TITLE

Novel *KITLG/SCF* regulatory variants are associated with lung function in African American children with asthma

AUTHORS

Angel CY Mak[1,*], Satria Sajuthi[2,^], Jaehyun Joo[3,^], Shujie Xiao[4,^], Patrick M Sleiman[5,6,^], Marquitta J White[1,^], Eunice Y Lee[1], Benjamin Saef[3], Donglei Hu[1], Hongsheng Gui[4], Kevin L Keys[1,7], Fred Lurmann[8], Deepti Jain[9], Gonçalo Abecasis[10], Hyun Min Kang[10], Deborah A. Nickerson[11,12,13], Soren Germer[14], Michael C Zody[14], Lara Winterkorn[14], Catherine Reeves[14], Scott Huntsman[1], Celeste Eng[1], Sandra Salazar[1], Sam S Oh[1], Frank D Gilliland[15], Zhanghua Chen[15], Rajesh Kumar[16], Fernando D Martínez[17], Ann Chen Wu[18], Elad Ziv[1], Hakon Hakonarson[5,6,#], Blanca E Himes[3,#], L Keoki Williams[4,#], Max A Seibold[3,#], Esteban G. Burchard[1,19]

[1] Department of Medicine, University of California San Francisco, San Francisco, CA, USA

[2] Center for Genes, Environment, and Health, National Jewish Health, Denver, CO, USA

[3] Department of Biostatistics, Epidemiology, and Informatics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA

[4] Center for Individualized and Genomic Medicine Research, Department of Internal Medicine, Henry Ford Health System, Detroit, MI, USA

[5] Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, PA, USA

19 [6] Department of Human Genetics, Children's Hospital of Philadelphia, Philadelphia, PA, USA

20 [7] Berkeley Institute for Data Science, University of California, Berkeley, CA, USA

21 [8] Sonoma Technology Inc, Petaluma, CA, USA

22 [9] Department of Biostatistics, University of Washington, Seattle, WA, USA

23 [10] Center for Statistical Genetics, University of Michigan, Ann Arbor, MI, USA

24 [11] Department of Genome Sciences, University of Washington, Seattle, WA, USA

25 [12] Northwest Genomics Center, Seattle, WA, USA

26 [13] Brotman Baty Institute, Seattle, WA, USA

27 [14] New York Genome Center, New York, NY, USA

28 [15] Department of Preventive Medicine, Division of Environmental Health, Keck School of Medicine,

29 University of Southern California, Los Angeles, CA, USA

30 [16] Ann and Robert H. Lurie Children's Hospital of Chicago, Chicago, IL, USA

31 [17] Asthma and Airway Disease Research Center, University of Arizona, Tucson, AZ, USA

32 [18] Precision Medicine Translational Research (PRoMoTeR) Center, Department of Population

33 Medicine, Harvard Medical School and Pilgrim Health Care Institute, Boston, MA, USA

34 [19] Department of Bioengineering and Therapeutic Sciences, University of California, San Francisco,

35 CA, USA

36 ^, # These authors contributed equally to this work

37 * Corresponding author

38 Angel CY Mak

39 Address: Box 2911, 1550 4th Street, San Francisco, CA 94112

40 Tel: 415-514-9931

41 Email: angelcymak@gmail.com

42 AUTHOR CONTRIBUTIONS

43 ACYM, SSaj, MJW, SSO, FG, ZC, RK, FDM, ACW, EZ, and EGB designed and supervised the study.
44 ACYM, SSaj, JJ, SX, PS, MJW, EYL, BS, DH, SH, BEH, HH, LKW, and MAS supervised/performed
45 analyses and/or interpreted results. CE processed biospecimens for whole genome sequencing.
46 FL estimated air pollution exposure. DJ, GA, HMK, DAN, and SG contributed to TOPMed WGS data
47 generation and/or analysis. MZ, LW, and CR contributed to CCDG WGS data generation. SSal
48 coordinated study recruitment and processed phenotype data ACYM, SSaj, JJ, SX, PS, MJW, EYL,
49 BS, HG, KLK, SG, BEH, HH, LKW, MAS, and EGB wrote and/or critically reviewed the manuscript.

50 SHORT TITLE

51 *KITLG* associated with FEV_1 in AA youths

52 KEYWORDS

53 GWAS, African American, FEV_1, gene-by-environment interaction, GxE, air pollution, KITLG, SCF

ABSTRACT

Baseline lung function, quantified as forced expiratory volume in the first second of exhalation (FEV_1), 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 TOPMed project, we identified a novel genetic association with FEV_1 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 9 variants as the most likely candidates responsible for the association with FEV_1. Hi-C data and eQTL analysis demonstrated that these variants physically interacted with *KITLG (aka SCF)* and their minor alleles were associated with increased expression of *KITLG* gene in nasal epithelial cells. Gene-by-air-pollution interaction analysis found that the candidate variant rs58475486 interacted with past-year SO_2 exposure ($p = 0.003$, $\beta = 0.32$). This study identified a novel protective genetic association with FEV_1, possibly mediated through *KITLG*, in African American children with asthma.

INTRODUCTION

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).

74 Specifically, African American children carry a higher asthma disease burden compared to their
75 European American counterparts (Akinbami *et al.* 2014; Akinbami 2015). Forced expiratory
76 volume in the first second (FEV_1), a measurement of lung function, is a vital clinical trait used by
77 physicians to assess overall lung health and diagnose pulmonary diseases such as asthma
78 (Johnson and Theurer. 2014). We have previously shown that genetic ancestry plays an important
79 role in FEV_1 variation and that African Americans have lower FEV_1 compared to European
80 Americans regardless of asthma status (Kumar *et al.* 2010; Pino-Yanes *et al.* 2015). The disparity
81 in lung function between populations may explain disparities in asthma disease burden.
82 Understanding the factors that influence FEV_1 variation among individuals with asthma could
83 lead to improved patient care and therapeutic interventions.

84 Twin and family-based studies estimate that the heritability of FEV_1 ranged from 26% to 81%,
85 supporting the combined contribution by genetic and environmental factors in FEV_1 variation
86 (Chatterjee and Das. 1995; Chen *et al.* 1996; Hukkinen *et al.* 2011; Palmer *et al.* 2001; Sillanpaa
87 *et al.* 2017; Tian *et al.* 2017; Yamada *et al.* 2015). Genome-wide association studies (GWAS) of
88 FEV_1, including among individuals with asthma, have identified many variants that contribute to
89 lung function (Li *et al.* 2013; Liao *et al.* 2014; Repapi *et al.* 2010; Soler Artigas *et al.* 2011; Soler
90 Artigas *et al.* 2015; Wain *et al.* 2017). A search in NHGRI-EBI GWAS Catalog (version e98_r2020-
91 03-08) on baseline lung function (FEV_1) alone revealed 349 associations (Buniello *et al.* 2019).
92 Most of these previous GWAS, however, were performed in adult populations of European
93 descent, and their results may not generalize across populations or across the life span of an
94 individual (Carlson *et al.* 2013; Martin, A. R. *et al.* 2017; Wojcik *et al.* 2019). Previous GWAS results

are also limited due to their reliance on genotyping arrays. In particular, variation in non-coding 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 (Kim, M. S. *et al.* 2018; Zhang and Lupski. 2015). Whole genome sequencing (WGS) is a newer technology that captures nearly all common variation from coding and non-coding regions of the genome and is unencumbered by genotype array design constraints and differences in linkage disequilibrium patterns among populations. To date, no large-scale WGS studies of lung function have been performed in African American children with asthma (Martin *et al.* 2017).

In addition to genetics, FEV_1 is a complex trait that is significantly influenced by both genetic variation and environmental factors, such as air pollution (Chatterjee and Das. 1995; Hukkinen *et al.* 2011; Palmer *et al.* 2001; Sillanpaa *et al.* 2017; Tian *et al.* 2017; Yamada *et al.* 2015). Exposure to ambient air pollution has been consistently associated with poor respiratory outcomes, including reduced FEV_1 (Barraza-Villarreal *et al.* 2008; Brunekreef and Holgate. 2002; Ierodiakonou *et al.* 2016; Wise 2019). We previously showed that exposure to sulfur dioxide (SO_2), an air pollutant emitted by the burning of fossil fuels, is significantly associated with reduced FEV_1 in African American children with asthma in the SAGE II study (Neophytou *et al.* 2016). Because the genetic variants associated with FEV_1 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

115 common genetic variants associated with FEV_1 in African American children with asthma in SAGE
116 II and investigated the effect of GxE (SO_2) interactions on FEV_1 associations.

117 METHODS

118 *Study population*

119 This study examined African American children between 8-21 years of age with physician-
120 diagnosed asthma from the Study of African Americans, Asthma, Genes & Environments (SAGE
121 II). All SAGE II participants were recruited from the San Francisco Bay Area. The inclusion and
122 exclusion are previously described in detailed (Oh *et al.* 2012; White *et al.* 2016) . Briefly,
123 participants were eligible if they were 8-21 years of age and self-identified as African American
124 and had four African American grandparents. Study exclusion criteria included the following: 1)
125 any smoking within one year of the recruitment date; 2) 10 or more pack-years of smoking; 3)
126 pregnancy in the third trimester; 4) history of lung diseases other than asthma (for cases) or
127 chronic illness (for cases and controls). Baseline lung function defined as forced expiratory
128 volume in the first second (FEV_1) was measured by spirometry prior to administering albuterol as
129 previously described (Oh *et al.* 2012).

130 *TOPMed whole genome sequencing data*

131 SAGE II DNA samples were sequenced as part of the Trans-Omics for Precision Medicine (TOPMed)
132 whole genome sequencing (WGS) program (Taliun *et al.* 2019). WGS was performed at the New
133 York Genome Center and Northwest Genomics Center on a HiSeq X system (Illumina, San Diego,
134 CA) using a paired-end read length of 150 base pairs (bp), with a minimum of 30x mean genome

135 coverage. DNA sample handling, quality control, library construction, clustering and sequencing,
136 read processing and sequence data quality control are described in detail in the TOPMed website
137 (TOPMed 2019). Variant calls were obtained from TOPMed data freeze 8 VCF files corresponding
138 to the GRCh38 assembly. Variants with a minimal read depth of 10 (DP10) were used for analysis
139 unless otherwise stated.

140 *Genetic principal components, global ancestry, and kinship estimation*

141 Genetic principal components (PCs), global ancestry, and kinship estimation on genetic
142 relatedness were computed using biallelic single nucleotide polymorphisms (SNPs) with a PASS
143 flag from TOPMed freeze 8 DP10 data. PCs and kinship estimates were computed using the PC-
144 Relate function from the GENESIS R package (Conomos *et al.* 2015; Conomos *et al.* 2016) using a
145 workflow available from the Summer Institute in Statistical Genetics Module 17 course website
146 (Summer Institute in Statistical Genetics 2019). African global ancestry was computed using the
147 ADMIXTURE package (Alexander *et al.* 2009) in supervised mode using European (CEU), African
148 (YRI) and Native American (NAM) reference panels as previously described (Mak, A. C. Y. *et al.*
149 2018).

150 *FEV_1 GWAS*

151 Non-normality of the distribution of FEV_1 values was tested with the Shapiro-Wilk test in R using
152 the shapiro.test function. Since FEV_1 was not normally distributed ($p = 1.41 \times 10^{-8}$ for FEV_1 and p
153 $= 1.05 \times 10^{-8}$ for $\log_{10} FEV_1$), FEV_1 was regressed on all covariates (age, sex, height, controller
154 medications, sequencing centers, and the first 5 genetic PCs) and the residuals were inverse-

155 normalized. These inverse-normalized residuals (FEV_1.res.rnorm) were the main outcome of the
156 discovery GWAS. The controller medication covariate included the use of inhaled corticosteroids
157 (ICS), long-acting beta-agonists (LABA), leukotriene inhibitors and/or an ICS/LABA combo in the
158 2 weeks prior to the recruitment date.

