Gene Promoter Hypermethylation Detected in Sputum Predicts FEV₁ Decline and All-Cause Mortality in Smokers

Shuguang Leng^{1,2,3}, Brenda Diergaarde^{4,5,6}, Maria A. Picchi¹, David O. Wilson⁷, Frank D. Gilliland⁸, Jian-Min Yuan^{9,10}, Jill M. Siegfried¹¹, and Steven A. Belinsky^{1,2,12}

¹Lung Cancer Program, Lovelace Respiratory Research Institute, Albuquerque, New Mexico; ²Cancer Control Research Program and ¹²Cancer Genetics and Epigenetics Program, University of New Mexico Comprehensive Cancer Center, Albuquerque, New Mexico; ³School of Public Health, Qingdao University, Qingdao, Shandong, China; ⁴Department of Human Genetics and ⁹Department of Epidemiology, Graduate School of Public Health, and ⁷Division of Pulmonary, Allergy, and Critical Care Medicine, Department of Medicine, University of Pittsburgh, Pittsburgh, Pennsylvania; ⁵Cancer Epidemiology and Prevention Program, ⁶Lung Cancer Program, and ¹⁰Division of Cancer Control and Population Sciences, University of Pittsburgh Cancer Institute, Pittsburgh, Pennsylvania; ⁸Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, California; and ¹¹Department of Pharmacology, University of Minnesota, Minneapolis, Minnesota

Abstract

Rationale: Gene promoter hypermethylation detected in sputum assesses the extent of field cancerization and predicts lung cancer (LC) risk in ever-smokers. A rapid decline of FEV₁ is a major driver for development of airway obstruction.

Objectives: To assess the effects of methylation of 12 genes on FEV₁ decline and of FEV₁ decline on subsequent LC incidence using two independent, longitudinal cohorts (i.e., LSC [Lovelace Smokers Cohort] and PLuSS [Pittsburgh Lung Screening Study]).

Methods: Gene methylation was measured in sputum using two-stage nested methylation-specific PCR. The linear mixed effects model was used to assess the effects of studied variables on FEV₁ decline.

Measurements and Main Results: A dose-dependent relationship between number of genes methylated and FEV₁ decline was identified, with smokers with three or more methylated genes having 27.8% and 10.3% faster FEV₁ decline than smokers with zero to two methylated genes in the LSC and PLuSS cohort, respectively (all $P < 0.01$). High methylation in sputum was associated with a shorter latency for LC incidence (log-rank $P = 0.0048$) and worse all-cause mortality (log-rank $P < 0.0001$). Smokers with subsequent LC incidence had a more rapid annual decline of FEV₁ (by 5.2 ml, $P = 0.038$) than smoker control subjects.

Conclusions: Gene methylation detected in sputum predicted FEV₁ decline, LC incidence, and all-cause mortality in smokers. Rapid FEV₁ decline may be a risk factor for LC incidence in smokers, which may explain a greater prevalence of airway obstruction seen in patients with LC.

Keywords: epigenetics; spirometry; longitudinal studies; lung neoplasms; cause of death

Lung cancer (LC) is the leading cause of cancer-related mortality in the United States (1), and chronic obstructive pulmonary disease (COPD) is the third leading cause of death next to cardiovascular and malignant

diseases (2). COPD is an established risk factor for LC (2, 3). Moreover, COPD and LC share etiological factors, such as age, cigarette smoking, exposure to air pollution, and genetic susceptibility variants (2, 4, 5).

The development of COPD and LC in ever-smokers (i.e., current and former smokers) likely requires decades of repetitive exposure of airways to cigarette smoke. Extensive field cancerization,

(Received in original form August 20, 2017; accepted in final form February 12, 2018)

Supported by National Cancer Institute grants R01 CA097356, R01 CA164782, P50 CA090440, P30 CA047904, and P30 CA118100 and by the State of New Mexico as a direct appropriation from the Tobacco Settlement Fund. The NIH and the State of New Mexico had no influence in study design; collection, analysis, and interpretation of data; in writing the report; or in the decision to submit the report for publication.

Author Contributions: S.L., J.M.S., and S.A.B. conceived of and designed the study; B.D., M.A.P., D.O.W., and J.-M.Y. performed the data collection and management; S.L. conducted the data analyses and tabulated the results; S.L. and B.D. interpreted the results and drafted the manuscript; and S.L., B.D., F.D.G., J.-M.Y., J.M.S., and S.A.B. critically edited the manuscript. All authors have read the manuscript and approved its submission.

Correspondence and requests for reprints should be addressed to Shuguang Leng, M.D., Ph.D., Lung Cancer Program, Lovelace Respiratory Research Institute, Albuquerque, NM 87108. E-mail: sleng@lrii.org.

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

Am J Respir Crit Care Med Vol 198, Iss 2, pp 187–196, Jul 15, 2018

Copyright © 2018 by the American Thoracic Society

Originally Published in Press as DOI: 10.1164/rccm.201708-1659OC on February 13, 2018

Internet address: www.atsjournals.org

At a Glance Commentary

Scientific Knowledge on the

Subject: Several studies identified associations between methylation of blood leukocyte DNA and prevalent chronic obstructive pulmonary disease, FEV₁, or FEV₁ decline. However, because of cell type-specific epigenetic regulation, methylation assessed in leukocyte DNA may not reflect methylation in the lungs. Promoter hypermethylation of cancer-associated genes detected in sputum assesses the extent of field cancerization and is a validated risk biomarker for lung cancer (LC) in ever-smokers (i.e., current and former smokers).

What This Study Adds to the

Field: Using two geographically independent, longitudinal smoker cohorts, we identified and replicated an association between high methylation (three or more vs. zero to two genes methylated) in sputum at baseline and FEV₁ decline. Gene methylation also predicted subsequent LC incidence and all-cause mortality in a prospective setting. Smokers who developed LC had a more rapid decline of FEV₁ (by 5.2 ml/yr) before LC incidence than smokers who remained LC free. Thus, in addition to being a validated LC risk biomarker, this methylation panel may also serve as a biomarker associated with predisposition for cigarette smoke-induced accelerated aging of overall lung health in ever-smokers.

characterized as the acquisition of genetic and epigenetic changes in oncogenes and tumor suppressor genes throughout the lung epithelium, can be detected in the lungs of smokers (6–8). Transcriptional silencing by promoter hypermethylation affects genes involved in all aspects of normal cell function and rivals genetic changes that affect the coding sequence as a crucial trigger for neoplastic development and progression (6). The detection of gene methylation in exfoliated lung cells present in sputum provides an assessment of the extent of field cancerization and is a validated biomarker for LC risk in smokers (6, 9–11).

