RESEARCH PAPER

Metastable DNA methylation sites associated with longitudinal lung function decline and aging in humans: an epigenome-wide study in the NAS and KORA cohorts

Juan Jose Carmona^{a,b,c}, Richard T. Barfield^{*d}, Tommaso Panni^e, Jamaji C. Nwanaji-Enwerem^{a,f}, Allan C. Just^g, John N. Hutchinson^h, Elena Colicinoⁱ, Stefan Karrasch^{e,j,k}, Simone Wahl^l, Sonja Kunze^{e,l}, Nadereh Jafari^m, Yinan Zhengⁿ, Lifang Houⁿ, Dawn L. DeMeo^o, Augusto A. Litonjua^o, Pantel S. Vokonas^p, Annette Peters^{c,e}, Xihong Lin^{h,q}, Joel Schwartz^c, Holger Schulz^{†e,k}, and Andrea A. Baccarelli^{†c,r}

^aDepartment of Environmental Health and Program in Quantitative Genomics, Harvard T.H. Chan School of Public Health, Boston, MA, USA; ^bCenter for Bioethics, Harvard Medical School, Boston, MA, USA; ^cOffice of Research, Dana-Farber/Harvard Cancer Institute, Boston, MA, USA; ^dPublic Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, WA, USA; ^eInstitute of Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany; ^fMD-PhD Program, Harvard Medical School, Boston, MA, USA; ^gPreventive Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, USA; ^hDepartment of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA, USA; ⁱDepartments Environmental Medicine & Public Health, Division of Biostatistics, Icahn School of Medicine at Mount Sinai, New York, NY, USA; ^jInstitute and Outpatient Clinic for Occupational, Social and Environmental Medicine, Ludwig-Maximilians-Universität, Munich, Germany; ^kComprehensive Pneumology Center Munich, Member of the German Center for Lung Research, Munich, Germany; ^lResearch Unit of Molecular Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany; ^mCenter for Genetic Medicine, Feinberg School of Medicine, Northwestern University, Chicago, IL, USA; ⁿCenter for Population Epigenetics, Robert H. Lurie Comprehensive Cancer Center and Department of Preventive Medicine, Feinberg School of Medicine, Northwestern University, Chicago, IL, USA; ^oChanning Division of Network Medicine and the Division of Pulmonary and Critical Care Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA; ^pVA Normative Aging Study, Veterans Affairs Boston Healthcare System and the Department of Medicine, Boston University School of Medicine, Boston, MA, USA; ^qDepartment of Statistics, Harvard University, Cambridge, MA, USA; ^rDepartments of Epidemiology and Environmental Health Sciences, Columbia University Mailman School of Public Health, New York, NY, USA

ABSTRACT

DNA methylation is an epigenetic regulator of gene transcription, which has been found to be both metastable and variable within human cohort studies. Currently, few studies have been done to identify metastable DNA methylation biomarkers associated with longitudinal lung function decline in humans. The identification of such biomarkers is important for screening vulnerable populations. We hypothesized that quantifiable blood-based DNA methylation alterations would serve as metastable biomarkers of lung function decline and aging, which may help to discover new pathways and/or mechanisms related to pulmonary pathogenesis. Using linear mixed models, we performed an Epigenome Wide Association Study (EWAS) between DNA methylation at CpG dinucleotides and longitudinal lung function (FVC, FEV₁, FEF_{25–75%}) decline and aging with initial discovery in the Normative Aging Study, and replication in the Cooperative Health Research in the Region of Augsburg cohort. We identified two metastable epigenetic loci associated with either poor lung function and aging, cg05575921 (*AHRR* gene), or lung function independently of aging, cg06126421 (*IER3* gene). These loci may inform basic mechanisms associated with pulmonary function, pathogenesis, and aging. Human epigenomic variation, may help explain features of lung function decline and related pathophysiology not attributable to DNA sequence alone, such as accelerated pulmonary decline in smokers, former smokers, and perhaps non-smokers. Our EWAS across two cohorts, therefore, will likely have implications for the human population, not just the elderly.

ARTICLE HISTORY

Received 6 March 2018
Revised 10 September 2018
Accepted 22 September 2018

KEYWORDS

Biomarker; DNA methylation; pulmonary function; lung function decline

Introduction

A number of genome-wide association studies have paved the way for identifying human genetic variants linked to lung function [1–3], yet

nucleotide-level polymorphisms can only explain a limited amount of disease risk, suggesting that other, gene-regulatory mechanisms may also be at play [4]. Lung function decline fits well into the

CONTACT Richard T. Barfield rbarfiel@fredhutch.org Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, WA, USA

*Designates equal contribution first author

†Equal senior authors

This article has been republished with minor changes. These changes do not impact the academic content of the article.

Supplemental data for this article can be accessed [here](#).

© 2018 Informa UK Limited, trading as Taylor & Francis Group

epigenetic hypothesis of disease [5], which posits that epigenetic variation across individuals may be an important intermediary between genetic variation and other factors that may ‘program’ the DNA [6,7]. Epigenetic mechanisms like DNA methylation may provide further explanation for poor lung function and accelerated decline and aging-related associations that are not explained entirely by the DNA sequence alone. This could be things such as the variable susceptibility to develop lung disease in smokers and the sustained, elevated risk for lung function decline years after smoking cessation [2,8–12]. Epigenetics studies can shed novel insights into the pathogenesis and susceptibility to lung function and pulmonary disease.

DNA methylation (DNAm) commonly occurs at cytosine residues arranged within ‘CpG’ (cytosine-phosphate-guanine dinucleotide) sites. When DNAm does not change over time we denote the methylation at that CpG site as metastable. DNAm is the best-studied examples of DNA-level epigenetic mechanisms across species and research is ongoing to better map and understand how human DNAm shapes health, disease and disparities across persons, with transformative uses in the clinical setting [7,13–15]. DNAm has been found to be associated with biological aging as well as a potential driver of biological aging [16]. Therefore, the exploration of non-invasive, blood-based DNAm diagnostics to screen for and help identify persons at risk for poor lung function and/or accelerated decline is a public health priority, as decreased pulmonary function is highly associated with mortality [17–20].

In humans, lung function increases after birth until about age 20, which is when it peaks and begins to plateau up to about age 30, defining the starting point for the followed age-related decline, which typically is accelerated in smokers [2]. In this study, we focused on three spirometric indices that provide slightly different information on lung function. Forced vital capacity (FVC, liters), is determined by the size of the gas exchanging region, the alveolar region of the lung. FVC decreases in cases of fibrosis as well as when the maximal respirable volume decreases for example in chronic obstructive pulmonary disease (COPD) particularly due to emphysema. Forced expiratory flow between 25–75% of

FVC ($FEF_{25-75\%}$, liters per minute), measures the conducting airway function, and is determined by the size of the airways. A low flow rate of $FEF_{25-75\%}$ reflects airflow limitation usually corresponding to airway narrowing or obstruction of the airways. Finally, forced expiratory volume in 1 second (FEV_1 , liters), which is affected by airway narrowing, is the volume that is expired in 1 second. It is often reduced in COPD and asthma but can also be reduced when the lung size is affected due to a decrease in maximal respirable volume, such as in fibrosis and emphysema.

Classically, smoking is a predominant environmental risk factor for lung function decline and relevant pathophysiological conditions, such as COPD, which is a leading cause of death in the USA [21]. While there are DNA methylation biomarkers that are associated with smoking exposures [22,23] and correlated with adverse lung function [24,25], we sought to account for smoking in a way that would allow us to consider DNAm that is variable and/or stable over time in individuals, and could inform pulmonary function independent of smoking. This allows for using biomarkers that are informative more broadly across individuals, and, hopefully, less susceptible to influence by other variables that may confound their utility within clinical and research settings.

We used linear mixed models to perform an epigenome-wide association study of the association between DNAm at CpG sites and longitudinal lung function decline in the Normative Aging Study (NAS), a cohort of elderly men residing in the greater-Boston metropolitan area. The initial discovery phase was in the NAS cohort followed by validation in the Cooperative Health Research in the Region of Augsburg (KORA) cohort. Lung function decline in this paper is defined as a decrease in any of the three studied pulmonary function parameters (FVC, $FEF_{25-75\%}$, or FEV_1) between visits. We assess the association of DNAm effect with this decline by modeling its interaction effect with time of observation. Finally, given that lung function decline is often indicative of aging and/or risk of mortality in vulnerable individuals [26–29], we assessed the correlation between methylation of particular CpG sites of interest, pulmonary function, and an age-difference variable constructed from ‘DNA methylation age’ (DNAm age) [30].