159 Genome-wide single variant analysis was performed on the ENCORE server (https://github.com
160 /statgen/encore) using the linear Wald test (q.linear) originally implemented in EPACTS
161 (https://genome.sph.umich.edu /wiki/EPACTS) and TOPMed freeze 8 data (DP0 PASS) with a
162 MAF filter of 0.1%. All pairwise relationships with degree 3 or more relatedness (kinship values >
163 0.044) were identified, and one participant of the related pair was subsequently chosen at
164 random and removed prior to analysis. All covariates used to obtain FEV_1.res.rnorm were also
165 included as covariates in the GWAS as recommended in a recent publication (Sofer *et al.* 2019).
166 The association analysis was repeated using untransformed FEV_1 and FEV_1 percent predicted
167 (FEV_1.perc.predicted). FEV_1 percent predicted was defined as the percentage of measured FEV_1
168 relative to predicted FEV_1 estimated by the Hankinson lung function prediction equation for
169 African Americans (Hankinson *et al.* 1999). A secondary analysis that included smoking-related
170 covariates (smoking status and number of smokers in the family) was performed in PLINK 1.9
171 (version 1p9_2019_0304_dev) (Chang *et al.* 2015; Purcell and Chang. 2013). To study whether
172 association with FEV_1 is specific to SAGE II participants with asthma, we repeated the association
173 analysis adjusting for age, sex, height and the first 5 genetic PCs in SAGE II participants without
174 asthma on the ENCORE server. All of these participants were sequenced in the same center.
175 Regional association results were plotted using LocusZoom 1.4 (Pruim *et al.* 2010) with a 500

176 kilobase (Kb) flanking region. Linkage disequilibrium (R^2) was estimated in PLINK 1.9. LD plot was
177 generated using recoded genotype files (plink --recode 12) in Haploview (Barrett *et al.* 2005) .

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

184 The following WGS quality control steps were applied to all reported variants from ENCORE to
185 ensure WGS variant quality: (1) The variant had VCF FILTER = PASS; (2) Variant quality was
186 confirmed via manual inspection on the BRAVO server based on TOPMed freeze 5 data
187 (University of Michigan and NHLBI TOPMed. 2018); (3) Variants were reanalyzed with linear
188 regression using PLINK 1.9 by applying the arguments --mac 5 --geno 0.1 --hwe 0.0001 using
189 TOPMed freeze 8 DP10 PASS data.

190 To determine if the rs73429450 association with FEV_1 was only identifiable using whole genome
191 sequencing data, we repeated the linear regression association analysis on signals that passed
192 the genome-wide significance threshold using PLINK 1.9 and genotype data generated with
193 Axiom Genome-Wide LAT 1 array (Affymetrix, Santa Clara, CA, dbGaP phs000921.v1.p1). These
194 array genotype data were imputed into the following reference panels: 1000 Genomes phase 3
195 version 5, Haplotype Reference Consortium (HRC) r1.1, the Consortium on Asthma among

196 African-ancestry Populations in the Americas (CAAPA) and the TOPMed phase 5 panels on the
197 Michigan Imputation Server (Das *et al.* 2016). It should be noted that 500 SAGE II subjects were
198 part of the TOPMed freeze 5 reference panel.

199 A total of 349 GWAS FEV_1-associated entries were retrieved from the NHGRI-EBI GWAS Catalog
200 version 1.0.2-associations_e98_r2020-03-08 (Buniello *et al.* 2019) using the trait names "Lung
201 function (FEV_1)", "FEV1", "Lung function (forced expiratory volume in 1 second)" or
202 "Prebronchodilator FEV1". After adding 100 Kb flanking regions to each of the 349 entries, a total
203 of 230 non-overlapping region were obtained. To look up whether we replicated previously
204 GWAS loci while control for multiple testing penalties, we only used 279,495 common variants
205 (MAF >= 0.01) that overlapped with the 230 regions. The 279,495 common variants is equivalent
206 to 17,755 effective test based on CODA and 5.63×10^{-5} (1/17,755) was used as suggestive p-value
207 threshold for replication.

208 *Conditional analysis*

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

216 *Region-based association analysis*

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

237 *Functional annotations and prioritization of genetic variants*

238 The Hi-C Unifying Genomic Interrogator (HUGIN) (Ay *et al.* 2014; Martin, J. S. *et al.* 2017; Schmitt
239 *et al.* 2016) was used to assign potential gene targets to each variant. HUGIN uses the Hi-C data
240 generated from the primary human tissues from four donors used in the Roadmap Epigenomics
241 Project (Schmitt *et al.* 2016). ENCODE annotations (ENCODE Project Consortium 2011; ENCODE
242 Project Consortium 2012) were based on overlap of the variants with functional data downloaded
243 from the UCSC Table Browser (Karolchik *et al.* 2004). These data included DNAase I
244 hypersensitivity peak clusters (hg38 wgEncodeRegDnaseClustered table), transcription factor
245 ChIP-Seq clusters (hg38 encRegTfbsClustered table) and histone modification ChIP-Seq peaks
246 (hg19 wgEncodeBroadHistone<cell type><histone>StdPk tables). For DNase I hypersensitivity
247 and transcription factor binding sites, we focused on blood, bone marrow, lung and embryonic
248 cells. For histone modification ChIP-Seq, we focused on H3K27ac and H3K4me3 modifications in
249 human blood (GM12878), bone marrow (K562), lung fibroblast (NHLF), and embryonic stem cells
250 (H1-hESC). LiftOver tool (Hinrichs *et al.* 2006) was used to convert genomic coordinates from
251 hg19 to hg38. Candidate cis-regulatory elements (ccREs) were a subset of representative DNase
252 hypersensitivity sites with epigenetic activity further supported by histone modification
253 (H3K4me3 and H3K27ac) or CTCF-binding data from the ENCODE project. Overlap of variants with
254 ccREs were detected using the Search Candidate cis-Regulatory Elements by ENCODE (SCREEN)
255 web interface (ENCODE Project Consortium 2011; ENCODE Project Consortium 2012).

256 Prioritization of genetic variants was based on the presence of statistical, functional and/or
257 bioinformatic evidence as described in the Diverse Convergent Evidence (DiCE) prioritization

258 framework (Ciesielski *et al.* 2014). The priority score of each variant was obtained by counting
259 the number of statistical, functional, and/or bioinformatic evidences that support potential
260 biological function for that variant.

261 *Replication of GWAS associations*

262 All replication analyses were performed in subjects with asthma. Replication of GWAS FEV_1
263 associations was attempted on TOPMed whole genome sequencing data generated from four
264 cohorts. These cohorts included Puerto Rican (n=1,109) and Mexican American (n=649) children
265 in the Genes-Environments and Admixture in Latino Americans (GALA II) study (Oh *et al.* 2012),
266 African American adults in the Study of Asthma Phenotypes and Pharmacogenomic Interactions
267 by Race-Ethnicity (SAPPHIRE, n=3,428) (Levin *et al.* 2014) and African American children in
268 Genetics of Complex Pediatric Disorders (GCPD-A, n=1,464) study (Ong *et al.* 2013). Age, sex,
269 height, controller medications and the first 5 PCs were used as covariates.

270 Additionally, replication of GWAS FEV_1 associations was attempted using data of black UK
271 Biobank subjects who had asthma (n=627) while adjusting for age, sex, height and the first 5
272 principal components. Asthma status was defined by ICD code or self-reported asthma. UK
273 Biobank genotype data was generated on Affymetrix UK BiLEVE axiom or UK Biobank Axiom array
274 and imputed into the Haplotype Reference Consortium, 1000G and UK 10K projects (Bycroft *et*
275 *al.* 2018; Canela-Xandri *et al.* 2018). Additional details on the UK Biobank study and the
276 replication procedures are available in File S1 (Supplementary Text 2).

277 *RNA sequencing and expression quantitative trait loci (eQTL) analysis*

278 Whole-transcriptome libraries of 370 nasal brushings from GALA II Puerto Rican children with
279 asthma were constructed by using the Beckman Coulter FX automation system (Beckman Coulter,
280 Fullerton, CA). Libraries were sequenced with the Illumina HiSeq 2500 system. Raw RNA-Seq
281 reads were trimmed using Skewer (Jiang *et al.* 2014) and mapped to human reference genome
282 hg38 using Hisat2 (Kim, D. *et al.* 2015). Reads mapped to genes were counted with htseq-count
283 and using the UCSC hg38 GTF file as reference (Anders *et al.* 2015). Cis-expression quantitative
284 trait locus (eQTL) analysis of *KITLG* was performed as described in the Genotype-Tissue
285 Expression (GTEx) project version 7 protocol (GTEx Consortium *et al.* 2017) using age, sex, BMI,
286 global African and European ancestries and 60 PEER factors as covariates.

287 *Gene-by-air-pollution interaction analysis*

288 We hypothesized that the effect of genetic variation on lung function in our study population
289 may differ by the levels of exposure to SO_2 (Neophytou *et al.* 2016). To test for an interaction
290 between a genetic variant and SO_2, an additional multiplicative interaction term (variant x $S0_2$
291 exposure) was included in the original GWAS model (see Method Section "FEV_1 GWAS"). The SO_2
292 estimates used in the interaction analysis were first-year, past-year, and lifetime exposure to
293 ambient of SO_2, which were estimated as described previously (Neophytou *et al.* 2016). Briefly,
294 we obtained regional ambient daily air pollution data from the U.S. Environmental Protection
295 Agency Air Quality System. SO_2 estimates for the participant's residential geographic coordinate
296 were calculated as the inverse distance-squared weighted average from the four closest air
297 pollution monitoring stations within 50 km of the participant's residence. We estimated yearly

298 exposure at the reported residential address by averaging all available daily measures (daily
299 average of 1-hour SO_2) in a given year. If the participant had a change of residential address in a
300 given year, we estimate yearly exposure as a time-weighted estimate based on the number of
301 months spent at each different address in that year. Average lifetime exposures were estimated
302 using all available yearly average estimates over the lifetime of the participant until the day of
303 spirometry testing. Since not all pollutants were measured daily, there are location- and
304 pollutant-dependent missing values. Residuals of FEV_1 were plotted against exposure to SO_2 and
305 stratified by the number of copies of the minor allele. Residuals of FEV_1 were obtained as
306 described in the Methods Section "FEV_1 GWAS".

307 *Data availability*

308 Local institutional review boards approved the studies (IRB# 10-02877). All subjects and legal
309 guardians provided written informed consent. TOPMed whole genome sequencing and
310 phenotype data from SAGE II are available on dbGaP under accession number phs000921.v4.p1.
311 Normalized gene count data for *KITLG* and supplemental materials are available at figshare.

312 RESULTS

313 *Novel lung function associations*

314 Subject characteristics of the 867 African American children with asthma included in this study
315 are shown in Table 1, and the distribution of their FEV_1 measurements (mean = 2.56 L, standard
316 deviation = 0.79 L) is in Figure S1. The CODA-adjusted statistical significance thresholds $2.10 \times 10^{-}$

317 ^{8} and 4.19 x 10^{-7} were used as the genome-wide and suggestive significance thresholds,
318 respectively. According to this threshold, one SNP in chromosome 12 (chr12:88846435,
319 rs73429450, G>A) was associated with FEV_1.res.rnorm (Figure 1, p = 9.01 x 10^{-9}, β = 0.801) at
320 genome-wide significance. The association between rs73429450 and lung function remained
321 statistically significant when the association was repeated using untransformed FEV_1 (p = 1.26 x
322 10^{-8}, β = 0.302) as the outcome variable. The association between rs73429450 and lung function
323 was suggestive using FEV_1.perc.predicted (p = 1.69 x 10^{-7}, β = 0.100). Twenty suggestive
324 associations corresponding to 4 tag signals are reported in Supplementary File S2. None of the
325 suggestive associations overlapped with any of the previously reported FEV_1-associated loci.
326 When considering only common variants and applying a p-value threshold of 5.63 x 10^{-5}, we
327 found replicated in 6 out of 230 previously reported FEV_1 associations (Table S1). Our top FEV_1
328 association, rs73429450, did not overlap with any previously reported loci and it is a novel
329 association with FEV_1 in this study population.

330 Secondary analysis that included covariates correcting for smoking status and number of smokers
331 in the family showed that smoking-related factors were not significantly associated with FEV_1 in
332 our pediatric SAGE cohort: using 657 out of 867 individuals with available smoking-related
333 covariates, the FEV_1.res.rnorm association p-values before and after including the smoking-
334 related covariates were 2.01 x 10^{-6} and 1.89 x 10^{-6}. Both p-values of the covariates smoking status
335 (p = 0.27) and number of smokers in the family (p = 0.54) were not significant.

336 Conditional analysis was performed on 45 variants with association p < 1 x 10^{-4} located within 1
337 Mb of the strongest association signal (rs73429450). Two weaker independent signals

338 (rs17016065, rs58475486) were identified (Table S2). None of the 45 variants showed association
339 with FEV_1.res.rnom in 251 SAGE II children without asthma (Table S3).

