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

\*Carmona, Juan Jose,1,2,3 carmona@mail.harvard.edu

\*#Barfield, Richard T.,4 rbarfiel@fredhutch.org

Panni, Tommaso,5 tommaso.panni@gmail.com

Nwanaji-Enwerem, Jamaji, C.,1,6 jnwanajienwerem@g.harvard.edu

Just, Allan C.,7 allan.just@mssm.edu

Hutchinson, John N.,8 jhutchin@hsph.harvard.edu

Colicino, Elena,9 elena.colicino@mssm.edu

Karrasch, Stefan,5,10,11  stefan.karrasch@helmholtz-muenchen.de

Wahl, Simone,12 simwahl@gmail.com

Kunze, Sonja, 5,12  sonja.kunze@helmholtz-muenchen.de

Jafari, Nadereh,13 nadereh.jafari@gmail.com

Zheng, Yinan,14  y-zheng@northwestern.edu

Hou, Lifang,14 l-hou@northwestern.edu

DeMeo, Dawn L.,15 redld@channing.harvard.edu

Litonjua, Augusto A.,15 augusto.litonjua@channing.harvard.edu

Vokonas, Pantel S.,16 pantel.vokonas@va.gov

Peters, Annette,1,5 peters@helmholtz-muenchen.de

Lin, Xihong,8,17 xlin@hsph.harvard.edu

Schwartz, Joel,1 jschwrtz@hsph.harvard.edu

†Schulz, Holger,5,11 schulz@helmholtz-muenchen.de

†Baccarelli, Andrea A.,1,18 ab4303@cumc.columbia.edu

\*Designates equal contribution first author

# Designates corresponding author

†Equal senior authors

1. Department of Environmental Health and Program in Quantitative Genomics, Harvard T.H. Chan School of Public Health, Boston, MA, USA
2. Center for Bioethics, Harvard Medical School, Boston, MA, USA
3. Dana-Farber/Harvard Cancer Institute, Boston, MA, USA
4. Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, WA, USA
5. Institute of Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany
6. MD-PhD Program, Harvard Medical School, Boston, MA, USA
7. Preventive Medicine, Icahn School of Medicine at Mount Sinai, New York City, NY, USA
8. Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA, USA
9. Departments Environmental Medicine & Public Health, Division of Biostatistics, Icahn School of Medicine at Mount Sinai, New York City, NY, USA
10. Institute and Outpatient Clinic for Occupational, Social and Environmental Medicine, Ludwig-Maximilians-Universität, Munich, Germany
11. Comprehensive Pneumology Center Munich, Member of the German Center for Lung Research, Munich, Germany
12. Research Unit of Molecular Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany
13. Center for Genetic Medicine, Feinberg School of Medicine, Northwestern University, Chicago, IL, USA
14. Center for Population Epigenetics, Robert H. Lurie Comprehensive Cancer Center and Department of Preventive Medicine, Feinberg School of Medicine, Northwestern University, Chicago, IL, USA
15. Channing 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
16. VA Normative Aging Study, Veterans Affairs Boston Healthcare System and the Department of Medicine, Boston University School of Medicine, Boston, MA, USA
17. Department of Statistics, Harvard University, Cambridge, MA, USA
18. Departments of Epidemiology and Environmental Health Sciences, Columbia University Mailman School of Public Health, New York City, 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 performed 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 Epigenetic Wide Association Study (EWAS) between DNA methylation at CpG dinucleotides and longitudinal lung function (FVC, FEV1, FEF25-75%) decline and aging with initial discovery in the NAS (Normative Aging Study), and replication in the KORA (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.

**KEYWORDS**

Biomarker; DNA methylation; pulmonary function; lung function decline

**Introduction**

A number of genome-wide association studies (GWAS) 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 epigenetic hypothesis of dfisease5, 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 aging16. 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, given that 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. With respect to lung function in this study, we focused on three spirometric indices that provide slightly different information. 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 (FEF25-75% , liters per minute), measures the conducting airway function, and is determined by the size of the airways. A low flow rate of FEF25-75% reflects airflow limitation usually corresponding to airway narrowing or obstruction of the airways. Finally, forced expiratory volume in 1 second (FEV1, 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 employed linear mixed models to perform an epigenome-wide association study (EWAS) of the association between DNAm at CpG dinucleotides and longitudinal lung function decline in the Normative Aging Study (NAS), a cohort of elderly men residing in the greater-Boston metropolitan area. We used a traditional test-replication approach: initial discovery phase 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, FEF25-75, or FEV1) 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 performed an analyses to consider any correlative relationships between methylation of particular CpG dinucleotides of interest, pulmonary function, and an age-difference variable constructed from “DNA methylation age” (DNAm age) 30.