Results

In NAS, there were 633 men who contributed 2245 observations (average 3.5 observations/participant) over about a decade. The mean age was 72 years old at time of blood draw, with 28 current smokers, 426 former smokers, and 179 never smokers. KORA consisted of 868 people with age between 41–63 and 408 men (47%). Table 1 displays descriptive statistics for the time varying variables in our study. It displays up to four visits prior to blood draw in NAS and up to one visit after blood draw in KORA as participants had varying amount of follow up time. Table 2 displays the variables that were held constant in the analysis. These are variables such as cell type proportion, age at first visit, and height at first non-missing observation.

We report in Table 2 body mass index at blood draw, though we do not adjust for it in the model. In NAS, there were 2 underweight (.3%), 136 normal weight (21.5%), 340 overweight (53.7%), and 155 obese (24.5%) people in NAS. In KORA, there was 1 underweight (.1%), 278 with normal weight (32%), 344 overweight (39.6%), and 245 obese (28.2%) people. In NAS, the mean age at first visit was 63 in NAS, and 54 in KORA. The lung function measures saw a decline across the four visits (Table 1). The varying smoking status indicate varying levels of data being available. In NAS for example, of the 28 current smokers at time of blood draw, only 22 had up to three observations prior to time of blood draw (Table 1). In KORA, of the 180 current smokers, 120 had a secondary lung function observation.

Table 1. Time-varying characteristics obtained from the NAS and KORA participants.

Cohort variables	NAS*				KORA†	
	At 3 visits prior to DNAm	At 2 visits prior to DNAm	At 1 visit prior to DNAm	At time of DNAm	At time of DNAm (F4)	At 1 visit after DNAm (F4L)
Number of observations	447	555	610	633	868	695
Follow-up time, mean $\pm$ SD, yr	-	2.44 $\pm$ 1.22	5.11 $\pm$ 2	8.41 $\pm$ 2.57	-	3.18 $\pm$ 0.47
Proportion of men	1.0	1.0	1.0	1.0	0.47	0.47
Weight, mean $\pm$ SD, kg	83.82 $\pm$ 12.52	84.55 $\pm$ 12.94	85.11 $\pm$ 13.68	84.95 $\pm$ 14.05	79.52 $\pm$ 16.92	80.84 $\pm$ 18.30
Education, mean $\pm$ SD, yr	15.05 $\pm$ 2.99	15.01 $\pm$ 2.99	15.06 $\pm$ 2.99	15.06 $\pm$ 2.99	12.04 $\pm$ 2.64	12.15 $\pm$ 2.66
Smoking status, number of participants (%)						
Never	130 (0.29)	164 (0.30)	173 (0.28)	179 (0.28)	323 (0.37)	271 (0.39)
Current	22 (0.05)	27 (0.05)	28 (0.05)	28 (0.04)	180 (0.21)	120 (0.17)
Former	295 (0.66)	364 (0.66)	409 (0.67)	426 (0.67)	365 (0.42)	304 (0.44)
Pack-years (current or former smokers only), mean $\pm$ SD	27.43 $\pm$ 22.66	28.54 $\pm$ 23.56	28.27 $\pm$ 24.64	28.94 $\pm$ 25.23	13.19 $\pm$ 20.09	12.69 $\pm$ 19.78
Participants taking medication ¹ , n (%)	80 (0.18)	94 (0.17)	85 (0.14)	71 (0.11)	62 (0.074)	47 (0.07)
Vitamin C intake, milligrams, not from supplements	149	162	166	161	-	-
FEV ₁ , mean $\pm$ SD, L in 1 sec (# missing)	2.87 $\pm$ 0.58 (4)	2.74 $\pm$ 0.58 (3)	2.63 $\pm$ 0.58 (6)	2.54 $\pm$ 0.58 (53)	3.28 $\pm$ 0.79 (0)	3.26 $\pm$ 0.81 (0)
FVC, mean $\pm$ SD, L (# missing)	3.78 $\pm$ 0.67 (4)	3.65 $\pm$ 0.65 (3)	3.49 $\pm$ 0.66 (5)	3.36 $\pm$ 0.65 (53)	4.23 $\pm$ 0.98 (0)	4.20 $\pm$ 1.01 (0)
FEF _{25%-75%} , mean $\pm$ SD, L/min (# missing)	2.82 $\pm$ 1.06 (3)	2.67 $\pm$ 1.10 (4)	2.63 $\pm$ 1.13 (6)	2.50 $\pm$ 1.11 (53)	2.92 $\pm$ 1.03 (0)	2.80 $\pm$ 1.04 (0)
Season, n (%)						
Spring, March-May	129 (0.29)	168 (0.30)	162 (0.27)	152 (0.24)	175 (0.20)	34 (0.05)
Summer, June-August	111 (0.25)	143 (0.26)	180 (0.30)	170 (0.27)	135 (0.16)	353 (0.51)
Autumn, September-November	105 (0.23)	129 (0.23)	150 (0.25)	199 (0.31)	286 (0.33)	308 (0.44)
Winter, December-February	102 (0.23)	115 (0.21)	118 (0.19)	112 (0.18)	272 (0.31)	-
Day of Week, n (%)						
Monday	139 (0.31)	190 (0.34)	170 (0.28)	44 (0.07)	182 (0.21)	119 (0.17)
Tuesday	-	-	17 (0.028)	152 (0.24)	165 (0.19)	136 (0.20)
Wednesday	152 (0.34)	172 (0.31)	219 (0.36)	282 (0.45)	152 (0.18)	109 (0.16)
Thursday	156 (0.35)	193 (0.35)	204 (0.33)	155 (0.24)	151 (0.17)	147 (0.21)
Friday	-	-	-	-	146 (0.17)	152 (0.22)
Saturday	-	-	-	-	72 (0.08)	32 (0.05)

*In NAS cohort, men are participating in 2,245 visits from 1999–2007; in our retrospective study design, NAS participants may have up to 3 visits prior to the latest visit at which blood was obtained for DNAm assessment, *i.e.* up to 4 visits total. †In the KORA cohort, women and men are participating in 1,563 visits from 2006–2010; given our prospective study design, KORA participants can have up to 1 additional visit after the blood is obtained for DNAm analysis, *i.e.* up to 2 visits total. ¹Medicine taken denotes either Corticosteroids (inhaled (ICS) or systemic), beta-2-adrenergic agonists, typical LABA sympathomimetic alpha AHFS, or Long Acting Muscarinic Antagonist (LAMA). FVC: forced vital capacity, FEV₁: forced expiratory volume in 1 second, FEF_{25–75%}: forced expiratory flow between at 25–75% of FVC, ICS: inhaled corticosteroids, LABA: long acting beta-2-adrenergic agonists

Table 2. Non-time-varying and other characteristics obtained from participants in the NAS and KORA cohorts: blood-cell proportions, baseline age, and height.

Cohort characteristics	NAS	KORA
Cell proportions <i>via</i> Houseman <i>et al.</i> at time of DNAm		
CD8 ⁺ T-cells	0.023	0.009
CD4 ⁺ T-cells	0.133	0.123
Natural Killer Cells	0.069	0.052
B-cells	0.034	0.004
Monocytes	0.099	0.113
Granulocytes	0.664	0.648
Age at first non-missing visit, mean $\pm$ SD, yr	62.78 $\pm$ 7.00	53.77 $\pm$ 4.46
Height at first non-missing visit, mean $\pm$ SD, cm	174.128 $\pm$ 6.66	169.291 $\pm$ 3.632
COPD (FEV ₁ /FVC $\leq$ 0.7)	176	91
Asthma	47	78
BMI at time of blood draw, mean $\pm$ SD, kg/m ²	27.96 $\pm$ 4.01	27.64 $\pm$ 4.96

We analyzed DNAm's association with pulmonary function in NAS assuming that DNAm was meta-stable (i.e. does not change with time). DNAm was collected at the last time point and then treated as non-time varying when regressing on lung function measures collected prior to DNAm. We analyzed FEV₁, FVC, and FEF_{25–75%} separately in 477,927 probes. We tested for two different associations: association with lung function decline (interaction between DNAm and follow up time from first spirometric measure, β_{LD}) and its cross sectional association (association between DNAm and lung function when follow up time is 0, β_{CS}). The statistical model is provided in the methods section under biostatistical analysis. Variables included in the model were: season, follow up time, day of the week, plate, chip, mean centered weight, mean centered weight squared, the natural log of height, baseline age, vitamin C intake not from supplements, maximum years of education and, taking any of the following drugs-corticosteroids (inhaled (ICS) or systemic), beta-2-adrenergic agonists (typical LABA), sympathomimetic alpha AHFS, or long acting muscarinic antagonist (LAMA). We also adjusted for cell type composition [31]. Significance was assessed either via a Holm adjusted p-value less than 0.05 or an FDR adjusted p-value less than 0.10 [32,33].

