Epigenome-wide association study of lung function level and its change

Medea Imboden[1,2]*, Matthias Wielscher[3,4]*, Faisal I. Rezwan[5]*, André F.S. Amaral[3,6]*, Emmanuel Schaffner[1,2], Ayoung Jeong[1,2], Anna Beckmeyer-Borowko[1,2], Sarah E. Harris[7,8], John M Starr[7,9], Ian J. Deary[7,10], Claudia Flexeder[11], Melanie Waldenberger[11,12], Annette Peters[11,12], Holger Schulz[11,13], Su Chen[14], Shadia Khan Sunny[15], Wilfried J.J. Karmaus[15], Yu Jiang[15], Gertraud Erhart[16], Florian Kronenberg[16], Ryan Arathimos[17,18], Gemma C. Sharp[17,18,19], Alexander John Henderson[18], Yu Fu[20], Päivi Piirilä[21], Kirsi H. Pietiläinen[22,23], Miina Ollikainen[20], Asa Johansson[24], Ulf Gyllensten[24], Maaike de Vries[25,26], Diana A. van der Plaat[25,26], Kim de Jong[25,26], H. Marike Boezen[25,26], Ian P. Hall[27], Martin D. Tobin[28,29], Marjo-Riitta Jarvelin[4,30,31,32,33]*, John W. Holloway[5]*, Deborah Jarvis[3,6]*, and Nicole M. Probst-Hensch[1,2]*.

* These authors have contributed equally.

Affiliations:
[1] Swiss Tropical and Public Health Institute, Basel, Switzerland
[2] University of Basel, Switzerland
[3] MRC-PHE Centre for Environment and Health, Imperial College London, London, UK
[4] Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, London, UK
[5] Human Development and Health, Faculty of Medicine, University of Southampton, Southampton, UK
[6] Population Health and Occupational Disease, NHLI, Imperial College London, London, UK
[7] Centre for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh, 7 George Square, Edinburgh EH8 9JZ, UK
[8] Medical Genetics Section, University of Edinburgh Centre for Genomic and Experimental Medicine and MRC Institute of Genetics and Molecular Medicine, Western General Hospital, Crewe Road, Edinburgh EH4 2XU, UK
[9] Alzheimer Scotland Dementia Research Centre, University of Edinburgh, 7 George Square, Edinburgh EH8 9JZ, UK
[10] Department of Psychology, University of Edinburgh, 7 George Square, Edinburgh EH8 9JZ, UK
[11] Institute of Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany
[12] Research Unit Molecular Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany
[13] Comprehensive Pneumology Center Munich (CPC-M), Member of the German Center for Lung Research (DZL)
[14] Department of Mathematical Sciences, University of Memphis, 373 Dunn Hall, Memphis, TN, 38152, USA
[15] Division of Epidemiology, Biostatistics, and Environmental Health, School of Public Health, University of Memphis, 236A Robison Hall, Memphis, TN, 38152, USA
[16] Division of Genetic Epidemiology, Department of Medical Genetics, Molecular and Clinical Pharmacology, Medical University of Innsbruck, Innsbruck, Austria
[17] MRC Integrative Epidemiology Unit, University of Bristol, Bristol, UK
[18] Department of Population Health Sciences, Bristol Medical School, University of Bristol, UK
[19] Bristol Dental School, University of Bristol, Bristol, UK

[20] Institute for Molecular Medicine Finland, FIMM, University of Helsinki, Helsinki, Finland
[21] Unit of Clinical Physiology, HUS Medical Imaging Center, Helsinki University Central Hospital and University of Helsinki, Helsinki, Finland
[22] Obesity Research Unit, Reserch Programs Unit, University of Helsinki, Helsinki, Finland
[23] Abdominal Center, Endocrinology, Helsinki University Hospital and University of Helsinki, Helsinki, Finland
[24] Department of Immunology, Genetics and Pathology, Science for Life Laboratory, Uppsala University, Sweden
[25] University of Groningen, University Medical Center Groningen, Department of Epidemiology, Groningen, Netherlands
[26] Groningen Research Institute for Asthma and COPD (GRIAC), University of Groningen, University Medical Center Groningen, Groningen, Netherlands
[27] Division of Respiratory Medicine, University of Nottingham, Nottingham, UK; National Institute for Health Research, Nottingham Biomedical Research Centre, Nottingham University Hospitals, Nottingham, UK
[28] Department of Health Sciences, University of Leicester, Leicester, UK
[29] National Institute of Health Research Biomedical Research Centre, University of Leicester, Leicester, UK
[30] Center for Life Course Health Research, Faculty of Medicine, University of Oulu, Finland
[31] Biocenter Oulu, University of Oulu, Oulu, Finland
[32] Unit of Primary Health Care, Oulu University Hospital, Finland
[33] Department of Life Sciences, College of Health and Life Sciences, Brunel University London, UK

(Word Count: 3989)

Corresponding authors:
Dr. Medea Imboden, Chronic Disease Epidemiology Unit, Department of Epidemiology and Public Health, Swiss Tropical and Public Health Institute, Socinstrasse 57, 4002 Basel, Switzerland.
e-mail: medea.imboden@swisstph.ch, Tel: +41 61 284 82 40

Prof. Dr. Nicole Probst-Hensch, Chronic Disease Epidemiology Unit, Department of Epidemiology and Public Health, Swiss Tropical and Public Health Institute, Socinstrasse 57, 4002 Basel, Switzerland.
e-mail: nicole.probst@swisstph.ch, Tel: +41 61 284 83 78

Take home message

An agnostic association study on lung function using longitudinal population-based cohort data shows that differentially methylated genomic sites related to smoking are strongly associated with lung function in adults.

Abstract (248 words)

Previous reports link differential DNA methylation (DNAme) to environmental exposures which are associated with lung function. Direct evidence on lung function DNAme is however limited. We undertook an agnostic epigenome-wide association study (EWAS) on pre-bronchodilation lung function and its change in adults.

In a discovery-replication EWAS design, DNAme in blood and spirometry were measured twice, six-to-15 years apart, in the same participants of three adult population-based discovery cohorts (n=2,043). Associated DNAme markers ($P<5x10^{-7}$) were tested in seven replication cohorts (adult: n=3,327; childhood: n=420). Technical-bias adjusted residuals of a regression of the normalized absolute beta-values on control-probe-derived principle components were regressed on level and change of FEV_1, FEV_1/FVC and FVC in covariate-adjusted discovery EWAS. Inverse-variance weighted meta-analyses were performed on results from discovery and replication samples in all participants and never smokers.

EWAS signals were enriched for smoking-related DNAme. We replicated 57 lung function DNAme in adult, but not childhood samples, all previously associated with smoking. Markers not previously associated with smoking failed replication. cg05575921 (*AHRR*) showed the statistically most significant association with cross-sectional lung function (FEV_1/FVC: $P_{discovery}=3.96x10^{-21}$ and $P_{combined}=7.22x10^{-50}$). A score combining ten DNAme markers previously reported to mediate the smoking effect on lung function was associated with lung function (FEV_1/FVC: $P=2.65x10^{-20}$).

Our results reveal that lung function associated methylation signals in adults are predominantly smoking-related and possibly of clinical utility in identifying poor lung function and accelerated decline. Larger studies with more repeat time points are needed to identify lung function DNAme in never smokers and in children.

Introduction

Lung function has an estimated heritability between 30 and 70% [1]. The variance in phenotype remains incompletely explained by genetic variation, but the impact of environmental exposure on respiratory health and lung function over the life course is well recognized. In particular, pro-inflammatory and oxidative inhalants such as cigarette and environmental tobacco smoke, air pollution, and occupational exposures are important contributors to the increased risk of respiratory symptoms, accelerated lung function decline in adults and poor lung growth in children. DNA methylation (DNAme) has been associated with a large variety of traits and chronic diseases.

A large body of evidence including results from epigenome-wide association analyses (EWAS) shows differentially methylated 5'-cytosine-phosphate-guanine-3' di-nucleotide (CpG) sites throughout the genome in response to environmental exposures in particular cigarette smoking [2-4]. In contrast, reports of DNAme associated with respiratory diseases and lung function show inconsistent findings [5, 6]. Most recently independent reports however pointed to the consistent association of DNAme in *AHRR* gene, cg05575921, with lung function in adults [4, 6, 7].

The current study aimed at agnostically identifying lung function-specific DNAme signals. We undertook a covariate-adjusted EWAS using questionnaire data, spirometry, and peripheral blood samples collected in the same participants (discovery cohorts: ECRHS, NFBC1966, SAPALDIA; see online supplement for cohort description) at two time points six-to-15 years apart. EWAS analyses were performed on lung function parameters of forced expiratory volume in one second (FEV_1), forced vital capacity (FVC) and their ratio (FEV_1/FVC). The analyses focused on cross-sectional associations at different time points and on identifying DNAme markers predicting change in lung function. We tested discovery-identified CpGs ($P<5x10^{-7}$ for at least one lung function parameter) for replication in adult samples from five adult cohorts (LBC1936, KORA, LifeLines, NSPHS, FTC) and in childhood samples from two birth cohorts (ALSPAC, IOWBC).

