2022

Pulmonary Function and Blood DNA Methylation: A Multiancestry Epigenome-Wide Association Meta-Analysis

Mikyeong Lee

Tianxiao Huan

Daniel L. McCartney

Geetha Chittoor

Maaike de Vries

See next page for additional authors

Original Publication Citation

Lee, M., Huan, T., McCartney, D. L., Chittoor, G., de Vries, M., Lahousse, L., Nguyen, J. N., Brody, J. A., Castillo-Fernandez, J., Terzikhan, N., Qi, C., Joehanes, R., Min, J. L., Smilnak, G. J., Shaw, J. R., Yang, C. X., Colicino, E., Hoang, T. T., Bermingham, M. L., . . . London, S. J. (2022). Pulmonary function and blood DNA methylation: A multiancestry epigenome-wide association meta-analysis. *American Journal of Respiratory and Critical Care Medicine*, *206*(3), 321-336. https://doi.org/10.1164/rccm.202108-1907OC

ORIGINAL ARTICLE

Pulmonary Function and Blood DNA Methylation
A Multiancestry Epigenome-Wide Association Meta-analysis

Mikyeong Lee[1], Tianxiao Huan[2,3], Daniel L. McCartney[4], Geetha Chittoor[5,6], Maaike de Vries[7,8], Lies Lahousse[9,10], Jennifer N. Nguyen[11], Jennifer A. Brody[12], Juan Castillo-Fernandez[13], Natalie Terzikhan[10], Cancan Qi[8,14], Roby Joehanes[3], Josine L. Min[15,16], Gordon J. Smilnak[1], Jessica R. Shaw[17], Chen Xi Yang[18], Elena Colicino[19], Thanh T. Hoang[1], Mairead L. Bermingham[4], Hanfei Xu[20], Anne E. Justice[5,6], Cheng-Jian Xu[21,22,23,24], Stephen S. Rich[11], Simon R. Cox[25], Judith M. Vonk[7,8], Ivana Prokić[10], Nona Sotoodehnia[26], Pei-Chien Tsai[13,27,28], Joel D. Schwartz[29,30,31], Janice M. Leung[18], Sinjini Sikdar[1,32], Rosie M. Walker[4], Sarah E. Harris[25], Diana A. van der Plaat[7,33], David J. Van Den Berg[34,35], Traci M. Bartz[36], Tim D. Spector[13], Pantel S. Vokonas[37], Riccardo E. Marioni[4], Adele M. Taylor[25], Yongmei Liu[38], R. Graham Barr[39,40], Leslie A. Lange[17,41], Andrea A. Baccarelli[42], Ma'en Obeidat[43], Myriam Fornage[44,45], Tianyuan Wang[46], James M. Ward[46], Alison A. Motsinger-Reif[47], Gibran Hemani[15,16], Gerard H. Koppelman[8,14], Jordana T. Bell[13], Sina A. Gharib[12,48], Guy Brusselle[10,49,50], H. Marike Boezen[7,8,†], Kari E. North[6], Daniel Levy[3], Kathryn L. Evans[4], Josée Dupuis[20], Charles E. Breeze[51], Ani Manichaikul[11], Stephanie J. London[1]; and the Biobank-based Integrative Omics Study (BIOS) Consortium, Genetics of DNA Methylation Consortium (GoDMC)

Abstract

Rationale: Methylation integrates factors present at birth and modifiable across the lifespan that can influence pulmonary function. Studies are limited in scope and replication.

Objectives: To conduct large-scale epigenome-wide meta-analyses of blood DNA methylation and pulmonary function.

Methods: Twelve cohorts analyzed associations of methylation at cytosine-phosphate-guanine probes (CpGs), using Illumina 450K or EPIC/850K arrays, with FEV_1, FVC, and FEV_1/FVC. We performed multiancestry epigenome-wide meta-analyses (total of 17,503 individuals; 14,761 European, 2,549 African, and 193 Hispanic/Latino ancestries) and interpreted results using integrative epigenomics.

Measurements and Main Results: We identified 1,267 CpGs (1,042 genes) differentially methylated (false discovery rate, <0.025) in relation to FEV_1, FVC, or FEV_1/FVC, including 1,240 novel and 73 also related to chronic obstructive pulmonary disease (1,787 cases). We found 294 CpGs unique to European or African ancestry and 395 CpGs unique to never or ever smokers. The majority of significant CpGs correlated with nearby gene expression in blood. Findings were enriched in key regulatory elements for gene function, including accessible chromatin elements, in both blood and lung. Sixty-nine implicated genes are targets of investigational or approved drugs. One example novel gene highlighted by integrative epigenomic and druggable target analysis is *TNFRSF4*. Mendelian randomization and colocalization analyses suggest that epigenome-wide association study signals capture causal regulatory genomic loci.

Conclusions: We identified numerous novel loci differentially methylated in relation to pulmonary function; few were detected in large genome-wide association studies. Integrative analyses highlight functional relevance and potential therapeutic targets. This comprehensive discovery of potentially modifiable, novel lung function loci expands knowledge gained from genetic studies, providing insights into lung pathogenesis.

Keywords: spirometry; epigenetics; respiratory function tests; chronic obstructive pulmonary disease

Pulmonary function traits, including FEV_1, FVC, and their ratio (FEV_1/FVC), assess the physiologic state of the lungs and provide the basis for diagnosing chronic obstructive pulmonary disease (COPD). They predict morbidity and mortality in the general population after accounting for other risk factors, even within the normal range (1, 2). The mechanisms for these associations remain largely unknown.

Adult pulmonary function reflects environment and genetics. Various exposures, most notably cigarette smoking, reduce lung function. Large-scale genome-wide association studies (GWASs)

(Received in original form August 18, 2021; accepted in final form May 9, 2022)

†Deceased.

Am J Respir Crit Care Med Vol 206, Iss 3, pp 321–336, Aug 1, 2022

Originally Published in Press as DOI: 10.1164/rccm.202108-1907OC on May 10, 2022
Internet address: www.atsjournals.org

[1]Epidemiology Branch, [46]Integrative Bioinformatics Support Group, and [47]Biostatistics and Computational Biology Branch, National Institute of Environmental Health Sciences, National Institutes of Health, U.S. Department of Health and Human Services, Research Triangle Park, North Carolina; [2]Department of Ophthalmology, University of Massachusetts Medical School, Worcester, Massachusetts; [3]Framingham Heart Study, National Heart, Lung, and Blood Institute, National Institutes of Health, U.S. Department of Health and Human Services, Framingham, Massachusetts; [4]Centre for Genomic and Experimental Medicine, Institute of Genetics and Cancer and [25]Lothian Birth Cohorts Group, Department of Psychology, The University of Edinburgh, Edinburgh, United Kingdom; [5]Department of Population Health Sciences, Geisinger, Danville, Pennsylvania; [6]Department of Epidemiology, School of Public Health, University of North Carolina, Chapel Hill, North Carolina; [7]Department of Epidemiology, [8]Groningen Research Institute for Asthma and COPD, and [14]Department of Pediatric Pulmonology and Pediatric Allergology, Beatrix Children's Hospital, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands; [9]Department of Bioanalysis, Ghent University, Ghent, Belgium; [10]Department of Epidemiology and [49]Department of Respiratory Medicine, Erasmus Medical Center Rotterdam, Rotterdam, The Netherlands; [11]Center for Public Health Genomics, University of Virginia, Charlottesville, Virginia; [12]Cardiovascular Health Research Unit, Division of Cardiology, Department of Medicine, [26]Cardiovascular Health Research Unit, Division of Cardiology, Department of Epidemiology, [36]Cardiovascular Health Research Unit, Department of Biostatistics, [48]Computational Medicine Core, Center for Lung Biology, Division of Pulmonary, Critical Care and Sleep Medicine, University of Washington, Seattle, Washington; [13]Department of Twin Research and Genetic Epidemiology, King's College London, London, United Kingdom; [15]Medical Research Council Integrative Epidemiology Unit and [16]Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, United Kingdom; [17]Division of Biomedical Informatics and Personalized Medicine, Department of Medicine, and [41]Department of Epidemiology, University of Colorado, Aurora, Colorado; [18]Centre for Heart Lung Innovation, Department of Medicine, University of British Columbia, Vancouver, British Columbia, Canada; [19]Department of Environmental Medicine and Public Health, Icahn School of Medicine at Mount Sinai, New York, New York; [20]Department of Biostatistics, Boston University School of Public Health, Boston, Massachusetts; [21]Centre for Individualised Infection Medicine, a joint venture between Hannover Medical School and the Helmholtz Centre for Infection Research, Hannover, Germany; [22]Research Group Bioinformatics and Computational Genomics, TWINCORE, Centre for Experimental and Clinical Infection Research, a joint venture between the Helmholtz Centre for Infection Research and the Hannover Medical School, Hannover, Germany; [23]Department of Gastroenterology, Hepatology and Endocrinology, Hannover Medical School, Hannover, Germany; [24]Department of Internal Medicine, Radboud Center for Infectious Diseases, Radboud University Medical Center, Nijmegen, The Netherlands; [27]Department of Biomedical Sciences, Chang Gung University, Taoyuan, Taiwan; [28]Genomic Medicine Core Laboratory, Chang Gung Memorial Hospital, Linkou, Taiwan; [29]Department of Environmental Health and [30]Department of Epidemiology, Harvard School of Public Health, Boston, Massachusetts; [31]Channing Laboratory, Harvard Medical School, Boston, Massachusetts; [32]Department of Mathematics and Statistics, Old Dominion University, Norfolk, Virginia; [33]National Heart and Lung Institute, Imperial College London, London, United Kingdom; [34]Department of Preventive Medicine and [35]Center for Genetic Epidemiology, Keck School of Medicine, University of Southern California, Los Angeles, California; [37]Veterans Affairs Boston Healthcare System, School of Medicine and School of Public Health, Boston University, Boston, Massachusetts; [38]Division of Cardiology, Department of Medicine, Duke University, Durham, North Carolina; [39]Department of Medicine and [40]Department of Epidemiology, Columbia University Medical Center, New York, New York; [42]Department of Environmental Health Sciences, Mailman School of Public Health, Columbia University, New York, New York; [43]Centre for Heart Lung Innovation, The University of British Columbia, St. Paul's Hospital, Vancouver, British Columbia, Canada; [44]Brown Foundation Institute of Molecular Medicine, McGovern Medical School, and [45]Human Genetics Center, School of Public Health, University of Texas Health Science Center, Houston, Texas; [50]Department of Respiratory Medicine, Ghent University Hospital, Ghent, Belgium; and [51]Altius Institute for Biomedical Sciences, Seattle, Washington

ORCID IDs: 0000-0002-3036-3684 (M.L.); 0000-0003-3242-0360 (D.L.M.C.); 0000-0001-6443-6012 (G.C.); 0000-0001-7210-8174 (M.d.V.); 0000-0002-3494-4363 (L.L.); 0000-0001-8509-148X (J.A.B.); 0000-0002-0034-8029 (J.C.-F.); 0000-0001-5549-9054 (R.J.); 0000-0003-4456-9824 (J.L.M.); 0000-0001-7519-2180 (J.R.S.); 0000-0002-8033-4769 (C.X.Y.); 0000-0002-8498-0020 (T.T.H.); 0000-0002-0931-0051 (M.L.B.); 0000-0003-1586-4672 (C.-J.X.); 0000-0003-3872-7793 (S.S.R.); 0000-0003-4036-3642 (S.R.C.); 0000-0002-0370-1473 (I.P.); 0000-0003-1575-9020 (P.-C.T.); 0000-0001-6168-378X (J.D.S.); 0000-0003-1230-5162 (S.S.); 0000-0002-4941-5106 (S.E.H.); 0000-0002-0361-357X (D.A.v.d.P.); 0000-0002-5404-5218 (D.J.V.D.B.); 0000-0003-4430-4260 (R.E.M.); 0000-0002-6562-1858 (Y.L.); 0000-0002-4220-6766 (L.A.L.); 0000-0002-5443-2752 (M.O.); 0000-0002-3970-0771 (T.W.); 0000-0003-1346-2493 (A.A.M.-R.); 0000-0003-0920-1055 (G.H.); 0000-0001-8567-3252 (G.H.K.); 0000-0002-3858-5986 (J.T.B.); 0000-0002-2480-4367 (S.A.G.); 0000-0002-8903-0366 (K.E.N.); 0000-0002-7884-5877 (K.L.E.); 0000-0003-2871-3603 (J.D.); 0000-0002-5294-915X (C.E.B.); 0000-0002-5998-795X (A.M.); 0000-0003-4911-5290 (S.J.L.).