NAS lung function decline results (β_{LD})

We first assessed the interaction between follow up time and DNA methylation. As lung function decreases after the age of 30, the association between

follow up time and lung function will be negative. The interaction between DNAm and follow up time then should denote the contribution of DNAm to lung function decline. When adjusting for smoking, nine unique CpG probes were significantly associated with the rate of decline in pulmonary function tests (PFTs) at the Holm significance level of 0.05 [33], (Table 3-Group 1). Not adjusting for smoking, ten unique probes were associated with rate of decline (Table 3-Group 1). There were seven probes that were significant regardless of adjustment for smoking, three probes significant when not adjusting for smoking, and two probes significant only when adjusting for smoking (Table 3-Group 1). The parameter estimates between the smoking adjusted and unadjusted models remained the same for the seven probes (Table 3-Group 1). Quantile-quantile plots (QQplots) of the analysis showed some inflation for this analysis in the FEV₁ and FVC analyses (Figure 1, subplot A1, B1, C1, genomic inflation factor in caption of Table 3-Group 1). There were 1490 significant associations (mapping to 1374 unique probes) when adjusting for smoking at the false discovery rate (FDR) level of 0.1 and 986 (954 unique probes) when we did not adjust for smoking [32]. A list of significant probes is provided in Supplement Table 1.

We reran the analysis excluding current smokers (n = 28) to ensure they were not influencing results (Table 3-Group 2). Five probes were still associated with rate of decline. Three (cg14292220, cg05644990 and cg26468478) were significant regardless of adjustment for smoking, while the other two (cg01249054 and cg09995068) were significant only when not adjusting for smoking. Probe cg14292220 was only significant in the whole sample (Table 3-Group 1) when not adjusting for smoking, while in the subgroup analysis it was significant irrespective of adjustment for smoking. There were 489 (477 unique probes) significant associations at FDR level of 0.1 when adjusting for smoking, and 368 (361 unique probes) when not adjusting for smoking. A list of significant probes and association is provided in Supplement Table 2.

NAS probes associated with cross sectional association (β_{CS})

When adjusting for smoking there were two probes associated with cross-sectional lung function via the

Table 3. DNAm and lung function decline (β_{LD}): Results for β_{LD} association in NAS, the measure of effect on DNAm on the rate of decline in lung function. The interaction between DNAm and follow up time, unique DNA methylation sites associated with a given Pulmonary Function Test (PFT) model organized by smoking analysis, unadjusted (left side) and adjusted for smoking (right side).

Probe (Gene)	Sample**	PFT†	Unadjusted for smoking		Adjusted for smoking		
			Effect Estimate	P-value	Effect Estimate	P-value	ICC* (95% CI)
cg02721176 (<i>C10ORF96</i>)	Group 1	FEV ₁	0.03	4.13e-08	0.03	1.99e-07 ^{NS}	0.81 (0.79–0.86)
cg09995068		FEV ₁	−0.41	2.68e-09	−0.34	9.65e-07 ^{NS}	0.99 (0.99–0.99)
cg14292220		FEV ₁	0.23	5.14e-10	0.19	2.74e-07 ^{NS}	0.95 (0.94–0.96)
cg18476993 (<i>MIR671;CHPF2</i>)		FEV ₁	0.13	7.87e-07 ^{NS}	0.14	6.81e-08	0.64 (0.61–0.73)
cg05644990 (<i>GPBP1</i>)		FVC	0.83	2.08e-08	0.85	1.37e-08	7.9e-14 (7.9e-14–1.1e-13)
cg12565126 (<i>MFSD2B</i>)		FVC	0.08	8.04e-08	0.08	8.69e-08	0.88 (0.88–0.92)
cg13532885 (<i>SYN1</i>)		FVC	0.06	4.26e-08	0.06	3.9e-08	0.93 (0.93–0.95)
cg21850722 (<i>ROR1</i>)		FVC	0.10	1.04e-07	0.10	7.67e-08	0.55 (0.52–0.66)
cg26468478 (<i>CELSR3</i>)		FVC	0.89	2.55e-08	0.89	3.38e-08	0.031 (1.3e-13–0.2)
cg16574073 (<i>RUNX3</i>)		FVC	0.13	5.47e-08	0.13	4.82e-08	0.67 (0.65–0.76)
cg03867607 (<i>MYL6</i>)		FVC	0.39	1.10e-07 ^{NS}	0.40	4.45e-08	0.76 (0.75–0.83)
cg05191655		FVC	0.15	9.36e-08	0.15	3.05e-08	0.73 (0.71–0.80)
cg09995068	Group 2	FEV ₁	−0.41	1.31e-09	−0.36	1.79e-07 ^{NS}	0.99 (0.99–0.99)
cg14292220		FEV ₁	0.23	1.14e-10	0.21	2.43e-08	0.95 (0.94–0.96)
cg01249054		FEV ₁	0.64	4.24e-08	0.57	1.83e-06 ^{NS}	0.78 (0.76–0.84)
cg05644990 (<i>GPBP1</i>)		FVC	0.81	8.04e-08	0.82	6.38e-08	7.9e-14 (7.9e-14–1.1e-13)
cg26468478 (<i>CELSR3</i>)		FVC	0.91	3.05e-08	0.91	4.25e-08	0.031 (1.3e-13–0.2)

*ICC denotes the intraclass correlation coefficient, which describes how stably DNAm at a given locus resembles itself within samples of an individual over time. **Sample: Group 1 included former, current, and never smokers, n = 633 participants across all three smoking classes, never (179 participants), former (426), and current smokers (28). Group 2 consists of just former and never smokers; this is restricting analysis to never (179 participants) and former (426) smokers. The genomic inflation lambda (λ) values, with no adjustment for smoking for the longitudinal association (β_{LD}), were 1.25, 1.23, and 0.96 for the FEV₁, FVC, and FEF_{25–75%} models, respectively; after adjusting for smoking, the λ values were 1.38, 1.28, and 1.02 for FEV₁, FVC, and FEF_{25–75%}, respectively. The genomic inflation λ values for the longitudinal association in Group 2 (β_{LD}) with no adjustment for smoking were: 1.19, 1.28, and 0.78 for FEV₁, FVC, and FEF_{25–75%}, respectively. After adjusting for smoking-related covariates, the λ values were as follows: 1.31, 1.31, and 0.85 for FEV₁, FVC, and FEF_{25–75%}, respectively. P-values denoted with an ^{NS} indicate that the probe was not Holm significant in that analyses.

Holm correction (Table 4-Group 1). Probe cg05575921 (aryl-hydrocarbon receptor repressor gene *AHRR*) gene, was associated with, FEV₁ and FEF_{25–75%}. The other, cg06126421 (immediate early response 3 gene, *IER3*), was associated with FEV₁. This probe also had a FDR significant association with FVC. The cross-sectional analysis had some under inflation (Table 4), yet p-values appeared to follow a uniform distribution (Figure 1 subplot A2, B2, C2, genomic inflation factor in caption of Table 4). There were eight significant probes when not adjusting for smoking (Table 4-Group 1). The β_{CS} estimates magnitudes decreased after adjustment for smoking. There were 5 (3 unique probes) associations when adjusting for smoking at the FDR level of 0.1 and 20 (9 unique probes) associations when not adjusting for smoking. When excluding current smokers, cg05575921 and cg06126421 still had their respective associations when adjusting for smoking (Table 4-Group2). There were 3 (2 unique probes) associations when adjusting for smoking and 14 (9 unique probes) associations when not adjusting for smoking the FDR level of 0.1.

There is a positive association between DNAm at cg05575921 and the lung function measures (FEV₁, FEF_{25–75%}) that is modified by smoking in our sample (Figure 2(a,b)). In current smokers, there is a positive linear association between lung function measures and DNAm (FEV₁ r^2 = 0.283, FEF_{25–75%} r^2 = 0.302). For former and non-smokers we do not see as strong a linear relationship (FEV₁ r^2 = 0.081, FEF_{25–75%} r^2 = 0.082). This strong difference by smoking status was not observed for cg06126421 and FEV₁ (Figure 2, subplot C, r^2 0.13 and 0.07 for current smokers and non-current smokers). There was a similar pattern on the M-value scale (Supplement Figure 1).