Several studies have assessed the associations between methylation of blood leukocyte DNA and lung function (12–14). An epigenome-wide association study identified 349 CpGs whose methylation was associated with the presence and severity of COPD in members from the International COPD Genetics Network and the Boston Early-Onset COPD Study (12). Moreover, these differentially methylated CpGs were enriched in the immune and inflammatory system pathways, responses to stress, and external stimuli, as well as wound healing and coagulation cascades. In the Normative Aging Study, methylation of the promoter of candidate genes (*CRAT*, *F3*, *TLR2*, *IFN γ* , and *IL6*) and repetitive Alu elements was associated with lung function (13, 14). These findings provide some evidence that COPD as a systemic disease can affect the methylome in circulating blood cells as an epigenetic readout of the aggregate inflammatory impact of advanced lung disease and smoking. To date, only one study assessed the association of DNA methylation with lung function decline. Using the Normative Aging Study, long interspersed element-1 hypomethylation was associated with a more rapid lung function decline in 301 individuals who had a second spirometry measurements collected an average of 4 years after baseline (13). However, because of cell type-specific epigenetic regulation, DNA methylation assessed in peripheral blood cells may not reflect DNA methylation in the lungs and thus may not be an informative marker of a causal relationship between gene methylation and lung diseases. Several studies have identified gene methylation in exfoliated lung cells collected in sputum associated with different COPD subphenotypes (e.g., FEV₁% predicted, chronic mucus hypersecretion, and airway obstruction) in the LSC (Lovelace Smokers Cohort) using cross-sectional design (5, 15). However, no studies have evaluated the impact of gene methylation in sputum and lung function change over time.

Acceleration of age-related decline in lung function, as assessed by the FEV₁, due to cigarette smoke is a major determinant of COPD pathogenesis. In this study, we assessed the association between the methylation status of 12 genes in sputum selected based on their association with LC and COPD subphenotypes and FEV₁ decline in 1,111 and 447 non-Hispanic white (NHW) ever-smokers from two

geographically independent longitudinal cohorts (i.e., LSC and PLuSS [Pittsburgh Lung Screening Study]). Moreover, the associations between gene methylation in sputum and subsequent LC incidence and all-cause mortality were assessed in the PLuSS. Finally, the association between FEV₁ decline and risk for subsequent LC incidence was assessed in the PLuSS.

Methods

LSC

The LSC has been actively enrolling ever-smokers from the Albuquerque, New Mexico metropolitan area since 2001 to conduct longitudinal studies of risk or diagnostic biomarkers for LC and other chronic pulmonary diseases. Detailed introduction of study protocols was available in references (16, 17) and supplemental material. Among the more than 2,500 LSC members enrolled to date, 1,111 NHW smokers with two or more longitudinal spirometry tests and complete methylation data for the 12 evaluated genes in baseline sputum samples were included in this study.

PLuSS

The PLuSS cohort was established in 2002 to support translational studies of the University of Pittsburgh Lung Cancer Specialized Programs of Research Excellence. Detailed introduction of study protocols is available in Reference 18 and the online supplement. Among the 3,638 PLuSS members, subjects ($n = 1,151$) who were in the highest tertile of LC risk on the basis of age and smoking history were enrolled in the sputum collection protocol, among whom 447 NHW smokers with two or more longitudinal spirometry tests after first sputum collection and complete methylation data for the 12 evaluated genes were included in this study.

Gene Promoter Methylation in Sputum

Twelve genes were selected for methylation analysis in cytologically adequate sputum samples on the basis of their prevalence in lung tumors ($\geq 25\%$), diversity of function (e.g., cell cycle, DNA repair, inflammation, apoptosis, rat sarcoma signaling and invasion), timing for inactivation during LC development when known, causality for tumor initiation (19–22), our previous

studies establishing their association with risk for LC and COPD subphenotypes (e.g., FEV₁% predicted, chronic mucus hypersecretion, and airway obstruction), and their specificity to methylation in lung epithelial cells (the 12 genes are listed in Table E1 in the online supplement) (5, 10, 15). In addition, integrating the composite methylation index (MI) of these 12 genes detected in sputum in the LC risk assessment model provided the most optimal discrimination by improving the prediction accuracy by 15% (10, 11). Given the low percentage (<3%) of lung epithelial cells in sputum samples, which also varied significantly between individuals, a two-stage nested methylation-specific PCR that can reproducibly detect one methylated allele in a background of 10,000 unmethylated alleles was used to detect methylated alleles (9), and methylation status for each individual gene was scored as 0 (unmethylated) or 1 (methylated). The genes and primer sequences for methylation-specific PCR are listed in our previous publication (23). A composite MI was created that summed the number of genes methylated based on the 12 genes to assess its association with outcomes.

Statistical Analysis

First, the linear mixed effects (LME) model with a subject-specific random intercept was used to associate MI of the 12 genes with repeated measures of FEV₁ to estimate the effect of MI on FEV₁ decline by including an interaction term between time and MI in the LME model. Variable “time” is the number of years between first sputum collection and each spirometry. Fixed effects of baseline values of FEV₁, age, sex, height, body mass index (BMI), smoking status and pack-years, and time between first sputum collection and each spirometry as a time-varying variable were included in the model. Baseline FEV₁ was included as an independent variable in this analysis to control for the interindividual difference in baseline lung function. Our analyses were confined to NHWs, because Hispanics had higher risk for gene methylation in sputum and slower FEV₁ decline than NHW individuals in the LSC (23, 24), and NHWs are dominant (>97%) in the PLuSS cohort, although inclusion of Hispanics in the analysis did not change the results observed in NHWs in the LSC (Table E2). Moreover, this analysis was conducted in cohort members with two or more FEV₁

measurements. Analyses were also conducted by including subjects with only one FEV₁ measurement and identified associations almost identical to those seen in subjects with two or more FEV₁ measurements (not shown).

