

Lung Function in African American Children with Asthma Is Associated with Novel Regulatory Variants of the KIT Ligand *KITLG/SCF* and Gene-By-Air-Pollution Interaction

Angel C. Y. Mak,^{*1} Satria Sajuthi,^{†,2} Jaehyun Joo,^{‡,2} Shujie Xiao,^{§,2} Patrick M. Sleiman,^{**††,2} Marquitta J. White,^{*,2} Eunice Y. Lee,^{*} Benjamin Saef,[†] Donglei Hu,^{*} Hongsheng Gui,[§] Kevin L. Keys,^{**} Fred Lurmann,^{§§} Deepti Jain,^{***} Gonçalo Abecasis,^{†††} Hyun Min Kang,^{†††} Deborah A. Nickerson,^{†††,§§§,****} Soren Germer,^{††††} Michael C. Zody,^{††††} Lara Winterkorn,^{††††} Catherine Reeves,^{††††} Scott Huntsman,^{*} Celeste Eng,^{*} Sandra Salazar,^{*} Sam S. Oh,^{*} Frank D. Gilliland,^{††††} Zhanghua Chen,^{††††} Rajesh Kumar,^{§§§§} Fernando D. Martínez,^{*****} Ann Chen Wu,^{†††††} Elad Ziv,^{*} Hakon Hakonarson,^{**††,‡,3} Blanca E. Himes,^{‡,3} L. Keoki Williams,^{§,‡,3} Max A. Seibold,^{†,‡,3} and Esteban G. Burchard^{*,†††††}

^{*}Department of Medicine and ^{†††††}Department of Bioengineering and Therapeutic Sciences, University of California, San Francisco, California 94143, [†]Center for Genes, Environment, and Health, National Jewish Health, Denver, Colorado 80206, [‡]Department of Biostatistics, Epidemiology, and Informatics and ^{††}Division of Human Genetics, Department of Pediatrics, The Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania 19104, [§]Center for Individualized and Genomic Medicine Research, Department of Internal Medicine, Henry Ford Health System, Detroit, Michigan 48202, ^{**}Center for Applied Genomics, Children's Hospital of Philadelphia, Pennsylvania, 19104, ^{††}Berkeley Institute for Data Science, University of California, Berkeley, California 94720, ^{§§}Sonoma Technology, Petaluma, California 94954, ^{***}Department of Biostatistics and ^{†††}Department of Genome Sciences, University of Washington, Seattle, Washington 98195, ^{††††}Center for Statistical Genetics, University of Michigan, Ann Arbor, Michigan 48109, ^{§§§}Northwest Genomics Center, Seattle, Washington, 98195, ^{****}Brotman Baty Institute for Precision Medicine, Seattle, Washington, 98195, ^{†††††}New York Genome Center, New York, 10013, ^{†††††}Department of Preventive Medicine, Division of Environmental Health, Keck School of Medicine, University of Southern California, Los Angeles, California 90033, ^{§§§§}Ann and Robert H. Lurie Children's Hospital of Chicago, Chicago, Illinois 60611, ^{*****}Asthma and Airway Disease Research Center, University of Arizona, Tucson, Arizona 85721, ^{†††††}Precision Medicine Translational Research (PRoMoTeR) Center, Department of Population Medicine, Harvard Medical School and Pilgrim Health Care Institute, Boston, Massachusetts 02215

ORCID IDs: 0000-0002-5372-4198 (A.C.Y.M.); 0000-0003-4886-6666 (K.L.K.); 0000-0002-2815-6037 (S.S.O.)

ABSTRACT Baseline lung function, quantified as forced expiratory volume in the first second of exhalation (FEV₁), is a standard diagnostic criterion used by clinicians to identify and classify lung diseases. Using whole-genome sequencing data from the National Heart, Lung, and Blood Institute Trans-Omics for Precision Medicine project, we identified a novel genetic association with FEV₁ on chromosome 12 in 867 African American children with asthma ($P = 1.26 \times 10^{-8}$, $\beta = 0.302$). Conditional analysis within 1 Mb of the tag signal (rs73429450) yielded one major and two other weaker independent signals within this peak. We explored statistical and functional evidence for all variants in linkage disequilibrium with the three independent signals and yielded nine variants as the most likely candidates responsible for the association with FEV₁. Hi-C data and expression QTL analysis demonstrated that these variants physically interacted with *KITLG* (KIT ligand, also known as *SCF*), and their minor alleles were associated with increased expression of the *KITLG* gene in nasal epithelial cells. Gene-by-air-pollution interaction analysis found that the candidate variant rs58475486 interacted with past-year ambient sulfur dioxide exposure ($P = 0.003$, $\beta = 0.32$). This study identified a novel protective genetic association with FEV₁, possibly mediated through *KITLG*, in African American children with asthma. This is the first study that has identified a genetic association between lung function and *KITLG*, which has established a role in orchestrating allergic inflammation in asthma.

KEYWORDS GWAS; African American; FEV₁ gene-by-environment interaction; GxE; air pollution; KITLG; SCF

ASTHMA, a chronic pulmonary condition characterized by reversible airway obstruction, is one of the hallmark diseases of childhood in the United States (World Health

Organization 2017). Asthma is also the most disparate common disease in the pediatric clinic, with significant variation in prevalence, morbidity, and mortality among U.S.

racial/ethnic groups (Oh *et al.* 2016). Specifically, African American children carry a higher asthma disease burden compared to their European American counterparts (Akinbami *et al.* 2014; Akinbami 2015). Forced expiratory volume in the first second (FEV₁), a measurement of lung function, is a vital clinical trait used by physicians to assess overall lung health and diagnose pulmonary diseases such as asthma (Johnson and Theurer 2014). We have previously shown that genetic ancestry plays an important role in FEV₁ variation and that African Americans have lower FEV₁ compared to European Americans, regardless of asthma status (Kumar *et al.* 2010; Pino-Yanes *et al.* 2015). The disparity in lung function between populations may explain disparities in asthma disease burden. Understanding the factors that influence FEV₁ variation among individuals with asthma could lead to improved patient care and therapeutic interventions.

Twin and family-based studies estimate that the heritability of FEV₁ ranges from 26 to 81%, supporting the combined contribution by genetic and environmental factors in FEV₁ variation (Chatterjee and Das 1995; Chen *et al.* 1996; Palmer *et al.* 2001; Hukkinen *et al.* 2011; Yamada *et al.* 2015; Sillanpaa *et al.* 2017; Tian *et al.* 2017). Genome-wide association studies (GWAS) of FEV₁, including among individuals with asthma, have identified many variants that contribute to lung function (Repapi *et al.* 2010; Soler Artigas *et al.* 2011, 2015; Li *et al.* 2013; Liao *et al.* 2014; Wain *et al.* 2017). A search in the National Human Genome Research Institute-European Bioinformatics Institute (NHGRI-EBI) GWAS Catalog (version e98_r2020-03-08) on baseline lung function (FEV₁) alone revealed 349 associations (Buniello *et al.* 2019). However, most of these previous GWAS were performed in adult populations of European descent, and their results may not generalize across populations or across the life span of an individual (Carlson *et al.* 2013; A. R. Martin *et al.* 2017; Wojcik *et al.* 2019). Previous GWAS results are also limited due to their reliance on genotyping arrays. In particular, variation in noncoding regions of the genome is not adequately covered by many genotyping arrays because they were not designed to account for the population-specific genetic variability of all populations (Zhang and Lupski 2015; Kim *et al.* 2018). Whole-genome sequencing (WGS) is a newer technology that captures nearly all common variation from coding and noncoding regions of the genome, and is unencumbered by genotype array design constraints and differences in linkage disequilibrium (LD) patterns among populations. To date, no large-scale WGS studies of lung function

have been performed in African American children with asthma (A. R. Martin *et al.* 2017).

In addition to genetics, FEV₁ is a complex trait that is significantly influenced by both genetic variation and environmental factors, such as air pollution (Chatterjee and Das. 1995; Palmer *et al.* 2001; Hukkinen *et al.* 2011; Yamada *et al.* 2015; Tian *et al.* 2017; Sillanpaa *et al.* 2017). Exposure to ambient air pollution has been consistently associated with poor respiratory outcomes, including reduced FEV₁ (Brunekreef and Holgate 2002; Barraza-Villarreal *et al.* 2008; Ierodiakonou *et al.* 2016; Wise 2019). We previously showed that exposure to sulfur dioxide (SO₂), an air pollutant emitted by the burning of fossil fuels, is significantly associated with reduced FEV₁ in African American children with asthma in the Study of African Americans, Asthma, Genes & Environments (SAGE II) study (Neophytou *et al.* 2016). Because the genetic variants associated with FEV₁ thus far do not account for the majority of its estimated heritability, considering gene–environment (GxE) interactions, specifically gene-by-air-pollution, may improve our understanding of lung function genetics (Moore 2005; Moore and Williams. 2009). Here, we performed a genome-wide association analysis using WGS data to identify common genetic variants associated with FEV₁ in African American children with asthma in SAGE II and investigated the effect of GxE (SO₂) interactions on FEV₁ associations.

Materials and Methods

Study population

This study examined African American children between 8–21 years of age with physician-diagnosed asthma from the SAGE II study. All SAGE II participants were recruited from the San Francisco Bay Area. The inclusion and exclusion criteria were previously described in detail (Oh *et al.* 2012; White *et al.* 2016). Briefly, participants were eligible if they were 8–21 years of age, self-identified as African American, and had four African American grandparents. Study exclusion criteria included the following: (1) any smoking within 1 year of the recruitment date; (2) 10 or more pack-years of smoking; (3) pregnancy in the third trimester; and (4) history of lung diseases other than asthma (for cases) or chronic illness (for cases and controls). Baseline lung function defined as FEV₁ was measured by spirometry prior to administering albuterol, as previously described (Oh *et al.* 2012).

Trans-Omics for Precision Medicine WGS data

SAGE II DNA samples were sequenced as part of the Trans-Omics for Precision Medicine (TOPMed) WGS program (Taliun *et al.* 2019 preprint). WGS was performed at the New York Genome Center and Northwest Genomics Center on a HiSeq X system (Illumina, San Diego, CA) using a paired-end read length of 150 bp, with a minimum of 30× mean genome coverage. DNA sample handling, quality control, library construction, clustering and sequencing, read processing, and sequence data quality control are described in detail

Copyright © 2020 by the Genetics Society of America

doi: <https://doi.org/10.1534/genetics.120.303231>

Manuscript received January 21, 2020; accepted for publication April 22, 2020; published Early Online April 23, 2020.

Available freely online through the author-supported open access option.

Supplemental material available at figshare: <https://doi.org/10.25386/genetics.12152196>.

¹Corresponding author: University of California San Francisco, Box 2911, 1550 4th St., San Francisco, CA 94112. E-mail: angelcymak@gmail.com

²These authors contributed equally to this work.

³These senior authors contributed equally to this work.

in the TOPMed website (National Heart, Lung and Blood Institute Trans-Omics for Precision Medicine (TOPMed) Program 2019). Variant calls were obtained from TOPMed data freeze 8 variant call format (VCF) files corresponding to the GRCh38 human genome assembly. Variants with a minimal read depth of 10 (DP10) were used for analysis unless otherwise stated.

Genetic principal components, global ancestry, and kinship estimation

Genetic principal components (PCs), global ancestry, and kinship estimation on genetic relatedness were computed using biallelic single-nucleotide polymorphisms (SNPs) with a PASS flag from TOPMed freeze 8 DP10 data. PCs and kinship estimates were computed using the PC-Relate function from the GENESIS R package (Conomos *et al.* 2015, 2016) using a workflow available from the Summer Institute in Statistical Genetics Module 17 course website (Summer Institute in Statistical Genetics 2019). African global ancestry was computed using the ADMIXTURE package (Alexander *et al.* 2009) in supervised mode using European (CEU), African (YRI), and Native American (NAM) reference panels as previously described (Mak *et al.* 2018).

FEV₁ GWAS

Nonnormality of the distribution of FEV₁ values was tested with the Shapiro–Wilk test in R using the `shapiro.test` function. Since FEV₁ was not normally distributed ($P = 1.41 \times 10^{-8}$ for FEV₁ and $P = 1.05 \times 10^{-8}$ for \log_{10} FEV₁), FEV₁ was regressed on all covariates (age, sex, height, controller medications, sequencing centers, and the first five genetic PCs) and the residuals were inverse-normalized. These inverse-normalized residuals (FEV₁.res.rnorm) were the main outcome of the discovery GWAS. The controller medication covariate included the use of inhaled corticosteroids (ICS), long-acting β -agonists (LABA), leukotriene inhibitors, and/or an ICS/LABA combo in the 2 weeks prior to the recruitment date.

Genome-wide single-variant analysis was performed on the ENCORE server (<https://github.com/statgen/encore>) using the linear Wald test (`q.linear`) originally implemented in EPACTS (<https://genome.sph.umich.edu/wiki/EPACTS>) and TOPMed freeze 8 data (DP0 PASS), with a minor allele frequency (MAF) filter of 0.1%. All pairwise relationships with degree three or more relatedness (kinship values > 0.044) were identified, and one participant of the related pair was subsequently chosen at random and removed prior to analysis. All covariates used to obtain FEV₁.res.rnorm were also included as covariates in the GWAS as recommended in a recent publication (Sofer *et al.* 2019). The association analysis was repeated using untransformed FEV₁ and FEV₁ percent predicted (FEV₁.perc.predicted). FEV₁ percent predicted was defined as the percentage of measured FEV₁ relative to predicted FEV₁, estimated by the Hankinson lung function prediction equation for African Americans (Hankinson *et al.* 1999). A secondary analysis that included smoking-related

covariates (smoking status and number of smokers in the family) was performed in PLINK 1.9 (version 1p9_2019_0304_dev) (Purcell and Chang, 2013; Chang *et al.* 2015). To study whether association with FEV₁ is specific to SAGE II participants with asthma, we repeated the association analysis adjusting for age, sex, height, and the first five genetic PCs in SAGE II participants without asthma on the ENCORE server. All of these participants were sequenced in the same center. Regional association results were plotted using LocusZoom 1.4 (Pruim *et al.* 2010) with a 500 kb flanking region. LD (R^2) was estimated in PLINK 1.9. An LD plot was generated using recoded genotype files (`plink-recode 12`) in Haploview (Barrett *et al.* 2005).