Methods

Study design and participants

The discovery sample (n=2,043) comprised three population-based cohort studies, part of the Aging Lungs in European Cohorts (ALEC) project. ECRHS (n=470, European Community Respiratory Health Study) and SAPALDIA (n=962, Swiss Study on Air Pollution Heart and Lung Disease in Adults) are adult cohorts designed to investigate respiratory health. NFBC1966 (n=611, Northern Finland Birth Cohort 1966) is a birth cohort with follow-up to adult age. The replication sample consisted of five adult cohorts, KORA (n= 628, Cooperative Health Research in the Augsburg Region Study), LifeLines (n=1,622, LifeLines cohort study), NSPHS (n=535, North Sweden Population Health Study), LBC1936 (n=449, Lothian Birth Cohort 1936, adult inception birth cohort), FTC (n=93, Finnish Twin Cohort study), and two childhood birth cohorts ALSPAC (n=258, Avon longitudinal Study of Parents and Children) and IOWBC (n=162, Isle of Wight Birth Cohort). Replication data from two time points was available only for KORA, LBC1936 (adult), and ALSPAC, IOWBC (childhood) (for cohort details and contribution to analysis see online supplement text and figure S1). All cohorts comply with the Declaration of Helsinki and ethical approval was obtained from the respective national and regional ethical review committees.

Procedures

In the discovery cohorts, DNAme measurement using Illumina® Infinium technology was obtained from peripheral blood samples collected at two consecutive follow-up surveys several years apart. The 450K BeadChip was used for samples of 984 SAPALDIA participants from both time points and of 732 NFBC1966 participants collected at time point 1. The EPIC BeadChip was used for samples of 509 ECRHS participants from both time points and of 716 NFBC1966 participants collected at time point 2. For cohort-specific EWAS analyses, we used all autosomal markers available for each time point and cohort-specific EWAS marker results were meta-analysed without restriction to markers

common to both arrays. DNAme data used for replication was restricted to discovery-identified (sentinel) CpGs and analyzed on various arrays.

Epidemiological data, including covariate information at subject level, was collected by interview-assisted questionnaires and objective measures. Pre-bronchodilation spirometric data was obtained by performing ATS/ERS-compliant spirometry (see online supplement).

Statistical analyses

Epigenome-wide methylation data was analysed in R (version3.4.3). Differential blood cell count was estimated using a reference dataset and the minfi R package [8, 9]. DNAme used as predictors in the statistical models for the adult cohorts were obtained by deriving residuals from linear regression of the normalized absolute DNA methylation (β-values) on the Illumina control probe derived 30 first principal components to correct for correlation structures within the data, including technical bias. Thus here reported effect sizes of the association are not comparable to elsewhere reported effect sizes using normalized β-values as predictor. In the childhood data, batch effect was corrected at the analysis level by regressing the DNAme values against the technical covariates.

Epigenome-wide covariate-adjusted linear regression was performed to assess the association of single CpG markers with forced expiratory volume in one second (FEV_1, (L)), forced vital capacity (FVC, (L)), their ratio (FEV_1/FVC) and their change during follow-up. This multi-level EWAS design tested different models in all and never-smoking participants (figure 1). First, cross-sectional EWAS were examined separately at time point 1 (EWAS1) and time point 2 (EWAS2) assessing the consistency of the association over follow-up time. Second, the association of DNAme at the first time point (DNAme1) with change in lung function during follow-up was assessed (prediction EWAS ($EWAS_{predict}$)). Covariate-adjusted mixed linear regressions with a random intercept on the subject were undertaken using data from both time points (repeat cross-sectional analysis ($EWAS_{repeat}$)).

All associations were adjusted for a set of *a priori* selected covariates known to influence respiratory outcomes from previous research conducted by SAPALDIA and ECRHS. The covariate-adjusted model M_{base} included age, age^2, height, squared deviation of height from the mean, sex and interaction terms of sex with four covariates (age, age^2, height and squared deviation of height), education, body mass index, spirometer type, study center as well as estimated cell composition (CD8-, CD4-, natural killer cells, B-cells, monocytes, eosinophils and neutrophils). Analyses in all participants were run without (M_{base}) and with additional smoking adjustment including smoking status and packyears (M_{smok}). In never smokers M_{base}-covariate adjustment was applied. Prediction associations of DNAme1 were additionally adjusted for lung function at time point 1. The same covariate adjustment was applied in adult replication analyses, whereas childhood covariates did not include squared terms.

Cohort-specific EWAS results were summarized by inverse-variance weighted meta-analyses using METAL [10]. Meta-analysis results were not controlled for genomic inflation after confirming its negligible influence. Epigenome-wide significance level was set to P-value<$1x10^{-7}$ (Bonferroni correction, 450,000 tests). The selection criteria for replication of sentinel CpGs was less stringent (P-value<$5x10^{-7}$). Successful replication was defined as P-value below outcome-specific Bonferroni-correction threshold.

Replicated CpGs were characterized by enrichment, pathway, and functional analyses and additional *post hoc* analyses were performed (see details in online supplement). First, a two-sample Mendelian randomization analysis based on publicly available data was applied to investigate the causality of replicated CpGs associations. Second, a replication of a recently published mediation analysis [4] evidencing ten smoking-related CpGs mediating the smoking effect on lung function was undertaken in one discovery cohort (SAPALDIA). Third, to assess the combined effects of smoking-related CpGs on lung function in three discovery cohorts, we built two different DNAme smoking indices based based on CpGs a) predicting lung function effects of smoking [4] and b) located in GWAS-identified

lung function genes [2]. These smoking indices were tested for association with lung function in covariate-adjusted linear regression analyses, in all participants and in subgroups stratified by smoking status.

Role of funding sources

This EWAS was funded by European Union's H2020 research programme. The funding agency had no role in the design, data collection and analysis of the data. Cohort-specific funding details provided in the online supplement.

Data availability statement

Statistical codes, and full discovery and replication EWAS effect estimates (meta-analysed and cohort-specific) are made publically available with no end date on the public repository DRYAD (http://datadryad.org/) at the time of publication. Access restrictions apply to the individual methylome data underlying the analysis. Contact details for data requests to the contributing cohorts can be found in the supplement material.

Results

Differences in the cohorts' age structure and smoking habits are shown in table 1. Mean age was highest for LBC1936 (69.9 years) and youngest (30.4 years) for FTC. Self-report of current smoking status was lowest in LBC1936 (5.8%) and highest in LifeLines (43.5% due to over-sampling of current smokers for DNAme typed subset).

Across all discovery EWAS meta-analyses, we identified 111 CpG markers for replication ($P<5x10^{-7}$; 74 for FEV_1; 47 for FEV_1/FVC and 16 for FVC; online supplement tables S1 & S2). We present here the results for FEV_1/FVC (for FEV_1 and FVC see online supplement).

Cross-sectional associations without smoking adjustment

In study-specific and meta-analyzed discovery EWAS, the number of lung function associated DNAme increased from first to second cross-sectional time point in the same participants, despite age-adjustment (figure 2). We therefore meta-analyzed cross-sectional discovery and replication results from the older participants' age time point available. We observed 29 cross-sectional CpG associations with FEV_1/FVC. 27 of them replicated formally (Bonferroni correction, $P<0.0011$, 47 tests on FEV_1/FVC; table 2; online supplement table S3). All replicated CpG-lung function associations were exclusively DNAme previously associated with smoking [2]. Successful replication was observed for cg05575921 (*AHRR*) showing the strongest signal for FEV_1 and FEV_1/FVC (FEV_1/FVC: P-value combining discovery and replication cohorts ($P_{combined}$)=$7.22x10^{-50}$) among all identified lung function DNAme markers. Methylation at this CpG, previously shown to be hypo-methylated with increased smoking, showed positive cross-sectional lung function association. The top ten CpGs associated with FEV_1/FVC (table 2) were located in six loci: cg03636183 (*F2RL3*), cg21566642, cg01940273 and cg03329539 (vicinity of *ALPPL2*), cg05575921 and cg21161138 (*AHRR*), cg23771366 and cg11660018 (*PRSS23*), cg21611682 (*LRP5*), and cg15342087 (*IER3*). The same CpGs, along with cg19572487 (*RARA*) were also among the top 11 markers cross-sectionally associated with FEV_1. Formal replication of cross-sectional associations with FEV_1 was observed for 44 CpGs and with FVC for

three CpGs (online supplement tables S4 & S5). Similar results were found for repeat cross-sectional analyses ($EWAS_{repeat}$, online supplement table S6, figure S3).

Cross-sectional smoking-adjusted associations

The smoking-adjusted EWAS (M_{smok}) resulted in fewer genome-wide significant results (figure 2). Yet, despite adjustment for self-report of smoking history, the top five CpGs were known smoking-related CpGs. DNAme at cg05575921 (*AHRR)* remained the top cross-sectional association signal for FEV_1/FVC ($P_{combined}=2.21x10^{-11}$; table S7).