Second, the Cox proportional hazards model was used to assess the effect of promoter methylation status of the 12 evaluated genes on LC risk and all-cause mortality in the PLuSS cohort. Finally, the LME model was used to assess whether PLuSS members with subsequent LC incidence had a more rapid FEV₁ decline before diagnosis. The first analysis was conducted in 124 cases with LC with two or more spirometry measurements collected before diagnosis and 1,068 smoker control subjects with two or more spirometry measurements without subsequent LC incidence (Table E3), all from the PLuSS cohort. However, because PLuSS subjects with subsequent LC incidence were more often men and continuous current smokers and had higher pack-years and lower FEV₁ at baseline than the subjects without (Table E3), and these four factors were also associated with FEV₁ decline (not shown), we were concerned whether the association between FEV₁ decline and subsequent LC incidence was confounded by the uneven distribution of these four factors between the two study groups. Thus, we further used a random sampling-based approach to create 1,000 control groups, with each group having 549 subjects who were frequency matched to the 124 cases by pack-years (9.2–60 and 60–120 pack-years), and FEV₁ (0.65–2.295 and 2.295–4.41 L) collected at study entry, sex, and smoking behavior change. The ratio was specifically determined by available sample size within each subgroup that could provide the largest power for the analysis from each round of random sampling. This approach completely removed the potential confounding effects from pack-years and FEV₁ at baseline, sex, and smoking behavior change. Such random selection was conducted for 1,000 times, and for each selection, the LME model was run to assess whether PLuSS members with subsequent LC incidence had a more rapid FEV₁ decline before diagnosis. There were a total of 1,000 estimates, SEs, and *P* values generated from the 1,000 random samplings. On the basis of the results from 1,000 random samplings, the association with the median

P value was identified and the estimate, SE, and *P* value from this association was used as the representative for the 1,000 associations. Directed acyclic graphs were drawn to assess the relationship among study variables and outcomes (25), and no evidence of inappropriate adjustment of covariates was identified in all major analyses (Figure E1). All statistical analyses were conducted using SAS 9.4 (SAS Institute Inc.).

Results

This study included 1,111 and 447 NHW smokers with available information on the methylation status of all 12 evaluated genes and longitudinal spirometry measurements from the LSC and PLuSS, respectively (Table 1). Study subjects from the PLuSS were more often men and current smokers, were older, and had a higher number of pack-years, longer follow-up time, and more spirometry measurements than subjects from the LSC. Accordingly, study subjects from the PLuSS had lower spirometry measurements at baseline and greater decline rate than the LSC. The MI and individual gene methylation prevalence were comparable between the two cohorts, except for *JPH3* and *P16* (Tables 1 and E1).

Association between MI and FEV₁ Decline

We first assessed the shape of the dose-response for the association between MI as an ordinal variable and FEV₁ decline in the LSC because of its abundant sample size. The results suggested a more heterogeneous effect; that is, the most rapid decline appears to occur when three or more genes are methylated (Table 2). Thus, we converted MI into a binary variable (three or more vs. zero to two genes methylated) and identified that subjects with three or more genes methylated had much faster FEV₁ decline (by 5.8 ± 1.4 ml/yr, $P < 0.0001$) than subjects with zero to two methylated genes in the LSC (Table 2 and Figure 1A). The result on the basis of MI defined as a binary variable is further replicated in the PLuSS cohort, with three or more methylated genes associated with faster FEV₁ decline by 5.0 ± 2.4 ml/yr ($P = 0.035$) (Table 2 and Figure 1B). The association between MI defined as a binary variable and FEV₁/FVC decline was not significant in either the LSC ($P = 0.93$) or the PLuSS cohorts ($P = 0.20$) (Table E4).

Table 1. Characteristics of Study Populations in the Lovelace Smokers Cohort and Pittsburgh Lung Screening Study

Characteristic	LSC	PLuSS
N	1,111	447
Baseline variables*		
Age, yr, mean $\pm$ SD	57.6 $\pm$ 9.6	65.3 $\pm$ 5.7
Sex, male, n (%)	310 (23.0)	195 (43.6)
Ethnicity, NHW, n (%)	1,111 (100)	447 (100)
Current smoker, n (%)	577 (51.9)	295 (66.0)
Pack-years, median (Q1–Q3)	36.0 (26.8–50.0)	59 (48–74)
Methylation index, n (%)		
0	212 (19.1)	119 (26.6)
1	247 (22.2)	100 (22.4)
2	218 (19.6)	88 (19.7)
3	154 (13.9)	62 (13.9)
4	112 (10.1)	34 (7.6)
5	68 (6.1)	16 (3.6)
6	45 (4.1)	11 (2.5)
7	29 (2.6)	10 (2.2)
8	17 (1.5)	6 (1.3)
9	6 (0.5)	1 (0.2)
10	3 (0.3)	0
Baseline spirometry, mean $\pm$ SD		
FEV ₁ , L	2.6 $\pm$ 0.8	2.1 $\pm$ 0.7
FVC, L	3.6 $\pm$ 1.0	3.1 $\pm$ 0.9
FEV ₁ /FVC, %	72.2 $\pm$ 10.4	67.4 $\pm$ 11.4
Longitudinal data		
No. of spirometry tests, median (Q1–Q3)	5 (3–8)	8 (4–10)
Follow-up period, yr, median (Q1–Q3)	3.4 (1.4–7.1)	7 (4–9)
Decline rate, [†] mean $\pm$ SD		
FEV ₁ , ml/yr	–22.6 $\pm$ 0.6	–50.2 $\pm$ 1.1
FVC, ml/yr	–10.4 $\pm$ 0.9	–38.6 $\pm$ 1.5
FEV ₁ /FVC, %/yr	–0.45 $\pm$ 0.01	–0.74 $\pm$ 0.02

Definition of abbreviations: LSC = Lovelace Smokers Cohort; NHW = non-Hispanic white; PLuSS = Pittsburgh Lung Screening Study; Q = quartile.

*Baseline refers to the visit when the first sputum samples were collected in the PLuSS.

[†]Baseline values of age, sex, height, body mass index, smoking status, pack-years, and spirometry were included in the linear mixed effects model for covariate adjustment.

