An Epigenome-Wide Study of DNA Methylation Profiles and Lung Function among American Indians in the Strong Heart Study

Arce Domingo-Relloso, Angela Riffo-Campos, Martha Powers, Maria Tellez-Plaza, Karin Haack, Robert Brown, Jason Umans, M. Daniele Fallin, Shelley A. Cole, Ana Navas-Acien, Tiffany R. Sanchez

Abstract

Background: Epigenetic modifications, including DNA methylation (DNAm), are often related to environmental exposures, and are increasingly recognized as key processes in the pathogenesis of chronic lung disease. American Indian communities have a high burden of lung disease compared to the national average. The objective of this study was to investigate the association of DNA methylation and lung function in the Strong Heart Study (SHS).

Methods: We conducted a cross-sectional study of American Indian adults, 45-74 years of age who participated in SHS in 1989-1991. DNAm was measured using the Illumina Infinium Human MethylationEPIC platform at baseline. Lung function was measured via spirometry at visit 2 (1993-1995). Airflow obstruction was defined as FEV1 <70% predicted and FEV1/FVC <0.7, restriction was defined as FEV1/FVC>0.7 and FVC <80% predicted, and normal spirometry defined as FEV1/FVC >0.7, FEV1>70% predicted, FVC >80% predicted. We used elastic-net models to select relevant CpGs for lung function and spirometry-defined lung disease. We next assessed associations between CpGs selected by elastic-net, lung function and spirometry-defined lung disease. We also examined protein-protein interactions for obstructive and restrictive pattern spirometry phenotypes.

Results: Among 1677 participants, 21.2% had spirometry-defined airflow limitation and 13.6% had spirometry-defined restrictive pattern lung function. Elastic-net models selected 1118 CpGs as predictors of airflow limitation and 1385 for restrictive pattern lung function. Many of the CpGs targeted genes with biological roles related to lung function (e.g. protein kinases).

Conclusion: We found several differentially methylated CpG sites associated with chronic lung disease. Further experimental and longitudinal studies are needed to assess whether DNA methylation has a causal role in lung disease.

Key words: DNA methylation, lung function, lung disease, epigenetics, American Indians

Introduction

Between 1980 and 2014, the mortality rate for chronic respiratory disease, including chronic obstructive pulmonary disease (COPD) and interstitial lung disease (ILD), increased by 29.7% in the U.S.[1] COPD is defined by airflow limitation that is not fully reversible,[2] whereas ILD is defined by the presence of cellular proliferation, infiltration and/or fibrosis of the lung not due to infection or neoplasia and resembles a restrictive spirometry pattern.[3] The development of chronic lung disease is associated with both environmental and genetic risk factors. Although cigarette smoking is one of the main risk factors for chronic lung disease development, not every smoker will develop chronic lung disease.

Epigenetic modifications, including DNA methylation (DNAm), are often related to environmental exposures, and are increasingly recognized as key processes in the pathogenesis of chronic lung disease.[4–8] In a systematic review examining the association of lung function with global, epigenome-wide, and locus-specific DNAm in peripheral blood from population-based studies, five of the six included studies showed evidence that DNAm profiles were differentially associated with lung function, including loci associated with the *SERPINA1*, *ORC4*, *WT1*, and *FXYD1* genes.[9] *SERPINA1,* for example, encodes alpha-1-antitrypsin, and alpha-1-antitrypsin deficiency has been shown to cause degenerative pulmonary disease through unregulated tissue breakdown.[10] Evidence suggests that DNAm alterations could play a role in the predisposition to or pathogenetic mechanism of lung disease. While there is a growing number of studies that evaluate the association of lung disease and differential DNAm profiles, epidemiologic studies examining lung disease-related DNAm profiles of American Indian communities are scarce.

The objective of this study was to investigate the association of DNA methylation with lung function and spirometry-defined lung disease in the Strong Heart Study. We used elastic-net models to select relevant CpGs, and conducted bioinformatic analyses to evaluate the biological plausibility of the findings.

Methods

Study population

The Strong Heart Study (SHS) is a prospective cohort study funded by the National Heart, Lung and Blood Institute and the National Institute of Environmental Health Sciences to investigate cardiovascular disease and its risk factors in American Indian adults.[11] In 1989-1991, 4549 men and women aged 45–75 years and members of 13 tribes based in Arizona, Oklahoma, and North Dakota and South Dakota who were free of cardiovascular disease enrolled in the study. DNA methylation was measured in 2,351 participants at the SHS baseline visit (1989-1991). Details regarding inclusion criteria for blood DNA methylation measurements have been described elsewhere.[12] Among eligible participants with DNA methylation, participants without a valid spirometry test at visit 2 (1993-1995) were excluded (N=648), as were individuals missing relevant covariate information, leaving a total of 1677 participants in this study.

Participant characteristics

At baseline, trained and certified nurses and medical examiners administered a standardized questionnaire and physical examination including collecting information on sociodemographic (age, sex, study region, education level), lifestyle (smoking status), medical history (prior tuberculosis infection) and anthropometric (height and weight) factors. A fasting blood sample was also collected during the physical examination.

Spirometry and self-reported lung disease

Pre-bronchodilator spirometry testing was conducted by centrally trained and certified nurses and technicians. Maneuvers were considered acceptable according to then-current American Thoracic Society recommendations.[13] Reference values for SHS participants have been previ-

ously derived.[14] Spirometry endpoints include continuous lung function measures of FEV1, FVC and FEV1/FVC and fixed ratio-defined airflow limitation (FEV1/FVC<0.70) and restriction (FVC<80% predicted, FEV1/FVC<0.70). For fixed-ratio defined lung disease endpoints, participants with FEV1/FVC>0.7 and FVC>80% predicted served as the reference group.

