Multi-omic association study implicates *PPP1R13B* in DNA methylation-mediated genotype and smoking exposure effects on lung function in children living in urban settings

Matthew Dapas[1], Emma E. Thompson[1], William Wentworth-Sheilds[1], Selene M. Clay[1], Cynthia M. Visness[2], Agustin Calatroni[2], Joanne E. Sordillo[3], Diane R. Gold[4,5], Robert A. Wood[6], Melanie Makhija[7], Gurjit K. Khurana Hershey[8,9], Michael G. Sherenian[8,9], Rebecca S. Gruchalla[10], Michelle A. Gill[11], Andrew H. Liu[12], Haejin Kim[13], Meyer Kattan[14], Leonard B. Bacharier[15], Deepa Rastogi[16], Matthew C. Altman[17], William W. Busse[18], Patrice M. Becker[19], Dan L. Nicolae[20], George T. O'Connor[21], James E. Gern[18], Daniel J. Jackson[18], Carole Ober[1]

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

[2] Rho Inc., Durham, NC

[3] Department of Population Medicine, Harvard Medical School, Boston, MA

[4] Department of Environmental Health, Harvard T.H. Chan School of Public Health, Boston, MA

[5] Channing Division of Network Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA

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

[7] Division of Allergy and Immunology, Ann & Robert H. Lurie Children's Hospital, Chicago, IL

[8] Department of Pediatrics, University of Cincinnati College of Medicine, Cincinnati, OH

[9] Division of Asthma Research, Cincinnati Children's Hospital Medical Center, Cincinnati, OH

[10] Department of Internal Medicine, University of Texas Southwestern Medical Center, Dallas, TX

[11] Department of Pediatrics, Washington University School of Medicine, St. Louis, MO

[12] Department of Allergy and Immunology, Children's Hospital Colorado, University of Colorado School of Medicine, Aurora, CO

[13] Department of Medicine, Henry Ford Health System, Detroit, MI

[14] Columbia University College of Physicians and Surgeons, New York, NY

[15] Monroe Carell Jr. Children's Hospital at Vanderbilt University Medical Center, Nashville, TN

[16] Children's National Health System, Washington, DC

[17] Department of Allergy and Infectious Diseases, University of Washington, Seattle, WA

[18] Department of Pediatrics and Medicine, University of Wisconsin School of Medicine and Public Health, Madison, WI

[19] National Institute of Allergy and Infectious Diseases, Bethesda, MD

[20] Department of Statistics, University of Chicago, Chicago, IL

[21] Pulmonary Center, Boston University School of Medicine, Boston, MA

ABSTRACT

Impaired lung function in early life is associated with the subsequent development of chronic respiratory disease. Most genetic associations with lung function have been identified in adults of European descent and therefore may not represent those most relevant to pediatric populations and populations of different ancestries. In this study, we performed genome-wide association analyses of lung function in a multiethnic cohort of children (n=1035) living in low-income urban neighborhoods. We identified one novel locus at the *TDRD9* gene in chromosome 14q32.33 associated with percent predicted forced expiratory volume in one second (FEV_1) ($p=2.4x10^{-9}$; β_z= -0.31, 95% CI= -0.41- -0.21). Mendelian randomization and mediation analyses revealed that this genetic effect on FEV_1 was partially mediated by DNA methylation levels at this locus in airway epithelial cells, which were also associated with environmental tobacco smoke exposure (p=0.015). Promoter-enhancer interactions in airway epithelial cells revealed chromatin interaction loops between FEV_1-associated variants in *TDRD9* and the promoter region of the *PPP1R13B* gene, a stimulator of p53-mediated apoptosis. Expression of *PPP1R13B* in airway epithelial cells was significantly associated the FEV_1 risk alleles ($p=1.26x10^{-5}$; β=0.12, 95% CI=0.06- 017). These combined results highlight a potential novel mechanism for reduced lung function in urban youth through increased airway epithelial apoptosis resulting from both genetics and smoking exposure.

INTRODUCTION

Reduced lung function is a hallmark of asthma and chronic obstructive pulmonary disease (COPD). Lung function measures, such as forced expiratory volume in one second (FEV_1) and forced vital capacity (FVC), are strong predictors of future all-cause mortality[1-6]. Impairment of lung function often begins in early life[7-10], with lower lung function in infancy being a risk factor for the development of asthma in childhood[11] and COPD in late adulthood[12].

Genetic factors contribute to differences in lung function among individuals, with heritability estimates ranging from 0.50 for FEV_1 to 0.66 for FEV_1/FVC ratio[13]. The many genome-wide association studies (GWAS) of lung function measures[14-25] have implicated pathways related to lung development[20,26-28], inflammation[26], and tissue repair[29], among others[29]. Lung function is also affected by environmental exposures, such as smoking[30-32] and air pollution[33], which can disrupt lung growth and development in early life, increasing the risk of chronic respiratory diseases[8-12,34,35]. For example, exposure to second hand smoke in utero and through childhood is associated with increased risk of childhood asthma[36], lower lung function in adolescence[37], and larger declines in lung function later in life[38,39]. Such adverse exposures are known to alter the epigenetic landscape in exposed individuals[40,41], potentially mediating downstream biological effects[42-44] and modifying genetic associations with lung function[45,46].

Environmental risk factors disproportionally affect socioeconomically disadvantaged children, particularly those living in urban environments[47,48]. In fact, socioeconomic effects contribute to disparities in lung health[49], including the higher burden of chronic respiratory disease among Black and Hispanic children compared to non-Hispanic white children[49-52]. Most genetic association studies of lung function, however, have been limited to adults of European descent. Therefore, genetic risk factors discovered to date may not reflect those most relevant to high-risk populations, which can further exacerbate health disparities[53,54]. Identifying genetic variants and epigenetic variation associated with lung function in high-risk, multiethnic, pediatric populations may provide more direct insights into the early development of impaired lung function.

In this study, we analyzed measures of lung function from the Asthma Phenotypes in the Inner City (APIC)[55,56] and Urban Environment and Childhood Asthma (URECA) cohorts[57], which consist of children living in low-income neighborhoods in 10 U.S. cities. We performed whole-genome sequencing (WGS) on 1,035 participants from APIC and URECA (ages 5-17 years; 67% non-Hispanic Black, 25% Hispanic; 66% with doctor-diagnosed asthma) and performed a GWAS with FEV_1 and the FEV_1/FVC ratio. We then performed expression quantitative trait locus (eQTL) and methylation quantitative trait locus (meQTL) mapping in airway epithelial cells and peripheral blood mononuclear cells (PBMCs) from a subset of the URECA children. We further tested for

genotype and DNA methylation interactions with smoking exposure. We aimed to identify methylation-mediated genetic and smoking exposure associations with lung function, linking environmental effects, epigenetic modifications, and specific genetic risk alleles to reduced pulmonary health in urban youth.

RESULTS

Genetic variants at the *TDRD9* locus are associated with lung function

We completed WGS and variant calling on 1,035 participants from the APIC and URECA studies (APIC=508, URECA=527; **Table 1**). The mean sequencing depth was 31.6x per sample (**Fig S1A**). On average, 95.3%, 90.3% and 62.6% of each genome was mapped with at least 10x, 20x and 30x sequencing read depth, respectively (**Fig S1B**). Approximately 3.8 million high-confidence autosomal variants were called per sample. Variant call concordance between replicate sample pairs (n=3) was >99.9% for single nucleotide polymorphisms (SNPs) and was 98.9% for insertions and deletions (**Table S1**).

The sequenced cohort included 696 (67%) participants who self-identified as non-Hispanic Black and 258 (25%) who self-identified as Hispanic (**Table 1**). Principal component and admixture analyses using genotypes were used to characterize the ancestry of the participants (**Fig 1**). This revealed that the genetic ancestry of our sample was 66% African, 26% European, 7% Native American, and 1% East Asian. The cohort was 54% male and included 681 (66%) children diagnosed with asthma (**Table 1**).

Using the WGS variant calls for 14.1 million variants with minor allele frequency (MAF) ≥0.01, we performed a GWAS of two lung function traits: FEV_1 (% predicted) and FEV_1/FVC (Z-scores), measured between ages 5-17 (**Table 1**, **Fig S2**). The FEV_1 GWAS included 896 participants from APIC (n=504) and URECA (n=392), and the FEV_1/FVC GWAS included 886 participants from APIC (n=497) and URECA (n=389). We identified one locus on chromosome 14q32.33 that was associated with FEV_1 at genome-wide significance ($p<2.5x10^{-8}$); no other variants were associated with FEV_1 and no variants were associated with FEV_1/FVC at genome-wide levels of significance (**Fig 2**). The FEV_1 locus on chromosome 14 consisted of a 200 kb region of associated variants in high linkage disequilibrium (LD) across the *TDRD9* (Tudor Domain Containing 9) gene (**Fig 3**, **Table S2**). The minor allele at the lead SNP (rs10220464; MAF=0.30) was significantly associated with lower FEV_1 ($p=2.4x10^{-9}$; β_z= -0.31, 95% confidence interval (CI)= -0.41- -0.21) and nominally associated with lower FEV_1/FVC ($p=1.1x10^{-3}$; β_z= -0.17, 95% CI= -0.28- -0.07).

Table 1. Demographic characteristics of sequenced APIC and URECA participants.

Characteristic	All	APIC	URECA
Number	1035	508	527
Age, years, mean (SD)	10.3 (2.5)	10.9 (3.1)	9.6 (1.1)
Female sex	477 (46%)	216 (43%)	261 (50%)
Race/Ethnicity			
Black (non-Hispanic)	696 (67%)	319 (63%)	377 (72%)
White (non-Hispanic)	14 (1%)	7 (1%)	7 (1%)
Hispanic	258 (25%)	153 (30%)	105 (20%)
Other/mixed	64 (6%)	26 (5%)	38 (7%)
Unknown	3 (<1%)	2 (<1%)	1 (<1%)
Site			
Baltimore	234 (23%)	85 (17%)	149 (28%)
Boston	189 (18%)	65 (13%)	124 (23%)
Chicago	62 (6%)	62 (12%)	-
Cincinnati	45 (4%)	45 (9%)	-
Dallas	38 (4%)	38 (9%)	-
Denver	59 (6%)	59 (12%)	-
Detroit	50 (5%)	50 (10%)	-
New York	164 (16%)	64 (13%)	100 (19%)
St. Louis	155 (15%)	-	155 (29%)
Washington, D.C.	39 (4%)	39 (8%)	-
Household income < $15k	550 (54%)	273 (54%)	277 (54%)
Caretaker completed HS	756 (73%)	364 (72%)	392 (74%)
Caretaker smokes*	315 (33%)	123 (27%)	192 (39%)
Asthma	681 (75%)	508 (100%)	173 (43%)
BMI, Z-score, mean (SD)	0.9 (1.2)	1.0 (1.2)	0.8 (1.1)
FEV_1, % predicted, mean (SD)	94.9 (16.3)	91.9 (17.6)	98.5 (14.5)
FEV/FVC, mean (SD)	0.80 (0.09)	0.78 (0.10)	0.83 (0.07)

Results are presented as counts and percentages or as means with standard deviations. Missing data were not included in percentage calculations. Ages for URECA correspond to the year the genome-wide association study lung function data were collected. *Caretaker smoking status in URECA was collected at age 10. APIC: Asthma Phenotypes in the Inner City; BMI: body mass index; FEV_1: forced expiratory volume in one second; FEV_1/FVC: ratio of FEV_1 to forced vital capacity; HS: high school; URECA: Urban Environment and Childhood Asthma.

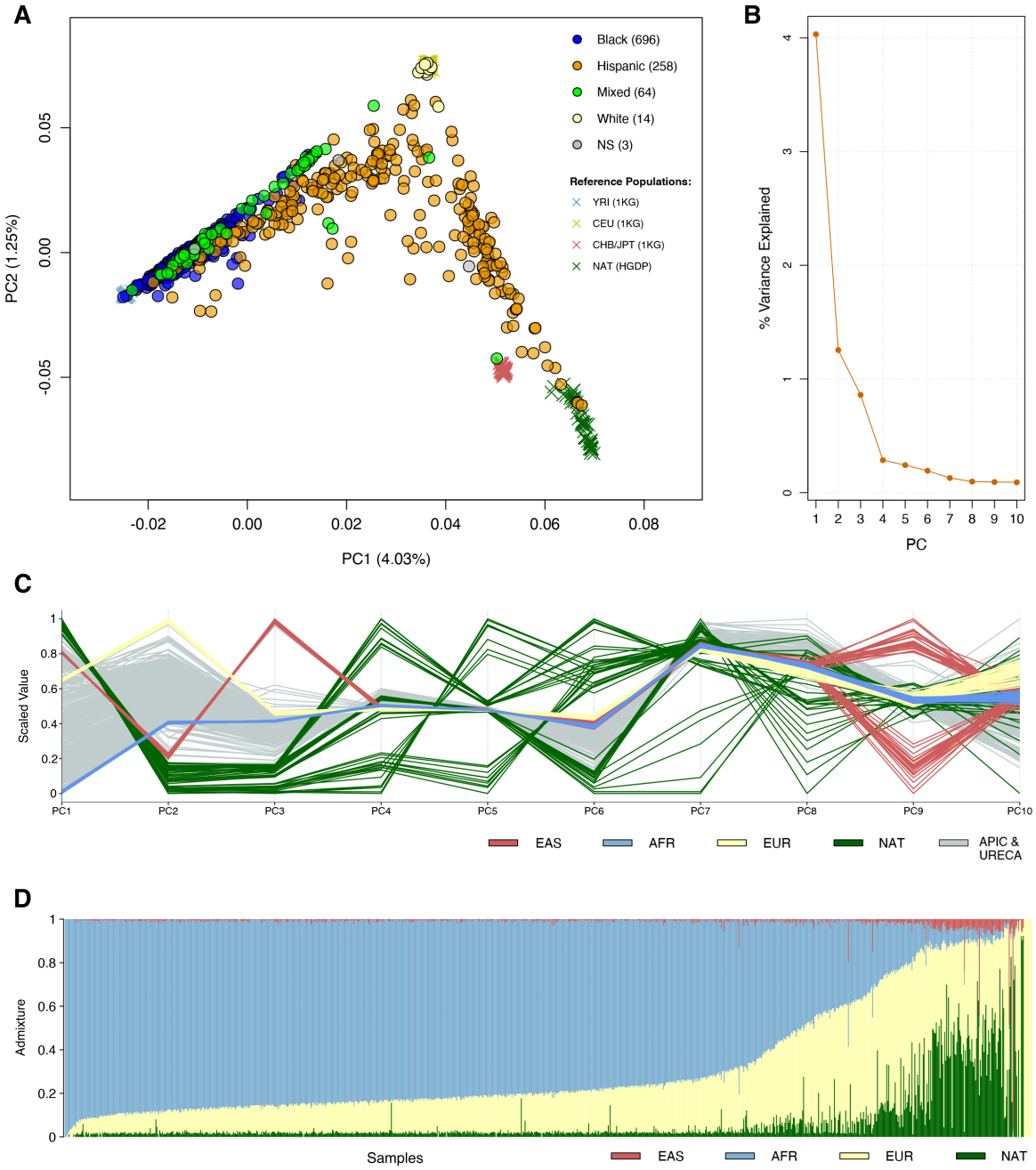

Fig 1. Ancestry composition of sequenced APIC & URECA participants. A) The top two principal components (PCs) of ancestry are plotted for sequenced APIC & URECA participants, colored by self-identified ancestry, along with the four ancestral reference populations used for determining ancestry. NS= not specified. **B)** The proportion of genetic variance explained by each of the top 10 PCs. **C)** The relative values of the top 10 PCs are plotted for each sample, colored by reference population. **D)** The estimated proportion of admixture from each ancestral population is shown for each sequenced APIC & URECA participant. Each vertical line corresponds to one sample. 1KG, 1000 Genomes project; HGDP, Human Genome Diversity Project; YRI, Yoruba in Ibadan, Nigeria; CEU, Utah residents with Northern and Western European ancestry; CHB, Han Chinese in Beijing, China; JPT, Japanese in Tokyo, Japan; NAT, Native Americans from HGDP; EAS, East Asian ancestry; AFR, African ancestry; EUR, European ancestry.

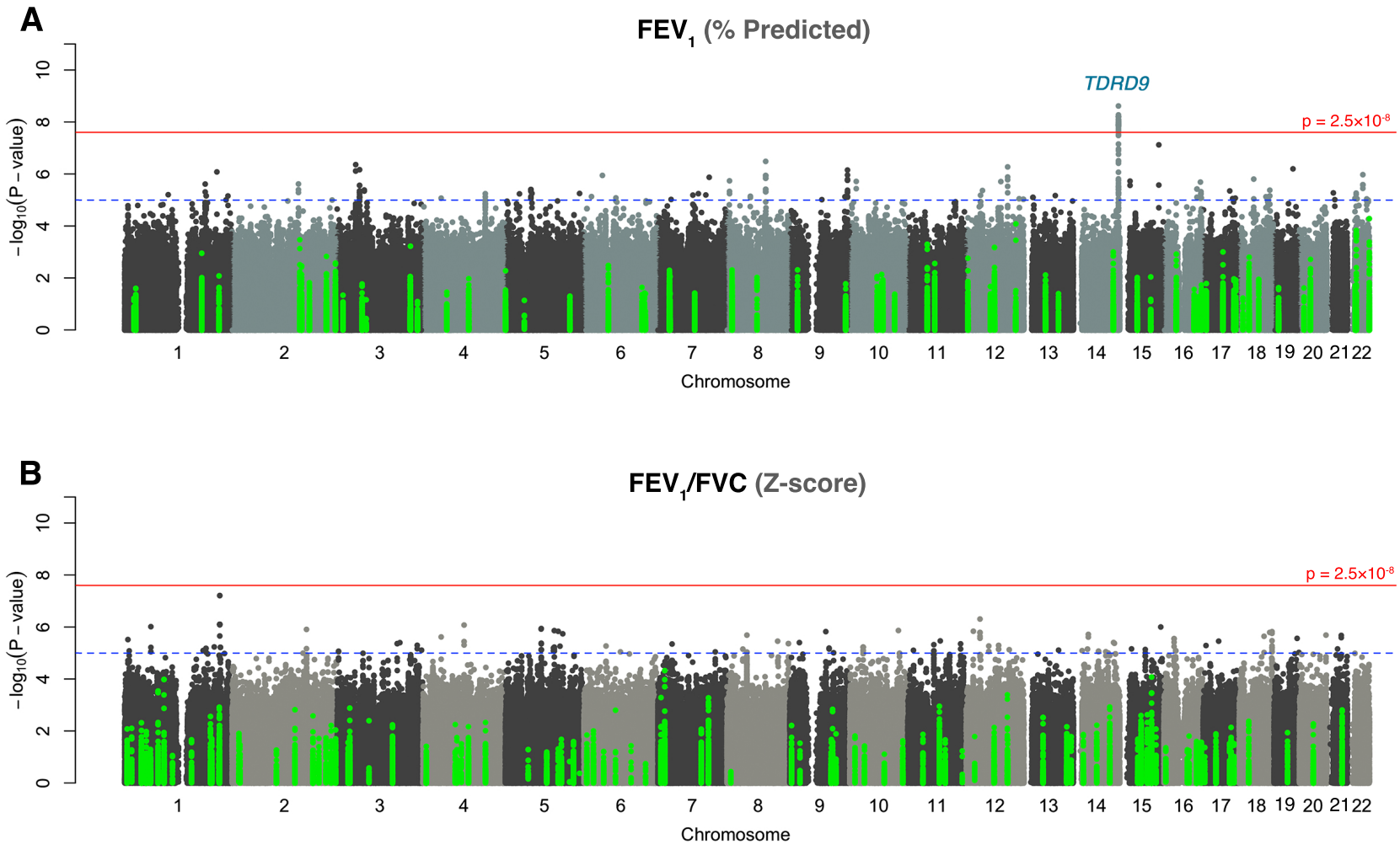

Fig 2. Genome-wide association results. GWAS Manhattan plots for **A)** FEV_1 and **B)** FEV_1/FVC ratio. The horizontal red line indicates genome-wide significance ($p \leq 2.5\times10^{-8}$). The dotted horizontal blue line indicates $p = 1\times10^{-5}$. Variants colored in green are in previously identified GWAS loci[23]. FEV_1, forced expiratory volume in one second; FEV_1/FVC, ratio of FEV_1 to forced vital capacity.