Supported in part by the Intramural Research Program of the National Institutes of Health, National Institute of Environmental Health Sciences (Z01-ES043012). Infrastructure for the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium is supported in part by the National Heart, Lung, and Blood Institute (HL105756). Additional study-specific funding statements can be found in the online supplement.

Author Contributions: M.L., L.L., J.N.N., S.S., and S.J.L. conceived and designed the study. M.L., T.H., D.L.M., G.C., A.E.J., M.d.V., L.L., J.N.N., J.A.B., J.C.-F., N.T., J.R.S., C.X.Y., E.C., and A.M. conducted study-specific analyses. M.L., C.Q., R.J., C.E.B., T.T.H., T.W., J.M.W., A.A.M.-R., and G.J.S. performed and/or supervised follow-up analyses. J.M.L., E.C., M.L.B., H.X., C.-J.X., S.S.R., S.R.C., J.M.V., I.P., N.S., P.-C.T., J.D.S., R.M.W., S.E.H., D.A.v.d.P., D.J.V.D.B., T.M.B., T.D.S., P.S.V., R.E.M., A.M.T., Y.L., R.G.B., L.A.L., A.A.B., M.O., M.F., G.H.K., J.T.B., S.A.G., G.B., H.M.B., K.E.N., D.L., K.L.E., J.D., A.M., and S.J.L. contributed to study design and/or supervised data analysis in each cohort. M.L. performed meta-analyses. T.H. confirmed the meta-analysis results. M.L., with input from T.H., D.L.M., R.J., L.L., J.D., A.M., and S.J.L., wrote the first draft of the manuscript. All authors contributed to interpretation of the results and/or critical revision of the manuscript. All authors approved the final version.

Correspondence and requests for reprints should be addressed to Stephanie J. London, M.D., Dr.P.H., National Institute of Environmental Health Sciences, P.O. Box 12233, MD A3-05, Research Triangle Park, NC 27709. E-mail: london2@niehs.nih.gov.

This article has a related editorial.

This article has an online supplement, which is accessible from this issue's table of contents at www.atsjournals.org.

At a Glance Commentary

Scientific Knowledge on the Subject: DNA methylation can influence pulmonary function. Data on blood DNA methylation and pulmonary function are relatively few with minimal replication.

What This Study Adds to the Field: This large-scale multiancestry study of epigenome-wide DNA methylation and pulmonary function identified many novel loci, mostly not discovered by genetic studies. Various integrative analyses enhance the functional and clinical relevance of our findings, including potential therapeutic targets.

have implicated more than 300 loci (3, 4); much of the variability remains unexplained. Epigenetic DNA modifications reflect genetics and exposures over the life course and can identify genes influencing pulmonary function. Methylation is the most studied epigenetic modification due to high-throughput, reproducible platforms with reasonable genome-wide coverage. Several epigenome-wide association studies (EWASs) using Illumina 27K (5), 450K (6–8), and EPIC/850K (9) platforms have identified pulmonary function–related cytosine-phosphate-guanine probes (CpGs). However, replication has been limited. Most were studies in European ancestry populations; no large-scale multiancestry study has been published.

We performed a meta-analysis of coordinated EWAS results from 16 separate analyses from 12 cohorts (17,503 individuals, including 14,761 European, 2,549 African, and 193 Hispanic/Latino ancestries) to identify CpGs differentially methylated in relation to pulmonary function. To provide insight into functional impacts, we evaluated associations between identified CpGs and nearby gene expression in paired blood DNA methylation and total blood RNA transcriptome data. Using integrative epigenomic methods, we assessed enrichment of regulatory elements in our blood findings across tissue types, including lung. Mendelian randomization (MR) and colocalization analyses were used to provide further functional interpretation. For clinical implications, we explored whether implicated genes were targets of drugs approved or under investigation. This large-scale multiancestry study of blood DNA methylation and pulmonary function identified numerous loci not found in GWASs, increasing our understanding of mechanisms regulating pulmonary function. Some preliminary results of this study were previously reported in the form of an abstract (10).

Methods

Further details regarding the methods are provided in the online supplement. The population included 17,503 adults (≥40 yr old) from ALHS (Agricultural Lung Health Study), ARIC (Atherosclerosis Risk in Communities Study), CHS (Cardiovascular Health Study), FHS (Framingham Heart Study), GS (Generation Scotland), LifeLines, LBC (Lothian Birth Cohort), MESA (Multi-Ethnic Study of Atherosclerosis), RS (Rotterdam Study), and TwinsUK within the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium framework. Trained staff in each study measured prebronchodilator pulmonary function (FEV_1 and FVC in ml) (11). Methylation was assessed in blood drawn at the same visit at spirometry, except for CHS, where the difference was 1 year. Ten studies used the Illumina HumanMethylation450 (~480,000 CpGs), and six used the newer MethylationEPIC (~850,000 CpGs). Each cohort retained autosomal CpGs after preprocessing and filtering.

Association Analyses

Each study assessed effects of methylation on spirometric traits after adjusting for age, age squared, sex, height, height squared, weight (for FVC only), smoking (never, former, or current), pack-years, and estimated cell type proportions (12) using robust linear regression to help account for potential heteroskedasticity and influential outliers. In addition, studies adjusted for analytic batch, ancestry principal components (calculated from genome-wide genotypes), study site, and selection factor or accounted for family structure when appropriate. Cohorts with more than one ancestry group performed analyses separately by ancestry. We combined study-specific results using inverse variance–weighted fixed-effects models (13, 14). We conducted two separate EWAS meta-analyses (450K and EPIC-unique), setting a genome-wide significance threshold of false discovery rate (FDR) <0.025 (0.05/2 meta-analyses) (15). Unless otherwise noted, genome-wide significant CpGs refer to those with FDR <0.025 from either meta-analysis. We examined differentially methylated regions using DMRcate (16).

CpG Annotation and Filtering

We used Illumina (17, 18), Zhou and colleagues (19), and Homer version 4.9.1 (20) (genome build GRCh37/hg19) for CpG annotation. From CpGs meeting genome-wide significance in meta-analyses, we removed those previously reported as potentially problematic (19). Using leave-one-out meta-analyses, we identified and removed CpGs with associations driven by a single study. Specifically, we did not consider a CpG significant genome-wide when the association did not meet at least nominal significance ($P < 0.05$) in meta-analysis, with consistent direction, after leaving out one study. Remaining CpGs were subjected to downstream analyses.

Functional Downstream Analyses

We tested for enrichment of genomic features (CpG islands, shores, shelves, promoters, and transcription factor [TF] binding sites). Using the eFORGE integrative epigenomics approach (21–23), we explored whether our lung function–associated CpGs were enriched in regulatory elements from the Roadmap Epigenomics Consortium (24) across more than 20 tissue types. To gain further biological insights, we conducted pathway analyses (25, 26).

Cis-expression quantitative trait methylation analysis

We assessed whether our significant CpGs associate with transcription of nearby genes using paired whole-blood 450K methylation and transcriptome data from FHS (27) and the BIOS (Biobank-based Integrative Omics Study) Consortium (28).

Methylation quantitative trait loci, MR, and colocalization

To examine whether our significant CpGs were methylation quantitative trait loci (mQTLs), we used the Genetics of DNA

Table 1. Characteristics of Participating Studies ($N = 17,503$ Participants)

Study	Country	Methylation array	Number of Participants by Ancestry		
			European	African	Hispanic/Latino
ALHS	United States	EPIC	2,268		
ARIC	United States	450K	787	2,261	
CHS	United States	450K	218	181	
FHS	United States	450K	3,205		
GS Set 1	United Kingdom	EPIC	1,700		
GS Set 2	United Kingdom	EPIC	2,954		
LBC1921	United Kingdom	450K	435		
LBC1936	United Kingdom	450K	905		
LifeLines	The Netherlands	450K	1,155		
MESA	United States	EPIC	246	107	193
RS	The Netherlands	450K	716		
TwinsUK	United Kingdom	450K	172		
Total			14,761	2,549	193

Definition of abbreviations: ALHS = Agricultural Lung Health Study; ARIC = Atherosclerosis Risk in Communities Study; CHS = Cardiovascular Health Study; FHS = Framingham Heart Study; GS = Generation Scotland; LBC = Lothian Birth Cohort; MESA = Multi-Ethnic Study of Atherosclerosis; RS = Rotterdam Study.

Methylation Consortium (GoDMC) database (29). To investigate causality, we performed MR (30) and colocalization (31) analyses.

Druggable targets

To explore clinical relevance, we searched for approved or experimental drugs targeting genes implicated in our meta-analyses using the ChEMBL database (32).

COPD

We examined whether CpGs significantly related to pulmonary function were associated in our data with COPD, defined using prebronchodilator spirometry: FEV_1 $<80\%$ predicted (33) and $FEV_1/FVC < 0.7$. Noncases had $FEV_1 \geq 80\%$ predicted and $FEV_1/FVC \geq 0.7$ (34, 35).

Replication and Validation

To replicate our findings, we looked up our significant CpGs in recent EWASs of lung function (6, 8). To validate previous findings (5–9), we examined CpGs reported as related to pulmonary function, after multiple-testing correction and smoking adjustment, in our meta-analyses. We looked up CpGs related to COPD in lung tissue (36) in our lung function meta-analyses.

Additional Analyses

To assess whether methylation findings were driven by genetic variants for lung function, we examined our significant findings after additional adjustment for polygenic risk scores (4, 37) in ALHS. To address possible residual confounding by smoking, we evaluated whether our significant CpGs overlapped with CpGs related to current smoking (38). We also adjusted for cg05575921 (*AHRR*), a biomarker of lifetime smoking (39), and two additional biomarkers: cg13039251 (*PDZD2*) and cg03636183 (*F2RL3*) (40, 41). We conducted meta-analyses separately for never smokers ($n = 8,830$) and ever smokers ($n = 8,673$) to evaluate whether pulmonary function–related methylation differs by smoking. In separate meta-analyses by European ($n = 14,761$) and African ancestry ($n = 2,549$), we considered consistency across ancestries and explored ancestry-specific signals.