KORA replication

In NAS, there were 1377 probes with a FDR significant association (0.1 level) with FEV₁, FVC, or FEF_{25–75%}, either via β_{CS} or β_{LD} after adjusting for smoking. KORA had 1052 of these 1377 probes. Four probes replicated in KORA (FDR-significant in KORA and association in the same direction-

Figure 1. QQ plots of P-values for each spirometric model: those associated with β_{LD} are located to the left (a, c, and e) and those with β_{CS} on the right (b, d, and f).

Table 5, more information Supplement Table 3). These four probes replicated in the analysis restricted to KORA men. Only three replicated within the non-sex stratified analysis adjusting for sex. These three probes (cg05575921, cg06126421, and cg15342087) all had significant β_{CS} associations. These three probes association with FEV₁ and FEF_{25–75%} also replicated in older men in KORA (55 and up). Probe, cg01086847, was associated with lung function decline, but upon closer examination has a SNP (single nucleotide polymorphism) directly under it. Supplement Figure 2 displays cg15342087 which has a similar

pattern of association as cg06126421 in its association with FEV₁ (r^2 in current smokers 0.09 and 0.03 in non-current smokers in NAS) and is in the same loci (IER3). The *AHRR* probe also replicated in the analysis restricted to KORA women.

Sensitivity analysis

We next performed a sensitivity analysis where we removed probes with nearby SNPs and also individual outliers (over 1.5 times the interquartile range away from the 1st and 3rd quartile) for each probe. This was done adjusting for medications-

Table 4. Cross sectional Associations of DNAm on lung function (β_{CS}): Results for β_{CS} association in NAS, the cross-sectional Association of DNA. Table shows unique DNA methylation sites associated with a given Pulmonary Function Test (PFT) model organized by smoking analysis, unadjusted (left side) and adjusted for smoking (right side).

Probe (Gene)	Sample*	PFT†	Unadjusted for smoking		Adjusted for smoking		
			Effect Estimate	P-value	Effect Estimate	P-value	ICC (95% CI)
cg03636183 (<i>F2RL3</i>)	Group 1	FEF _{25–75%}	1.03	6.07e-09	0.75	8.31e-05 ^{NS}	0.83 (0.81–0.88)
cg05575921 (<i>AHRR</i>)		FEV ₁	1.81	9.20e-17	1.58	2.33e-11	0.91 (0.90–0.93)
		FVC	1.40	2.74e-09	1.29	8.82e-07 ^{NS}	
		FEF _{25–75%}	1.28	2.05e-19	1.17	3.57e-13	
cg05951221 (<i>ALPL2</i>)		FEF _{25–75%}	0.95	4.38e-08	0.67	2.97e-04 ^{NS}	0.89 (0.88–0.92)
cg06126421 (<i>IER3</i>)		FEV ₁	1.49	1.43e-12	1.26	9.83e-09	0.93 (0.92–0.95)
		FVC	1.35	2.02e-09	1.24	1.85e-07 ^{NS}	
		FEF _{25–75%}	0.89	2.84e-10	0.68	3.87e-06 ^{NS}	
cg21566642		FEV ₁	1.27	8.60e-09	1.00	1.62e-05 ^{NS}	0.9 (0.89–0.93)
		FEF _{25–75%}	0.90	6.37e-10	0.70	6.26e-06 ^{NS}	
cg15342087	Group 2	FEV ₁	3.24	1.79e-09	2.77	2.99e-07 ^{NS}	0.67 (0.64–0.75)
		FVC	3.15	2.90e-08	2.86	7.53e-07 ^{NS}	
cg14753356		FEV ₁	2.45	4.68e-08	1.94	2.02e-05 ^{NS}	0.69 (0.67–0.77)
cg24859433		FEV ₁	2.80	9.57e-09	2.34	1.82e-06 ^{NS}	0.69 (0.66–0.77)
cg03636183 (<i>F2RL3</i>)		FEF _{25–75%}	1.05	1.04e-07	0.74	3.02e-04 ^{NS}	0.83 (0.81–0.88)
cg05575921 (<i>AHRR</i>)		FEV ₁	1.96	7.41e-15	1.74	8.79e-11	0.91 (0.90–0.93)
		FVC	1.47	6.18e-08	1.39	2.13e-06 ^{NS}	
		FEF _{25–75%}	1.46	1.66e-18	1.26	3.76e-12	
cg06126421 (<i>IER3</i>)		FEV ₁	1.41	4.44e-10	1.21	2.14e-07 ^{NS}	0.93 (0.92–0.95)
		FEF _{25–75%}	0.87	7.28e-09	0.66	2.16e-05 ^{NS}	
cg21566642		FEF _{25–75%}	0.94	2.90e-09	0.72	1.09e-05 ^{NS}	0.9 (0.89–0.93)

* Sample: Group 1 included former, current, and never smokers, n = 633 participants across all three smoking classes, never (179 participants), former (426), and current smokers (28). Group 2 consists of just former and never smokers; this is restricting analysis to never (179 participants) and former (426) smokers. † The genomic inflation lambda (λ) values, with no adjustment for smoking for the cross-sectional association (β_{CS}), were 0.67, 0.91, and 0.82 for the FEV₁, FVC, and FEF_{25–75%} models, respectively; after adjusting for smoking, the λ values were 0.68, 0.94, and 0.83 for FEV₁, FVC, and FEF_{25–75%}, respectively. The genomic inflation λ values for the cross-sectional association (β_{CS}) in Group 2 with no adjustment for smoking were: 0.63, 0.89, and 0.93 for FEV₁, FVC, and FEF_{25–75%}, respectively. After adjusting for smoking-related covariates, the λ values were as follows: 0.63, 0.91, and 0.94 for FEV₁, FVC, and FEF_{25–75%}, respectively. P-values denoted with a ^{NS} indicate that the probe was not Holm significant in that analyses.

corticosteroids, beta adrenergic agonists, and muscarinic receptor antagonists. There were now 384,010 probes. This led to a different set of significant probes (Supplement Tables 4 and 5). There were 1527 (1344 unique probes) significant associations when adjusting for smoking and including current smokers. For β_{LD} , 10 probes were significant via Holm adjustment (Supplement Tables 4 and 5). Eight were significant in the previous analysis either via Holm or FDR adjustment. The two new probes were cg00437258 (*DAGLB* gene) and cg14219256 (*MYST4* gene). For β_{CS} , the two Holm significant probes adjusting for smoking were cg05575921 and cg06126421. For β_{CS} , there were 6 associations over 3 unique probes (cg05575921, cg06126421, and cg15342087). For β_{LD} , the associations were for 1341 unique probes, 704 of which were not significant in the initial analysis (Supplement Tables 4 and 5). Of the 1377 probes found significant in the original analysis adjusting for smoking,

281 were removed either due to having a SNP nearby or being cross-reactive.

Metastability

A subset of NAS men had a secondary DNAm measure which was used (in addition to individuals with only one DNAm observation) to assess our metastability assumption via estimating the intraclass correlation coefficient (ICC) in our 1377 probes. Probes cg05575921 (ICC 0.91) and cg06126421 (ICC of 0.93), had high ICCs suggestive of a level of metastability across time (Table 4). These two probes have been found stable in the literature, though low heritability was reported for cg05575921 [34]. Plotting DNAm beta-values for these probes at the next available time point suggested metastability (Figure 3). In cg15342087 we see less metastability (ICC estimate of 0.67), but the DNAm seen similar between time points (Figure 3). Supplement

Figure 2. DNAm association at cross-sectional probes of interest in relation to spirometric function and smoking status: (a) cg05575921 and the FEV₁ lung function model; (b) cg05575921 and FEV_{25–75%}; and (c) cg06126421 and FEV₁.

Table 6 gives the ICC for each probe significant in the original analysis (sheet 1) or in the sensitivity analysis (sheet 2).

Association with difference in DNAm age and chronological age

We next tested for the correlation between DNAm and Δ_{age} (DNAm age minus chronological age). In the 1377 probes, 663 were FDR significant (adjusting for 1377 tests, at FDR 0.05 level [32]). DNAm at cg05575921 had a negative correlation with Δ_{age} (Figure 4, quadrant IV). Of the 11 probes Holm significant in the NAS analysis adjusting for smoking, six were significantly associated with Δ_{age} (Supplement Table 7). In KORA, of the 663 probes associated with Δ_{age} in NAS, 497 were available. There were

36 probes that replicated in the KORA analysis (FDR-significant at 0.05 and the same Pearson correlation sign, Supplement Table 8). Probe cg05575921 was not significantly associated with Δ_{age} in KORA. In NAS, correlation between DNAm and Δ_{age} /age acceleration (residuals of regression of Δ_{age} and chronological age) were similar and are displayed in Supplement Table 9.