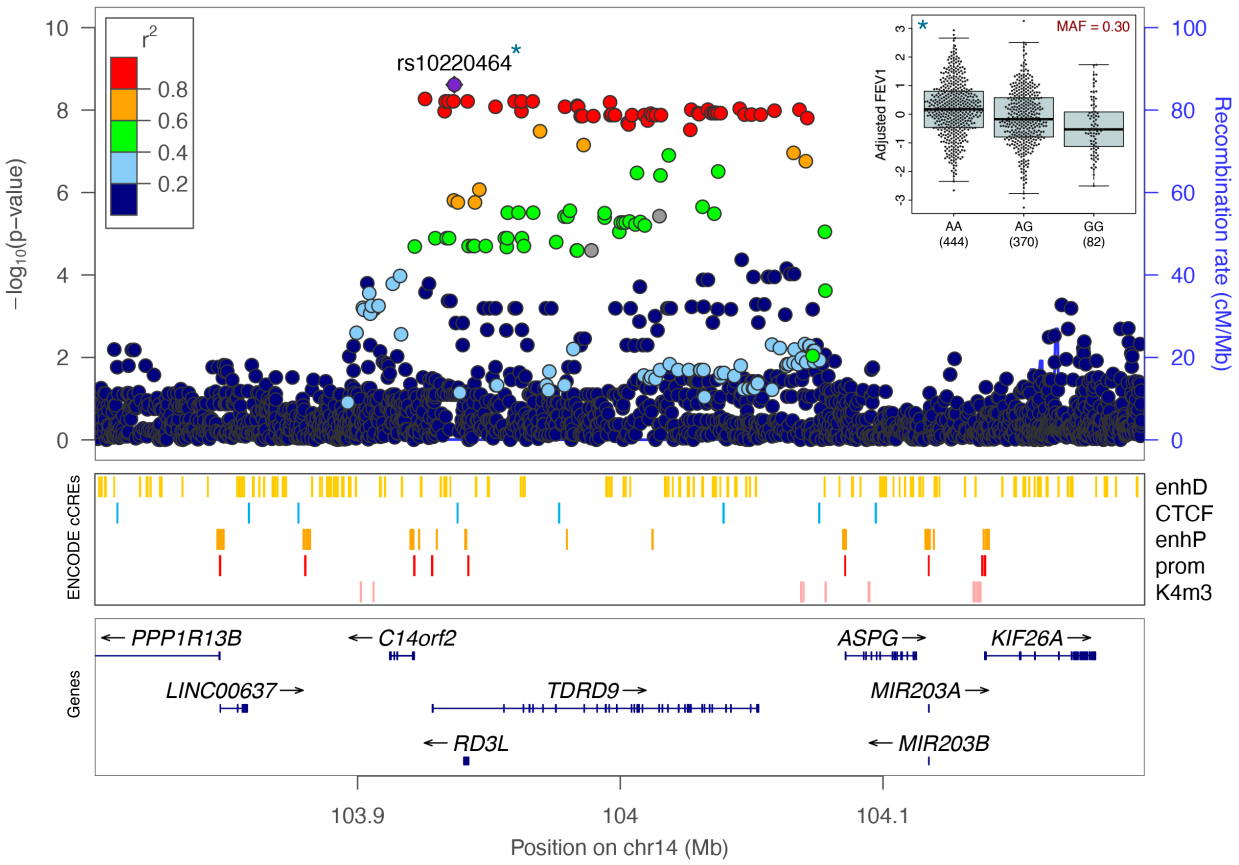

Fig 3. FEV_1-associated variants on chromosome 14q32.33. FEV_1 association results are shown at the *TDRD9* gene locus. Each variant is plotted according to its position and -$\log_{10}$ p-value, colored by linkage disequilibrium to the lead variant, rs10220464, within the sample. Candidate cis-Regulatory Elements (cCREs) from ENCODE[58] are also shown for the region. The inset panel in the upper right shows the distribution of adjusted FEV_1 values by rs10220464 genotype. FEV_1, forced expiratory volume in one second; MAF, minor allele frequency; EnhD, distal enhancer-like signature; CTCF, CCCTC-binding factor sites; enhP, proximal enhancer-like signature; prom, promoter-like signature; K4m3, trimethylation of histone H3 at lysine 4.

We examined association results for the previously identified FEV_1 and FEV_1/FVC loci reported in the meta-analysis by Shrine and colleagues[23], which included 70 loci for FEV_1 and 117 for FEV_1/FVC. Of these, 64 of the lead SNPs for FEV_1 and 112 for FEV_1/FVC were genotyped in the APIC and URECA sample. Only one SNP, for FEV_1, replicated with false discovery rate (FDR) $q<0.05$ (rs9610955; $p=1.0\times10^{-4}$; $\beta_z= -0.38$, 95% CI= -0.58- -0.19; **Fig S3, Fig S4**). Cumulatively, 56% (n=36) and 54% (n=60) of these SNPs demonstrated consistent directions of effect for FEV_1 and FEV_1/FVC, respectively, with effect size correlations of 0.29 (95% CI= 0.05- 0.50; p=0.020) for FEV_1 and 0.42 (95% CI=0.25- 0.56; $p=4.2\times10^{-6}$) for FEV_1/FVC.

Lung function risk alleles are associated with DNA methylation at the *TDRD9* locus in airway epithelial cells

The majority of complex trait-associated variants exert effects by altering gene regulatory networks[59-61]. These changes are often marked by quantitative differences in DNA methylation levels[62-64]. We therefore investigated correlations between the FEV_1-associated allele at *TDRD9* and DNA methylation at the locus in upper airway (nasal) epithelial cells (NECs) from URECA children at age 11 (n=286). We tested for associations between the FEV_1 genotype, as tagged by rs10220464, and DNA methylation levels at 796 CpG sites within 10 kb of the region containing FEV_1-associated variants ($p<1x10^{-5}$, n=82). The rs10220464 genotype was an meQTL for 5 CpG sites at an FDR <0.05 (**Table S3**). DNA methylation levels at only one of these CpG sites, cg03306306 ($p=2.3\ x10^{-4}$; $\beta=0.07$, 95% CI=0.03-0.10; **Fig 4a**), was also significantly associated with FEV_1 at age 10 in URECA (p=0.011; $\beta= -11.48$, 95% CI= -20.27- -2.69; **Fig 4b**). The rs10220464 genotype accounted for 4.7% of residual variation in cg03306306 methylation, and cg03306306 methylation explained 2.4% of residual variation in FEV_1.

We then analyzed cg03306306 methylation in PBMCs collected at age 7 (n=169)[65] from URECA children to evaluate whether the genotype and lung function associations observed in NECs were shared with blood cells. In PBMCs, we observed no correlation between the rs10220464 risk allele and cg03306306 methylation (**Fig 4c**), nor was there an association between cg03306306 methylation and FEV_1 (**Fig 4d**). These results indicate that cg03306306 methylation dynamics in the airway epithelium are not present in peripheral blood cells.

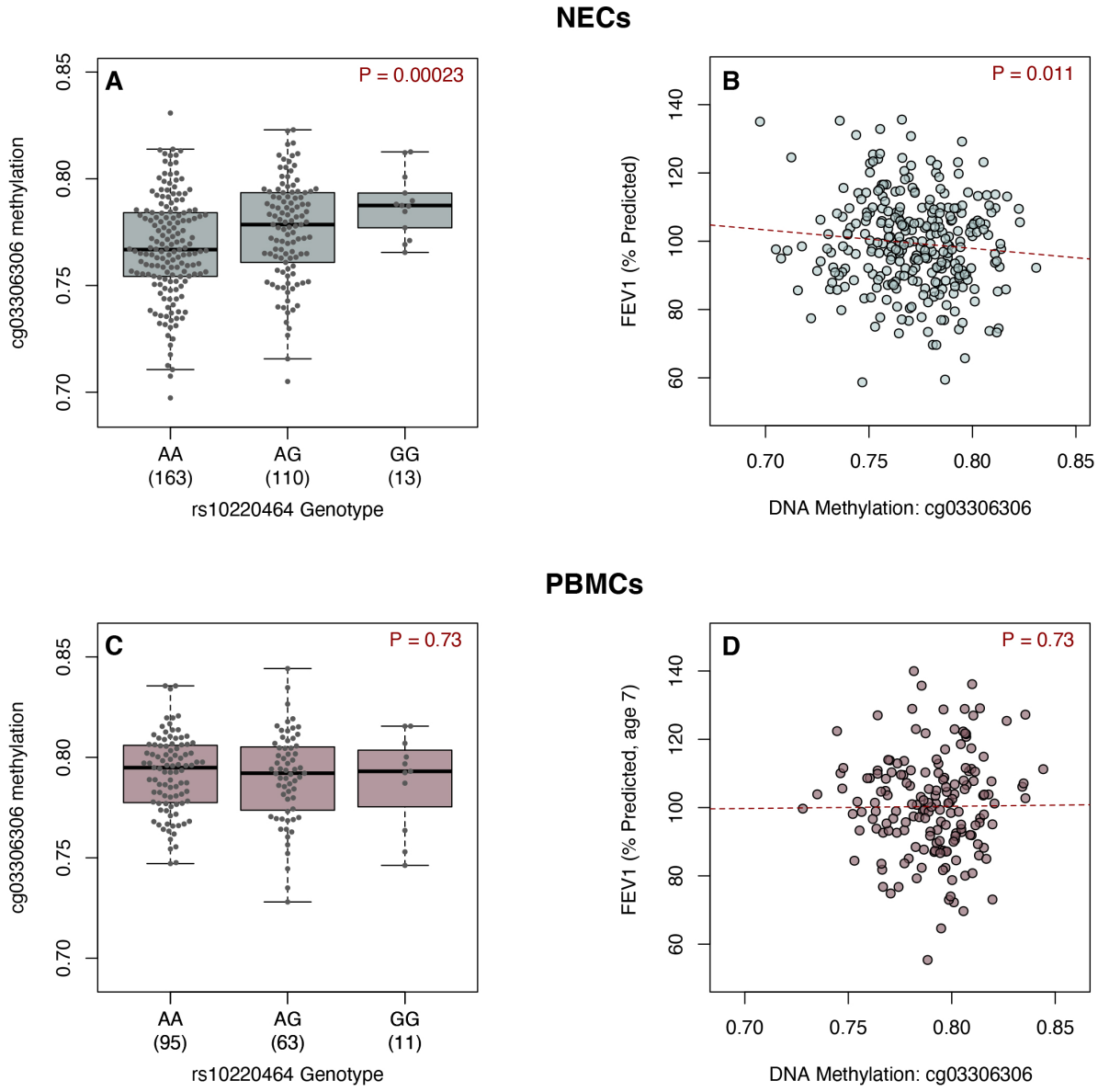

Fig 4. Genotype and FEV_1 associations with DNA methylation. DNA methylation levels at cg03306306 are shown by rs10220464 genotype and FEV_1 measures are plotted against cg03306306 methylation levels in NECs at age 11 **(A, B)**, and PBMCs at age 7 **(C, D)** from URECA. FEV_1, forced expiratory volume in one second; NECs, nasal epithelial cells; PBMCs, peripheral blood mononuclear cells; URECA, Urban Environment and Childhood Asthma study.

Smoking exposure is associated with DNA methylation at the *TDRD9* locus

DNA methylation at the *TDRD9* locus had previously been associated with maternal smoking during pregnancy[66,67]. Therefore, we tested for associations between smoking exposure (**Fig S5**) and DNA methylation at this locus in the URECA children. Methylation at cg03306306 in NECs was significantly associated with nicotine metabolite (cotinine) levels in urine collected at ages 7-10 years (p=0.015; β=0.03, 95% CI=0.01-0.05; **Fig 5**). Methylation at cg03306306 in PBMCs from age 7 was not associated with urine cotinine levels.

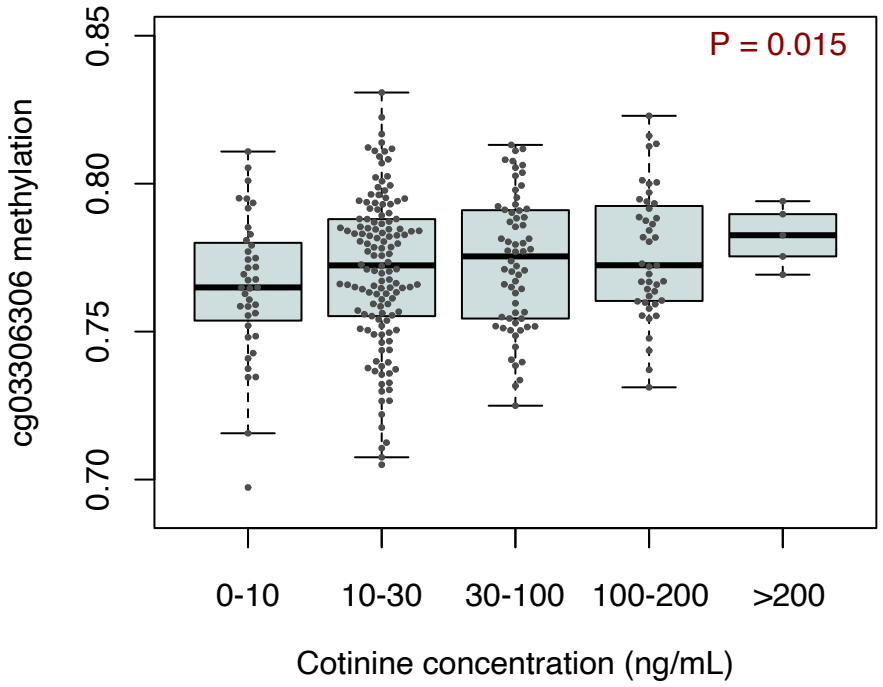

Fig 5. DNA methylation association with smoking exposure. DNA methylation at cg03306306 in nasal epithelial cells at age 11 are plotted against urine cotinine levels from URECA at ages 8-10 as measured using the NicAlert assay (n=285). URECA, Urban Environment and Childhood Asthma study.

To determine if there was an interaction effect between genotype and smoking exposure on DNA methylation and/or lung function, we repeated the cotinine association tests in URECA with the addition of an interaction term to assess if the genotype effect differed between individuals with low and high exposures to smoking. There were no significant genotype-by-smoking exposure interaction effects on methylation levels in NECs or PBMCs in URECA, nor were there any significant methylation-by-smoking effects on FEV_1 (**Fig S6**). There was modest evidence for a genotype-by-smoking exposure interaction effect on FEV_1 in the combined APIC and URECA sample, but this did not reach statistical significance (p=0.06, **Fig S7**).

Genetic effects on lung function are mediated by DNA methylation

To determine if DNA methylation at the *TDRD9* locus had a causal effect on lung function, we performed a Mendelian randomization (MR) analysis using two-stage least squares (2SLS) regression. In the first stage, cg03306306 methylation levels in NECs were regressed on an instrument composed of four meQTLs for cg03306306 (rs11160777, rs137961671, rs7143936, rs11160776; **Methods**). In the second stage, FEV_1 was regressed on the predicted DNA methylation values generated from the first stage regression, thereby yielding a causal effect estimate of cg03306306 methylation on FEV_1. Urine cotinine levels were included as a covariate in both stages. The causal effect of cg03306306 methylation on FEV_1 was statistically significant (p=0.020). We further performed a bootstrapped mediation analysis to test whether the rs10220464 risk allele effect on FEV_1 was mediated by DNA methylation. The indirect effect of rs10220464 on FEV_1 via cg03306306 methylation was significant (β=-0.61, 95% CI=-1.58- -0.04). These results indicate that the effect of the FEV_1-associated genotype at the *TDRD9* locus is partially mediated through its impact on nearby DNA methylation levels.

Gene expression and promoter-enhancer interactions implicate *PPP1R13B*

Trait-associated variants and DNA methylation often affect the transcriptome by influencing the expression of one or more neighboring genes[68,69]. Identifying these correlations can help infer causal mechanisms[70]. Therefore, we next explored the relationship between the genotype for the lead FEV_1 variant rs10220464 and the expression of genes within 1 Mb in NECs and PBMCs from URECA children. Notably, the rs10220464 genotype was not associated with *TDRD9* expression levels in these cells (NECs: p=0.60, β= 0.116; PBMCs: p=0.91, β= 0.014). Of the 27 genes that were evaluated (**Table S4**), rs10220464 was significantly associated with the expression of only one gene, *PPP1R13B* (Protein Phosphatase 1 Regulatory Subunit 13B; FDR $q=2.77.x10^{-4}$; $p=1.26x10^{-5}$; β=0.12, 95% CI=0.06- 017; **Fig 6a**), in NECs. *PPP1R13B* expression levels were also the most strongly associated of the 27 genes with methylation at cg03306306 in NECs (p=0.018; β=0.10, 95% CI=0.02- 018; **Fig 6b**). *PPP1R13B* expression in NECs, however, was not associated with FEV_1 or smoking exposure (**Fig S8**).

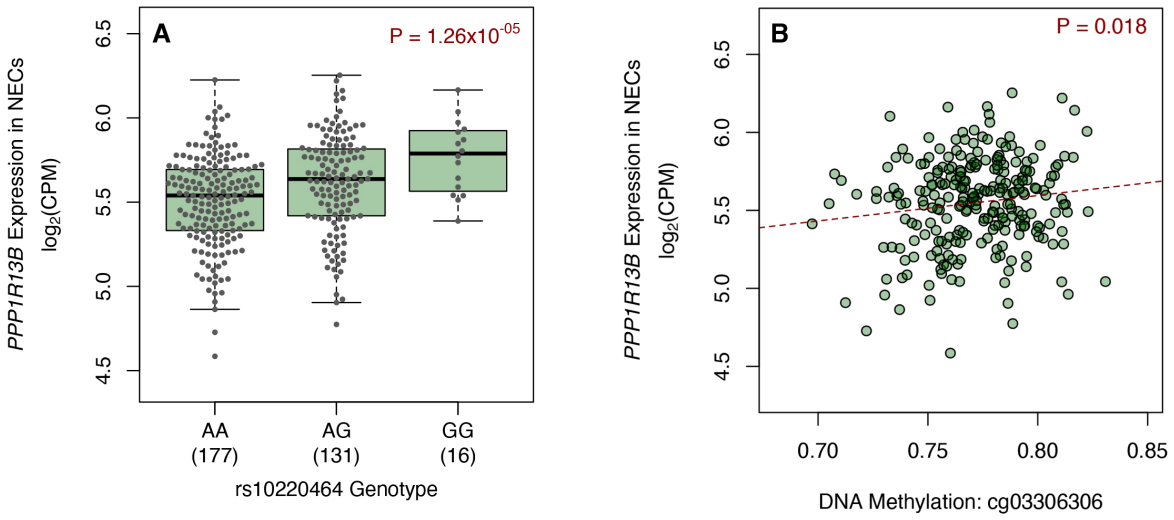

Fig 6. *PPP1R13B* gene expression in NECs. *PPP1R13B* gene expression in NECs at age 11 are plotted against A) rs10220464 genotype (n=324) and B) DNA methylation at cg03306306 in NECs at age 11 (n=254). NECs, nasal epithelial cells; CPM, counts per million.