Blood DNA methylation determinations

Details of microarray DNA methylation measurements at the baseline visit of the SHS have been published elsewhere.[12] Briefly, DNA methylation from white blood cells was measured using the Illumina MethylationEPIC BeadChip (850K). Individuals with low detection p-values, cross-hybridizing probes, probes located in sex chromosomes and SNPs (Single Nucleotide Polymorphisms) with minor allele frequency > 0.05 were excluded. Single sample noob normalization and regression on correlated probes normalization were conducted following Illumina's recommendations for preprocessing (*Minfi* R package).[15] Blood cell proportions (CD8T, CD4T, NK cells, B cells, monocytes and neutrophils) were estimated using the *FlowSorted.Blood.EPIC* R package.[16] Beta-values, which range from 0 to 1 and represent the proportion of unconverted cytosines (Cs) in bisulfite-converted DNA at specific locations, were calculated using the R package *minfi*.[15] We used all cell types except neutrophils (the most common cell type) as adjustment variables in regression models. We detected and corrected for potential batch effects by sample plate, sample row, and DNA isolation time with the combat function (*sva* R package).[17] We conducted annotation of CpGs to the nearest gene according to the Infinium MethylationEPIC Manifest File v1.0b4.[18,19] The preprocessing resulted in data from 1677 individuals and 788,368 CpG sites in our analyses.

Statistical Analysis

Differentially Methylated Positions (DMPs) analysis by Elastic Net: In total, we examined five outcomes: 1) FEV1 as a continuous variable, 2) FVC as a continuous variable, 3) FEV1/FVC as a continuous variable, 4) obstructive vs. normal lung function, and 5) restrictive vs. normal lung function. Given that many smoking-related genes were found to be DMPs for obstructive disease, we repeated the analysis among never smokers. In contrast to traditional one-by-one liner regression CpG modeling approaches, which are limited in accounting for large numbers of predictors or correlated data, we used elastic-net. Elastic net methods have recently become very popular in Epigenome-Wide and Genome-Wide Association Studies[20–22] as the elastic-net method is robust to limitations of the Lasso method such as multicollinearity in very high-dimensional settings.[23,24] Specifically, when the correlations among predictors are high, the elastic-net method exceeds the predictive accuracy of the Lasso [25]. Elastic-net has previously shown to be able to select relevant predictors in differential DNA-methylation analysis and has been used to construct methylation-based risk-scores that have shown great promise for disease prediction based in epigenetic data.[20,26,27]

We used elastic-net to select DMPs (simultaneously modeled independent variables) that were associated with lung function and disease (dependent variables). Using the DMPs selected by elastic-net, we then ran traditional linear regression models (for continuous outcomes) and logistic regression models (for dichotomous outcomes) to obtain effect estimates and 95% CI-s. Models were adjusted for smoking status (never, former, current), cumulative smoking (cigarette pack-years), age, sex, BMI, study center (Arizona, Oklahoma or North and South Dakota), smoking status, eGFR, prior tuberculosis diagnosis, cell counts (CD8T, CD4T, NK, B cells and monocytes) and five genetic PCs.[28] For continuous lung function measures, we also adjusted for height. Multiple comparisons were accounted separately using the Benjamini and Hochberg method for false discovery rates (FDR).

Protein-protein interaction network. From the DMPs selected in the elastic-net model, we created two sets of unique protein-coding genes. The first set represents an obstructive-phenotype using the following 3 outcomes: FEV1, FEV1/FVC, and obstructive vs. normal lung function. The second set represents a restrictive phenotype using the following 3 outcomes: FEV1, FVC, and restrictive vs. normal lung function. The protein interaction information was obtained from the STRING database v11.0.[29] The STRING database provides a confidence score (from 0 to 1) obtained from the estimated likelihood of each annotated interaction between a given pair of proteins being biologically meaningful, specific and reproducible.[29] The protein interaction networks were analyzed and displayed using the yfiles Organic layout by Cytoscape v. 3.7.2.[30] In the resultant networks, we only kept connections obtained from experimental studies with a minimum confidence score of 0.4. The unconnected nodes were excluded from the network.

Results

1,677 participants were included in this study (Figure 1). Participant's characteristics are presented in Table 1. At baseline (time of blood collection), all subjects were 44–75 years of age, with an average age of 55 years. 61% of participants were female, and 32% had never smoked. The elastic-net model selected 838 DMPs for FEV1, 762 DMPs for FVC, 557 DMPs for FEV1/FVC, 1118 DMPs for airflow limitation and 1385 for restrictive pattern lung function. 322 (48%) of the DMPs selected for FEV1 overlapped with selected DMPs for FVC elastic-net models, whereas 20 DMPs overlapped with FEV1/FVC (Figure 2). Airflow limitation shared 31 DMPs with FEV1, 142 DMPs with FEV1/FVC and 6 DMPs overlapped together with FEV1, FEV1/FVC and airflow limitation (Figure 2). Restrictive pattern lung function shared 48 DMPs with FEV1, 62 DMPs each with FVC and 30 DMPs overlapped together with FEV1, and FVC and restrictive pattern lung function (Figure 2). While 12 DMPs overlapped between airflow limitation and restrictive pattern, no DMPs overlapped with all 5 outcomes.

Table 3 shows the top five DMPs selected by the elastic-net models and the mean differences (95% CIs) for continuous lung function measures (FEV1, FVC, FEV1/FVC) calculated using linear regression models. Table 4 shows the top five DMPs selected by the elastic-net models and the Odds Ratios (95% CIs) for airflow limitation and restrictive pattern calculated using logistic regression models. A list of all DMPs selected by elastic-net models and respective effect estimates for each of the 5 lung function outcomes studied are included in the Supplementary Excel Tables.

In the protein-protein interaction networks, the obstructive phenotype network (FEV1, FEV1/FVC and airflow limiation vs normal lung function) included 1965 unique genes associated with 2326 DMPs identified by elastic-net models. Of these, 1467 ncRNA genes or unconnected nodes were discarded (Figure 5, network 1). The protein-protein interaction network for obstruction included 498 nodes and 829 interactions (Figure 6 and Supplementary Material). *EGFR*, *MAPK1* and *PRPF8* were the most connected nodes in the network with 32, 22 and 19 interactions, respectively. The restrictive phenotype (FEV1, FVC, and restrictive pattern vs normal lung function) included 2156 unique genes associated with 2583 DMPs identified by elastic-net models. Of these, 1551 ncRNA genes or unconnected nodes were discarded (Figure 5, network 2). The protein-protein interaction network for restrictive pattern included 605 nodes and 1101 interactions (Figure 7). *UBA52*, *CREBBP*, *SRC* and *EGFR* were the most connected nodes with 38, 34, 29 and 27 interactions, respectively.