Results

We performed a meta-analysis of data from 17,503 participants (16 separate analyses from 12 cohorts; 12 European ancestry, 3 African ancestry, and 1 Hispanic/Latino ancestry) (Table 1; *see* study-level characteristics in Tables E1–E3 in the online supplement). We included up to 865,971 CpGs analyzed in at least three studies: 473,215 (93% also on EPIC) in the 450K EWAS meta-analysis and 392,756 in the EPIC-unique meta-analysis (workflow in Figure 1).

Pulmonary Function–related CpGs

We identified 1,267 CpGs (1,042 genes) significantly differentially methylated (FDR, <0.025) in relation to pulmonary function (Table E4), including 164 from EPIC-unique meta-analyses and 85% associated with only one trait (Table 2, Figure E1). Of the 1,042 implicated genes, 24% contained multiple genome-wide significant CpGs. Tables 3–5 display the top 30 CpGs for each trait. Of 1,451 genome-wide significant associations (FDR, <0.025) with any of the three traits (Tables E5–E7), after removing 70 driven by a single study, 165 met Bonferroni correction (Table E8). We provide graphic representation of EWAS meta-analysis results: Miami (Figures 2 and E2) and QQ plots (λ values 1–1.3, supporting minimal inflation; Figure E3). For significant CpGs, we plotted leave-one-out meta-analysis results (Figure E4), study-specific results (forest plots in Figure E5), and distributions (Figure E6). Using DMRcate (16), we identified 2,806 differentially methylated regions associated (FDR, <0.01) with FEV_1, FVC, or FEV_1/FVC (Table E9); $\sim$25% contained a genome-wide significant CpGs.

Functional Impacts

Notably, in our significant findings, TF binding sites and promoter regions (for FEV_1, the trait with the largest number of findings) were enriched (Table E10), supporting potential impacts on transcription. Integrative epigenomic analyses (eFORGE) for FEV_1 highlighted enrichment of DNase I hotspots in blood and lung (Figure 3A). Enrichment in blood and fetal lung were distinct signals (Figures 3B and 3C). FEV_1-associated genetic variants also show enrichment for lung DNase I hotspots (Figure 3D) for a different set of loci when compared with FEV_1-associated CpGs (42). Because DNase I hotspots represent broad regions of accessible chromatin containing various regulatory elements, these analyses highlight functional implications.

Pathway Enrichment

Pathways relevant to pulmonary function were enriched (FDR, <0.05) (Figure 4). Several pathways overlapped across the traits, including Wnt signaling, a key developmental pathway involved in lung pathogenesis, and inflammatory pathways such as cytokine–cytokine receptor interaction.

Correlation with Expression

Linking our significant CpGs on the 450K to paired blood gene expression and 450K methylation data (27, 28), 97% (of 1,103 CpGs available) had at least one transcript within ± 250 kb. At FDR <0.05, 56% were related to gene expression (Table E11), and

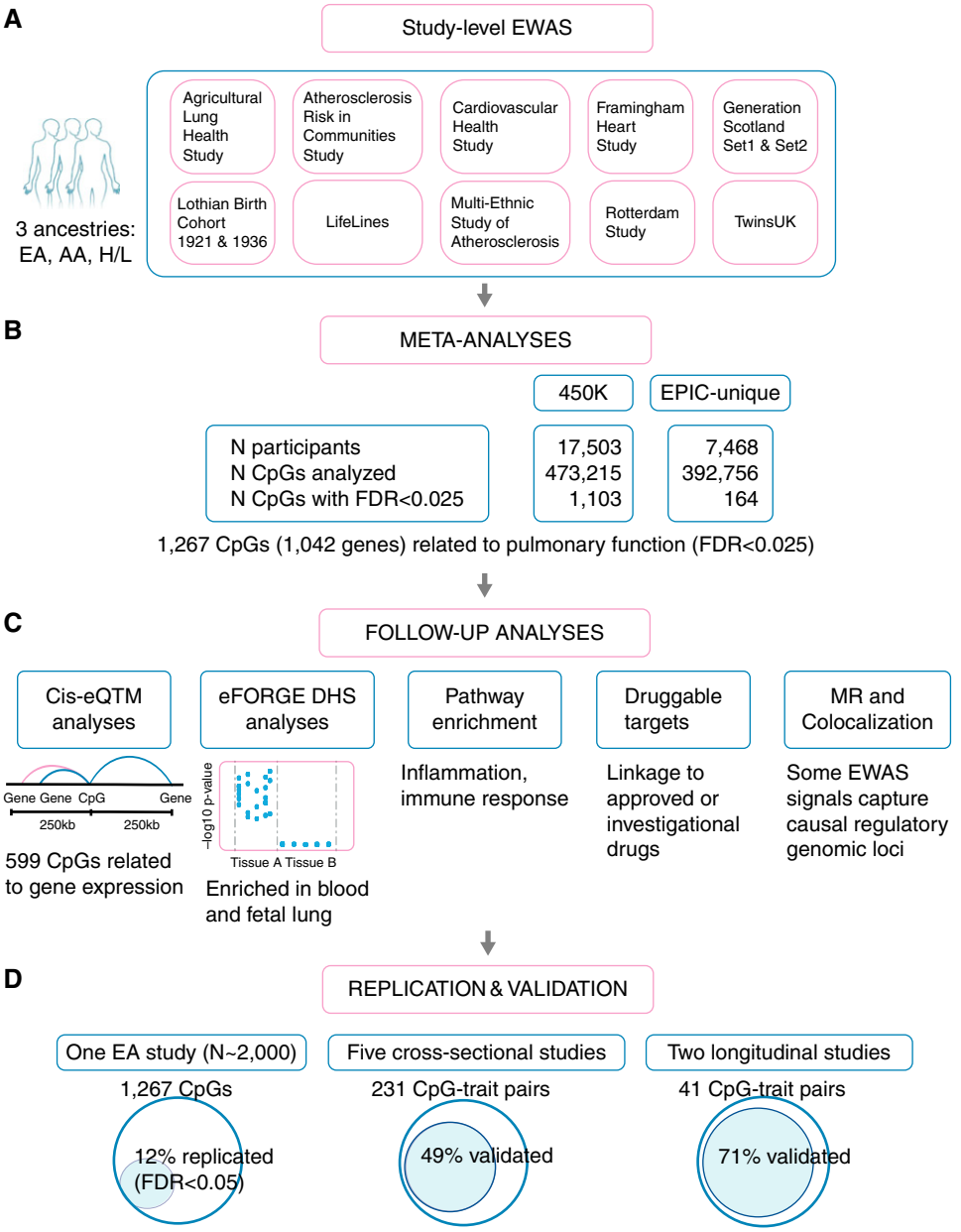

Figure 1. Overview of our epigenome-wide association study (EWAS) meta-analyses on pulmonary function. (*A*) Each study examined associations between DNA methylation and pulmonary function. Participating studies were ALHS (Agricultural Lung Health Study), ARIC (Atherosclerosis Risk in Communities), CHS (Cardiovascular Health Study), FHS (Framingham Heart Study), GS (Generation Scotland), LifeLines, LBC (Lothian Birth Cohort), MESA (Multi-Ethnic Study of Atherosclerosis), RS (Rotterdam Study), and TwinsUK. Our EWAS meta-analyses included datasets from three ancestries: European ancestry (EA), African ancestry (AA), and Hispanic/Latino ancestry (H/L). (*B*) Two separate meta-analyses were conducted: 450K EWAS meta-analysis (17,503 individuals) and EPIC-unique EWAS meta-analysis (7,468 individuals). (*C*) Functional follow-up analyses included *cis*-expression quantitative trait methylation (eQTM) analyses using paired blood methylation and transcriptome data from FHS and the BIOS (Biobank-based Integrative Omics Study) Consortium, eFORGE DNase I hypersensitive site (DHS) analysis using Roadmap Epigenomics data, pathway analyses using gene sets from Kyoto Encyclopedia of Genes and Genomes, druggable targets analysis, and Mendelian randomization (MR) and colocalization analyses. (*D*) We replicated 12% of our significant findings in an EWAS of pulmonary function (N≅2,000 EA participants) (false discovery rate [FDR], <0.05); an additional 49% were significant at $P<0.05$. Previously reported cytosine-phosphate-guanine probes (CpGs) were validated: 113 (49%) of 231 CpG–trait pairs from five cross-sectional studies and 29 (71%) of 41 CpG–trait pairs from the two longitudinal studies.

at $P < 0.05$, 75% were related to gene expression ($P_{\text{enrichment}} < 2.2 \times 10^{-16}$), supporting functional impacts on gene regulation.

mQTL, colocalization, and MR analyses

Using GoDMC (29), 798 of our 1,103 significant CpGs had at least one mQTL. The mQTLs for our EWAS findings were associated with lung function more than expected by chance: mQTLs for the FEV_1-related CpGs associated more strongly with

Table 2. Number of Cytosine-Phosphate-Guanine Probes Differentially Methylated (False Discovery Rate <0.025) in Relation to FEV_1, FVC, and FEV_1/FVC in Multiancestry Meta-Analyses

Trait	450K and EPIC-Unique Meta-Analyses (*N* = 17,503)	450K Meta-Analyses (*N* = 17,503)	EPIC-Unique Meta-Analyses (*n* = 7,468)
FEV_1 only	935	783	152
FVC only	47	46	1
FEV_1/FVC only	101	98	3
Both FEV_1 and FVC	168	160	8
Both FEV_1 and FEV_1/FVC	15	15	0
Both FVC and FEV_1/FVC	1	1	0
Total	1,267	1,103	164

FEV_1 than other mQTLs ($P_{enrichment} = 7.1 \times 10^{-7}$; controlling for linkage disequilibrium, minor allele frequency, and removing MHC region) (Table E12). Observed enrichment could reflect EWAS findings capturing relevant regulatory regions for the traits or potentially causal effects of the CpGs. Using MR to investigate causality, we found 78 significant associations (FDR, <0.05) (Figure E7, Table E13). Because lung function is polygenic, these associations might reflect distinct causal variants for the outcomes in linkage disequilibrium with the mQTLs.