The transcription start site of *PPP1R13B* resides 87 kb from rs10220464 and 152 kb from cg03306306, suggesting long-range interactions between the FEV_1-associated genotype and the promoter of *PPP1R13B*. To determine whether any of the FEV_1-associated GWAS variants at the *TDRD9* locus resided in regions that physically interacted with the promoters of *cis*-genes, we evaluated chromatin interactions in lower airway (bronchial) epithelial cells (BECs)[71], assessed by promoter-capture Hi-C. Forty-two of the GWAS variants resided in regions that interacted with the promoters of 9 different genes expressed in NECs (**Fig 7**; **Table S5**). The gene most frequently mapped to these variants was *PPP1R13B*, with 15 variants located in 3 different interaction loops. Moreover, the strongest observed interaction was between a region containing 4 FEV_1-associated variants and the *PPP1R13B* promoter (CHiCAGO score = 9.38; **Table S5**), suggesting that this region is an enhancer for *PPP1R13B* expression. This putative enhancer region is located just 2.21 kb from cg03306306.

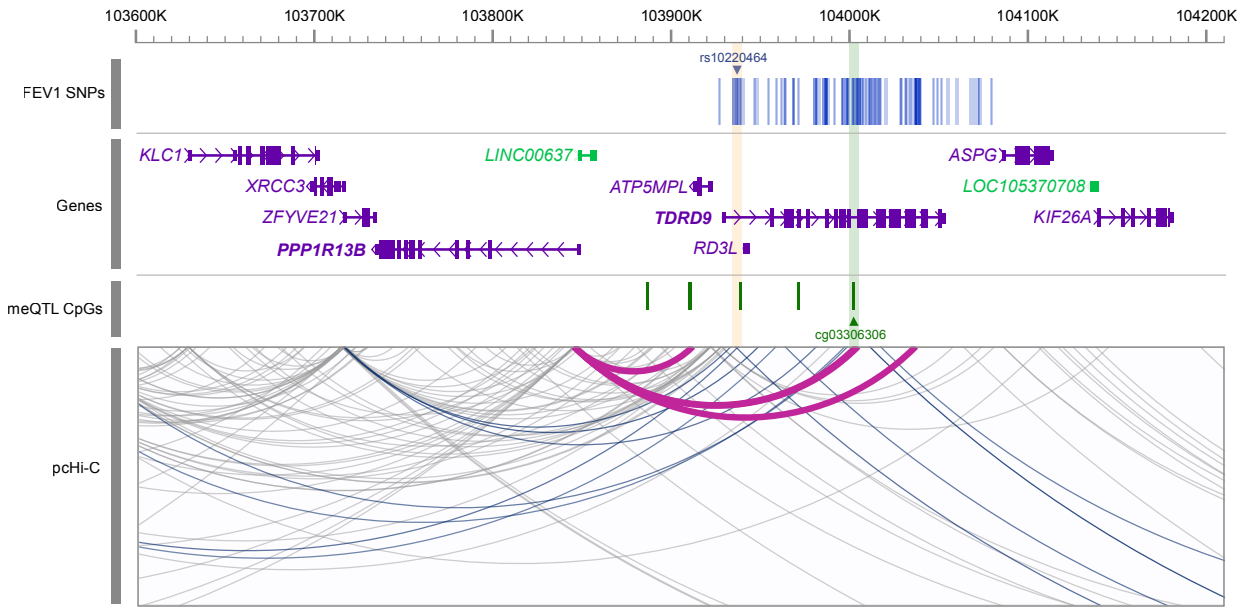

Fig 7. Promoter-enhancer interactions at *TDRD9* locus in nasal epithelial cells. Promoter-to-enhancer chromatin interactions captured by Hi-C in nasal epithelial cells from URECA at age 11 are displayed as grey arcs. SNPs associated with FEV_1 ($p<1x10^{-5}$) are marked by blue lines in the top row according to their genomic position. The lead FEV_1 SNP, rs1022464, is highlighted in yellow. CpG sites associated with rs1022464 (FDR<0.05) are displayed as green markers below the genes, with cg03306306 highlighted in green. Chromatin Interactions containing SNPs associated with FEV_1 ($p<1x10^{-5}$) are highlighted in blue. Magenta arcs highlight interactions between the *PPP1R13B* promoter and regions containing FEV_1 SNPs and/or rs1022464-associated CpG sites. FEV_1, forced expiratory volume in one second; SNPs, single nucleotide polymorphisms; meQTL, methylation quantitative trait locus; pcHi-C, promoter capture Hi-C.

DISCUSSION

Using whole-genome sequence variant calls in an asthma-enriched cohort of predominantly African-American children raised in urban environments, we identified a genotype at the *TDRD9* locus associated with lower FEV_1 % predicted. This genotype effect was partially mediated by DNA methylation in airway epithelial cells, which were also correlated with smoking exposure. Data from RNA-sequencing and promoter-capture Hi-C in airway epithelial cells suggested that these FEV_1-associated genetic and epigenetic variations influence the expression of the *PPP1R13B* gene through long-range interactions.

The *PPP1R13B* gene encodes a protein that promotes apoptosis, a form of programmed cell death, via its interaction with the tumor suppressor p53 and is often referred to by its alias ASPP1 (apoptosis-stimulating protein of p53 1)[72]. In response to oncogenic stress, PPP1R13B translocates to the nucleus, where it enhances the transcriptional activity of p53 on specific target genes relevant to apoptosis[73,74]. Exposure to smoking and fine particulate matter induces epithelial apoptosis in the lung via p53[75-77]. PPP1R13B may also promote apoptosis in a p53-independent manner by inhibiting autophagy in response to upregulation by EGR-1 (early growth response protein 1)[78]. EGR-1 mediates stress-induced proinflammatory responses in the airway epithelium and contributes to the pathogenesis of COPD[79-84]. Within the lung, *PPP1R13B* is indeed predominantly expressed in epithelial cells, particularly in alveolar type 2 cells, and less so in immune cells and fibroblasts[85,86]. However, Cheng and colleagues studied PPP1R13B function in lung fibroblasts and found that it was upregulated following SiO_2 exposure, where it promoted fibroblast proliferation and migration through endoplasmic reticulum stress and autophagy pathways[87]. Overall, these studies suggest that *PPP1R13B* plays a key role in maintaining tissue homeostasis by regulating apoptosis and autophagy in response to environmental stimuli[73,88,89]. The specific function(s) of this gene in the airway epithelium and its potential impact on lung development remain to be elucidated. *PPP1R13B* expression was not directly associated with lung function or urine cotinine levels in the URECA children, but its cofactors[78,90] have been found previously to be upregulated in smokers with COPD[80,91]. Given its association with lung function alleles in our study, its expression in the lung epithelium, and its purported functions in autophagy and apoptosis pathways, additional study of *PPP1R13B* in lung development is warranted, particularly in the context of adverse environmental stimuli, many of which are enriched in low-income urban environments.

In NECs, *PPP1R13B* gene expression was significantly associated with DNA methylation levels at the cg03306306 CpG site in *TDRD9*. Methylation at the *TDRD9* locus was previously reported to correlate with specific environmental exposures[66,67,92] and with *TDRD9* expression in blood[66,93]. *TDRD9* is lowly expressed in the lung but is detected in alveolar macrophages and in

monocytes[85,86]. Interestingly, the gene was among the most differentially expressed genes in alveolar macrophages in smokers relative to non-smokers[94], and its knockdown in TDRD9-expressing lung carcinomas resulted in increased apoptosis[95]. Its expression was not correlated with the rs10220464 genotype in URECA NECs or PBMCs, but rs10220464 is an eQTL for *TDRD9* expression in whole blood in GTEx data[96], with the minor allele associated with lower *TDRD9* expression. Although evidence from this study points to *PPP1R13B* in the airway epithelium, we can't exclude the possibility that *TDRD9* or other genes could contribute to the locus' influence on lung function via other tissues.

The correlations of FEV_1 and rs10220464 with cg03306306 methylation and *PPP1R13B* expression in NECs were absent in PBMCs. Although global DNA methylation patterns between tissues are highly correlated[97], tissue-specific differentially methylated regions are more likely to be functional, particularly if they are positively correlated with gene expression[98]. The *TDRD9* locus has not been identified in epigenome-wide association studies of lung function[44,99-103], but these measured DNA methylation from blood, which may be an insufficient proxy for methylation in the lungs[104]. Indeed, previous studies have found that DNA methylation profiles in NECs are significantly more predictive of pediatric asthma than those in PBMCs[105,106]. Furthermore, epigenetic biomarkers can change with age. For example, epigenetic markers for lung function in adults do not replicate in children[100].

Mendelian randomization and mediation analyses indicated that methylation at cg03306306 in NECs mediated the rs10220464 genotype effect on FEV_1, but there was residual correlation between rs10220464 and FEV_1, signifying that the genotype effect was only partially mediated by cg03306306. The FEV_1 association signal at the *TDRD9* locus included many variants in high LD across a 200 kb region that could be independently contributing to function. Some of the variants lie in different long-range enhancers (**Fig 7**; **Table S5**)[58]. It is also possible that one or more correlated variants were not included because they failed quality control standards. In addition, due to the limited sample size of the WGS cohort, we excluded rare variants (MAF<0.01) from consideration, which could contribute to the signal at this locus. Additional fine mapping and/or functional studies are needed to identify the causal variant(s) and full mechanism of action.

If increased methylation at cg03306306 leads to lower FEV_1 in response to environmental stressors, one might expect to see interaction effects with smoking exposure on FEV_1. We did not detect any such interactions, although our analyses in that regard were likely underpowered given our observed effects and sample sizes[36]. Furthermore, because this study was limited to children living in low-income urban neighborhoods, environmental risk factors are likely to be more prevalent than in the general population[55-57]. Additionally, such exposures are not necessarily ubiquitous across all the different neighborhoods and communities represented in this sample, and although environmental

tobacco smoke exposure was examined, many other environmental factors were not considered in this study.

Another limitation of our study was the relatively small size for a GWAS. This likely contributed to the lack of statistically significant replication for previously identified lung function loci[23], considering that the observed effects were correlated with results of prior GWAS. However, the APIC and URECA cohorts represent understudied, high-risk, pediatric populations that likely harbor distinct genetic and environmental risk factors compared to older, primarily European ancestry cohorts included in previous GWAS of lung function[14-20,23]. The findings of this study have yet to be replicated in an independent cohort, and should therefore be considered preliminary; however, it is possible that these associations would differ in populations with dissimilar ancestry, age, exposures, and/or asthma risk.

There are additional caveats to consider when interpreting our findings. First, this study integrated data from two cohorts with different recruitment criteria, asthma definitions, and ancestral compositions. To control for potential population stratification, we used the first ten principal components (PCs) of ancestry to adjust lung function values and then included the ancestry PCs as fixed effects in the GWAS models (**Methods**). The linear mixed models also included a genetic relatedness matrix as a random effect to account for residual population structure. Because children with asthma have lower lung function overall (**Table 1**) and their lung function is more affected by certain environmental exposures[107-109], we adjusted for asthma status in the GWAS, as others have done previously[110-113]. The likelihood of discovering lung function variants with consistent effects in asthmatics and non-asthmatics was thereby increased, although genetic determinants of lung function may differ by asthma status[114]. The significant genotype effect at the *TDRD9* locus, however, remained when asthma was excluded as a covariate. Second, some of the analyses used data collected at different timepoints. For example, most of the urine cotinine and spirometry measures were collected at age 10, but the samples used for the NEC DNA methylation and RNA-seq analyses were collected at age 11. Because DNA methylation and gene expression can change over time[40,115-117], their values at age 11 may not be fully representative of exposures at age 10. Finally, the promoter-capture Hi-C data were from lower airway (bronchial) epithelial cells, whereas the DNA methylation and RNA-seq data were generated from upper airway (nasal) epithelial cells. Although there are transcriptomic differences between epithelial cells from each compartment, their respective profiles are highly correlated[118-122], and the use of NECs as a proxy for the lower airway epithelium has been validated for both gene expression and epigenetic studies[120-123].

Our study identified a novel avenue through which genetic risk and environmental exposures can affect lung development in children raised in low-income urban neighborhoods. Further research into this pathway may yield mechanistic insights into the early development of impaired lung

function, perhaps leading to interventions that can help reduce the high incidence and morbidity of chronic respiratory diseases in socioeconomically disadvantaged children.

METHODS

Study Population and Phenotypes

We analyzed samples and phenotypes from two National Institutes of Allergy and Infectious Diseases (NIAID)-funded asthma studies conducted by the Inner-City Asthma Consortium (ICAC)[124]: the Asthma Phenotypes in the Inner City (APIC) study[55,56] and the Urban Environment and Childhood Asthma (URECA) birth cohort study [57]. The APIC study was a 1-year, prospective, epidemiological investigation of children and adolescents with asthma (ages 6-17) living in low-income areas (≥20% of residents below poverty level) in nine U.S. cities (Baltimore, MD; Boston, MA; Chicago, IL; Cincinnati, OH; Dallas, TX; Denver, CO; Detroit, MI; New York, NY; Washington, DC). The APIC participants were required to have a diagnosis of asthma by a physician and to have had at least two episodes requiring bronchodilator administration within the past year [55]. The URECA study enrolled pregnant women living in low-income areas of four U.S. cities (Baltimore, MD; Boston, MA; New York, NY; St. Louis, MO) who reported that either or both parents of the index pregnancy had a history of asthma or allergic diseases [57]. This prospective, longitudinal study followed each child through adolescence, periodically collecting samples and clinical and environmental exposure data. Institutional review board approval was received from all participating sites, and written informed consent was obtained from legal guardians of all participating children, who also assented.

Lung function was assessed using spirometry. Lung function measures used in this study for APIC participants were taken at the study entry visit (V0). For URECA, measurements from age 10 were used when available; otherwise, the most recent measurement after age 5 was used (**S6 Table**). Asthma status was assigned according to study-specific criteria. For APIC, asthma was defined by a doctor's diagnosis of asthma and short-acting beta-agonist use in the year prior [55]. For URECA, asthma status was determined either by doctor diagnosis, lung function reversibility, or symptom recurrence [125]. The 2012 Global Lung Initiative reference equations [126] were applied to generate percent predicted estimates for FEV_1 and Z-scores for FEV_1/FVC ratio. Urine cotinine levels were measured using NicAlert immunochromatographic assays, which report results on a scale of 0-6 according to different cotinine concentration ranges [127]. For URECA, urine cotinine results were available at age 10 for most participants (n=391); otherwise, assays from age 8 (n=29) or age 7 (n=2) were used. This study utilized DNA methylation and RNA-seq data generated for other URECA studies; therefore, the number of samples included in each analysis varied and was limited by data availability (**S7 Table**).

Whole-Genome Sequencing and Data Processing

DNA was extracted from peripheral blood (APIC, URECA) or cord blood (URECA) and quantified using an Invitrogen Qubit 3 Fluorometer. DNA quality was assessed using the Thermo Scientific NanoDrop One spectrophotometer and confirmed using an Agilent TapeStation system. DNA was processed in batches of 60 using the Illumina Nextera DNA Flex library prep kit with unique dual adaptors. Each set of 60 libraries was sequenced over two NovaSEQ S4 flowcells. Whole-genome sequencing was performed by the University of Chicago Genomics Facility using the Illumina NovaSEQ6000, which generated 150 bp paired-end reads. Sequencing data processing followed the Broad Institute's Genome Analysis Toolkit (GATK) best practices for germline short variant discovery, as implemented in the harmonized pipeline used by the New York Genome Center for TOPMed [128,129]. Reads were aligned to the GRCh38 human reference genome (including alternate loci and decoy contigs) using BWA-MEM (Burrows-Wheeler Aligner; v0.7.17). Aligned reads further underwent duplicate removal (Picard MarkDuplicates; v2.8.1) and base quality score recalibration (GATK BaseRecalibrator; v3.8) against known sites (dbSNP138, known indels, and Mills and 1KG gold standard indels) provided in the GATK resource bundle. Read alignment metrics were calculated using Picard CollectWgsMetrics (v2.8.1) for all aligned reads and for aligned reads with base quality and mapping quality ≥ 20. DNA contamination levels were estimated using VerifyBamID2 (v1.0.6) [130]. Samples with estimated DNA contamination >0.05 were removed from consideration. Samples with poor coverage (<50% of the genome with ≥20x depth) were also removed from further consideration. To identify potential sample swaps, WGS samples were validated using independent genotyping arrays.

QC Array for Sample Validation

To identify potential WGS sample swaps, we independently genotyped the APIC and URECA participants using the Illumina QC Array-24 BeadChip. SNPs were tested for Hardy-Weinberg Equilibrium (HWE) within each self-identified ancestry group using the chi-square test and removed if they deviated from HWE (Bonferroni-adjusted $p<0.05$) within at least one ancestry. SNPs with call rates <0.98 were also removed. Samples with total variant call rates <0.95 were not used. Array data with incorrect or indeterminate sex according to X-chromosome heterozygosity rates (Plink v1.90) were also not used [131]. For fourteen of the sequenced URECA samples, we used results from the Illumina Infinium CoreExome+Custom array for sample validation, which were generated and controlled for quality as described by McKennan and colleagues [132]. WGS and array genotypes were tested for concordance using VerifyBamID (v1.1.3) [133]. WGS samples that were not validated with array data were not included in genetic analyses.

Variant Calling and Quality Control

Variant calls were generated using GATK HaplotypeCaller (v4.1.3.0), accounting for contamination estimates, for single nucleotide variants and short insertions, deletions, and substitutions. Sample genotypes were joined using GATK GenomicsDBImport and GenotypeGVCFs over the genomic intervals defined in the GATK WGS calling region interval list provided in the GATK resource bundle. Genotypes with read depth (DP) <10 or quality scores (GQ) <20 were set as missing. Sites with ≥0.1 missingness were then removed from consideration. Variants with minor allele frequencies >0.05 were tested for accordance with HWE, accounting for population structure [134]. Sites with common variants that deviated from structural HWE ($P<1x10^{-6}$) were removed from consideration. Sites with quality by depth ratios (QD) <4 or >34 were also removed, as we observed declines in variant transition/transversion (TS/TV) ratios beyond these bounds (**S10 Fig**). Variant site quality was further evaluated using machine-learning-based Variant Quality Score Recalibration (VQSR). First, SNPs were modeled using GATK VariantRecalibrator (v4.1.3.0) with Hapmap 3 and with Omni 2.5M SNP chip array as truth resources, 1000G as a training resource, and dbSNP138 as a known sites resource. InDels were likewise trained with the Mills and 1KG gold standard indels dataset as a truth resource and dbSNP138 as a known sites resource. SNPs and InDels with resultant predicted true positive probabilities below 0.997 and 0.990, respectively, were removed from consideration. Variant call accuracy was assessed by comparing call concordance between three replicate sequencing samples using VCFtools (v0.1.14) vcf-compare [135]. Variant call file manipulation was conducted using BCFtools (v1.10.2) [136].