**Results**

Briefly, the analysis consisted of 633 male study participants in NAS who contributed 2245 observations (average 3.5 observations/participant) over about one 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 study participants ranging from 41-63 years with 47% male. More detailed study population characteristics are provided in Tables 1 and 2. Table 1 displays the covariates and outcomes over the up to four visits observed for study participants in NAS and the up to two visits observed in KORA. Table 2 displays the variables that were held constant for each individual in the analysis. This included variables such as cell type proportion, age at first visit, and height at first non-missing observation. In NAS, the mean age at first visit was 63 while in KORA it was 54. The weight did not appear to vary much across the time points. The lung function measures saw a decline across the four visits (Table 1). The varying smoking status indicate varying levels of data being available. For 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). Similar interpretation for KORA, only going forward in time.

Briefly, we analyzed DNA methylation’s association with pulmonary function in NAS assuming that DNA methylation was metastable (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. Overall, we analyzed 477,927 probes. We analyzed three different lung function measures separately: forced expiratory volume in 1 second (FEV1), forced vital capacity (FVC), and forced expiratory flow between 25-75% of FVC (FEF25-75%). We tested for two different associations of DNA methylation on pulmonary function: its association with lung function decline (interaction between DNAm and follow up time from first spirometric measure, denoted as ***βLD***) and it’s cross sectional association (association between DNAm and lung function when follow up time is 0, denoted as ***βCS***).These two terms ***βLD*** and ***βCS*** are shown in the equation of the biostatistical analysis section of the methods. In each model we adjusted for 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 without dietary intake, maximum years of education, 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 included estimates of the cell type composition31.

***NAS Lung Function Decline Results (βLD)***

We first examined 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.In the NAS analysis 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 0.0532 (Table 3-Group 1). Not adjusting for smoking, ten unique probes were associated with rate of decline (Table 3-Group 1). The eight probes that were significant regardless of adjustment for smoking were cg18476993, cg05644990, cg12565126, cg13532885, cg21850722, cg26468478, cg16574073, and

cg05191655. There were three probes that were only significant when not adjusting for smoking (cg02721176, cg09995068, cg14292220) while two probes were only significant when adjusting for smoking (cg03867607 and cg18476993). The parameter estimates between the smoking adjusted and unadjusted models remained the same for the Holm significant 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 smoking33. A detailed list of significant probes is provided in Supplement Table 1 (including all individuals). We also repeated the analysis only adjusting for medications typically used for COPD or asthma therapy, i.e. corticosteroids, beta adrenergic agonists, and muscarinic receptor antagonist. Results are provided in Supplement Tables 3 and 4.

To ensure that the few current smokers (28) in the NAS were not influencing our results, we reran the analyses excluding current smokers (Table 3-Group 2). There were five probes (cg09995068, cg05644990, cg26468478, cg14292220, and cg01249054) associated with rate of decline were still found to be significant in this analysis. Of these five, 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 sub-group analysis it was significant irrespective of adjustment for smoking. One new probe was observed (cg01249054) and was associated with lung function decline before adjustment for smoking. Probe cg01249054 was FDR significant associated with FEV1 in the analysis including all smokers when did not adjust for smoking. There were 489 (477 unique probes) significant associations at the FDR level of 0.1 when adjusting for smoking, and 368 (361 unique probes) significant associations when no adjustment for smoking. A detailed list of significant probes and association is provided in Supplement Table 2.

***NAS Probes associated with Cross Sectional Association (βCS)***

In the analyses adjusting for smoking there were only two probes associated with lung function cross-sectional via the Holm procedure (Table 4-Group 1, ***βCS***). One, cg05575921, which maps to the aryl-hydrocarbon receptor repressor (*AHRR*) gene, was associated with, FEV1 and FEF25-75%. The other was cg06126421 which maps to the immediate early response 3 gene, *IER3,* and was associated withFEV1*.* This analysis had some under inflation (Table 4), yet p-values for the associations appeared to be following a uniform distribution, as expected (Figure 1 subplot A2, B2, C2, genomic inflation factor in caption of Table 4). When not adjusting for smoking, eight probes were found to be significant (Table 4-Group 1). In these probes there was a decrease in association between not-adjusting for smoking and adjusting for smoking (Table 4-Group 1). Overall, there were 5 (3 unique probes) significant associations when adjusting for smoking at the FDR level of 0.1 and 20 (9 unique probes) significant associations when not adjusting for smoking. Results from analyses only adjusting for medications-corticosteroids, beta adrenergic agonists, and muscarinic receptor antagonists are provided in the supplement (Supplement Tables 3 and 4).