Discussion

In summary, in NAS there were 11 unique DNAm sites significant via Holm correction after adjusting for smoking-related covariates (Tables 3 and 4 Group 1). In KORA, 3 CpG sites reproduced (cg05575921, cg06126421, and cg15342087; comparison of effects in Table 5). We evaluated the scaled-effect estimate for one-unit change in the

Table 5. Replicated findings across the NAS and KORA cohorts with scaled parameter estimates.

Model	CpG Probe		NAS	KORA all	KORA men	KORA oldest men (55 and up)	KORA women	NAS ICC (95% CI)
β_{CS}								
FEF _{25–75%}	cg05575921	P-value	2.33E-13	1.48E-3	2.89E-03	3.02E-04	0.03	0.91 (0.90–0.93)
		Parameter Estimate	1.17	1.17	1.69	3.32	1.04	
FEV ₁	cg05575921	P-value	2.33E-11	1.90E-02	7.86E-03	8.00E-04	3.12E-02	0.91 (0.90–0.93)
		Parameter Estimate	1.59	0.49	0.84	1.71	0.53	
	cg06126421	P-value	9.83E-09	2.55E-02	4.34E-04	1.02E-02	0.185 ^{NS}	0.93 (0.92–0.95)
		Parameter Estimate	1.26	0.50	1.23	1.43	0.34	
FVC	cg15342087	P-value	2.99E-07	0.009	0.00052	0.03	0.368 ^{NS}	0.67 (0.64–0.75)
		Parameter Estimate	2.77	1.58	3.07	3.14	0.65	
	cg06126421	P-value	1.85E-07	0.585 ^{NS}	0.08	0.648 [*]	0.319 ^{NS}	0.67 (0.64–0.75)
		Parameter Estimate	1.24	0.13	0.66	0.27	0.28	
β_{LD}								
FVC	cg01086847	P-value	1.31E-04	0.00083 ^{NS}	5.10E-04	0.0113 ^{NS}	0.462 ^{NS}	0.98 (0.98–0.98)
		Parameter Estimate	−0.241	−1.48	−2.60	−3.10	−0.30	

P-values denoted with an * indicate that the probe did not replicate in that KORA sample.

standard deviation of this data for cg05575921 and cg0612642. We did not evaluate cg15342087 as it was within 200 BP of cg06126421. For the AHRR probe, it had a similar magnitude of association in its FEV₁ association and KORA men/older men (NAS: 0.141, KORA Men and older men: 0.1 and 0.21 respectively). We did not see much similarities for its FEF_{25–75%} association, though they were still all in the same direction upon scaling. In cg06126421 there was a similar pattern for its association with FEV₁ (NAS: 0.119, KORA men and older men: 0.127 and 0.146 respectively) but not as much so for its FVC association. The similarity of the scaled associations between NAS and KORA men adds strength to the credibility of the associations in addition to them replicating. In evaluating the correlation of these DNAm marks with biological aging [30], 6 of the 11 NAS loci were significantly associated with Δ_{age} (Supplement Table 9). None of these 6 CpG sites overlapped with the CpGs in the algorithm to compute DNAm age [30]. Collectively, this data provide additional evidence suggesting the utility of DNAm age as not only a predictor of chronological age in humans, but also as a potential biomarker of aging-related characteristics, such as lung function.

Examining the β_{LD} probes associated in NAS, there were 12 probes significantly associated with FVC or FEV₁. There were seven probes associated with FVC regardless of adjustment for smoking (Table 3). These seven probes had varying evidence of metastability (ICC 7.9e-14 to 0.99). One probe with high ICC was cg13532885 (ICC 0.93 SYN1) which was also associated with Δ_{age} . The ICC did not appear to vary by smoking status (0.9 and 0.94 in never and former smokers). Another probe, cg05644990, had very low ICC (7.9e-14, 4.34E-14, and 7.02e-17 in overall, never, and former respectively) and was also associated with Δ_{age} . These two probes may implicate different biological functions that influence PFTs. These findings for cg05644990 and cg13532885 did not reproduce in KORA. This failure to reproduce may be due to: the older age of the NAS, the smaller time interval studied in KORA, or the lack of metastability. In FEV₁, β_{LD} significant probes were much more likely to be affected by smoking variables as there were no probes significant regardless of smoking adjustment. Probe cg18476993 (ICC of 0.64, never vs former: 0.51 vs 0.67) was the only probe significant for FEV₁ after adjusting for smoking in the larger group. The probes associated with FEV₁ were not associated with FVC suggesting an association with

Figure 3. Visualization of DNAm metastability across two visits in the NAS. Examining for cg05575921 (a) cg06126421 (b), and cg15342087 (c), on study participants with a second DNAm observation. X-axis displays DNAm at first visit, y-axis at second visit.

airway function and less with the alveolar region of these probes. This agrees with the pathophysiologic understanding of cigarette smoke related COPD. The remaining probes associated with FEV₁, were likely being driven by the small number of smokers in our study, as they were not significant in the analyses in just former and never smokers (Table 3-Group 2).

For the cross-sectional analysis (β_{CS}), results were mainly observed for FEV₁ and FEF_{25–75%}, with little association in FVC (Table 4-Group 1). This suggests these probes are related to airway function. Upon adjustment for smoking we saw the magnitude of association attenuate (Table 4). Observed hits were in the three probes cg06126421, cg15342087, and cg05575921. These

probes were associated with FEV₁. All three were associated with FVC when not accounting for smoking status, with only cg06126421 still significant after adjustment. The AHRR probe was the only one associated with FEF_{25–75%}. It should be noted that there has been evidence of SNPs near cg05575921, though these SNPs did not appear to confound the analysis in a reported study on maternal smoking [35]. Probe cg15342087, while not HOLM significant, did pass the FDR threshold of 0.1 with FEV₁ and did replicate in KORA. The other probes in Table 4 are likely being driven by smoking status.

Poor lung function is either due to a low plateau or a rapid age-related decline and is a predictor for lung morbidity and mortality. In our analysis,

Figure 4. Scatterplot of DNAm beta estimates and Pearson correlation coefficients for CpGs significantly associated with lung function and significantly correlated with Δ_{age} in the NAS. In the bottom right quadrant, cg05575921 (*AHRR* gene) is highlighted in red, as it was the only CpG whose association with lung function persisted in both the NAS and KORA, while also being significantly correlated with Δ_{age} in the NAS.

there was no overlap between probes associated with cross sectional or longitudinal decline in FEV₁, FVC, and FEF_{25–75%}. This suggests that the β_{CS} significant probes are mainly determined or associated with the growth phase of an individual, i.e. the leveling off of lung function around the age of 20 as discussed in the introduction and are only slightly modulated by the age-related decline. We can see this in the differences in the associations, with the probes with a significant β_{LD} association being solely with measures of volume related spirometric indices (FVC or FEV₁). Meanwhile, the probes found with a significant β_{CS} association, were mainly associated with lung function parameters that are reflective of airway function, with only one being associated with FVC.

While our two top hits, cg05575021 and cg06126421, appeared metastable (high ICC and reported in the literature [34]) with our study design it is impossible to prove our metastable assumptions. These associations may be driven by a persistent impact of smoking. We ran a stratified analysis of the three cross-sectional probes that replicated in KORA by former or never smokers in NAS. Testing DNA methylation at cg05575921 with FEV₁ and FEF_{25–75%} in never smokers was nominally significant (p-value of 0.04 and 0.07 respectively). DNAm at cg06126421 was not associated with FEV₁ and FVC in never

smokers (p-value of 0.33 and 0.49 respectively). In addition, the test for DNAm at cg15342087 and FEV₁ was not significant (p-value of 0.26). In the analysis restricted to former smokers, we continued to adjust for pack-years and in addition include time since quitting. DNAm at cg05575921 was still significantly associated with FEV₁ and FEF_{25–75%} (p-values of 5.3e-08 and 2.3e-07 respectively), as was cg06126421 with FEV₁ and FVC (p-values of 4.28e-06 and 1.03e-04 respectively), and cg15342087 (p-value of 4.2e-05 for FEV₁). However, we cannot rule out persistent smoking effects that are not adequately modeled by pack years or time since quitting. There could be an issue of power as there are fewer never smokers in our analysis than former smokers (n = 179/426 never vs former). We believe the analysis is still useful as it highlights potential biological pathways related to lung function that future studies can elucidate.