Given that several smoking-related genes (*AHRR, F2RL3, PRSS23, RARA*) were found to be DMPs for airflow limitation, we repeated those analysis restricted to never smokers. There were N=562 never smokers in our study, of which 168 presented airflow limitation. 468 CpGs were

selected by elastic-net as DMPs, including CpGs annotated to the genes *AHRR, PRSS23* and *RARA* (Supplementary Excel Table).

Discussion

We conducted an epigenome-wide association study investigating the association between DNA methylation and lung function and explored common epigenetic signatures between lung function and disease. Using robust methods for high-dimensional correlated data, we found 1118 DMPs associated with airflow limitation and 1385 associated with restrictive pattern lung function. The biological functions of the top genes, as well as the most connected nodes on the protein-protein interaction network, were related to biological processes associated with lung disease.

Several top genes and highly connected nodes in the protein-protein interaction networks for FEV1 (*PIM1*), FVC (*CDK5*), FEV1/FVC (*NTRK2*), airflow limitation (*CPPED1*) and restrictive pattern (*EGFR, MAPK1*) are protein kinases. Protein kinases play a role in many key pulmonary cellular responses, including mediating inflammatory signals and airway remodeling. Thus, they have been proposed as therapeutic targets for several lung diseases such as chronic obstructive pulmonary disease and asthma.[31,32] *PIM1* and *CKD5*, top genes associated with FEV1 and FVC, respectively, are serine/threonine protein kinases. An animal study reported evidence of high-tidal volume ventilation increasing pulmonary fibrosis in acute lung injury via the serine/threonine protein kinase B.[33] Also, the *MAPK1* gene (mitogen-activated protein kinase 1) was a highly connected node in the obstruction protein-protein interaction network. Lung endothelial barrier function is regulated by multiple signaling pathways, including mitogen-activated protein kinases (MAPK).[34] MAPK kinases might contribute to ameliorate the lung endothelial barrier-disruptive effects.[35]

On the other hand, the *UBA52* gene (Ubiquitin A-52 Residue Ribosomal Protein Fusion Product 1) was a highly connected node in the restrictive pattern protein-protein interaction network. Ubiquitination regulates the proteins that modulate the alveolocapillary barrier and the inflammatory response, therefore playing an important role in acute lung injury. [36] The *ADARB2* gene (Adenosine Deaminase RNA Specific B2) was the top differentially methylated position for restrictive pattern. An animal model showed that adenosine deaminase deficiency might lead to pulmonary fibrosis.[37] Our work provides further evidence that these biological processes are involved in lung disease, however, experimental studies are needed to disentangle whether DNA methylation changes influence these biological pathways or, conversely, alterations in these pathways lead to DNA methylation dysregulations.

Importantly, hypomethylation of several smoking-related genes was associated with obstruction in our study (*AHRR, F2RL3, PRSS23, RARA*), whereas none of those were associated with restriction (Supplementary Excel Table). When running the obstructive disease analysis only among never smokers, all genes remained significant except *F2RL3*. This suggests that epigenetic dysregulation in these genes might be related to adverse lung outcomes independent of smoking. This is consistent with previous literature pointing that hypomethylation in the gene *AHRR* is associated with lower lung function and respiratory symptoms,[38,39] and with DNA methylation dysregulations in *AHRR* being associated with lung function in two multi-cohort epigenome-wide association studies in adults.[40,41] Further studies are needed to investigate the potential causal role of the *AHRR* in lung disease.

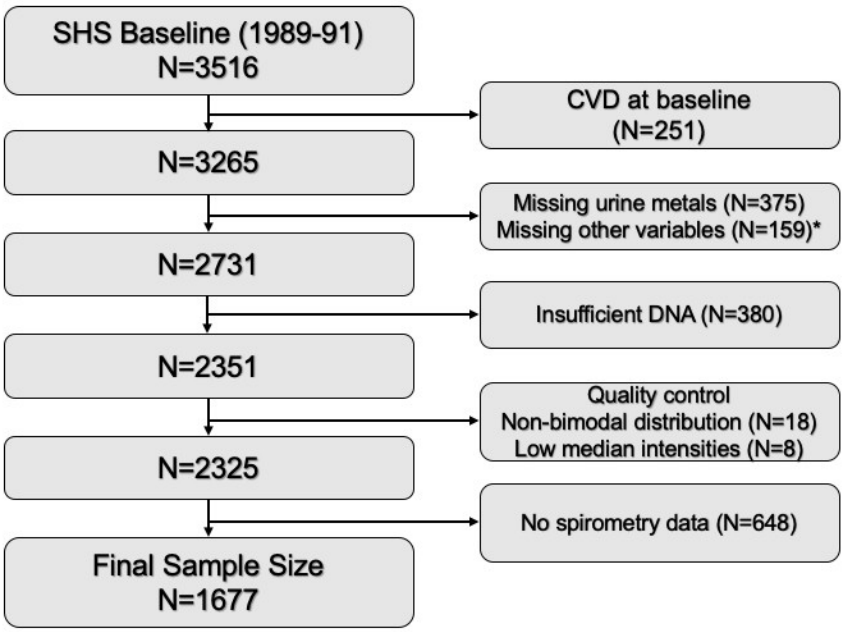
This is, to our knowledge, the first epigenome-wide association study with the main focus on lung function conducted in a population of American Indians. We found four previous epigenome-wide association studies of lung function with spirometry measurements in adults.

One did not report any significant associations.[42] The second one was conducted in 2012 in a population of female twins, and only found one DMR associated with FEV1 and FVC annotated to the gene *WT1*, which was not replicated in our population.[43] In 2018, another EWAS of lung function was conducted in two cohorts. Three CpG sites associated with lung function were consistently found in the two cohorts.[40] Only one (cg05575921, annotated to *AHRR*) was replicated in our study. Last, in 2019, another EWAS was conducted in eight cohorts (three discovery cohorts and five replication cohorts),[41] and our results were highly consistent with the findings. Most of the genes associated with lung function were replicated in our population. Although several other epigenome-wide association studies in lung function have been conducted, they were conducted in specific populations such as children[44] , individuals with chronic obstructive pulmonary disease,[45] individuals with HIV[46] or never smokers.[47] Findings for these specific populations might not be generalizable. Nevertheless, many differentially methylated positions found in these studies overlapped with our findings. For instance, eight of the top sites found in the never-smokers EWAS of lung function were replicated in our population, which might indicate that the epigenomic signature of lung function is also stable across different population groups. However, the findings of the meta-analysis conducted by Machin et al. were not replicated in our population, which suggests that part of the epigenomic signature of lung function might also be specific to populations.[9]

This work has some limitations. First, only 1677 of the SHS participants were included, which might induce some bias among those who were excluded due to not meeting spirometry quality standards. In addition, we only have one measure of spirometry, therefore, we cannot discard potential measurement errors. Strengths of this work include using one of the largest arrays currently available in microarray technology (the EPIC array), the large sample size in an indigenous population, measurement of spirometry-defined lung disease, and innovative statistical methods that allow evaluating the effect of methylation sites jointly instead of individually.