In the analysis on former and never smokers, after adjustment for smoking the *AHRR* probe (cg05575921) was still found to be significant for the FEV1 and FEF25-75% models (Table 4-Group 2). The probe cg06126421 was also still found to be associated with FEV1 (Table 4-Group2). In the model not adjusting for smoking both cg05575921 and cg06126421 were significant (Table 4- Group 2) as were cg03636183 and cg21566642. There were 3 (2 unique probe) significant associations when adjusting for smoking at the FDR level of 0.1 and 14 (9 unique probes) significant associations when not adjusting for smoking. Results for when adjusted for medications typically used for COPD are provided in Supplementary Table 4.

Examining cg05575921, there appears to be a positive association between DNAm and the lung function measures (FEV1, FEF25-75) that is modified by smoking (Figures 2A and 2B). For current smokers, there is a positive linear association between the lung function measure and DNAm (FEV1 r2= 0.283, FEF25-75% r2= 0.302). As DNAm at cg05575921 increases, lung function increases as well. For former and non-smokers there does not appear to be a linear relationship (FEV1 r2= 0.081, FEF25-75% r2= 0.082). This effect modification by smoking was not observed for cg06126421 shown for FEV1 (Figure 2, subplot C). The r2between FEV1 and cg06126421 was 0.13 and 0.07 for current smokers and non-current smokers respectively. A similar pattern is observed on the M-value scale (Supplement Figure 1).

***KORA replication***

Overall, there were 1377 probes with a FDR significant association at one of the three lung function measures, either via the cross-sectional association (***βCS***) or the interaction with follow up time (***βLD***) after adjusting for smoking. KORA had 1052 of these 1377 probes. Only four probes replicated in KORA (FDR-significant in KORA *and* with the association in the same direction), as shown in Table 5 (More information provided in Supplement Table 5). Three of these were probes with cross-sectional association, cg05575921, cg06126421, and cg15342087. The other probe, cg01086847, was associated with the rate of lung function decline. Upon closer examination of the cg01086847 probe, it had a SNP directly under the probe. Figure showing cg05575921 and cg06126421 and their association with the respective lung function measures in NAS are shown in Figure 2. Supplement Figure 2 displays cg15342087 which has a similar pattern of association as cg06126421 in its association with FEV1 (r2 in current smokers 0.09 and 0.03 in non-current smokers in NAS). Probe cg06126421’s FVC association also reproduced but only in KORA Men.

***Sensitivity analysis***

We next removed probes with nearby single nucleotide polymorphisms (SNPs) and also individual outliers (defined by 1.5 times the interquartile range away from the 1st and 3rd quartile) for each probe, leading to a different set of probes found to be significant (Supplement Tables 6 and 7 for all FDR significant associations). This was done on the set adjusting for medications-corticosteroids, beta adrenergic agonists, and muscarinic receptor antagonists. The analysis was now on 384010 probes. There were in total 1527 (1344 unique probes) significant associations when adjusting for smoking and including current smokers. There was 1 significant FEF25-75% ***βCS*** association, 3 with FEV1 ***βCS***, 461 with FEV1 ***βLD***, 2 with FVC ***βCS***, and 1060 with a significant FVC ***βLD***. Of the ***βCS***, these associations were over 3 unique probes (cg05575921, cg06126421, and cg15342087). For ***βLD***, the associations were for 1341 unique probes, 704 of which were new. A list of these new probes is provided in Supplement Table 6 and 7.

Examining just the Holm significant sites (Supplement Tables 6), there were 10 significant probes with ***βLD***(Supplement Table 6). three were Holm significant in the original analysis (cg12565126, cg18476993, cg05644990). Five of the remaining seven were FDR significant at 0.1 in the original analysis (cg01417714, cg07273125, cg12396368, cg14503168, and cg19273694). The two new probes were cg00437258 (chromosome 7 in the 3'UTR of *DAGLB*) and cg14219256 (chromosome 10, body of *MYST4*). For ***βCS***, there were only two Holm significant probes associated with ***βCS***upon adjustment for smoking (cg05575921 and cg06126421). 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.

***Metastabiliy***

To confirm metastability, we calculated the intraclass correlation coefficient (ICC) on the subset of individuals with a second DNAm measure. Probe cg15342087 had relatively low ICC 0.67. Probes cg05575921 and cg06126421, had high ICCs indicating a level of metastability across time (Table 4). In addition, these two probes had been found to be metastable in the literature 34, though Shah et al did find low heritability for cg05575921. As confirmation, plotting percent methylation for each of these probes, at the next available time point, did indicate metastability (Figure 3). Figure 3 looks as expected given that cg05575921 and cg06126421 had high ICC estimates (0.91 and 0.93 respectively). For cg15342087 we see less metastability (ICC estimate of 0.67), but for the most parts the DNA methylation points seen similar between time points.Supplement Table 8 displays the ICC for each probe found significant in either the original analysis (ICC-1 sheet 1), in the sensitivity only adjusting for corticosteroids, beta adrenergic agonists, and muscarinic receptor antagonist (ICC-2 sheet 2), or in the stringent sensitivity analysis (ICC-3 sheet 3).