Ancestry Estimation

Ancestry principal components (PCs) were calculated on the intersect of high quality single-nucleotide variants (SNVs) genotyped in the WGS data and several reference panels from the 1000 Genomes Project (1KG; n=156) [137] and the Human Genome Diversity Project (HGDP; n=52) [138]. Native American reference samples consisted of 52 samples from the HGDP with <5% non-native ancestry, according to an analysis of roughly 2 million markers using the program ADMIXTURE (v1.3.0) [139]. These samples were filtered for site quality (missingness 5%; ExcHet<60; VQSLOD≥8.3929), genotype quality (GQ≥20) and depth (DP≥10), MAF >0.02, and HWE (p>0.001) [138]. European, West African, and East Asian reference samples were randomly selected from CEU (n=52), YRI (n=52), JPT (n=26), and CHB (n=26) samples in the phase 3 1KG reference panel [137]. The combined genotypes were pruned for linkage disequilibrium (LD) ≤0.1 within 1Mb intervals. Ancestry PCs were calculated, accounting for subject relatedness, using PC-Air [140] and PC-Relate [141]. Initial kinship estimates were produced using KING[142]. Kinship and PCs were iteratively estimated using PC-Relate and PC-Air, respectively, until estimates for the top 5 PCs stabilized (n=3). Reference population admixture estimates were estimated for each WGS sample with ADMIXTURE

(v1.3.0), using the 1KG and HGDP reference samples for supervised analysis [139]. Because sample relatedness can lead to biased admixture estimates [140,143], admixture was estimated for each WGS sample separately.

Quantitative Trait Association Testing

Quantitative traits were adjusted for confounding factors and normalized using a two-stage approach [144,145]. First, each trait was regressed on age, sex, asthma status, and the first 10 PCs of ancestry. The residuals were then rank-normalized using an inverse normal transformation. In the second stage, the normalized residuals were considered outcome variables in the GWAS, adjusting for the same covariates as in the first stage. Genome-wide association testing was performed for all high-quality common variant calls (MAF≥0.01) using a linear mixed model, as implemented in GEMMA [146], with subject relatedness included as a random effect. Individuals without a confirmed asthma status (n=127) were excluded from trait association testing. The threshold we applied for genome-wide significance was $P \leq 2.5 \times 10^{-8}$, based on a 5×10^{-8} GWAS threshold and further accounting for two tests.

DNA Methylation Analysis

DNA from NECs was collected at age 11 from 287 URECA participants and assessed for genome-wide methylation patterns using the Illumina Infinium Human Methylation EPIC Beadchip. DNA methylation levels from PBMCs at age 7 in URECA were collected and processed as previously described [65]. MeQTL analysis was performed using Matrix eQTL [147]. NEC DNA methylation levels were adjusted globally for sex, array, plate, collection site, DNA concentration, percent ciliated epithelial cells, percent squamous cells, and ancestry PCs 1-3. Principal components analysis was then performed on the residual methylation levels, and the first three PCs were included as covariates in the meQTL association tests. Additional methylation PCs were not included in association tests, as they were significantly correlated with asthma phenotypes. Associations with FDR-adjusted $P<0.05$ were considered significant. MeQTL analysis with the PBMC data included sex, collection site, plate, ancestry PCs 1-3, and eight latent factors [148] (protecting for FEV_1 at age 7) as covariates.

To test CpG site methylation associations with lung function in NECs, we performed linear regressions on the most recent FEV_1 measures, with age, sex, ancestry PCs 1-3, and methylation PCs 1-3 as covariates. For the PBMC analysis, we set FEV_1 at age 7 as the dependent variable, with sex, collection site, plate, ancestry PCs 1-3, and latent factors included as covariates.

For association testing with smoking exposures, we ran linear regressions for DNA methylation and lung function in NECs and PBMCs, as described above, with the addition of cotinine concentrations as a predictor. We further tested for smoking-by-genotype interaction effects

on DNA methylation and lung function using these models by adding an interaction term (cotinine concentration : rs10220464 genotype). Proportions of explained variance were calculated by squaring partial correlation coefficients of regression model predictors [149]. One sample from one sibling pair was removed from all methylation analyses to prevent confounding due to relatedness.

Mendelian Randomization and Mediation Analysis

To assess the causal effects of DNA methylation on lung function, we performed one-sample Mendelian randomization analysis. We applied a two-stage least squares (2SLS) regression to URECA samples with WGS and DNA methylation data (n=285) using ivreg [150]. DNA methylation levels in NECs at the cg03306306 CpG site were first adjusted for methylation PCs 1-3 and used as the endogenous, exposure variable. The adjusted and normalized FEV_1 values from the GWAS were set as the dependent outcome variable. Urine cotinine levels were included as an exogenous covariate (included in both stages). The instrumental variables were chosen from a set of candidate SNPs that were at least nominally associated with cg03306306 methylation with $p<0.15$. Clustering of pairwise linkage disequilibrium values between these SNPs revealed six distinct haplotypes (**S9 Fig**). To ensure instrument exogeneity, each candidate SNP was tested for association with FEV_1 after conditioning on cg03306306 methylation and urine cotinine, and SNPs associated with $p<0.05$ were removed from consideration. Of the remaining candidate SNPs, one was chosen from each haplotype, resulting in an instrument composed of 4 SNPs (rs11160777, rs137961671, rs7143936, rs11160776). Instrument relevance was validated using the F test, endogeneity using the Wu-Hausman test, and instrument exogeneity using the Sargon test. Mediation analysis was conducted with ROBMED [151]. FEV_1 (% predicted) was set as the dependent variable, adjusted cg03306306 methylation as the mediator, and rs10220464 as the independent variable. Age at FEV1 measurement, sex, asthma status, ancestry PCs 1-3, and urine cotinine levels were included as covariates. The indirect effect of rs10220464 on FEV_1 via cg03306306 methylation was estimated using 100,000 bootstrap resamples.

Gene Expression Analysis

We analyzed gene expression in NECs and PBMCs from the URECA birth cohort using RNA-seq. The NEC data were derived from 324 children (156 females, 168 males) at age 11 years at the time of sample collection, and the PBMC data were derived from 132 (78 males, 54 females) PBMC children aged 2 years at the time of collection. Sequencing reads were mapped and quantified using STAR (v2.6.1) [152] and samples underwent trimmed means of M-value (TMM) normalization and voom transformation [153]. Genes with <1 count per million mapped reads (CPM) were removed from analysis. For eQTL association testing in NECs we corrected for sex, the first three ancestral PCs,

collection site, epithelial cell proportion, sequencing batch, and 14 latent factors [148] using limma [154]. In PBMCs, we corrected for sex, the first three ancestral PCs, collection site, and 19 latent factors.

Chromatin Interaction Analysis

Chromatin interactions were assessed using promoter capture Hi-C [155,156] in ex vivo human BECs from 8 adult lung donors, including 4 with asthma. The data was processed and analyzed as previously described [71,157]. Chromosomal interactions were evaluated using the CHiCAGO algorithm [158]. Interactions with CHiCAGO scores ≥5 were considered significant [158]. Genetic variants within 1 kb of a given interacting fragment were considered part of the chromatin loop. Genes that were not expressed in NECs were not included in the analysis.

Acknowledgements: We are grateful to all participants and their families who took part in these studies. We would like to extend special thanks to Pieter Faber and the University of Chicago Genomics Facility, as well as to Petra LeBeau and Rebecca Z. Krouse, formerly of Rho Inc., for their respective contributions.

Funding: This work was supported by NIH grants U19 AI62310, HHSN272200900052C, HHSN272201000052I, UM1 AI114271, UG3 OD023282, and UM1 AI160040. Site data collection was supported by the following NIH grants: RR00052, UL1 TR001079 (Baltimore); M01 RR00533, UL1 RR025771, 1 UL1 TR001430 (Boston); UL1 TR000150 (Chicago); UL1 TR000451, UL1 TR001105 (Dallas); Ul1 RR025780 (Denver); UL1 TR000040, M01 RR00071, UL1 RR024156 (New York); UL1 TR000075 (D.C.); UL1 TR000077 (Cincinnati). M.D. was supported by TL1 TR002388 and T32 HL007605.

Author contributions: M.D., W.W.B., P.B., D.N., G.T.O., J.E.G., D.J.J., C.O. made contributions to the conception and/or design of the study. M.D., E.T., and S.C. performed computational analyses. E.T. prepared samples for 'omic analyses. W.W-S. oversaw 'omic data management. C.M.V. and A.C. oversaw clinical data management. J.E.S., D.G., G.T.O., J.E.G., D.J.J., C.O. contributed to interpretation of results. R.A.W., M.M., G.N.K.H., M.G.S., R.S.G, M.A.G., A.H.L., H.K., M.K., L.B.B., D.S., M.C.A. contributed patient samples and/or data. M.D. wrote the manuscript. All authors read and approved the manuscript.

Disclosures:
All authors, with the exception of M. Altman and P. Becker, report grants from NIH/NIAID during the conduct of study. M. Dapas, C. Visness, A. Calatroni and P. Becker have nothing to disclose outside the submitted work. C. Ober reports personal fees from American Association of Asthma, Allergy and Immunology, outside the submitted work. M. Altman reports personal fees from Sanofi-Regeneron outside the submitted work. W. Busse reports personal fees from Boston Scientific, Novartis, Glaxo SmithKline, Genentech, Sanofi/Genzyme, AstraZeneca, Teva, Regeneron and Elsevier outside the submitted work. M. Gill reports an honorarium for and support for travel to the 2017 AAAAI meeting during the conduct of study and monetary compensation from the American Academy of Pediatrics for her work teaching the biannual Pediatrics board review course, PREP The Course. K. Hershey reports grants from Adare, during the conduct of the study. D. Jackson reports personal fees from Novartis, Boehringer Ingelheim, Pfizer, Regeneron, AstraZeneca, Sanofi and Vifor Pharma, grants and personal fees from GlaxoSmithKline and grants from NIH/NHLBI, outside the submitted work. M. Kattan reports personal fees from Regeneron, outside the submitted work. R.

Gruchalla reports government employment from Center for Biologics Evaluation and Research as well as personal fees from Consulting Massachusetts Medical Society, outside the submitted work. A. Liu reports personal fees from Phadia ThermoFisher as consulting honoraria, grants and non-financial support from ResMed/Propeller Health, non-financial support from Revenio, grants and personal fees from Avillion and personal fees from Labcorp, outside the submitted work. L. Bacharier reports personal fees from GlaxoSmithKline, Genentech/Novartis, DBV Technologies, Teva, Boehringer Ingelheim, AstraZeneca, WebMD/Medscape, Sanofi/Regeneron, Vectura and geCircassia and personal fees and non-financial support from Merckoutside the submitted work. J. Gern reports personal fees from AstraZeneca and Gossamer Bio and personal fees and stock options from Meissa Vaccines Inc, outside the submitted work. In addition, Dr. Gern has a patent Methods of Propagating Rhinovirus C in Previously Unsusceptible Cell Lines issued, and a patent Adapted Rhinovirus C issued. G. O'Connor reports personal fees from AstraZeneca and grants from Janssen Pharmaceuticals, outside the submitted work. R. Wood reports grants from DBV, Aimmune, Astellas, HAL-Allergy and Regeneron and royalties from Up to Date, outsite the submitted work. Dr. Becker's co-authorship of this publication does not necessarily constitute endorsement by the National Institute of Allergy and Infectious Diseases, the National Institutes of Health or any other agency of the United States government.

REFERENCES

1. Hole, D.J. *et al.* Impaired lung function and mortality risk in men and women: findings from the Renfrew and Paisley prospective population study. *BMJ* **313**, 711-5; discussion 715-6 (1996).
2. Schunemann, H.J., Dorn, J., Grant, B.J., Winkelstein, W., Jr. & Trevisan, M. Pulmonary function is a long-term predictor of mortality in the general population: 29-year follow-up of the Buffalo Health Study. *Chest* **118**, 656-64 (2000).
3. Mannino, D.M., Holguin, F., Pavlin, B.I. & Ferdinands, J.M. Risk factors for prevalence of and mortality related to restriction on spirometry: findings from the First National Health and Nutrition Examination Survey and follow-up. *Int J Tuberc Lung Dis* **9**, 613-21 (2005).
4. Chinn, S., Gislason, T., Aspelund, T. & Gudnason, V. Optimum expression of adult lung function based on all-cause mortality: results from the Reykjavik study. *Respir Med* **101**, 601-9 (2007).
5. Miller, M.R., Pedersen, O.F., Lange, P. & Vestbo, J. Improved survival prediction from lung function data in a large population sample. *Respir Med* **103**, 442-8 (2009).
6. Agusti, A., Noell, G., Brugada, J. & Faner, R. Lung function in early adulthood and health in later life: a transgenerational cohort analysis. *Lancet Respir Med* **5**, 935-945 (2017).
7. Stern, D.A., Morgan, W.J., Wright, A.L., Guerra, S. & Martinez, F.D. Poor airway function in early infancy and lung function by age 22 years: a non-selective longitudinal cohort study. *Lancet* **370**, 758-64 (2007).
8. Stocks, J., Hislop, A. & Sonnappa, S. Early lung development: lifelong effect on respiratory health and disease. *Lancet Respir Med* **1**, 728-42 (2013).
9. McGeachie, M.J. *et al.* Patterns of Growth and Decline in Lung Function in Persistent Childhood Asthma. *N Engl J Med* **374**, 1842-1852 (2016).
10. Agusti, A. & Faner, R. Lung function trajectories in health and disease. *Lancet Respir Med* **7**, 358-364 (2019).
11. Bisgaard, H., Jensen, S.M. & Bonnelykke, K. Interaction between asthma and lung function growth in early life. *Am J Respir Crit Care Med* **185**, 1183-9 (2012).
12. Duijts, L., Reiss, I.K., Brusselle, G. & de Jongste, J.C. Early origins of chronic obstructive lung diseases across the life course. *Eur J Epidemiol* **29**, 871-85 (2014).
13. Klimentidis, Y.C. *et al.* Heritability of pulmonary function estimated from pedigree and whole-genome markers. *Front Genet* **4**, 174 (2013).
14. Wilk, J.B. *et al.* A genome-wide association study of pulmonary function measures in the Framingham Heart Study. *PLoS Genet* **5**, e1000429 (2009).

15. Repapi, E. *et al.* Genome-wide association study identifies five loci associated with lung function. *Nat Genet* **42**, 36-44 (2010).

16. Hancock, D.B. *et al.* Meta-analyses of genome-wide association studies identify multiple loci associated with pulmonary function. *Nat Genet* **42**, 45-52 (2010).

17. Soler Artigas, M. *et al.* Genome-wide association and large-scale follow up identifies 16 new loci influencing lung function. *Nat Genet* **43**, 1082-90 (2011).

18. Yao, T.C. *et al.* Genome-wide association study of lung function phenotypes in a founder population. *J Allergy Clin Immunol* **133**, 248-55 e1-10 (2014).

19. Loth, D.W. *et al.* Genome-wide association analysis identifies six new loci associated with forced vital capacity. *Nat Genet* **46**, 669-77 (2014).

20. Wain, L.V. *et al.* Genome-wide association analyses for lung function and chronic obstructive pulmonary disease identify new loci and potential druggable targets. *Nat Genet* **49**, 416-425 (2017).

21. Burkart, K.M. *et al.* A Genome-Wide Association Study in Hispanics/Latinos Identifies Novel Signals for Lung Function. The Hispanic Community Health Study/Study of Latinos. *Am J Respir Crit Care Med* **198**, 208-219 (2018).

22. Wyss, A.B. *et al.* Multiethnic meta-analysis identifies ancestry-specific and cross-ancestry loci for pulmonary function. *Nat Commun* **9**, 2976 (2018).

23. Shrine, N. *et al.* New genetic signals for lung function highlight pathways and chronic obstructive pulmonary disease associations across multiple ancestries. *Nat Genet* **51**, 481-493 (2019).

24. Akenroye, A.T. *et al.* Genome-wide association study of asthma, total IgE, and lung function in a cohort of Peruvian children. *J Allergy Clin Immunol* (2021).

25. Zhu, Z. *et al.* A large-scale genome-wide association analysis of lung function in the Chinese population identifies novel loci and highlights shared genetic aetiology with obesity. *Eur Respir J* **58**(2021).

26. Obeidat, M. *et al.* Molecular mechanisms underlying variations in lung function: a systems genetics analysis. *Lancet Respir Med* **3**, 782-95 (2015).

27. Gharib, S.A. *et al.* Integrative pathway genomics of lung function and airflow obstruction. *Hum Mol Genet* **24**, 6836-48 (2015).

28. Portas, L. *et al.* Lung Development Genes and Adult Lung Function. *Am J Respir Crit Care Med* **202**, 853-865 (2020).

29. Kheirallah, A.K., Miller, S., Hall, I.P. & Sayers, I. Translating Lung Function Genome-Wide Association Study (GWAS) Findings: New Insights for Lung Biology. *Adv Genet* **93**, 57-145 (2016).

30. Aschard, H. *et al.* Evidence for large-scale gene-by-smoking interaction effects on pulmonary function. *Int J Epidemiol* **46**, 894-904 (2017).

31. Park, B. *et al.* Effect of 6p21 region on lung function is modified by smoking: a genome-wide interaction study. *Sci Rep* **10**, 13075 (2020).

32. Kim, W. *et al.* Smoking Interaction with a Polygenic Risk Score for Reduced Lung Function. *medRxiv*, 2021.03.26.21254415 (2021).

33. Melbourne, C.A. *et al.* Genome-wide gene-air pollution interaction analysis of lung function in 300,000 individuals. *medRxiv*, 2021.06.03.21256376 (2021).

34. Miller, M.D. & Marty, M.A. Impact of environmental chemicals on lung development. *Environ Health Perspect* **118**, 1155-64 (2010).

35. Decrue, F., Gorlanova, O., Usemann, J. & Frey, U. Lung functional development and asthma trajectories. *Semin Immunopathol* **42**, 17-27 (2020).

36. He, Z. *et al.* The association between secondhand smoke and childhood asthma: A systematic review and meta-analysis. *Pediatr Pulmonol* **55**, 2518-2531 (2020).

37. Thacher, J.D. *et al.* Tobacco smoke exposure in early life and adolescence in relation to lung function. *Eur Respir J* **51**(2018).

38. Dratva, J. *et al.* Early Life Origins of Lung Ageing: Early Life Exposures and Lung Function Decline in Adulthood in Two European Cohorts Aged 28-73 Years. *PLoS One* **11**, e0145127 (2016).

39. Savran, O. & Ulrik, C.S. Early life insults as determinants of chronic obstructive pulmonary disease in adult life. *Int J Chron Obstruct Pulmon Dis* **13**, 683-693 (2018).

40. Zeilinger, S. *et al.* Tobacco smoking leads to extensive genome-wide changes in DNA methylation. *PLoS One* **8**, e63812 (2013).