Predictive associations without smoking adjustment

The prediction EWAS results (table 3, figure 3) revealed that DNAme at time point 1 (DNAme1) at six of nine sentinel CpGs ($P<5x10^{-7}$) associated with change in FEV_1/FVC were replicated (cg05575921 and cg21161138 (*AHRR*), cg21566642, cg01940273 and cg03329539 (vicinity of *ALPPL2*), and cg03636183 (*F2RL3*)). These six replicated CpGs were smoking-related markers. They were also associated with cross-sectional FEV_1/FVC and four of them also with predicting change in FEV_1 (*AHRR (*cg05575921), *ALPPL2 (*cg05951221, cg01940273), and *F2RL3* (cg03636183), online supplement table S8).

Associations in never smokers

The agnostic discovery EWAS (M_{base}) in never smokers, similarly to the entire sample, showed more statistically significant associations at time point 2 (older age). Eight CpGs were cross-sectionally associated with FEV_1/FVC in never smokers ($P<5x10^{-7}$), but none replicated (table 4, online supplement figure S5). The CpG, cg14366110 (*FIBCD1*) showed predictive association of DNAme1 with change in FEV_1/FVC ($P_{discovery}=4.2x10^{-9}$, $P_{combined}=3.6x10^{-9}$) in never smokers, but it did not replicate in KORA and LBC1936 ($P_{replication}=0.439$, replication cohorts with lung function at two time points). The direction of effect however was consistent (table 5, online supplement table S9 for

cross-sectional associations; table S10 for prediction associations) in discovery and replication cohorts.

Characterization of replicated CpGs

None of the not-smoking-related discovery-identified sentinel CpGs (n=25) were confirmed by replication. In contrast, 78% of the sentinel CpGs (n=86) had previously been identified as smoking-related and 57 of these (mapping to 43 loci) formally replicated across all models and lung function outcomes tested (online supplement table S11). They were used jointly for functional annotation and pathway analyses (online supplement tables S12-to-S16). Briefly, these 57 lung function associated CpGs displayed enrichment for transcription factors, such as *RELA* (P_{FDR}=0.002) and *EP300* (P_{FDR}=0.004) and suggestive enrichment (P_{FDR}<0.1) for chromatin state model of flanking active transcription start sites, of transcription at gene 5' and 3', and of enhancers. Using IPA database, or GO term enrichment no significant pathways were revealed. Transcriptional misregulation in cancer, pathways in cancer and regulation of actin cytoskeleton were identified (P_{FDR}<0.05) using KEGG pathways enrichment.

Using the weighted Kolmogorov Smirnov test on the entire EWAS discovery results, we noted statistically significant enrichment for smoking-related CpGs among the lung function associated CpGs. This enrichment was also present in the smoking-adjusted EWAS and even in the EWAS restricted to never smokers (online supplement table S17).

Association of adult lung function CpG markers with childhood lung function

Using the same scheme of analysis as for the adult replication cohorts, none of the sentinel CpGs showed associations with FEV_1, FEV_1/FVC and FVC in the childhood replication cohorts (ALSPAC, IOWBC) (online supplement table S18). The strongest associations observed in children (P<0.01) were for five CpGs not known to be smoking-related DNAme markers and one smoking-related CpG, cg00310412 (*SEMA7A*).

Comparison with published DNAme – lung function association reports

Our agnostic results were compared with previously reported lung function- [4, 6, 7, 11] or COPD- [12, 13] specific DNAme. We retrieved all CpGs reported being associated with lung function (n=376) for a look-up in the cross-sectional FEV_1, FEV_1/FVC and FVC associations at time point 2. Only 12 out of 376 CpGs showed evidence for association (Bonferroni correction for 376 tests: $P<1.3x10^{-4}$, online supplement table S19). Notably, the most recently reported CpG markers [4, 6] – having also been related to smoking - showed consistent associations with lung function e.g. cg05575921, cg21161138 (*AHRR*), cg05951221 (near *ALPPL2*) and cg06126421 (*IER3*). They were among our top replicated lung function association signals.

Two-sample Mendelian Randomisation investigation

To assess the causality of replicated DNAme-lung function association, we conducted a *post-hoc* Mendelian randomization (MR) look-up using publicly available databases [14, 15]. Genetic instruments were identified for 12 replicated CpGs. For seven CpGs a two sample MR on cross-sectional lung function could be completed (see online supplement table S20). Results support causal effects for cg23771366, cg11660018 (*PRSS23*); cg21990700 (*C1RL*); cg00073460 (*ZC3H12D*) on FEV_1 and for cg00073460 (*ZC3H12D*); cg24086068 (*SHROOM3*) on FVC.

Integration of DNAme into a smoking index

A recent smoking EWAS followed-up by a mediation analysis identified ten CpGs as mediators of the smoking-lung function association [4]. Eight of these mediating CpGs were among our replicated lung-function associated CpGs (see online table S21). In a *post hoc* mediation analysis in SAPALDIA, we showed statistically significant average causal mediation on lung function for nine of these mediating CpGs (FEV_1/FVC: table 6; FEV_1 and FVC: online supplement table S22).

To assess the combined effect of these smoking-exposure-mediating CpGs on lung function, we constructed a mediation-smoking index (Mediation-SI). Its association with lung function by smoking

status was tested in covariate-adjusted regression models in the discovery cohorts and following EWAS models (SAPALDIA, ECRHS, NFBC1966). Meta-analyzed results of the Mediation-SI showed strong association with cross-sectional FEV_1/FVC in all participants and ever smokers (table 7, figure 4; FEV_1 and FVC, online supplement table S23). The Mediation-SI association in all participants was more pronounced for cross-sectional (coefficient (β) (standard error (SE))=-1.2 (0.13), P =2.65 $x10^{-20}$) than for prediction association (β(SE)=-0.03% (0.01), P =0.0072). We noted comparable associations of the Mediation-SI and of packyears with lung function (figure 5). Both were inversely associated with level of FEV_1/FVC. Adding the Mediation-SI or self-reported smoking history (smoking status and packyears) to the different M_{base}-adjusted statistical models showed a comparable increase in total adjusted R^2. Highest total adjusted R^2 was obtained when including both, DNAme score and self-reported smoking history. Covariate-adjusted mean Mediation-SI values decreased from never to former to current smokers and from more distant to more recent smoking exposure - with increase in packyears in current smokers; with fewer years since quitting in former smokers (figure 6).

The assessment of a second DNAme-smoking score (lung-function-genes-SI), based on smoking-related CpGs located in 18 GWAS-identified lung function candidate genes (online supplement table S24), showed less prominent associations with lung function (strongest association observed in ever smokers for FEV_1: β (SE)=-0.196 (0.053), P=0.0002; online supplement table S25).

Discussion

The understanding of how environmental exposure and disease are related to site-specific DNAme status is growing [16, 17]. Our agnostic EWAS on lung function contributes to this body of evidence. Lung function-associated DNAme markers were strongly enriched for smoking-associated loci. More than 50 known smoking CpGs were consistently and in several cases causally associated with lung function and its change in adults. The current agnostic approach converges with recent results of DNAme-lung function studies [4, 6, 7] that were *a priori* focusing on smoking related loci and included pyrosequencing in blood [7] and lung tissue [4] of some of our strongest association signals, including *AHRR* hypomethylation at cg05575921 and cg21161138, cg05951221 and cg21566642 (*ALPPL2)* and cg06126421 (*IER3)*. A methylation index integrating ten DNAme that reportedly mediates the effect of smoking on lung function [4] was associated with lung function level and its change in adults.

Smoking is an important risk factor for poor lung function and accelerated decline. Several EWAS identified a large number of differentially methylated CpG markers to be associated with smoking [2-4]. In particular, the hypo-methylation of cg05575921, a CpG located in the third intron of the aryl hydrocarbon receptor repressor (*AHRR*) gene, investigated for lung function and respiratory symptoms [4], stands out as a robust indicator of smoking status and smoking history [18]. Given the consistency of the associations observed for cg05575921 and the smoking index containing it in this study, the latter may have potential as a biomarker of clinical utility in predicting smoking-related morbidity and mortality [18, 19]. The positive direction of effects observed in identified DNAme-lung function association is in accordance with the reported hypo-methylation of smoking-related DNAme sites. The identified lung function-associated CpGs in this study have been previously reported to be associated with smoking-related molecular phenotypes [20], with increased risk of non-communicable disease, including cancer [18, 21], and with epigenetically defined accelerated ageing [22].