In conclusion, we found several differentially methylated positions for FEV1, FVC, FEV1/FVC, obstructive pattern and restrictive pattern, with several genes pointing to biological pathways related to lung disease including protein kinases, which are therapeutic targets for lung disease. Further studies are needed to investigate the potential mechanistic role of DNA methylation in lung disease.

Figure 1. Flowchart of included participants



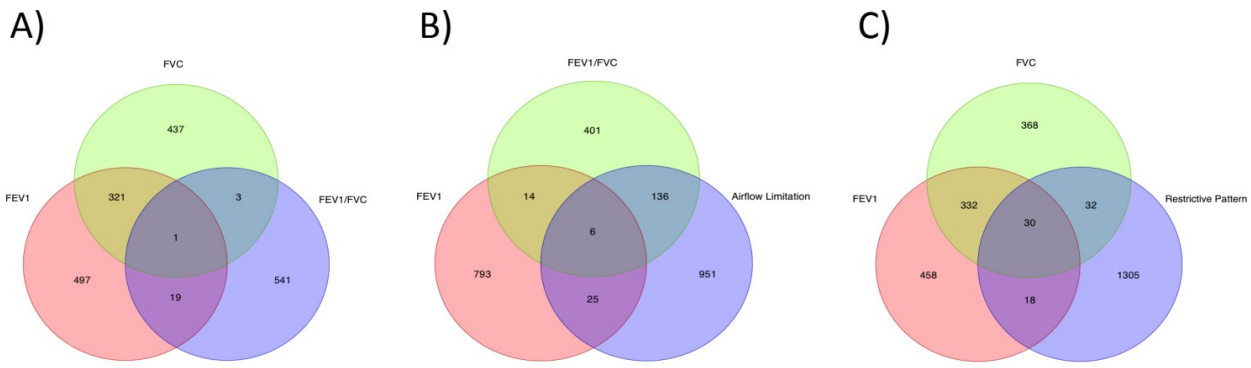
* 5 participants missing education, 2 smoking status, 11 BMI, 52 LDL cholesterol, 14 hypertension treatment, 111 eGFR, 30 diabetes

Table 1. Participant characteristics

	Overall (N=1677)
Female, *n (%)*	1015 (61)
Age (years), *median (IQR)*	55 (49, 61)
Height (cm), *median (IQR)*	165 (159, 173)
BMI, *median (IQR)*	29.7 (26.3, 33.8)
Smoking, *n (%)*	
Never	531 (32)
Former	503 (30)
Current	643 (38)
Center, *n (%)*	
Arizona	237 (14)
Dakotas	726 (43)
Oklahoma	714 (43)
eGFR, *median (IQR)*	80.47 (70.56, 92.61)
Prior TB diagnosis, *n (%)*	223 (13)
FEV1 (L), *median (IQR)*	2.47 (2.04, 3.03)
FVC (L), *median (IQR)*	3.27 (2.69, 4.02)
FEV1/FVC, *median (IQR)*	76.5 (71.1, 75.3)

Abbreviations: eGFR, estimated glomerular filtration rate; TB, tuberculosis.

Figure 2. Venn diagram of elastic-net selected DMPs.



A) Venn diagram of lung function outcomes. B) Venn diagram of airflow limation C) Venn diagram of restrictive pattern.

Table 2. Top five Differentially Methylated Positions for continuous lung function measures (FEV1, FVC, and FEV1/FVC)

CpG	Chr	Gene	Function	MD (95% CI)	P-value
FEV1					
cg25325512	6	*PIM1*	Serine/Threonine-Protein Kinase Pim-1. Cell proliferation and survival	20.9 (-33.5, -8.3)	7.40E-09
cg26058502	1	*CERS2*	Regulation of cell growth and lipid metabolism	-27.8 (-43.6, -11.9)	1.14E-07
cg01641754	5	*LOC100289230*	Uncharacterized	17.9 (8.8, 27.0)	2.66E-07
cg15167811	9	*PTBP3*	Regulator of cell differentiation	-8.3 (-14.5, -2.0)	3.86E-07
cg07941411	3	*CD80*	T-cell proliferation and cytokine production	-7.1 (-11.2, -2.9)	8.75E-07
FVC					
cg02203833	7	*CDK5*	Serine/threonine kinase. Synaptic plasticity and neuronal migration	2.3 (1.3, 3.2)	4.34E-06
cg19265480	1	*NBPF8*	Associated with developmental and neurogenetic diseases	-0.9 (-1.3, -0.5)	5.86E-06
cg18140268	1	*NBPF8*	Associated with developmental and neurogenetic diseases	-0.7 (-1.0, -0.4)	5.93E-06
cg07343418	12	*GRIN2B*	Brain development, synaptic plasticity	-1.0 (-1.4, -0.6)	7.05E-06
cg03725414	16	*CRISPLD2*	Promotes matrix assembly	2.0 (1.1, 2.9)	7.91E-06
FEV1/FVC					
cg25001882	14	*NRXN3*	Cell adhesion in the nervous system, synaptic plasticity	33.0 (21.3, 44.6)	3.32E-08
cg16771344	9	*NTRK2*	Tyrosine receptor kinase. Phosphorylates itself and members of the MAPK signaling pathway	29.9 (18.5, 41.2)	2.99E-07
cg12420787	2	*CACNB4*	Calcium channel. Plasticity on the brain	21.7 (13.4, 30.1)	3.84E-07
cg03636183	19	*F2RL3*	Blood coagulation, inflammation and response to pain. Associated with smoking and lung cancer	18.1 (11.0, 25.2)	5.80E-07
cg27127773	16	*SLX4*	Repair of specific types of DNA lesions	-42.6 (-59.2, -25.9)	6.17E-07

Top five DMPs selected by elastic-net models. Mean differences reported from linear models comparing p90 vs p10 of methylation. Models were adjusted for age, sex, BMI, smoking status (never, former, current), cumulative smoking (pack-years), study center (Arizona, Oklahoma, or North and South Dakota), cell counts (CD8T, CD4T, NK, B cells and monocytes) and five genetic PCs.