41. Rider, C.F. & Carlsten, C. Air pollution and DNA methylation: effects of exposure in humans. *Clin Epigenetics* **11**, 131 (2019).

42. Jamieson, E. *et al.* Smoking, DNA Methylation, and Lung Function: a Mendelian Randomization Analysis to Investigate Causal Pathways. *Am J Hum Genet* **106**, 315-326 (2020).

43. Kwak, S.Y., Park, C.Y. & Shin, M.J. Smoking May Affect Pulmonary Function through DNA Methylation: an Epigenome-Wide Association Study in Korean Men. *Clin Nutr Res* **9**, 134-144 (2020).

44. Sunny, S.K. *et al.* Sex-specific longitudinal association of DNA methylation with lung function. *ERJ Open Res* **7**(2021).

45. Mukherjee, N. *et al.* DNA methylation and genetic polymorphisms of the Leptin gene interact to influence lung function outcomes and asthma at 18 years of age. *Int J Mol Epidemiol Genet* **7**, 1-17 (2016).

46. Zhang, H. *et al.* The interplay of DNA methylation over time with Th2 pathway genetic variants on asthma risk and temporal asthma transition. *Clin Epigenetics* **6**, 8 (2014).

47. Munoz-Pizza, D.M., Villada-Canela, M., Reyna, M.A., Texcalac-Sangrador, J.L. & Osornio-Vargas, A.R. Air pollution and children's respiratory health: a scoping review of socioeconomic status as an effect modifier. *Int J Public Health* **65**, 649-660 (2020).

48. Hajizadeh, M. & Nandi, A. The socioeconomic gradient of secondhand smoke exposure in children: evidence from 26 low-income and middle-income countries. *Tob Control* **25**, e146-e155 (2016).

49. Martinez, C.H., Mannino, D.M., Curtis, J.L., Han, M.K. & Diaz, A.A. Socioeconomic Characteristics Are Major Contributors to Ethnic Differences in Health Status in Obstructive Lung Disease: An Analysis of the National Health and Nutrition Examination Survey 2007-2010. *Chest* **148**, 151-158 (2015).

50. Thakur, N. *et al.* Socioeconomic status and childhood asthma in urban minority youths. The GALA II and SAGE II studies. *Am J Respir Crit Care Med* **188**, 1202-9 (2013).

51. Oraka, E., Iqbal, S., Flanders, W.D., Brinker, K. & Garbe, P. Racial and ethnic disparities in current asthma and emergency department visits: findings from the National Health Interview Survey, 2001-2010. *J Asthma* **50**, 488-96 (2013).

52. Keet, C.A., Matsui, E.C., McCormack, M.C. & Peng, R.D. Urban residence, neighborhood poverty, race/ethnicity, and asthma morbidity among children on Medicaid. *J Allergy Clin Immunol* **140**, 822-827 (2017).

53. Manrai, A.K. *et al.* Genetic Misdiagnoses and the Potential for Health Disparities. *N Engl J Med* **375**, 655-65 (2016).

54. Landry, L.G., Ali, N., Williams, D.R., Rehm, H.L. & Bonham, V.L. Lack Of Diversity In Genomic Databases Is A Barrier To Translating Precision Medicine Research Into Practice. *Health Aff (Millwood)* **37**, 780-785 (2018).

55. Pongracic, J.A. *et al.* Distinguishing characteristics of difficult-to-control asthma in inner-city children and adolescents. *J Allergy Clin Immunol* **138**, 1030-1041 (2016).

56. Zoratti, E.M. *et al.* Asthma phenotypes in inner-city children. *J Allergy Clin Immunol* **138**, 1016-1029 (2016).

57. Gern, J.E. *et al.* The Urban Environment and Childhood Asthma (URECA) birth cohort study: design, methods, and study population. *BMC Pulm Med* **9**, 17 (2009).

58. Consortium, E.P. *et al.* Expanded encyclopaedias of DNA elements in the human and mouse genomes. *Nature* **583**, 699-710 (2020).

59. Ma, M. *et al.* Disease-associated variants in different categories of disease located in distinct regulatory elements. *BMC Genomics* **16 Suppl 8**, S3 (2015).

60. Boyle, E.A., Li, Y.I. & Pritchard, J.K. An Expanded View of Complex Traits: From Polygenic to Omnigenic. *Cell* **169**, 1177-1186 (2017).

61. Watanabe, K. *et al.* A global overview of pleiotropy and genetic architecture in complex traits. *Nat Genet* **51**, 1339-1348 (2019).

62. Schulz, H. *et al.* Genome-wide mapping of genetic determinants influencing DNA methylation and gene expression in human hippocampus. *Nat Commun* **8**, 1511 (2017).

63. Huan, T. *et al.* Genome-wide identification of DNA methylation QTLs in whole blood highlights pathways for cardiovascular disease. *Nat Commun* **10**, 4267 (2019).

64. Hannon, E. *et al.* Leveraging DNA-Methylation Quantitative-Trait Loci to Characterize the Relationship between Methylomic Variation, Gene Expression, and Complex Traits. *Am J Hum Genet* **103**, 654-665 (2018).

65. McKennan, C. *et al.* Longitudinal data reveal strong genetic and weak non-genetic components of ethnicity-dependent blood DNA methylation levels. *Epigenetics* **16**, 662-676 (2021).

66. Joubert, B.R. *et al.* DNA Methylation in Newborns and Maternal Smoking in Pregnancy: Genome-wide Consortium Meta-analysis. *Am J Hum Genet* **98**, 680-96 (2016).

67. Sikdar, S. *et al.* Comparison of smoking-related DNA methylation between newborns from prenatal exposure and adults from personal smoking. *Epigenomics* **11**, 1487-1500 (2019).

68. Nicolae, D.L. *et al.* Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS. *PLoS Genet* **6**, e1000888 (2010).

69. Yao, D.W., O'Connor, L.J., Price, A.L. & Gusev, A. Quantifying genetic effects on disease mediated by assayed gene expression levels. *Nat Genet* **52**, 626-633 (2020).

70. Hormozdiari, F. *et al.* Colocalization of GWAS and eQTL Signals Detects Target Genes. *Am J Hum Genet* **99**, 1245-1260 (2016).

71. Helling, B.A. *et al.* Altered transcriptional and chromatin responses to rhinovirus in bronchial epithelial cells from adults with asthma. *Commun Biol* **3**, 678 (2020).

72. Samuels-Lev, Y. *et al.* ASPP proteins specifically stimulate the apoptotic function of p53. *Mol Cell* **8**, 781-94 (2001).

73. Aylon, Y. *et al.* The Lats2 tumor suppressor augments p53-mediated apoptosis by promoting the nuclear proapoptotic function of ASPP1. *Genes Dev* **24**, 2420-9 (2010).

74. Wang, Y. *et al.* ASPP1 and ASPP2 bind active RAS, potentiate RAS signalling and enhance p53 activity in cancer cells. *Cell Death Differ* **20**, 525-34 (2013).

75. Xue, H. & Li, M.X. MicroRNA-150 protects against cigarette smoke-induced lung inflammation and airway epithelial cell apoptosis through repressing p53: MicroRNA-150 in CS-induced lung inflammation. *Hum Exp Toxicol* **37**, 920-928 (2018).

76. Xu, F. *et al.* PM2.5 exposure induces alveolar epithelial cell apoptosis and causes emphysema through p53/Siva-1. *Eur Rev Med Pharmacol Sci* **24**, 3943-3950 (2020).

77. Song, Q., Zhou, Z.J., Cai, S., Chen, Y. & Chen, P. Oxidative stress links the tumour suppressor p53 with cell apoptosis induced by cigarette smoke. *Int J Environ Health Res*, 1-11 (2021).

78. Zhao, K. *et al.* EGR-1/ASPP1 inter-regulatory loop promotes apoptosis by inhibiting cyto-protective autophagy. *Cell Death Dis* **8**, e2869 (2017).

79. Reynolds, P.R., Cosio, M.G. & Hoidal, J.R. Cigarette smoke-induced Egr-1 upregulates proinflammatory cytokines in pulmonary epithelial cells. *Am J Respir Cell Mol Biol* **35**, 314-9 (2006).

80. Chen, Z.H. *et al.* Egr-1 regulates autophagy in cigarette smoke-induced chronic obstructive pulmonary disease. *PLoS One* **3**, e3316 (2008).

81. Shen, N. *et al.* Cigarette smoke-induced pulmonary inflammatory responses are mediated by EGR-1/GGPPS/MAPK signaling. *Am J Pathol* **178**, 110-8 (2011).

82. Wang, S.B. *et al.* Early growth response factor 1 is essential for cigarette smoke-induced MUC5AC expression in human bronchial epithelial cells. *Biochem Biophys Res Commun* **490**, 147-154 (2017).

83. Xu, F. *et al.* Early growth response gene 1 is essential for urban particulate matter-induced inflammation and mucus hyperproduction in airway epithelium. *Toxicol Lett* **294**, 145-155 (2018).

84. Golebski, K. *et al.* EGR-1 as a potential biomarker in asthma and proinflammatory responses in airway epithelium. *European Respiratory Journal* **58**, PA2041 (2021).

85. Wang, A. *et al.* Single-cell multiomic profiling of human lungs reveals cell-type-specific and age-dynamic control of SARS-CoV2 host genes. *Elife* **9**(2020).

86. Karlsson, M. *et al.* A single-cell type transcriptomics map of human tissues. *Sci Adv* **7**(2021).

87. Cheng, Y. *et al.* CircRNA-012091/PPP1R13B-mediated Lung Fibrotic Response in Silicosis via Endoplasmic Reticulum Stress and Autophagy. *Am J Respir Cell Mol Biol* **61**, 380-391 (2019).

88. Vigneron, A.M., Ludwig, R.L. & Vousden, K.H. Cytoplasmic ASPP1 inhibits apoptosis through the control of YAP. *Genes Dev* **24**, 2430-9 (2010).

89. Manfredi, J.J. An identity crisis for a cancer gene: subcellular location determines ASPP1 function. *Cancer Cell* **18**, 409-10 (2010).

90. Fogal, V. *et al.* ASPP1 and ASPP2 are new transcriptional targets of E2F. *Cell Death Differ* **12**, 369-76 (2005).

91. Zhou, S.J. *et al.* Expression variations of connective tissue growth factor in pulmonary arteries from smokers with and without chronic obstructive pulmonary disease. *Sci Rep* **5**, 8564 (2015).

92. Eguchi, A. *et al.* An Altered DNA Methylation Status in the Human Umbilical Cord Is Correlated with Maternal Exposure to Polychlorinated Biphenyls. *Int J Environ Res Public Health* **16**(2019).

93. Pierce, B.L. *et al.* Co-occurring expression and methylation QTLs allow detection of common causal variants and shared biological mechanisms. *Nat Commun* **9**, 804 (2018).

94. Philibert, R.A. *et al.* Coordinated DNA methylation and gene expression changes in smoker alveolar macrophages: specific effects on VEGF receptor 1 expression. *J Leukoc Biol* **92**, 621-31 (2012).

95. Guijo, M., Ceballos-Chavez, M., Gomez-Marin, E., Basurto-Cayuela, L. & Reyes, J.C. Expression of TDRD9 in a subset of lung carcinomas by CpG island hypomethylation protects from DNA damage. *Oncotarget* **9**, 9618-9631 (2018).

96. Consortium, G.T. The Genotype-Tissue Expression (GTEx) project. *Nat Genet* **45**, 580-5 (2013).

97. Lokk, K. *et al.* DNA methylome profiling of human tissues identifies global and tissue-specific methylation patterns. *Genome Biol* **15**, r54 (2014).

98. Wan, J. *et al.* Characterization of tissue-specific differential DNA methylation suggests distinct modes of positive and negative gene expression regulation. *BMC Genomics* **16**, 49 (2015).

99. den Dekker, H.T. *et al.* Newborn DNA-methylation, childhood lung function, and the risks of asthma and COPD across the life course. *Eur Respir J* **53**(2019).

100. Imboden, M. *et al.* Epigenome-wide association study of lung function level and its change. *Eur Respir J* **54**(2019).

101. Mukherjee, N. *et al.* DNA methylation at birth is associated with lung function development until age 26 years. *Eur Respir J* **57**(2021).

102. Wang, T. *et al.* Genome-wide DNA methylation analysis of pulmonary function in middle and old-aged Chinese monozygotic twins. *Respir Res* **22**, 300 (2021).

103. Herrera-Luis, E. *et al.* Epigenome-wide association study of lung function in Latino children and youth with asthma. *Clin Epigenetics* **14**, 9 (2022).

104. Cosin-Tomas, M., Bustamante, M. & Sunyer, J. Epigenetic association studies at birth and the origin of lung function development. *Eur Respir J* **57**(2021).

105. Yang, I.V., Lozupone, C.A. & Schwartz, D.A. The environment, epigenome, and asthma. *J Allergy Clin Immunol* **140**, 14-23 (2017).

106. Lin, P.I., Shu, H. & Mersha, T.B. Comparing DNA methylation profiles across different tissues associated with the diagnosis of pediatric asthma. *Sci Rep* **10**, 151 (2020).

107. Li, Y.F. *et al.* Effects of in utero and environmental tobacco smoke exposure on lung function in boys and girls with and without asthma. *Am J Respir Crit Care Med* **162**, 2097-104 (2000).

108. Gilliland, F.D., Berhane, K., Li, Y.F., Rappaport, E.B. & Peters, J.M. Effects of early onset asthma and in utero exposure to maternal smoking on childhood lung function. *Am J Respir Crit Care Med* **167**, 917-24 (2003).

109. Schultz, E.S., Litonjua, A.A. & Melen, E. Effects of Long-Term Exposure to Traffic-Related Air Pollution on Lung Function in Children. *Curr Allergy Asthma Rep* **17**, 41 (2017).

110. Kreiner-Moller, E., Bisgaard, H. & Bonnelykke, K. Prenatal and postnatal genetic influence on lung function development. *J Allergy Clin Immunol* **134**, 1036-42 e15 (2014).

111. Lee, E.Y. *et al.* Whole-Genome Sequencing Identifies Novel Functional Loci Associated with Lung Function in Puerto Rican Youth. *Am J Respir Crit Care Med* **202**, 962-972 (2020).

112. Goddard, P.C. *et al.* Integrative genomic analysis in African American children with asthma finds three novel loci associated with lung function. *Genet Epidemiol* **45**, 190-208 (2021).

113. Akenroye, A.T. *et al.* Genome-wide association study of asthma, total IgE, and lung function in a cohort of Peruvian children. *J Allergy Clin Immunol* **148**, 1493-1504 (2021).

114. Imboden, M. *et al.* Genome-wide association study of lung function decline in adults with and without asthma. *J Allergy Clin Immunol* **129**, 1218-28 (2012).

115. Jones, M.J., Goodman, S.J. & Kobor, M.S. DNA methylation and healthy human aging. *Aging Cell* **14**, 924-32 (2015).

116. Ambatipudi, S. *et al.* Tobacco smoking-associated genome-wide DNA methylation changes in the EPIC study. *Epigenomics* **8**, 599-618 (2016).

117. Tommasi, S., Zheng, A. & Besaratinia, A. Exposure of mice to secondhand smoke elicits both transient and long-lasting transcriptional changes in cancer-related functional networks. *Int J Cancer* **136**, 2253-63 (2015).

118. Sridhar, S. *et al.* Smoking-induced gene expression changes in the bronchial airway are reflected in nasal and buccal epithelium. *BMC Genomics* **9**, 259 (2008).

119. Zhang, X. *et al.* Similarities and differences between smoking-related gene expression in nasal and bronchial epithelium. *Physiol Genomics* **41**, 1-8 (2010).

120. Brugha, R. *et al.* DNA methylation profiles between airway epithelium and proxy tissues in children. *Acta Paediatrica* **106**, 2011-2016 (2017).

121. Imkamp, K. *et al.* Nasal epithelium as a proxy for bronchial epithelium for smoking-induced gene expression and expression Quantitative Trait Loci. *J Allergy Clin Immunol* **142**, 314-317 e15 (2018).

122. Kicic, A. *et al.* Assessing the unified airway hypothesis in children via transcriptional profiling of the airway epithelium. *J Allergy Clin Immunol* **145**, 1562-1573 (2020).

123. Bergougnoux, A., Claustres, M. & De Sario, A. Nasal epithelial cells: a tool to study DNA methylation in airway diseases. *Epigenomics* **7**, 119-26 (2015).

124. Gergen, P.J., Teach, S.J., Togias, A. & Busse, W.W. Reducing Exacerbations in the Inner City: Lessons from the Inner-City Asthma Consortium (ICAC). *J Allergy Clin Immunol Pract* **4**, 22-6 (2016).

125. O'Connor, G.T. *et al.* Early-life home environment and risk of asthma among inner-city children. *J Allergy Clin Immunol* **141**, 1468-1475 (2018).

126. Quanjer, P.H. *et al.* Multi-ethnic reference values for spirometry for the 3-95-yr age range: the global lung function 2012 equations. *Eur Respir J* **40**, 1324-43 (2012).

127. Bernert, J.T., Harmon, T.L., Sosnoff, C.S. & McGuffey, J.E. Use of continine immunoassay test strips for preclassifying urine samples from smokers and nonsmokers prior to analysis by LC-MS-MS. *J Anal Toxicol* **29**, 814-8 (2005).

128. Regier, A.A. *et al.* Functional equivalence of genome sequencing analysis pipelines enables harmonized variant calling across human genetics projects. *Nat Commun* **9**, 4038 (2018).

129. Kowalski, M.H. *et al.* Use of >100,000 NHLBI Trans-Omics for Precision Medicine (TOPMed) Consortium whole genome sequences improves imputation quality and detection of rare variant associations in admixed African and Hispanic/Latino populations. *PLoS Genet* **15**, e1008500 (2019).

130. Zhang, F. *et al.* Ancestry-agnostic estimation of DNA sample contamination from sequence reads. *Genome Res* **30**, 185-194 (2020).

131. Chang, C.C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015).

132. McKennan, C. *et al.* Longitudinal data reveal strong genetic and weak non-genetic components of ethnicity-dependent blood DNA methylation levels. *Epigenetics*, 1-15 (2020).

133. Jun, G. *et al.* Detecting and estimating contamination of human DNA samples in sequencing and array-based genotype data. *Am J Hum Genet* **91**, 839-48 (2012).

134. Hao, W. & Storey, J.D. Extending Tests of Hardy-Weinberg Equilibrium to Structured Populations. *Genetics* **213**, 759-770 (2019).

135. Danecek, P. *et al.* The variant call format and VCFtools. *Bioinformatics* **27**, 2156-8 (2011).

136. Danecek, P. *et al.* Twelve years of SAMtools and BCFtools. *Gigascience* **10**(2021).

137. Genomes Project, C. *et al.* A global reference for human genetic variation. *Nature* **526**, 68-74 (2015).

138. Bergstrom, A. *et al.* Insights into human genetic variation and population history from 929 diverse genomes. *Science* **367**(2020).

139. Alexander, D.H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res* **19**, 1655-64 (2009).

140. Conomos, M.P., Miller, M.B. & Thornton, T.A. Robust inference of population structure for ancestry prediction and correction of stratification in the presence of relatedness. *Genet Epidemiol* **39**, 276-93 (2015).