An exploratory analysis for association between individual gene methylation and FEV₁ decline identified three genes (*DAPK*, *GATA4*, and *PAX5α*; all $P \leq 0.11$) associated with faster FEV₁ decline in both cohorts (Table E5).

Promoter Methylation Predicting Subsequent LC Incidence

Among the 718 PLuSS members studied in our previous genome-wide association study (26), 36 subjects had a subsequent LC incidence, with a median time interval between sputum collection and cancer diagnosis of 7.0 years (Table E6). The remaining 682 subjects were censored on the date of the most recent contact or death. Using MI as an ordinal variable in analysis, again a more heterogeneous effect is seen; that is, the highest risk for LC appears to occur when three or more genes are

methyated (Figure E2). We further converted MI into a binary variable (three or more vs. zero to two genes methylated), and Kaplan-Meier estimator showed that subjects with high methylation had a greater risk for subsequent LC incidence (log-rank test, $P = 0.0048$, Figure 2A). In addition, having three or more methylated genes was associated with an 89% increased probability of having subsequent LC incidence (hazard ratio, 1.89; 95% confidence interval, 1.26–2.83; $P = 0.0021$) than having zero to two methylated genes in a Cox proportional hazards model with adjustment for age, sex, smoking status, and pack-years.

Promoter Methylation Predicting All-Cause Mortality

Because methylation of 12 genes was associated with LC risk and FEV₁ decline,

two major factors associated with higher mortality in smokers (1, 2, 27), additional analysis was conducted to assess whether gene methylation also affected mortality in the PLuSS. Among the 718 PLuSS members with methylation data (26), 101 subjects died during follow-up, with 75.2% deaths due to smoking-related diseases (e.g., lung cancer, COPD, and cardiovascular diseases) (Table E7). Using MI as an ordinal variable in analysis, a more heterogeneous effect is seen; that is, the highest risk for all-cause mortality appears to occur when three or more genes are methylated (Figure E3). We further converted MI into a binary variable (three or more vs. zero to two genes methylated), and Kaplan-Meier estimator showed that subjects with high methylation (three or more genes methylated) had a greater risk for death during follow-ups than subjects with zero to two genes methylated (log-rank test, $P < 0.0001$; Figure 2B). The Cox proportional hazards model identified an 87% increased probability of death during follow-ups (hazard ratio, 1.87; 95% confidence interval, 1.25–2.81; $P = 0.0023$) in subjects with three or more methylated genes compared with zero to two methylated genes, with adjustment for age, sex, smoking status, and pack-years (Table 3). Disease-specific mortality analysis identified significant associations of methylation status with LC mortality and non-LC mortality, although the association for methylation status with LC deaths was much stronger than that seen for non-LC deaths (Figures 2C and 2D and Table 3).

Faster FEV₁ Decline as a Risk Factor for Subsequent Incident LC

Promoter methylation of cancer-associated genes detected in sputum may represent an expanding field of precancerous epigenetic changes in the aerodigestive tract of smokers; it was an intermediate biomarker for lung cancer and was associated with FEV₁ decline (6, 9–11). Moreover, lung cancer cases have higher prevalence of COPD. Thus, we hypothesized that faster FEV₁ decline may be a risk factor for subsequent LC incidence during follow-up in smokers. This analysis was only conducted in the PLuSS because of the rigorous identification and verification process of incident LC cases during follow-up (28) and was based on FEV₁ measurements collected before LC diagnosis. Among the 321 incident LC cases

Table 2. Association between Methylation Index and FEV₁ Decline in the Lovelace Smokers Cohort and Pittsburgh Lung Screening Study

Methylation Index	Time	MI	Time × MI
LSC (n = 1,111)			
MI as an ordinal variable*			
1 gene methylated	−20.3 ± 1.5	10.3 ± 16.9	−0.04 ± 2.1
P value	<0.0001	0.55	0.99
2 genes methylated	−20.3 ± 1.5	11.4 ± 17.5	−2.0 ± 2.2
P value	<0.0001	0.51	0.37
3 genes methylated	−20.3 ± 1.5	14.0 ± 19.2	−6.2 ± 2.4
P value	<0.0001	0.47	0.010
≥4 genes methylated	−20.3 ± 1.5	25.1 ± 16.7	−6.5 ± 2.1
P value	<0.0001	0.13	0.0018
P for trend			<0.0001
MI as a binary variable†			
High methylation	−20.9 ± 0.9	13.6 ± 11.2	−5.8 ± 1.4
P value	<0.0001	0.22	<0.0001
PLuSS (n = 477)			
MI as a binary variable†			
High methylation	−48.6 ± 1.3	−69.2 ± 25.9	−5.0 ± 2.4
P value	<0.0001	0.0075	0.035

Definition of abbreviations: LSC = Lovelace Smokers Cohort; MI = methylation index; PLuSS = Pittsburgh Lung Screening Study cohort.

Baseline values of age, sex, height, body mass index, smoking status, pack-years, and spirometry were adjusted in the linear mixed effects model. Baseline refers to the visit when the first sputum samples were collected in the PLuSS.

*Compared with zero methylated genes. Because of dramatically reduced number of subjects with MI greater than 4, these subjects were merged with subjects with MI of 4 to provide a reliable assessment of the dose–response relationship between MI and FEV₁ decline.

†Compared with low methylation (zero to two methylated genes). High methylation was defined as having three or more methylated genes.

identified to date, 124 cases had two or more spirometry measurements collected before diagnosis (Table E3). The time interval between the last visit and LC

incidence ranged from 0.24 to 7.28 years, with a median of 1.07 years. A total of 1,068 PLuSS members with two or more spirometry measurements without

subsequent LC incidence were used as control subjects. As shown in Figure 3A and Table 4, PLuSS subjects with subsequent LC incidence had a lower average FEV₁ (by 166.8 ml) and a more rapid FEV₁ decline (by 5.4 ml/yr) than smoker control subjects. However, because PLuSS subjects with subsequent LC incidence were more often male smokers and continuous current smokers, had higher pack-years, and had lower FEV₁ at baseline, four factors associated with FEV₁ decline compared with the subjects without, additional analyses on the basis of random sampling were conducted to minimize the confounding effects of these factors on the associations seen (Figure 3). For FEV₁, the estimate for the interaction term between time and subsequent LC incidence from the random sampling–based data analyses was −5.3 ml, with an SE of 2.5 and statistical significance of 0.037 (Figures 3B and 3C and Table 4). This suggests that the annual FEV₁ decline rate in smokers with subsequent lung cancer diagnosis was 5.3 ml faster than in smokers without.