***Association with difference in DNAm age and chronological age***

Of the 1377 probes found significant in NAS, 663 had significant FDR-adjusted (adjusting for 1377 tests, at FDR 0.05 level33) associations with Δage in the NAS. Δage was defined as DNAm age minus chronological age. DNAm at the cg05575921 probe had a negative correlation with Δage (Figure 4, quadrant IV). Of the eleven probes found Holm significant in the NAS analysis adjusting for smoking, six had an association with Δage (Supplement Table 9). In KORA, of the 663 probes that came back as significant, 497 were in the KORA replication analysis. There were 36 probes that replicated in the KORA analysis (FDR-significant at 0.05 and the Pearson correlation sign was in the same direction, Supplement Table 10). The *AHRR* probe (cg05575921) did not have a significant association with Δage in KORA. In NAS, correlation between Δage  and age acceleration (residuals of regression of Δage and chronological age) were similar and displayed in Supplement Table 11 for these 1377 probes.

**Discussion**

In our study, there were seven probes that were associated with lung function decline in FVC (***βLD***) regardless of smoking adjustment or the study group examined (Table 3, cg05644990, cg12565126, cg13532885, cg21850722, cg16574073, cg05191655, and cg26468478). This suggests that these probes are independent of smoking exposure and may be also inhaled environmental noxious agents like air pollution. Some of these probes appeared metastable (cg12565126 and cg13532885 with ICC of 0.88 and 0.93 respectively), while others did not appear metastable at all (ICC for cg05644990 and cg26468478 of 7.9e-14 and 0.031 respectively). Biological aging may be a driving mechanism of the association for these two probes (cg05644990 and cg26468478). Indeed, in Supplement Table 9, we see that cg05644990 is associated with Δage, though cg26468478 is not. The two metastable probes, cg12565126 and cg13532885, were both associated with Δage. Unfortunately, these findings for cg05644990, cg12565126, 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. Comparing the ICC for these three probes between never and former smokers (too low of numbers in current smokers) it was 4.34E-14 vs 7.02e-17 for cg05644990, 0.85 vs 0.90 for cg12565126, and 0.9 and 0.94 for cg13532885.

Three probes associated with changes in FEV1 were significant only in the unadjusted for smoking analysis in the larger group, cg09995068 (ICC of 0.99, never vs former: 0.75 vs 0.99), cg02721176 (ICC of 0.81 never vs former: 0.83 vs 0.80), and cg14292220 (ICC of 0.9 never vs former: 0.41 vs 0.96) are likely being driven by cigarette smoking, as they were not significant in the analysis adjusting for smoking status. These probes only being associated with FEV1 suggest an association with airway function and less with the alveolar region of the lung (FVC was not associated). This is in line with the pathophysiologic understanding of cigarette smoke related COPD. 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 remaining probes that were detected only in the overall group, were likely being driven by the small number of smokers in our study, as they did not show up in the analyses restricting to former and never smokers.

For the cross-sectional analysis (***βCS***), results were mainly observed for FEV1 and FEF25-75%, with little association in FVC when adjusting for smoking suggesting that these probes are mostly associated with airway function. The probes that were significant for FVC (only when not adjusting for smoking) were the *AHRR* probe cg05575921 (ICC of 0.91, never vs former: 0.83 vs 0.93), the *IER3* probe cg06126421 (ICC of 0.93, never vs former: 0.89 vs 0.93), and the probe cg15342087 (ICC of 0.67, never vs former: 0.57 vs 0.66). Upon adjustment for smoking, this association shrank. *AHRR* is significant for FEV1 and FEF25-75% in both study groups before and after adjusting for smoking ­ and also is associated with Δage (Supplement Table 9). It should be noted that there has been evidence of SNPs near our *AHRR* probe, though these SNPs did not appear to confound the analysis in a reported study of maternal smoking35. Further, cg06126421is associated with FEV1 and FVC upon adjustment for smoking. The DNAm at probe cg06126421 association with FVC did overlap with KORA Men. Probe cg15342087, while not HOLM significant, did pass the FDR threshold of 0.1 with FEV1 and did replicate in KORA. This probe is close to cg06126421 and was associated with Δage (correlated -0.148, p-value 1.8E-4). Based on the shrinkage of other associations reported in Table 4, the other probes are likely being driven by smoking. For the three probes (cg06126421, cg15342087, and cg05575921**)** the effect size is somewhat lower, about 25 %, when adjusting for subjects with chronic lung diseases (ever asthma and/or COPD defined by FEV1/FVC<0.7) but still highly significant (Supplement Table 12). This suggests that the observed associations are not primarily driven by chronic lung diseases but appear to be amplified by these diseases.