The function `effectiveSize` in the R package CODA was used to estimate the actual effective number of independent tests, and CODA-adjusted statistical and suggestive significance P -value thresholds were defined as 0.05 and 1, divided by the effective number of tests, respectively (Duggal *et al.* 2008). We compared the CODA-adjusted statistical significance threshold and the widely used 5×10^{-8} GWAS genome-wide significance threshold (Pe'er *et al.* 2008), and selected the more stringent threshold for genome-wide significance.

The following WGS quality control steps were applied to all reported variant association results from the ENCORE server to ensure WGS variant quality: (1) The variant had `VCF FILTER = PASS`; (2) variant quality was confirmed via manual inspection on the BRAVO server based on TOPMed freeze 5 data (University of Michigan and National Heart, Lung and Blood Institute Trans-Omics for Precision Medicine (TOPMed) Program 2018); and (3) variants were reanalyzed with linear regression using PLINK 1.9 by applying the arguments `-mac 5 -geno 0.1 -hwe 0.0001` using TOPMed freeze 8 DP10 PASS data.

To determine if the rs73429450 association with FEV₁ was only identifiable using WGS data, we repeated the linear regression association analysis on signals that passed the genome-wide significance threshold using PLINK 1.9 and genotype data generated with the Axiom Genome-Wide LAT 1 array (Affymetrix, Santa Clara, CA; dbGaP phs000921.v1.p1). These array genotype data were imputed into the following reference panels: 1000 Genomes Project (1000G) phase 3 version 5, Haplotype Reference Consortium (HRC) r1.1, the Consortium on Asthma among African-ancestry Populations in the Americas (CAAPA), and the TOPMed phase 5 panels on the Michigan Imputation Server (Das *et al.* 2016). It should be noted that 500 SAGE II subjects were part of the TOPMed freeze 5 reference panel.

A total of 349 GWAS FEV₁-associated entries were retrieved from the NHGRI-EBI GWAS Catalog version 1.0.2-associations_e98_r2020-03-08 (Buniello *et al.* 2019) using the trait names “Lung function (FEV₁)”, “FEV₁,” “Lung function (forced expiratory volume in 1 sec),” or “Prebronchodilator FEV₁.” After adding 100-kb flanking regions to each of the 349 entries, a total of 230 nonoverlapping region were obtained. To look up whether we replicated previous GWAS loci while control for multiple testing penalties, we only used 279,495 common variants (MAF ≥ 0.01) that overlapped

with the 230 regions. The 279,495 common variants were equivalent to 17,755 effective tests based on CODA and 5.63×10^{-5} ($1/17,755$) was used as the suggestive P -value threshold for replication.

Conditional analysis

Conditional analysis was performed to identify all independent signals in a GWAS peak using PLINK 1.9. All TOPMed freeze 8 DP10 variants within 1 Mb of the tag association signal, and with association P -value of 1×10^{-4} or smaller in the discovery GWAS, were included in the analysis. Variants were first ordered by ascending P -value. A variant was considered to be an independent signal if the association P -value after conditioning (conditional P -value) on the tag signal was < 0.05 . Newly identified independent signals were included with the tag signal for conditioning on the next variant.

Region-based association analysis

Region-based association analyses were performed in 1-kb sliding windows with 500-bp increments in a 1-Mb flanking region of the tag GWAS signal using the SKAT_CommonRare function from the SKAT R package v1.3.2.1 (Ionita-Laza *et al.* 2013). Default settings were used with method = "C" and test.type = "Joint." A MAF threshold of 0.01 was used as the cutoff to distinguish rare and common variants. Variants were annotated in TOPMed using the Whole Genome Sequencing Annotator (WGSA) pipeline (Liu *et al.* 2016). Since SKAT imputes missing genotypes by default by assigning mean genotype values (impute.method="fixed"), we chose to use low-coverage genotypes instead of SKAT imputation, and hence TOPMed freeze 8 DP0 variants with a VCF FILTER of PASS were included in the analysis. The function effectiveSize in the R package CODA (Plummer *et al.* 2006) was used to estimate the effective number of independent hypothesis tests for accurate Bonferroni multiple testing corrections. P -value thresholds for statistical significance and suggestive significance were defined as 0.05 and 1 divided by the effective number of tests, respectively (Duggal *et al.* 2008). If a region was suggestively significant, region-based analyses were repeated with functional variants and/or rare variants ($MAF \leq 0.01$) to assess contributions of common, rare, and/or functional variants. Region-based analyses using rare variants only were performed using SKAT-O (Lee *et al.* 2012). The WGSA annotation filters used to define functional variants are provided in Supplemental Material, File S1. To study the contribution of individual variants to a region-based association P -value, drop-one variant analysis was performed by repeating the region-based analysis multiple times and dropping one variant only at a time.

Functional annotations and prioritization of genetic variants

The Hi-C Unifying Genomic Interrogator (HUGIN) (Ay *et al.* 2014; Schmitt *et al.* 2016; J. S. Martin *et al.* 2017) was used to assign potential gene targets to each variant. HUGIN uses the Hi-C data generated from the primary human tissues from four

donors used in the Roadmap Epigenomics Project (Schmitt *et al.* 2016). Encyclopedia of DNA Elements (ENCODE) annotations (ENCODE Project Consortium 2011, 2012) were based on overlap of the variants with functional data downloaded from the University of California, Santa Cruz (UCSC) Table Browser (Karolchik *et al.* 2004). These data included DNase I hypersensitivity peak clusters (hg38 wgEncodeRegDnaseClustered table), transcription factor chromatin immunoprecipitation-sequencing (ChIP-Seq) clusters (hg38 encRegTfbsClustered table), and histone modification ChIP-Seq peaks (hg19 wgEncodeBroadHistone <cell type> <histone> StdPk tables). For DNase I hypersensitivity and transcription factor binding sites, we focused on blood, bone marrow, lung, and embryonic cells. For histone modification ChIP-Seq, we focused on H3K27ac and H3K4me3 modifications in human blood (GM12878), bone marrow (K562), lung fibroblast (NHLF), and embryonic stem cells (¹H-hESC). The LiftOver tool (Hinrichs *et al.* 2006) was used to convert genomic coordinates from hg19 to hg38. Candidate *cis*-regulatory elements (ccREs) were a subset of representative DNase hypersensitivity sites with epigenetic activity further supported by histone modification (H3K4me3 and H3K27ac) or CTCF-binding data from the ENCODE project. Overlaps of variants with ccREs were detected using the Search Candidate *cis*-Regulatory Elements by ENCODE (SCREEN) web interface (ENCODE Project Consortium 2011, 2012).

Prioritization of genetic variants was based on the presence of statistical, functional, and/or bioinformatic evidence as described in the Diverse Convergent Evidence prioritization framework (Ciesielski *et al.* 2014). The priority score of each variant was obtained by counting the number of pieces of statistical, functional, and/or bioinformatic evidence that supported a potential biological function for that variant.

Replication of GWAS associations

All replication analyses were performed in subjects with asthma. Replication of GWAS FEV₁ associations was attempted on TOPMed WGS data generated from four cohorts. These cohorts included Puerto Rican ($n = 1109$) and Mexican American ($n = 649$) children in the Genes-Environments and Admixture in Latino Americans (GALA II) study (Oh *et al.* 2012), African American adults in the Study of Asthma Phenotypes and Pharmacogenomic Interactions by Race-Ethnicity (SAPPHIRE, $n = 3428$) (Levin *et al.* 2014), and African American children in the Genetics of Complex Pediatric Disorders (GCPD-A, $n = 1464$) study (Ong *et al.* 2013). Age, sex, height, controller medications, and the first five PCs were used as covariates.

Additionally, replication of GWAS FEV₁ associations was attempted using data of black UK Biobank subjects who had asthma ($n = 627$) while adjusting for age, sex, height, and the first five PCs. Asthma status was defined by International Statistical Classification of Diseases and Related Health Problems (ICD) code of 493 or self-reported asthma. UK Biobank genotype data were generated on the Affymetrix UK BiLEVE axiom or UK Biobank Axiom array and imputed into the HRC,

1000G, and UK 10K projects (Bycroft *et al.* 2018; Canela-Xandri *et al.* 2018). Additional details on the UK Biobank study and the replication procedures are available in File S1 (Text 2).

RNA sequencing and expression QTL analysis

Whole-transcriptome libraries of 370 nasal brushings from GALA II Puerto Rican children with asthma were constructed by using the Beckman Coulter FX automation system (Beckman, Fullerton, CA). Libraries were sequenced with the Illumina HiSeq 2500 system. Raw RNA sequencing (RNA-Seq) reads were trimmed using Skewer (Jiang *et al.* 2014) and mapped to human reference genome hg38 using Hisat2 (Kim *et al.* 2015). Reads mapped to genes were counted with htseq-count and using the UCSC hg38 Gene Transfer Format (GTF) file as reference (Anders *et al.* 2015). *Cis*-expression QTL (eQTL) analysis of *KITLG* (KIT ligand) was performed as described in the Genotype-Tissue Expression (GTEx) project version 7 protocol (GTEx Consortium *et al.* 2017) using age, sex, body mass index, global African and European ancestries, and 60 probabilistic estimation of expression residual (PEER) factors as covariates.

Gene-by-air-pollution interaction analysis

We hypothesized that the effect of genetic variation on lung function in our study population may differ by the levels of exposure to SO₂ (Neophytou *et al.* 2016). To test for an interaction between a genetic variant and SO₂, an additional multiplicative interaction term (variant × SO₂ exposure) was included in the original GWAS model (see section *FEV₁ GWAS*). The SO₂ estimates used in the interaction analysis were first-year, past-year, and lifetime exposure to ambient SO₂, which were estimated as described previously (Neophytou *et al.* 2016). Briefly, we obtained regional ambient daily air pollution data from the U.S. Environmental Protection Agency Air Quality System. SO₂ estimates for each participant's residential geographic coordinate were calculated as the inverse distance-squared weighted average from the four closest air pollution monitoring stations within 50 km of the participant's residence. We estimated yearly exposure at the reported residential address by averaging all available daily measures (daily average of 1-hr SO₂) in a given year. If the participant had a change of residential address in a given year, we estimate yearly exposure as a time-weighted estimate based on the number of months spent at each different address in that year. Average lifetime exposures were estimated using all available yearly average estimates over the lifetime of the participant until the day of spirometry testing. Since not all pollutants were measured daily, there are location- and pollutant-dependent missing values. Residuals of FEV₁ were plotted against exposure to SO₂ and stratified by the number of copies of the minor allele of a variant. Residuals of FEV₁ were obtained as described in section *FEV₁ GWAS*.

Data availability

Local institutional review boards approved the studies (number 10-02877). All subjects and legal guardians provided

written informed consent. TOPMed WGS and phenotype data from SAGE II are available on dbGaP under accession number phs000921.v4.p1. Supplemental material and normalized gene count data for *KITLG* available at figshare: <https://doi.org/10.25386/genetics.12152196>.

Results

Novel lung function associations

Subject characteristics of the 867 African American children with asthma included in this study are shown in Table 1, and the distribution of their FEV₁ measurements (mean = 2.56 L, SD = 0.79 L) is in Figure S1. The CODA-adjusted statistical significance thresholds 2.10×10^{-8} and 4.19×10^{-7} were used as the genome-wide and suggestive significance thresholds, respectively. According to this threshold, one SNP in chromosome 12 (chromosome 12:88846435, rs73429450, G > A) was associated with FEV₁.res.rnorm (Figure 1, $P = 9.01 \times 10^{-9}$, $\beta = 0.801$) at genome-wide significance. The association between rs73429450 and lung function remained statistically significant when the association was repeated using untransformed FEV₁ ($P = 1.26 \times 10^{-8}$, $\beta = 0.302$) as the outcome variable. The association between rs73429450 and lung function was suggestive using FEV₁.perc.predicted ($P = 1.69 \times 10^{-7}$, $\beta = 0.100$). Twenty suggestive associations corresponding to four tag signals are reported in File S2. None of the suggestive associations overlapped with any of the previously reported FEV₁-associated loci. When considering only common variants and applying a P -value threshold of 5.63×10^{-5} , we found replication in 6 out of 230 previously reported FEV₁ associations (Table S1). Our top FEV₁ association, rs73429450, did not overlap with any previously reported loci and it is a novel association with FEV₁ in this study population.

Secondary analysis that included covariates correcting for smoking status and number of smokers in the family showed that smoking-related factors were not significantly associated with FEV₁ in our pediatric SAGE II cohort: using 657 out of 867 individuals with available smoking-related covariates, the FEV₁.res.rnorm association P -values before and after including the smoking-related covariates were 2.01×10^{-6} and 1.89×10^{-6} . Both P -values of the covariates smoking status ($P = 0.27$) and number of smokers in the family ($P = 0.54$) were not significant.

Conditional analysis was performed on 45 variants with association $P < 1 \times 10^{-4}$ located within 1 Mb of the strongest association signal (rs73429450). Two weaker independent signals (rs17016065 and rs58475486) were identified (Table S2). None of the 45 variants showed association with FEV₁.res.rnorm in 251 SAGE II children without asthma (Table S3).