Table 3. Top 10 Differentially Methylated Positions for Airflow Limitation and Restrictive Pattern Lung Function

CpG	Chr	Gene	Function	OR (95% CI)	P-value
Airflow Limitation					
cg20304857	2	*GALNT14*	Catalyze transfer of N-acetyl-D-galactosamin to hydroxyl groups on serines and threonines	2.8 (1.9, 4.3)	0.000001
cg07842459	1	*CD84*	Homophilic adhesion molecule	2.8 (1.8, 4.3)	0.000002
cg17916980	16	*CPPED1*	Serine/Threonine-Protein Phosphatase. Dephosphorylates AKT family kinase	2.5 (1.7, 3.7)	0.000002
cg03647068	2	*FMNL2*	Morphogenesis, cytokinesis, cell polarity	3.0 (1.9, 4.8)	0.000002
cg06100532	16	*LMF1*	Involved in the maturation of specific proteins in the endoplasmic reticulum	2.8 (1.8, 4.3)	0.000003
Restrictive Pattern					
cg05504535	10	*ADARB2*	Regulatory role in RNA editing	0.3 (0.2, 0.5)	1.86E-07
cg04890495	11	*MCAM*	Cell adhesion, cohesion of the endothelial monolayer at intercellular junctions in vascular tissue	0.3 (0.2, 0.5)	5.92E-07
cg03320255	19	*ZNF540*	Transcriptional repressor	0.3 (0.2, 0.5)	8.72E-07
cg20024687	8	*TOP1MT*	Role in the modification of DNA topology	0.4 (0.2, 0.5)	3.00E-06
cg19693031	1	*TXNIP*	Protects cells from oxidative stress	0.4 (0.3, 0.6)	4.00E-06

Top 10 DMPs selected by elastic-net models. Odds Ratios reported from regression models comparing p90 vs p10 of methylation. Models were adjusted for age, sex, BMI, smoking status (never, former, current), cumulative smoking (pack-years), study center (Arizona, Oklahoma, or North and South Dakota), cell counts (CD8T, CD4T, NK, B cells and monocytes) and five genetic PCs.

Figure 5. Workflow for Protein-Protein Interaction Networks

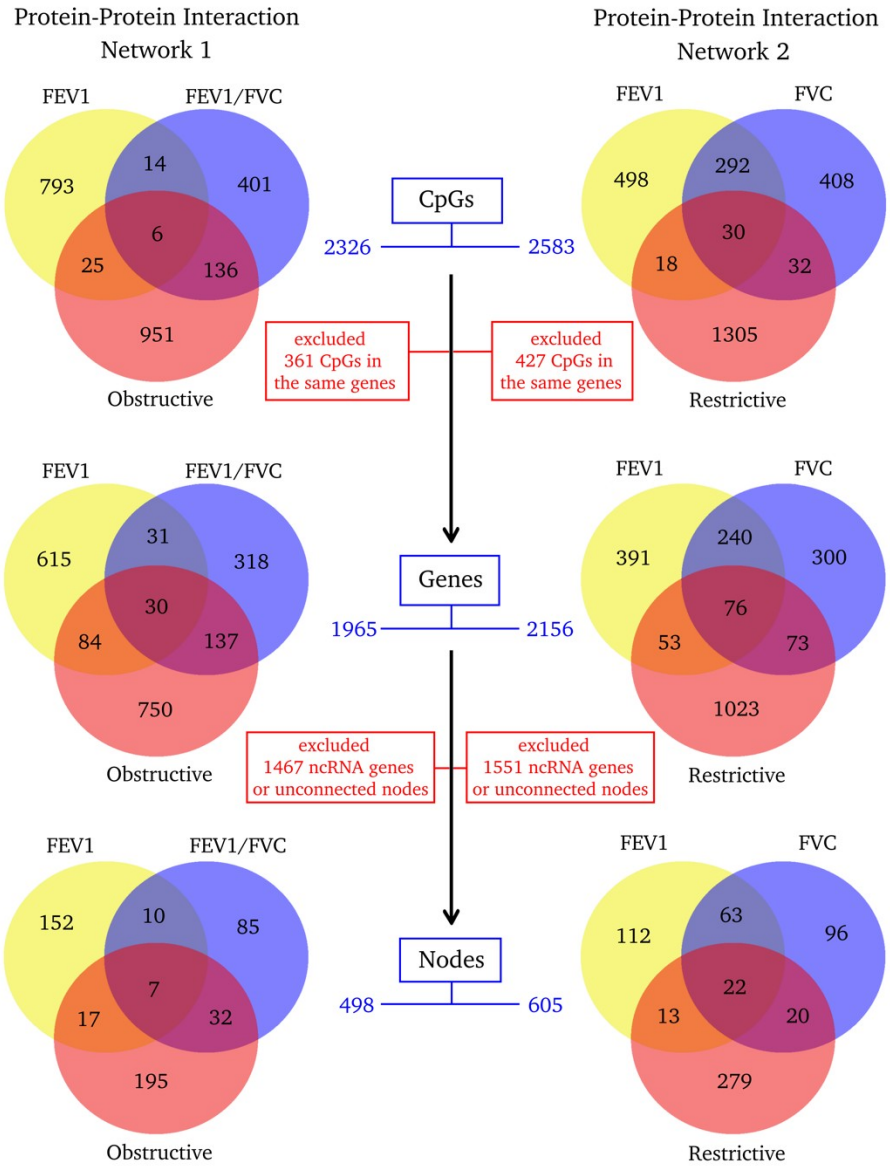


Figure 6. Protein-protein interaction network for obstructive lung function phenotype: FEV1, FEV1/FVC and obstructive vs normal lung function