340 The minor allele frequency of rs73429450 in continental populations from the 1000 Genomes
341 Project (1000G) is 3% in Africans (AFR) and < 1% in Admixed Americans (AMR), Europeans (EUR)
342 and Asians (EAS and SAS) (1000 Genomes Project Consortium *et al.* 2015). Rs73429450 was not
343 included on the Affymetrix LAT1 genotyping array where SAGE participants were previously
344 genotyped. To determine if the rs73429450 association with FEV_1 was only identifiable using
345 whole genome sequencing data, we attempted to reproduce our results by imputing the
346 genotype of rs73429450 in 851 SAGE participants with available array data using 1000G phase 3
347 (n = 2,504), HRC r1.1 (n = 32,470), CAAPA (n = 883) and TOPMed freeze 5 (n = 62,784) reference
348 panels. Our results remained statistically significant when using the 1000G phase 3 ($p = 4.97 \times 10^{-}$
349 ^{8}, $\beta = 0.79$, imputation $R^2 = 0.95$) and TOPMed freeze 5 ($p = 1.22 \times 10^{-8}$, $\beta = 0.80$, imputation $R^2 =$
350 0.98) reference panels, but lost statistical significance when rs73429450 genotypes were
351 imputed using the HRC ($p = 4.35 \times 10^{-7}$, $\beta = 0.68$, imputation $R^2 = 0.94$) and CAAPA ($p = 1.95 \times 10^{-}$
352 ^{7}, $\beta = 0.80$, imputation $R^2 = 0.71$) reference panels.

353 Region-based association analysis including all variants conditioned on the association signal
354 from rs73429450 was performed in its 1 Mb flanking region (chr12:87846435-89846435). No
355 windows were significantly associated after Bonferroni multiple testing correction ($p < 2.80 \times 10^{-}$
356 ^{4}, Figure S2), but 20 windows were suggestively associated with FEV_1.res.rnorm ($p < 5.60 \times 10^{-3}$,
357 Table S5). Two of 20 windows re-tested using only functional variants were suggestively
358 significant (region 4 and 16). Both of these windows were no longer suggestively significant after

359 removing the common variants, indicating that association signal from these regions was mostly
360 driven by common variants. Further investigation on region 16 using drop-one analysis on the 2
361 rare and 1 common function variants confirmed the major contribution by the common variant,
362 rs1895710, as shown by the major increase in p-value (Table S6). The signal was also slightly
363 driven by the singleton, rs990979778. Drop-one analysis was not performed on region 4 because
364 there were only 1 common and 1 rare variants.

365 A Hi-C assay couples a chromosome conformation capture (3C) assay with next-generation
366 sequencing to capture long-range interactions in the genome. We identified a statistically
367 significant long-range chromatin interaction between the GWAS peak and the KIT ligand (*KITLG*,
368 also known as stem cell factor, *SCF*) gene in human fetal lung fibroblast cell line IMR90 (Table S7).
369 The long-range interaction detected in human primary lung tissue was not significant, implying
370 that the potential long-range interactions are specific to tissue type or developmental stage.

371 *Potential regulatory role of FEV_1-associated variants on KITLG expression*

372 To further elucidate potential regulatory relationships between the GWAS association peak and
373 *KITLG*, we analyzed whether variants in the peak were eQTL of *KITLG* in previously published
374 whole blood RNA-Seq data available from the same study participants (Mak, Angel CY *et al.* 2016) .
375 The whole blood RNA-Seq data, however, did not yield evidence of expressed *KITLG*, consistent
376 with results in GTEx. We subsequently used RNA-Seq data from nasal epithelial cells of 370 Puerto
377 Rican children with asthma from the GALA II study, and found that five out of 45 variants were
378 eQTL of *KITLG* (Table S8). While Puerto Ricans are a different population than African Americans,

379 they are both admixed populations with substantial African genetic ancestry, and therefore could
380 share eQTLs. All five eQTLs corresponded to one signal in a region with strong linkage
381 disequilibrium ($r^2 > 0.8$, Figure S3).

382 *Replication of genetic association with FEV_1*

383 Subject characteristics of our four replication cohorts (SAPPHIRE, GCPD-A, UK Biobank and GALA
384 II) are shown in Table S9. We attempted to replicate the association of the 45 SNPs in our primary
385 FEV_1 GWAS in each cohort. We used 0.05 as the suggestive p-value threshold and 0.0167 as the
386 Bonferroni-corrected p-value threshold after correcting for 3 independent signals (see
387 conditional analysis in Results Section). A total of 20 variants were replicated at $p < 0.05$ with
388 consistent direction of effect in black UK Biobank participants; 14 variants in SAPPHIRE and 2
389 variants in GCPD-A were significant but had an opposite direction of effect (Table S10).
390 We attempted to replicate the FEV_1.res.rnorm association in Mexican American (n = 649) and
391 Puerto Rican (n = 1,109) children with asthma from the GALA II study. In Mexican Americans, we
392 excluded 19 variants with MAF < 0.1% and associations for the remaining 26 variants did not
393 replicate (Table S11). In Puerto Ricans, the associations were not replicated (Table S11).

394 *Incorporating statistical and functional evidence for candidate variant prioritization*

395 We combined and summarized all functional evidence for the top 45 variants, along with eQTL
396 findings from nasal epithelial RNA-Seq and replication results (Figure 2, Table 2 and S12). To
397 facilitate interpretation of the variant association with FEV_1, the effect sizes and p-values of both
398 FEV_1 (β and p) and FEV_1.res.rnorm (β_{norm} and p_{norm}) associations are also reported. CADD
399 functional prediction score and ENCODE histone modification ChIP-Seq peaks in embryonic,

400 blood, bone marrow, and lung-related tissues were also examined but not reported because
401 none of the variants had a CADD score greater than 10 and none overlapped with histone
402 modification sites. Rs73440122 received the highest priority score of 3 based on replication in
403 the UK Biobank, overlap with a DNase I hypersensitivity site in B-lymphoblastoid cells (GM12865)
404 and overlap with an SPI1 binding site in acute promyelocytic leukemia cells. Eight other variants
405 were prioritized with score > 2 or evidence of being an eQTL for *KITLG* in nasal epithelial cells
406 (Table 2, score marked with ^ or # respectively). These nine candidate variants were selected for
407 gene-by-air-pollution interaction analyses.

408 *Gene-by-air-pollution interaction of rs58475486*

409 We previously found that first year of life and lifetime exposure to SO_2 were associated with FEV_1
410 in African American children (Neophytou *et al.* 2016). We investigated whether the effect of the
411 nine prioritized genetic variants associated with lung function varied by SO_2 exposure (first year
412 of life, past year, and lifetime exposure). Since the nine variants represent three independent
413 signals (see conditional analysis in the Results Section), the Bonferroni-corrected p-value
414 threshold was set to p = 0.0056 (correction for nine tests; three signals and three exposure
415 periods to SO_2). We observed a single statistically significant interaction between the T allele of
416 rs58475486 and past year exposure to SO_2 that was positively associated with FEV_1 (p = 0.003, β
417 = 0.32, Table 3, Figure 3A). This interaction remains significant (p = 0.003, β = 0.32) in secondary
418 analyses adjusted for smoking status or a multiplicative interaction term of rs58475486 and
419 smoking status as additional covariates. Interestingly, six of the remaining eight variants also
420 displayed interaction effects with past year exposure to SO_2 that were suggestively associated (p

421 < 0.05) with FEV_1 (Table 3). We also found a suggestive interaction of the C allele of rs73440122
422 with first year exposure to SO_2 that was associated with decreased FEV_1 ($p = 0.045$, $\beta = -0.32$,
423 Figure 3B). The same allele also showed interaction with past year of exposure to SO_2 that was
424 suggestively associated with FEV_1 in the opposite direction ($p = 0.051$, $\beta = 0.39$).

425 DISCUSSION

426 Variant rs73429450 (MAF = 0.030) was identified as the strongest association signal with FEV_1.
427 Each additional copy of the protective A allele of rs73429450 was associated with a 0.3 L increase
428 of FEV_1. We did not find any statistically significant contribution of rare variants to the association
429 signal from a 1 Kb sliding window analyses in the 1 MB flanking region centered on rs73429450.
430 We were surprised to identify a novel common variant (MAF = 0.030) associated with lung
431 function using whole genome sequence data in a population that was previously analyzed for
432 associations with lung function using genotype array data. Further investigation revealed that
433 our discovered variant, rs73429450, was not captured by the LAT 1 genotyping array, and the
434 association with lung function depended on the reference panel used to impute the variant into
435 our population. More surprisingly, our statistically significant finding was only found to be
436 suggestively significant using data imputed from the CAAPA reference panel ($p = 1.95 \times 10^{-7}$, $\beta =$
437 0.80). Of the imputation reference panels that we assessed, CAAPA is one of the more relevant
438 reference panels for our study population because it is based on African populations in the
439 Americas. However, we note that the effect size estimated from CAAPA-imputed data was
440 comparable to that generated from WGS data. While whole genome sequencing data is usually
441 praised for enabling analysis of rare-variant contributions to phenotype variability, our results

442 show the utility of whole genome sequencing data for the reliable analysis of common variants
443 as well in the absence of relevant imputation panels.

444 Although rs73429450 had the lowest p-value from our whole genome sequencing association
445 analysis, we did not find the required amount of functional evidence to prioritize this marker for
446 inclusion in downstream gene-by-air-pollution analyses. Another variant, rs73440122, was in
447 moderate to strong linkage disequilibrium (r^2 = 0.76) with rs7349450 and had a similar MAF
448 (0.027) in our study population, but was only suggestively associated with FEV_1 in our association
449 analysis ($p = 2.08 \times 10^{-7}$, Table2). In contrast to rs73429450, there were multiple lines of evidence
450 suggesting the functional relevance of rs73440122: rs73440122 received the highest priority
451 score based on its replicated FEV_1 association in black UK Biobank participants and overlap with
452 ENCODE gene regulatory regions, making it one of the most likely drivers of FEV_1 variability
453 among individuals, possibly mediated through *KITLG*.

454 Bioinformatic interrogation of rs73440122 revealed that the variant overlapped with a ccRE
455 (SCREEN accession EH37E0279310), DNase I hypersensitivity site, and SPI1 ChIP-Seq clusters that
456 were indicative of a candidate open chromatin gene regulatory region (Table S12). The binding
457 evidence of SPI1 is highly relevant to the role of KITLG in type 2 inflammation (see below). Variant
458 rs73440122 is located in a region that physically interacted with *KITLG* based on Hi-C data in fetal
459 lung fibroblast cells. Additionally, five neighboring FEV_1 associated variants were identified as
460 eQTLs of *KITLG*, although they appeared to be an independent signal ($r^2 < 0.2$). Overall, these
461 results support regulatory interactions between our novel locus and *KITLG*.

462 Atopic or type 2 high asthma is the most common form of asthma in children (Comberiati *et al.*
463 2017). *KITLG,* more commonly known as stem cell factor (*SCF*), is a ligand of the KIT tyrosine
464 kinase receptor. It plays an important role in type 2 inflammation in atopic asthma, especially in
465 inflammatory processes mediated through mast cells, IgE and group 2 innate lymphoid cells (Da
466 Silva and Frossard. 2005; Da Silva *et al.* 2006; Fonseca *et al.* 2019; Oliveira and Lukacs. 2003) . In
467 the airways, *KITLG* is expressed in bronchial epithelial cells, lung fibroblasts, bronchial smooth
468 muscle cells, endothelial cells, peripheral blood eosinophils, dendritic cells and mast cells (Hsieh
469 *et al.* 2005; Kassel *et al.* 1999; Oriss *et al.* 2014; Valent *et al.* 1992; Wen *et al.* 1996). *KITLG* is a
470 major growth factor of mast cells (Reviewed in Broudy 1997; Da Silva *et al.* 2006; Galli *et al.* 1994;
471 Galli *et al.* 1995). It promotes recruitment of mast cell progenitors into tissues (Reviewed in
472 Oliveira and Lukacs. 2003), prevents mast cell apoptosis (Iemura *et al.* 1994; Mekori *et al.* 1993)
473 and promotes release of inflammatory mediators such as proteases, histamine, chemotactic
474 factors, cytokines (Reviewed in Amin 2012; Borish and Joseph. 1992). While KITLG promotes the
475 production of cytokines like IL-13 upon IgE-receptor crosslinking on the surface of mast cells
476 (Kobayashi *et al.* 1998), IL-13 was also reported to up-regulate *KITLG* (Rochman *et al.* 2015).
477 Consistent with the critical role of KITLG for mast cells and type 2 inflammation, we found our
478 prioritized variant, rs73440122, overlapped with a SPI1 (aka PU.1) ChIP-Seq cluster. The
479 transcription factor SPI1 was demonstrated in SPI1 knockout mice to be necessary for the
480 development of B cells, T cells, neutrophils, macrophages, dendritic cells, and mast cells
481 (Anderson *et al.* 2000; Guerriero *et al.* 2000; McKercher *et al.* 1996; Scott *et al.* 1994; Scott *et al.*
482 1997; Walsh *et al.* 2002). It plays an essential role in macrophage differentiation in asthmatic and

483 other allergic inflammation (Qian *et al.* 2015; Yashiro *et al.* 2019). It was also shown to regulate
484 the cell fate between mast cells and monocytes (Ito *et al.* 2005; Ito *et al.* 2009; Nishiyama,
485 Nishiyama, Ito, Masaki, Maeda *et al.* 2004; Nishiyama, Nishiyama, Ito, Masaki, Masuoka *et al.*
486 2004). The presence of a SPI1 binding site in a candidate regulatory region of KITLG is therefore
487 highly relevant given the critical role of KITLG in mast cell survival and activation.