The minor allele frequency of rs73429450 in continental populations from the 1000G is 3% in Africans and < 1% in Admixed Americans, Europeans, Asians (1000 Genomes Project Consortium *et al.* 2015). SNP rs73429450 was not included on the Affymetrix LAT 1 genotyping array where

Table 1 Descriptive characteristics of 867 African American children with asthma included in this study

Characteristic	African American (n = 867)
Age	
Mean (SD)	14.1 (3.64)
Median (25%, 75%)	13.8 (10.98, 17.11)
Sex	
Male	439 (50.6%)
Female	428 (49.4%)
Height (m)	
Mean (SD)	1.58 (0.145)
Median (25%, 75%)	1.60 (1.47, 1.68)
Any control medications* in last 2 weeks	
No	543 (62.6%)
Yes	324 (37.4%)
ICS in last 2 weeks	
No	211 (24.3%)
Yes	306 (35.3%)
Missing	350 (40.4%)
LABA in last 2 weeks	
No	5 (0.6%)
Yes	94 (10.8%)
Missing	768 (88.6%)
Leukotriene inhibitor in last 2 weeks	
No	11 (1.3%)
Yes	68 (7.8%)
Missing	788 (90.9%)
African ancestry	
Mean (SD)	0.792 (0.129)
Median (25%, 75%)	0.826 (0.759, 0.869)
Smoking status	
Never	793 (91.5%)
Past	72 (8.3%)
Current	0 (0%)
Missing	2 (0.2%)
Number of smokers in family	
0	469 (54.1%)
1	137 (15.8%)
2	42 (4.8%)
3+	10 (1.2%)
Missing	209 (24.1%)
SO ₂ first year exposure (ppb)	
Mean (SD)	1.59 (0.961)
Median (25%, 75%)	1.50 (1.24, 1.87)
Missing	227 (26.2%)
SO ₂ past year exposure (ppb)	
Mean (SD)	1.10 (0.302)
Median (25%, 75%)	1.08 (0.910, 1.27)
Missing	206 (23.8%)
SO ₂ lifetime exposure (ppb)	
Mean (SD)	1.50 (0.371)
Median (25%, 75%)	1.47 (1.40, 1.54)
Missing	206 (23.8%)
TOPMed sequencing center and phase	
Phase 1, CAAPA	6 (0.7%)
Phase 1, NYGC	460 (53.1%)
Phase 3, NW	401 (46.3%)

25% and 75%, 25th and 75th percentiles. Control medications include ICS, LABA, leukotriene inhibitor, and/or ICS/LABA combo. SO₂ exposures are hourly exposures averaged over the specified time periods before spirometry testing, as previously described in Neophytou *et al.* (2016). CAAPA, Consortium on Asthma among African-ancestry Populations in the Americas; NYGC, New York Genome Center; NW, Northwest Genomics Center; ICS, inhaled corticosteroid; LABA, long-acting β -agonist; ppb, parts per billion or $\mu\text{g}/\text{m}^3$; TOPMed, Trans-Omics for Precision Medicine.

SAGE II participants were previously genotyped. To determine if the rs73429450 association with FEV₁ was only identifiable using WGS data, we attempted to reproduce our results by imputing the genotype of rs73429450 in 851 SAGE II participants with available array data using 1000G phase 3 ($n = 2504$), HRC r1.1 ($n = 32,470$), CAAPA ($n = 883$), and TOPMed freeze 5 ($n = 62,784$) reference panels. Our results remained statistically significant when using the 1000G phase 3 ($P = 4.97 \times 10^{-8}$, $\beta = 0.79$, imputation $R^2 = 0.95$) and TOPMed freeze 5 ($P = 1.22 \times 10^{-8}$, $\beta = 0.80$, imputation $R^2 = 0.98$) reference panels, but lost statistical significance when rs73429450 genotypes were imputed using the HRC ($P = 4.35 \times 10^{-7}$, $\beta = 0.68$, imputation $R^2 = 0.94$) and CAAPA ($P = 1.95 \times 10^{-7}$, $\beta = 0.80$, imputation $R^2 = 0.71$) reference panels.

Region-based association analysis including all variants conditioned on the association signal from rs73429450 was performed in its 1-Mb flanking region (chromosome 12:87846435-89846435). No windows were significantly associated after Bonferroni multiple testing correction ($P < 2.80 \times 10^{-4}$, Figure S2), but 20 windows were suggestively associated with FEV₁.res.rnorm ($P < 5.60 \times 10^{-3}$, Table S5). Two of 20 windows retested using only functional variants were suggestively significant (regions 4 and 16). Both of these windows were no longer suggestively significant after removing the common variants, indicating that association signals from these regions were mostly driven by common variants. Further investigation of region 16 using drop-one analysis of the two rare and one common function variants confirmed the major contribution by the common variant, rs1895710, as shown by the major increase in P -value (Table S6). The signal was also slightly driven by the singleton, rs990979778. Drop-one analysis was not performed on region 4 because there was only one common and one rare variant.

A Hi-C assay couples a chromosome conformation capture assay with next-generation sequencing to capture long-range interactions in the genome. We identified a statistically significant long-range chromatin interaction between the GWAS peak and the *KITLG* [also known as stem cell factor (*SCF*)] gene in human fetal lung fibroblast cell line IMR90 (Table S7). The long-range interaction detected in human primary lung tissue was not significant, implying that the potential long-range interactions are specific to tissue type or developmental stage.

Potential regulatory role of FEV₁-associated variants on *KITLG* expression

To further elucidate potential regulatory relationships between the GWAS association peak and *KITLG*, we analyzed whether variants in the peak were eQTL of *KITLG* in previously published whole-blood RNA-Seq data available from the same study participants (Mak *et al.* 2016). However, the whole blood RNA-Seq data did not yield evidence of expressed *KITLG*, consistent with results in GTEx. We subsequently used RNA-Seq data from nasal epithelial cells of

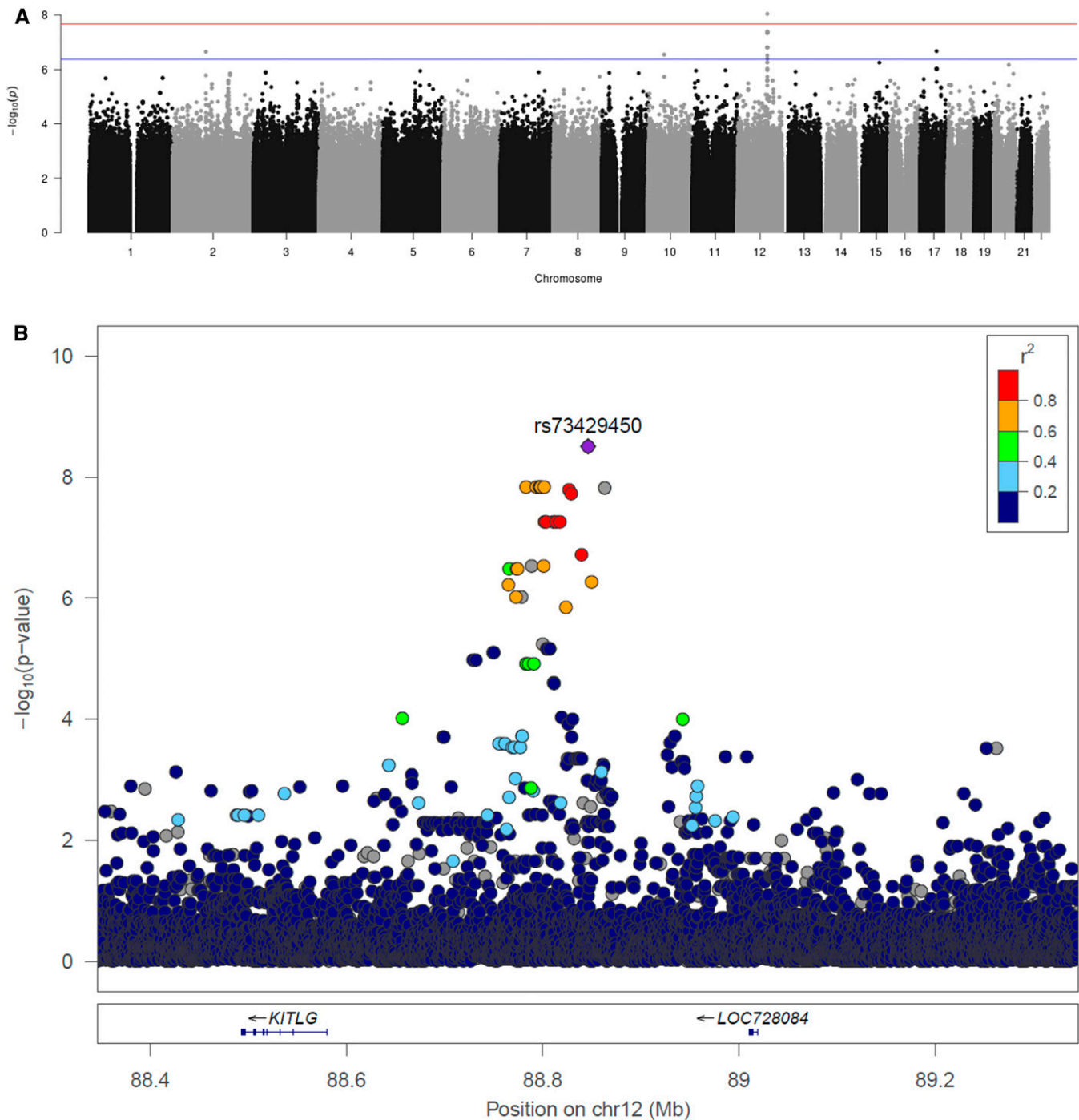


Figure 1 Manhattan and LocusZoom plots from genome-wide association study of lung function. (A) Manhattan plot from genome-wide association study of lung function using linear regression on the ENCORE server. FEV_{1.res.rnorm} was used as the phenotype for the association testing. Red horizontal line: CODA-adjusted genome-wide significance P -value of 2.10×10^{-8} . Blue horizontal line: CODA-adjusted suggestive significance P -value of 4.19×10^{-7} . (B) LocusZoom plot of rs73429450 (chr12: 88846435) and 500-kb flanking region. Colors show linkage disequilibrium in the study population. chr12, chromosome 12; FEV₁, forced expiratory volume in the first second of exhalation.

370 Puerto Rican children with asthma from the GALA II study, and found that 5 out of 45 variants were eQTL of *KITLG* (Table S8). While Puerto Ricans are a different population compared with African Americans, they are both admixed populations with substantial African genetic ancestry, and therefore could share eQTLs. All five eQTLs

corresponded to one signal in a region with strong LD ($r^2 > 0.8$, Figure S3).

Replication of genetic association with FEV₁

Subject characteristics of our four replication cohorts (SAPPHIRE, GCPD-A, UK Biobank, and GALA II) are shown in Table S9.

We attempted to replicate the association of the 45 SNPs in our primary FEV₁ GWAS in each cohort. We used 0.05 as the suggestive *P*-value threshold and 0.0167 as the Bonferroni-corrected *P*-value threshold after correcting for three independent signals (see conditional analysis in the *Results* section). A total of 20 variants were replicated at *P* < 0.05 with consistent direction of effect in black UK Biobank participants; 14 variants in SAPHIRE and 2 variants in GCPD-A were significant but had an opposite direction of effect (Table S10).

We attempted to replicate the FEV₁.res.rnorm association in Mexican American (*n* = 649) and Puerto Rican (*n* = 1109) children with asthma from the GALA II study. In Mexican Americans, we excluded 19 variants with MAF < 0.1% and associations for the remaining 26 variants did not replicate (Table S11). In Puerto Ricans, the associations were not replicated but we observed the same protective effect in 38 of the 45 variants in the locus (Table S11).

Incorporating statistical and functional evidence for candidate variant prioritization

We combined and summarized all functional evidence for the top 45 variants, along with eQTL findings from nasal epithelial RNA-Seq and replication results (Figure 2, Table 2, and Table S12). To facilitate interpretation of the variant association with FEV₁, the effect sizes and *P*-values of both FEV₁ (β and *P*) and FEV₁.res.rnorm (β_{norm} and P_{norm}) associations are also reported. Combined Annotation Dependent Depletion (CADD) functional prediction score and ENCODE histone modification ChIP-Seq peaks in embryonic, blood, bone marrow, and lung-related tissues were also examined, but not reported because none of the variants had a CADD score > 10 and none overlapped with histone modification sites. SNP rs73440122 received the highest priority score of 3 based on replication in the UK Biobank, overlap with a DNase I hypersensitivity site in B lymphoblastoid cells (GM12865), and overlap with an SPI1-binding site in acute promyelocytic leukemia cells. Eight other variants were prioritized with a score > 2 or evidence of being an eQTL for *KITLG* in nasal epithelial cells (Table 2, score marked with ^ or #, respectively). These nine candidate variants were selected for gene-by-air-pollution interaction analyses.

Gene-by-air-pollution interaction of rs58475486

We previously found that first year of life and lifetime exposure to SO₂ were associated with FEV₁ in African American children (Neophytou *et al.* 2016). We investigated whether the effect of the nine prioritized genetic variants associated with lung function varied by SO₂ exposure (first year of life, past-year, and lifetime exposure). Since the nine variants represent three independent signals (see conditional analysis in the *Results* section), the Bonferroni-corrected *P*-value threshold was set to *P* = 0.0056 (correction for nine tests; three signals and three exposure periods to SO₂). We observed a single statistically significant interaction between the T allele of rs58475486 and past-year exposure to SO₂ that was positively associated with FEV₁ (*P* = 0.003, β = 0.32; Figure 3A and Table 3). This

interaction remained significant (*P* = 0.003, β = 0.32) in secondary analyses adjusted for smoking status, or a multiplicative interaction term of rs58475486 and smoking status as additional covariates. Interestingly, six of the remaining eight variants also displayed interaction effects with past-year exposure to SO₂ that were suggestively associated (*P* < 0.05) with FEV₁ (Table 3). We also found a suggestive interaction of the C allele of rs73440122 with first-year exposure to SO₂ that was associated with decreased FEV₁ (*P* = 0.045, β = -0.32; Figure 3B). The same allele also showed interaction with past year of exposure to SO₂ that was suggestively associated with FEV₁ in the opposite direction (*P* = 0.051, β = 0.39).

Discussion

Variant rs73429450 (MAF = 0.030) was identified as the strongest association signal with FEV₁. Each additional copy of the protective A allele of rs73429450 was associated with a 0.3 L increase of FEV₁. We did not find any statistically significant contribution of rare variants to the association signal from a 1-kb sliding window analyses in the 1-MB flanking region centered on rs73429450. We were surprised to identify a novel common variant (MAF = 0.030) associated with lung function using WGS data in a population that was previously analyzed for associations with lung function using genotype array data. Further investigation revealed that our discovered variant, rs73429450, was not captured by the LAT 1 genotyping array, and the association with lung function depended on the reference panel used to impute the variant into our population. More surprisingly, our statistically significant finding was only found to be suggestively significant using data imputed from the CAAPA reference panel (*P* = 1.95×10^{-7} , β = 0.80). Of the imputation reference panels that we assessed, CAAPA is one of the more relevant reference panels for our study population because it is based on African populations in the Americas. However, we note that the effect size estimated from CAAPA-imputed data was comparable to that generated from WGS data. While WGS data are usually praised for enabling analysis of rare-variant contributions to phenotype variability, our results show the utility of WGS data for the reliable analysis of common variants as well as in the absence of relevant imputation panels.