Another potential limitation to our study is potential bias due to unmeasured diseases. To examine this, we analyzed the three probes (cg06126421, cg15342087, and cg05575921) adjusting for chronic lung diseases (either asthma and/or COPD defined by FEV₁/FVC < 0.7). This led to a decrease in effect size, about 25% lower, but still highly significant association (Supplement Table 10). This suggests that the observed

associations are not primarily driven by chronic lung diseases but appear to be amplified by these diseases. In addition, the differing study designs of KORA and NAS may be the reason we failed to replicate any lung function decline associations within KORA. The associations in NAS may be indicative of previous smoking exposure, while in KORA it may be interactive associations.

This difference in study design is also a strength as the NAS cohort was assessed retrospectively (*i.e.* blood samples were drawn at the end of a series of lung function testing at multiple times over 10 plus years for each individual), whereas KORA was assessed prospectively (*i.e.* blood samples were taken at baseline and followed by lung function testing at multiple times over a similar time-span) (Table 1). This design helped in facilitating: (i.) the discovery of potentially metastable DNAm sites retrospectively for evaluating lung function decline within an elderly population (NAS cohort); and (ii.) replication and confirmation of these methylation sites that may precede lung function decline in younger humans, too (KORA). However, this differing study designs of KORA and NAS may be the reason we failed to replicate any lung function decline associations within KORA. The associations in NAS may be indicative of previous smoking exposure, while in KORA it may be interactive associations. For our top three probes that replicated in KORA there was strong evidence of metastability in our sample (Table 4). Throughout our work, metastability was assessed by calculating the ICC of DNAm from blood samples taken 3–5 years apart (Tables 3 and 4); and reported in the literature [34].

All of these findings led us to a novel working model, which can serve to integrate and contextualize our results. As discussed in the literature [22,23], DNAm status is highly sensitivity to smoking status (seen in our study Figure 2, subplot A and B). Here we show that it is also associated with PFTs, and, in some settings, with aging (Δ_{age}). These findings are plausible as *AHRR* has been implicated in the detoxification of tobacco smoke components protecting against oxidative stress and inflammation as well as having tumor suppressing effects [36]. Mechanistically, *AHRR* has been shown to be induced eg. by wood smoke exposure leading to a reduction of AHR-mediated anti-

inflammatory action of the arachidonic acid pathway [37]. The data (Figures 2 and 4 and Supplement Table 8) shows that high DNAm status at cg05575921 *in vivo* correlates with greater pulmonary function and decreased aging (Figure 4). To our knowledge, this is the first time that directional changes in aging have been associated with DNAm status at cg05575921 in the literature. The other consistent probe, cg06126421 (*IER3*) did not appear to have an effect modification with smoking or be associated with Δ_{age} . This probe has been found to be associated with smoking and all cause mortality in previous studies [38–41]. These two CpG loci may inform pulmonary function studies in complementary and nuanced ways, given that they may represent different epigenetic networks *in vivo*. The probe cg15342087 is near cg06126421 and has been found as well to be associated with all-cause mortality as well [41]. *IER3* has been described to be involved in the immune response and inflammation, however, it is induced by a variety of external and internal factors [42] and a possibly more specific role for the respiratory system remains to be elucidated. Further investigation into DNAm at the *AHRR* gene may help to uncover tools/strategies to better dissect the link between lung function, aging, and smoking. DNAm at the *IER3* gene, however, may be more directly linked to lung function and respiratory outcomes irrespective of factors such as smoking/aging.

Materials and methods

Normative aging study cohort (NAS)

The study included 633 participants from the NAS, a closed cohort of community-dwelling men living within the Greater Boston metropolitan area [43]. We obtained and analyzed DNA samples specifically from 1999 to 2007. As detailed in Table 1, participants underwent up to 4 total examinations (or less) with a 3- to 5-year space between visits. Blood was isolated at the last visit and a number of measures were collected, such as age, education, height, weight, and medication status (*e.g.*, use of sympathomimetic alpha and beta, anticholinergics, *etc.*). Pathophysiological lung conditions – *e.g.*, asthma,

chronic bronchitis, emphysema, *etc.* – were evaluated by a physician, and smoking data was collected *via* the American Thoracic Society (ATS) questionnaire [44]. We used FVC, FEV₁, and FEF_{25–75%} to assess pulmonary function. These spirometric parameters were measured following the ATS guidelines as previously reported [45,46], and were measured over time going back at most four visits for our study. Further, a methacholine challenge test was administered to assist in the diagnosis of asthma [47]. This study was approved by the Institutional Review Boards of all relevant institutions, and all participants provided signed informed consent.

Cooperative health research in the region of Augsburg cohort (KORA)

The present work included participants from the population-based study KORA F4 and its follow-up study, KORA F4L. Blood for buffy coat was sampled in an EDTA-Falcon-Tube (9 ml) and collected via venepuncture from the cubital fossa. Experimental design details about these studies has been reported previously [48–50]. Information on the epigenetic design and DNA methylation processing has been published in Zeilinger et al and Panni et al [23,51]. Briefly, spirometry was performed in 1321 human participants aged 41–63 years from the Augsburg region between 2006–2008 in KORA F4. Spirometry was then re-measured after approximately 3.1 years in participants of the KORA F4L study. Spirometry was performed in accordance with European Respiratory Society and ATS guidelines [45,49,52], and information about the presence of respiratory diseases, medication and smoking status was assessed by a standardized interview and/or questionnaire (as published previously). The Ethics Committee of the Bavarian Medical Association approved both studies, KORA F4 and KORA F4L, and informed consent was obtained from all research study participants [51]. Of this 1321 study participants, 868 had DNA methylation and passed quality control.

Detection and analysis of DNA methylation

Buffy coat DNA was isolated from each sample *via* the QIAamp DNA Blood Kit (QIAGEN, Valencia,

CA) and a 0.5 µg aliquot was bisulfite converted with the EZ-96 DNA Methylation Kit (Zymo Research, Orange, CA). In the NAS, this was done on blood collected between 1999 and 2007. In the NAS, DNA methylation was detected by the Infinium HumanMethylation450 BeadChip platform (Northwestern University, Feinberg School of Medicine, Center for Genetic Medicine). Technical effects due to the plate/chip were minimized by utilizing a two-stage age-stratified algorithm to randomize the samples, thereby ensuring comparable age distribution across plates/chips. Specifically, we arranged the samples by stratified randomization onto blocks for the Illumina BeadChips (grouped in sets of 12) and analytic plates (sets of 8 BeadChips), ensuring that repeated DNA measures from a person were located on the same BeadChip and stratified by age quartile; we also employed statistical validations to confirm evenness across plates/chips by age and other cohort characteristics.

Quality control samples consisted of replicate pairs and a single sample that was run within and between plates/chips to help detect batch effects. Analytic plates were run consecutively, by the same technician, and processed and read on the same scanner. Quality control approaches also included the detection and removal of 15 DNA samples and 949 probes *via* the pfilter command in the Bioconductor watermelon package [53], which excluded DNA samples containing >1% of probes with detection P-values >0.05 and probes having >1% of samples with detection P-value >0.05 (after omitting samples excluded above). Furthermore, we also excluded probes with specific design and/or annotation, namely: 65 with genotyping function; 3091 used for detecting CpH methylation (CpH are non CpG methylation sites); and 3688 containing a SNP in the last 10 bases with a minor allele frequency greater than 0.01 in the CEU reference set. A number of these probes were already excluded by the pfilter command, so after these steps we finally obtained 477,927 probes, *i.e.* ~98.4% out of 485,512, which were used to obtain DNA methylation. Lastly, we applied a 3-part, pre-processing pipeline to our data: (i.) background correction *via* the out-of-band (noob) method by Triche *et al* [54].; (ii.)

dye-bias adjustment by the Bioconductor methylumi package [55]; and (iii.) probe-type correction with BMIQ according to Teschendorff *et al.* (2013) [56], as provided by watermelon [53]. DNA samples from the KORA cohort were analyzed similarly to the detailed scheme that is provided above for the NAS, as is reported in the literature [51].