488 Higher levels of KITLG (Al-Muhsen *et al.* 2004; Da Silva *et al.* 2006; Tayel *et al.* 2017) and an
489 increased number of mast cells in the lung (Cruse and Bradding. 2016; Fajt and Wenzel. 2013;
490 Mendez-Enriquez and Hallgren. 2019) were detected in individuals with asthma. The percentage
491 of a subpopulation of circulating blood mast cell progenitors (Lin^{+} $CD34^{hi}$ $CD117^{int/hi}$ $Fc\varepsilon RI^{+}$) was
492 higher in individuals with a reduced lung function (Dahlin *et al.* 2016). These findings suggested
493 that higher *KITLG* expression and/or number of mast cells may be a contributing factor to lower
494 lung function. This notion was inconsistent with the association of our novel locus with higher
495 *KITLG* expression and increased lung function in SAGE II children with asthma. Interestingly, a
496 study of 20 subjects with severe asthma found that increased in the number of chymase-positive
497 mast cells in the small airway was associated with increased in lung function (Balzar *et al.* 2005).
498 Overall, while there is still controversy on the direction of effect, previous findings support the
499 association of our novel *KITLG* locus with lung function, especially in patients with allergic asthma.
500 Our novel locus likely represents part of a complex regulatory mechanism that modulates
501 immune cell differentiation, survival, and activation in highly cell-specific and context-dependent
502 manners. Further studies are required to study how this locus is regulated in different airway and
503 immune cells to affect lung function outcome in the context of asthma.

504 GxE interactions likely account for a portion of the "missing" heritability of many complex
505 phenotypes (Moore and Williams. 2009). We previously found that lung function in SAGE II
506 participants was associated with first year of life and lifetime exposures to SO_2 (1.66% decrease
507 [95% CI = -2.92 to -0.37] for first year of life and 5.30% decrease [95% CI = -8.43 to -2.06] for
508 lifetime exposures in FEV_1 per 1 ppb increases in SO_2) (Neophytou *et al.* 2016). We hypothesized
509 that a significant portion of the heritability of lung function was due in part to gene-by-air-
510 pollution (SO_2) interaction effects. The interaction between rs58475486 and past year exposure
511 to SO_2 that was significantly associated with lung function supports our hypothesis. The T allele
512 of rs58475486 is common (8-14%) in African populations and showed a protective effect on lung
513 function in the presence of past year SO_2 exposure. SNP rs58475486 is located in a ccRE (SCREEN
514 accession EH37E0279296) and a FOXA1 binding site in the A549 lung adenocarcinoma cell line.
515 FOXA1 has a known compensatory role with FOXA2 during lung morphogenesis in mice (Wan *et*
516 *al.* 2005). Deletion of both FOXA1 and FOXA2 inhibited cell proliferation, epithelial cell
517 differentiation, and branching morphogenesis in fetal lung tissue. Further functional validation
518 on the effect of rs58475486 on binding affinity of FOXA1 is necessary to confirm whether the role
519 of FOXA1 in this ccRE is important for *KITLG* regulatory and lung function.

520 The higher frequency of the protective alleles of both rs73440122 and rs58465486 in African
521 populations appears to contradict previous findings that African ancestry was associated with
522 lower lung function (Kumar *et al.* 2010). One possible explanation for this seeming inconsistency
523 is that FEV_1 is a complex trait whose variation is influenced by many genetic variants of small to
524 moderate effect sizes whose influence on lung function may vary by exposure to environmental

525 factors. We found suggestive evidence that the interaction between rs73440122 and first year
526 exposure to SO_2 reverses the positive association of rs73440122 with lung function to a negative
527 one (Table 3). When assessed independently, our genetic association analysis showed that the
528 protective A allele of rs73440122 was associated with higher lung function. However, with
529 increasing levels of SO_2 exposure in the first year of life, increasing copies of the A allele of
530 rs73440122 were associated with decreased lung function. Air pollution is known to negatively
531 impact lung function, and we have previously shown that the deleterious effects of air pollution
532 on lung phenotypes may be significantly increased in African American children compared to
533 other populations experiencing the same amount of exposure (Nishimura *et al.* 2013). It has also
534 been reported that Latino and African American populations often live in neighborhoods with
535 high levels of air pollution (Mott 1995). The increased susceptibility to negative pulmonary effects
536 from air pollution exposure coupled with the disproportionate exposure to air pollution
537 experienced by the African American population may also contribute to the lower lung function
538 seen in this population despite the presence of protective alleles. The overlap of the SPI1 binding
539 site with rs73440122 further supports gene-by-SO_2 interaction at this locus, since SPI1 played a
540 critical role in the development of type 2 inflammation in the airways through macrophage
541 polarization (Qian *et al.* 2015). We noted that the rs73440122 A allele also showed an interaction
542 approaching suggestive threshold with past year exposure to SO_2 that was positively associated
543 with FEV_1. The difference is not surprising because age of exposure may significantly impact the
544 effect of air pollution on lung function (Reviewed in Usemann *et al.* 2019) . Further studies are
545 required to better understand the effect of this suggestive interaction on lung function.

546 One strength of this study is the interrogation of independent lung function associated signals at
547 our novel locus. We identified evidence of three independent signals: the replicated signal that
548 showed evidence of regulatory functions (an open chromatin region with a SPI1/PU.1 binding
549 site), one signal that showed a statistically significant gene-by-SO_2 interaction on lung function,
550 and one signal that represents to *KITLG* eQTLs in the nasal epithelial cells together with suggestive
551 gene-by-SO_2 interaction. Our results demonstrated a glimpse of the complicated genetic
552 architecture behind complex traits.

553 One limitation of this study is that the FEV_1 genetic association and the eQTL analyses with *KITLG*
554 were performed in different populations due to data availability constraints. Although we did not
555 have RNA-Seq data from lung tissues from our study subjects, we previously demonstrated that
556 there is a high degree of overlap in gene expression profiles between nasal and bronchial
557 epithelial cells (Poole *et al.* 2014). The direction of effect of the association was the same in GALA
558 II Puerto Rican children with asthma but not statistically significant. This may in part due to the
559 significantly lower African Ancestry in Puerto Ricans compared to African Americans.

560 We replicated 20 of 45 variants in black UK Biobank subjects and observed conflicting "flip-flop"
561 associations in African Americans from the SAPPHIRE and GCPD-A studies. In the past, flip-flop
562 associations were deemed as spurious results. Traditional association testing approach studies
563 the effect of each variant on phenotype independently and increases the chance of flip-flop
564 associations detected between studies. Differences in study design, sampling variation that leads
565 to variation in LD patterns, and lack of consideration of other disease influencing genetic and/or
566 environmental factors are all potential causes of flip-flop associations (Kraft *et al.* 2009; Lin *et al.*

567 2007). Hence, it is not surprising to observe flip-flop associations when gene and environment
568 interactions were detected at our FEV_1 GWAS locus. It was previously shown that flip-flop
569 associations can occur between and within populations even in the presence of a genuine genetic
570 effect (Kraft *et al.* 2009; Lin *et al.* 2007). Further functional analysis is thus required to validate
571 the relationship between the candidate variants, *KITLG* and FEV_1. This may include reporter
572 assays to validate potential enhancer or repressor activity and CRISPR-based editing assays to
573 validate the regulatory role of the candidate variants on *KITLG*. Although literature exists
574 describing KIT signaling for lung function in mice (Lindsey *et al.* 2011), additional knockout
575 experiments in a model animal system may be necessary to study how *KITLG* contributed to
576 variation in lung function.

577 The average concentration of ambient SO_2 exposure in our participants (Table 1) was lower than
578 the National Ambient Air Quality Standards. It is possible that SO_2 acted as a surrogate for other
579 unmeasured toxic pollutants emitted from local point sources. Major sources of SO_2 in San
580 Francisco Bay Area during the recruitment years of 2006 to 2011 include airports, petroleum
581 refineries, gas and oil plants, calcined petroleum coke plants, electric power plants, cement
582 manufacturing factories, chemical plants, and landfills (United States Environmental Protection
583 Agency 2008; United States Environmental Protection Agency 2011). The Environmental
584 Protection Agency's national emissions inventory data also showed that these facilities emit
585 Volatile Organic Compounds, heavy metals (lead, mercury, chromium, arsenic), formaldehyde,
586 ethyl benzene, acrolein, 1,3-butadiene, 1,4-dichlorobenzene, and tetrachloroethylene into the
587 air along with SO_2. These chemicals are highly toxic and inhaling even a small amount may

588 contribute to poor lung function. Another possibility is that exposure to SO_2 captured
589 unmeasured confounding socioeconomic factors.
590 This study identified a novel protective allele for lung function in African American children with
591 asthma. The protective association with lung function intensified with increased past year
592 exposure to SO_2. Our findings showcase the complexity of the relationship between genetic and
593 environmental factors impacting variation in FEV_1, highlights the utility of WGS data for genetic
594 research of complex phenotypes, and underscores the importance of including diverse study
595 populations in our exploration of the genetic architecture underlying lung function.

596 ACKNOWLEDGEMENTS

597 The Genes-Environments and Admixture in Latino Americans (GALA II) Study, the Study of African
598 Americans, Asthma, Genes and Environments (SAGE) Study and E.G.B. were supported by the
599 Sandler Family Foundation, the American Asthma Foundation, the RWJF Amos Medical Faculty
600 Development Program, the Harry Wm. and Diana V. Hind Distinguished Professor in
601 Pharmaceutical Sciences II, the National Heart, Lung, and Blood Institute (NHLBI) [R01HL117004,
602 R01HL128439, R01HL135156, X01HL134589]; the National Institute of Environmental Health
603 Sciences [R01ES015794]; the National Institute on Minority Health and Health Disparities
604 (NIMHD) [P60MD006902, R01MD010443], the National Human Genome Research Institute
605 [U01HG009080] and the Tobacco-Related Disease Research Program [24RT-0025]. MJW was
606 supported by the NHLBI [K01HL140218]. JJ and BEH were supported by the NHLBI [R01HL133433,
607 R01HL141992]. KLK was supported by the NHLBI [R01HL135156-S1], the UCSF Bakar Institute, the
608 Gordon and Betty Moore Foundation [GBMF3834], and the Alfred P. Sloan Foundation [2013-10-

609 27] grant to UC Berkeley through the Moore-Sloan Data Science Environment Initiative. ACW was
610 supported by the Eunice Kennedy Shriver National Institute of Child Health and Human
611 Development [1R01HD085993-01].

612 The SAPPHIRE study was supported by the Fund for Henry Ford Hospital, the American Asthma
613 Foundation, the NHLBI [R01HL118267, R01HL141485, X01HL134589], the National Institute of
614 Allergy and Infectious Diseases [R01AI079139], and the National Institute of Diabetes and
615 Digestive and Kidney Diseases [R01DK113003].

616 The GCPD-A study was supported by an Institutional award from the Children's Hospital of
617 Philadelphia and by the NHLBI [X01HL134589].

618 Part of this research was conducted using the UK Biobank Resource under Application Number
619 40375. We would like to thank UK Biobank participants and researchers who contributed or
620 collected data.