Whether most smoking-related DNAme markers are only markers of exposure or indirectly associated with lung function [7] or whether some inform on causal disease pathways cannot be answered conclusively by the current study. First, DNAme may just be a more precise measure of smoking exposure than self-reporting, as *AHRR* DNAme was previously shown to correlate with the genetic smoking dependency [18]. Second, DNAme identified by previous smoking EWAS [2, 4] may not exclusively have picked up methylation effects of smoking, but methylation related to phenotypes also affected by smoking. In this case, the observed DNAme-lung function associations may result from comorbidity between lung function and other smoking-related phenotypes. However some of the results are consistent with a causal disease pathway. First, MR results support causal effects from some DNAme. Unfortunately no genetic instrument was available for the top ranked AHRR signal. Second, our report confirms nine CpGs, including cg05575921 (*AHRR*), previously shown to mediate the smoking effect of lung function [4]. The observation that many smoking DNAme-lung function associations withstood smoking adjustment is consistent with the mediating role of DNAme between smoking behaviour (more distant predictor) and lung function. Third, smoking was also observed to influence methylation in lung tissue at several lung function CpGs including at cg05575921 in *AHRR* and these methylation levels correlated with *AHRR* gene expression [23] and gene expression of other genes [4]. Hypotheses for a mediating and causal role of smoking-related DNAme include altered *AHRR* DNAme inducing altered phase-2 enzyme activity and toxicant metabolism and altered inflammatory pathways in the lung [7]. Other inhalants impacting on the same pathways could in part explain the observed enrichment for smoking DNAme among never smokers. Methylation of *AHRR* cg05575921 was previously associated with lung function, and chronic bronchitis in never smokers [7]. Maternal smoking, passive smoking, and environmental exposures other than cigarette smoking (e.g. air pollution) are known to modify DNAme patterns across the genome [24-30]. Maternal smoking during pregnancy has been shown to alter offspring's DNA markers in a number of genes known to contain smoking-related CpGs [25, 26], and some of these epigenetic patterns including in *AHRR* persist to adulthood [27].

From our findings in two well-characterized childhood birth cohorts, there was no evidence for shared common epigenetic mechanisms underlying lung function in adult- and childhood. The comparison was driven by results from the lung function EWAS in adults, given sample size limitations in the available birth cohorts. Lung function in childhood versus adulthood is expected to be influenced in part by different biological processes. The non-replication of the mostly smoking-related lung function DNAme signals might reflect the non-smoking status of the children and adolescents. Our findings in SAPALDIA point to a dose-response effect of smoking history and intensity on the smoking index. Effects of maternal exposure in utero, passive smoking or other inhalants on smoking DNAme are likely smaller than the effects of active smoking [28]. Our EWAS findings generally showed an age-related increase in number and strength of DNAme-lung function associations in adults, despite covariate adjustment for age, as also observed by others [6]. This result is consistent with the observed dose-response effect of smoking and possibly other inhalants on DNAme. But the inherent interdependency of lung function decline, cumulative smoking exposure, and DNAme with aging prohibits attributing associations to single factors.

A systematic review of peripheral DNAme associated with lung function in population-based cohorts pointed to the lack of consistent evidence [5]. Epigenome-wide DNAme profiling studies of lung tissue suggested DNAme in genes like *NOS1AP*, *TNFAIP2*, and *CHRM1* to be associated with COPD [12, 13]. An EWAS meta-analysis, adjusted for smoking status and packyears, identified differential DNAme related to COPD and lung function in Koreans. Five loci (*CTU2, USP36, ZNF516, KLK10* and *CPT1B*) were associated with at least two respiratory traits [11]. Evidence of associations in the current EWAS was only observed for 12 of 376 CpGs associated with lung function phenotypes in these previous studies. This inconsistency may be due to differences in population ancestry, disease status, exposure status, tissue-specific methylation or covariate adjustment. Furthermore limited sample size and false discovery findings could contribute to non-replication as could the absence of post-bronchodilation lung function in the current EWAS. However, our results confirm the associations of two most recently published population-based reports [4, 6] investigating smoking,

DNAme and lung function. Both reports and our results reveal the same smoking CpGs as prominent signals.

The strength of this EWAS investigation is the robust and extensive study design with availability of repeat measures of DNAme and spirometry data in the same cohort participants, as well as its population–based design. The utilization of a multi-level analysis scheme - including cross-sectional and longitudinal EWAS analyses at two time points in the same participants, and EWAS with and without smoking adjustment in all participants and in never smokers - allowed for a better understanding of lung function DNAme being affected by aging and smoking. The lung function associated smoking index derived is building on robust evidence that DNAme in blood is correlated with DNAme and gene expression in lung tissue [4, 21, 31] and that it is a valid biomarker for capturing the effect of smoking on DNA methylation in the lung [7, 18].

There are several limitations to this study. Limitations in sample size may explain the inability to find association signals in never smokers and therefore signals common to lung function in childhood and adulthood. The estimation of decline in lung function from only two spirometry time points is likely to misclassify decline. Additionally, not all replication cohorts had data available for more than one time point. Pre-bronchodilation lung function is less robust than post-bronchodilator values and may increase variability of the findings. The meta-analyzed EWAS results of the cross-sectional analyses showed evidence of inflation (λ>1.1) indicating insufficient genomic control. Yet, adjusting for genomic inflation did not alter our main results. The relevance of the smoking index derived from CpGs in or close to lung function GWAS genes can be questioned given evidence on the complex trans-regulation of gene expression [32].

In conclusion, our agnostic investigation shows that DNAme at CpGs related to smoking behaviour are the predominant signals associated cross-sectionally and prospectively with lung function in adults. The findings stimulate further research into the involvement of smoking-related CpGs in lung function relevant mechanism and potentially their role as exposure markers beyond active smoking.

From our EWAS results it has become clear that larger samples are required to confidently identify CpGs involved in lung function and its age-related decline in persons who never smoked.

Acknowledgements

This work has been conducted within the Aging Lungs in European Cohorts (ALEC) project, funded from the European Union's Horizon 2020 research and innovation programme under grant agreement No 633212. For a full list of acknowledgments for each cohort, please see online supplement. Also we gratefully acknowledge the contribution of co-author Professor John M. Starr, who died prior to the publication of this manuscript.

table 1: Characteristics of discovery cohorts.

	SAPALDIA 2 time point 1	SAPALDIA 3 time point 2	ECRHS 2 time point 1	ECRHS 3 time point 2	NFBC1966 age31 time point 1	NFBC1966 age46 time point 2
N	962	962	470	470	611	611
Female, %	53.5	53.5	56	56	55.3	55.3
Age (years), mean (SD)	50.5 (11.3)	58.8 (11.3)	43.6 (6.8)	54.5 (6.8)	31.0 (0.3)	46.3(0.4)
Height (cm), mean (SD)	169.4 (9.2)	168.7 (9.4)	170.0 (9.2)	169.2 (9.3)	171 (8.8)	171 (8.9)
Weight (kg), mean (SD)	74.2 (14.7)	75.5 (15.4)	72.6 (14.6)	76.2 (15.5)	71.3 (13.6)	78.7 (16.3)
Body mass index, (kg/m2) mean (SD)	25.8 (4.4)	26.5 (4.6)	25.0 (4.0)	26.5 (4.4)	24.2 (3.7)	26.7(4.8)
Smoking status, %						
Never smoker*	41.7	41.1	43.2	41.7	54.5	54.5
Former smoker	30.0	37.0	31.1	40.4	21.3	30.2
Current smoker	28.3	21.9	25.7	17.9	24.1	15.3
Packyears, mean (SD)	20.4 (20.2)	22.6 (22.1)	16.6 (16.9)	20.0 (21.3)	7.7 (5.9)	11.0 (9.6)
Education†, %						
Low	5.4	5.4	11.5	11.5	0.65	0.65
Intermediate	65.7	65.7	29.2	29.2	55.9	55.9
High	28.9	28.9	59.3	59.3	43.3	43.3
FVC (L), mean (SD) ‡	4.4 (1.0)	4.1 (1.1)	4.3 (1.0)	3.9 (1.0)	4.8 (1.0)	4.5 (0.9)
FEV_1 (L), mean (SD) ‡	3.3 (0.8)	3.0 (0.8)	3.4 (0.7)	3.0 (0.8)	4.0 (0.8)	3.5 (0.7)
FEV_1/FVC, mean (SD) ‡	0.75 (0.07)	0.73 (0.08)	0.78 (0.06)	0.75 (0.06)	0.83 (0.06)	0.77 (0.06)
Airflow obstruction (FEV_1/FVC<0.7) ‡, %	20.4	29.5	8.9	16.2	1.8	10.8
Airflow obstruction (FEV_1/FVC<LLN) ‡¶, %	12.9	14.1	8.7	10.4	3.27	9.49
Doctor-diagnosed asthma, %	13.8	16.5	14.3	16.8	10.7	15.8
Respiratory medication, % (% missing values)	22.2 (0.8)	23.7 (0.3)	13.4	14.2	n.a.	n.a.

Footnote to table 1A:

* self-reported lifetime non-smoking.† The categorical variable "education" is defined differently in cohorts. In SAPALDIA low corresponds to primary education; intermediate to secondary, middle, or vocational school; and high to technical college or university. In ECRHS and NFBC1966 information of age reached at end of studies is used to define low : ≤16 years; intermediate : 17-19 years and high : ≥20 years.

‡ Values derived from pre-bronchodilation spirometry. Lung function values corrected for spirometer device change in SAPALDIA at time point 2.

¶ LLN: lower limit of normal values estimated using GLI2012 reference equations. n.a. not assessed.

table 1 continued: Characteristics of adult replication cohorts.