141. Conomos, M.P., Reiner, A.P., Weir, B.S. & Thornton, T.A. Model-free Estimation of Recent Genetic Relatedness. *Am J Hum Genet* **98**, 127-48 (2016).

142. Manichaikul, A. *et al.* Robust relationship inference in genome-wide association studies. *Bioinformatics* **26**, 2867-73 (2010).

143. Shringarpure, S.S., Bustamante, C.D., Lange, K. & Alexander, D.H. Efficient analysis of large datasets and sex bias with ADMIXTURE. *BMC Bioinformatics* **17**, 218 (2016).

144. Sofer, T. *et al.* A fully adjusted two-stage procedure for rank-normalization in genetic association studies. *Genet Epidemiol* **43**, 263-275 (2019).

145. McCaw, Z.R., Lane, J.M., Saxena, R., Redline, S. & Lin, X. Operating characteristics of the rank-based inverse normal transformation for quantitative trait analysis in genome-wide association studies. *Biometrics* **76**, 1262-1272 (2020).

146. Zhou, X. & Stephens, M. Genome-wide efficient mixed-model analysis for association studies. *Nat Genet* **44**, 821-4 (2012).

147. Shabalin, A.A. Matrix eQTL: ultra fast eQTL analysis via large matrix operations. *Bioinformatics* **28**, 1353-8 (2012).

148. McKennan, C. & Nicolae, D. Accounting for unobserved covariates with varying degrees of estimability in high-dimensional biological data. *Biometrika* **106**, 823-840 (2019).

149. Zhang, D. A Coefficient of Determination for Generalized Linear Models. *The American Statistician* **71**, 310-316 (2017).

150. Fox, J., Kleiber, C. & Zeileis, A. ivreg: Two-Stage Least-Squares Regression with Diagnostics. (https://john-d-fox.github.io/ivreg/, 2021).

151. Alfons, A., Ateş, N.Y. & Groenen, P.J.F. A Robust Bootstrap Test for Mediation Analysis. *Organizational Research Methods* **0**, 1094428121999096.

152. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21 (2013).

153. Law, C.W., Chen, Y., Shi, W. & Smyth, G.K. voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol* **15**, R29 (2014).

154. Ritchie, M.E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* **43**, e47 (2015).

155. Jager, R. *et al.* Capture Hi-C identifies the chromatin interactome of colorectal cancer risk loci. *Nat Commun* **6**, 6178 (2015).

156. Mifsud, B. *et al.* Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C. *Nat Genet* **47**, 598-606 (2015).

157. Montefiori, L.E. *et al.* A promoter interaction map for cardiovascular disease genetics. *Elife* **7**(2018).

158. Cairns, J. *et al.* CHiCAGO: robust detection of DNA looping interactions in Capture Hi-C data. *Genome Biol* **17**, 127 (2016).

SUPPORTING INFORMATION

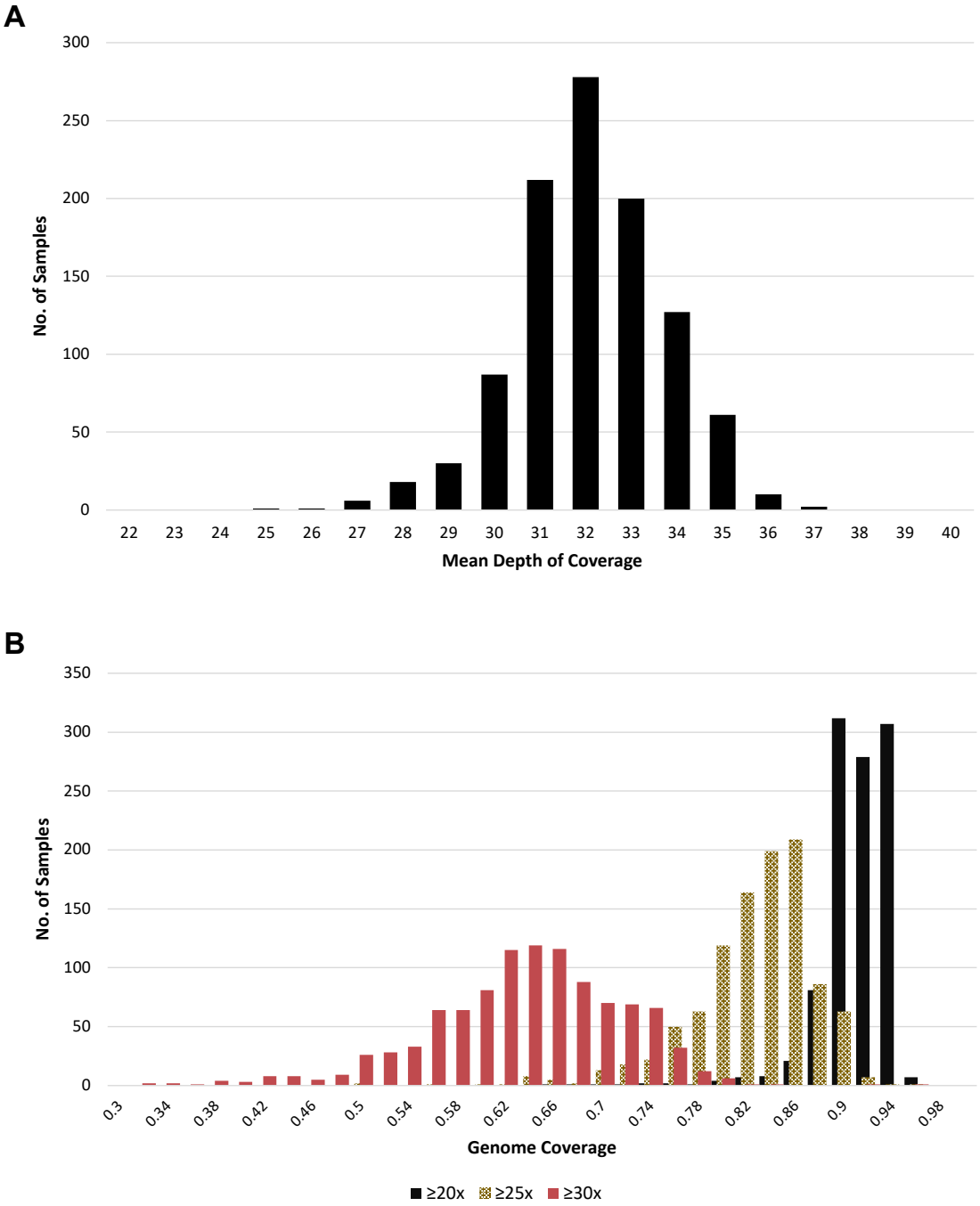

Fig S1. Whole-genome sequencing depth and coverage. A) Histogram of 1,035 whole-genome sequencing (WGS) samples from APIC and URECA by mean depth of coverage. **B**) Histogram of WGS samples based on proportion of genome covered at 20x, 25x, and 30x depth. APIC, Asthma Phenotypes in the Inner City study; URECA, Urban Environment and Childhood Asthma study.

Table S1. Post-QC sequencing call concordance between replicates.

Variant Type	Variants (mean)	Call concordance
All Variants	4,284,228	0.9985
SNPs	3,793,694	0.9997
Common SNPs (MAF > 0.05)	3,244,661	0.9997
Low Frequency SNPs (0.01 > MAF ≤ 0.05)	383,918	0.9995
Rare SNPs (MAF ≤0.01)	165,115	0.9995
InDels	490,534	0.9892
Common InDels (MAF > 0.05)	428,621	0.9896
Low Frequency InDels (0.01 > MAF ≤ 0.05)	39,572	0.9845
Rare InDels (MAF ≤0.01)	22,340	0.9911

Variant call concordance between three pairs of replicate samples, by variant type and cohort allele frequency. SNPs, single nucleotide polymorphisms; MAF, minor allele frequency; InDels, insertions and deletions.

Table S2. FEV_1-associated variants in chr14q32.33

Pos	rsID	N	Ref	Alt	MAF	Beta	SE	P	Type	Gene
103936818	**rs10220464**	**896**	**A**	**G**	**0.30**	**-0.311**	**0.052**	**2.42 x10^{-9}**	**intronic**	***TDRD9***
103926863	rs11850186	896	G	A	0.29	-0.307	0.052	5.34 x10^{-9}	intergenic	*ATP5MPL; TDRD9*
103934654	rs4906387	896	C	T	0.29	-0.305	0.052	6.20 x10^{-9}	intronic	*TDRD9*
103935736	rs9285602	896	C	T	0.29	-0.305	0.052	6.20 x10^{-9}	intronic	*TDRD9*
103937635	rs9324069	896	G	A	0.29	-0.305	0.052	6.20 x10^{-9}	intronic	*TDRD9*
103942966	.	896	GA	G	0.29	-0.305	0.052	6.20 x10^{-9}	intronic	*TDRD9*
103960917	rs4900604	896	G	A	0.29	-0.305	0.052	6.20 x10^{-9}	intronic	*TDRD9*
103963565	rs28574832	896	G	T	0.29	-0.305	0.052	6.20 x10^{-9}	intronic	*TDRD9*
103967867	rs4900605	896	G	T	0.29	-0.305	0.052	6.20 x10^{-9}	intronic	*TDRD9*
103997188	rs72712900	896	G	A	0.29	-0.304	0.052	6.51 x10^{-9}	intronic	*TDRD9*
103984806	rs11847797	896	G	A	0.29	-0.303	0.052	7.73 x10^{-9}	intronic	*TDRD9*
103980079	rs58289480	896	A	G	0.29	-0.303	0.052	8.28 x10^{-9}	intronic	*TDRD9*
103985095	rs4906396	896	A	C	0.29	-0.303	0.052	8.28 x10^{-9}	intronic	*TDRD9*
103953683	rs72712884	896	C	G	0.29	-0.303	0.052	8.28 x10^{-9}	intronic	*TDRD9*
104046654	rs143594477	896	C	CT	0.29	-0.302	0.052	9.12 x10^{-9}	intronic	*TDRD9*
104028226	rs10137997	896	T	C	0.29	-0.301	0.052	9.73 x10^{-9}	intronic	*TDRD9*
104034493	rs9944163	896	G	T	0.29	-0.301	0.052	9.73 x10^{-9}	intronic	*TDRD9*
104069368	rs72714937	896	C	T	0.29	-0.302	0.052	9.80 x10^{-9}	intergenic	*TDRD9; ASPG*
104059833	rs11160785	896	T	A	0.29	-0.300	0.052	1.03 x10^{-8}	intergenic	*TDRD9; ASPG*
103934222	rs28461376	896	A	G	0.29	-0.301	0.052	1.06 x10^{-8}	intronic	*TDRD9*
103963484	.	896	TC	T	0.29	-0.301	0.052	1.06 x10^{-8}	intronic	*TDRD9*
104037291	rs4906408	896	G	T	0.29	-0.299	0.052	1.14 x10^{-8}	intronic	*TDRD9*
104030861	rs10135338	896	G	A	0.29	-0.299	0.052	1.18 x10^{-8}	intronic	*TDRD9*
104035861	rs10134394	896	G	A	0.29	-0.299	0.052	1.18 x10^{-8}	intronic	*TDRD9*
104036696	rs11160780	896	G	A	0.29	-0.299	0.052	1.18 x10^{-8}	intronic	*TDRD9*
104037342	rs4900608	896	G	A	0.29	-0.299	0.052	1.18 x10^{-8}	intronic	*TDRD9*
104038178	rs4906409	896	T	C	0.29	-0.299	0.052	1.18 x10^{-8}	intronic	*TDRD9*
104039412	rs10144682	896	G	A	0.29	-0.299	0.052	1.18 x10^{-8}	intronic	*TDRD9*
104012687	rs11851441	896	A	G	0.29	-0.299	0.052	1.23 x10^{-8}	intronic	*TDRD9*
104031206	rs10135507	896	C	T	0.29	-0.298	0.052	1.25 x10^{-8}	exonic	*TDRD9*
103998027	rs55938939	896	G	A	0.29	-0.299	0.052	1.26 x10^{-8}	intronic	*TDRD9*
104048481	rs61244535	896	C	G	0.29	-0.299	0.052	1.29 x10^{-8}	intronic	*TDRD9*
104051127	rs28644198	896	C	A	0.29	-0.299	0.052	1.29 x10^{-8}	intronic	*TDRD9*
104054541	rs28377615	896	G	A	0.29	-0.299	0.052	1.29 x10^{-8}	intergenic	*TDRD9; ASPG*
103997438	rs4906397	896	C	T	0.29	-0.298	0.052	1.31 x10^{-8}	intronic	*TDRD9*
103999081	rs4906399	896	C	T	0.29	-0.298	0.052	1.31 x10^{-8}	intronic	*TDRD9*
104005628	rs12147936	896	C	T	0.29	-0.298	0.052	1.31 x10^{-8}	intronic	*TDRD9*
104010259	rs145867870	896	C	CT	0.29	-0.298	0.052	1.31 x10^{-8}	intronic	*TDRD9*
104013496	rs11851723	896	C	T	0.29	-0.298	0.052	1.31 x10^{-8}	intronic	*TDRD9*
104014746	rs11160779	896	T	C	0.29	-0.298	0.052	1.31 x10^{-8}	exonic	*TDRD9*
104016537	rs72714909	896	C	A	0.29	-0.298	0.052	1.31 x10^{-8}	intronic	*TDRD9*
103986068	rs10143030	896	C	G	0.29	-0.300	0.052	1.41 x10^{-8}	intronic	*TDRD9*
103986303	rs10143389	896	G	A	0.29	-0.300	0.052	1.41 x10^{-8}	exonic	*TDRD9*
103986825	rs61248168	896	T	A	0.29	-0.300	0.052	1.41 x10^{-8}	intronic	*TDRD9*
103990858	rs1957518	896	G	A	0.29	-0.300	0.052	1.41 x10^{-8}	intronic	*TDRD9*
104072329	rs111576189	896	G	A	0.29	-0.298	0.052	1.54 x10^{-8}	intergenic	*TDRD9; ASPG*
104011541	rs28522352	896	T	A	0.29	-0.296	0.052	1.79 x10^{-8}	intronic	*TDRD9*
104003975	rs8010286	896	T	C	0.29	-0.293	0.052	2.06 x10^{-8}	intronic	*TDRD9*
104004349	rs28391043	896	G	A	0.29	-0.294	0.052	2.22 x10^{-8}	intronic	*TDRD9*
104027784	rs28725314	896	C	T	0.29	-0.290	0.052	2.96 x10^{-8}	intronic	*TDRD9*
103970594	rs11851097	896	C	T	0.36	-0.267	0.048	3.28 x10^{-8}	exonic	*TDRD9*

Pos	rsID	N	Ref	Alt	MAF	Beta	SE	P	Type	Gene
103987132	rs200735321	880	A	G	0.26	-0.311	0.057	6.98×10^{-8}	intronic	*TDRD9*
104067149	rs10782497	896	T	C	0.39	-0.265	0.049	1.08×10^{-7}	intergenic	*TDRD9; ASPG*
104019632	rs74089113	896	G	C	0.44	-0.246	0.046	1.25×10^{-7}	intronic	*TDRD9*
104071795	rs35775205	896	A	G	0.37	-0.264	0.050	1.72×10^{-7}	intergenic	*TDRD9; ASPG*
104038358	rs877009	896	T	C	0.43	-0.240	0.046	3.06×10^{-7}	intronic	*TDRD9*
104007407	rs12100528	896	G	A	0.43	-0.240	0.047	3.32×10^{-7}	intronic	*TDRD9*
104016337	rs7158223	896	G	T	0.42	-0.238	0.046	3.83×10^{-7}	intronic	*TDRD9*
103947548	rs7144813	896	G	C	0.35	-0.244	0.049	8.48×10^{-7}	intronic	*TDRD9*
103937851	rs10141985	896	G	A	0.34	-0.242	0.050	1.53×10^{-6}	intronic	*TDRD9*
103939314	rs7148877	896	A	G	0.34	-0.241	0.050	1.73×10^{-6}	intronic	*TDRD9*
103945800	rs12147460	896	T	G	0.34	-0.241	0.050	1.73×10^{-6}	intronic	*TDRD9*
104032500	rs9944058	896	C	T	0.42	-0.235	0.049	2.22×10^{-6}	intronic	*TDRD9*
103981884	rs4906395	896	T	C	0.42	-0.220	0.047	2.71×10^{-6}	intronic	*TDRD9*
103958246	rs10149734	896	A	C	0.41	-0.221	0.047	3.04×10^{-6}	intronic	*TDRD9*
103962412	rs72712889	896	T	C	0.41	-0.221	0.047	3.04×10^{-6}	intronic	*TDRD9*
103967964	rs4906392	896	T	C	0.41	-0.221	0.047	3.04×10^{-6}	intronic	*TDRD9*
103995217	rs10132153	896	C	A	0.41	-0.220	0.047	3.16×10^{-6}	intronic	*TDRD9*
104036896	rs7156122	896	G	A	0.41	-0.217	0.046	3.28×10^{-6}	intronic	*TDRD9*
104014863	rs397949797	896	A	AT	0.41	-0.216	0.046	3.75×10^{-6}	intronic	*TDRD9*
103979872	rs8022833	896	A	G	0.41	-0.219	0.047	3.87×10^{-6}	intronic	*TDRD9*
103981060	rs1957517	896	A	G	0.41	-0.219	0.047	3.87×10^{-6}	intronic	*TDRD9*
103995157	rs10132127	896	C	G	0.41	-0.218	0.047	4.02×10^{-6}	intronic	*TDRD9*
104004783	rs9671921	896	A	G	0.49	-0.214	0.046	4.98×10^{-6}	intronic	*TDRD9*
104008697	rs4906401	896	A	G	0.41	-0.215	0.047	5.23×10^{-6}	intronic	*TDRD9*
104001342	rs61319604	896	C	T	0.41	-0.215	0.047	5.35×10^{-6}	intronic	*TDRD9*
104002208	rs72714904	896	T	C	0.41	-0.215	0.047	5.35×10^{-6}	intronic	*TDRD9*
104003206	rs10135296	896	G	T	0.41	-0.215	0.047	5.35×10^{-6}	intronic	*TDRD9*
104006916	rs4900607	896	A	G	0.49	-0.212	0.046	5.90×10^{-6}	intronic	*TDRD9*
104010388	rs4906402	896	G	A	0.41	-0.213	0.047	6.22×10^{-6}	intronic	*TDRD9*
104000834	rs7160557	896	G	A	0.48	-0.207	0.046	8.97×10^{-6}	intronic	*TDRD9*
104078948	rs4906415	896	G	A	0.19	-0.268	0.060	9.00×10^{-6}	intergenic	*TDRD9; ASPG*

All variants in chr14q32.33 associated with FEV_1 (% predicted) with $p<1\times10^{-5}$ (n=82) in GWAS of 896 participants from APIC & URECA. N, number of genotyped individuals. MAF, minor allele frequency; 95% CI, 95% confidence interval; SE, standard error; P, P-value (Wald); FEV_1, forced expiratory volume in one second; APIC, Asthma Phenotypes in the Inner City study; URECA, Urban Environment and Childhood Asthma study.