621 Whole genome sequencing (WGS) for the Trans-Omics in Precision Medicine (TOPMed) program
622 was supported by the National Heart, Lung and Blood Institute (NHLBI). WGS for "NHLBI TOPMed:
623 Gene-Environment, Admixture and Latino Asthmatics Study" (phs000920) and "NHLBI TOPMed:
624 Study of African Americans, Asthma, Genes and Environments" (phs000921) was performed at
625 the New York Genome Center (3R01HL117004-02S3) and the University of Washington
626 Northwest Genomics Center (HHSN268201600032I). WGS for "NHLBI TOPMed: Study of Asthma
627 Phenotypes & Pharmacogenomic Interactions by Race-Ethnicity" (phs001467) and "Genetics of
628 Complex Pediatric Disorders - Asthma" (phs001661) was performed at the University of

629 Washington Northwest Genomics Center (HHSN268201600032I). Centralized read mapping and
630 genotype calling, along with variant quality metrics and filtering were provided by the TOPMed
631 Informatics Research Center (3R01HL-117626-02S1; contract HHSN268201800002I). Phenotype
632 harmonization, data management, sample-identity QC, and general study coordination were
633 provided by the TOPMed Data Coordinating Center (3R01HL-120393-02S1; contract
634 HHSN268201800001I). We gratefully acknowledge the studies and participants who provided
635 biological samples and data for TOPMed.

636 WGS of part of GALA II was performed by New York Genome Center under The Centers for
637 Common Disease Genomics of the Genome Sequencing Program (GSP) Grant (UM1 HG008901).
638 The GSP Coordinating Center (U24 HG008956) contributed to cross-program scientific initiatives
639 and provided logistical and general study coordination. GSP is funded by the National Human
640 Genome Research Institute, the National Heart, Lung, and Blood Institute, and the National Eye
641 Institute.

642 The TOPMed imputation panel was supported by the NHLBI and TOPMed study investigators who
643 contributed data to the reference panel. The panel was constructed and implemented by the
644 TOPMed Informatics Research Center at the University of Michigan (3R01HL-117626-02S1;
645 contract HHSN268201800002I). The TOPMed Data Coordinating Center (3R01HL-120393-02S1;
646 contract HHSN268201800001I) provided additional data management, sample identity checks,
647 and overall program coordination and support. We gratefully acknowledge the studies and
648 participants who provided biological samples and data for TOPMed.

649 The authors wish to acknowledge the following GALA II and SAGE study collaborators: Shannon
650 Thyne, UCSF; Harold J. Farber, Texas Children's Hospital; Denise Serebrisky, Jacobi Medical Center;
651 Rajesh Kumar, Lurie Children's Hospital of Chicago; Emerita Brigino-Buenaventura, Kaiser
652 Permanente; Michael A. LeNoir, Bay Area Pediatrics; Kelley Meade, UCSF Benioff Children's
653 Hospital, Oakland; William Rodríguez-Cintrón, VA Hospital, Puerto Rico; Pedro C. Ávila,
654 Northwestern University; Jose R. Rodríguez-Santana, Centro de Neumología Pediátrica; Luisa N.
655 Borrell, City University of New York; Adam Davis, UCSF Benioff Children's Hospital, Oakland;
656 Saunak Sen, University of Tennessee.

657 The authors acknowledge the families and patients for their participation and thank the
658 numerous health care providers and community clinics for their support and participation in
659 GALA II and SAGE. In particular, the authors thank the recruiters who obtained the data: Duanny
660 Alva, MD; Gaby Ayala-Rodríguez; Lisa Caine, RT; Elizabeth Castellanos; Jaime Colón; Denise
661 DeJesus; Blanca López; Brenda López, MD; Louis Martos; Vivian Medina; Juana Olivo; Mario
662 Peralta; Esther Pomares, MD; Jihan Quraishi; Johanna Rodríguez; Shahdad Saeedi; Dean Soto;
663 and Ana Taveras.

664 The authors thank María Pino-Yanes for providing feedback on this study and Thomas W
665 Blackwell for providing critical review on this manuscript.

666 The content is solely the responsibility of the authors and does not necessarily represent the
667 official views of the National Institutes of Health.

668

669 LITERATURE CITED

670 1000 Genomes Project Consortium, A. Auton, L. D. Brooks, R. M. Durbin, E. P. Garrison *et al*,
671 2015 A global reference for human genetic variation. Nature **526:** 68-74.

672 Akinbami, L. J., 2015 *Asthma Prevalence, Health Care use and Mortality: United States, 2003-05*.
673 [Online] Available at: http://www.cdc.gov/nchs/data/hestat/asthma03-05/asthma03-05.htm.
674 [Accessed 2020 Jan 8].

675 Akinbami, L. J., J. E. Moorman, A. E. Simon and K. C. Schoendorf, 2014 Trends in racial
676 disparities for asthma outcomes among children 0 to 17 years, 2001-2010. J. Allergy Clin.
677 Immunol. **134:** 547-553.e5.

678 Alexander, D. H., J. Novembre and K. Lange, 2009 Fast model-based estimation of ancestry in
679 unrelated individuals. Genome Res. **19:** 1655-1664.

680 Al-Muhsen, S. Z., G. Shablovsky, R. Olivenstein, B. Mazer and Q. Hamid, 2004 The expression of
681 stem cell factor and c-kit receptor in human asthmatic airways. Clin. Exp. Allergy **34:** 911-916.

682 Amin, K., 2012 The role of mast cells in allergic inflammation. Respir. Med. **106:** 9-14.

683 Anders, S., P. T. Pyl and W. Huber, 2015 HTSeq--a python framework to work with high-
684 throughput sequencing data. Bioinformatics **31:** 166-169.

685 Anderson, K. L., H. Perkin, C. D. Surh, S. Venturini, R. A. Maki *et al*, 2000 Transcription factor
686 PU.1 is necessary for development of thymic and myeloid progenitor-derived dendritic cells. J.
687 Immunol. **164:** 1855-1861.

688 Ay, F., T. L. Bailey and W. S. Noble, 2014 Statistical confidence estimation for hi-C data reveals
689 regulatory chromatin contacts. Genome Res. **24:** 999-1011.

690 Balzar, S., H. W. Chu, M. Strand and S. Wenzel, 2005 Relationship of small airway chymase-
691 positive mast cells and lung function in severe asthma. Am. J. Respir. Crit. Care Med. **171:** 431-
692 439.

693 Barraza-Villarreal, A., J. Sunyer, L. Hernandez-Cadena, M. C. Escamilla-Nunez, J. J. Sienra-Monge
694 *et al*, 2008 Air pollution, airway inflammation, and lung function in a cohort study of mexico city
695 schoolchildren. Environ. Health Perspect. **116:** 832-838.

696 Barrett, J. C., B. Fry, J. Maller and M. J. Daly, 2005 Haploview: Analysis and visualization of LD
697 and haplotype maps. Bioinformatics **21:** 263-265.

698 Borish, L., and B. Z. Joseph, 1992 Inflammation and the allergic response. Med. Clin. North Am.
699 **76:** 765-787.

700 Broudy, V. C., 1997 Stem cell factor and hematopoiesis. Blood **90:** 1345-1364.

701 Brunekreef, B., and S. T. Holgate, 2002 Air pollution and health. Lancet **360:** 1233-1242.

702 Buniello, A., J. A. L. MacArthur, M. Cerezo, L. W. Harris, J. Hayhurst *et al*, 2019 The NHGRI-EBI
703 GWAS catalog of published genome-wide association studies, targeted arrays and summary
704 statistics 2019. Nucleic Acids Res. **47:** D1005-D1012.

705 Bycroft, C., C. Freeman, D. Petkova, G. Band, L. T. Elliott *et al*, 2018 The UK biobank resource
706 with deep phenotyping and genomic data. Nature **562:** 203-209.

707 Canela-Xandri, O., K. Rawlik and A. Tenesa, 2018 An atlas of genetic associations in UK biobank.
708 Nat. Genet. **50:** 1593-1599.

709 Carlson, C. S., T. C. Matise, K. E. North, C. A. Haiman, M. D. Fesinmeyer *et al*, 2013
710 Generalization and dilution of association results from european GWAS in populations of non-
711 european ancestry: The PAGE study. PLoS Biol. **11:** e1001661.

712 Chang, C. C., C. C. Chow, L. C. Tellier, S. Vattikuti, S. M. Purcell *et al*, 2015 Second-generation
713 PLINK: Rising to the challenge of larger and richer datasets. Gigascience **4:** 7-8. eCollection
714 2015.

715 Chatterjee, S., and N. Das, 1995 Lung function in indian twin children: Comparison of genetic
716 versus environmental influence. Ann. Hum. Biol. **22:** 289-303.

717 Chen, Y., S. L. Horne, D. C. Rennie and J. A. Dosman, 1996 Segregation analysis of two lung
718 function indices in a random sample of young families: The humboldt family study. Genet.
719 Epidemiol. **13:** 35-47.

720 Ciesielski, T. H., S. A. Pendergrass, M. J. White, N. Kodaman, R. S. Sobota *et al*, 2014 Diverse
721 convergent evidence in the genetic analysis of complex disease: Coordinating omic, informatic,
722 and experimental evidence to better identify and validate risk factors. BioData Min. **7:** 10-10.
723 eCollection 2014.

724 Comberiati, P., M. E. Di Cicco, S. D'Elios and D. G. Peroni, 2017 How much asthma is atopic in
725 children? Front. Pediatr. **5:** 122.

726 Conomos, M. P., M. B. Miller and T. A. Thornton, 2015 Robust inference of population structure
727 for ancestry prediction and correction of stratification in the presence of relatedness. Genet.
728 Epidemiol. **39:** 276-293.

729 Conomos, M. P., A. P. Reiner, B. S. Weir and T. A. Thornton, 2016 Model-free estimation of
730 recent genetic relatedness. Am. J. Hum. Genet. **98:** 127-148.

731 Cruse, G., and P. Bradding, 2016 Mast cells in airway diseases and interstitial lung disease. Eur.
732 J. Pharmacol. **778:** 125-138.

733 Da Silva, C. A., and N. Frossard, 2005 Regulation of stem cell factor expression in inflammation
734 and asthma. Mem. Inst. Oswaldo Cruz **100 Suppl 1:** 145-151.

735 Da Silva, C. A., L. Reber and N. Frossard, 2006 Stem cell factor expression, mast cells and
736 inflammation in asthma. Fundam. Clin. Pharmacol. **20:** 21-39.

737 Dahlin, J. S., A. Malinovschi, H. Ohrvik, M. Sandelin, C. Janson *et al*, 2016 Lin- CD34hi
738 CD117int/hi FcepsilonRI+ cells in human blood constitute a rare population of mast cell
739 progenitors. Blood **127:** 383-391.

740 Das, S., L. Forer, S. Schonherr, C. Sidore, A. E. Locke *et al*, 2016 Next-generation genotype
741 imputation service and methods. Nat. Genet. **48:** 1284-1287.

742 Duggal, P., E. M. Gillanders, T. N. Holmes and J. E. Bailey-Wilson, 2008 Establishing an adjusted
743 p-value threshold to control the family-wide type 1 error in genome wide association studies.
744 BMC Genomics **9:** 516-516.

745 ENCODE Project Consortium, 2012 An integrated encyclopedia of DNA elements in the human
746 genome. Nature **489:** 57-74.

747 ENCODE Project Consortium, 2011 A user's guide to the encyclopedia of DNA elements
748 (ENCODE). PLoS Biol. **9:** e1001046.

749 Fajt, M. L., and S. E. Wenzel, 2013 Mast cells, their subtypes, and relation to asthma
750 phenotypes. Ann. Am. Thorac. Soc. **10 Suppl:** 158.

751 Fonseca, W., A. J. Rasky, C. Ptaschinski, S. H. Morris, S. K. K. Best *et al*, 2019 Group 2 innate
752 lymphoid cells (ILC2) are regulated by stem cell factor during chronic asthmatic disease.
753 Mucosal Immunol. **12:** 445-456.

754 Galli, S. J., K. M. Zsebo and E. N. Geissler, 1994 The kit ligand, stem cell factor. Adv. Immunol.
755 **55:** 1-96.

756 Galli, S. J., M. Tsai, B. K. Wershil, S. Y. Tam and J. J. Costa, 1995 Regulation of mouse and human
757 mast cell development, survival and function by stem cell factor, the ligand for the c-kit
758 receptor. Int. Arch. Allergy Immunol. **107:** 51-53.