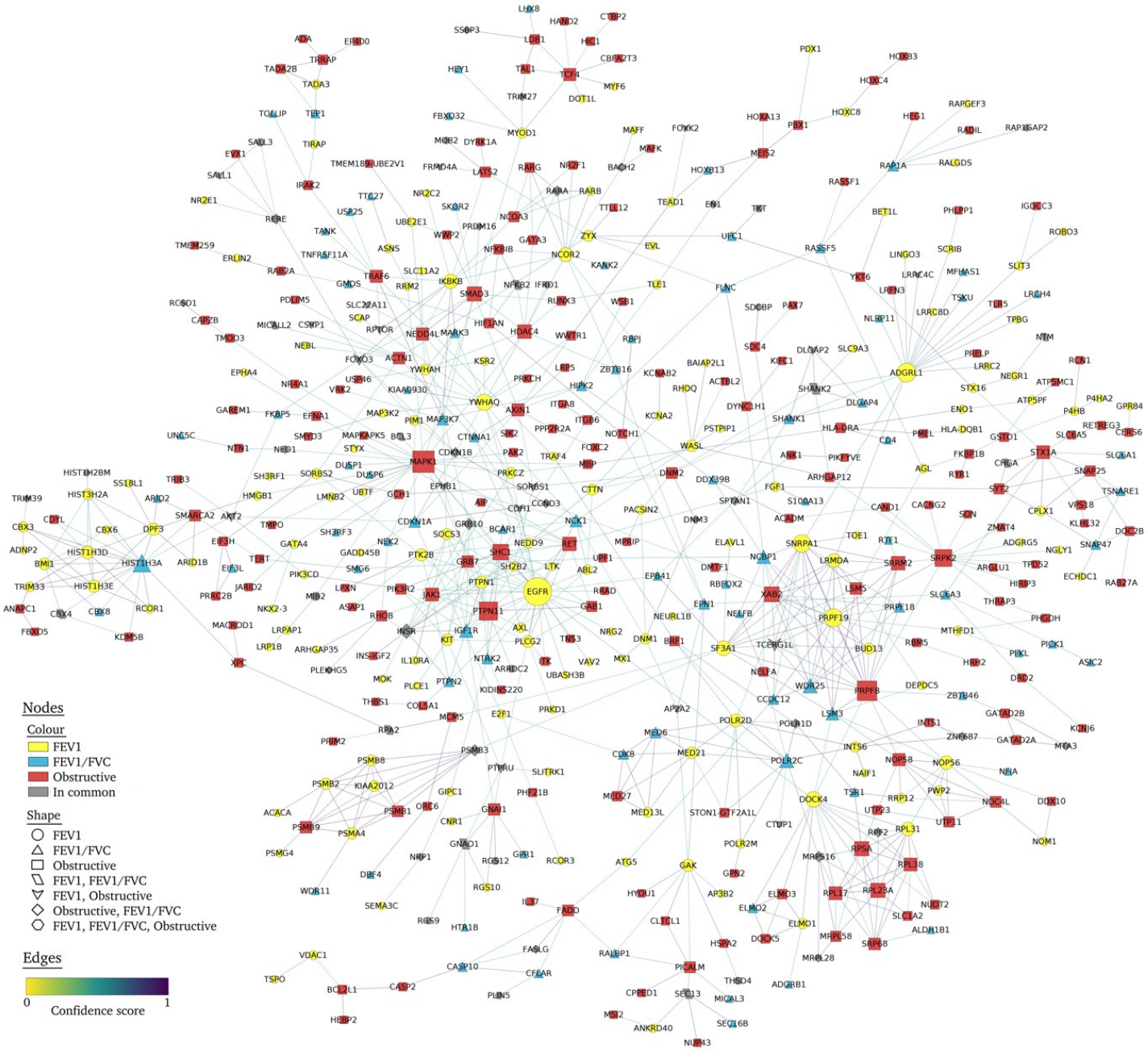
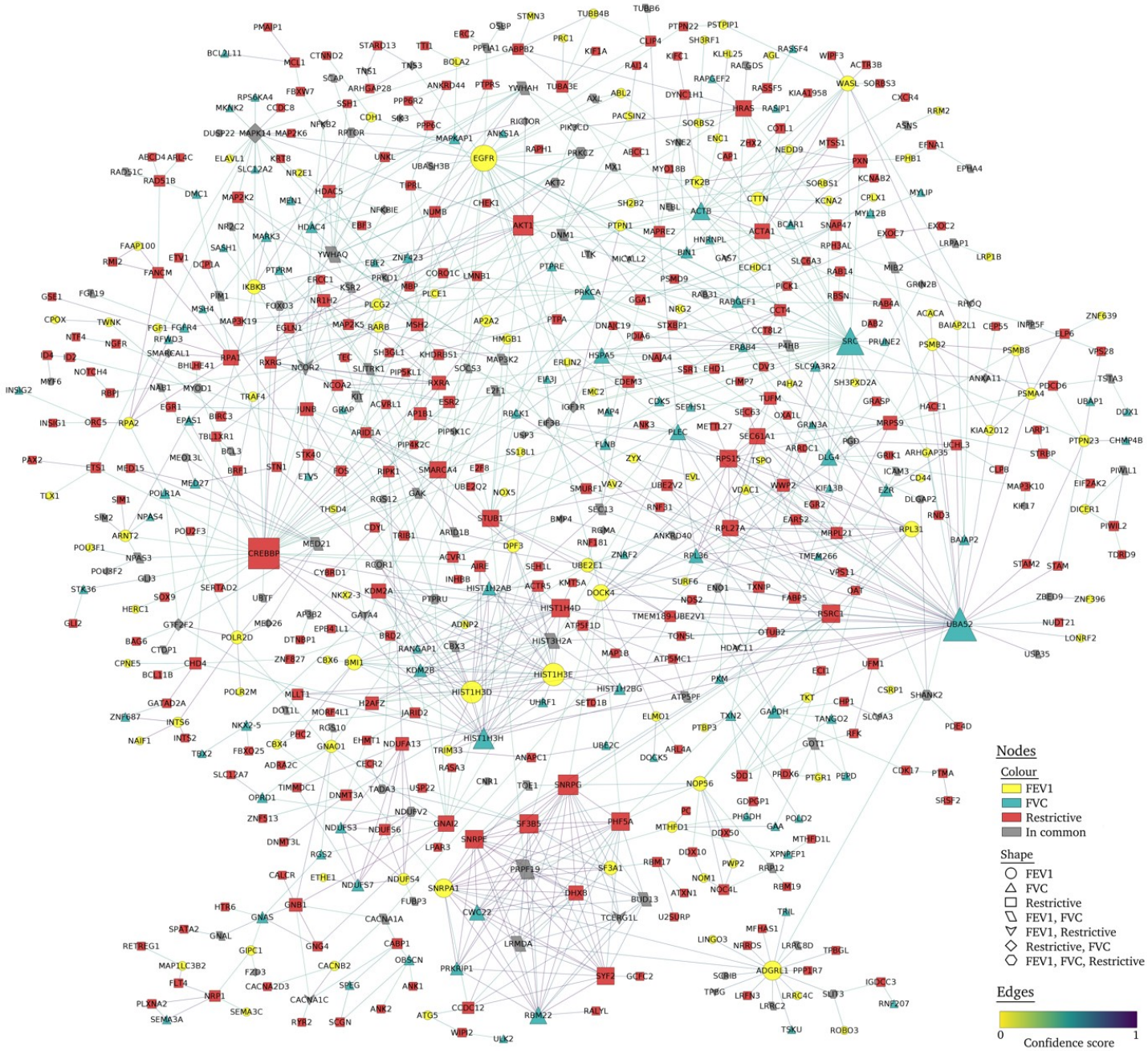