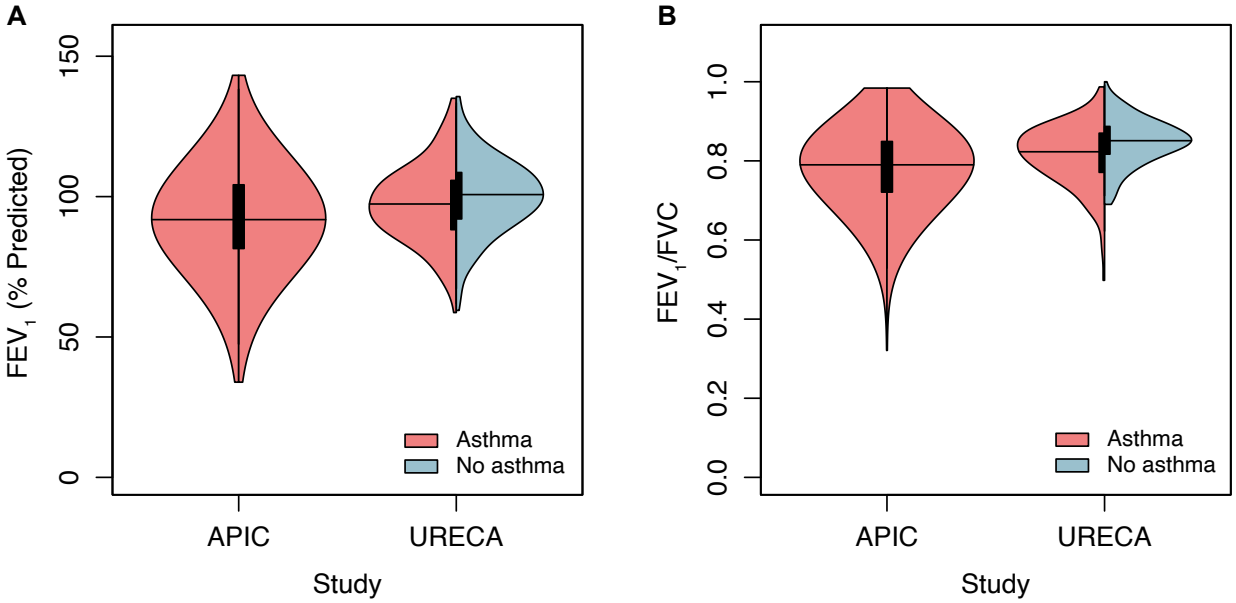

Fig S2. **Distribution of lung function measures by study. A)** Distribution of FEV_1 (% predicted) in APIC and URECA. **B)** Distribution of FEV_1/FVC in APIC and URECA. APIC, Asthma Phenotypes in the Inner City study; URECA, Urban Environment and Childhood Asthma study. FEV_1, forced expiratory volume in one second; FVC, forced vital capacity.

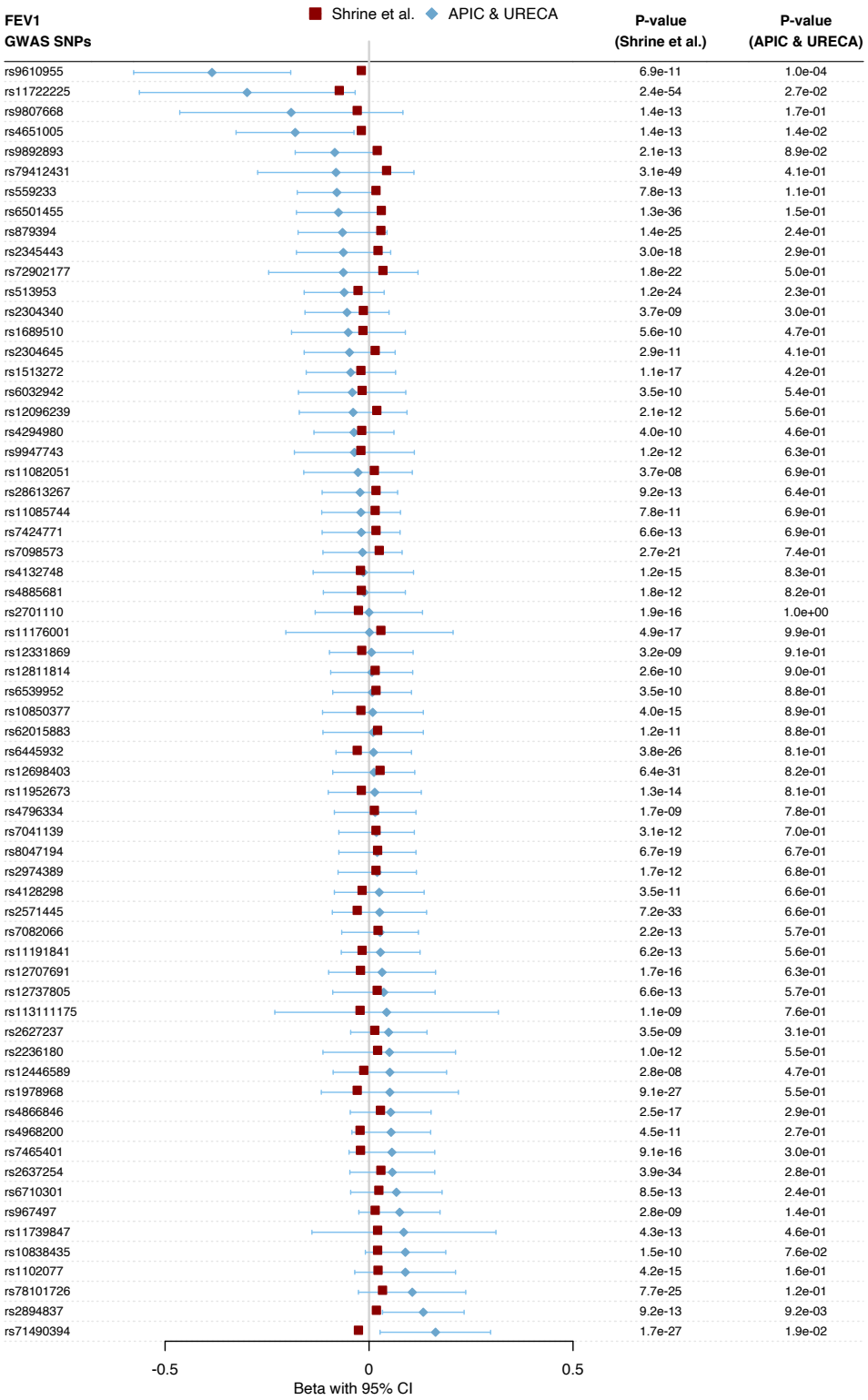

Fig S3. Replication of FEV_1 GWAS SNPs. Association statistics for previously identified FEV_1 GWAS SNPs [23]. 64 out of 70 previously identified SNPs were genotyped in APIC & URECA. GWAS, genome-wide association study; SNP, single nucleotide polymorphism; APIC, Asthma Phenotypes in the Inner City study; URECA, Urban Environment and Childhood Asthma study. FEV_1, forced expiratory volume in one second.

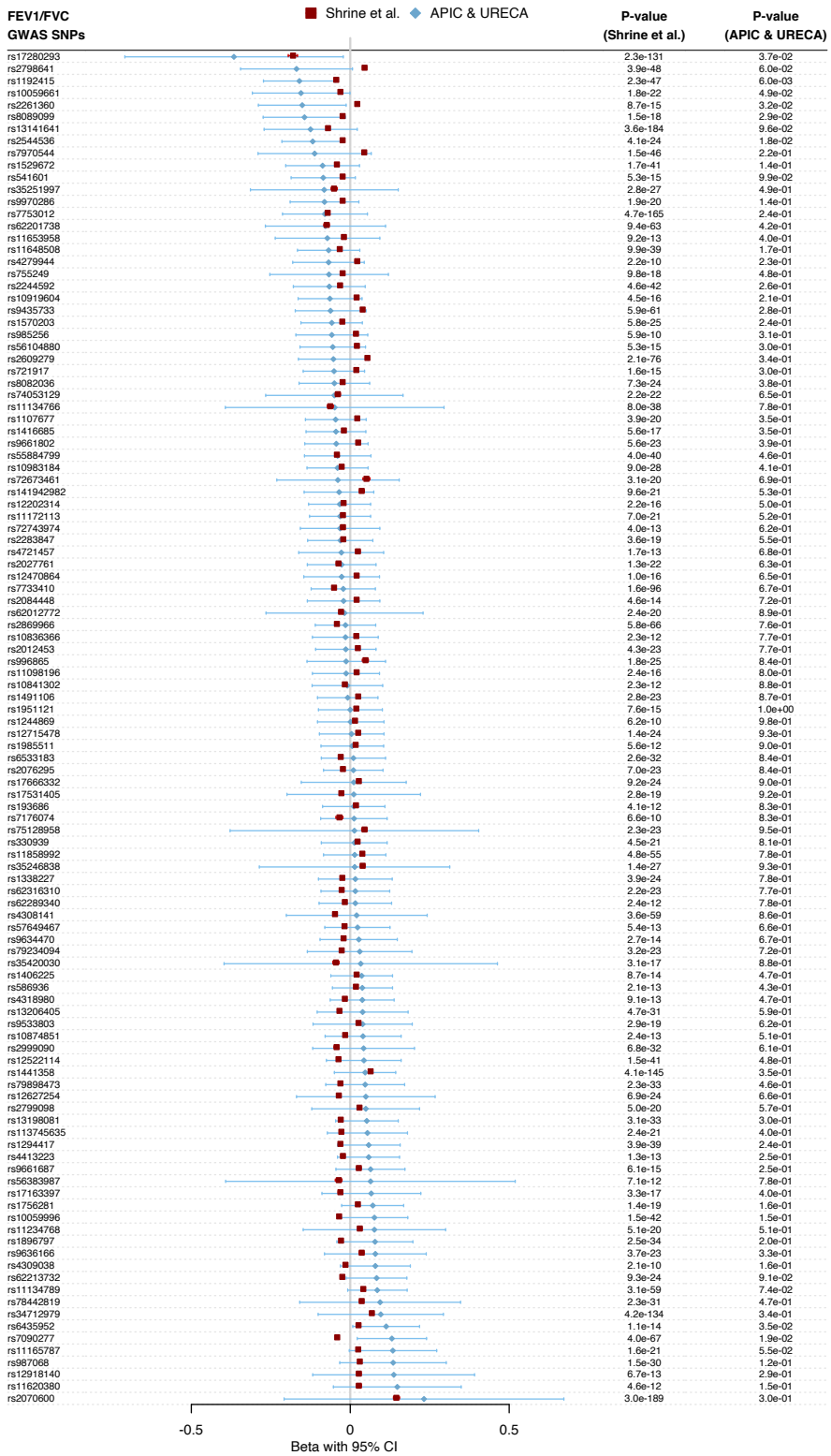

Fig S4. Replication of FEV_1/FVC GWAS SNPs. Association statistics for previously identified FEV_1/FVC GWAS SNPs [23]. 112 out of 117 previously identified SNPs were genotyped in APIC & URECA. GWAS, genome-wide association study; SNP, single nucleotide polymorphism; APIC, Asthma Phenotypes in the Inner City study; URECA, Urban Environment and Childhood Asthma study. FEV_1, forced expiratory volume in one second; FVC, forced vital capacity.

Table S3. MeQTL analysis results and associations with FEV_1

CpG Site	Position (strand), hg38 (chr14)	rs10220464			FEV_1	
		Beta [95% CI]	P	FDR Q	Beta [95% CI]	P
cg21567958	103909862 (-)	0.10 [0.05, 0.14]	1.79×10^{-5}	0.01	-1.5 [-8.8, 5.7]	0.68
cg12183467	103885907 (-)	0.10 [0.05, 0.14]	1.31×10^{-4}	0.04	5.3 [-1.5, 12.1]	0.13
cg16820107	103970549 (-)	0.08 [0.04, 0.12]	1.62×10^{-4}	0.04	-5.4 [-13.4, 2.5]	0.18
cg03306306	**104001397 (+)**	**0.07 [0.03, 0.11]**	**2.28×10^{-4}**	**0.04**	**-11.5 [-20.3, -2.7]**	**0.01**
cg17298714	103938355 (+)	0.14 [0.07, 0.21]	2.36×10^{-4}	0.04	-4.1 [-8.7, 0.4]	0.07

All CpG sites where DNA methylation levels in NECs at age 11 in URECA were associated with rs10220464 at FDR<0.05 are shown with their corresponding associations with FEV_1. The FDR-adjusted P-values (FDR Q) correspond to a 5% false-discovery rate. FDR, false discovery rate; 95% CI, 95% confidence interval; FEV_1, forced expiratory volume in one second; URECA, Urban Environment and Childhood Asthma study.

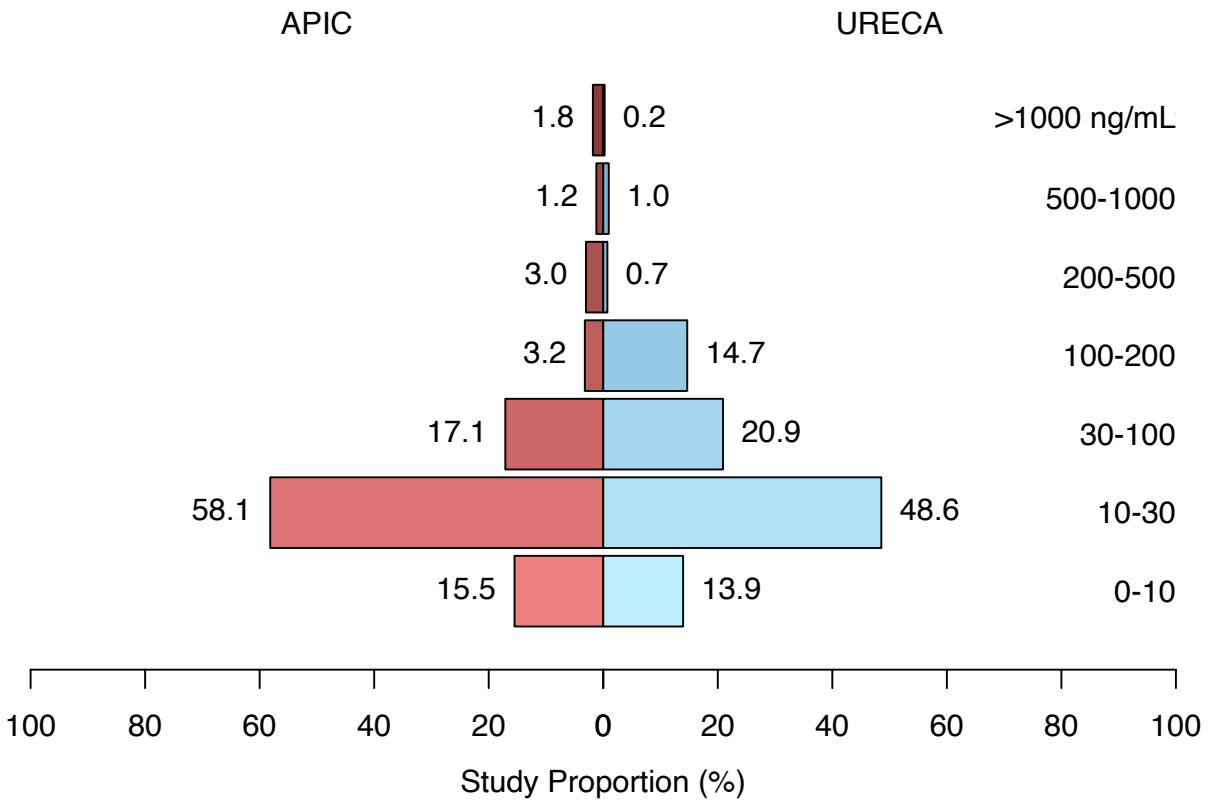

Fig S5. NicAlert Results by Study. Distribution of urine cotinine levels, as measured using NicAlert immunochromatographic assays, which report results on a scale of 0-6 according to the labeled concentration ranges. Proportions were calculated relative to the number of samples with available NicAlert results. APIC, Asthma Phenotypes in the Inner City study; URECA, Urban Environment and Childhood Asthma study.

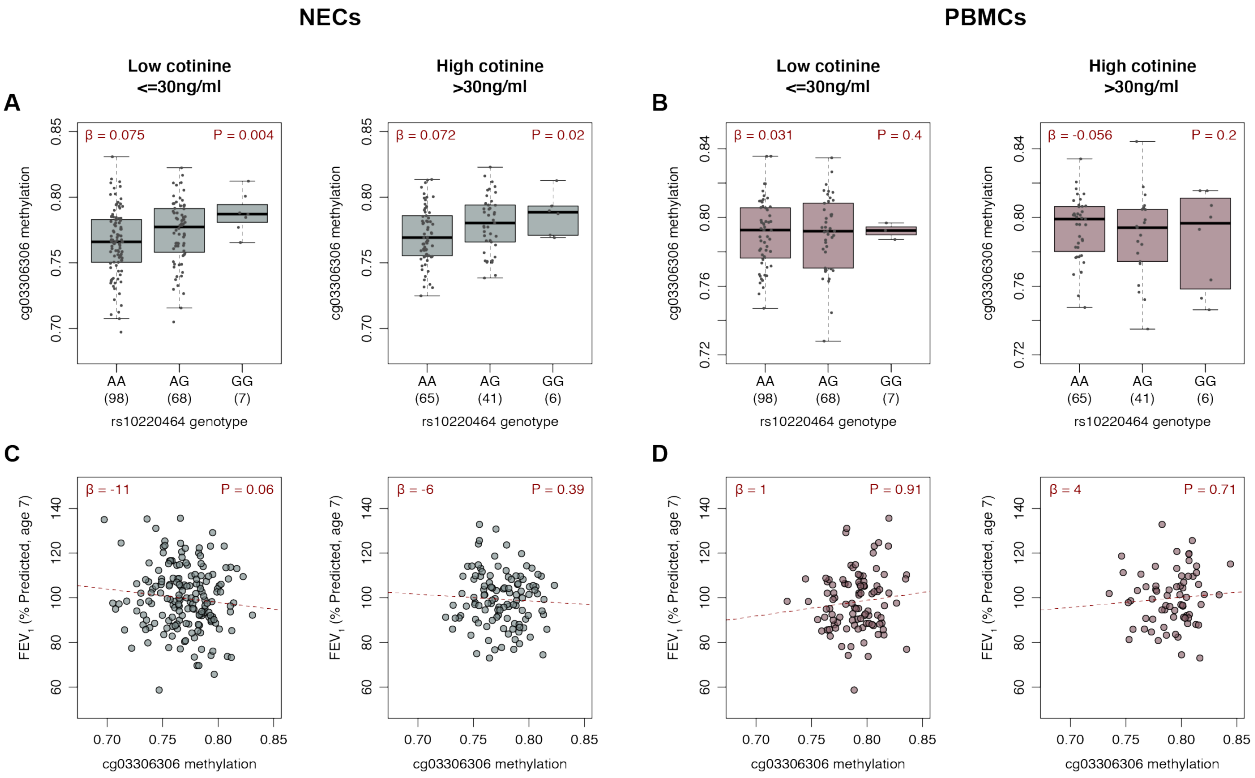

Fig S6. DNA methylation at cg03306306 by smoking exposure. DNA methylation levels at cg03306306 are shown by rs10220464 genotype in URECA participants with low and high smoking exposures in (**A**) NECs at age 11 and (**B**) PBMCs at age 7. FEV_1 (% predicted) are also shown by cg03306306 DNA methylation levels in URECA participants with low and high smoking exposures in (**C**) NECs at age 11 and (**D**) PBMCs at age 7. NECs, nasal epithelial cells; PBMCs, peripheral blood mononuclear cells; FEV_1, forced expiratory volume in one second; URECA, Urban Environment and Childhood Asthma study.