Related to the observation that poor lung function, either due to a low plateau or a rapid age-related decline, is a predictor for lung morbidity and mortality, we examined whether there was an overlap between cross-sectional probes and longitudinal decline probes. However, comparing the results from the ***βCS*** and the ***βLD***(Tables 3 & 4), we see no overlap between the probes, suggesting different pulmonary insight. This likely suggests that the ***βCS***significant probes are mainly determined or associated with the growth phase of an individual, i.e. the leveling off at the plateau, and only slightly modulated by the age-related decline. We can see this in the differences in the associations, with the probes with a significant ***βLD*** association being solely with measures of volume related spirometric indices. Meanwhile, the probes found with a significant ***βCS*** association, were mainly associated with lung function parameters that are reflective of airway function.

The strength of our experimental design rests upon the fact that 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+ 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 timespan) (Table 1). This design feature helped in facilitating: (i.) the discovery of metastable DNAm sites retrospectively for evaluating lung function decline within a vulnerable elderly population (NAS cohort); and (ii.) replication and confirmation of these methylation sites that may *precede* lung function decline in younger humans, too (KORA). Throughout our work, metastability was confirmed by calculating the ICC of DNAm from blood samples taken 3-5 years apart (Table 3 and 4); this metastability assessment of DNAm metastability has only been evaluated in a small amount of literature within the human genome 34. While this competing study design is a strength, it may be the reason we failed to replicate any lung function decline associations within KORA. The associations in NAS may be indicative of just previous smoking exposure, while in KORA there actually be interactive associations.

In summary, in NAS we identified 18 unique DNAm sites using all smoking classes (never, current, and former smokers) prior to adjusting for smoking and 11 unique DNAm sites after adjusting for smoking-related covariates (Tables 3 and 4). When we re-ran our analyses *excluding* the current smokers yet maintaining never and former smokers; this approach yielded 9 unique DNAm sites before adjusting for smoking-related covariates (smoking status and pack-years, given the presence of former smokers) and 4 DNAm sites after adjustment (Tables 3 and 4). In KORA, 3 specific CpG loci reproduced (cg05575921, cg06126421, and cg15342087), so we next evaluated the scaled-effect estimate for one-unit change in the standard deviation of these data for cg05575921 and cg0612642. We did not evaluate for cg15342087 as it was within 200 BP of cg06126421. The one-unit change in the standard deviation effect of cg05575921 with FEF25-75% in NAS was 0.104, KORA all 0.135, KORA men 0.202, KORA older men, 0.399, and KORA women 0.057. For cg05575921 with FEV1 it was 0.141, 0.056, 0.1, 0.21, and 0.05 for NAS, KORA All, KORA men, KORA older mean, and KORA women respectively. For cg06126421, these were 0.119 for NAS, 0.051 for KORA all, 0.127 for KORA men, 0.146 for KORA older men, and 0.034 for KORA women. Finally, for cg06126421 association with FVC (which only replicated in KORA men) we had 0.12 in NAS, 0.01 in KORA all, 0.07 in KORA Men, 0.03 in KORA Older Men, 0.03 in KORA women. While the scaled estimates are slightly different, all effects are in the same direction. More importantly, the scaled-effect estimates for NAS were similar to that in the KORA men. Regardless, the unscaled effect estimates were all in the same direction (Table 5).

In evaluating the correlation of these DNAm marks with biological aging30, we found that 6 of the 11 NAS loci (Supplement Table 9) were significantly associated with Δage. None of these 6 CpG loci (whose DNAm was associated with PFTs in the NAS) overlapped with any of the CpGs queried by the algorithm to compute DNAm age 30. Collectively, these data provide additional evidence suggesting the utility of DNAm age as not only a predictor of chronological age in humans, but also as an important biomarker of aging-related characteristics, such as altered lung function.

Collectively, all of these key findings led us to a novel working model, which can serve to integrate and contextualize our results. We propose that DNAm status at the *AHRR* probe is highly sensitive to smoking status (Figure 2, subplots A and B and in the literature 22, 23), highly associated with PFTs (FEV1 and FEF25-75%), and, in some settings, with aging (Δage). This seems plausible since AHRR has been described to be involved 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 an association with 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. This probe has been reported numerous times in the literature as associated with smoking 22, 23. Moreover, although DNAm at cg06126421 (*IER3*) is also correlated with lung function, it does not appear to have an effect modification associated with smoking or aging in our data (Figure 2, subplot C). This probe has however been found to be associated with smoking and with mortality in other 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 that in addition replicated, near cg06126421 and has been found as well to be associated with all-cause mortality41.

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.  IER3 has been described to be involved in the immune response and inflammation, however, it is induced by a variety of external and internal factors42 and a possibly more specific role for the respiratory system remains to be elucidated. It should be noted that there has been evidence of SNPs near our *AHRR* probe, though these SNPs did not appear to confound the analysis in a reported study of maternal smoking35.