Sensitivity Analysis

Longitudinal smoking behavior was characterized based on self-reported smoking status collected at each visit or contact in both cohorts. As expected, continuous current smokers had the fastest FEV₁ decline, whereas continuous abstainers had the slowest FEV₁ decline in both cohorts (Figure E4). No significant three-way interaction was identified among

Figure 1. Association between gene methylation and FEV₁ decline. High and low methylation groups were defined as smokers with three or more or zero to two genes methylated in sputum, respectively. The high methylation group has a much faster annual FEV₁ decline than the low methylation group in (A) the LSC (Lovelace Smokers Cohort) (by 5.8 ± 1.4 ml, $P < 0.0001$) and (B) PLuSS (Pittsburgh Lung Screening Study) (by 5.0 ± 2.4 ml, $P = 0.035$). Estimate and P value were calculated for LSC members who were male current smokers with baseline values of FEV₁ of 2.58 L, age of 57.74 years, height of 166.5 cm, body mass index (BMI) of 27.85, and pack-years of 40.27. Estimate and P value were calculated for PLuSS members who were male current smokers with baseline values of FEV₁ of 2.28 L, age of 65.01 years, height of 168.4 cm, BMI of 28.3, and pack-years of 61.96.

Figure 2. Kaplan-Meier curve for lung cancer (LC) latency and all-cause mortality in PLUSS (Pittsburgh Lung Screening Study) smokers by methylation status. High ($n = 195$) and low ($n = 523$) methylation groups were defined as smokers with three or more or zero to two genes methylated in sputum, respectively. The high methylation group has (A) shorter latency between sputum collection and LC incidence (log-rank $P = 0.0048$) and (B) greater all-cause mortality (log-rank $P < 0.0001$). In addition, the high methylation group also has greater (C) LC mortality (log-rank $P = 0.0005$) and (D) non-LC mortality (log-rank $P = 0.0022$).

MI, time, and longitudinal smoking status change (all $P > 0.61$ for both cohorts) in the LME models, suggesting the association between MI and FEV₁ decline did not vary by longitudinal smoking status change.

Discussion

This is the first study to assess the association between promoter methylation in exfoliated lung epithelial cells and subsequent FEV₁ decline in NHW smokers.

A more rapid decline in FEV₁ was discovered in ever-smokers with three or more genes methylated than those with zero to two genes methylated in the LSC, which was further replicated in a subset of the PLUSS cohort with much higher risk for lung cancer, suggesting that the expanding field of precancerous epigenetic changes in the aerodigestive tract may be associated with acceleration of the age-related FEV₁ decline in smokers. Because MI was not associated with pack-years in moderate and heavy smokers, the

association between MI and FEV₁ decline was not caused by MI simply reflecting smoking dose. A causal relationship between acquisition of gene promoter methylation and FEV₁ decline exceeding the age-related slope in smokers may be proposed based on the following criteria. First, a temporal relationship was observed between MI measured in baseline sputum samples and FEV₁ decline assessed using spirometry measured in the follow-up visits. Removal of the first spirometry data from the outcome did not change the

Table 3. Associations of Methylation Status (High vs. Low) with Lung-Cancer and Non-Lung-Cancer Mortality in the Pittsburgh Lung Screening Study Cohort

Mortality	Methylation Status*	N	Death	Person-Years	HR (95% CI) [†]	P Value [‡]
All-cause	High	195	46	1,343.3	1.87 (1.25–2.81)	0.0023
	Low	523	55	3,484.4		
LC [‡]	High	195	12	1,343.3	4.00 (1.52–10.51)	0.0049
	Low	523	7	3,484.4		
Non-LC [§]	High	195	34	1,343.3	1.56 (0.99–2.45)	0.056
	Low	523	48	3,484.4		

Definition of abbreviations: CI = confidence interval; HR = hazard ratio; LC = lung cancer.

*High methylation: three or more methylated genes; low methylation: zero to two methylated genes.

[†]Age, sex, smoking status, and pack-years were included in Cox proportional hazards model for covariate adjustment.

[‡]Deaths due to non-LC causes were censored at the date of death.

[§]Deaths due to LC were censored at the date of death.

associations observed in both cohorts (not shown). Second, a dose-dependent association between MI and FEV₁ decline was identified in both cohorts. Third, the significant association between MI and FEV₁ decline was independent of baseline spirometry in both cohorts, and LSC contained very low percentage (1.5%) of subjects with severe airway obstruction, strongly indicating that this association was less likely to be confounded by severe airway obstruction.

Repeated exposure of the aerodigestive tract to the carcinogens in tobacco smoke causes the development of field cancerization in the lungs of smokers that

was histologically characterized as hyperplasia and dysplasia affecting the alveolar and bronchial epithelium (6, 29). Field cancerization is driven by many important biological events of carcinogenesis, including epigenetic alterations, gene mutations, chromosome loss and rearrangement, and transcriptome changes. Genes and pathways affected by these events may together create a lung microenvironment with compromised defense capacity against cigarette smoke insults and more severe inflammatory conditions, which as drivers contribute to lung function decline and LC incidence (8, 30, 31). Moreover, this “pre-malignant

burden” forms the basis for how early lung cancer could be predicted, and in our case provides the signal for detection in sputum by our validated gene methylation panel. In addition to the excellent reproducibility regarding the performance of genes within this biomarker panel across six independent case-control studies (9, 10, 21, 32, 33), the prediction of this 12-gene panel for subsequent LC risk was also replicated in a prospective setting in the PLuSS in this study.