Table 3. Top 30 Cytosine-Phosphate-Guanine Probes Differentially Methylated (False Discovery Rate <0.025) in Relation to FEV_1 in Multiancestry Meta-Analysis of Content on the 450K Array (16 Separate Analyses from 12 Cohorts; 17,503 Individuals) or Unique to EPIC (6 Studies; 7,468 Individuals), Sorted by Chromosomal Position

Chromosomal Position*	CpG Probe	Regression Coefficient	SE	*P* Value	Mean Methylation†	Gene Name‡	Meta-Analysis
1:2161049	cg05603985	7.828	1.168	2.1×10^{-11}	0.252	*SKI*	450K
1:55353706	cg17901584	4.256	0.614	4.1×10^{-12}	0.469	*DHCR24*	450K
1:109757585	cg03725309	7.625	1.062	7.0×10^{-13}	0.138	*SARS*	450K
1:120255992	cg14476101	3.533	0.538	5.2×10^{-11}	0.608	*PHGDH*	450K
1:145441552	cg19693031	4.441	0.661	1.8×10^{-11}	0.729	*TXNIP*	450K
2:65225988	cg23831876	7.772	1.169	3.0×10^{-11}	0.833	*SLC1A4*	450K
2:233284661	cg21566642	3.786	0.552	6.8×10^{-12}	0.500	*ECEL1P1*§	450K
3:101901234	cg12992827	5.112	0.793	1.2×10^{-10}	0.702	*ZPLD1*§	450K
3:185538892	cg24960291	4.853	0.727	2.4×10^{-11}	0.597	*IGF2BP2*	450K
4:57947735	cg15696506	4.127	0.610	1.4×10^{-11}	0.505	*IGFBP7*	450K
4:139162808	cg06690548	6.039	0.694	3.3×10^{-18}	0.849	*SLC7A11*	450K
5:373378	cg05575921	5.051	0.581	3.6×10^{-18}	0.815	*AHRR*	450K
5:159428643	cg18394552	−4.348	0.657	3.7×10^{-11}	0.597	*TTC1*‖	450K
6:166970252	cg17501210	4.626	0.687	1.7×10^{-11}	0.704	*RPS6KA2*	450K
8:121597619	cg01198738	6.235	0.908	6.5×10^{-12}	0.478	*SNTB1*	EPIC-unique
11:68607622	cg00574958	16.308	2.023	7.6×10^{-16}	0.069	*CPT1A*	450K
11:68607737	cg17058475	9.571	1.307	2.5×10^{-13}	0.081	*CPT1A*	450K
11:102189303	cg19120513	9.894	1.390	1.1×10^{-12}	0.516	*BIRC3*	EPIC-unique
12:11898284	cg07986378	4.130	0.614	1.8×10^{-11}	0.605	*ETV6*	450K
12:104853274	cg06647068	4.751	0.677	2.3×10^{-12}	0.285	*CHST11*	450K
13:79968324	cg16969872	5.421	0.752	5.8×10^{-13}	0.703	*RBM26*	450K
14:74227441	cg10919522	5.497	0.822	2.3×10^{-11}	0.216	*C14orf43*	450K
16:75079000	cg08761535	8.906	1.358	5.4×10^{-11}	0.793	*ZNRF1*	450K
17:76354621	cg18181703	6.180	0.704	1.7×10^{-18}	0.454	*SOCS3*	450K
17:76354934	cg11047325	4.617	0.706	6.3×10^{-11}	0.570	*SOCS3*	EPIC-unique
19:45252955	cg26470501	6.869	0.861	1.4×10^{-15}	0.504	*BCL3*	450K
19:47287778	cg22304262	5.543	0.829	2.2×10^{-11}	0.735	*SLC1A5*	450K
19:47287964	cg02711608	9.412	1251	5.2×10^{-14}	0.173	*SLC1A5*	450K
22:31686097	cg08548559	4.534	0.681	2.9×10^{-11}	0.210	*PIK3IP1*	450K
22:50327986	cg09349128	6.998	0.931	5.7×10^{-14}	0.280	*CITF22-49E9.3*§	450K

Definition of abbreviation: CpG = cytosine-phosphate-guanine probe.
Top 30 CpGs based on meta-analysis *P* values. Individual study results were obtained from robust linear regression with methylation as the predictor and FEV_1 as the outcome. Covariates included age, sex, height, age squared, height squared, smoking status (never, former, or current), pack-years, and estimated cell type proportions. Study-specific covariates included study center, selection factor, ancestry principal components, batch variables, and family structure when appropriate. Regression coefficients represent milliliter differences in FEV_1 per 1% difference in methylation. Table E5 contains complete results (false discovery rate, <0.025).
*Genome build GRCh37/hg19.
†Weighted average methylation across participating studies at the specified CpG.
‡Gene names from the Illumina annotation (17, 18), Zhou and colleagues (19), or Homer version 4.9.1 (20).
§Gene names from Zhou and colleagues (19).
‖Gene names (within ±2 Mb) from Homer version 4.9.1 (20).

When we tested this using genetic colocalization at the mQTLs exhibiting a strong methylation–outcome association, 46 of 78 MR associations showed evidence for colocalization (posterior probability >0.8 for the hypothesis of one shared genetic variant for methylation and trait) (Figure E8, Table E13).

Druggable targets

Among the 1,042 implicated genes, 261 had bioactive molecules with druglike properties in ChEMBL, including 69 with at least one approved or candidate drug for respiratory or other conditions. Sixty-one genes had not been identified by GWAS of pulmonary function (Table E14).

COPD associations

Of the 1,267 CpGs differentially methylated in relation to pulmonary function, 73 associated with COPD in our data (1,787 cases and 11,824 noncases) at FDR <0.05, and 323 met nominal significance ($P_{enrichment} < 2.2 \times 10^{-16}$) (Table E15). Directions matched expectation: CpGs positively associated with pulmonary function were negatively associated with COPD and vice versa.

Replication

In a published EWAS of lung function in ~2,000 European ancestry participants (8), 12% of our significant CpGs were associated with any of the three traits (FDR, <0.05), and an additional 49% were nominally significant ($P_{enrichment}$ for both $< 2.2 \times 10^{-16}$) (Table E16). In a small Korean study (6), 14% were related to any of the three traits ($P < 0.05$; $P_{enrichment} < 2.2 \times 10^{-16}$) (Table E17).

Validating Published CpGs

Five cross-sectional studies (5–9) reported 134 CpGs associated with at least one spirometric trait at genome-wide significance. Of 231 CpG–trait associations available, 36% met FDR <0.05 and had directions of association consistent with previous reports (Table E18). At nominal $P < 0.05$, an additional 59 CpGs were associated with any traits. Overall, we validated 70% of previously reported CpGs in our results ($P_{enrichment} < 2.2 \times 10^{-16}$).

Table 4. Top 30 Cytosine-Phosphate-Guanine Probes Differentially Methylated (False Discovery Rate <0.025) in Relation to FVC in Multiancestry Meta-Analysis of Content on the 450K Array (16 Separate Analyses from 12 Cohorts; 17,503 Individuals) or Unique to EPIC (6 Studies; 7,468 Individuals), Sorted by Chromosomal Position

Chromosomal Position*	CpG Probe	Regression Coefficient	SE	*P* Value	Mean Methylation†	Gene Name‡	Meta-Analysis
1:120255992	cg14476101	4.021	0.625	1.3×10^{-10}	0.608	*PHGDH*	450K
1:145441552	cg19693031	5.244	0.784	2.3×10^{-11}	0.729	*TXNIP*	450K
4:139162808	cg06690548	6.804	0.838	4.6×10^{-16}	0.849	*SLC7A11*	450K
6:36326677	cg03149958	5.440	0.953	1.1×10^{-8}	0.783	*ETV7*§	450K
6:166970252	cg17501210	5.355	0.805	2.9×10^{-11}	0.704	*RPS6KA2*	450K
7:71800412	cg00277397	5.943	1.056	1.8×10^{-8}	0.739	*CALN1*	450K
8:103937374	cg19589396	5.074	0.854	2.8×10^{-9}	0.688	*KB-1507C5.2;RPL5P24*§	450K
8:121597619	cg01198738	6.141	1.055	5.9×10^{-9}	0.478	*SNTB1*	EPIC-unique
8:134066590	cg17088014	5.510	0.978	1.8×10^{-8}	0.348	*SLA;TG*	450K
9:111885602	cg13661827	4.390	0.781	1.9×10^{-8}	0.429	*TMEM245*‖	450K
11:68607622	cg00574958	19.982	2.483	8.5×10^{-16}	0.069	*CPT1A*	450K
11:68607737	cg17058475	10.018	1.573	1.9×10^{-10}	0.081	*CPT1A*	450K
11:102189303	cg19120513	10.088	1.587	2.1×10^{-10}	0.516	*BIRC3*	EPIC-unique
12:104853274	cg06647068	4.343	0.773	2.0×10^{-8}	0.285	*CHST11*	450K
13:79968324	cg16969872	5.571	0.875	1.9×10^{-10}	0.703	*RBM26*	450K
15:40620444	cg04847110	8.201	1.440	1.2×10^{-8}	0.789	*INAFM2*‖	450K
15:59587546	cg24263283	6.125	1.053	6.0×10^{-9}	0.815	*MYO1E*	450K
15:64290807	cg07037944	8.433	1.502	2.0×10^{-8}	0.199	*DAPK2*	450K
15:91455407	cg11183227	−5.691	0.996	1.1×10^{-8}	0.808	*MAN2A2*	450K
16:3030649	cg02386244	−35.453	6.066	5.1×10^{-9}	0.019	*PKMYT1*	450K
16:30410051	cg00711896	−7.818	1.338	5.2×10^{-9}	0.89	*ZNF48*	450K
16:75079000	cg08761535	9.763	1.613	1.4×10^{-9}	0.793	*ZNRF1*	450K
17:27333185	cg04614997	−17.937	3.132	1.0×10^{-8}	0.03	*SEZ6*	450K
17:76354621	cg18181703	6.510	0.821	2.1×10^{-15}	0.454	*SOCS3*	450K
17:76354934	cg11047325	4.605	0.817	1.7×10^{-8}	0.57	*SOCS3*	EPIC-unique
18:47901430	cg16196758	−10.705	1.831	5.0×10^{-9}	0.016	*SKA1*	450K
19:1423902	cg00994936	−7.423	1.249	2.8×10^{-9}	0.831	*DAZAP1*	450K
19:45252955	cg26470501	6.801	0.998	9.3×10^{-12}	0.504	*BCL3*	450K
19:47287964	cg02711608	9.360	1.455	1.2×10^{-10}	0.173	*SLC1A5*	450K
22:50327986	cg09349128	6.838	1.072	1.8×10^{-10}	0.280	*CITF22-49E9.3*§	450K

For definition of abbreviations, *see* Table 3.

Top 30 CpGs based on meta-analysis *P* values. Individual study results were obtained from robust linear regression with methylation as the predictor and FVC as the outcome. Covariates included age, sex, height, age squared, height squared, weight, smoking status (never, former, or current), pack-years, and estimated cell type proportions. Study-specific covariates included study center, selection factor, ancestry principal components, batch variables, and family structure when appropriate. Regression coefficients represent milliliter differences in FVC per 1% difference in methylation. Table E6 contains complete results (false discovery rate, <0.025).

*Genome build GRCh37/hg19.

†Weighted average methylation across participating studies.

‡Gene names from the Illumina annotation (17, 18), Zhou and colleagues (19), or Homer version 4.9.1 (20).

§Gene names from Zhou and colleagues (19).

‖Gene names (within ±2 Mb) from Homer version 4.9.1 (20).

For the remaining CpGs, there was another CpG annotated to the same gene that showed associations in our data ($P < 0.05$) (Table E19). In addition, of 535 CpGs previously related to COPD in lung (36), 28% were associated with pulmonary function at nominal significance ($P_{enrichment} < 2.2 \times 10^{-16}$) (Table E20). Given the sparse EWAS literature on longitudinal decline in pulmonary function (7, 8), we considered findings from one study that did not adjust for smoking (8). Of 31 CpGs (7, 8) available, 24 were nominally associated with lung function in our cross-sectional data ($P_{enrichment} < 2.2 \times 10^{-16}$) (Table E21), including 10 showing genome-wide significance.