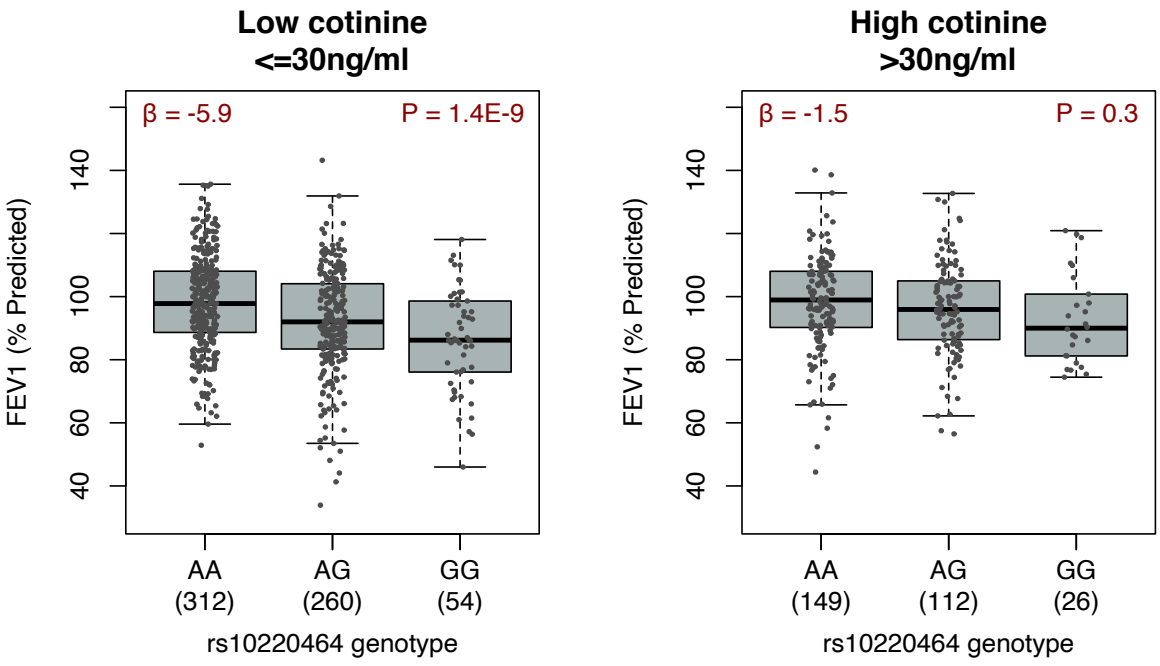

Fig S7. Genotype associations with FEV_1 by smoking exposure. FEV_1 (% predicted) are shown by rs10220464 genotype in APIC & URECA participants with low and high smoking exposures according to urine cotinine levels. FEV_1, forced expiratory volume in one second; APIC, Asthma Phenotypes in the Inner City study; URECA, Urban Environment and Childhood Asthma study.

Table S4. rs10220464 eQTL analysis results

Gene	NECs			PBMCs	
	Beta [95% CI]	P	FDR Q	Beta [95% CI]	P
PPP1R13B	**0.12 [0.06, 0.17]**	**1.26 x10^{-5}**	**2.77 x10^{-4}**	0.01 [-0.08, 0.09]	0.87
BAG5	-0.06 [-0.02, -0.10]	0.0054	0.059	0.03 [-0.03, 0.08]	0.31
ASPG	0.12 [-0.02, 0.26]	0.089	0.48	NOT EXPRESSED	
CKB	-0.04 [0.01, -0.09]	0.10	0.48	0.13 [-0.03, 0.30]	0.12
KLC1	-0.04 [0.01, -0.08]	0.12	0.48	0.02 [-0.04, 0.07]	0.53
TNFAIP2	-0.05 [0.02, -0.12]	0.15	0.48	-0.11 [0.06, -0.28]	0.22
MARK3	-0.03 [0.01, -0.07]	0.15	0.48	-0.02 [0.02, -0.06]	0.40
C14orf2	0.01 [-0.01, 0.03]	0.23	0.61	-0.03 [0.03, -0.09]	0.35
XRCC3	0.08 [-0.06, 0.22]	0.25	0.61	0.00 [0.06, -0.07]	0.92
ZFYVE21	0.02 [-0.01, 0.04]	0.28	0.61	-0.02 [0.03, -0.07]	0.49
ADSSL1	0.04 [-0.04, 0.11]	0.33	0.61	0.04 [-0.09, 0.17]	0.54
PLD4	0.07 [-0.07, 0.20]	0.33	0.61	0.07 [-0.12, 0.26]	0.47
SIVA1	-0.01 [0.01, -0.03]	0.36	0.61	0.01 [-0.06, 0.09]	0.77
EXOC3L4	-0.08 [0.14, -0.29]	0.47	0.69	NOT EXPRESSED	
TRMT61A	-0.02 [0.04, -0.09]	0.48	0.69	-0.04 [0.04, -0.12]	0.32
CEP170B	-0.02 [0.05, -0.09]	0.51	0.69	0.02 [-0.11, 0.14]	0.78
APOPT1	-0.02 [0.04, -0.08]	0.56	0.69	-0.03 [0.03, -0.09]	0.37
EIF5	-0.01 [0.02, -0.04]	0.60	0.69	-0.03 [0.05, -0.12]	0.44
TDRD9	-0.07 [0.21, -0.36]	0.60	0.69	0.01 [-0.23, 0.26]	0.91
ZBTB42	-0.02 [0.05, -0.08]	0.63	0.69	-0.09 [0.01, -0.19]	0.089
INF2	-0.01 [0.08, -0.10]	0.84	0.85	0.04 [-0.06, 0.13]	0.45
AKT1	0.00 [0.03, -0.03]	0.85	0.85	-0.02 [0.04, -0.08]	0.57
AL049840.1	NOT EXPRESSED			0.00 [0.08, -0.09]	0.93
AL049840.4	NOT EXPRESSED			0.01 [-0.05, 0.08]	0.65
AL049840.5	NOT EXPRESSED			0.02 [-0.08, 0.13]	0.68
AL133367.1	NOT EXPRESSED			0.09 [-0.02, 0.20]	0.12
LINC00638	NOT EXPRESSED			0.03 [-0.11, 0.18]	0.66

Results of eQTL analyses in NECs and PBMCs with rs10220464 for all genes within 1 Mb in URECA. Gene expression was measured in counts per million mapped reads. The FDR-adjusted P-values (FDR Q) correspond to a 5% false-discovery rate. FDR, false discovery rate; 95% CI, 95% confidence interval; NECs, nasal epithelial cells; PBMCs, peripheral blood mononuclear cells; URECA, Urban Environment and Childhood Asthma study.

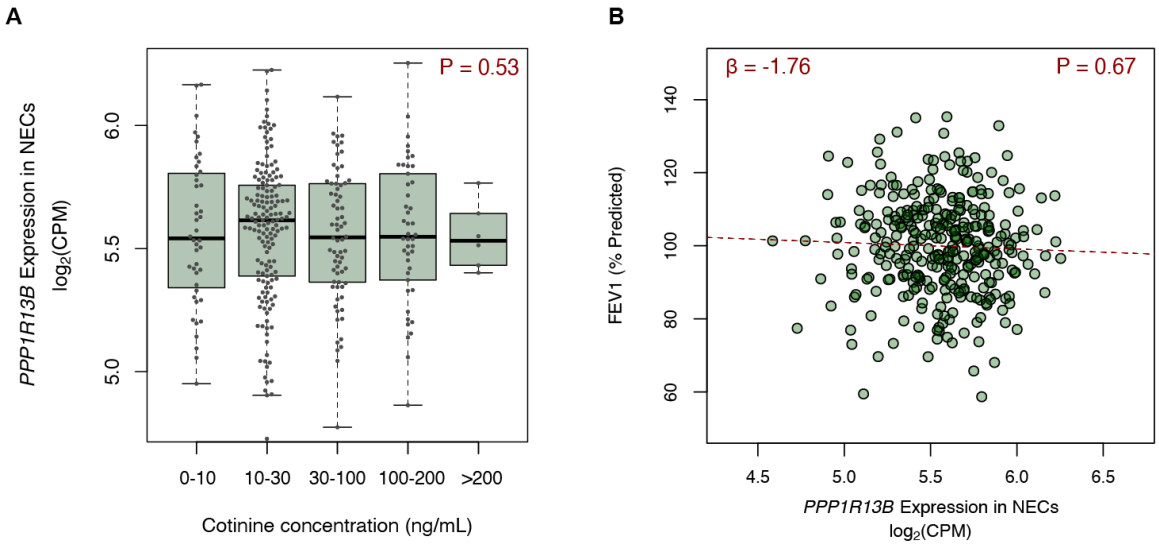

Fig S8. *PPP1R13B* expression in NECs vs. smoking exposure, FEV_1. *PPP1R13B* expression in NECs at age 11 was not associated with smoking exposure at age 10 (**A**) nor with FEV_1 (% predicted) at age 10 (**B**) in URECA. NECs, nasal epithelial cells; FEV_1, forced expiratory volume in one second; URECA, Urban Environment and Childhood Asthma.

Table S5. Chromatin interactions with FEV_1-associated SNPs

Promoter	Bait Fragment	Target Fragment	Strand	Distance	FEV_1 SNPs	CHiCAGO Score [158]
PPP1R13B	**103845355-103845955**	**104003512-104004312**	**-**	**157557**	**4**	**9.38**
CDCA4	105020395-105021827	104010864-104011532	-	1008863	3	7.65
PPP1R13B	**103845355-103845955**	**104004616-104004922**	**-**	**158661**	**4**	**7.16**
ZFYVE21	103716777-103717467	103948470-103949743	+	231003	1	6.57
ZFYVE21	103716777-103717467	103933835-103934479	+	216368	2	6.49
C14orf79	104985639-104987255	103935625-103936182	+	1049457	3	6.10
MARK3	103385080-103385712	103937160-103938502	+	551448	4	5.89
ZFYVE21	103715624-103716777	103981359-103981909	+	264582	2	5.89
PLD4	104927266-104927937	103963380-103964018	+	963248	3	5.69
ZFYVE21	103715624-103716777	103933397-103933835	+	216620	2	5.62
ADSSL1	104720007-104724747	104017055-104017482	+	702525	2	5.52
APOPT1	103562490-103563614	103999031-103999343	+	435417	1	5.48
CKB	103521551-103522109	103997308-103998030	-	475199	3	5.26
MARK3	103385881-103386341	103958731-103959336	+	572390	1	5.19
PPP1R13B	**103845355-103845955**	**104036454-104037811**	**-**	**190499**	**7**	**5.00**

Bait and target fragments refer to mapped Hi-C restriction fragments on chr14 (hg38) for gene promoters and putative enhancers, respectively. FEV_1 SNPs refer to number of FEV_1-associated variants ($p<1x10^{-5}$) within 1kb of target fragment. SNPs, single nucleotide polymorphisms; FEV_1, forced expiratory volume in one second.

Table S6. Age at used lung function measure in URECA

Age (years)	**Count**	
	FEV_1	**FEV_1/FVC**
10	382	372
9	26	30
8	10	11
7	8	5
6	6	10
5	9	10

URECA, Urban Environment and Childhood Asthma study; FEV_1, forced expiratory volume in one second; FVC, forced vital capacity.

Table S7. Study samples

Cohort	Sample	Cell Type	Age Collected	N
APIC	WGS	-	-	508
URECA	WGS	-	-	527
URECA	DNA Methylation	NECs	11	286
URECA	DNA Methylation	PBMCs	7	169
URECA	RNA-Seq	NECs	11	324
URECA	RNA-Seq	PBMCs	2	132

APIC, Asthma Phenotypes in the Inner City study; URECA, Urban Environment and Childhood Asthma study; WGS, whole-genome sequencing; NECs, nasal epithelial cells; PBMCs, peripheral blood mononuclear cells.

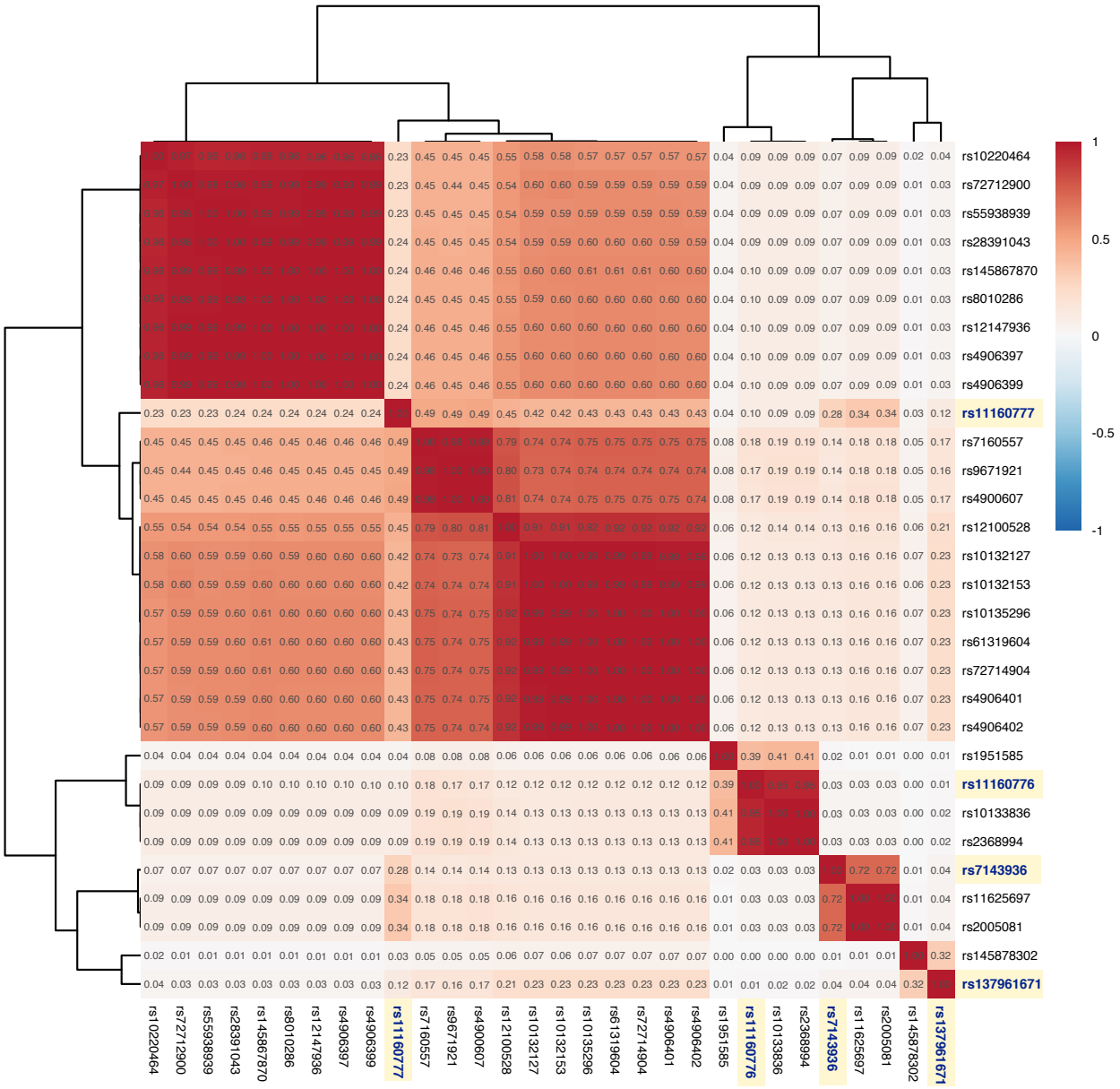

Fig S9. Intercorrelation of Mendelian randomization candidate instrument SNPs in URECA. Instrumental variables were chosen from a set of candidate SNPs that were at least nominally associated with cg03306306 methylation with p<0.15. The correlation values between these SNPs are shown, clustered using Ward's method. The four SNPs used for the instrument are highlighted. URECA, Urban Environment and Childhood Asthma.

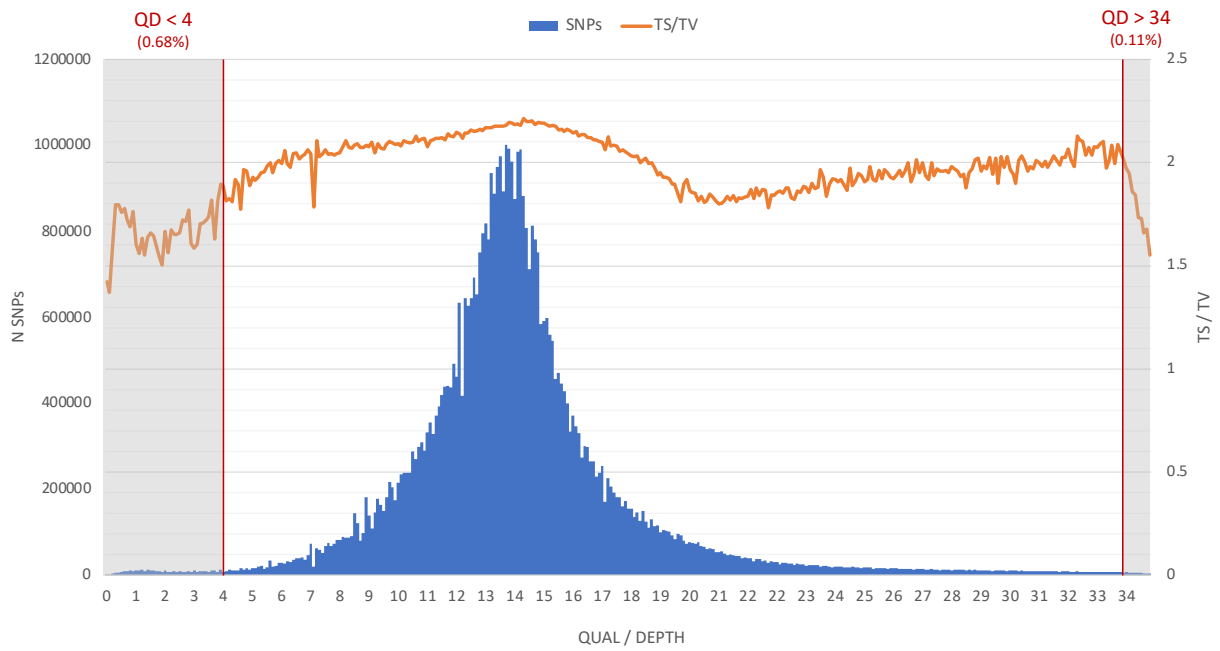

Fig S10. Transitions/transversions vs. quality/depth in WGS variant calls. The transition/transversion ratio (TS/TV) is plotted against the variant call quality/depth metric (QD) across all WGS SNP calls in APIC & URECA. Sites with QD less than 4 or greater than 34 were removed from consideration in this study. SNPs, single nucleotide polymorphisms; WGS, whole-genome sequencing; APIC, Asthma Phenotypes in the Inner City study; URECA, Urban Environment and Childhood Asthma study.