Although rs73429450 had the lowest *P*-value from our WGS association analysis, we did not find the required amount of functional evidence to prioritize this marker for inclusion in downstream gene-by-air-pollution analyses. Another variant, rs73440122, was in moderate-to-strong LD (r^2 = 0.76) with rs73429450 and had a similar MAF (0.027) in our study population, but was only suggestively associated with FEV₁ in our association analysis (*P* = 2.08×10^{-7} ; Table 2). In contrast to rs73429450, there were multiple lines of evidence suggesting the functional relevance of rs73440122: rs73440122 received the highest priority score based on its replicated FEV₁ association in black UK Biobank participants and overlap with ENCODE gene regulatory

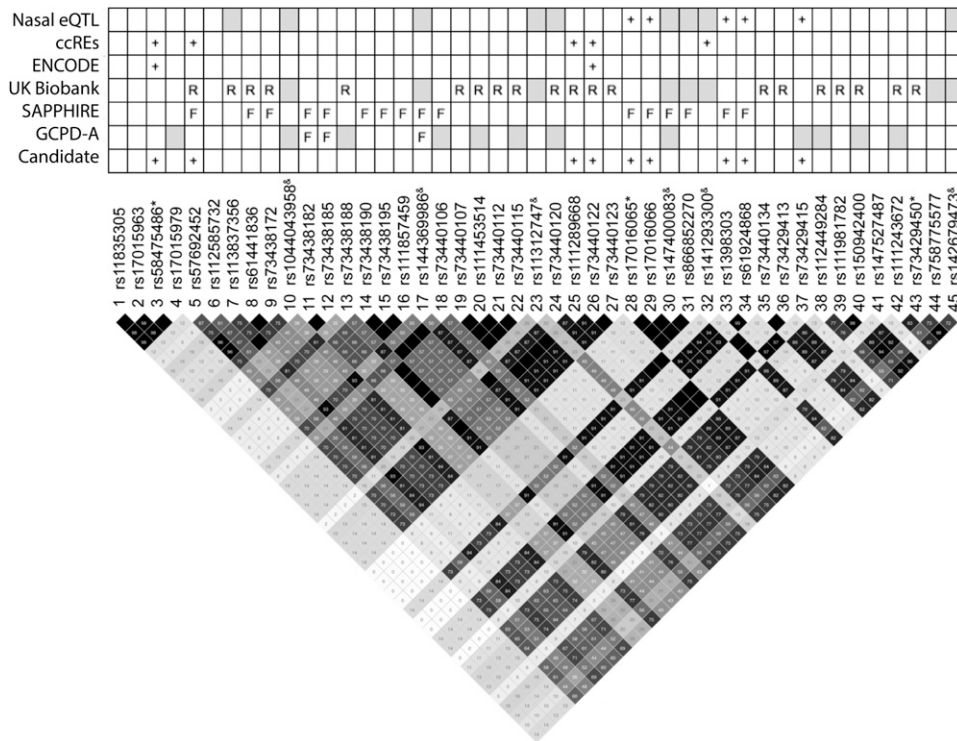


Figure 2 Integration of statistical and functional evidence for variant prioritization. Numbers and different shades of black in the LD plot represent LD in R^2 . The three independent signals identified in the conditional analysis are marked with “*”. Insertions/deletions are marked with “&”. Nasal eQTL, eQTLs of *KITLG* in nasal epithelial cells. ENCODE, DNase I hypersensitivity site and/or transcription factor ChIP-Seq peaks overlapping with the variants. UK Biobank, SAPHIRE, and GCPD-A: replication results using blacks in the UK Biobank and African Americans in the SAPHIRE and GCPD-A cohorts (R = replicated at $P < 0.05$; F = flip-flop association at $P < 0.05$). Candidate, candidate variants prioritized because of presence of two or more pieces of evidence, or because they are a nasal eQTL. + indicates presence of evidence. Boxes in the top panel were shaded gray if results were not available. ccREs, candidate *cis*-regulatory elements in SCREEN registry; ChIP-Seq, chromatin immunoprecipitation sequencing; ENCODE, Encyclopedia of DNA Elements; eQTL, expression QTL; GCPD-A, Genetics of

Complex Pediatric Disorders study; LD, linkage disequilibrium; SAPHIRE, Study of Asthma Phenotypes and Pharmacogenomic Interactions by Race-Ethnicity; SCREEN, Search Candidate *cis*-Regulatory Elements by ENCODE.

regions, making it one of the most likely drivers of FEV₁ variability among individuals, possibly mediated through *KITLG*.

Bioinformatic interrogation of rs73440122 revealed that the variant overlapped with a ccRE (SCREEN accession EH37E0279310), DNase I hypersensitivity site, and SPI1 ChIP-Seq clusters that were indicative of a candidate open chromatin gene regulatory region (Table S12). The binding evidence of SPI1 is highly relevant to the role of *KITLG* in type 2 inflammation (see below). Variant rs73440122 is located in a region that physically interacts with *KITLG* based on Hi-C data in fetal lung fibroblast cells. Additionally, five neighboring FEV₁ associated variants were identified as eQTLs of *KITLG*, although they appeared to be an independent signal ($r^2 < 0.2$). Overall, these results support regulatory interactions between our novel locus and *KITLG*.

Atopic or type 2 high asthma is the most common form of asthma in children (Comberiati *et al.* 2017). *KITLG*, more commonly known as *SCF*, is a ligand of the KIT tyrosine kinase receptor. It plays an important role in type 2 inflammation in atopic asthma, especially in inflammatory processes mediated through mast cells, IgE, and group 2 innate lymphoid cells (Oliveira and Lukacs 2003; Da Silva and Frossard 2005; Da Silva *et al.* 2006; Fonseca *et al.* 2019). In the airways, *KITLG* is expressed in bronchial epithelial cells, lung fibroblasts, bronchial smooth muscle cells, endothelial cells, peripheral blood eosinophils, dendritic cells, and mast cells (Valent *et al.* 1992; Wen *et al.* 1996; Kassel *et al.* 1999; Hsieh *et al.* 2005; Oriss *et al.* 2014). *KITLG* is a major growth factor of mast cells [reviewed in Galli *et al.* (1994, 1995), Broudy

(1997), and Da Silva *et al.* (2006)]. It promotes recruitment of mast cell progenitors into tissues [reviewed in Oliveira and Lukacs (2003)], prevents mast cell apoptosis (Mekori *et al.* 1993; Iemura *et al.* 1994), and promotes release of inflammatory mediators such as proteases, histamine, chemotactic factors, and cytokines [reviewed in Borish and Joseph (1992) and Amin (2012)]. While *KITLG* promotes the production of cytokines like IL-13 upon IgE-receptor cross-linking on the surface of mast cells (Kobayashi *et al.* 1998), IL-13 has also been reported to upregulate *KITLG* (Rochman *et al.* 2015). Consistent with the critical role of *KITLG* for mast cells and type 2 inflammation, we found our prioritized variant, rs73440122, overlapped with an SPI1 (also known as PU.1) ChIP-Seq cluster. The transcription factor SPI1 was demonstrated in SPI1 knockout mice to be necessary for the development of B cells, T cells, neutrophils, macrophages, dendritic cells, and mast cells (Scott *et al.* 1994, 1997; McKercher *et al.* 1996; Anderson *et al.* 2000; Guerriero *et al.* 2000; Walsh *et al.* 2002). It plays an essential role in macrophage differentiation in asthmatic and other allergic inflammation (Qian *et al.* 2015; Yashiro *et al.* 2019). It was also shown to regulate the cell fate between mast cells and monocytes (Nishiyama *et al.* 2004a,b; Ito *et al.* 2005, 2009). The presence of an SPI1 binding site in a candidate regulatory region of *KITLG* is therefore highly relevant given the critical role of *KITLG* in mast cell survival and activation.

Higher levels of *KITLG* (Al-Muhsen *et al.* 2004; Da Silva *et al.* 2006; Tayel *et al.* 2017) and an increased number of

Table 2 Genome-wide lung function association in SAGE II children with asthma

M	rsID	Alt	β	P	β_{norm}	P_{norm}	MAF	1000 Genomes				Score
								ALL	AFR	AMR	EUR	
1	rs11835305	T	0.126	1.93E-05	0.320	3.69E-05	0.104	0.036	0.119	0.012	0.001	0
2	rs17015963	C	0.126	1.93E-05	0.320	3.69E-05	0.104	0.036	0.120	0.012	0.001	0
3	rs58475486 ^a	T	0.127	1.45E-05	0.323	2.81E-05	0.105	0.037	0.123	0.012	0.001	2 [^]
4	rs17015979	T	0.127	1.45E-05	0.323	2.81E-05	0.105	0.037	0.123	0.012	0.001	0
5	rs57692452	C	0.245	1.63E-06	0.654	1.06E-06	0.033	0.011	0.030	0.006	0.001	2 [^]
6	rs112585732	T	0.235	4.35E-07	0.625	3.06E-07	0.041	0.016	0.050	0.006	0.001	0
7	rs113837356	T	0.270	3.44E-06	0.719	2.49E-06	0.025	0.008	0.027	0.006	0.001	1
8	rs61441836	G	0.252	8.19E-07	0.671	5.46E-07	0.033	0.010	0.030	0.006	0.001	1
9	rs73438172	A	0.252	8.19E-07	0.671	5.46E-07	0.033	0.010	0.030	0.006	0.001	1
10	rs1044043958 ^b	A	0.270	3.44E-06	0.719	2.49E-06	0.025	—	—	—	—	0
11	rs73438182	G	0.138	1.68E-04	0.378	9.18E-05	0.064	0.020	0.068	0.006	0.001	0
12	rs73438185	A	0.138	1.68E-04	0.378	9.18E-05	0.064	0.020	0.068	0.006	0.001	0
13	rs73438188	A	0.297	6.97E-08	0.792	4.42E-08	0.028	0.010	0.027	0.006	0.001	1
14	rs73438190	C	0.181	1.86E-05	0.486	1.22E-05	0.048	0.016	0.047	0.006	0.001	0
15	rs73438195	A	0.181	1.86E-05	0.486	1.22E-05	0.048	0.016	0.047	0.006	0.001	0
16	rs111857459	T	0.181	1.86E-05	0.486	1.22E-05	0.048	0.016	0.047	0.006	0.001	0
17	rs144369986 ^b	T	0.285	1.21E-06	0.756	9.44E-07	0.025	0.008	0.026	0.006	0.001	0
18	rs73440106	G	0.181	1.86E-05	0.486	1.22E-05	0.048	0.016	0.047	0.006	0.001	0
19	rs73440107	A	0.297	6.97E-08	0.792	4.42E-08	0.028	0.010	0.027	0.006	0.001	1
20	rs111453514	C	0.297	6.97E-08	0.792	4.42E-08	0.028	0.010	0.027	0.006	0.001	1
21	rs73440112	T	0.297	6.97E-08	0.792	4.42E-08	0.028	0.010	0.027	0.006	0.001	1
22	rs73440115	G	0.297	6.97E-08	0.792	4.42E-08	0.028	0.011	0.028	0.006	0.001	1
23	rs11312747 ^b	A	0.133	1.43E-05	0.357	8.51E-06	0.100	0.036	0.121	0.010	0.001	0
24	rs73440120	A	0.285	1.21E-06	0.756	9.44E-07	0.025	0.008	0.026	0.006	0.001	1
25	rs111289668	G	0.297	6.97E-08	0.792	4.42E-08	0.028	0.010	0.027	0.006	0.001	2 [^]
26	rs73440122	C	0.292	2.08E-07	0.775	1.55E-07	0.027	0.011	0.030	0.006	0.001	3 [^]
27	rs73440123	G	0.292	2.08E-07	0.775	1.55E-07	0.027	0.011	0.030	0.006	0.001	1
28	rs17016065 ^a	G	0.112	3.19E-06	0.296	2.54E-06	0.177	0.075	0.217	0.017	0.009	1#
29	rs17016066	A	0.112	3.19E-06	0.296	2.54E-06	0.177	0.075	0.217	0.017	0.009	1#
30	rs147400083 ^b	T	0.112	3.19E-06	0.296	2.54E-06	0.177	—	—	—	—	0
31	rs866852270	T	0.112	3.19E-06	0.296	2.54E-06	0.177	—	—	—	—	0
32	rs141293300 ^b	C	0.292	2.08E-07	0.775	1.55E-07	0.027	0.011	0.030	0.006	0.001	1
33	rs1398303	A	0.104	1.22E-05	0.274	1.12E-05	0.186	0.077	0.223	0.020	0.009	1#
34	rs61924868	T	0.104	1.24E-05	0.275	1.14E-05	0.185	0.078	0.223	0.020	0.009	1#
35	rs73440134	T	0.292	2.08E-07	0.775	1.55E-07	0.027	0.011	0.030	0.006	0.001	1
36	rs73429413	G	0.292	2.08E-07	0.775	1.55E-07	0.027	0.011	0.030	0.006	0.001	1
37	rs73429415	A	0.096	5.13E-05	0.253	4.84E-05	0.189	0.078	0.225	0.022	0.009	1#
38	rs112449284	T	0.242	4.64E-06	0.640	3.87E-06	0.031	0.012	0.035	0.007	0.001	1
39	rs111981782	C	0.296	5.78E-08	0.786	4.09E-08	0.029	0.012	0.033	0.006	0.001	1
40	rs150942400	T	0.293	6.01E-08	0.780	4.01E-08	0.029	0.012	0.034	0.006	0.002	1
41	rs147527487	C	0.086	8.49E-05	0.226	8.14E-05	0.205	0.095	0.249	0.016	0.005	0
42	rs111243672	A	0.258	6.41E-07	0.690	3.99E-07	0.032	0.014	0.037	0.007	0.004	1
43	rs73429450 ^a	A	0.302	1.26E-08	0.801	9.01E-09	0.031	0.012	0.033	0.009	0.002	1
44	rs758775577	C	0.217	2.22E-06	0.574	1.85E-06	0.041	—	—	—	—	0
45	rs142679473 ^b	C	0.285	6.30E-08	0.756	4.62E-08	0.031	0.012	0.033	0.009	0.002	0

Score, priority score based on statistical and functional evidence, which is reported in Table S12. M, marker number that corresponds to those in Figure 2 and Table S12. Candidate variants were prioritized if they had a priority score of > 2 ([^]) or if they were expression QTL of *KITLG* in nasal epithelial cells (#). β (P) and β_{norm} (P_{norm}) are the effect sizes (P-values) of the genetic associations of the alternate allele (Alt) with FEV₁ and FEV_{1,res.norm}, respectively. ALL/AFR/AMR/EUR, 1000 Genomes minor allele frequency from all/African/American/European populations. —, not available. MAF, minor allele frequency.