Adjustment for Polygenic Risk Scores

To evaluate whether our methylation findings reflect genetic susceptibility, in ALHS, we also adjusted for a polygenic risk score (4, 37) specific for each trait. Each risk score was significantly related to its trait (Table E22). Of our genome-wide significant signals available in ALHS, 54% had $P < 0.05$, and ~95% remained nominally significant after adjustment for the risk score (Table E23). Regression coefficients were virtually unchanged by the adjustment (Pearson correlation, 0.998), suggesting that the EWAS meta-analysis results provide information complementary to genetics of pulmonary function.

Examination of Potential Residual Confounding by Smoking

Overlap previously reported between CpGs related to pulmonary function and those related to cigarette smoking (8) could reflect residual confounding because smoking-related CpGs are strong biomarkers that better capture lifetime smoking history than questionnaire data

Table 5. Top 30 Cytosine-Phosphate-Guanine Probes Differentially Methylated (False Discovery Rate <0.025) in Relation to FEV_1/FVC in Multiancestry Meta-Analysis of Content on the 450K Array (16 Separate Analyses from 12 Cohorts; 17,503 Individuals) or Unique to EPIC (6 Studies; 7,468 Individuals), Sorted by Chromosomal Position

Chromosomal Position*	CpG Probe	Regression Coefficient	SE	*P* Value	Mean Methylation†	Gene Name‡	Meta-Analysis
1:92947588	cg09935388	0.00045	0.00008	9.9×10^{-9}	0.726	*GFI1*	450K
2:8343710	cg23079012	0.00116	0.00019	1.7×10^{-9}	0.94	*LINC00298;LINC00299*§	450K
2:70008161	cg05155595	0.00075	0.00013	1.6×10^{-8}	0.641	*ANXA4*	450K
2:233284661	cg21566642	0.00058	0.00009	1.2×10^{-11}	0.5	*ECEL1P1*§	450K
3:98251294	cg19859270	0.00171	0.00027	3.5×10^{-10}	0.887	*GPR15*	450K
4:8174148	cg09390241	−0.00059	0.00010	4.0×10^{-9}	0.667	*ABLIM2*‖	450K
4:109038130	cg12623364	0.00086	0.00015	1.3×10^{-8}	0.188	*LEF1*	450K
5:373378	cg05575921	0.00112	0.00009	9.1×10^{-35}	0.815	*AHRR*	450K
5:393347	cg17287155	0.00120	0.00021	9.7×10^{-9}	0.885	*AHRR*	450K
5:393366	cg04551776	0.00096	0.00017	4.0×10^{-8}	0.771	*AHRR*	450K
5:150161299	cg14580211	0.00071	0.00012	4.0×10^{-9}	0.682	*C5orf62*	450K
6:30720203	cg24859433	0.00139	0.00024	6.6×10^{-9}	0.834	*IER3*‖	450K
6:167536184	cg05094429	−0.00063	0.00012	1.0×10^{-7}	0.673	*CCR6*	450K
9:108005349	cg01692968	0.00052	0.00009	9.7×10^{-9}	0.323	*SLC44A1*‖	450K
9:134280803	cg14264316	0.00050	0.00009	1.7×10^{-8}	0.601	*PRRC2B*§	450K
11:44626750	cg01199327	−0.00064	0.00012	7.5×10^{-8}	0.842	*CD82*	450K
11:86510915	cg11660018	0.00070	0.00012	3.0×10^{-9}	0.518	*PRSS23*	450K
12:54677008	cg02583484	0.00072	0.00013	6.1×10^{-8}	0.295	*HNRNPA1;HNRPA1L-2*	450K
14:92979577	cg26829189	0.00102	0.00017	1.3×10^{-9}	0.542	*RIN3*	EPIC-unique
14:92981121	cg03345232	0.00059	0.00011	2.7×10^{-8}	0.578	*RIN3*	450K
14:92981227	cg12072028	0.00033	0.00006	1.1×10^{-7}	0.699	*RIN3*	450K
14:94547496	cg20554312	−0.00568	0.00084	1.4×10^{-11}	0.021	*DDX24;IFI27L1*	450K
15:45028270	cg10439456	0.00043	0.00008	5.5×10^{-8}	0.351	*TRIM69*	450K
16:8985593	cg08065963	−0.00091	0.00016	2.0×10^{-8}	0.694	*CARHSP1*‖	450K
16:8985638	cg05946118	−0.00095	0.00017	2.5×10^{-8}	0.708	*CARHSP1*‖	450K
17:80872461	cg10310700	−0.00074	0.00014	1.0×10^{-7}	0.817	*TBCD*	450K
19:17000585	cg03636183	0.00076	0.00010	1.1×10^{-13}	0.655	*F2RL3*	450K
20:5931325	cg20225569	0.00315	0.00057	4.2×10^{-8}	0.018	*TRMT6;MCM8*	EPIC-unique
21:43656587	cg06500161	0.00088	0.00016	2.3×10^{-8}	0.593	*ABCG1*	450K
22:30639979	cg23635663	−0.00163	0.00022	3.9×10^{-13}	0.908	*LIF*	450K

For definition of abbreviations, *see* Table 3.

Top 30 CpGs based on meta-analysis *P* values. Individual study results were obtained from robust linear regression with methylation as the predictor and FEV_1/FVC as the outcome. Covariates included age, sex, height, age squared, height squared, smoking status (never, former, or current), pack-years, and estimated cell type proportions. Study-specific covariates included study center, selection factor, ancestry principal components, batch variables, and family structure when appropriate. Methylation values were between 0 (unmethylated) and 1 (methylated). Regression coefficients represent differences in FEV_1/FVC ratio per 1% difference in methylation. Table E7 contains complete results (false discovery rate, <0.025).

*Genome build GRCh37/hg19.

†Weighted average methylation across participating studies.

‡Gene names from the Illumina annotation (17, 18), Zhou and colleagues (19), or Homer version 4.9.1 (20).

§Gene names from Zhou and colleagues (19).

‖Gene names (within ±2 Mb) from Homer version 4.9.1 (20).

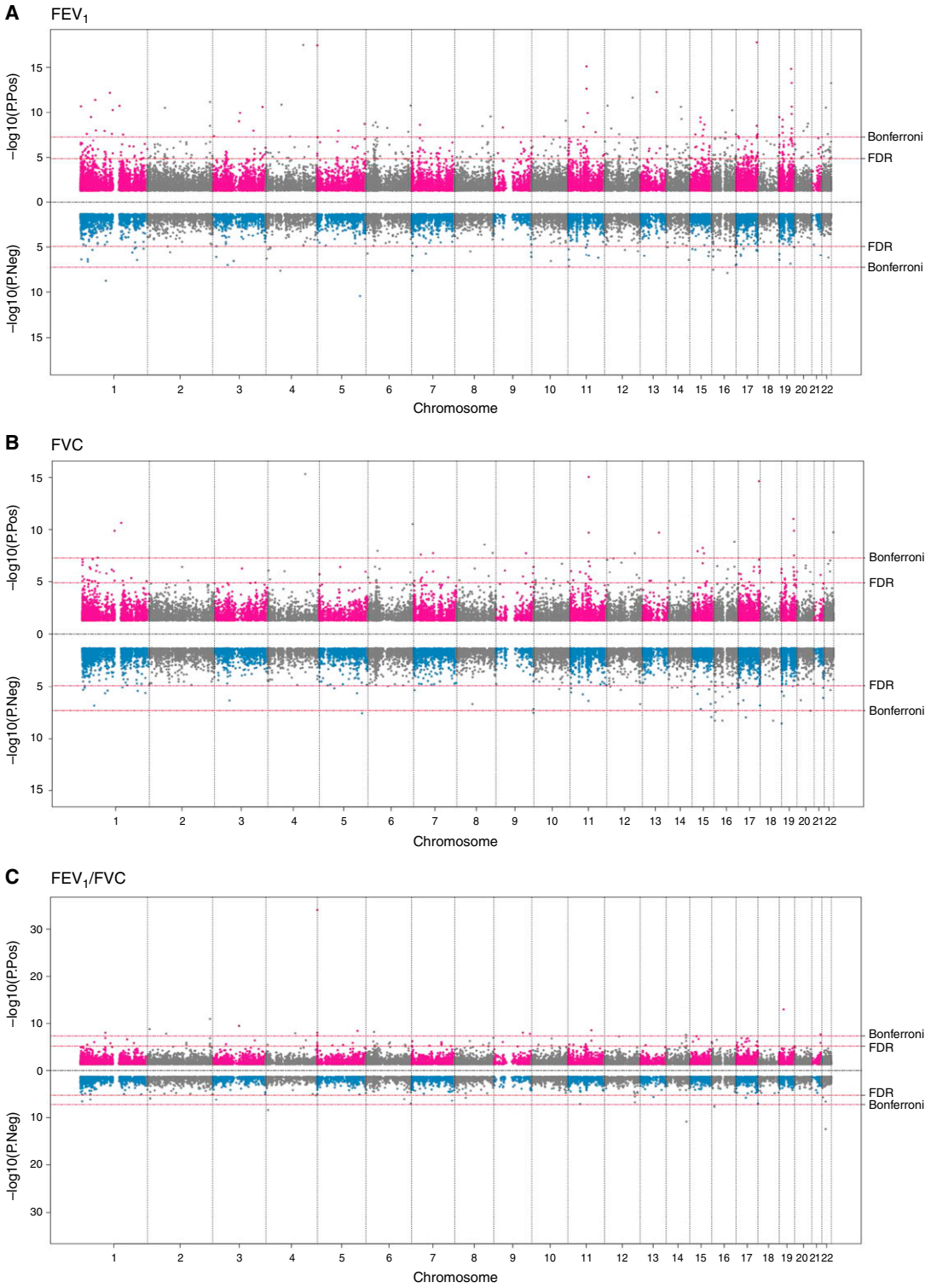

Figure 2. Visualization of 450K epigenome-wide association study meta-analysis results ($N = 17{,}503$ participants). (*A*) Miami plot for FEV_1, with each dot representing the $-\log_{10}(P \text{ value})$ of a single cytosine-phosphate-guanine probe (CpG). Each plot has two panels: upper (P.Pos) for association results with positive regression coefficients and lower (P.Neg) for association results with negative regression coefficients, with $-\log_{10}(P \text{ value})$ on the *y*-axis and 22 chromosomes on the *x*-axis. Horizontal lines depict *P* value cutoffs for statistical significance after multiple-testing correction: Bonferroni and Benjamini-Hochberg false discovery rate (FDR). CpGs having uncorrected $P > 0.05$ were not displayed. (*B*) Same for FVC. (*C*) Same for FEV_1/FVC. JAK-STAT = Janus kinase/signal transducer and activator of transcription; MAPK = mitogen-activated protein kinase.