Biostatistical analyses

In the NAS, we analyzed $FEF_{25-75\%}$, FEV_1 , and FVC. We fit a mixed-effect model to analyze the data, holding the DNAm at the probe constant and using lung function measures at the time of blood drawn and the three prior visits of blood being drawn as the dependent variable. DNAm was the independent variable of interest. Doing this allowed us to account for the correlation between lung function measures from the same individual. $FEF_{25-75\%}$ was square root transformed to better approximate a normal distribution. The predictor for lung function was percent of DNAm (*i.e.* the beta-value). This analysis was done using the lme4 package in R [57]. Given the large number of observations, we treated the resulting t-statistic as normal to perform inference. The model is given below for the DNA methylation at probe j ($M_{i,j}$) for spirometric measure at time k for individual i ($Y_{i,k}$). X_i is the set of confounders. The follow up time (time between first included visit and current visit) is $t_{i,k}$, with $t_{i,k} = 0$ if $k = 1$. A study participant could have k up to 4.

$$E(Y_{i,k}|X, M, t) = \beta_0 + X_i\beta_C + t_{i,k}\beta_t + M_{i,j}\beta_{CS} + t_{i,k}M_{i,j}\beta_{LD}$$

We included an interaction between the beta-value and follow-up time. There are two associations of interest: (i.) the *cross-sectional association* of DNAm on lung function, β_{CS} (when follow-up time is 0, CS for cross sectional), and (ii.) the *interaction* between DNAm and follow-up time, β_{LD} (LD for longitudinal decline). Furthermore, we examined models without and with adjustment for smoking-related confounders (smoking status and pack years). General covariates common to all models were: season, follow up time, day of the week, plate, chip, weight mean centered, weight mean centered squared, the natural log of height,

baseline age, vitamin C intake not from vitamin pills, maximum years of education, and taking any of the following drugs-corticosteroids (inhaled (ICS) or systemic), beta-2-adrenergic agonists (typical LABA), sympathomimetic alpha AHFS, or long acting muscarinic antagonist (LAMA) (Tables 1 and 2). Cell-type proportions were estimated *via* Houseman *et al* [31], (Table 2), and were included as additional covariates. To see if results were influenced by the (current) smokers, we reran the models restricted to the former and never smokers. For each DNAm probe, we again tested for the cross-sectional (β_{CS}) and interaction (β_{LD}) terms defined above.

A probe was considered genome-wide-significant if it had a Holm adjusted p-value less than 0.05 [33] or significant at an FDR level of 0.10 [32]. Holm procedure is a more conservative approach than FDR and controls the family wise error rate. Replication was performed in the KORA cohort. Replication in KORA was done in: (A) use all participants/ages (controlling for sex); (B) restricted to men; (C) only the oldest men (55 and up, to mirror the NAS); (D) restricted to women. This methodology was adopted to ensure comparability of our results, as the NAS is an all-male cohort. We adjusted our KORA models for cigarette use (smoking status and pack-years).

To compute DNAm metastability, we utilized second observations of methylation on a small proportion of NAS individuals while still using the individuals' observations with one observation. We computed the intraclass correlation coefficient (ICC) using random effects model. for each probe of interest to assess DNAm stability over time. The ICC was calculated using random effect modeling and measuring the proportion of total variation coming from the shared variance. The ICC gives a measure of the validity of the metastability assumption. A value closer to one indicates metastability.

For sensitivity, we re-ran the analysis removing cross-reactive probes, probes that were associated with a SNP with MAF greater 5%, non-specific probes, polymorphic CpG sites, and probes with a SNP within 10BP of the target site [58]. We also dropped sympathomimetic alpha AHFS from our definition of medicine taking as it is not often taken with COPD. This brought the number of probes

down to 384,010. For each probe, we reran the analysis removing any outliers where the DNAm beta-value was greater than 3 interquartile ranges outside the 25% and 75% quartiles for that probe.

Calculation of dnam age, Δ_{age} , and relationship with lung function associated CpGs

DNAm age is a predictor of chronological age calculated by using DNAm measurements from 353 CpG sites shared by both the Illumina HumanMethylation27 and HumanMethylation450 BeadChip platforms [30]. This biocalculator has been shown to maintain predictive accuracy across human tissues [30]. We calculated an age-difference variable, ' Δ_{age} ', by subtracting a participant's DNAm age (predicted age given their epigenetic profile) from his/her chronological age (their actual age). A positive Δ_{age} is illustrative of enhanced/increased aging, and *vice versa*. We tested if any of the significant probes (β_{CS} or β_{LD}) from the lung function models, were also significantly associated with Δ_{age} . This was done via a Pearson correlation test between the two measures. We also examined age acceleration which are the residuals after regression biological age from DNAm age.

Acknowledgments

We thank all of the human participants involved in both the NAS and KORA cohorts: we are grateful for their participation and cooperation. We also thank the clinical staff and scientific teams associated with these cohorts based in the USA and Germany for their excellent care of all participants. The NAS computations in this paper were run on the Odyssey cluster supported by the FAS Division of Science, Research Computing Group at Harvard University. Lastly, we also thank the Genomics Core Facility at Northwestern University for their technical assistance regarding the Infinium HumanMethylation450K BeadChip technology.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This study was funded by the National Institute of Environmental Health Sciences (NIEHS) (R01ES025225;

R01ES021733 and R01ES015172). JJC was supported by a Ruth L. Kirschstein National Research Service Award (NRSA) for Individual Postdoctoral Fellows F32ES024068 from the National Institute of Environmental Health Science, National Institutes of Health, and by Training Grant T32ES007069 in Environmental Epidemiology from the National Institute of Environmental Health Science, National Institutes of Health. RTB was funded by R35CA197449, T32GM074897, T32ES007142, and T32CA094880. The US Department of Veterans Affairs (VA) Normative Aging Study (NAS) is supported by the VA Cooperative Studies Program/ERIC, Massachusetts Veterans Epidemiology Research and Information Center (MAVERIC). Additional support to the VA-NAS was provided by the US Department of Agriculture, Agricultural Research Service (contract 53-K06-510). AS3 is supported by a VA Clinical Sciences Research and Development Senior Research Career Scientist award. Funding sources was not involved in study design, collection, analysis and interpretation of data; in the writing of the manuscript; and in the decision to submit the article for publication. KORA work was also supported by the Comprehensive Pneumology Center Munich (CPC-M) as member of the German Center for Lung Research and the Competence Network Asthma and COPD (ASCONET), network COSYCONET (subproject 2, BMBF FKZ 01GI0882) funded by the German "Bundesministerium für Bildung und Forschung" (Federal Ministry of Education and Research, BMBF). The KORA study was initiated and financed by the Helmholtz Zentrum München – German Research Center for Environmental Health, which was funded by the German Federal Ministry of Education and Research (BMBF) and by the State of Bavaria. J.N-E is also supported by a NIH/NIA Ruth L. Kirschstein National Research Service Award (1 F31AG056124-01A1). Work by JNH at the Harvard Chan Bioinformatics Core was partly supported by Award Number P30 ES00002 from the National Institute of Environmental Health Sciences. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of Environmental Health Sciences or the National Institutes of Health.

Author contributions

J.J.C. and R.T.B. contributed equally to this work as first authors; J.J.C. and R.T.B. conceived of the study, participated in study design and coordination, and wrote the manuscript. R.T.B. analyzed the data and generated figures. R.T.B., T.P., and J.N.H., designed and performed bioinformatic analyses. J.N-E., A.C.J., E.C., S.K., S.W., S.K., N.J., Y.Z., L.H., D.L.D., A.A.L., P.S.V., and A.P. participated in study design and expert input regarding data analysis throughout the project's life course. As senior faculty mentors, X.L., J.S., H.S., and A. A.B. co-conceived of the study with the co-first authors,

participated in study design, helped to coordinate the study, and supervised all aspects of the work.

Declarations

Information pertaining to ethical approval statements is listed in the Materials and methods section. All authors read and approved the manuscript.