^a The three independent signals identified in the conditional analyses.

^b Insertions/deletions.

mast cells in the lung (Fajt and Wenzel. 2013; Cruse and Bradding. 2016; Méndez-Enriquez and Hallgren 2019) were detected in individuals with asthma. The percentage of a subpopulation of circulating blood mast cell progenitors (Lin⁺ CD34^{hi} CD117^{int/hi} FcεRI⁺) was higher in individuals with reduced lung function (Dahlin *et al.* 2016). These findings suggested that higher *KITLG* expression and/or number of mast cells may be a contributing factor to lower lung

function. This notion was inconsistent with the association of our novel locus with higher *KITLG* expression and increased lung function in SAGE II children with asthma. Interestingly, a study of 20 subjects with severe asthma found that an increase in the number of chymase-positive mast cells in the small airway was associated with increased lung function (Balzar *et al.* 2005). Overall, while there is still controversy on the direction of effect, previous findings support the

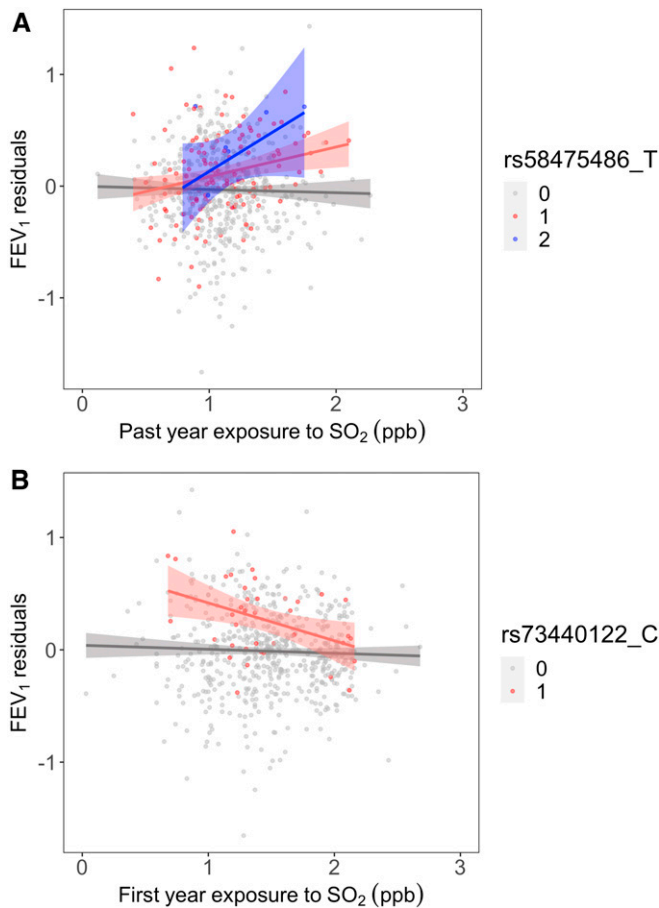


Figure 3 Gene-by-environment interaction analysis on FEV₁. FEV₁ residuals, residuals after FEV₁ was regressed on the covariates age, sex, height, controller medications, sequencing centers, and the first five genetic PCs. FEV₁ residuals were plotted against (A) past-year exposure to SO₂ stratified by the number of copies of T allele of rs58475486 and (B) first year of life exposure to SO₂ stratified by the number of copies of C allele of rs73440122. FEV₁, forced expiratory volume in the first second of exhalation; PC, principal component.

association of our novel *KITLG* locus with lung function, especially in patients with allergic asthma. Our novel locus likely represents part of a complex regulatory mechanism that modulates immune cell differentiation, survival, and activation in highly cell-specific and context-dependent manners. Further studies are required to study how this locus is regulated in different airway and immune cells to affect lung function outcome in the context of asthma.

GxE interactions likely account for a portion of the “missing” heritability of many complex phenotypes (Moore and Williams 2009). We previously found that lung function in SAGE II participants was associated with first year of life and lifetime exposures to SO₂ [1.66% decrease (95% C.I. = -2.92 to -0.37) for first year of life and 5.30% decrease (95% C.I. = -8.43 to -2.06) for lifetime exposures in FEV₁ per 1 ppb increases in SO₂] (Neophytou *et al.* 2016). We hypothesized that a significant portion of the heritability of lung function was due, in part, to gene-by-air-pollution (SO₂)

interaction effects. The interaction between rs58475486 and past-year exposure to SO₂ that was significantly associated with lung function supports our hypothesis. The T allele of rs58475486 is common (8–14%) in African populations and showed a protective effect on lung function in the presence of past-year SO₂ exposure. SNP rs58475486 is located in a ccRE (SCREEN accession EH37E0279296) and a FOXA1 binding site in the A549 lung adenocarcinoma cell line. FOXA1 has a known compensatory role with FOXA2 during lung morphogenesis in mice (Wan *et al.* 2005). Deletion of both FOXA1 and FOXA2 inhibited cell proliferation, epithelial cell differentiation, and branching morphogenesis in fetal lung tissue. Further functional validation of the effect of rs58475486 on the binding affinity of FOXA1 is necessary to confirm whether the role of FOXA1 in this ccRE is important for *KITLG* regulatory and lung function.

The higher frequencies of the protective alleles of both rs73440122 and rs58465486 in African populations appear to contradict previous findings that African ancestry is associated with lower lung function (Kumar *et al.* 2010). One possible explanation for this seeming inconsistency is that FEV₁ is a complex trait, whose variation is influenced by many genetic variants of small-to-moderate effect sizes whose influences on lung function may vary by exposure to environmental factors. We found suggestive evidence that the interaction between rs73440122 and first-year exposure to SO₂ reverses the positive association of rs73440122 with lung function to a negative one (Table 3). When assessed independently, our genetic association analysis showed that the protective A allele of rs73440122 was associated with higher lung function. However, with increasing levels of SO₂ exposure in the first year of life, increasing copies of the A allele of rs73440122 were associated with decreased lung function. Air pollution is known to negatively impact lung function, and we have previously shown that the deleterious effects of air pollution on lung phenotypes may be significantly increased in African American children compared to other populations experiencing the same amount of exposure (Nishimura *et al.* 2013). It has also been reported that Latino and African American populations often live in neighborhoods with high levels of air pollution (Mott 1995). The increased susceptibility to negative pulmonary effects from air pollution exposure coupled with the disproportionate exposure to air pollution experienced by the African American population may also contribute to the lower lung function seen in this population, despite the presence of protective alleles. The overlap of the SPI1 binding site with rs73440122 further supports gene-by-SO₂ interaction at this locus, since SPI1 plays a critical role in the development of type 2 inflammation in the airways through macrophage polarization (Qian *et al.* 2015). We noted that the rs73440122 A allele also showed an interaction approaching suggestive threshold with past-year exposure to SO₂ that was positively associated with FEV₁. The difference is not surprising because age of exposure may significantly impact the effect of air pollution on lung function [reviewed in Usemann *et al.* (2019)]. Further studies are

Table 3 Gene-and-environment analysis on FEV₁

Variant	Exposure	n	Variant		Exposure		GxE	
			β	P	β	P	β	P
rs58475486_T ¹	SO ₂ first year	640	0.13	6.62E-06	-0.05	0.003	-0.03	0.658
rs57692452_C ²	SO ₂ first year	640	0.25	1.16E-06	-0.05	0.003	-0.24	0.091
rs111289668_G ²	SO ₂ first year	640	0.31	2.53E-08	-0.05	0.003	-0.27	0.079
rs73440122_C ^{2a}	SO ₂ first year	640	0.31	7.78E-08	-0.05	0.003	-0.32	0.045
rs17016065_G ³	SO ₂ first year	640	0.11	8.82E-06	-0.05	0.003	-0.08	0.108
rs17016066_A ³	SO ₂ first year	640	0.11	8.82E-06	-0.05	0.003	-0.08	0.108
rs1398303_A ³	SO ₂ first year	640	0.10	3.15E-05	-0.05	0.003	-0.09	0.082
rs61924868_T ³	SO ₂ first year	640	0.10	3.30E-05	-0.05	0.003	-0.08	0.088
rs73429415_A ³	SO ₂ first year	640	0.09	1.09E-04	-0.05	0.003	-0.09	0.069
rs58475486_T ^{1b}	SO ₂ past year	661	0.13	6.62E-06	0.05	0.362	0.32	0.003
rs57692452_C ²	SO ₂ past year	661	0.25	1.16E-06	0.05	0.362	0.29	0.100
rs111289668_G ^{2a}	SO ₂ past year	661	0.31	2.53E-08	0.05	0.362	0.41	0.037
rs73440122_C ²	SO ₂ past year	661	0.31	7.78E-08	0.05	0.362	0.39	0.051
rs17016065_G ^{3a}	SO ₂ past year	661	0.11	8.82E-06	0.05	0.362	0.20	0.026
rs17016066_A ^{3a}	SO ₂ past year	661	0.11	8.82E-06	0.05	0.362	0.20	0.026
rs1398303_A ^{3a}	SO ₂ past year	661	0.10	3.15E-05	0.05	0.362	0.21	0.021
rs61924868_T ^{3a}	SO ₂ past year	661	0.10	3.30E-05	0.05	0.362	0.21	0.023
rs73429415_A ^{3a}	SO ₂ past year	661	0.09	1.09E-04	0.05	0.362	0.20	0.026
rs58475486_T ¹	SO ₂ lifetime	661	0.13	6.62E-06	-0.13	0.001	0.26	0.173
rs57692452_C ²	SO ₂ lifetime	661	0.25	1.16E-06	-0.13	0.001	0.47	0.221
rs111289668_G ²	SO ₂ lifetime	661	0.31	2.53E-08	-0.13	0.001	0.32	0.444
rs73440122_C ²	SO ₂ lifetime	661	0.31	7.78E-08	-0.13	0.001	0.29	0.489
rs17016065_G ³	SO ₂ lifetime	661	0.11	8.82E-06	-0.13	0.001	-0.19	0.143
rs17016066_A ³	SO ₂ lifetime	661	0.11	8.82E-06	-0.13	0.001	-0.19	0.143
rs1398303_A ³	SO ₂ lifetime	661	0.10	3.15E-05	-0.13	0.001	-0.17	0.184
rs61924868_T ³	SO ₂ lifetime	661	0.10	3.30E-05	-0.13	0.001	-0.17	0.199
rs73429415_A ³	SO ₂ lifetime	661	0.09	1.09E-04	-0.13	0.001	-0.16	0.207

n, sample sizes for the gene-by-SO₂ interaction analysis. Superscript 1 to 3 in the variant column, variants that are in linkage disequilibrium with the three independent signals, rs58475486, rs73429415, and rs17016065, respectively. β (P), effect sizes (P - values) from the main effects of the variants, exposure, and GxE interaction. GxE, gene-environment interaction.

^a GxE P < 0.05.

^b GxE P < Bonferroni P-value of 0.0056.

required to better understand the effect of this suggestive interaction on lung function.

One strength of this study is the interrogation of independent lung function-associated signals at our novel locus. We identified evidence of three independent signals: the replicated signal that showed evidence of regulatory functions (an open chromatin region with an SPI1/PU.1 binding site), one signal that showed a statistically significant gene-by-SO₂ interaction on lung function, and one signal that represents *KITLG* eQTLs in the nasal epithelial cells together with suggestive gene-by-SO₂ interaction. Our results demonstrate a glimpse of the complicated genetic architecture behind complex traits.

One limitation of this study is that the FEV₁ genetic association and the eQTL analyses with *KITLG* were performed in different populations due to data availability constraints. Although we did not have RNA-Seq data from lung tissues from our study subjects, we previously demonstrated that there is a high degree of overlap in gene expression profiles between nasal and bronchial epithelial cells (Poole *et al.* 2014). The direction of effect of the association was the same in GALA II Puerto Rican children with asthma but not statistically significant. This may be due to: (1) the lower minor allele frequency in Puerto Ricans with significantly lower African

Ancestry compared to African Americans, and (2) the modest sample size of the replication study and the weak effect of the protective allele can lead to undetectable true associations (Altshuler *et al.* 2000).

We replicated 20 of 45 variants in black UK Biobank subjects and observed conflicting “flip-flop” associations in African Americans from the SAPPHERE and GCPD-A studies. In the past, flip-flop associations were deemed as spurious results. The traditional association-testing approach studies the effect of each variant on phenotype independently and increases the chance of flip-flop associations detected between studies. Differences in study design, sampling variation that leads to variation in LD patterns, and lack of consideration of other disease-influencing genetic and/or environmental factors are all potential causes of flip-flop associations (Lin *et al.* 2007; Kraft *et al.* 2009). Hence, it is not surprising that flip-flop associations were observed when gene and environment interactions were detected at our FEV₁ GWAS locus. It was previously shown that flip-flop associations can occur between and within populations, even in the presence of a genuine genetic effect (Lin *et al.* 2007; Kraft *et al.* 2009). Further functional analysis is thus required to validate the relationship between the candidate variants *KITLG* and FEV₁. This may include reporter assays to validate

potential enhancer or repressor activity, and clustered regularly interspaced short palindromic repeats-based editing assays to validate the regulatory role of the candidate variants on *KITLG*. Although literature exists describing KIT signaling for lung function in mice (Lindsey *et al.* 2011), additional knockout experiments in a model animal system are necessary to study how *KITLG* contributes to variation in lung function.