Of the 12 genes analyzed, *PAX5α* methylation demonstrated the strongest and most consistent association with FEV₁ decline in the two cohorts. The *PAX5* gene plays an important role in cell differentiation and embryonic development and is located on chromosome 9p13 (34). This locus has frequent chromosomal translocations and contains two distinct promoters resulting in two alternative 5' exons (α and β) that are spliced to common coding sequences of exons 2 to 10 (35). Both of these genes are methylated in approximately 65% of adenocarcinomas and squamous cell carcinomas (36). As a known tumor suppressor gene, ectopic expression of *PAX5* in silenced LC cell lines inhibited their colony formation and cell viability, arrested the cell cycle at the G2 phase, and suppressed cell migration/invasion and tumorigenicity in nude mice (37). Among these oncogenic mechanisms, reexpression of *PAX5* increased the tissue inhibitors of metalloproteinase 2 (TIMP2) mRNA level

Figure 3. PLuSS (Pittsburgh Lung Screening Study) smokers with subsequent lung cancer (LC) incidence had a more rapid FEV₁ decline. (A) Significantly lower FEV₁ at baseline (-145.9 ± 54.3 , $P = 0.0072$) and faster FEV₁ decline slope (-5.0 ± 2.5 , $P = 0.047$) were identified in PLuSS members with subsequent LC incidence ($n = 124$) compared with smoker control subjects ($n = 1,068$). (B and C) Estimates and P values were calculated for PLuSS members who were male current smokers with age of 61.95 years, height of 170.0 cm, body mass index of 28.3, and pack-years of 58.73. The distributions of (B) estimates and (C) P values for the interaction term between time and subsequent LC incidence from 1,000 random samplings were presented. A significant interaction ($P < 0.05$) was identified in 65% of random samplings. In addition, the association with the median P value (0.036) was identified, and the estimate for the interaction term between time and subsequent LC incidence on the basis of this random sampling was -5.2 ml, with an SE of 2.5.

Table 4. Associations between Subsequent Lung Cancer Incidence and Average Level and Decline of FEV₁ in Pittsburgh Lung Screening Study Members (124 Cases and 1,068 Control Subjects)

Spirometry	Full Dataset*		Random Sampling†	
	Estimate ± SE	P Value	Estimate ± SE	P Value
Average level of FEV ₁ , ml [‡]				
LC incidence (yes vs. no)	−173.5 ± 53.4	0.0012	−148.7 ± 58.6	0.011
FEV ₁ decline, ml/yr [§]				
Time, yr	−42.8 ± 0.6	<0.0001	−42.9 ± 0.8	<0.0001
LC incidence (yes vs. no)	−4.3 ± 21.2	0.84	−17.0 ± 22.3	0.45
Time × LC incidence	−5.4 ± 2.5	0.027	−5.3 ± 2.5	0.037

Definition of abbreviation: LC = lung cancer.

*Full dataset included 124 cases and 1,068 control subjects.

†Each random sampling dataset included 124 cases and 549 control subjects.

‡Baseline values of age, sex, height, body mass index, smoking behavior change, and pack-years were adjusted in the linear mixed effects model.

§Baseline values of age, sex, height, body mass index, smoking behavior change, pack-years, and FEV₁ were adjusted in the linear mixed effects model.

and inhibited the transcription of *MMP2*, *MMP7*, and *MMP9* genes, suggesting that PAX5 can inhibit tumor invasion and metastasis through modulating these tissue-remodeling enzymes. The degradation of the extracellular matrix by specialized proteolytic enzymes, such as matrix metalloproteinases, has been shown to play a key pathogenic role in both emphysema and LC (2).

The epigenetic mechanism underlying the link between COPD and LC has begun to emerge. Greater promoter methylation of *CCDC37*, presumably an essential regulator of ciliary motility, was identified in sputum from LC-free smokers and in tumor tissues from patients with non-small-cell LC with COPD versus those without COPD (38). Thus, epigenetic silencing of *CCDC37* in human lungs may result in the accumulation of mucus, leading to increased pulmonary inflammation and likely enhanced oxidative tissue damage, which in turn could play a role in further reducing lung function. Interestingly, *CCDC37* methylation is significantly correlated with MI of the 12-gene panel in sputum ($n = 206$, Spearman $r = 0.23$, $P_s = 0.0009$) in the LSC, suggesting promoter methylation of genes not in the 12-gene panel may functionally contribute to the observed association because of their correlation with the 12-gene panel. Comparison of the methylome in small airway epithelial cells from former smokers with and without airway obstruction identified three small airway biology pathways enriched for genes predominantly

hypermethylated in subjects with airway obstruction, which included G protein-coupled receptor signaling, aryl hydrocarbon receptor signaling, and cAMP-mediated signaling (39). Most importantly, cAMP-mediated signaling and G protein-coupled receptor signaling pathways are also the two most affected pathways enriched for genes whose promoters acquire hypermethylation in LC tissues using The Cancer Genome Atlas data (S. Leng and colleagues, unpublished results). Thus, these findings suggest that epigenetic silencing of genes associated with COPD may contribute to lung carcinogenesis. However, delineation of the temporal relationship between acquisition of gene promoter hypermethylation and COPD requires longitudinal follow-up of the smokers with relatively healthy lung function at study entry.

Accumulating evidence has shown that methylation profiles of CpGs in blood DNA identified as a proxy of biological aging or chronic exposure to cigarette smoke were associated with all-cause mortality and mortality attributable to cancer or cardiovascular diseases (40–43). These findings suggested that methylation in blood DNA could quantify the heritable epigenetic damage resulting from the host and environment exposure interactions and serve as a biomarker predicting all-cause mortality. The current study is the first to show that, among moderate and heavy smokers, promoter hypermethylation of cancer-associated genes in sputum could also serve as a biomarker predicting all-

cause mortality, with the majority (75.2%) of deaths caused by smoking-related diseases. Furthermore, disease-specific mortality analyses also suggested higher methylation associated with LC mortality and non-LC mortality.

The translational significance of this study was further strengthened by showing a more rapid decline of FEV₁ observed in smokers with subsequent LC incidence in the PLuSS. This is consistent with the facts that patients with asymptomatic lung tumors have a greater prevalence of COPD (28) and lower FEV₁ was associated with higher subsequent LC incidence in a prospective setting (3). More rapid decline in lung function has also been shown to be independently associated with a modest increased risk of hospital admissions and deaths from COPD in an elderly cohort of U.S. participants (27).

The major strengths of this study include the use of two geographically independent prospective smoker cohorts with ample length of follow-up period assessing FEV₁ decline and assessment for methylation of a 12-gene panel that have been replicated multiple times in several independent studies in terms of their associations with LC risk. In addition, a random sampling approach was used to properly control the potential confounding effects in assessing FEV₁ decline by subsequent LC incidence. Furthermore, sensitivity analysis was conducted for assessing the association between MI and FEV₁ decline, and the results suggested that there was little likelihood of confounding due to longitudinal smoking behavior change.