Figure 7. Protein-protein interaction networks for restrictive lung function phenotype: FEV1, FVC and restrictive vs normal lung function



References:

1. Dwyer-Lindgren L, Bertozzi-Villa A, Stubbs RW, et al. Trends and Patterns of Differences in Chronic Respiratory Disease Mortality Among US Counties, 1980-2014. *JAMA*. 2017;318(12):1136. doi:10.1001/jama.2017.11747
2. Vestbo J, Hurd SS, Agustí AG, et al. Global Strategy for the Diagnosis, Management, and Prevention of Chronic Obstructive Pulmonary Disease. *Am J Respir Crit Care Med*. 2013;187(4):347-365. doi:10.1164/rccm.201204-0596PP
3. Rosas IO, Dellaripa PF, Lederer DJ, Khanna D, Young LR, Martinez FJ. Interstitial Lung Disease: NHLBI Workshop on the Primary Prevention of Chronic Lung Diseases. *Ann Am Thorac Soc*. 2014;11(Supplement 3):S169-S177. doi:10.1513/AnnalsATS.201312-429LD
4. Yang I V., Schwartz DA. Epigenetic control of gene expression in the lung. *Am J Respir Crit Care Med*. 2011;183(10):1295-1301. doi:10.1164/rccm.201010-1579PP
5. Mortaz E, Masjedi MR, Barnes PJ, Adcock IM. *Epigenetics and Chromatin Remodeling Play a Role in Lung Disease*.; 2011.
6. Helling BA, Yang I V. Epigenetics in lung fibrosis: From pathobiology to treatment perspective. *Curr Opin Pulm Med*. 2015;21(5):454-462. doi:10.1097/MCP.0000000000000191
7. Adcock IM, Tsaprouni L, Bhavsar P, Ito K. Epigenetic regulation of airway inflammation. *Curr Opin Immunol*. 2007;19(6):694-700. doi:10.1016/j.coi.2007.07.016
8. Berndt A, Leme AS, Shapiro SD. Emerging genetics of COPD. *EMBO Mol Med*. 2012;4(11):1144-1155. doi:10.1002/emmm.201100627
9. Machin M, Amaral AFS, Wielscher M, et al. Systematic review of lung function and COPD with peripheral blood DNA methylation in population based studies. *BMC Pulm Med*. 2017;17(1):54. doi:10.1186/s12890-017-0397-3
10. Laurell C-B, Eriksson S. The Electrophoretic α 1 -Globulin Pattern of Serum in α 1 -Antit-

rypsin Deficiency. *Scand J Clin Lab Invest*. 1963;15:132-140. doi:10.3109/15412555.2013.771956

11. Lee ET, Welty TK, Fabsitz R, et al. The strong heart study a study of cardiovascular disease in american indians: design and methods. *Am J Epidemiol*. 1990;132(6):1141-1155. doi:10.1093/oxfordjournals.aje.a115757
12. Domingo-Relloso A, Riffo-Campos AL, Haack K, et al. Cadmium, Smoking, and Human Blood DNA Methylation Profiles in Adults from the Strong Heart Study. *Environ Health Perspect*. 2020;128(6):067005. doi:10.1289/EHP6345
13. ATS. Standardization of Spirometry, 1994 Update. American Thoracic Society. *Am J Respir Crit Care Med*. 1995;152(3):1107-1136. doi:10.1164/ajrccm.152.3.7663792
14. Marion MS, Leonardson GR, Rhoades ER, Welty TK, Enright PL. Spirometry reference values for American Indian adults: results from the Strong Heart Study. *Chest*. 2001;120(2):489-495. doi:10.1378/chest.120.2.489
15. Aryee MJ, Jaffe AE, Corrada-Bravo H, et al. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics*. 2014;30(10):1363-1369. doi:10.1093/bioinformatics/btu049
16. Salas LA, Koestler DC, Butler RA, et al. An optimized library for reference-based deconvolution of whole-blood biospecimens assayed using the Illumina HumanMethylationEPIC BeadArray. *Genome Biol*. 2018;19(1):64. doi:10.1186/s13059-018-1448-7
17. Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics*. 2012;28(6):882-883. doi:10.1093/bioinformatics/bts034
18. Illumina Inc. Infinium MethylationEPIC Product Files.
19. Fortin J-P, Triche TJ, Hansen KD. Preprocessing, normalization and integration of the Illumina HumanMethylationEPIC array with minfi. *Bioinformatics*. 2016;33(4):btw691. doi:10.1093/bioinformatics/btw691

20. Benton MC, Sutherland HG, Macartney-Coxson D, Haupt LM, Lea RA, Griffiths LR. Methylome-wide association study of whole blood DNA in the Norfolk Island isolate identifies robust loci associated with age. *Aging (Albany NY)*. 2017;9(3):753-768. doi:10.18632/aging.101187

21. Abraham G, Kowalczyk A, Zobel J, Inouye M. SparSNP: Fast and memory-efficient analysis of all SNPs for phenotype prediction. *BMC Bioinformatics*. 2012;13(1). doi:10.1186/1471-2105-13-88

22. Waldmann P, Mészáros G, Gredler B, Fuerst C, Sölkner J. Evaluation of the lasso and the elastic net in genome-wide association studies. *Front Genet*. 2013;4. doi:10.3389/fgene.2013.00270

23. Fan J, Lv J. Sure independence screening for ultrahigh dimensional feature space. *J R Stat Soc Ser B (Statistical Methodol*. 2008;70(5):849-911. doi:10.1111/j.1467-9868.2008.00674.x

24. Hasinur M, Khan R, Ewart J, Shaw H. *VARIABLE SELECTION FOR SURVIVAL DATA WITH A CLASS OF ADAPTIVE ELASTIC NET TECHNIQUES*.; 2013.