	KORA time point 1	KORA time point 2	LCB1936 time point 1	LCB1936 time point 2	Lifelines time point 1	NSPHS time point 1	FTC time point 1
N	628	628	449	449	1622	535	93
Female, %	53.2	53.2	46.8	46.8	42.8	53.1	47.3
Age (years), mean (SD)	53.6 (4.5)	60.1 (4.5)	69.6 (0.9)	76.3 (0.7)	46.7 (10.8)	55.1 (16.0)	30.4 (3.8)
Height (cm), mean (SD)	169.5 (9.3)	168.7 (9.4)	167.2 (8.8)	166.1 (8.81)	176.9 (9.1)	163.8 (9.8)	173.03 (10.5)
Weight (kg), mean (SD)	79.0 (16.7)	79.9 (17.3)	77.2 (14.6)	76.5 (14.8)	82.1 (14.7)	74.0 (15.2)	82.04 (18.8)
Body mass index (kg/m2), mean (SD)	27.4 (4.7)	28.0 (5.1)	27.5 (4.3)	27.7 (4.6)	26.2 (3.9)	27.5 (4.7)	27.30 (5.4)
Smoking status, %							
Never smoker*	38.2	38.2	52.3	52.3	56.6	83.2	53.8
Former smoker	43.8	45.5	40.8	41.9	0¶	n.a.¶	26.9
Current smoker	18.0	16.2	6.9	5.8	43.5	16.5	19.4
Packyears, mean (SD)	12.8 (19.3)	13.5 (20.2)	13.9 (24.0)	14.1 (24.6)	21.0 (11.7)	8.1 (21.6)	n.a.
Education† (%)							
Low	47.6	47.6	49.7	49.7	23.1	n.a.	1.1
Intermediate	26.43	26.43	32.3	32.3	40.8	n.a.	38.6
High	25.96	25.96	18.0	18.0	35.4	n.a.	60.2
FVC (L), mean (SD) ‡	4.3(1.0)	3.9 (1.0)	3.2 (0.9)	2.8 (0.9)	4.7 (1.1)	3.4 (1.1)	4.8 (1.1)
FEV_1 (L), mean (SD) ‡	3.3 (0.8)	3.0 (0.7)	2.5 (0.7)	2.1 (0.7)	3.5 (0.9)	2.8 (0.9)	3.9 (0.9)
FEV_1/FVC, mean (SD) ‡	0.78 (0.06)	0.75 (0.07)	0.79 (0.09)	0.76 (0.12)	0.73 (0.09)	0.83 (0.09)	0.81 (0.07)
Airflow obstruction (FEV_1/FVC<0.7) ‡, %	8.12	20.06	15.4	26.3	38.4	8.8	5.0
Airflow obstruction (FEV_1/FVC<LLN)‡, %	5.0	9.6	7.6	14.9	27.5	4.3	11.3
Doctor-diagnosed asthma, %	7.2	8.6	4.5	7.1	9.9	14.2	0
Respiratory medication, % (% missing values)	3.3	4.9	6.7	11.8	8.0	7.7	0

Footnote to table 1 continued:

* self-reported lifetime non-smoking.

† The categorical variable education is defined differently in cohorts.

‡ Values derived from pre-bronchodilation spirometry. LLN: lower limit of normal values estimated using GLI2012 reference equations.

¶ LifeLines: non-random selection of samples for DNA methylation typing (current smokers versus never smokers. NSPHS: information obtained on current smoking status (yes/no).

Abbreviations: n.a. not assessed.

table 2: Combined EWAS meta-analyses of cross-sectional associations* of CpG markers with FEV_1/FVC in all participants, base model covariate adjusted EWAS (M_{base}†). Meta-analyses of cross-sectional associations obtained using data from the oldest time point available: time point 2 of ECRHS, NFBC1966, SAPALDIA, LBC1936 and time point 1 of KORA, LifeLines and NSPHS.

CpG ID	chr	position (hg19)	Locus	beta* (SE)	P-value meta-analysis	direction of effects¶	P-value between study hetero-geneity	replicated P-value <0.0011	*previously reported smoking CpG‡*	*previously reported smoking FDR P-value ‡*	*previously reported smoking association direction of effect ‡*
cg05575921	5	373378	*AHRR*	0.124 (0.008)	**7.22E-50**	+++++++	0.023	**yes**	*yes***	*6.10E-22*	*(-)*
cg03636183	19	17000585	*F2RL3*	0.201 (0.015)	**4.50E-43**	+++++++	0.008	**yes**	*yes***	*5.70E-17*	*(-)*
cg21566642	2	233284661	*ALPPL2*	0.151 (0.011)	**5.02E-43**	+++++++	0.043	**yes**	*yes***	*4.50E-21*	*(-)*
cg01940273	2	233284934	*ALPPL2*	0.206 (0.015)	**4.09E-41**	+++++++	0.031	**yes**	*yes***	*9.80E-30*	*(-)*
cg03329539	2	233283329	*ALPPL2*	0.257 (0.023)	**5.58E-30**	+++++++	0.628	**yes**	*yes*	*9.70E-16*	*(-)*
cg21161138	5	399360	*AHRR*	0.243 (0.021)	**9.72E-30**	+++++++	0.152	**yes**	*yes***	*7.90E-13*	*(-)*
cg23771366	11	86510998	*PRSS23*	0.233 (0.022)	**5.38E-27**	+++++++	0.286	**yes**	*yes*	*1.90E-14*	*(-)*
cg11660018	11	86510915	*PRSS23*	0.238 (0.023)	**3.40E-26**	+++++++	0.318	**yes**	*yes*	*4.40E-21*	*(-)*
cg21611682	11	68138269	*LRP5*	0.309 (0.03)	**1.26E-25**	+++++++	0.049	**yes**	*yes*	*4.20E-15*	*(-)*
cg15342087	6	30720209	*IER3*	0.359 (0.036)	**5.44E-24**	+++++++	0.169	**yes**	*yes*	*3.90E-14*	*(-)*
cg26703534	5	377358	*AHRR*	0.266 (0.026)	**7.34E-24**	+++++++	0.101	**yes**	*yes*	*7.20E-18*	*(-)*
cg25648203	5	395444	*AHRR*	0.25 (0.026)	**9.84E-22**	+++++++	0.194	**yes**	*yes*	*2.70E-11*	*(-)*
cg19572487	17	38476024	*RARA*	0.196 (0.021)	**8.87E-21**	+++++++	0.018	**yes**	*yes*	*1.60E-16*	*(-)*
cg00310412	15	74724918	*SEMA7A*	0.261 (0.028)	**4.01E-20**	+++++++	0.275	**yes**	*yes*	*1.20E-13*	*(-)*
cg24859433	6	30720203	*IER3*	0.303 (0.034)	**2.05E-19**	+++++++	0.067	**yes**	*yes***	*2.20E-09*	*(-)*
cg09935388	1	92947588	*GFI1*	0.105 (0.012)	**7.05E-19**	+++++++	0.034	**yes**	*yes***	*7.00E-14*	*(-)*
cg14753356	6	30720108	*IER3*	0.189 (0.021)	**9.08E-19**	+++++++	0.405	**yes**	*yes*	*2.30E-14*	*(-)*
cg04885881	1	11123118	*SRM/EXOSC10*	0.168 (0.02)	**5.66E-18**	+++++++	0.670	**yes**	*yes*	*2.70E-11*	*(-)*
cg25949550	7	145814306	*CNTNAP2*	0.335 (0.039)	**6.04E-18**	+++++++	0.013	**yes**	*yes*	*9.30E-21*	*(-)*
cg19859270	3	98251294	*GPR15*	0.467 (0.055)	**2.80E-17**	+++++++	0.029	**yes**	*yes*	*6.30E-17*	*(-)*
cg03450842	10	80834947	*ZMIZ1*	0.265 (0.031)	**2.92E-17**	+++++++	0.003	**yes**	*yes*	*2.40E-11*	*(-)*
cg03707168	19	49379127	*PPP1R15A*	0.206 (0.025)	**1.27E-16**	+++++++	0.668	**yes**	*yes*	*3.50E-07*	*(-)*
cg17087741	2	233283010	*ALPPL2*	0.161 (0.02)	**4.48E-16**	++++++-	<0.001	**yes**	*yes*	*6.10E-07*	*(-)*
cg21140898	1	51442318	*CDKN2C*	0.12 (0.017)	**4.46E-13**	+++++++	0.103	**yes**	*yes*	*3.70E-08*	*(-)*
cg01899089	5	369969	*AHRR*	0.172 (0.027)	**1.47E-10**	+++++++	0.005	**yes**	*yes*	*1.80E-12*	*(-)*
cg08763102	4	3079751	*HTT*	0.225 (0.039)	**1.20E-08**	++++++-	0.001	**yes**	*yes*	*3.80E-15*	*(-)*
cg21282907	6	74289980	*SLC17A5*	0.176 (0.031)	**1.28E-08**	++++++-	0.003	no	*yes*	*1.28E-02*	*(-)*
cg20853880	2	10184444	*KLF11*	0.077 (0.014)	**6.05E-08**	+++++++	0.052	no	*yes*	*3.70E-07*	*(-)*

cg16391678	16	30485597	*ITGAL*	0.164 (0.031)	**1.15E-07**	++++++-	0.003	**yes**	*yes*	*3.00E-11*	*(-)*	

Footnote to table 2:

For complete results for FEV_1/FVC associations see Supplementary Material table S3. See online supplement table S4 and table S5 analogous results for FEV_1 or FVC, respectively.

