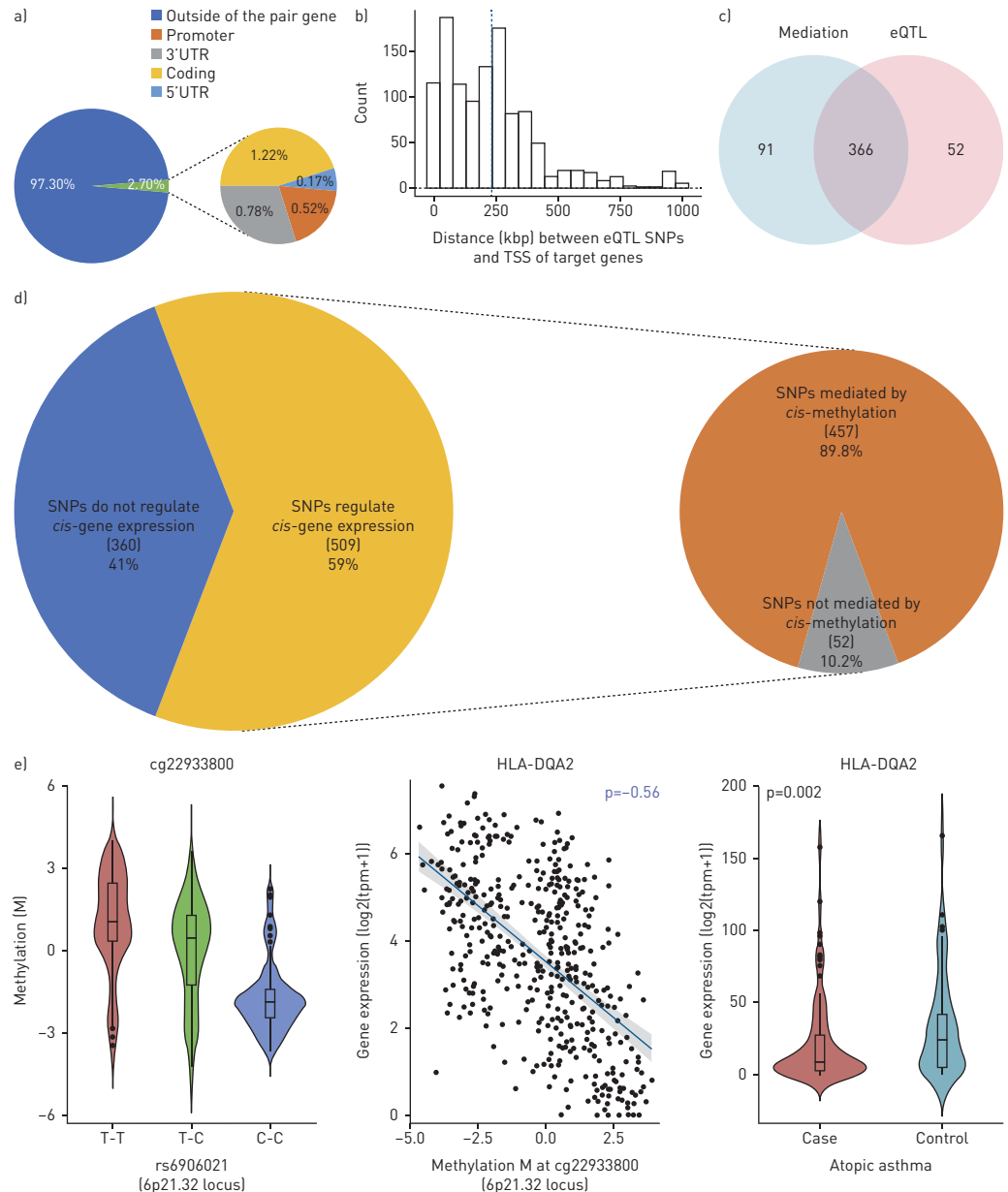
FIGURE 1 Expression quantitative trait loci (eQTL) and mediation analysis of 869 Trans-National Asthma Genetic Consortium (TAGC) single nucleotide polymorphisms (SNPs). a) Percentage of eQTL SNPs that are located outside (blue) or inside (orange) of their paired gene region (promoter: 5'UTR). When SNPs 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. 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 genome-wide association study 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. Left: violin plots of the relationship between the eQTL SNP and an associated methylation probe (red, homozygous major allele; green, heterozygous; blue, homozygous minor allele) Middle: scatter plots of the relationship between DNA methylation and expression of the gene. Right: violin plots of data from atopic asthmatic *versus* those from non-atopic non-asthmatic controls against differential gene expression.