The average concentration of ambient SO₂ exposure in our participants (Table 1) was lower than National Ambient Air Quality Standards. It is possible that SO₂ acted as a surrogate for other unmeasured toxic pollutants emitted from local point sources. Major sources of SO₂ in the San Francisco Bay Area during the recruitment years of 2006–2011 include airports, petroleum refineries, gas and oil plants, calcined petroleum coke plants, electric power plants, cement manufacturing factories, chemical plants, and landfills (United States Environmental Protection Agency 2008, 2011). The Environmental Protection Agency's national emissions inventory data also show that these facilities emit volatile organic compounds, heavy metals (lead, mercury, chromium, and arsenic), formaldehyde, ethyl benzene, acrolein, 1,3-butadiene, 1,4-dichlorobenzene, and tetrachloroethylene into the air along with SO₂. These chemicals are highly toxic and inhaling even a small amount may contribute to poor lung function. Another possibility is that exposure to SO₂ captures unmeasured confounding socioeconomic factors.

This study identified a novel protective allele for lung function in African American children with asthma. The protective association with lung function intensified with increased past-year exposure to SO₂. Our findings showcase the complexity of the relationship between genetic and environmental factors impacting variation in FEV₁, highlights the utility of WGS data for genetic research of complex phenotypes, and underscores the importance of including diverse study populations in our exploration of the genetic architecture underlying lung function.

Acknowledgments

The authors thank the studies and participants who provided biological samples and data for TOPMed; the recruiters who obtained the data: Duanny Alva, Gaby Ayala-Rodríguez, Lisa Caine, Elizabeth Castellanos, Jaime Colón, Denise DeJesus, Blanca López, Brenda López, Louis Martos, Vivian Medina, Juana Olivo, Mario Peralta, Esther Pomares, Jihan Quraishi, Johanna Rodríguez, Shahdad Saeedi, Dean Soto, and Ana Taveras; María Pino-Yanes for providing feedback on this study; Thomas W Blackwell for providing critical review on this manuscript; Biobank participants and researchers who contributed or collected data; and the Editors and Reviewers for providing valuable feedback on this manuscript. They authors also acknowledge the following GALA II and SAGE II study collaborators: Shannon Thyne, University of California, San Francisco (UCSF); Harold

J. Farber, Texas Children's Hospital; Denise Serebrisky, Jacobi Medical Center; Rajesh Kumar, Lurie Children's Hospital of Chicago; Emerita Brigino-Buenaventura, Kaiser Permanente; Michael A. LeNoir, Bay Area Pediatrics; Kelley Meade, UCSF Benioff Children's Hospital, Oakland; William Rodríguez-Cintrón, VA Hospital, Puerto Rico; Pedro C. Ávila, Northwestern University; Jose R. Rodríguez-Santana, Centro de Neumología Pediátrica; Luisa N. Borrell, City University of New York; Adam Davis, UCSF Benioff Children's Hospital, Oakland; and Saunak Sen, University of Tennessee; the families and patients for their participation; and thank the numerous healthcare providers and community clinics for their support, and participation, in GALA II and SAGE II. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The GALA II study, the SAGE II study, and E.G.B. were supported by the Sandler Family Foundation, the [American Asthma Foundation](#), the RWJF Amos Medical Faculty Development Program, the Harry W. and Diana V. Hind Distinguished Professorship in Pharmaceutical Sciences II, the [National Heart, Lung, and Blood Institute \(NHLBI\)](#) (R01 HL-117004, R01 HL-128439, R01 HL-135156, and X01 HL-134589), the [National Institute of Environmental Health Sciences](#) (R01 ES-015794), the National Institute on [Minority Health and Health Disparities](#) (P60 MD-006902 and R01 MD-010443), the [National Human Genome Research Institute \(NHGRI\)](#) (U01 HG-009080), and the [Tobacco-Related Disease Research Program](#) (24RT-0025). M.J.W. was supported by the [NHLBI](#) (K01 HL-140218). J.J. and B.E.H. were supported by the [NHLBI](#) (R01 HL-133433 and R01 HL-141992). K.L.K. was supported by the [NHLBI](#) (R01 HL-135156-S1), the UCSF Bakar Institute, the [Gordon and Betty Moore Foundation](#) (GBMF3834), and the [Alfred P. Sloan Foundation](#) (2013-10-27) grant to the University of California, Berkeley through the Moore-Sloan Data Science Environment Initiative. A.C.W. was supported by the [Eunice Kennedy Shriver National Institute of Child Health and Human Development](#) (1R01 HD-085993-01). The SAPHIRE study was supported by the Fund for Henry Ford Hospital, the American Asthma Foundation, the [NHLBI](#) (R01 HL-118267, R01 HL-141485, and X01 HL-134589), the [National Institute of Allergy and Infectious Diseases](#) (R01 AI-079139), and the [National Institute of Diabetes and Digestive and Kidney Diseases](#) (R01 DK-113003). The GCPD-A study was supported by an Institutional award from the [Children's Hospital of Philadelphia](#) and by the [NHLBI](#) (X01 HL-134589). Part of this research was conducted using the UK Biobank Resource under application number 40375. WGS for the TOPMed program was supported by the [NHLBI](#). WGS for "NHLBI TOPMed: Gene-Environment, Admixture and Latino Asthmatics Study" (phs000920) and "NHLBI TOPMed: Study of African Americans, Asthma, Genes and Environments" (phs000921) was performed at the New York Genome Center (3R01 HL-117004-02S3) and the University of Washington Northwest Genomics Center

(HHSN268201600032I). WGS for “NHLBI TOPMed: Study of Asthma Phenotypes & Pharmacogenomic Interactions by Race-Ethnicity” (phs001467) and “Genetics of Complex Pediatric Disorders - Asthma” (phs001661) was performed at the University of Washington Northwest Genomics Center (HHSN268201600032I). Centralized read mapping and genotype calling, along with variant quality metrics and filtering, were provided by the TOPMed Informatics Research Center (3R01 HL-117626-02S1; contract HHSN268201800002I). Phenotype harmonization, data management, sample identity quality control, and general study coordination were provided by the TOPMed Data Coordinating Center (3R01 HL-120393-02S1; contract HHSN268201800001I). WGS of part of GALA II was performed by New York Genome Center under The Centers for Common Disease Genomics (CCDG) of the Genome Sequencing Program (GSP) grant (UM1 HG-008901). The GSP Coordinating Center (U24 HG-008956) contributed to cross-program scientific initiatives, and provided logistical and general study coordination. GSP is funded by the NHGRI, the NHLBI, and the National Eye Institute. The TOPMed imputation panel was supported by the NHLBI and TOPMed study investigators who contributed data to the reference panel. The panel was constructed and implemented by the TOPMed Informatics Research Center at the University of Michigan (3R01 HL-117626-02S1; contract HHSN268201800002I). The TOPMed Data Coordinating Center (3R01 HL-120393-02S1; contract HHSN268201800001I) provided additional data management, sample identity checks, and overall program coordination and support.

Author contributions: A.C.Y.M., S. Sajuthi, M.J.W., S.S.O., F.G., Z.C., R.K., F.D.M., A.C.W., E.Z., and E.G.B. designed and supervised the study. A.C.Y.M., S. Sajuthi, J.J., S.X., P.M.S., M.J.W., E.Y.L., B.S., D.H., S.H., B.E.H., H.H., L.K.W., and M.A.S. supervised/performed analyses and/or interpreted results. C.E. processed biospecimens for WGS. F.L. estimated air pollution exposure. D.J., G.A., H.M.K., D.A.N., and S.G. contributed to TOPMed WGS data generation and/or analysis. M.Z., L.W., and C.R. contributed to CCDG WGS data generation. S. Salazar coordinated study recruitment and processed phenotype data. A.C.Y.M., S. Sajuthi, J.J., S.X., P.M.S., M.J.W., E.Y.L., B.S., H.G., K.L.K., S.G., B.E.H., H.H., L.K.W., M.A.S., and E.G.B. wrote and/or critically reviewed the manuscript.

Literature Cited

- Akinbami, L. J., 2015 Asthma prevalence, health care use and mortality: United States, 2003–05. Available at: <https://www.cdc.gov/nchs/data/hestat/asthma03-05/asthma03-05.htm>. Accessed January 8, 2020.
- Akinbami, L. J., J. E. Moorman, A. E. Simon, and K. C. Schoendorf, 2014 Trends in racial disparities for asthma outcomes among children 0 to 17 years, 2001–2010. *J. Allergy Clin. Immunol.* 134: 547–553.e5. <https://doi.org/10.1016/j.jaci.2014.05.037>
- Alexander, D. H., J. Novembre, and K. Lange, 2009 Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 19: 1655–1664. <https://doi.org/10.1101/gr.094052.109>
- Al-Muhsen, S. Z., G. Shablovsky, R. Olivenstein, B. Mazer, and Q. Hamid, 2004 The expression of stem cell factor and c-kit receptor in human asthmatic airways. *Clin. Exp. Allergy* 34: 911–916. <https://doi.org/10.1111/j.1365-2222.2004.01975.x>
- Altshuler, D., J. N. Hirschhorn, M. Klannemark, C. M. Lindgren, M. C. Vohl *et al.*, 2000 The common PPARgamma Pro12Ala polymorphism is associated with decreased risk of type 2 diabetes. *Nat. Genet.* 26: 76–80. <https://doi.org/10.1038/79216>
- Amin, K., 2012 The role of mast cells in allergic inflammation. *Respir. Med.* 106: 9–14. <https://doi.org/10.1016/j.rmed.2011.09.007>
- Anders, S., P. T. Pyl, and W. Huber, 2015 HTSeq—a python framework to work with high-throughput sequencing data. *Bioinformatics* 31: 166–169. <https://doi.org/10.1093/bioinformatics/btu638>
- Anderson, K. L., H. Perkin, C. D. Surh, S. Venturini, R. A. Maki *et al.*, 2000 Transcription factor PU.1 is necessary for development of thymic and myeloid progenitor-derived dendritic cells. *J. Immunol.* 164: 1855–1861. <https://doi.org/10.4049/jimmunol.164.4.1855>
- Ay, F., T. L. Bailey, and W. S. Noble, 2014 Statistical confidence estimation for hi-C data reveals regulatory chromatin contacts. *Genome Res.* 24: 999–1011. <https://doi.org/10.1101/gr.160374.113>
- Balzar, S., H. W. Chu, M. Strand, and S. Wenzel, 2005 Relationship of small airway chymase-positive mast cells and lung function in severe asthma. *Am. J. Respir. Crit. Care Med.* 171: 431–439. <https://doi.org/10.1164/rccm.200407-949OC>
- Barraza-Villarreal, A., J. Sunyer, L. Hernandez-Cadena, M. C. Escamilla-Nunez, J. J. Sienna-Monge *et al.*, 2008 Air pollution, airway inflammation, and lung function in a cohort study of Mexico City schoolchildren. *Environ. Health Perspect.* 116: 832–838. <https://doi.org/10.1289/ehp.10926>
- Barrett, J. C., B. Fry, J. Maller, and M. J. Daly, 2005 Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 21: 263–265. <https://doi.org/10.1093/bioinformatics/bth457>
- Borish, L., and B. Z. Joseph, 1992 Inflammation and the allergic response. *Med. Clin. North Am.* 76: 765–787. [https://doi.org/10.1016/S0025-7125\(16\)30325-X](https://doi.org/10.1016/S0025-7125(16)30325-X)
- Broudy, V. C., 1997 Stem cell factor and hematopoiesis. *Blood* 90: 1345–1364. <https://doi.org/10.1182/blood.V90.4.1345>
- Brunekreef, B., and S. T. Holgate, 2002 Air pollution and health. *Lancet* 360: 1233–1242. [https://doi.org/10.1016/S0140-6736\(02\)11274-8](https://doi.org/10.1016/S0140-6736(02)11274-8)
- Buniello, A., J. A. L. MacArthur, M. Cerezo, L. W. Harris, J. Hayhurst *et al.*, 2019 The NHGRI-EBI GWAS catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res.* 47: D1005–D1012. <https://doi.org/10.1093/nar/gky1120>
- Bycroft, C., C. Freeman, D. Petkova, G. Band, L. T. Elliott *et al.*, 2018 The UK biobank resource with deep phenotyping and genomic data. *Nature* 562: 203–209. <https://doi.org/10.1038/s41586-018-0579-z>
- Canela-Xandri, O., K. Rawlik, and A. Tenesa, 2018 An atlas of genetic associations in UK Biobank. *Nat. Genet.* 50: 1593–1599. <https://doi.org/10.1038/s41588-018-0248-z>
- Carlson, C. S., T. C. Matise, K. E. North, C. A. Haiman, M. D. Fesinmeyer *et al.*, 2013 Generalization and dilution of association results from European GWAS in populations of non-European ancestry: the PAGE study. *PLoS Biol.* 11: e1001661. <https://doi.org/10.1371/journal.pbio.1001661>
- Chang, C. C., C. C. Chow, L. C. Tellier, S. Vattikuti, S. M. Purcell *et al.*, 2015 Second-generation PLINK: Rising to the challenge of larger and richer datasets. *Gigascience* 4: 7. <https://doi.org/10.1186/s13742-015-0047-8>