* Presentation of CpG markers showing meta-analysis P-value < $5x10^{-7}$ in the combined meta-analysis. Note: DNAme predictors used were technical bias-adjusted, normalized residuals, thus effect size of the association (beta) are not directly comparable to elsewhere reported effect sizes using normalized %-methylation as predictor.

† Base model (M_{base}) EWAS were covariate adjusted for age, age squared, height, squared deviation from the mean of height, sex and interaction terms of age, age squared, height and squared deviation of height with sex, education (low, medium, high), body mass index, spirometer type, study center as well as cell composition.

‡ Smoking CpGs defined on the reported FDR corrected P-value <0.05 for association reported with smoking status and reported direction of effects for association with smoking.[2]

¶Order of cohorts: ECRHS, NFBC1966, SAPALDIA, KORA, LBC36, LifeLines, NSPHS. FTC was excluded from this meta-analysis, given the smaller sample size and lower mean age (30.4 years) compared to the other adult cohorts (ECRHS (mean age: 54.5 years), NFBC1966 (46.3 years), SAPALDIA (58.8 years), LBC1936 (76.3 years) and the single available time point for KORA (60.1 years), LifeLines (46.7 years) and NSPHS (55.1 years)).

** Smoking CpG previously reported to mediate the smoking effect on lung function. [4]

Abbreviations: beta – coefficient of association; chr – chromosome; hg19 – human genome build 19; n.a – not assessed; SE – standard error.

table 3: Combined meta-analyses of the prediction associations* of CpG markers on change in FEV_1/FVC ($year^{-1}$), in all participants, base model adjustment (Mbase†).

				Combined meta-analysis (ECRHS/NFBC/SAPALDIA/KORA/LBC1936)							
CpG ID	**chr**	**position (hg19)**	**Locus**	**beta* (SE)**	**P-value meta-analysis**	**direction of effect¶**	**P-value between study hetero-geneity**	**replicated P-value <0.0011**	*previously reported smoking CpG‡*	*previously reported smoking FDR P-value ‡*	*previously reported smoking association direction of effect ‡*
cg05575921	5	373378	*AHRR*	0.006(0.001)	**2.77E-13**	+++++	0.005	**yes**	*yes***	*6.10E-22*	*(-)*
cg21566642	2	233284661	*ALPPL2*	0.006(0.001)	**3.17E-11**	+++++	0.235	**yes**	*yes***	*4.50E-21*	*(-)*
cg01940273	2	233284934	*ALPPL2*	0.009(0.001)	**4.93E-11**	+++++	0.023	**yes**	*yes***	*9.80E-30*	*(-)*
cg21161138	5	399360	*AHRR*	0.011(0.002)	**5.81E-09**	+++++	0.103	**yes**	*yes***	*7.90E-13*	*(-)*
cg03636183	19	17000585	*F2RL3*	0.008(0.001)	**6.22E-09**	+++++	0.001	**yes**	*yes***	*5.70E-17*	*(-)*
cg01377124	2	237172609	*ASB18*	-0.018(0.003)	**7.38E-08**	--+-+	0.005	no	*no*	*n.a.*	*n.a.*
cg03329539	2	233283329	*ALPPL2*	0.011(0.002)	**7.66E-08**	+++++	0.015	**yes**	*yes*	*9.70E-16*	*(-)*
cg07222133	5	179499488	*RNF130*	-0.009(0.002)	**2.45E-07**	?-+-+	<0.001	no	*no*	*n.a.*	*n.a.*
cg14366110	9	133779382	*FIBCD1*	0.014(0.003)	9.62E-07	+++--	0.206	no	*no*	*n.a.*	*n.a.*

Footnote to table 3:

For complete results for FEV_1/FVC and analogous results for FEV_1 or FVC see Supplementary Material table S8.

* Predictive associations of DNA methylation at first time point (DNAme 1) with annual change in lung function during follow-up, defined as lung function at second time point – lung function at first time point divided by the time of follow-up in years. Presentation of CpG markers showing meta-analysis P-value < $5x10^{-7}$ at discovery or combined meta-analyses level. CpGs shown sorted by statistical significance of combined meta-analysis results. Note: DNAme predictors used were technical bias-adjusted, normalized residuals, thus effect size of the association (beta) are not directly comparable to elsewhere reported effect sizes using normalized %-methylation as predictor.

† Base model (M_{base}) EWAS were covariate adjusted for age, age squared, height, FEV_1/FVC at time point 1, squared deviation from the mean of height, sex and interaction terms of age, age squared, height and squared deviation of height with sex, education (low, medium, high), body mass index, spirometer type, study center as well as cell composition.

‡ Replication was defined for association if replication P-value<0.0011 (multiple testing correction, 47 tests for FEV_1/FVC).

¶ Smoking CpGs defined on the reported FDR corrected P-value <0.05 for association reported with smoking status and reported direction of effects for association with smoking. [2]

** Smoking CpG previously reported to mediate the smoking effect on lung function. [4]

Abbreviations: beta – coefficient of association; chr – chromosome; hg19 – human genome build 19; n.a – not assessed; SE – standard error.

table 4: Combined meta-analyses* of cross-sectional associations on FEV_1/FVC in never smokers only, obtained using data from time point T2 of ECRHS, NFBC1966, SAPALDIA, LBC1936 and from time point T1 of KORA, LifeLines and NSPHS.

				Combined meta-analysis (ECRHS/ NFBC1966/ SAPALDIA/KORA/ LBC1936/ Lifelines/ NSPHS)							
CpG ID	**chr**	**position (hg19)**	**Locus**	**beta* (SE)**	**P-value meta-analysis**	**direction of effects¶**	**P-value between study hetero-geneity**	**replicated P-value <0.0011**	*previously reported smoking CpG‡*	*previously reported smoking FDR P-value ‡*	*previously reported smoking association direction of effect ‡*
cg09884077	15	23086698	*NIPA1*	-0.308 (0.084)	0.0003	---+---	0.001	no	*no*	*n.a.*	*n.a.*
cg25758394	1	3623859	*TP73*	0.213 (0.083)	0.0107	??+--+-	<0.001	no	*no*	*n.a.*	*n.a.*
cg18664508	3	169487465	*ARPM1*	-0.308 (0.072)	2.02E-05	+------	<0.001	no	*no*	*n.a.*	*n.a.*
cg19268386	15	23086595	*NIPA1*	-0.263 (0.14)	0.0615	??-----	<0.001	no	*no*	*n.a.*	*n.a.*
cg15981995	3	169487311	*ARPM1*	-0.231 (0.073)	0.0016	??----+	<0.001	no	*no*	*n.a.*	*n.a.*
cg05785298	1	204654622	*LRRN2*	-0.423 (0.111)	1.41E-04	-+-+-+-	0.001	no	*no*	*n.a.*	*n.a.*
cg20278790	20	57583474	*CTSZ*	0.319 (0.07)	5.01E-06	-+++---	<0.001	no	*no*	*n.a.*	*n.a.*
cg13562246	8	33368277	*C8orf41*	0.349 (0.074)	2.67E-06	+++++-+	0.206	no	*no*	*n.a.*	*n.a.*

Footnote to table 4:

For complete results for FEV_1/FVC and for FEV_1 or FVC in never smokers see online supplement table S8.

* Presentation of CpG markers showing meta-analysis P-value < $5x10^{-7}$ at discovery level for cross-sectional association at time point 2, base model adjustment (M_{base}†), using data from time point T2 of ECRHS, NFBC1966, SAPALDIA, LBC1936 and from time point T1 of KORA, LifeLines and NSPHS. Note: DNAme predictors used were technical bias-adjusted, normalized residuals, thus effect size of the association (beta) are not directly comparable to elsewhere reported effect sizes using normalized %-methylation as predictor.

† Base model (M_{base}) EWAS were covariate adjusted for age, age squared, height, squared deviation from the mean of height, sex and interaction terms of age, age squared, height and squared deviation of height with sex, education (low, medium, high), body mass index, spirometer type, study center as well as cell composition.

‡ Smoking CpGs defined on the reported FDR corrected P-value <0.05 for association reported with smoking status and reported direction of effects for association with smoking.[2]

¶ Order of cohorts: for discovery: ECRHS, NFBC1966, SAPALDIA; for combined analysis: ECRHS, NFBC1966, SAPALDIA, KORA, LBC1936, LifeLines, NSPHS.

Abbreviations: beta – coefficient of association; chr – chromosome; hg19 – human genome build 19; n.a – not assessed; SE – standard error.

table 5: Combined meta-analyses of the prediction associations* of CpG markers on change in FEV_1/FVC ($year^{-1}$), in never smokers only, base model adjustment (M_{base}†).

				Combined meta-analysis (ECRHS/NFBC/SAPALDIA/KORA/LBC1936)							
CpG ID	**chr**	**position**	***Locus***	**beta* (SE)**	**P-value meta-analysis**	**direction of effects¶**	**P-value between study hetero-geneity**	**replicated P-value <0.0011**	*previously reported smoking CpG‡*	*previously reported smoking FDR P-value ‡*	*previously reported smoking association direction of effect ‡*
cg14366110	9	133779382	*FIBCD1*	0.017 (0.003)	**3.60E-09**	++-++	0.315	no	*no*	*n.a.*	*n.a.*
cg11216682	2	131113867	*PTPN18*	-0.017 (0.003)	**1.10E-07**	+-+--	0.282	no	*no*	*n.a.*	*n.a.*

Footnote to table 5:

For complete results for FEV_1/FVC and for analogous results for FEV_1 or FVC see Supplementary Material table S9.