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 www.enhanceratlas.org, finding that five of our 418 eQTL SNPs were located in enhancer regions of their nine 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 analysed 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 10 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 test of mediation effects, we identified 12 035 trios with significant mediation effects (FDR p<0.05), comprising 4571 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 summarise, 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 5119 SNP-gene expression associations in EVA-PR, 4571 (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% [3]. 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.

Soyeon Kim[1,2,6], **Erick Forno** [1,2,6], **Qi Yan**[1,2], **Yale Jiang**[1,3], **Rong Zhang**[4], **Nadia Boutaoui**[1,2], **Edna Acosta-Pérez**[5], **Glorisa Canino**[5], **Wei Chen** [1,2,7] **and Juan C. Celedón**[1,2,7]
[1]Division of Pediatric Pulmonary Medicine, UPMC Children's Hospital of Pittsburgh, University of Pittsburgh, Pittsburgh, PA, USA. [2]Dept of Pediatrics, School of Medicine, University of Pittsburgh, Pittsburgh, PA, USA. [3]School of Medicine, Tsinghua University, Beijing, China. [4]Dept of Statistics, University of Pittsburgh, Pittsburgh, PA, USA. [5]Behavioral Sciences Research Institute, University of Puerto Rico, San Juan, Puerto Rico. [6]Shared first authors. [7]Shared senior authors.

Correspondence: Juan C. Celedón, Division of Pediatric Pulmonary Medicine, UPMC Children's Hospital of Pittsburgh, 4401 Penn Avenue, Pittsburgh, PA 15224, USA. E-mail: juan.celedon@chp.edu

References

1   Shrine N, Portelli MA, John C, *et al.* Moderate-to-severe asthma in individuals of European ancestry: a genome-wide association study. *Lancet Respir Med* 2019; 7: 20–34.

2   Wyss AB, Sofer T, Lee MK, *et al.* Multiethnic meta-analysis identifies ancestry-specific and cross-ancestry loci for pulmonary function. *Nat Commun* 2018; 9: 2976.

3   Demenais F, Margaritte-Jeannin P, Barnes KC, *et al.* Multiancestry association study identifies new asthma risk loci that colocalize with immune-cell enhancer marks. *Nat Genet* 2018; 50: 42–53.

4   Yan Q, Brehm J, Pino-Yanes M, *et al.* A meta-analysis of genome-wide association studies of asthma in Puerto Ricans. *Eur Respir J* 2017; 49: 1601505.

5   Moffatt MF, Gut IG, Demenais F, *et al.* A large-scale, consortium-based genomewide association study of asthma. *N Engl J Med* 2010; 363: 1211–1221.

6   Brugha R, Lowe R, Henderson AJ, *et al.* DNA methylation profiles between airway epithelium and proxy tissues in children. *Acta Paediatr* 2017; 106: 2011–2016.

7   Forno E, Wang T, Qi C, *et al.* DNA methylation in nasal epithelium, atopy, and atopic asthma in children: a genome-wide study. *Lancet Respir Med* 2019; 7: 336–346.

8   Buniello A, MacArthur JAL, Cerezo M, *et al.* The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Research* 2018; 47: D1005–D1012.

9   Movahedi M, Moin M, Gharagozlou M, *et al.* Association of HLA class II alleles with childhood asthma and Total IgE levels. *Iran J Allergy Asthma Immunol* 2008; 7: 215–220.

10 Ried JS, Baurecht H, Stückler F, *et al.* Integrative genetic and metabolite profiling analysis suggests altered phosphatidylcholine metabolism in asthma. *Allergy* 2013; 68: 629–636.

11 Galanter JM, Gignoux CR, Torgerson DG, *et al.* Genome-wide association study and admixture mapping identify different asthma-associated loci in Latinos: The Genes-environments & Admixture in Latino Americans study. *J Allergy Clin Immunol* 2014; 134: 295–305.

12 Miller M, Broide DH. Why is ORMDL3 on chromosome 17q21 highly linked to asthma? *Am J Respir Crit Care Med* 2018; 199: 404–406.

13 Zhang X, Köhl J. A complex role for complement in allergic asthma. *Expert Rev Clin Immunol* 2010; 6: 269–277.

14 Calender A, Rollat Farnier PA, Buisson A, *et al.* Whole exome sequencing in three families segregating a pediatric case of sarcoidosis. *BMC Med Genomics* 2018; 11: 23–23.

15 Yang IV, Pedersen BS, Liu AH, *et al.* The nasal methylome and childhood atopic asthma. *J Allergy Clin Immunol* 2017; 139: 1478–1488.