- Chatterjee, S., and N. Das, 1995 Lung function in Indian twin children: comparison of genetic vs. environmental influence. *Ann. Hum. Biol.* 22: 289–303. <https://doi.org/10.1080/03014469500003962>
- Chen, Y., S. L. Horne, D. C. Rennie, and J. A. Dosman, 1996 Segregation analysis of two lung function indices in a random sample of young families: the Humboldt Family Study. *Genet. Epidemiol.* 13: 35–47. [https://doi.org/10.1002/\(SICI\)1098-2272\(1996\)13:1<35::AID-GEPI4>3.0.CO;2-5](https://doi.org/10.1002/(SICI)1098-2272(1996)13:1<35::AID-GEPI4>3.0.CO;2-5)
- Ciesielski, T. H., S. A. Pendergrass, M. J. White, N. Kodaman, R. S. Sobota *et al.*, 2014 Diverse convergent evidence in the genetic analysis of complex disease: coordinating omic, informatic, and experimental evidence to better identify and validate risk factors. *BioData Min.* 7: 10. <https://doi.org/10.1186/1756-0381-7-10>
- Comberiati, P., M. E. Di Cicco, S. D'Elisio, and D. G. Peroni, 2017 How much asthma is atopic in children? *Front Pediatr.* 5: 122. <https://doi.org/10.3389/fped.2017.00122>
- Conomos, M. P., M. B. Miller, and T. A. Thornton, 2015 Robust inference of population structure for ancestry prediction and correction of stratification in the presence of relatedness. *Genet. Epidemiol.* 39: 276–293. <https://doi.org/10.1002/gepi.21896>
- Conomos, M. P., A. P. Reiner, B. S. Weir, and T. A. Thornton, 2016 Model-free estimation of recent genetic relatedness. *Am. J. Hum. Genet.* 98: 127–148. <https://doi.org/10.1016/j.ajhg.2015.11.022>
- Cruse, G., and P. Bradding, 2016 Mast cells in airway diseases and interstitial lung disease. *Eur. J. Pharmacol.* 778: 125–138. <https://doi.org/10.1016/j.ejphar.2015.04.046>
- Dahlin, J. S., A. Malinowski, H. Ohrvik, M. Sandelin, C. Janson *et al.*, 2016 Lin- CD34hi CD117int/hi FcεRI+ cells in human blood constitute a rare population of mast cell progenitors. *Blood* 127: 383–391. <https://doi.org/10.1182/blood-2015-06-650648>
- Das, S., L. Forer, S. Schonherr, C. Sidore, A. E. Locke *et al.*, 2016 Next-generation genotype imputation service and methods. *Nat. Genet.* 48: 1284–1287. <https://doi.org/10.1038/ng.3656>
- Da Silva, C. A., and N. Frossard, 2005 Regulation of stem cell factor expression in inflammation and asthma. *Mem. Inst. Oswaldo Cruz* 100: 145–151. <https://doi.org/10.1590/S0074-02762005000900025>
- Da Silva, C. A., L. Reber, and N. Frossard, 2006 Stem cell factor expression, mast cells and inflammation in asthma. *Fundam. Clin. Pharmacol.* 20: 21–39. <https://doi.org/10.1111/j.1472-8206.2005.00390.x>
- Duggal, P., E. M. Gillanders, T. N. Holmes, and J. E. Bailey-Wilson, 2008 Establishing an adjusted p-value threshold to control the family-wide type 1 error in genome wide association studies. *BMC Genomics* 9: 516. <https://doi.org/10.1186/1471-2164-9-516>
- ENCODE Project Consortium, 2011 A user's guide to the encyclopedia of DNA elements (ENCODE). *PLoS Biol.* 9: e1001046. <https://doi.org/10.1371/journal.pbio.1001046>
- ENCODE Project Consortium, 2012 An integrated encyclopedia of DNA elements in the human genome. *Nature* 489: 57–74. <https://doi.org/10.1038/nature11247>
- Fajt, M. L., and S. E. Wenzel, 2013 Mast cells, their subtypes, and relation to asthma phenotypes. *Ann. Am. Thorac. Soc.* 10: S158–S164. <https://doi.org/10.1513/AnnalsATS.201303-064AW>
- Fonseca, W., A. J. Rasky, C. Ptaschinski, S. H. Morris, S. K. K. Best *et al.*, 2019 Group 2 innate lymphoid cells (ILC2) are regulated by stem cell factor during chronic asthmatic disease. *Mucosal Immunol.* 12: 445–456. <https://doi.org/10.1038/s41385-018-0117-1>
- Galli, S. J., K. M. Zsebo, and E. N. Geissler, 1994 The kit ligand, stem cell factor. *Adv. Immunol.* 55: 1–96.
- Galli, S. J., M. Tsai, B. K. Wershil, S. Y. Tam, and J. J. Costa, 1995 Regulation of mouse and human mast cell development, survival and function by stem cell factor, the ligand for the c-kit receptor. *Int. Arch. Allergy Immunol.* 107: 51–53. <https://doi.org/10.1159/000236928>
- 1000 Genomes Project Consortium, A. Auton, L. D. Brooks, R. M. Durbin, E. P. Garrison, *et al.*, 2015 A global reference for human genetic variation. *Nature* 526: 68–74. <https://doi.org/10.1038/nature15393>
- GTEx Consortium; Laboratory, Data Analysis & Coordinating Center (LDACC)—Analysis Working Group; Statistical Methods groups—Analysis Working Group; Enhancing GTEx (eGTEx) groups; NIH Common Fund, 2017 Genetic effects on gene expression across human tissues. *Nature* 550: 204–213 [corrigenda: *Nature* 553: 530 (2018)]. <https://doi.org/10.1038/nature24277>
- Guerrero, A., P. B. Langmuir, L. M. Spain, and E. W. Scott, 2000 PU.1 is required for myeloid-derived but not lymphoid-derived dendritic cells. *Blood* 95: 879–885. https://doi.org/10.1182/blood.V95.3.879.003k13_879_885
- Hankinson, J. L., J. R. Odencrantz, and K. B. Fedan, 1999 Spirometric reference values from a sample of the general U.S. population. *Am. J. Respir. Crit. Care Med.* 159: 179–187. <https://doi.org/10.1164/ajrccm.159.1.9712108>
- Hinrichs, A. S., D. Karolchik, R. Baertsch, G. P. Barber, G. Bejerano *et al.*, 2006 The UCSC genome browser database: update 2006. *Nucleic Acids Res.* 34: D590–D598. <https://doi.org/10.1093/nar/gkj144>
- Hsieh, F. H., P. Sharma, A. Gibbons, T. Goggans, S. C. Erzurum *et al.*, 2005 Human airway epithelial cell determinants of survival and functional phenotype for primary human mast cells. *Proc. Natl. Acad. Sci. USA* 102: 14380–14385. <https://doi.org/10.1073/pnas.0503948102>
- Hukkinen, M., J. Kaprio, U. Broms, A. Viljanen, D. Kotz *et al.*, 2011 Heritability of lung function: a twin study among never-smoking elderly women. *Twin Res. Hum. Genet.* 14: 401–407. <https://doi.org/10.1375/twin.14.5.401>
- Iemura, A., M. Tsai, A. Ando, B. K. Wershil, and S. J. Galli, 1994 The c-kit ligand, stem cell factor, promotes mast cell survival by suppressing apoptosis. *Am. J. Pathol.* 144: 321–328.
- Ierodiakonou, D., A. Zanobetti, B. A. Coull, S. Melly, D. S. Postma *et al.*, 2016 Ambient air pollution, lung function, and airway responsiveness in asthmatic children. *J. Allergy Clin. Immunol.* 137: 390–399. <https://doi.org/10.1016/j.jaci.2015.05.028>
- Ionita-Laza, I., S. Lee, V. Makarov, J. D. Buxbaum, and X. Lin, 2013 Sequence kernel association tests for the combined effect of rare and common variants. *Am. J. Hum. Genet.* 92: 841–853. <https://doi.org/10.1016/j.ajhg.2013.04.015>
- Ito, T., C. Nishiyama, M. Nishiyama, H. Matsuda, K. Maeda *et al.*, 2005 Mast cells acquire monocyte-specific gene expression and monocyte-like morphology by overproduction of PU.1. *J. Immunol.* 174: 376–383. <https://doi.org/10.4049/jimmunol.174.1.376>
- Ito, T., C. Nishiyama, N. Nakano, M. Nishiyama, Y. Usui *et al.*, 2009 Roles of PU.1 in monocyte- and mast cell-specific gene regulation: PU.1 transactivates CIITA pIV in cooperation with IFN-γ. *Int. Immunol.* 21: 803–816. <https://doi.org/10.1093/intimm/dxp048>
- Jiang, H., R. Lei, S. W. Ding, and S. Zhu, 2014 Skewer: a fast and accurate adapter trimmer for next-generation sequencing paired-end reads. *BMC Bioinformatics* 15: 182. <https://doi.org/10.1186/1471-2105-15-182>
- Johnson, J. D., and W. M. Theurer, 2014 A stepwise approach to the interpretation of pulmonary function tests. *Am. Fam. Physician* 89: 359–366.
- Karolchik, D., A. S. Hinrichs, T. S. Furey, K. M. Roskin, C. W. Sugnet *et al.*, 2004 The UCSC table browser data retrieval tool. *Nucleic Acids Res.* 32: D493–D496. <https://doi.org/10.1093/nar/gkh103>
- Kassel, O., F. Schmidlin, C. Duvernelle, B. Gasser, G. Massard *et al.*, 1999 Human bronchial smooth muscle cells in culture produce

- stem cell factor. *Eur. Respir. J.* 13: 951–954. <https://doi.org/10.1034/j.1399-3003.1999.13e04.x>
- Kim, D., B. Langmead, and S. L. Salzberg, 2015 HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* 12: 357–360. <https://doi.org/10.1038/nmeth.3317>
- Kim, M. S., K. P. Patel, A. K. Teng, A. J. Berens, and J. Lachance, 2018 Genetic disease risks can be misestimated across global populations. *Genome Biol.* 19: 179. <https://doi.org/10.1186/s13059-018-1561-7>
- Kobayashi, H., Y. Okayama, T. Ishizuka, R. Pawankar, C. Ra *et al.*, 1998 Production of IL-13 by human lung mast cells in response to fcepsilon receptor cross-linkage. *Clin. Exp. Allergy* 28: 1219–1227. <https://doi.org/10.1046/j.1365-2222.1998.00377.x>
- Kraft, P., E. Zeggini, and J. P. Ioannidis, 2009 Replication in genome-wide association studies. *Stat. Sci.* 24: 561–573. <https://doi.org/10.1214/09-STS290>
- Kumar, R., M. A. Seibold, M. C. Aldrich, L. K. Williams, A. P. Reiner *et al.*, 2010 Genetic ancestry in lung-function predictions. *N. Engl. J. Med.* 363: 321–330. <https://doi.org/10.1056/NEJMoa0907897>
- Lee, S., M. J. Emond, M. J. Bamshad, K. C. Barnes, M. J. Rieder *et al.*, 2012 Optimal unified approach for rare-variant association testing with application to small-sample case-control whole-exome sequencing studies. *Am. J. Hum. Genet.* 91: 224–237. <https://doi.org/10.1016/j.ajhg.2012.06.007>
- Levin, A. M., Y. Wang, K. E. Wells, B. Padhukasahasram, J. J. Yang *et al.*, 2014 Nocturnal asthma and the importance of race/ethnicity and genetic ancestry. *Am. J. Respir. Crit. Care Med.* 190: 266–273.
- Li, X., G. A. Hawkins, E. J. Ampleford, W. C. Moore, H. Li *et al.*, 2013 Genome-wide association study identifies TH1 pathway genes associated with lung function in asthmatic patients. *J. Allergy Clin. Immunol.* 132: 313–320.e15. <https://doi.org/10.1016/j.jaci.2013.01.051>
- Liao, S. Y., X. Lin, and D. C. Christiani, 2014 Genome-wide association and network analysis of lung function in the Framingham Heart Study. *Genet. Epidemiol.* 38: 572–578. <https://doi.org/10.1002/gepi.21841>
- Lin, P. I., J. M. Vance, M. A. Pericak-Vance, and E. R. Martin, 2007 No gene is an island: the flip-flop phenomenon. *Am. J. Hum. Genet.* 80: 531–538. <https://doi.org/10.1086/512133>
- Lindsey, J. Y., K. Ganguly, D. M. Brass, Z. Li, E. N. Potts *et al.*, 2011 C-kit is essential for alveolar maintenance and protection from emphysema-like disease in mice. *Am. J. Respir. Crit. Care Med.* 183: 1644–1652. <https://doi.org/10.1164/rccm.201007-1157OC>
- Liu, X., S. White, B. Peng, A. D. Johnson, J. A. Brody *et al.*, 2016 WGS: an annotation pipeline for human genome sequencing studies. *J. Med. Genet.* 53: 111–112. <https://doi.org/10.1136/jmedgenet-2015-103423>
- Mak, A. C. Y., M. J. White, C. Eng, D. Hu, S. Huntsman *et al.*, 2016 Whole genome sequencing to identify genetic variation associated with bronchodilator response in minority children with asthma. *Am. J. Respir. Crit. Care Med.* 193: A3709.
- Mak, A. C. Y., M. J. White, W. L. Eckalbar, Z. A. Szpiech, S. S. Oh *et al.*, 2018 Whole-genome sequencing of pharmacogenetic drug response in racially diverse children with asthma. *Am. J. Respir. Crit. Care Med.* 197: 1552–1564. <https://doi.org/10.1164/rccm.201712-2529OC>
- Martin, A. R., C. R. Gignoux, R. K. Walters, G. L. Wojcik, B. M. Neale *et al.*, 2017 Human demographic history impacts genetic risk prediction across diverse populations. *Am. J. Hum. Genet.* 100: 635–649. <https://doi.org/10.1016/j.ajhg.2017.03.004>
- Martin, J. S., Z. Xu, A. P. Reiner, K. L. Mohlke, P. Sullivan *et al.*, 2017 HUGIn: Hi-C unifying genomic interrogator. *Bioinformatics* 33: 3793–3795. <https://doi.org/10.1093/bioinformatics/btx359>
- McKercher, S. R., B. E. Torbett, K. L. Anderson, G. W. Henkel, D. J. Vestal *et al.*, 1996 Targeted disruption of the PU.1 gene results in multiple hematopoietic abnormalities. *EMBO J.* 15: 5647–5658. <https://doi.org/10.1002/j.1460-2075.1996.tb00949.x>
- Mekori, Y. A., C. K. Oh, and D. D. Metcalfe, 1993 IL-3-dependent murine mast cells undergo apoptosis on removal of IL-3. prevention of apoptosis by c-kit ligand. *J. Immunol.* 151: 3775–3784.
- Méndez-Enriquez, E., and J. Hallgren, 2019 Mast cells and their progenitors in allergic asthma. *Front. Immunol.* 10: 821. <https://doi.org/10.3389/fimmu.2019.00821>
- Moore, J. H., 2005 A global view of epistasis. *Nat. Genet.* 37: 13–14. <https://doi.org/10.1038/ng0105-13>
- Moore, J. H., and S. M. Williams, 2009 Epistasis and its implications for personal genetics. *Am. J. Hum. Genet.* 85: 309–320. <https://doi.org/10.1016/j.ajhg.2009.08.006>
- Mott, L., 1995 The disproportionate impact of environmental health threats on children of color. *Environ. Health Perspect.* 103: 33–35.
- National Heart, Lung and Blood Institute Trans-Omics for Precision Medicine (TOPMed) Program, 2019 TOPMed whole genome sequencing methods: freeze 8. Available at: <https://www.nhlbiwgs.org/topmed-whole-genome-sequencing-methods-freeze-8>. Accessed December 13, 2019.
- Neophytou, A. M., M. J. White, S. S. Oh, N. Thakur, J. M. Galanter *et al.*, 2016 Air pollution and lung function in minority youth with asthma in the GALA II (Genes-Environments and Admixture in Latino Americans) and SAGE II (Study of African Americans, Asthma, Genes, and Environments) studies. *Am. J. Respir. Crit. Care Med.* 193: 1271–1280. <https://doi.org/10.1164/rccm.201508-1706OC>
- Nishimura, K. K., J. M. Galanter, L. A. Roth, S. S. Oh, N. Thakur *et al.*, 2013 Early-life air pollution and asthma risk in minority children. the GALA II and SAGE II studies. *Am. J. Respir. Crit. Care Med.* 188: 309–318. <https://doi.org/10.1164/rccm.201302-0264OC>
- Nishiyama, C., M. Nishiyama, T. Ito, S. Masaki, N. Masuoka *et al.*, 2004a Functional analysis of PU.1 domains in monocyte-specific gene regulation. *FEBS Lett.* 561: 63–68. [https://doi.org/10.1016/S0014-5793\(04\)00116-4](https://doi.org/10.1016/S0014-5793(04)00116-4)
- Nishiyama, C., M. Nishiyama, T. Ito, S. Masaki, K. Maeda *et al.*, 2004b Overproduction of PU.1 in mast cell progenitors: its effect on monocyte- and mast cell-specific gene expression. *Biochem. Biophys. Res. Commun.* 313: 516–521. <https://doi.org/10.1016/j.bbrc.2003.11.145>
- Oh, S. S., H. Tcheurekdjian, L. A. Roth, E. A. Nguyen, S. Sen *et al.*, 2012 Effect of secondhand smoke on asthma control among black and Latino children. *J. Allergy Clin. Immunol.* 129: 1478–1483.e7. <https://doi.org/10.1016/j.jaci.2012.03.017>
- Oh, S. S., M. J. White, C. R. Gignoux, and E. G. Burchard, 2016 Making precision medicine socially precise. Take a deep breath. *Am. J. Respir. Crit. Care Med.* 193: 348–350. <https://doi.org/10.1164/rccm.201510-2045ED>
- Oliveira, S. H., and N. W. Lukacs, 2003 Stem cell factor: a hematopoietic cytokine with important targets in asthma. *Curr. Drug Targets Inflamm. Allergy* 2: 313–318. <https://doi.org/10.2174/1568010033483990>
- Ong, B. A., J. Li, J. M. McDonough, Z. Wei, C. Kim *et al.*, 2013 Gene network analysis in a pediatric cohort identifies novel lung function genes. *PLoS One* 8: e72899. <https://doi.org/10.1371/journal.pone.0072899>
- Oriss, T. B., N. Krishnamoorthy, P. Ray, and A. Ray, 2014 Dendritic cell c-kit signaling and adaptive immunity: implications for the upper airways. *Curr. Opin. Allergy Clin. Immunol.* 14: 7–12. <https://doi.org/10.1097/ACI.0000000000000019>
- Palmer, L. J., M. W. Knuiman, M. L. Divitini, P. R. Burton, A. L. James *et al.*, 2001 Familial aggregation and heritability of