* Predictive associations of DNA methylation at first time point (DNAme 1) with change in lung function during follow-up, defined as lung function at second time point – lung function at first time point divided by the time of follow-up in years. Presentation of CpG markers showing meta-analysis P-value < $5x10^{-7}$ at discovery or replication level. Note: DNAme predictors used were technical bias-adjusted, normalized residuals, thus effect size of the association (beta) are not directly comparable to elsewhere reported effect sizes using normalized %-methylation as predictor.

† Base model (M_{base}) EWAS were covariate adjusted for age, age squared, height, FEV_1/FVC at time point 1, squared deviation from the mean of height, sex and interaction terms of age, age squared, height and squared deviation of height with sex, education (low, medium, high), body mass index, spirometer type, study center as well as cell composition.

‡ Smoking CpGs defined on the reported FDR corrected P-value <0.05 for association reported with smoking status and reported direction of effects for association with smoking. [2]

¶ Order of cohorts: for discovery: ECRHS, NFBC1966, SAPALDIA; for combined analysis: ECRHS, NFBC1966, SAPALDIA, KORA, LBC1936.

Abbreviations: beta – coefficient of association; chr – chromosome; n.a – not assessed; SE – standard error.

table 6: Mediation* analysis on the role of previously reported CpGs in the smoking association with FEV_1/FVC, the SAPALDIA cohort.
For analogous results for FEV_1 or FVC see online supplementary table S22.

CpG†	Locus	ACME			ADE			Total effect			Proportion		
		Estimate	95%CI	p	Estimate	95%CI	p-value	Estimate	95%CI	p	Estimate	95%CI	p
cg01940273	*ALPPL2*	-0.0079	[-0.0119, -0.0041]	**<0.0001**	-0.0026	[-0.0129, 0.0077]	0.604	-0.0106	[-0.0203, -0.0014]	0.026	0.7313	[0.2616, 3.4325]	0.026
cg03636183	*F2RL3*	-0.0080	[-0.0122, -0.0040]	**<0.0001**	-0.0029	[-0.0126, 0.0062]	0.556	-0.0108	[-0.0197, -0.0021]	0.018	0.7312	[0.2819, 2.9097]	0.018
cg05575921	*AHRR*	-0.0102	[-0.0147, -0.0055]	**<0.0001**	-0.0008	[-0.0109, 0.0086]	0.870	-0.0110	[-0.0202, -0.0020]	0.012	0.9213	[0.3818, 4.0453]	0.012
cg05951221	*ALPPL2*	-0.0075	[-0.0122, -0.0030]	**0.002**	-0.0033	[-0.0131, 0.0062]	0.520	-0.0109	[-0.0197, -0.0022]	0.020	0.6836	[0.1942, 2.7656]	0.022
cg06126421	*IER3*	-0.0054	[-0.0093, -0.0017]	**<0.0001**	-0.0049	[-0.0148, 0.0049]	0.328	-0.0103	[-0.0194, -0.0012]	0.030	0.5233	[0.1050, 2.5558]	0.030
cg09935388	*GFI1*	-0.0033	[-0.0058, -0.0010]	**0.002**	-0.0073	[-0.0168, 0.0022]	0.122	-0.0105	[-0.0198, -0.0013]	0.034	0.3009	[0.0568, 1.4190]	0.036
cg21161138	*AHRR*	-0.0056	[-0.0089, -0.0025]	**<0.0001**	-0.0052	[-0.0146, 0.0043]	0.282	-0.0108	[-0.0194, -0.0020]	0.020	0.5127	[0.1647, 2.0961]	0.020
cg21566642	*ALPPL2*	-0.0098	[-0.0145, -0.0057]	**<0.0001**	-0.0014	[-0.0116, 0.0089]	0.796	-0.0112	[-0.0209, -0.0011]	0.024	0.8663	[0.3453, 4.6567]	0.024
cg22994830	*PRKAR1B*	-0.0002	[-0.0009, 0.0003]	0.542	-0.0103	[-0.0201, -0.0010]	0.028	-0.0105	[-0.0202, -0.0013]	0.024	0.0103	[-0.0470, 0.1595]	0.550
cg24859433	*IER3*	-0.0024	[-0.0053, 0.0002]	**0.068**	-0.0082	[-0.0179, 0.0013]	0.112	-0.0107	[-0.0201, -0.0014]	0.022	0.2186	[-0.0438, 1.2776]	0.090

Footnote to table 6:
* performed using R package mediation [33].
† previously reported candidate CpG for mediation of smoking effect on lung function [4].
Abbreviations: ACME – average causal mediation effect; ADE – average direct effect.

table 7: Meta-analyses* of discovery cohort specific association of Mediation smoking index (SI) with FEV_1/FVC (%), cross-sectionally at time point 2, and longitudinally predicting the annual change (%/year) during follow-up, base model adjustment (M_{base}†), in all study participant, ever and never smokers.

	Cross-sectional meta-analysis at time point 2*				**Prediction on change in lung function†**			
	beta (SE)	**P-value¶**	**Direction¶**	**P-value between study hetero-geneity**	**beta (SE)**	**P-value¶**	**Direction¶**	**P-value between study hetero-geneity**
All	-0.012 (0.0013)	1.05E-20	---	0.44	-0.0005 (0.0001)	8.66E-09	---	0.006
Ever smokers	-0.014 (0.0016)	3.28E-18	---	0.30	-0.0004 (0.0001)	4.94E-04	---	0.13
Never smokers	-0.0033 (0.0041)	0.423	--+	0.62	-0.0007 (0.0002)	1.73E-04	+-+	0.003

Footnotes to table 7:

*Cohort-specific association results for Mediation-SI were meta-analysed. The 10 CpGs contributing Mediation-SI are shown in online supplement table S21). For analogous results of associations of Mediation-SI with FEV_1 and FVC see online supplement table S23. Note: DNAme predictors used were technical bias-adjusted, normalized residuals, thus effect size of the association (beta) are not directly comparable to elsewhere reported effect sizes using normalized %-methylation as predictor.

† Base model covariate adjustment (M_{base}): age, age squared, height, squared deviation from the mean of height, sex and interaction terms of age, age squared, height and squared deviation of height with sex, education (low, medium, high), body mass index, spirometer type, study center as well as cell composition. Prediction models were additionally adjusted for FEV_1/FVC at time point 1.

‡ P-value of meta-analysis: P<0.008 was considered statistically significant, Bonferroni correction for 6 tests per lung function outcome Order of cohorts for direction of effects: ECRHS, NFBC1966, SAPALDIA.

Abbreviations: beta – coefficient of association; chr – chromosome; SE – standard error.

figure 1: Flow of the multi-level discovery design of the epigenome wide association study (EWAS) on lung function parameters FEV_1, FEV_1/FVC and FVC.
* Base model (M_{base}) EWAS were covariate adjusted for age, age squared, height, squared deviation from the mean of height, sex and interaction terms of age, age squared, height and squared deviation of height with sex, education (low, medium, high), body mass index, spirometer type, study center, and cell composition. ** Smoking model EWAS (M_{smok}) were additionally adjusted for smoking covariates: history of smoking intensity as pack years smoked up to the time point of data collection for regressions and for smoking status (current, former and never smoker). EWAS longitudinally predicting the change in lung function ($EWAS_{predict}$) was additionally adjusted for lung function at time point 1. DNAme1 – DNA methylation at time point 1; DNAme2 – DNA methylation at time point 2.