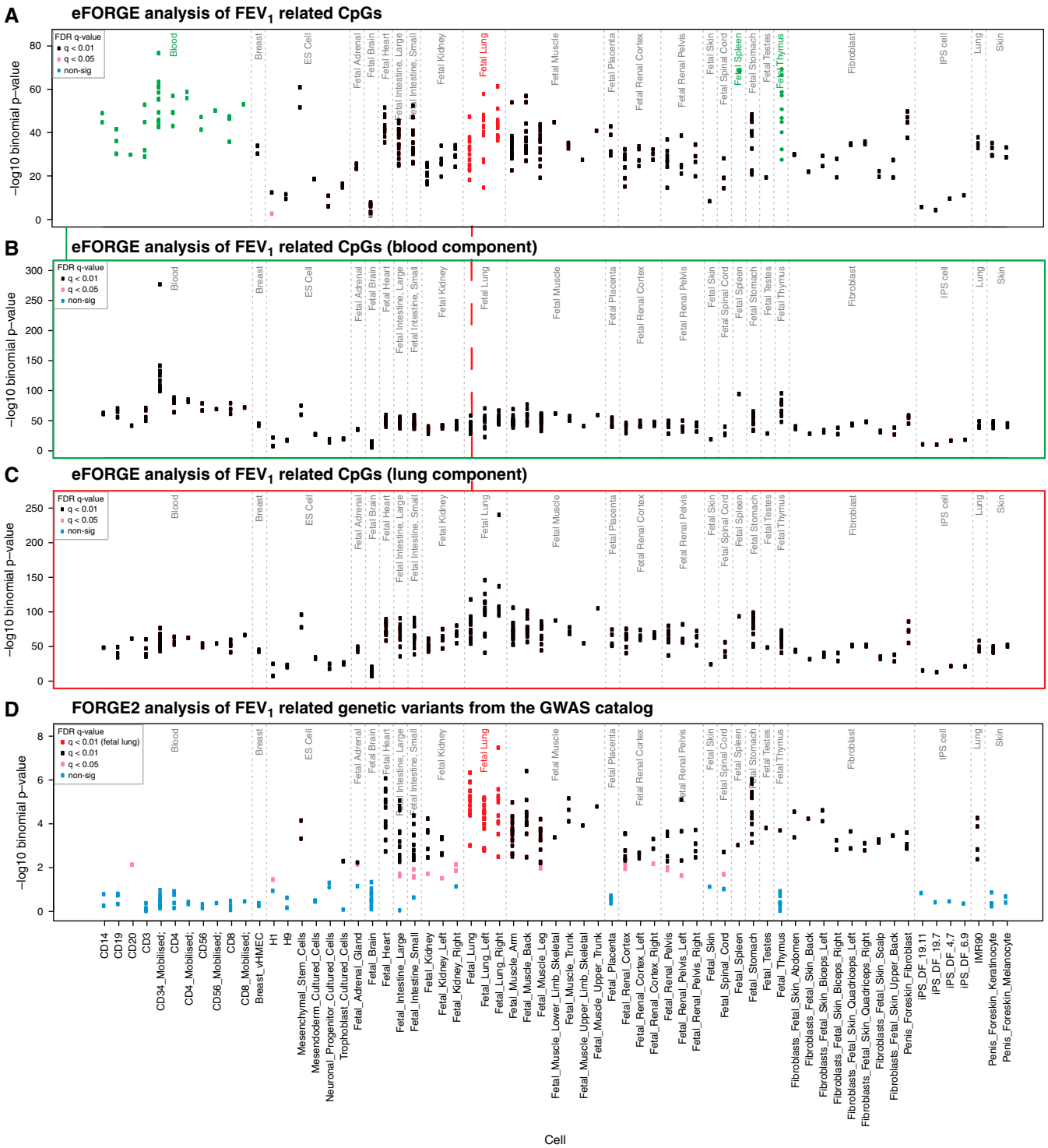

Figure 3. Integrative epigenomic analysis indicates potential effects on lung and blood and comparison with pulmonary function genome-wide association study (GWAS) loci. (*A*) FEV_1-related cytosine-phosphate-guanine probes (CpGs) (false discovery rate, <0.025): eFORGE analysis to quantify enrichment in DNase I hotspots. The *x*-axis represents tissue- and cell-type samples used in the analysis; the *y*-axis indicates enrichment ($-\log_{10} P$ value). (*B*) eFORGE DNase I hotspot analysis limited to FEV_1-related CpGs in the top blood component. (*C*) eFORGE DNase I hotspot analysis limited to FEV_1-related CpGs in the top fetal lung component. (*D*) FORGE2 DNase I hotspot analysis of FEV_1-related genetic variants from the GWAS catalog.

Figure 4. Heatmap of enriched pathways (false discovery rate [FDR], <0.05) for FEV_1, FVC, and FEV_1/FVC using the methylGSA R package. Significantly enriched (FDR, <0.05) pathways for at least one of the three traits are shown. The color spectrum is based on the *P* values corrected for multiple testing using Benjamini-Hochberg FDR. Darker shading indicates higher level of statistical significance. The heatmap was created in R version 3.6.1, using ComplexHeatmap package version 2.7.10. JAK-STAT = Janus kinase/signal transducer and activator of transcription; MAPK = mitogen-activated protein kinase.

(39, 43). Comparing our pulmonary function–related CpGs with those related to current smoking (FDR, <0.05) in a large 450K EWAS meta-analysis (38), we found substantial overlap (Tables E5–E7). In ALHS, we confirmed that the majority (61%) of these overlapping CpGs remained at least nominally significant after adjustment for lifetime smoking

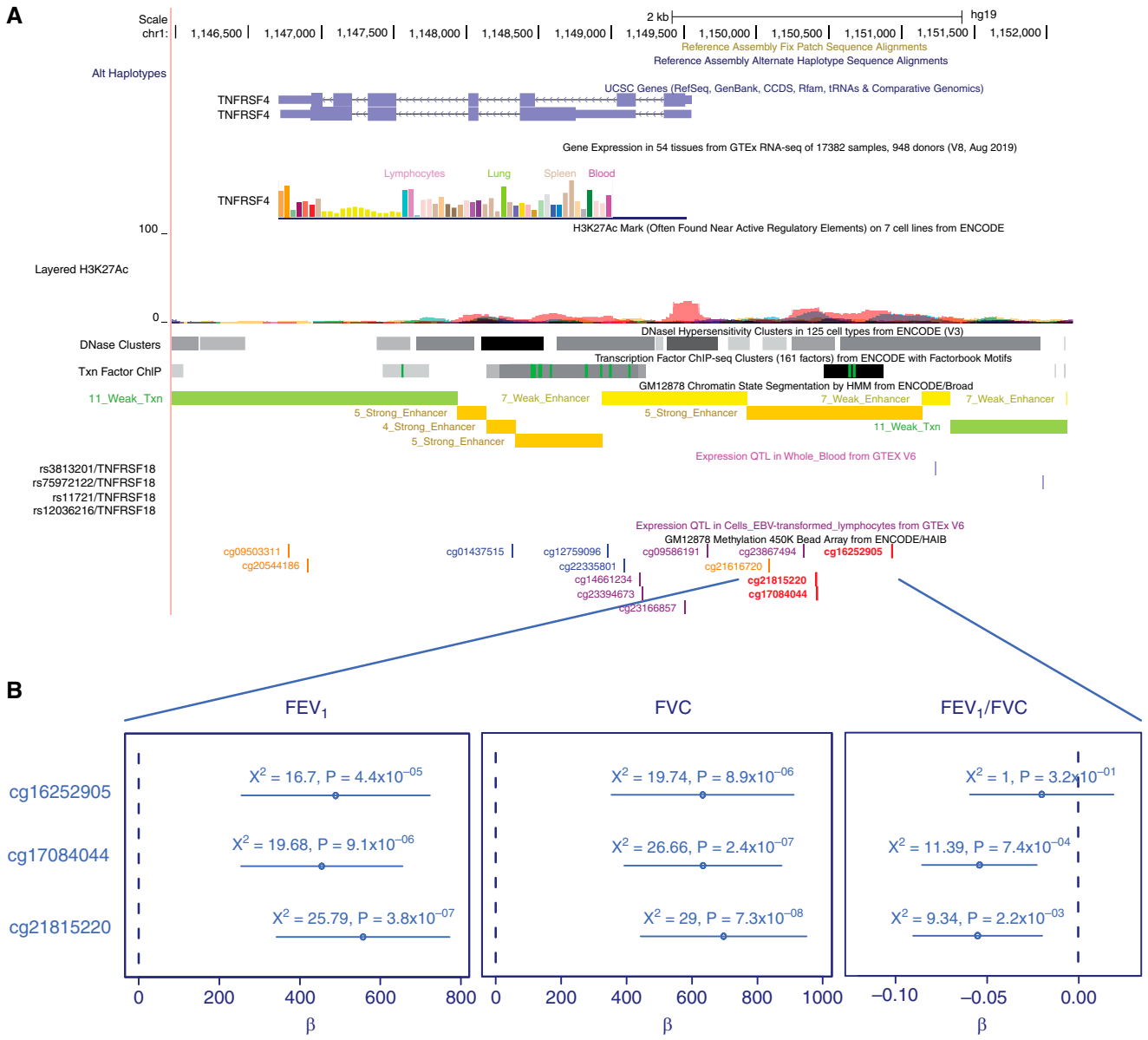

Figure 5. Pulmonary function–associated differentially methylated cytosine-phosphate-guanine probes (CpGs). (*A*) *TNFRSF4* gene browser shot displaying (in order, starting from the top): genomic coordinates, gene locations, GTEx RNA sequencing–based gene expression across different organs, H3K27ac peaks for seven ENCODE cell lines, ENCODE chromatin state segmentations and chromatin accessibility data, and coordinates for pulmonary function–associated CpGs cg21815220, cg16252905, and cg17084044. Both cg16252905 and cg17084044 are also located in DNase I hotspots in the eFORGE catalog. These data indicate that these three CpGs overlap with an immune enhancer near the promoter of *TNFRSF4*, a gene expressed in lung and immune tissue, and an enhancer and DNase I hypersensitive site from ENCODE, which were detected in immune cell samples. These three CpGs are also located within 1.5 kb of an H3K27ac peak and the transcription start site of HSP90AA1. This browser shot was generated using the University of California, Santa Cruz Genomics Institute genome browser (https://genome.ucsc.edu/) on human genome build hg19. (*B*) Forest plots for pulmonary function–associated CpGs cg21815220, cg16252905, and cg17084044 (linked to the genomic coordinates of these CpGs in *A*) indicating meta-analysis β-values and SEs across pulmonary function traits: FEV_1, FVC, and their ratio (FEV_1/FVC). Studies incorporated in the meta-analysis include ALHS (Agricultural Lung Health Study), ARIC (Atherosclerosis Risk in Communities) study, CHS (Cardiovascular Health Study), FHS (Framingham Heart Study), GS (Generation Scotland), LifeLines, LBC (Lothian Birth Cohort), MESA (Multi-Ethnic Study of Atherosclerosis), RS (Rotterdam Study), and TwinsUK.

biomarkers: *AHRR* cg05575921, *PDZD2* cg13039251, and *F2RL3* cg03636183 (39–41) (Table E24). Adjustment attenuated the magnitude of 71% of associations.