**Materials and methods**

***Human Participants and Lung Function***

***Normative Aging Study Cohort (NAS)***

The present work 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 questionnaire 44. We used FVC, FEV1, and FEF25-75% to assess pulmonary function. These spirometric parameters were measured following the American Thoracic Society’s (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 (IRBs) 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 al23, 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 (ERS) 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 477927 probes, *i.e*., ~98.4% out of 485512, 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 wateRmelon53. 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 three different lung function outcomes, FEF25-75%, FEV1, 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 take into account the correlation between lung function measures from the same individual. FEF25-75% 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 R57. 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 ( for spirometric measure at time k for individual i (. Here Xi is the set of confounders. The follow up time (time between first included visit and current visit) is , with if k=1. A study participant could have k up to 4.

In our longitudinal model, we included an interaction between the beta-value and follow-up time between visit and the visit at time point 1. There are two associations of interest: (i.) the *cross-sectional association* of DNAm on lung function, ***βCS*** (when follow-up time is 0, *CS* for cross sectional), and (ii.) the *interaction* between DNAm and follow-up time, ***βLD*** (*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 without dietary intake, 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 (***βCS***) and interaction (***βLD***) terms defined above.

A probe was considered genome-wide-significant if it had a Holm p-value of 0.05 or less 32 or significant at an FDR level of 0.10 33. Holm significance is a more conservative approach than FDR and controls the family wise error rate. Replication was performed in the KORA cohort. Replication in KORA proceeded as follows: (A) use all participants/ages (controlling for sex) *vs*. all of the men in KORA; then (B), analysis of only the oldest men in KORA (to simulate the NAS) *vs*. all of the 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. Ideally a probe with a small p-value will have a high ICC.

We in addition re-ran the analysis removing any 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 site58. This brought the number of probes down to 384010. For each probe, we reran the analysis removing any observations where the DNAm beta-value was greater than 3 interquartile ranges outside the 25% and 75% quartiles for that probe. This was to remove any potential outliers that could influence our study.

***Calculation of DNAm age, Δage, and Relationship with Lung Function Associated CpGs***

DNA methylation age (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 is a publically available tool (http://labs.genetics.ucla.edu/horvath/dnamage/), and has been shown to maintain predictive accuracy across many human tissues, including blood 30. We calculated an age-difference variable, “Δage,” by subtracting a participant’s DNAm age (predicted age given their epigenetic profile) from his/her chronological age (their actual age). A positive Δage is illustrative of enhanced/increased aging, and *vice versa*. We tested if any of the significant probes (***βCS*** or ***βLD***) from the lung function models, were also significantly associated with Δage. 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.

**Disclosure of relevant conflict of interests**

The authors have no conflict of interests to disclose.

**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.

**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 had no involvement 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 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.

**References**

1. Soler Artigas M, Wain LV, Miller S, Kheirallah AK, Huffman JE, Ntalla I, Shrine N, Obeidat M, Trochet H, McArdle WL, et al. Sixteen new lung function signals identified through 1000 Genomes Project reference panel imputation. Nat Commun 2015; 6:8658.

2. Weiss ST. Lung function and airway diseases. Nat Genet 2010; 42:14-6.

3. Wain LV, Shrine N, Artigas MS, Erzurumluoglu AM, Noyvert B, Bossini-Castillo L, Obeidat M, Henry AP, Portelli MA, Hall RJ, et al. Genome-wide association analyses for lung function and chronic obstructive pulmonary disease identify new loci and potential druggable targets. Nat Genet 2017; 49:416-25.

4. Maher B. Personal genomes: The case of the missing heritability. Nature 2008; 456:18-21.

5. Morrow JD, Cho MH, Hersh CP, Pinto-Plata V, Celli B, Marchetti N, Criner G, Bueno R, Washko G, Glass K, et al. DNA methylation profiling in human lung tissue identifies genes associated with COPD. Epigenetics 2016:1-10.

6. Obeidat M, Hao K, Bosse Y, Nickle DC, Nie Y, Postma DS, Laviolette M, Sandford AJ, Daley DD, Hogg JC, et al. Molecular mechanisms underlying variations in lung function: a systems genetics analysis. Lancet Respir Med 2015; 3:782-95.

7. Piletic K, Kunej T. MicroRNA epigenetic signatures in human disease. Arch Toxicol 2016; 90:2405-19.

8. Joehanes R, Just AC, Marioni RE, Pilling LC, Reynolds LM, Mandaviya PR, Guan W, Xu T, Elks CE, Aslibekyan S, et al. Epigenetic Signatures of Cigarette Smoking. Circ Cardiovasc Genet 2016.