The major limitation of our study is that it was conducted in NHW smokers only; thus, whether this result can be generalized to other ethnic groups with higher LC susceptibility and risk, such as Hispanics and African Americans, is unknown. Second, gene methylation was only measured in baseline sputum samples, and it is unknown whether its change over time will affect FEV₁ decline. Third, functional study is required to disentangle which gene in particular could contribute to the observed association between MI and FEV₁ decline in smokers.

In conclusion, our study provides convincing evidence that the development of field cancerization in the lungs of smokers as detected by methylation of cancer-associated genes in sputum was

associated with the rapid decline of lung function and worse all-cause and LC- and non-LC-specific mortality. Thus, in addition to being validated as a risk biomarker for LC, this 12-gene methylation panel may also serve as a biomarker associated with predisposition for cigarette smoke-induced accelerated aging of overall lung health in moderate and heavy smokers. All these exciting

findings provide strong evidence for assessing the efficacy of drugs, such as azacitidine and entinostat, or dietary nutrients, such as folate, niacin, vitamin D, and omega 3 fatty acids, that are being developed for epigenetic therapy or chemoprevention of LC (17, 44–47) for prevention of FEV₁ decline and accelerated aging of lung health in smokers. ■

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

Acknowledgment: The authors thank Dr. Joel L. Weissfeld, previously at the Department of Epidemiology, Graduate School of Public Health, University of Pittsburgh, for his contribution in data management and analysis at the initial stage of this study. They also thank Ms. Elise Calvillo at Lovelace Respiratory Research Institute for scientific editing of the figures.

Reference

1. Youliden DR, Cramb SM, Baade PD. The international epidemiology of lung cancer: Geographical distribution and secular trends. *J Thorac Oncol* 2008;3:819–831.
2. Houghton AM. Mechanistic links between COPD and lung cancer. *Nat Rev Cancer* 2013;13:233–245.
3. Fry JS, Hamling JS, Lee PN. Systematic review with meta-analysis of the epidemiological evidence relating FEV1 decline to lung cancer risk. *BMC Cancer* 2012;12:498.
4. Rice MB, Ljungman PL, Wilker EH, Dorans KS, Gold DR, Schwartz J, et al. Long-term exposure to traffic emissions and fine particulate matter and lung function decline in the Framingham heart study. *Am J Respir Crit Care Med* 2015;191:656–664.
5. Sood A, Petersen H, Blanchette CM, Meek P, Picchi MA, Belinsky SA, et al. Wood smoke exposure and gene promoter methylation are associated with increased risk for COPD in smokers. *Am J Respir Crit Care Med* 2010;182:1098–1104.
6. Belinsky SA. Gene-promoter hypermethylation as a biomarker in lung cancer. *Nat Rev Cancer* 2004;4:707–717.
7. Gomperts BN, Spira A, Massion PP, Walser TC, Wistuba II, Minna JD, et al. Evolving concepts in lung carcinogenesis. *Semin Respir Crit Care Med* 2011;32:32–43.
8. Steiling K, Ryan J, Brody JS, Spira A. The field of tissue injury in the lung and airway. *Cancer Prev Res (Phila)* 2008;1:396–403.
9. Belinsky SA, Liechty KC, Gentry FD, Wolf HJ, Rogers J, Vu K, et al. Promoter hypermethylation of multiple genes in sputum precedes lung cancer incidence in a high-risk cohort. *Cancer Res* 2006;66:3338–3344.
10. Leng S, Do K, Yingling CM, Picchi MA, Wolf HJ, Kennedy TC, et al. Defining a gene promoter methylation signature in sputum for lung cancer risk assessment. *Clin Cancer Res* 2012;18:3387–3395.
11. Leng S, Wu G, Klinge DM, Thomas CL, Casas E, Picchi MA, et al. Gene methylation biomarkers in sputum as a classifier for lung cancer risk. *Oncotarget* 2017;8:63978–63985.
12. 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.
13. Lange NE, Sordillo J, Tarantini L, Bollati V, Sparrow D, Vokonas P, et al. Alu and LINE-1 methylation and lung function in the normative ageing study. *BMJ Open* 2012;2:e001231.
14. Lepeule J, Baccarelli A, Motta V, Cantone L, Litonjua AA, Sparrow D, et al. Gene promoter methylation is associated with lung function in the elderly: the Normative Aging Study. *Epigenetics* 2012;7:261–269.
15. Bruse S, Petersen H, Weissfeld J, Picchi M, Willink R, Do K, et al. Increased methylation of lung cancer-associated genes in sputum DNA of former smokers with chronic mucous hypersecretion. *Respir Res* 2014;15:2.
16. Leng S, Stidley CA, Willink R, Bernauer A, Do K, Picchi MA, et al. Double-strand break damage and associated DNA repair genes predispose smokers to gene methylation. *Cancer Res* 2008;68:3049–3056.
17. Stidley CA, Picchi MA, Leng S, Willink R, Crowell RE, Flores KG, et al. Multivitamins, folate, and green vegetables protect against gene promoter methylation in the aerodigestive tract of smokers. *Cancer Res* 2010;70:568–574.
18. Wilson DO, Weissfeld JL, Fuhrman CR, Fisher SN, Balogh P, Landreneau RJ, et al. The Pittsburgh Lung Screening Study (PLUSS): outcomes within 3 years of a first computed tomography scan. *Am J Respir Crit Care Med* 2008;178:956–961.
19. Sharma S, Kelly TK, Jones PA. Epigenetics in cancer. *Carcinogenesis* 2010;31:27–36.
20. Feinberg AP, Ohlsson R, Henikoff S. The epigenetic progenitor origin of human cancer. *Nat Rev Genet* 2006;7:21–33.
21. Belinsky SA, Nikula KJ, Palmisano WA, Michels R, Saccomanno G, Gabrielson E, et al. Aberrant methylation of p16(INK4a) is an early event in lung cancer and a potential biomarker for early diagnosis. *Proc Natl Acad Sci USA* 1998;95:11891–11896.
22. Baylin SB, Herman JG, Graff JR, Vertino PM, Issa JP. Alterations in DNA methylation: a fundamental aspect of neoplasia. *Adv Cancer Res* 1998;72:141–196.
23. Leng S, Liu Y, Thomas CL, Gauderman WJ, Picchi MA, Bruse SE, et al. Native American ancestry affects the risk for gene methylation in the lungs of Hispanic smokers from New Mexico. *Am J Respir Crit Care Med* 2013;188:1110–1116.
24. Bruse S, Sood A, Petersen H, Liu Y, Leng S, Celedón JC, et al. New Mexican Hispanic smokers have lower odds of chronic obstructive pulmonary disease and less decline in lung function than non-Hispanic whites. *Am J Respir Crit Care Med* 2011;184:1254–1260.
25. Shrier I, Platt RW. Reducing bias through directed acyclic graphs. *BMC Med Res Methodol* 2008;8:70.
26. Leng S, Liu Y, Weissfeld JL, Thomas CL, Han Y, Picchi MA, et al. 15q12 variants, sputum gene promoter hypermethylation, and lung cancer risk: a GWAS in smokers. *J Natl Cancer Inst* 2015;107:djv035.
27. Mannino DM, Davis KJ. Lung function decline and outcomes in an elderly population. *Thorax* 2006;61:472–477.
28. Wilson DO, Weissfeld JL, Balkan A, Schragin JG, Fuhrman CR, Fisher SN, et al. Association of radiographic emphysema and airflow obstruction with lung cancer. *Am J Respir Crit Care Med* 2008;178:738–744.
29. Belinsky SA. Unmasking the lung cancer epigenome. *Annu Rev Physiol* 2015;77:453–474.
30. Pierrou S, Broberg P, O'Donnell RA, Pawlowski K, Virtala R, Lindqvist E, et al. Expression of genes involved in oxidative stress responses in airway epithelial cells of smokers with chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 2007;175:577–586.
31. Spira A, Beane JE, Shah V, Steiling K, Liu G, Schembri F, et al. Airway epithelial gene expression in the diagnostic evaluation of smokers with suspect lung cancer. *Nat Med* 2007;13:361–366.
32. Hubers AJ, Heideman DA, Duin S, Witte BI, de Koning HJ, Groen HJ, et al. DNA hypermethylation analysis in sputum of asymptomatic subjects at risk for lung cancer participating in the NELSON trial: argument for maximum screening interval of 2 years. *J Clin Pathol* 2017;70:250–254.
33. Palmisano WA, Divine KK, Saccomanno G, Gilliland FD, Baylin SB, Herman JG, et al. Predicting lung cancer by detecting aberrant promoter methylation in sputum. *Cancer Res* 2000;60:5954–5958.
34. Stapleton P, Weith A, Urbánek P, Kozmik Z, Busslinger M. Chromosomal localization of seven PAX genes and cloning of a novel family member, PAX-9. *Nat Genet* 1993;3:292–298.