Associations by Smoking Status

Separate meta-analyses by smoking identified 196 differential methylation signals in never smokers and 465 in ever smokers (FDR, <0.025) (Tables E25 and E26). In both smoking groups, 66% of signals were unique, defined as having genome-wide significance (FDR, <0.025) in one group but not reaching nominal significance ($P > 0.05$) in the other. For 92% of these unique signals,

effect estimates were different between never and ever smokers at $P < 0.05$. Miami, QQ, and leave-one-out meta-analyses and forest plots for smoking-stratified meta-analyses are shown in Figures E9–E12. Enriched pathways in smoking-stratified analyses (Figure E13) were similar to those overall; for FEV_1/FVC, some pathways were enriched only for never smoker findings (Figure 4). We explored potential confounding by underreported smoking and secondhand smoke exposure by also adjusting the 129 associations unique to never smokers for *AHRR* cg0557592 in ALHS. Approximately 50% showed nominal significance ($P < 0.05$); all but two remained significant after the adjustment (Table E27). Effect sizes were highly correlated (Pearson correlation, 0.99; $P < 2.2 \times 10^{-16}$), indicating minimal impact of unreported smoking on findings in never smokers.

Associations across Ancestries

Ancestry-specific analyses identified 564 differently methylated CpGs in European ancestry (Table E28) and 29 differentially methylated CpGs (Table E29) in the smaller African ancestry dataset. Figures E14–E17 contain ancestry-specific Miami, QQ, and leave-one-out meta-analyses and forest plots. Effect estimates from results in European ancestry were correlated with African ancestry results (Pearson correlation, 0.59; $P < 2.2 \times 10^{-16}$); 31% of available associations reached at least nominal significance in African ancestry–specific analyses. However, only four African ancestry findings had $P < 0.05$ and matching direction in Europeans, suggesting the remainder to be unique to African ancestry.

Discussion

To our knowledge, this is the first large-scale multiancestry study examining associations between blood DNA methylation and pulmonary function. We identified numerous novel CpGs differentially methylated in relation to pulmonary function. We found evidence of replication in European and Asian ancestry studies (6, 8). We also validated findings from published studies that mostly had not been replicated. Our cross-sectional findings overlapped with limited published data on lung function decline. Our large-scale study enabled identification of pulmonary function–related methylation signals unique to smoking status, an important influence on pulmonary function. Although most signals were consistent across ancestries, we also identified signals potentially unique to individuals of African ancestry. Many CpGs correlated with nearby gene expression and were enriched for key regulatory elements in both immune cells and lung, providing functional biologic relevance of our findings further supported by MR and colocalization analyses. Implicated genes include targets of approved or investigational drugs, providing potential clinical implications.

GWASs have identified >300 genetic loci for pulmonary function (3, 4). Our EWAS loci are largely distinct. Only 3% of genes we implicated were reported in the largest lung function GWAS (Table E30) (4). Adjustment for polygenic risk scores (4, 37) for pulmonary function made little change to our results, supporting the independence of our findings from GWAS signals.

Loci identified in cross-sectional GWASs in adults primarily reflect maximally attained lung growth and are skewed toward developmental genes. In contrast to genetic variants, epigenetic alterations occur throughout life in response to numerous factors impacting pulmonary function, including environmental exposures such as smoking (38), diet (44), and air pollution (45), and endogenous influences, including systemic inflammation (46) and adiposity (47). Thus, differentially methylated genes identified by EWASs complement findings from genetic studies to identify targets potentially modifiable by lifestyle or new therapeutic interventions.

Whether EWAS findings are causal for lung function is a key question but difficult to answer. We employed MR to address this. MR requires genetic instruments related to both lung function and our EWAS CpGs. Given the novelty of most loci compared with GWASs, MR has limited ability to interrogate our findings. Nonetheless, MR analysis of all the methylation–trait pairs with genetic instruments available revealed 46 CpGs that share a genetic factor with lung function, consistent with a causal effect of methylation on lung function. However, caution is required because these could reflect violation of MR pleiotropy assumptions, and few of the 46 CpGs had the multiple genetic instruments required to test this. Regardless, without inferring causality, the shared genetic variation could reflect methylation and lung function being common consequences of regulatory impacts of the genetic variants, such as influencing TF binding. Analyses showing that, compared with other mQTLs, mQTLs for our EWAS CpGs are substantially enriched for association with lung function in GWAS support a functional role for the genomic regions uncovered by the EWAS. Along with the finding in ALHS that adjustment for genetic risk scores for lung function do not alter associations, the MR analyses suggest that our EWAS findings do not reflect reverse causation (i.e., lung function influencing methylation).

Druggable target analysis enhances the clinical relevance of our findings. Several drugs identified via ChEMBL have epigenetically relevant mechanisms. For example, mocetinostat and fimepinostat, both in phase II trials, have activity against several histone deacetylase enzymes (48, 49). Notably, three candidate drugs were annotated to tumor necrosis factor receptor superfamily member 4 (*TNFRSF4*; also known as *OX40*), including the monoclonal antibody telazorlimab, a *TNFRSF4* antagonist currently in phase II trials for atopic eczema.

Integrative epigenomics identifying enrichment of DNase I hotspots across blood and lung highlights that our findings in blood can inform key epigenomic mechanisms in the lung. Analyses of TF binding motifs identified some expressed in immune tissue. Some of our significant CpGs reside near or within genes that play an important role in inflammation and immunity in the lung. One example is *TNFRSF4*, a gene targeted by several candidate drugs, expressed in both immune and lung tissue. Annotation for chromatin states shows overlap of lung function–associated CpGs cg21815220, cg16252905, and cg17084044 with an ENCODE immune cell "strong enhancer" located proximal to the transcription start site of *TNFRSF4* (Figure 5), and these CpGs also overlap with an immune cell–specific DNase I hypersensitive site. In eFORGE, both cg16252905 and cg17084044 reside in DNase I hotspots (22). In addition, cg21815220 and cg17084044 localize to the same TF motif for hypermethylated in cancer 1 (*HIC1*), an important epigenetically regulated gene (Figure E18). This motif is in an active enhancer proximal to *TNFRSF4*

that contains cg16252905. Finding differential DNA methylation in a regulatory element proximal to a lung- and blood-specific gene hints at a putative causal role in lung function for *TNFRSF4*, a gene not previously reported in GWAS of pulmonary function that has a role in lung-associated inflammation and immunity (50). Mechanistically, *TNFRSF4* expression is thought to sustain lung tissue inflammation (50), and *TNFRSF4* blockade in experimental models improves respiratory function (50). Further research is needed to confirm our findings, which constitute the first reported epidemiological association between *TNFRSF4* and lung function.

Substantial overlap between our pulmonary function–related findings and those for COPD in our data show that examination of quantitative traits in population-based studies reveals relationships relevant to clinical outcomes such as COPD. Furthermore, we validated COPD-related CpGs identified in the lung (36), indicating that blood methylation can reflect signals in target tissue. Our results could be applied in future studies with large numbers of COPD cases to examine whether a lung function methylation score predicts that outcome.

Many CpGs associated with pulmonary function in this study and another (8) also associate with smoking (38). This led to speculation regarding whether they mediate effects of smoking on lung function (8). However, many of these overlapping CpGs are strong biomarkers of smoking that capture lifetime smoking better than pack-years from questionnaires (38, 43). One CpG (*AHRR* cg05575921) so strongly captures exposure that it is patented as a biomarker of lifetime smoking for insurance applications (39). Attempts to identify whether differential methylation at a given CpG that is a strong smoking biomarker mediates the biologic effects of smoking on lung function will likely produce false-positive evidence for mediation (43, 51). An alternative explanation for overlap of CpGs related to both smoking and lung function is residual confounding by lifetime smoking when adjusting for exposure using questionnaire data. We attempted to address residual confounding by adjusting for three CpGs, strong biomarkers of lifetime smoking, in one of our larger studies (ALHS), and found some attenuation of effect estimates after adjustment. Although some biologic mediation of the effect of smoking on lung function by methylation biomarkers of smoking, such as *AHRR* cg05575921, is possible, our results are consistent with some residual confounding. To address possible residual confounding by quantitative smoking history in another way, we examined associations in lifelong never smokers. Signals unique to never smokers are less likely due to uncontrolled confounding. However, understanding whether CpGs that are strong biomarkers of smoking are truly involved in the pathogenesis of smoking-related impairment in lung function will require deciphering the fundamental mechanisms whereby smoking alters methylation at specific sites.

This study has limitations. Assessing methylation in blood limits inference to other tissue types. Our data are cross-sectional. However, a high proportion of CpGs identified in previous longitudinal studies (7, 8) associate with lung function in our data, suggesting that cross-sectional meta-analyses can shed light on methylation predictors of decline. Individuals with European ancestry, mainly from the United States, the United Kingdom, or Northern Europe, compose 84% of our population, limiting detection of ancestry-specific signals. Because only one study had Hispanic/Latino participants, most analyses focused on European and African ancestry, limiting generalizability to other populations. Most cohorts had few individuals under 40, so we excluded this potentially interesting group. Using more widely available prebronchodilator spirometry to classify COPD is another limitation. However, this approach has been taken in previous large-scale genomic meta-analyses (3, 4). Like researchers in those genomic studies, we used the actual spirometric values adjusted for factors used in prediction equations. Although this does not allow effects of height and age to differ by sex, values were highly correlated with percent predicted values using Global Lung Function Initiative equations (ALHS Spearman correlation, 0.97; $P < 2.2 \times 10^{-16}$ for both FEV_1 and FVC). We did not perform analyses by sex. Sex-specific analyses would be of interest in future studies with larger sample sizes required for reliable interaction testing (52).

Our study has several key strengths. This is the largest multiancestry study of DNA methylation and pulmonary function to date, including studies of African ancestry, Hispanic/Latino ancestry, and European ancestry populations. The multiancestry design provided evidence for ancestry-specific signals. Our large sample size enabled determination of signals unique to ever and never smokers. Results from the EPIC array identified CpGs and genes unique to this more comprehensive platform. We confirmed our findings were independent of polygenic risk of reduced lung function. Correlation of our findings with gene expression, integrative epigenomics approaches identifying key regulatory elements enriched in both blood and lung, and MR and colocalization analyses support a functional role for the genomic regions we uncovered in EWAS. Analyses of druggable targets highlighted potential clinical utility.

In conclusion, our large-scale study comprehensively identified epigenome-wide differential methylation in blood related to pulmonary function. It extends the current literature by including newer DNA methylation arrays, different ancestry populations, greatly increased sample size, implementation of state-of-the-art *in silico* integrative epigenomics methods, MR and colocalization, and analysis of druggable targets. We identified many novel genes related to lung function. These are potentially modifiable targets for development of preventive and therapeutic strategies. In addition, the results can be leveraged for development of epigenomic risk scores for predictive biomarkers of lung disease. These findings provide new insights into the pathogenesis of lung function impairment and respiratory disease. ■

Author disclosures are available with the text of this article at www.atsjournals.org.

Acknowledgment: We thank Drs. Huiling Li, Frank Day, and Kathryn Dalton of the National Institute of Environmental Health Sciences for providing expert assistance.