25. Zou H, Hao A, Zhang H. ON THE ADAPTIVE ELASTIC-NET WITH A DIVERGING NUMBER OF PARAMETERS. *Ann Stat*. 2009;37(4):1733-1751. doi:10.1214/08-AOS625

26. Liu J, Liang G, Siegmund KD, Lewinger JP. Data integration by multi-tuning parameter elastic net regression. *BMC Bioinformatics*. 2018;19(1). doi:10.1186/s12859-018-2401-1

27. Zhang Q, Vallerga CL, Walker RM, et al. Improved prediction of chronological age from DNA methylation limits it as a biomarker of ageing. *bioRxiv*. Published online May 2018:327890. doi:10.1101/327890

28. Barfield RT, Almli LM, Kilaru V, et al. Accounting for Population Stratification in DNA Methylation Studies. *Genet Epidemiol*. 2014;38(3):231. doi:10.1002/GEPI.21789

29. Szklarczyk D, Gable AL, Lyon D, et al. STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experi-

mental datasets. *Nucleic Acids Res*. 2019;47(D1):D607-D613. doi:10.1093/nar/gky1131

30. Shannon P, Markiel A, Ozier O, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*. 2003;13(11):2498-2504. doi:10.1101/gr.1239303

31. Defnet AE, Hasday JD, Shapiro P. Kinase inhibitors in the treatment of obstructive pulmonary diseases. *Curr Opin Pharmacol*. 2020;51:11-18. doi:10.1016/j.coph.2020.03.005

32. Dempsey EC, Cool CD, Littler CM. Lung disease and PKCs. *Pharmacol Res*. 2007;55(6):545-559. doi:10.1016/j.phrs.2007.04.010

33. Li L-F, Liao S-K, Huang C-C, Hung M-J, Quinn DA. Serine/threonine kinase-protein kinase B and extracellular signal-regulated kinase regulate ventilator-induced pulmonary fibrosis after bleomycin-induced acute lung injury: a prospective, controlled animal experiment. *Crit Care*. 2008;12(4):R103. doi:10.1186/cc6983

34. Birukova AA, Birukov KG, Gorshkov B, Liu F, Garcia JGN, Verin AD. MAP kinases in lung endothelial permeability induced by microtubule disassembly. *Am J Physiol Lung Cell Mol Physiol*. 2005;289(1):L75-84. doi:10.1152/ajplung.00447.2004

35. Schweitzer KS, Hatoum H, Brown MB, et al. Mechanisms of lung endothelial barrier disruption induced by cigarette smoke: role of oxidative stress and ceramides. *Am J Physiol Lung Cell Mol Physiol*. 2011;301(6):L836-46. doi:10.1152/ajplung.00385.2010

36. Magnani ND, Dada LA, Sznajder JI. Ubiquitin-proteasome signaling in lung injury. *Transl Res*. 2018;198:29-39. doi:10.1016/j.trsl.2018.04.003

37. Chunn JL, Molina JG, Mi T, Xia Y, Kellems RE, Blackburn MR. Adenosine-dependent pulmonary fibrosis in adenosine deaminase-deficient mice. *J Immunol*. 2005;175(3):1937-1946. doi:10.4049/jimmunol.175.3.1937

38. Kodal JB, Kobylecki CJ, Vedel-Krogh S, Nordestgaard BG, Bojesen SE. AHRR hypomethylation, lung function, lung function decline and respiratory symptoms. *Eur Respir J*. 2018;51(3). doi:10.1183/13993003.01512-2017

39. Bojesen SE, Timpson N, Relton C, Davey Smith G, Nordestgaard BG. AHRR (cg05575921) hypomethylation marks smoking behaviour, morbidity and mortality. *Thorax*. 2017;72(7):646-653. doi:10.1136/thoraxjnl-2016-208789

40. Carmona JJ, Barfield RT, Panni T, 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(10-11):1039-1055. doi:10.1080/15592294.2018.1529849

41. Imboden M, Wielscher M, Rezwan FI, et al. Epigenome-wide association study of lung function level and its change. *Eur Respir J*. 2019;54(1). doi:10.1183/13993003.00457-2019

42. Marioni RE, Shah S, McRae AF, et al. The epigenetic clock is correlated with physical and cognitive fitness in the Lothian Birth Cohort 1936. *Int J Epidemiol*. 2015;44(4):1388-1396. doi:10.1093/ije/dyu277

43. Bell JT, Tsai P-C, Yang T-P, et al. Epigenome-Wide Scans Identify Differentially Methylated Regions for Age and Age-Related Phenotypes in a Healthy Ageing Population. Li J, ed. *PLoS Genet*. 2012;8(4):e1002629. doi:10.1371/journal.pgen.1002629

44. Mukherjee N, Arathimos R, Chen S, et al. DNA methylation at birth is associated with lung function development until age 26 years. *Eur Respir J*. 2021;57(4):2003505. doi:10.1183/13993003.03505-2020

45. Casas-Recasens S, Noell G, Mendoza N, et al. Lung DNA Methylation in Chronic Obstructive Pulmonary Disease: Relationship with Smoking Status and Airflow Limitation Severity. *Am J Respir Crit Care Med*. 2021;203(1):129-134. doi:10.1164/rccm.201912-2420LE

46. Hernandez Cordero AI, Yang CX, Obeidat M, et al. DNA methylation is associated with airflow obstruction in patients living with HIV. *Thorax*. 2021;76(5):448-455. doi:10.1136/thoraxjnl-2020-215866

47. de Vries M, Nedeljkovic I, van der Plaat DA, et al. DNA methylation is associated with lung function in never smokers. *Respir Res*. 2019;20(1):268. doi:10.1186/s12931-019-1222-8