ORCID

Allan C. Just <http://orcid.org/0000-0003-4312-5957>

Augusto A. Litonjua <http://orcid.org/0000-0003-0422-5875>

References

- Soler Artigas M, Wain LV, Miller S, et al. Sixteen new lung function signals identified through 1000 Genomes Project reference panel imputation. *Nat Commun.* **2015**;6:8658.
- Weiss ST. Lung function and airway diseases. *Nat Genet.* **2010**;42:14–16.
- Wain LV, Shrine N, Artigas MS, et al. Genome-wide association analyses for lung function and chronic obstructive pulmonary disease identify new loci and potential druggable targets. *Nat Genet.* **2017**;49:416–425.
- Maher B. Personal genomes: the case of the missing heritability. *Nature.* **2008**;456:18–21.
- Morrow JD, Cho MH, Hersh CP, et al. DNA methylation profiling in human lung tissue identifies genes associated with COPD. *Epigenetics.* **2016**;11:730–739.
- Obeidat M, Hao K, Bosse Y, et al. Molecular mechanisms underlying variations in lung function: a systems genetics analysis. *Lancet Respir Med.* **2015**;3:782–795.
- Piletic K, Kunej T. MicroRNA epigenetic signatures in human disease. *Arch Toxicol.* **2016**;90:2405–2419.
- Joehanes R, Just AC, Marioni RE, et al. Epigenetic signatures of cigarette smoking. *Circ Cardiovasc Genet.* **2016**.
- LaCroix AZ, Omenn GS. Older adults and smoking. *Clin Geriatr Med.* **1992**;8:69–87.
- Kerstjens HA, Rijcken B, Schouten JP, et al. Decline of FEV1 by age and smoking status: facts, figures, and fallacies. *Thorax.* **1997**;52:820–827.
- Speizer FE, Tager IB. Epidemiology of chronic mucus hypersecretion and obstructive airways disease. *Epidemiol Rev.* **1979**;1:124–142.
- Godtfredsen NS, Lam TH, Hansel TT, et al. COPD-related morbidity and mortality after smoking cessation: status of the evidence. *Eur Respir J.* **2008**;32:844–853.
- Vercelli D. Does epigenetics play a role in human asthma? *Allergol Int.* **2016**;65:123–126.
- Gruzieva O, Merid SK, Melen E. An update on epigenetics and childhood respiratory diseases. *Paediatr Respir Rev.* **2014**;15:348–354.
- Comer BS, Ba M, Singer CA, et al. Epigenetic targets for novel therapies of lung diseases. *Pharmacol Ther.* **2015**;147:91–110.
- Pal S, Tyler JK. Epigenetics and aging. *Sci Adv.* **2016**;2:e1600584.
- Sin DD, Wu L, Man SF. The relationship between reduced lung function and cardiovascular mortality: a population-based study and a systematic review of the literature. *Chest.* **2005**;127:1952–1959.
- Hole DJ, Watt GC, Davey-Smith G, et al. Impaired lung function and mortality risk in men and women: findings from the Renfrew and Paisley prospective population study. *BMJ.* **1996**;313: 711–715. discussion 5–6.
- Lee HM, Le H, Lee BT, et al. Forced vital capacity paired with Framingham Risk Score for prediction of all-cause mortality. *Eur Respir J.* **2010**;36:1002–1006.
- Higgins M. Risk factors associated with chronic obstructive lung disease. *Ann N Y Acad Sci.* **1991**;624:7–17.
- Hoyert DL, Xu J. Deaths: preliminary data for 2011. *Natl Vital Stat Rep.* **2012**;61:1–51.
- Fasanelli F, Baglietto L, Ponzi E, et al. Hypomethylation of smoking-related genes is associated with future lung cancer in four prospective cohorts. *Nat Commun.* **2015**;6:10192.
- Zeilinger S, Kuhnel B, Klopp N, et al. Tobacco smoking leads to extensive genome-wide changes in DNA methylation. *PLoS One.* **2013**;8:e63812.
- Bauer T, Trump S, Ishaque N, et al. Environment-induced epigenetic reprogramming in genomic regulatory elements in smoking mothers and their children. *Mol Syst Biol.* **2016**;12:861.
- Gibbs K, Collaco JM, McGrath-Morrow SA. Impact of tobacco smoke and nicotine exposure on lung development. *Chest.* **2016**;149:552–561.
- Drummond MB, Hansel NN, Connett JE, et al. Spirometric predictors of lung function decline and mortality in early chronic obstructive pulmonary disease. *Am J Respir Crit Care Med.* **2012**;185:1301–1306.
- Donaldson GC, Seemungal TA, Bhowmik A, et al. Relationship between exacerbation frequency and lung function decline in chronic obstructive pulmonary disease. *Thorax.* **2002**;57:847–852.
- Mannino DM, Buist AS, Petty TL, et al. Lung function and mortality in the United States: data from the First National Health and Nutrition Examination Survey follow up study. *Thorax.* **2003**;58:388–393.
- Anthonisen NR, Connett JE, Murray RP. Smoking and lung function of Lung Health Study participants after 11 years. *Am J Respir Crit Care Med.* **2002**;166:675–679.
- Horvath S. DNA methylation age of human tissues and cell types. *Genome Biol.* **2013**;14:R115.

31. Houseman EA, Accomando WP, Koestler DC, et al. DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics*. 2012;13:86.
32. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J Royal Stat Soc Ser B (Methodological)*. 1995;57:289–300.
33. Holm S. A simple sequentially rejective multiple test procedure. *Scand J Stat*. 1979;6:65–70.
34. Shah S, McRae AF, Marioni RE, et al. Genetic and environmental exposures constrain epigenetic drift over the human life course. *Genome Res*. 2014;24:1725–1733.
35. Gonseth S, de Smith AJ, Roy R, et al. Genetic contribution to variation in DNA methylation at maternal smoking-sensitive loci in exposed neonates. *Epigenetics*. 2016;11:664–673.
36. Traboulsi H, Guerrina N, Iu M, et al. Inhaled pollutants: the molecular scene behind respiratory and systemic diseases associated with ultrafine particulate matter. *Int J Mol Sci*. 2017;18:243.
37. Awji EG, Chand H, Bruse S, et al. Wood smoke enhances cigarette smoke-induced inflammation by inducing the aryl hydrocarbon receptor repressor in airway epithelial cells. *Am J Respir Cell Mol Biol*. 2015;52:377–386.
38. Elliott HR, Tillin T, McArdle WL, et al. Differences in smoking associated DNA methylation patterns in South Asians and Europeans. *Clin Epigenetics*. 2014;6:4.
39. Shenker NS, Polidoro S, van Veldhoven K, et al. Epigenome-wide association study in the European Prospective Investigation into Cancer and Nutrition (EPIC-Turin) identifies novel genetic loci associated with smoking. *Hum Mol Genet*. 2013;22:843–851.
40. Zhang Y, Schottker B, Florath I, et al. Smoking-associated DNA methylation biomarkers and their predictive value for all-cause and cardiovascular mortality. *Environ Health Perspect*. 2016;124:67–74.
41. Zhang Y, Wilson R, Heiss J, et al. DNA methylation signatures in peripheral blood strongly predict all-cause mortality. *Nat Commun*. 2017;8:14617.
42. Arlt A, Schafer H. Role of the immediate early response 3 (IER3) gene in cellular stress response, inflammation and tumorigenesis. *Eur J Cell Biol*. 2011;90:545–552.
43. Bell B, Rose CL, Damon A. The Veterans Administration longitudinal study of healthy aging. *Gerontologist*. 1966;6:179–184.
44. Ferris BG. Epidemiology standardization project (American Thoracic Society). *Am Rev Respir Dis*. 1978;118:1–120.
45. Miller MR, Hankinson J, Brusasco V, et al. Standardisation of spirometry. *Eur Respir J*. 2005;26:319–338.
46. Sparrow D, O'Connor G, Colton T, et al. The relationship of nonspecific bronchial responsiveness to the occurrence of respiratory symptoms and decreased levels of pulmonary function. The normative aging study. *Am Rev Respir Dis*. 1987;135:1255–1260.
47. Chatham M, Bleecker ER, Norman P, et al. A screening test for airways reactivity. An abbreviated methacholine inhalation challenge. *Chest*. 1982;82:15–18.
48. Holle R, Happich M, Lowel H, et al. KORA—a research platform for population based health research. *Gesundheitswesen*. 2005;67 Suppl 1:S19–25.
49. Karrasch S, Flexeder C, Behr J, et al. Spirometric reference values for advanced age from a South german population. *Respiration*. 2013;85:210–219.
50. Tang W, Kowgier M, Loth DW, et al. Large-scale genome-wide association studies and meta-analyses of longitudinal change in adult lung function. *PLoS One*. 2014;9:e100776.
51. Panni T, Mehta AJ, Schwartz JD, et al. A genome-wide analysis of DNA methylation and fine particulate matter air pollution in three study populations: KORA F3, KORA F4, and the normative aging study. *Environ Health Perspect*. 2016;124:983.
52. Miller MR, Crapo R, Hankinson J, et al. General considerations for lung function testing. *Eur Respir J*. 2005;26:153–161.
53. Pidsley R, Wong CCY, Volta M, et al. A data-driven approach to preprocessing Illumina 450K methylation array data. *BMC Genomics*. 2013;14:293.
54. Triche TJ Jr., Weisenberger DJ, Van Den Berg D, et al. Low-level processing of Illumina Infinium DNA Methylation BeadArrays. *Nucleic Acids Res*. 2013;41:e90.
55. Davis S, Du P, Bilke S, et al. methylumi: handle Illumina methylation data. 2015.
56. Teschendorff AE, Marabita F, Lechner M, et al. A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data. *Bioinformatics*. 2013;29:189–196.
57. Bates D, McHler M, Bolker BM, et al. Fitting linear mixed-effects models using lme4. *J Stat Softw*. 2015;67:1–48.
58. Chen YA, Lemire M, Choufani S, et al. Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray. *Epigenetics*. 2013;8:203–209.