759 GTEx Consortium, Laboratory, Data Analysis &Coordinating Center (LDACC)-Analysis Working
760 Group, Statistical Methods groups-Analysis Working Group, Enhancing GTEx (eGTEx) groups,
761 NIH Common Fund *et al*, 2017 Genetic effects on gene expression across human tissues. Nature
762 **550:** 204-213.

763 Guerriero, A., P. B. Langmuir, L. M. Spain and E. W. Scott, 2000 PU.1 is required for myeloid-
764 derived but not lymphoid-derived dendritic cells. Blood **95:** 879-885.

765 Hankinson, J. L., J. R. Odencrantz and K. B. Fedan, 1999 Spirometric reference values from a
766 sample of the general U.S. population. Am. J. Respir. Crit. Care Med. **159:** 179-187.

767 Hinrichs, A. S., D. Karolchik, R. Baertsch, G. P. Barber, G. Bejerano *et al*, 2006 The UCSC genome
768 browser database: Update 2006. Nucleic Acids Res. **34:** 590.

769 Hsieh, F. H., P. Sharma, A. Gibbons, T. Goggans, S. C. Erzurum *et al*, 2005 Human airway
770 epithelial cell determinants of survival and functional phenotype for primary human mast cells.
771 Proc. Natl. Acad. Sci. U. S. A. **102:** 14380-14385.

772 Hukkinen, M., J. Kaprio, U. Broms, A. Viljanen, D. Kotz *et al*, 2011 Heritability of lung function: A
773 twin study among never-smoking elderly women. Twin Res. Hum. Genet. **14:** 401-407.

774 Iemura, A., M. Tsai, A. Ando, B. K. Wershil and S. J. Galli, 1994 The c-kit ligand, stem cell factor,
775 promotes mast cell survival by suppressing apoptosis. Am. J. Pathol. **144:** 321-328.

776 Ierodiakonou, D., A. Zanobetti, B. A. Coull, S. Melly, D. S. Postma *et al*, 2016 Ambient air
777 pollution, lung function, and airway responsiveness in asthmatic children. J. Allergy Clin.
778 Immunol. **137:** 390-399.

779 Ionita-Laza, I., S. Lee, V. Makarov, J. D. Buxbaum and X. Lin, 2013 Sequence kernel association
780 tests for the combined effect of rare and common variants. Am. J. Hum. Genet. **92:** 841-853.

781 Ito, T., C. Nishiyama, M. Nishiyama, H. Matsuda, K. Maeda *et al*, 2005 Mast cells acquire
782 monocyte-specific gene expression and monocyte-like morphology by overproduction of PU.1.
783 J. Immunol. **174:** 376-383.

784 Ito, T., C. Nishiyama, N. Nakano, M. Nishiyama, Y. Usui *et al*, 2009 Roles of PU.1 in monocyte-
785 and mast cell-specific gene regulation: PU.1 transactivates CIITA pIV in cooperation with IFN-
786 gamma. Int. Immunol. **21:** 803-816.

787 Jiang, H., R. Lei, S. W. Ding and S. Zhu, 2014 Skewer: A fast and accurate adapter trimmer for
788 next-generation sequencing paired-end reads. BMC Bioinformatics **15:** 182-182.

789 Johnson, J. D., and W. M. Theurer, 2014 A stepwise approach to the interpretation of
790 pulmonary function tests. Am. Fam. Physician **89:** 359-366.

791 Karolchik, D., A. S. Hinrichs, T. S. Furey, K. M. Roskin, C. W. Sugnet *et al*, 2004 The UCSC table
792 browser data retrieval tool. Nucleic Acids Res. **32:** 493.

793 Kassel, O., F. Schmidlin, C. Duvernelle, B. Gasser, G. Massard *et al*, 1999 Human bronchial
794 smooth muscle cells in culture produce stem cell factor. Eur. Respir. J. **13:** 951-954.

795 Kim, D., B. Langmead and S. L. Salzberg, 2015 HISAT: A fast spliced aligner with low memory
796 requirements. Nat. Methods **12:** 357-360.

797 Kim, M. S., K. P. Patel, A. K. Teng, A. J. Berens and J. Lachance, 2018 Genetic disease risks can be
798 misestimated across global populations. Genome Biol. **19:** 179-7.

799 Kobayashi, H., Y. Okayama, T. Ishizuka, R. Pawankar, C. Ra *et al*, 1998 Production of IL-13 by
800 human lung mast cells in response to fcepsilon receptor cross-linkage. Clin. Exp. Allergy **28:**
801 1219-1227.

802 Kraft, P., E. Zeggini and J. P. Ioannidis, 2009 Replication in genome-wide association studies.
803 Stat. Sci. **24:** 561-573.

804 Kumar, R., M. A. Seibold, M. C. Aldrich, L. K. Williams, A. P. Reiner *et al*, 2010 Genetic ancestry
805 in lung-function predictions. N. Engl. J. Med. **363:** 321-330.

806 Lee, S., M. J. Emond, M. J. Bamshad, K. C. Barnes, M. J. Rieder *et al*, 2012 Optimal unified
807 approach for rare-variant association testing with application to small-sample case-control
808 whole-exome sequencing studies. Am. J. Hum. Genet. **91:** 224-237.

809 Levin, A. M., Y. Wang, K. E. Wells, B. Padhukasahasram, J. J. Yang *et al*, 2014 Nocturnal asthma
810 and the importance of race/ethnicity and genetic ancestry. Am. J. Respir. Crit. Care Med. **190:**
811 266-273.

812 Li, X., G. A. Hawkins, E. J. Ampleford, W. C. Moore, H. Li *et al*, 2013 Genome-wide association
813 study identifies TH1 pathway genes associated with lung function in asthmatic patients. J.
814 Allergy Clin. Immunol. **132:** 313-20.e15.

815 Liao, S. Y., X. Lin and D. C. Christiani, 2014 Genome-wide association and network analysis of
816 lung function in the framingham heart study. Genet. Epidemiol. **38:** 572-578.

817 Lin, P. I., J. M. Vance, M. A. Pericak-Vance and E. R. Martin, 2007 No gene is an island: The flip-
818 flop phenomenon. Am. J. Hum. Genet. **80:** 531-538.

819 Lindsey, J. Y., K. Ganguly, D. M. Brass, Z. Li, E. N. Potts *et al*, 2011 C-kit is essential for alveolar
820 maintenance and protection from emphysema-like disease in mice. Am. J. Respir. Crit. Care
821 Med. **183:** 1644-1652.

822 Liu, X., S. White, B. Peng, A. D. Johnson, J. A. Brody *et al*, 2016 WGSA: An annotation pipeline
823 for human genome sequencing studies. J. Med. Genet. **53:** 111-112.

824 Mak, A. C. Y., M. J. White, W. L. Eckalbar, Z. A. Szpiech, S. S. Oh *et al*, 2018 Whole-genome
825 sequencing of pharmacogenetic drug response in racially diverse children with asthma. Am. J.
826 Respir. Crit. Care Med. **197:** 1552-1564.

827 Mak, A. C., M. J. White, C. Eng, D. Hu, S. Huntsman *et al*, 2016 *Whole Genome Sequencing to*
828 *Identify Genetic Variation Associated with Bronchodilator Response in Minority Children with*
829 *Asthma*.

830 Martin, A. R., C. R. Gignoux, R. K. Walters, G. L. Wojcik, B. M. Neale *et al*, 2017 Human
831 demographic history impacts genetic risk prediction across diverse populations. Am. J. Hum.
832 Genet. **100:** 635-649.

833 Martin, J. S., Z. Xu, A. P. Reiner, K. L. Mohlke, P. Sullivan *et al*, 2017 HUGIn: Hi-C unifying
834 genomic interrogator. Bioinformatics **33:** 3793-3795.

835 McKercher, S. R., B. E. Torbett, K. L. Anderson, G. W. Henkel, D. J. Vestal *et al*, 1996 Targeted
836 disruption of the PU.1 gene results in multiple hematopoietic abnormalities. EMBO J. **15:** 5647-
837 5658.

838 Mekori, Y. A., C. K. Oh and D. D. Metcalfe, 1993 IL-3-dependent murine mast cells undergo
839 apoptosis on removal of IL-3. prevention of apoptosis by c-kit ligand. J. Immunol. **151:** 3775-
840 3784.

841 Mendez-Enriquez, E., and J. Hallgren, 2019 Mast cells and their progenitors in allergic asthma.
842 Front. Immunol. **10:** 821.

843 Moore, J. H., 2005 A global view of epistasis. Nat. Genet. **37:** 13-14.

844 Moore, J. H., and S. M. Williams, 2009 Epistasis and its implications for personal genetics. Am. J.
845 Hum. Genet. **85:** 309-320.

846 Mott, L., 1995 The disproportionate impact of environmental health threats on children of
847 color. Environ. Health Perspect. **103 Suppl 6:** 33-35.

848 Neophytou, A. M., M. J. White, S. S. Oh, N. Thakur, J. M. Galanter *et al*, 2016 Air pollution and
849 lung function in minority youth with asthma in the GALA II (genes-environments and admixture
850 in latino americans) and SAGE II (study of african americans, asthma, genes, and environments)
851 studies. Am. J. Respir. Crit. Care Med. **193:** 1271-1280.

852 Nishimura, K. K., J. M. Galanter, L. A. Roth, S. S. Oh, N. Thakur *et al*, 2013 Early-life air pollution
853 and asthma risk in minority children. the GALA II and SAGE II studies. Am. J. Respir. Crit. Care
854 Med. **188:** 309-318.

855 Nishiyama, C., M. Nishiyama, T. Ito, S. Masaki, N. Masuoka *et al*, 2004 Functional analysis of
856 PU.1 domains in monocyte-specific gene regulation. FEBS Lett. **561:** 63-68.

857 Nishiyama, C., M. Nishiyama, T. Ito, S. Masaki, K. Maeda *et al*, 2004 Overproduction of PU.1 in
858 mast cell progenitors: Its effect on monocyte- and mast cell-specific gene expression. Biochem.
859 Biophys. Res. Commun. **313:** 516-521.

860 Oh, S. S., M. J. White, C. R. Gignoux and E. G. Burchard, 2016 Making precision medicine socially
861 precise. take a deep breath. Am. J. Respir. Crit. Care Med. **193:** 348-350.

862 Oh, S. S., H. Tcheurekdjian, L. A. Roth, E. A. Nguyen, S. Sen *et al*, 2012 Effect of secondhand
863 smoke on asthma control among black and latino children. J. Allergy Clin. Immunol. **129:** 1478-
864 83.e7.

865 Oliveira, S. H., and N. W. Lukacs, 2003 Stem cell factor: A hemopoietic cytokine with important
866 targets in asthma. Curr. Drug Targets Inflamm. Allergy **2:** 313-318.

867 Ong, B. A., J. Li, J. M. McDonough, Z. Wei, C. Kim *et al*, 2013 Gene network analysis in a
868 pediatric cohort identifies novel lung function genes. PLoS One **8:** e72899.

869 Oriss, T. B., N. Krishnamoorthy, P. Ray and A. Ray, 2014 Dendritic cell c-kit signaling and
870 adaptive immunity: Implications for the upper airways. Curr. Opin. Allergy Clin. Immunol. **14:** 7-
871 12.

872 Palmer, L. J., M. W. Knuiman, M. L. Divitini, P. R. Burton, A. L. James *et al*, 2001 Familial
873 aggregation and heritability of adult lung function: Results from the busselton health study. Eur.
874 Respir. J. **17:** 696-702.

875 Pe'er, I., R. Yelensky, D. Altshuler and M. J. Daly, 2008 Estimation of the multiple testing burden
876 for genomewide association studies of nearly all common variants. Genet. Epidemiol. **32:** 381-
877 385.

878 Pino-Yanes, M., N. Thakur, C. R. Gignoux, J. M. Galanter, L. A. Roth *et al*, 2015 Genetic ancestry
879 influences asthma susceptibility and lung function among latinos. J. Allergy Clin. Immunol. **135:**
880 228-235.

881 Plummer, M., N. Best, K. Cowles and K. Vines, 2006 CODA: Convergence diagnosis and output
882 analysis for MCMC. R News **6:** 7-11.

883 Poole, A., C. Urbanek, C. Eng, J. Schageman, S. Jacobson *et al*, 2014 Dissecting childhood asthma
884 with nasal transcriptomics distinguishes subphenotypes of disease. J. Allergy Clin. Immunol.
885 **133:** 670-8.e12.

886 Pruim, R. J., R. P. Welch, S. Sanna, T. M. Teslovich, P. S. Chines *et al*, 2010 LocusZoom: Regional
887 visualization of genome-wide association scan results. Bioinformatics **26:** 2336-2337.