- adult lung function: results from the Busselton Health Study. *Eur. Respir. J.* 17: 696–702. <https://doi.org/10.1183/09031936.01.17406960>
- Pe'er, I., R. Yelensky, D. Altshuler, and M. J. Daly, 2008 Estimation of the multiple testing burden for genomewide association studies of nearly all common variants. *Genet. Epidemiol.* 32: 381–385. <https://doi.org/10.1002/gepi.20303>
- Pino-Yanes, M., N. Thakur, C. R. Gignoux, J. M. Galanter, L. A. Roth *et al.*, 2015 Genetic ancestry influences asthma susceptibility and lung function among Latinos. *J. Allergy Clin. Immunol.* 135: 228–235. <https://doi.org/10.1016/j.jaci.2014.07.053>
- Plummer, M., N. Best, K. Cowles, and K. Vines, 2006 CODA: convergence diagnosis and output analysis for MCMC. *R News* 6: 7–11.
- Poole, A., C. Urbanek, C. Eng, J. Schageman, S. Jacobson *et al.*, 2014 Dissecting childhood asthma with nasal transcriptomics distinguishes subphenotypes of disease. *J. Allergy Clin. Immunol.* 133: 670–678.e12. <https://doi.org/10.1016/j.jaci.2013.11.025>
- Pruim, R. J., R. P. Welch, S. Sanna, T. M. Teslovich, P. S. Chines *et al.*, 2010 LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics* 26: 2336–2337. <https://doi.org/10.1093/bioinformatics/btq419>
- Purcell, S., and C. Chang, 2013 Plink 1.9. Available at: www.cog-genomics.org/plink/1.9/. Accessed March 2019.
- Qian, F., J. Deng, Y. G. Lee, J. Zhu, M. Karpurapu *et al.*, 2015 The transcription factor PU.1 promotes alternative macrophage polarization and asthmatic airway inflammation. *J. Mol. Cell Biol.* 7: 557–567. <https://doi.org/10.1093/jmcb/mjv042>
- Repapi, E., I. Sayers, L. V. Wain, P. R. Burton, T. Johnson *et al.*, 2010 Genome-wide association study identifies five loci associated with lung function. *Nat. Genet.* 42: 36–44. <https://doi.org/10.1038/ng.501>
- Rochman, M., A. V. Kartashov, J. M. Caldwell, M. H. Collins, E. M. Stucke *et al.*, 2015 Neurotrophic tyrosine kinase receptor 1 is a direct transcriptional and epigenetic target of IL-13 involved in allergic inflammation. *Mucosal Immunol.* 8: 785–798. <https://doi.org/10.1038/mi.2014.109>
- Schmitt, A. D., M. Hu, I. Jung, Z. Xu, Y. Qiu *et al.*, 2016 A compendium of chromatin contact maps reveals spatially active regions in the human genome. *Cell Rep.* 17: 2042–2059. <https://doi.org/10.1016/j.celrep.2016.10.061>
- Scott, E. W., M. C. Simon, J. Anastasi, and H. Singh, 1994 Requirement of transcription factor PU.1 in the development of multiple hematopoietic lineages. *Science* 265: 1573–1577. <https://doi.org/10.1126/science.8079170>
- Scott, E. W., R. C. Fisher, M. C. Olson, E. W. Kehrli, M. C. Simon *et al.*, 1997 PU.1 functions in a cell-autonomous manner to control the differentiation of multipotential lymphoid-myeloid progenitors. *Immunity* 6: 437–447. [https://doi.org/10.1016/S1074-7613\(00\)80287-3](https://doi.org/10.1016/S1074-7613(00)80287-3)
- Sillanpaa, E., S. Sipila, T. Tormakangas, J. Kaprio, and T. Rantanen, 2017 Genetic and environmental effects on telomere length and lung function: a twin study. *J. Gerontol. A Biol. Sci. Med. Sci.* 72: 1561–1568.
- Sofer, T., X. Zheng, S. M. Gogarten, C. A. Laurie, K. Grinde *et al.*, 2019 A fully adjusted two-stage procedure for rank-normalization in genetic association studies. *Genet. Epidemiol.* 43: 263–275. <https://doi.org/10.1002/gepi.22188>
- Soler Artigas, M., D. W. Loth, L. V. Wain, S. A. Gharib, M. Obeidat *et al.*, 2011 Genome-wide association and large-scale follow up identifies 16 new loci influencing lung function. *Nat. Genet.* 43: 1082–1090. <https://doi.org/10.1038/ng.941>
- Soler Artigas, M., L. V. Wain, S. Miller, A. K. Kheirallah, J. E. Huffman *et al.*, 2015 Sixteen new lung function signals identified through 1000 genomes project reference panel imputation. *Nat. Commun.* 6: 8658. <https://doi.org/10.1038/ncomms9658>
- Summer Institute in Statistical Genetics, 2019 PC-Relate. Available at: https://uw-gac.github.io/SISG_2019/pc-relate.html. Accessed July 25, 2019.
- Taliun, D., D. N. Harris, M. D. Kessler, J. Carlson, Z. A. Szpiech *et al.*, 2019 Sequencing of 53,831 diverse genomes from the NHLBI TOPMed program. *bioRxiv*. (Preprint posted March 6, 2019). <https://doi.org/doi:10.1101/563866>
- Tayel, S. I., S. M. El-Hefnway, E. M. Abd El Gayed, and G. A. Abdelaal, 2017 Association of stem cell factor gene expression with severity and atopic state in patients with bronchial asthma. *Respir. Res.* 18: 21–22. <https://doi.org/10.1186/s12931-017-0504-2>
- Tian, X., C. Xu, Y. Wu, J. Sun, H. Duan *et al.*, 2017 Genetic and environmental influences on pulmonary function and muscle strength: the Chinese twin study of aging. *Twin Res. Hum. Genet.* 20: 53–59. <https://doi.org/10.1017/thg.2016.97>
- United States Environmental Protection Agency, 2008 National Emissions Inventory (NEI) 2008 Data. Available at: <https://www.epa.gov/air-emissions-inventories/2008-national-emissions-inventory-nei-data>. Accessed January 8, 2020.
- United States Environmental Protection Agency, 2011 National Emissions Inventory (NEI) 2011 Data. Available at: <https://www.epa.gov/air-emissions-inventories/2011-national-emissions-inventory-nei-data>. Accessed January 8, 2020.
- University of Michigan, and National Heart, Lung and Blood Institute Trans-Omics for Precision Medicine (TOPMed) Program, 2018 BRAVO Variant Browser. Available at: <https://bravo.sph.umich.edu/freeze5/hg38/>. Accessed August 1, 2019.
- Usemann, J., F. Decrue, I. Korten, E. Proietti, O. Gorlanova *et al.*, 2019 Exposure to moderate air pollution and associations with lung function at school-age: a birth cohort study. *Environ. Int.* 126: 682–689. <https://doi.org/10.1016/j.envint.2018.12.019>
- Valent, P., E. Spanblochl, W. R. Sperr, C. Sillaber, K. M. Zsebo *et al.*, 1992 Induction of differentiation of human mast cells from bone marrow and peripheral blood mononuclear cells by recombinant human stem cell factor/kit-ligand in long-term culture. *Blood* 80: 2237–2245. <https://doi.org/10.1182/blood.V80.9.2237.2237>
- Wain, L. V., N. Shrine, M. S. Artigas, A. M. Erzurumluoglu, B. Noyvert *et al.*, 2017 Genome-wide association analyses for lung function and chronic obstructive pulmonary disease identify new loci and potential druggable targets. *Nat. Genet.* 49: 416–425. <https://doi.org/10.1038/ng.3787>
- Walsh, J. C., R. P. DeKoter, H. J. Lee, E. D. Smith, D. W. Lancki *et al.*, 2002 Cooperative and antagonistic interplay between PU.1 and GATA-2 in the specification of myeloid cell fates. *Immunity* 17: 665–676. [https://doi.org/10.1016/S1074-7613\(02\)00452-1](https://doi.org/10.1016/S1074-7613(02)00452-1)
- Wan, H., S. Dingle, Y. Xu, V. Besnard, K. H. Kaestner *et al.*, 2005 Compensatory roles of Foxa1 and Foxa2 during lung morphogenesis. *J. Biol. Chem.* 280: 13809–13816. <https://doi.org/10.1074/jbc.M414122200>
- Wen, L. P., J. A. Fahrni, S. Matsui, and G. D. Rosen, 1996 Airway epithelial cells produce stem cell factor. *Biochim. Biophys. Acta* 1314: 183–186. [https://doi.org/10.1016/S0167-4889\(96\)00138-3](https://doi.org/10.1016/S0167-4889(96)00138-3)
- White, M. J., O. Risse-Adams, P. Goddard, M. G. Contreras, J. Adams *et al.*, 2016 Novel genetic risk factors for asthma in African American children: precision medicine and the SAGE II study. *Immunogenetics* 68: 391–400. <https://doi.org/10.1007/s00251-016-0914-1>
- Wise, J., 2019 Air pollution is linked to infant deaths and reduced lung function in children. *BMJ* 366: l5772. <https://doi.org/10.1136/bmj.l5772>
- Wojcik, G. L., M. Graff, K. K. Nishimura, R. Tao, J. Haessler *et al.*, 2019 Genetic analyses of diverse populations improves

- discovery for complex traits. *Nature* 570: 514–518. <https://doi.org/10.1038/s41586-019-1310-4>
- World Health Organization, 2017 Asthma. Available at: <http://www.who.int/mediacentre/factsheets/fs307/en/>. Accessed January 8, 2020.
- Yamada, H., Y. Yatagai, H. Masuko, T. Sakamoto, H. Iijima *et al.*, 2015 Heritability of pulmonary function estimated from genome-wide SNPs in healthy Japanese adults. *Respir. Investig.* 53: 60–67. <https://doi.org/10.1016/j.resinv.2014.10.004>
- Yashiro, T., S. Nakano, K. Nomura, Y. Uchida, K. Kasakura *et al.*, 2019 A transcription factor PU.1 is critical for Ccl22 gene expression in dendritic cells and macrophages. *Sci. Rep.* 9: 1161–1169. <https://doi.org/10.1038/s41598-018-37894-9>
- Zhang, F., and J. R. Lupski, 2015 Non-coding genetic variants in human disease. *Hum. Mol. Genet.* 24: R102–R110. <https://doi.org/10.1093/hmg/ddv259>

Communicating editor: A. Long