9. LaCroix AZ, Omenn GS. Older adults and smoking. Clin Geriatr Med 1992; 8:69-87.

10. Kerstjens HA, Rijcken B, Schouten JP, Postma DS. Decline of FEV1 by age and smoking status: facts, figures, and fallacies. Thorax 1997; 52:820-7.

11. Speizer FE, Tager IB. Epidemiology of chronic mucus hypersecretion and obstructive airways disease. Epidemiol Rev 1979; 1:124-42.

12. Godtfredsen NS, Lam TH, Hansel TT, Leon ME, Gray N, Dresler C, Burns DM, Prescott E, Vestbo J. COPD-related morbidity and mortality after smoking cessation: status of the evidence. Eur Respir J 2008; 32:844-53.

13. Vercelli D. Does epigenetics play a role in human asthma? Allergol Int 2016; 65:123-6.

14. Gruzieva O, Merid SK, Melen E. An update on epigenetics and childhood respiratory diseases. Paediatr Respir Rev 2014; 15:348-54.

15. Comer BS, Ba M, Singer CA, Gerthoffer WT. Epigenetic targets for novel therapies of lung diseases. Pharmacol Ther 2015; 147:91-110.

16. Pal S, Tyler JK. Epigenetics and aging. Sci Adv 2016; 2:e1600584.

17. 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; 127:1952-9.

18. Hole DJ, Watt GC, Davey-Smith G, Hart CL, Gillis CR, Hawthorne VM. Impaired lung function and mortality risk in men and women: findings from the Renfrew and Paisley prospective population study. BMJ 1996; 313:711-5; discussion 5-6.

19. Lee HM, Le H, Lee BT, Lopez VA, Wong ND. Forced vital capacity paired with Framingham Risk Score for prediction of all-cause mortality. Eur Respir J 2010; 36:1002-6.

20. Higgins M. Risk factors associated with chronic obstructive lung disease. Ann N Y Acad Sci 1991; 624:7-17.

21. Hoyert DL, Xu J. Deaths: preliminary data for 2011. Natl Vital Stat Rep 2012; 61:1-51.

22. Fasanelli F, Baglietto L, Ponzi E, Guida F, Campanella G, Johansson M, Grankvist K, Johansson M, Assumma MB, Naccarati A, et al. Hypomethylation of smoking-related genes is associated with future lung cancer in four prospective cohorts. Nat Commun 2015; 6:10192.

23. Zeilinger S, Kuhnel B, Klopp N, Baurecht H, Kleinschmidt A, Gieger C, Weidinger S, Lattka E, Adamski J, Peters A, et al. Tobacco smoking leads to extensive genome-wide changes in DNA methylation. PLoS One 2013; 8:e63812.

24. Bauer T, Trump S, Ishaque N, Thurmann L, Gu L, Bauer M, Bieg M, Gu Z, Weichenhan D, Mallm JP, et al. Environment-induced epigenetic reprogramming in genomic regulatory elements in smoking mothers and their children. Mol Syst Biol 2016; 12:861.

25. Gibbs K, Collaco JM, McGrath-Morrow SA. Impact of Tobacco Smoke and Nicotine Exposure on Lung Development. Chest 2016; 149:552-61.

26. Drummond MB, Hansel NN, Connett JE, Scanlon PD, Tashkin DP, Wise RA. Spirometric predictors of lung function decline and mortality in early chronic obstructive pulmonary disease. Am J Respir Crit Care Med 2012; 185:1301-6.

27. Donaldson GC, Seemungal TA, Bhowmik A, Wedzicha JA. Relationship between exacerbation frequency and lung function decline in chronic obstructive pulmonary disease. Thorax 2002; 57:847-52.

28. Mannino DM, Buist AS, Petty TL, Enright PL, Redd SC. Lung function and mortality in the United States: data from the First National Health and Nutrition Examination Survey follow up study. Thorax 2003; 58:388-93.

29. 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; 166:675-9.

30. Horvath S. DNA methylation age of human tissues and cell types. Genome Biol 2013; 14:R115.

31. Houseman EA, Accomando WP, Koestler DC, Christensen BC, Marsit CJ, Nelson HH, Wiencke JK, Kelsey KT. DNA methylation arrays as surrogate measures of cell mixture distribution. BMC Bioinformatics 2012; 13:86.

32. Holm S. A Simple Sequentially Rejective Multiple Test Procedure. Scandinavian Journal of Statistics 1979; 6:65-70.

33. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society Series B (Methodological) 1995; 57:289-300.

34. Shah S, McRae AF, Marioni RE, Harris SE, Gibson J, Henders AK, Redmond P, Cox SR, Pattie A, Corley J, et al. Genetic and environmental exposures constrain epigenetic drift over the human life course. Genome Res 2014; 24:1725-33.