888 Purcell, S., and C. Chang, 2013 *Plink 1.9*. [Online] Available at: www.cog-
889 genomics.org/plink/1.9/. [Accessed 2019 Mar].

890 Qian, F., J. Deng, Y. G. Lee, J. Zhu, M. Karpurapu *et al*, 2015 The transcription factor PU.1
891 promotes alternative macrophage polarization and asthmatic airway inflammation. J. Mol. Cell.
892 Biol. **7:** 557-567.

893 Repapi, E., I. Sayers, L. V. Wain, P. R. Burton, T. Johnson *et al*, 2010 Genome-wide association
894 study identifies five loci associated with lung function. Nat. Genet. **42:** 36-44.

895 Rochman, M., A. V. Kartashov, J. M. Caldwell, M. H. Collins, E. M. Stucke *et al*, 2015
896 Neurotrophic tyrosine kinase receptor 1 is a direct transcriptional and epigenetic target of IL-13
897 involved in allergic inflammation. Mucosal Immunol. **8:** 785-798.

898 Schmitt, A. D., M. Hu, I. Jung, Z. Xu, Y. Qiu *et al*, 2016 A compendium of chromatin contact maps
899 reveals spatially active regions in the human genome. Cell. Rep. **17:** 2042-2059.

900 Scott, E. W., M. C. Simon, J. Anastasi and H. Singh, 1994 Requirement of transcription factor
901 PU.1 in the development of multiple hematopoietic lineages. Science **265:** 1573-1577.

902 Scott, E. W., R. C. Fisher, M. C. Olson, E. W. Kehrli, M. C. Simon *et al*, 1997 PU.1 functions in a
903 cell-autonomous manner to control the differentiation of multipotential lymphoid-myeloid
904 progenitors. Immunity **6:** 437-447.

905 Sillanpaa, E., S. Sipila, T. Tormakangas, J. Kaprio and T. Rantanen, 2017 Genetic and
906 environmental effects on telomere length and lung function: A twin study. J. Gerontol. A Biol.
907 Sci. Med. Sci. **72:** 1561-1568.

908 Sofer, T., X. Zheng, S. M. Gogarten, C. A. Laurie, K. Grinde *et al*, 2019 A fully adjusted two-stage
909 procedure for rank-normalization in genetic association studies. Genet. Epidemiol. **43:** 263-275.

910 Soler Artigas, M., L. V. Wain, S. Miller, A. K. Kheirallah, J. E. Huffman *et al*, 2015 Sixteen new
911 lung function signals identified through 1000 genomes project reference panel imputation. Nat.
912 Commun. **6:** 8658.

913 Soler Artigas, M., D. W. Loth, L. V. Wain, S. A. Gharib, M. Obeidat *et al*, 2011 Genome-wide
914 association and large-scale follow up identifies 16 new loci influencing lung function. Nat.
915 Genet. **43:** 1082-1090.

916 Summer Institute in Statistical Genetics, 2019 *PC-Relate*. [Online] Available at: https://uw-
917 gac.github.io/SISG_2019/pc-relate.html. [Accessed 2019 Jul 25].

918 Taliun, D., D. N. Harris, M. D. Kessler, J. Carlson, Z. A. Szpiech *et al*, 2019 Sequencing of 53,831
919 diverse genomes from the NHLBI TOPMed program. bioRxiv 563866.

920 Tayel, S. I., S. M. El-Hefnway, Abd El Gayed, E. M. and G. A. Abdelaal, 2017 Association of stem
921 cell factor gene expression with severity and atopic state in patients with bronchial asthma.
922 Respir. Res. **18:** 21-2.

923 Tian, X., C. Xu, Y. Wu, J. Sun, H. Duan *et al*, 2017 Genetic and environmental influences on
924 pulmonary function and muscle strength: The chinese twin study of aging. Twin Res. Hum.
925 Genet. **20:** 53-59.

926 TOPMed, 2019 *TOPMed Whole Gneome Sequencing Methods: Freeze 8*. [Online] Available at:
927 https://www.nhlbiwgs.org/topmed-whole-genome-sequencing-methods-freeze-8. [Accessed
928 2019 Dec 13].

929 United States Environmental Protection Agency, 2011 *National Emissions Inventory (NEI) 2011*
930 *Data*. [Online] Available at: https://www.epa.gov/air-emissions-inventories/2011-national-
931 emissions-inventory-nei-data. [Accessed 2020 Jan 8].

932 United States Environmental Protection Agency, 2008 *National Emissions Inventory (NEI) 2008*
933 *Data*. [Online] Available at: https://www.epa.gov/air-emissions-inventories/2008-national-
934 emissions-inventory-nei-data. [Accessed 2020 Jan 8].

935 University of Michigan, and NHLBI TOPMed, 2018 *BRAVO Variant Browser*. [Online] Available
936 at: https://bravo.sph.umich.edu/freeze5/hg38/. [Accessed 2019 Aug].

937 Usemann, J., F. Decrue, I. Korten, E. Proietti, O. Gorlanova *et al*, 2019 Exposure to moderate air
938 pollution and associations with lung function at school-age: A birth cohort study. Environ. Int.
939 **126:** 682-689.

940 Valent, P., E. Spanblochl, W. R. Sperr, C. Sillaber, K. M. Zsebo *et al*, 1992 Induction of
941 differentiation of human mast cells from bone marrow and peripheral blood mononuclear cells
942 by recombinant human stem cell factor/kit-ligand in long-term culture. Blood **80:** 2237-2245.

943 Wain, L. V., N. Shrine, M. S. Artigas, A. M. Erzurumluoglu, B. Noyvert *et al*, 2017 Genome-wide
944 association analyses for lung function and chronic obstructive pulmonary disease identify new
945 loci and potential druggable targets. Nat. Genet. **49:** 416-425.

946 Walsh, J. C., R. P. DeKoter, H. J. Lee, E. D. Smith, D. W. Lancki *et al*, 2002 Cooperative and
947 antagonistic interplay between PU.1 and GATA-2 in the specification of myeloid cell fates.
948 Immunity **17:** 665-676.

949 Wan, H., S. Dingle, Y. Xu, V. Besnard, K. H. Kaestner *et al*, 2005 Compensatory roles of Foxa1
950 and Foxa2 during lung morphogenesis. J. Biol. Chem. **280:** 13809-13816.

951 Wen, L. P., J. A. Fahrni, S. Matsui and G. D. Rosen, 1996 Airway epithelial cells produce stem cell
952 factor. Biochim. Biophys. Acta **1314:** 183-186.

953 White, M. J., O. Risse-Adams, P. Goddard, M. G. Contreras, J. Adams *et al*, 2016 Novel genetic
954 risk factors for asthma in african american children: Precision medicine and the SAGE II study.
955 Immunogenetics **68:** 391-400.

956 Wise, J., 2019 Air pollution is linked to infant deaths and reduced lung function in children. BMJ
957 **366:** l5772.

958 Wojcik, G. L., M. Graff, K. K. Nishimura, R. Tao, J. Haessler *et al*, 2019 Genetic analyses of
959 diverse populations improves discovery for complex traits. Nature **570:** 514-518.

960 World Health Organization, 2017 *Asthma*. [Online] Available at:
961 http://www.who.int/mediacentre/factsheets/fs307/en/. [Accessed 2020 Jan 8].

962 Yamada, H., Y. Yatagai, H. Masuko, T. Sakamoto, H. Iijima *et al*, 2015 Heritability of pulmonary
963 function estimated from genome-wide SNPs in healthy japanese adults. Respir. Investig. **53:** 60-
964 67.

965 Yashiro, T., S. Nakano, K. Nomura, Y. Uchida, K. Kasakura *et al*, 2019 A transcription factor PU.1
966 is critical for Ccl22 gene expression in dendritic cells and macrophages. Sci. Rep. **9:** 1161-9.

967 Zhang, F., and J. R. Lupski, 2015 Non-coding genetic variants in human disease. Hum. Mol. Genet.
968 **24:** 102.

969

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

	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%)

SO2 first year exposure (ppb)	
Mean (SD)	1.59 (0.961)
Median [25%, 75%]	1.50 [1.24, 1.87]
Missing	227 (26.2%)
SO2 past year exposure (ppb)	
Mean (SD)	1.10 (0.302)
Median [25%, 75%]	1.08 [0.910, 1.27]
Missing	206 (23.8%)
SO2 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%)

971 25% and 75%, 25 and 75 percentiles. Control medications include inhaled corticosteroid (ICS),

972 long-acting beta-agonist (LABA), leukotriene inhibitor and/or ICS/LABA combo. SO_2 exposure

973 are hourly exposure averaged over the specified time period before spirometry testing as

974 previously described in Neophytou *et al.* 2016. ppb, parts per billion or $\mu g/m^3$.

975

976 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 ALL	AFR	AMR	EUR	Score
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*	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[&]	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[&]	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[&]	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*	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[&]	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[&]	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	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[&]	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 evidences which are 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 greater than 2 (^) or if they are eQTL of KITLG in nasal epithelial cells (#). The three independent signals identified in the conditional analyses are marked with * near the rsID. Indels were marked with the superscript & near the rsID. β (p) and β_{norm}(p_{norm}) are the effect sizes (p-values) of the genetic associations of the alternate allele (alt) with FEV_1 and FEV_1.res.rnorm respectively. MAF minor allele frequency. ALL/AFR/AMR/EUR, 1000 Genomes minor allele frequency from all/African/American/European populations. -, not available.

986 Table 3. Gene-and-environment analysis on FEV_1

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[2]	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[1]	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[2]	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[3]	SO_2 past year	661	0.11	8.82E-06	0.05	0.362	0.20	0.026	*
rs17016066_A[3]	SO_2 past year	661	0.11	8.82E-06	0.05	0.362	0.20	0.026	*
rs1398303_A[3]	SO_2 past year	661	0.10	3.15E-05	0.05	0.362	0.21	0.021	*
rs61924868_T[3]	SO_2 past year	661	0.10	3.30E-05	0.05	0.362	0.21	0.023	*
rs73429415_A[3]	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	

987 *, $p < 0.05$; **, p < bonferroni p value of 0.0056 for GxE analysis. n, sample sizes for the gene-

988 by-SO_2 interaction analysis. Superscript 1 to 3 in the variant column, variants that are in LD with

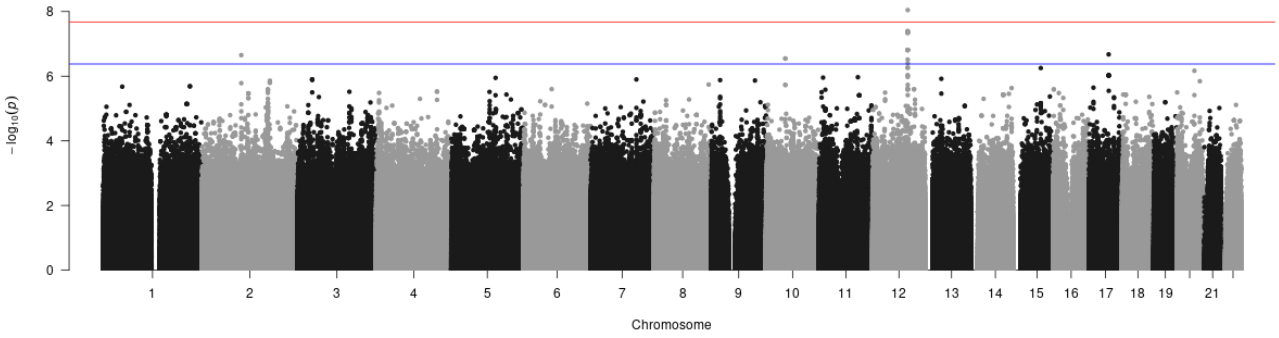
989 the 3 independent signals, rs58475486, rs73429450 and rs17016065, respectively. β, effect

990 sizes from the main effects of the variants, exposure and GxE interaction, respectively.