35. Busslinger M, Klix N, Pfeffer P, Graninger PG, Kozmik Z. Deregulation of PAX-5 by translocation of the Emu enhancer of the IgH locus adjacent to two alternative PAX-5 promoters in a diffuse large-cell lymphoma. *Proc Natl Acad Sci USA* 1996;93:6129–6134.
36. Palmisano WA, Crume KP, Grimes MJ, Winters SA, Toyota M, Esteller M, *et al*. Aberrant promoter methylation of the transcription factor genes PAX5 alpha and beta in human cancers. *Cancer Res* 2003;63:4620–4625.
37. Zhao L, Li S, Gan L, Li C, Qiu Z, Feng Y, *et al*. Paired box 5 is a frequently methylated lung cancer tumour suppressor gene interfering β -catenin signalling and GADD45G expression. *J Cell Mol Med* 2016;20:842–854.
38. Tessema M, Yingling CM, Picchi MA, Wu G, Liu Y, Weissfeld JL, *et al*. Epigenetic repression of CCDC37 and MAP1B links chronic obstructive pulmonary disease to lung cancer. *J Thorac Oncol* 2015;10:1181–1188.
39. Vucic EA, Chari R, Thu KL, Wilson IM, Cotton AM, Kennett JY, *et al*. DNA methylation is globally disrupted and associated with expression changes in chronic obstructive pulmonary disease small airways. *Am J Respir Cell Mol Biol* 2014;50:912–922.
40. Chen BH, Marioni RE, Colicino E, Peters MJ, Ward-Caviness CK, Tsai PC, *et al*. DNA methylation-based measures of biological age: meta-analysis predicting time to death. *Aging (Albany NY)* 2016;8:1844–1865.
41. Perna L, Zhang Y, Mons U, Holleczeck B, Saum KU, Brenner H. Epigenetic age acceleration predicts cancer, cardiovascular, and all-cause mortality in a German case cohort. *Clin Epigenetics* 2016;8:64.
42. Zhang Y, Schöttker B, Florath I, Stock C, Butterbach K, Holleczeck B, *et al*. Smoking-associated DNA methylation biomarkers and their predictive value for all-cause and cardiovascular mortality. *Environ Health Perspect* 2016;124:67–74.
43. Zhang Y, Wilson R, Heiss J, Breitling LP, Saum KU, Schöttker B, *et al*. DNA methylation signatures in peripheral blood strongly predict all-cause mortality. *Nat Commun* 2017;8:14617.
44. Juergens RA, Wrangle J, Vendetti FP, Murphy SC, Zhao M, Coleman B, *et al*. Combination epigenetic therapy has efficacy in patients with refractory advanced non-small cell lung cancer. *Cancer Discov* 2011;1:598–607.
45. Dimri M, Bommi PV, Sahasrabudhe AA, Khandekar JD, Dimri GP. Dietary omega-3 polyunsaturated fatty acids suppress expression of EZH2 in breast cancer cells. *Carcinogenesis* 2010;31:489–495.
46. Raynal NJ, Lee JT, Wang Y, Beaudry A, Madireddi P, Garriga J, *et al*. Targeting calcium signaling induces epigenetic reactivation of tumor suppressor genes in cancer. *Cancer Res* 2016;76:1494–1505.
47. Fetahu IS, Höbaus J, Kállay E. Vitamin D and the epigenome. *Front Physiol* 2014;5:164.