35. Gonseth S, de Smith AJ, Roy R, Zhou M, Lee ST, Shao X, Ohja J, Wrensch MR, Walsh KM, Metayer C, et al. Genetic contribution to variation in DNA methylation at maternal smoking-sensitive loci in exposed neonates. Epigenetics 2016; 11:664-73.

36. Traboulsi H, Guerrina N, Iu M, Maysinger D, Ariya P, Baglole CJ. Inhaled Pollutants: The Molecular Scene behind Respiratory and Systemic Diseases Associated with Ultrafine Particulate Matter. Int J Mol Sci 2017; 18.

37. Awji EG, Chand H, Bruse S, Smith KR, Colby JK, Mebratu Y, Levy BD, Tesfaigzi Y. 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; 52:377-86.

38. Elliott HR, Tillin T, McArdle WL, Ho K, Duggirala A, Frayling TM, Davey Smith G, Hughes AD, Chaturvedi N, Relton CL. 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, Sacerdote C, Ricceri F, Birrell MA, Belvisi MG, Brown R, Vineis P, Flanagan JM. 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; 22:843-51.

40. Zhang Y, Schottker B, Florath I, Stock C, Butterbach K, Holleczek B, Mons U, Brenner H. 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, Breitling LP, Saum KU, Schottker B, Holleczek B, Waldenberger M, Peters A, Brenner H. 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; 90:545-52.

43. Bell B, Rose CL, Damon A. The Veterans Administration longitudinal study of healthy aging. Gerontologist 1966; 6:179-84.

44. Ferris BG. Epidemiology Standardization Project (American Thoracic Society). Am Rev Respir Dis 1978; 118:1-120.

45. Miller MR, Hankinson J, Brusasco V, Burgos F, Casaburi R, Coates A, Crapo R, Enright P, van der Grinten CP, Gustafsson P, et al. Standardisation of spirometry. Eur Respir J 2005; 26:319-38.

46. Sparrow D, O'Connor G, Colton T, Barry CL, Weiss ST. 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; 135:1255-60.

47. Chatham M, Bleecker ER, Norman P, Smith PL, Mason P. A screening test for airways reactivity. An abbreviated methacholine inhalation challenge. Chest 1982; 82:15-8.

48. Holle R, Happich M, Lowel H, Wichmann HE, Group MKS. KORA--a research platform for population based health research. Gesundheitswesen 2005; 67 Suppl 1:S19-25.

49. Karrasch S, Flexeder C, Behr J, Holle R, Huber RM, Jorres RA, Nowak D, Peters A, Wichmann HE, Heinrich J, et al. Spirometric reference values for advanced age from a South german population. Respiration 2013; 85:210-9.

50. Tang W, Kowgier M, Loth DW, Soler Artigas M, Joubert BR, Hodge E, Gharib SA, Smith AV, Ruczinski I, Gudnason V, 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, Baccarelli AA, Just AC, Wolf K, Wahl S, Cyrys J, Kunze S, Strauch K, 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.

52. Miller MR, Crapo R, Hankinson J, Brusasco V, Burgos F, Casaburi R, Coates A, Enright P, van der Grinten CP, Gustafsson P, et al. General considerations for lung function testing. Eur Respir J 2005; 26:153-61.

53. Pidsley R, CC YW, Volta M, Lunnon K, Mill J, Schalkwyk LC. 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, Laird PW, Siegmund KD. Low-level processing of Illumina Infinium DNA Methylation BeadArrays. Nucleic Acids Res 2013; 41:e90.

55. Davis S, Du P, Bilke S, Triche T, Jr., Bootwalla M. methylumi: Handle Illumina methylation data. 2015.

56. Teschendorff AE, Marabita F, Lechner M, Bartlett T, Tegner J, Gomez-Cabrero D, Beck S. A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data. Bioinformatics 2013; 29:189-96.

57. Bates D, McHler M, Bolker BM, Walker SC. Fitting linear mixed-effects models using lme4. Journal of Statistical Software 2015; 67:1--48.

58. Chen YA, Lemire M, Choufani S, Butcher DT, Grafodatskaya D, Zanke BW, Gallinger S, Hudson TJ, Weksberg R. Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray. Epigenetics 2013; 8:203-9.

**Figure Legends**

**Figure 1:** QQ plots of P-values for each spirometric model: those associated with *LD* are located to the left (A, C, and E) and those with *CS* on the right (B, D, and F).

**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 FEF25-75%; and (C) cg06126421 and FEV1.

**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.

**Figure 4:** Scatterplot of DNAm beta estimates and Pearson correlation coefficients for CpGs significantly associated with lung function and significantly correlated with ∆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 ∆age in the NAS.

**Table Legends**

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

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

**Table 3:** Results for ***βLD*** 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).

**Table 4**: Cross sectional Associations of DNAm on lung function (*ßCS*): Results for *βCS* 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).

Table 5: Replicated findings across the NAS and KORA cohorts.