991

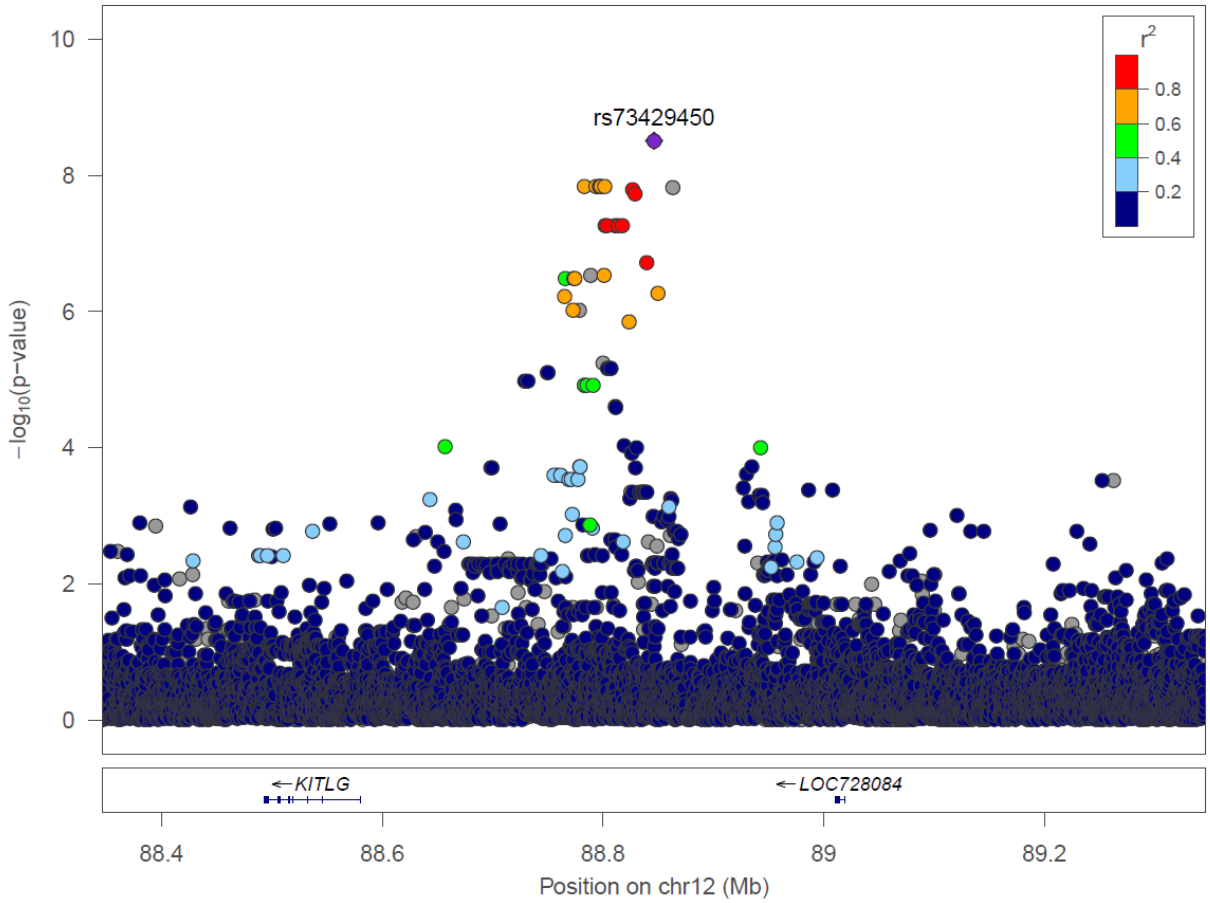
992 Figure 1.

993 A.



994

995 B.



996

997 Figure 1. Manhattan and LocusZoom plots from genome-wide association study of lung function*.

998 (A) Manhattan plot from genome-wide association study of lung function* using linear regression

999 in ENCORE. Red horizontal line: CODA-adjusted genome-wide significance p-value of 2.10×10^{-8}.

1000 Blue horizontal line: CODA-adjusted suggestive significance p-value of 4.19×10^{-7}. (B) LocusZoom

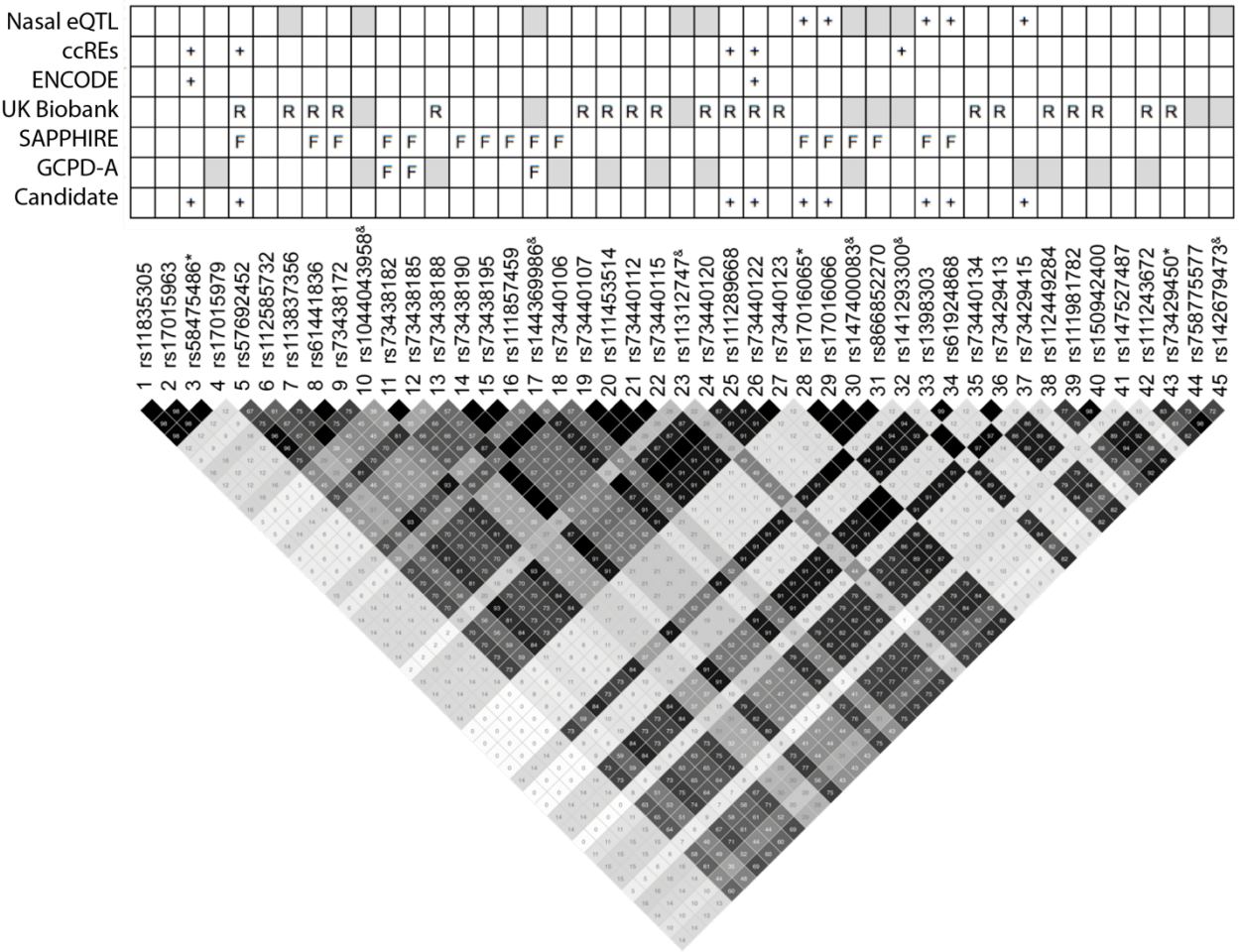
1001 plot of rs73429450 (chr12 : 88846435) and 500 Kb flanking region. Colors show linkage

1002 disequilibrium in the study population. * FEV_1.res.rnorm was used as the phenotype for the

1003 association testing.

1004

1005 Figure 2.



1006

1007 Figure 2. Integration of statistical and functional evidence for variant prioritization. Numbers and

1008 different shades of black in the LD plot represent LD in R^2. The three independent signals

1009 identified in the conditional analysis are marked with *. Indels are marked with &. Nasal eQTL,

1010 variants eQTL of *KITLG* in nasal epithelial cells. ccREs, candidate cis-regulatory elements in

1011 SCREEN registry. ENCORE, DNase I hypersensitivity site and/or transcription factor ChIP-Seq

1012 overlapping with the variants. UK Biobank, SAPPHIRE, GCPD-A, replication results using Blacks in

1013 UK Biobank and African Americans in the SAPPHIRE and GCPD-A cohorts (R = replicated at p <

1014 0.05; F = flip-flop association at $p < 0.05$). Candidate, candidate variants prioritized because of

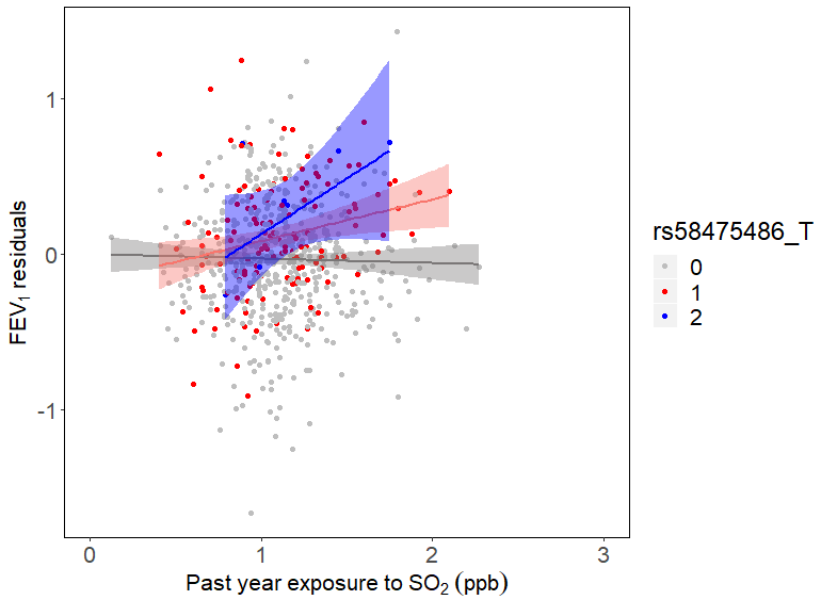
1015 presence of two or more evidence or is nasal eQTL. + indicates presence of evidence. Boxes in

1016 the top panel were shaded grey if results were not available.

1017

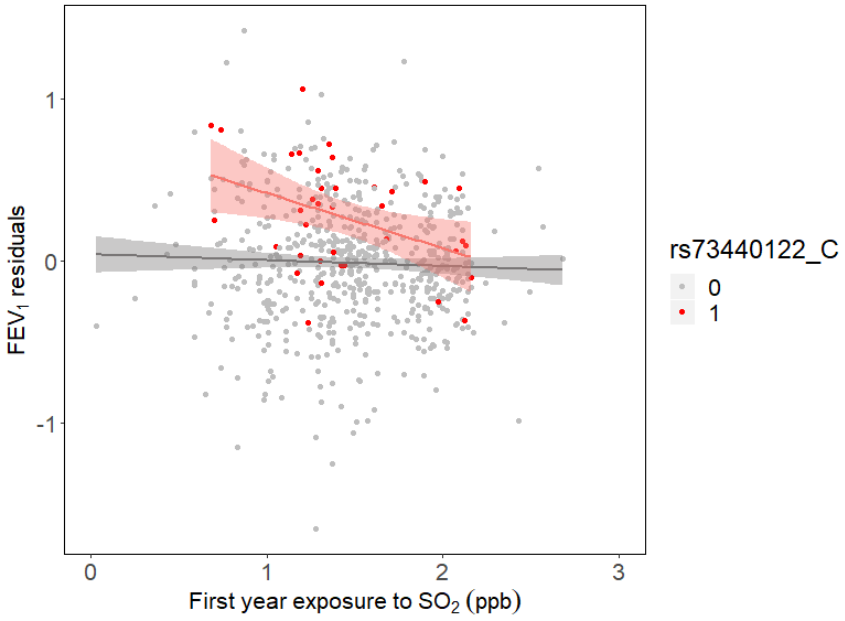
1018 Figure 3.

1019 A



1020

1021 B



1022

1023 Figure 3. Gene-by-environment interaction analysis on FEV_1. FEV_1 residuals, residuals after FEV_1

1024 was regressed on the covariates age, sex, height, controller medications, sequencing centers and

1025 the first 5 genetic PCs. FEV_1 residuals was plotted against (A) past year exposure to SO_2 stratified

1026 by the number of copies of T allele of rs58475486, (B) first year of life exposure to SO_2 stratified

1027 by the number of copies of C allele of rs73440122.