References

1. Young RP, Hopkins R, Eaton TE. Forced expiratory volume in one second: not just a lung function test but a marker of premature death from all causes. *Eur Respir J* 2007;30:616–622.
2. Schünemann HJ, Dorn J, Grant BJ, Winkelstein W Jr, Trevisan M. Pulmonary function is a long-term predictor of mortality in the general population: 29-year follow-up of the Buffalo Health Study. *Chest* 2000; 118:656–664.
3. Wyss AB, Sofer T, Lee MK, Terzikhan N, Nguyen JN, Lahousse L, *et al.* Multiethnic meta-analysis identifies ancestry-specific and cross-ancestry loci for pulmonary function. *Nat Commun* 2018; 9:2976.
4. Shrine N, Guyatt AL, Erzurumluoglu AM, Jackson VE, Hobbs BD, Melbourne CA, *et al.*; Understanding Society Scientific Group. New genetic signals for lung function highlight pathways and chronic obstructive pulmonary disease associations across multiple ancestries. *Nat Genet* 2019;51:481–493.
5. Qiu W, Baccarelli A, Carey VJ, Boutaoui N, Bacherman H, Klanderman B, *et al.* Variable DNA methylation is associated with chronic obstructive pulmonary disease and lung function. *Am J Respir Crit Care Med* 2012; 185:373–381.
6. Lee MK, Hong Y, Kim SY, Kim WJ, London SJ. Epigenome-wide association study of chronic obstructive pulmonary disease and lung function in Koreans. *Epigenomics* 2017;9:971–984.
7. Carmona JJ, Barfield RT, Panni T, Nwanaji-Enwerem JC, Just AC, Hutchinson JN, *et al.* Metastable DNA methylation sites associated with longitudinal lung function decline and aging in humans: an epigenome-wide study in the NAS and KORA cohorts. *Epigenetics* 2018;13:1039–1055.
8. Imboden M, Wielscher M, Rezwan FI, Amaral AFS, Schaffner E, Jeong A, *et al.* Epigenome-wide association study of lung function level and its change. *Eur Respir J* 2019;54:1900457.
9. Bermingham ML, Walker RM, Marioni RE, Morris SW, Rawlik K, Zeng Y, *et al.* Identification of novel differentially methylated sites with potential as clinical predictors of impaired respiratory function and COPD. *EBioMedicine* 2019;43:576–586.
10. Lee M, Terzikhan N, Lahousse L, De Vries M, Sikdar S, London S, *et al.*; CHARGE Consortium Epigenetics Working Group. Epigenome-wide association study of pulmonary function traits and chronic obstructive pulmonary disease: a multiethnic meta-analysis [abstract]. *Am J Respir Crit Care Med* 2019;199:A4867.
11. Graham BL, Steenbruggen I, Miller MR, Barjaktarevic IZ, Cooper BG, Hall GL, *et al.* Standardization of spirometry 2019 update. an official American Thoracic Society and European Respiratory Society technical statement. *Am J Respir Crit Care Med* 2019;200:e70–e88.
12. Houseman EA, Accomando WP, Koestler DC, Christensen BC, Marsit CJ, Nelson HH, *et al.* DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics* 2012;13:86.
13. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* 2010; 26:2190–2191.
14. Rice K, Higgins JPT, Lumley T. A re-evaluation of fixed effect(s) meta-analysis. *J R Stat Soc Ser A Stat Soc* 2018;181:205–227.
15. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Series B Stat Methodol* 1995;57:289–300.
16. Peters TJ, Buckley MJ, Statham AL, Pidsley R, Samaras K, V Lord R, *et al.* De novo identification of differentially methylated regions in the human genome. *Epigenetics Chromatin* 2015;8:6.
17. Illumina. Illumina Infinium HumanMethylation450 BeadChip Annotation [accessed 2015 Sep 18]. Available from: https://support.illumina.com/array/array_kits/infinium_humanmethylation450_beadchip_kit/downloads.html.
18. Illumina. Illumina Infinium MethylationEPIC BeadChip Annotation [accessed 2020 Aug 7]. Available from: https://support.illumina.com/array/array_kits/infinium-methylationepic-beadchip-kit/downloads.html.
19. Zhou W, Laird PW, Shen H. Comprehensive characterization, annotation and innovative use of Infinium DNA methylation BeadChip probes. *Nucleic Acids Res* 2017;45:e22.
20. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, *et al.* Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* 2010;38:576–589.
21. Breeze CE, Paul DS, van Dongen J, Butcher LM, Ambrose JC, Barrett JE, *et al.* eFORGE: a tool for identifying cell type-specific signal in epigenomic data. *Cell Rep* 2016;17:2137–2150.
22. Breeze CE, Reynolds AP, van Dongen J, Dunham I, Lazar J, Neph S, *et al.* eFORGE v2.0: updated analysis of cell type-specific signal in epigenomic data. *Bioinformatics* 2019;35:4767–4769.
23. Breeze CE. Cell type-specific signal analysis in EWAS [preprint]. *bioRxiv*; 2021 [accessed 2021 June 3]. Available from: https://doi.org/10.1101/2021.05.21.445209.
24. Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, Heravi-Moussavi A, *et al.*; Roadmap Epigenomics Consortium. Integrative analysis of 111 reference human epigenomes. *Nature* 2015;518:317–330.
25. Kolde R, Laur S, Adler P, Vilo J. Robust rank aggregation for gene list integration and meta-analysis. *Bioinformatics* 2012;28:573–580.
26. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA* 2005;102:15545–15550.
27. Huan T, Joehanes R, Song C, Peng F, Guo Y, Mendelson M, *et al.* Genome-wide identification of DNA methylation QTLs in whole blood highlights pathways for cardiovascular disease. *Nat Commun* 2019;10:4267.
28. Bonder MJ, Luijk R, Zhernakova DV, Moed M, Deelen P, Vermaat M, *et al.*; BIOS Consortium. Disease variants alter transcription factor levels and methylation of their binding sites. *Nat Genet* 2017;49: 131–138.
29. Min JL, Hemani G, Hannon E, Dekkers KF, Castillo-Fernandez J, Luijk R, *et al.*; BIOS Consortium. Genomic and phenotypic insights from an atlas of genetic effects on DNA methylation. *Nat Genet* 2021;53:1311–1321.
30. Hemani G, Zheng J, Elsworth B, Wade KH, Haberland V, Baird D, *et al.* The MR-Base platform supports systematic causal inference across the human phenome. *Elife* 2018;7:e34408.
31. Giambartolomei C, Vukcevic D, Schadt EE, Franke L, Hingorani AD, Wallace C, *et al.* Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet* 2014;10:e1004383.
32. Mendez D, Gaulton A, Bento AP, Chambers J, De Veij M, Félix E, *et al.* ChEMBL: towards direct deposition of bioassay data. *Nucleic Acids Res* 2019;47:D930–D940.
33. Hankinson JL, Odencrantz JR, Fedan KB. Spirometric reference values from a sample of the general U.S. population. *Am J Respir Crit Care Med* 1999;159:179–187.
34. Mannino DM, Buist AS. Global burden of COPD: risk factors, prevalence, and future trends. *Lancet* 2007;370:765–773.
35. Pauwels RA, Buist AS, Calverley PM, Jenkins CR, Hurd SS; GOLD Scientific Committee. Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease. NHLBI/WHO Global Initiative for Chronic Obstructive Lung Disease (GOLD) Workshop summary. *Am J Respir Crit Care Med* 2001;163:1256–1276.
36. Morrow JD, Cho MH, Hersh CP, Pinto-Plata V, Celli B, Marchetti N, *et al.* DNA methylation profiling in human lung tissue identifies genes associated with COPD. *Epigenetics* 2016;11:730–739.
37. Sikdar S, Wyss AB, Lee MK, Hoang TT, Richards M, Beane Freeman LE, *et al.* Interaction between Genetic Risk Scores for reduced pulmonary function and smoking, asthma and endotoxin. *Thorax* 2021;76:1219–1226.
38. Joehanes R, Just AC, Marioni RE, Pilling LC, Reynolds LM, Mandaviya PR, *et al.* Epigenetic signatures of cigarette smoking. *Circ Cardiovasc Genet* 2016;9:436–447.
39. Andersen AM, Ryan PT, Gibbons FX, Simons RL, Long JD, Philibert RA. A droplet digital PCR assay for smoking predicts all-cause mortality. *J Insur Med* 2018;47:220–229.
40. Maas SCE, Vidaki A, Wilson R, Teumer A, Liu F, van Meurs JBJ, *et al.*; BIOS Consortium. Validated inference of smoking habits from blood with a finite DNA methylation marker set. *Eur J Epidemiol* 2019;34: 1055–1074.
41. Zhang Y, Elgizouli M, Schöttker B, Holleczek B, Nieters A, Brenner H. Smoking-associated DNA methylation markers predict lung cancer incidence. *Clin Epigenetics* 2016;8:127.

42. Breeze CE, Haugen E, Reynolds A, Teschendorff A, van Dongen J, Lan Q, *et al.* Integrative analysis of 3604 GWAS reveals multiple novel cell type-specific regulatory associations. *Genome Biol* 2022; 23:13.
43. London SJ. Methylation, smoking, and reduced lung function. *Eur Respir J* 2019;54:1900920.
44. Do WL, Whitsel EA, Costeira R, Masachs OM, Le Roy CI, Bell JT, *et al.* Epigenome-wide association study of diet quality in the Women's Health Initiative and TwinsUK cohort. *Int J Epidemiol* 2021;50:675–684.
45. Lee MK, Xu CJ, Carnes MU, Nichols CE, Ward JM, Kwon SO, *et al.*; BIOS consortium. Genome-wide DNA methylation and long-term ambient air pollution exposure in Korean adults. *Clin Epigenetics* 2019;11:37.
46. Myte R, Sundkvist A, Van Guelpen B, Harlid S. Circulating levels of inflammatory markers and DNA methylation, an analysis of repeated samples from a population based cohort. *Epigenetics* 2019;14:649–659.
47. Geurts YM, Dugué PA, Joo JE, Makalic E, Jung CH, Guan W, *et al.* Novel associations between blood DNA methylation and body mass index in middle-aged and older adults. *Int J Obes* 2018;42: 887–896.
48. Zhang Q, Sun M, Zhou S, Guo B. Class I HDAC inhibitor mocetinostat induces apoptosis by activation of miR-31 expression and suppression of E2F6. *Cell Death Discov* 2016;2:16036.
49. Gunst JD, Kjaer K, Olesen R, Rasmussen TA, Østergaard L, Denton PW, *et al.* Fimepinostat, a novel dual inhibitor of HDAC and PI3K, effectively reverses HIV-1 latency *ex vivo* without T cell activation. *J Virus Erad* 2019;5:133–137.
50. Burrows KE, Dumont C, Thompson CL, Catley MC, Dixon KL, Marshall D. OX40 blockade inhibits house dust mite driven allergic lung inflammation in mice and in vitro allergic responses in humans. *Eur J Immunol* 2015;45:1116–1128.
51. Valeri L, Reese SL, Zhao S, Page CM, Nystad W, Coull BA, *et al.* Misclassified exposure in epigenetic mediation analyses. Does DNA methylation mediate effects of smoking on birthweight? *Epigenomics* 2017;9:253–265.
52. Koo HK, Morrow J, Kachroo P, Tantisira K, Weiss ST, Hersh CP, *et al.* Sex-specific associations with DNA methylation in lung tissue demonstrate smoking interactions. *Epigenetics* 2021;16:692–703.