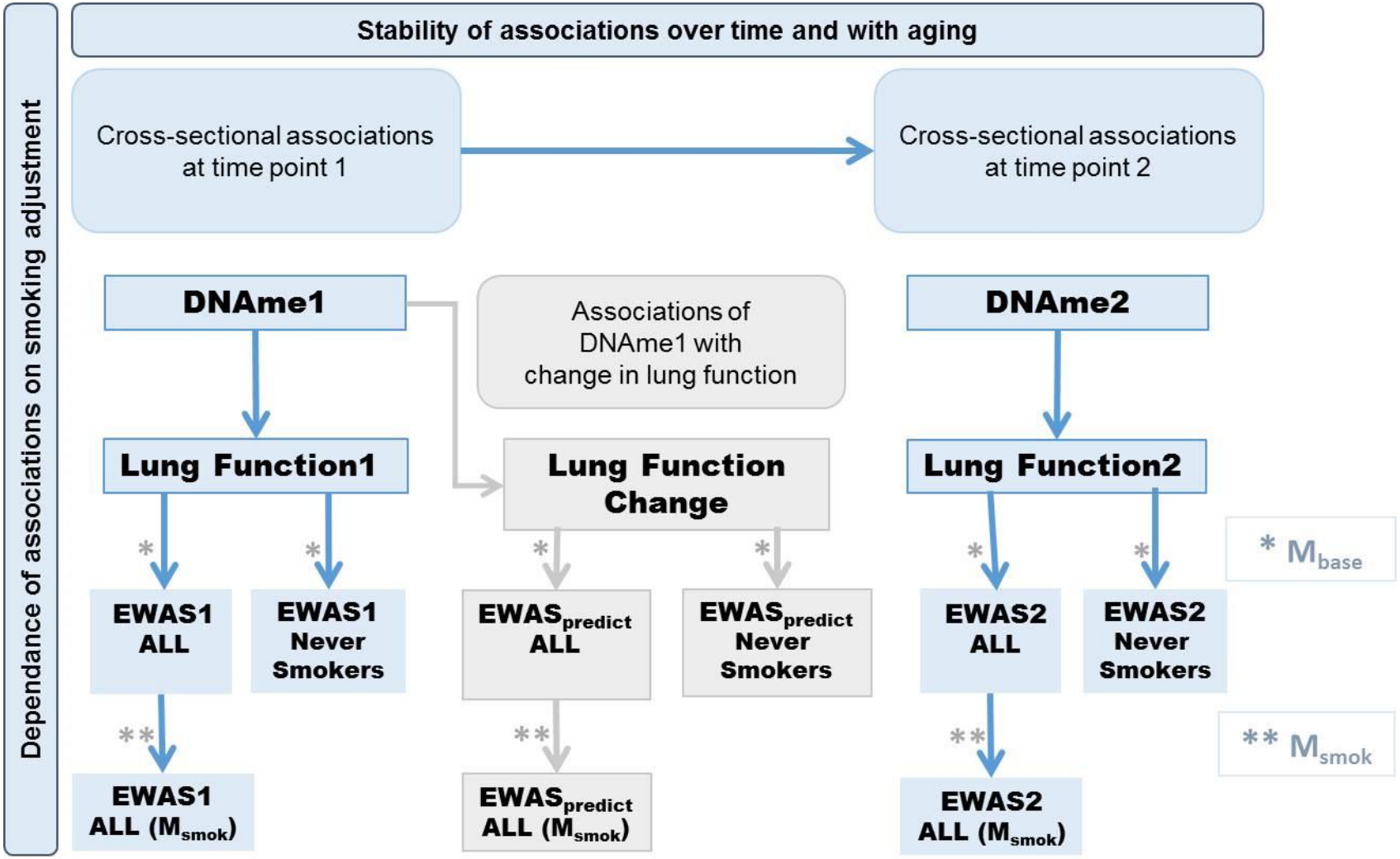
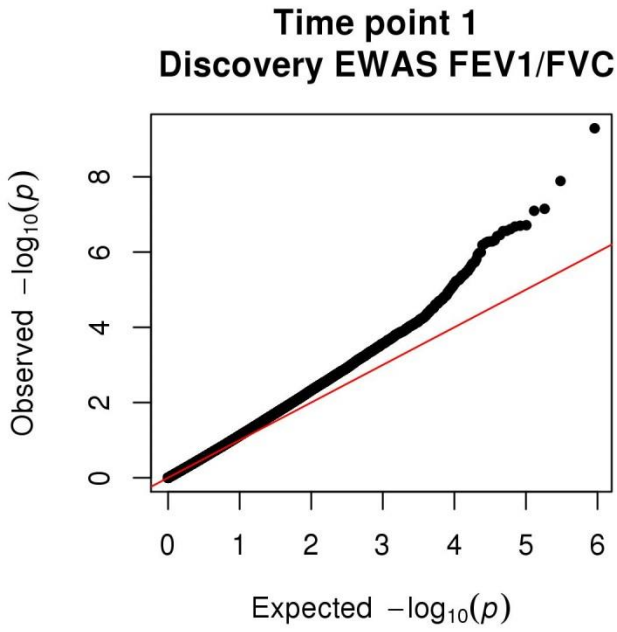

Dependance of associations on smoking adjustment
Stability of associations over time and with aging
Cross-sectional associations at time point 1
Cross-sectional associations at time point 2
DNAme1
Associations of DNAme1 with change in lung function
DNAme2
Lung Function1
Lung Function Change
Lung Function2
EWAS1 ALL
EWAS1 Never Smokers
EWAS predict ALL
EWAS predict Never Smokers
EWAS2 ALL
EWAS2 Never Smokers
EWAS1 ALL (M smok)
EWAS predict ALL (M smok)
EWAS2 ALL (M smok)
* M base
** M smok

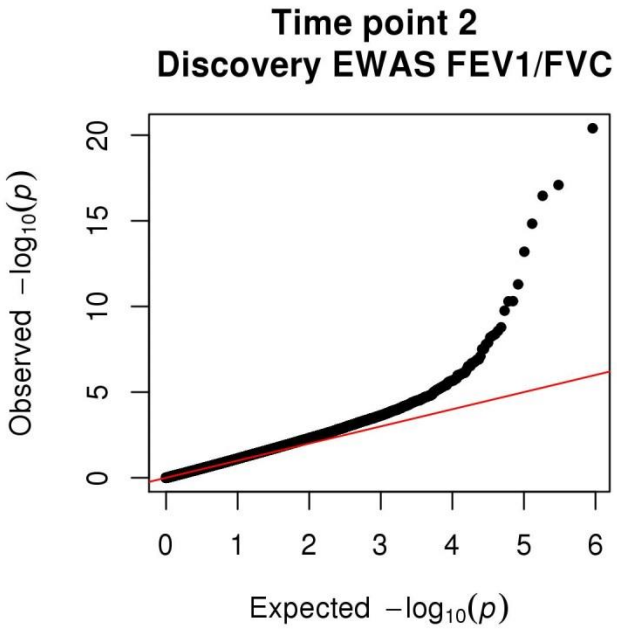
figure 2:

A - Effect of aging on the associations between DNAme and lung function: Increase in numbers of signals with aging. For FEV_1/FVC, we identified 21 CpGs at time point 2 compared to three CpGs at time point 1 to be statistically significant. Quantile-Quantile plots of cross-sectional covariate-adjusted EWAS (M_{base}*) on FEV_1/FVC at first and second time point, all participants. Meta-analyses were performed without genomic control (inflation factor λ for time point 1 (λ = 1.15) and for time point 2 (λ = 1.14)). Analogous figures for cross-sectional association with FEV_1 and with FVC, see online supplement figure S2.

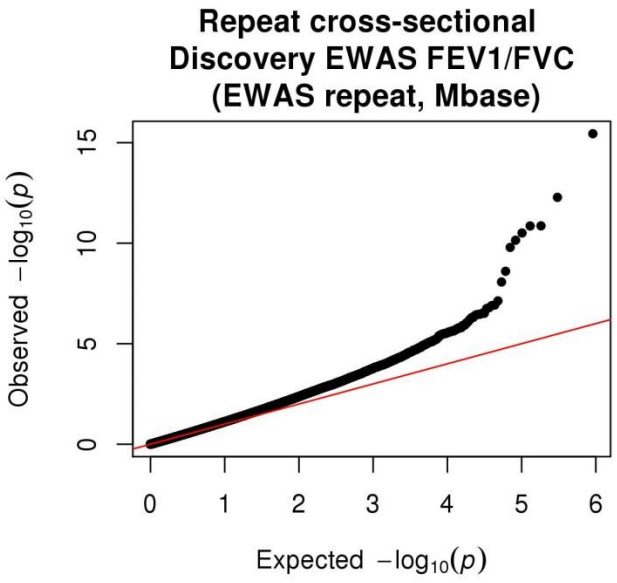
B - Effect of smoking adjustment on the associations between DNAme and lung function: Quantile-Quantile plots of the repeat cross-sectional covariate-adjusted EWAS (M_{base}*, λ = 1.13)) and additionally smoking adjusted (M_{smok}*, λ = 1.05), all participants. Decrease in numbers of signals after smoking adjustment.

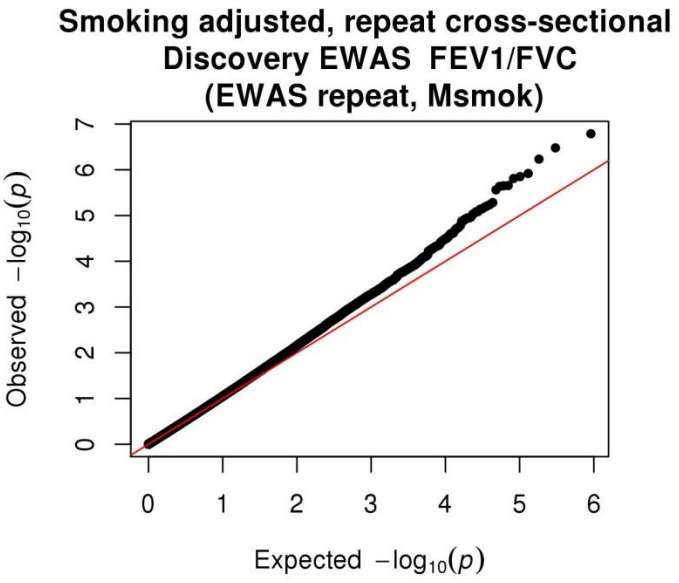
A





B

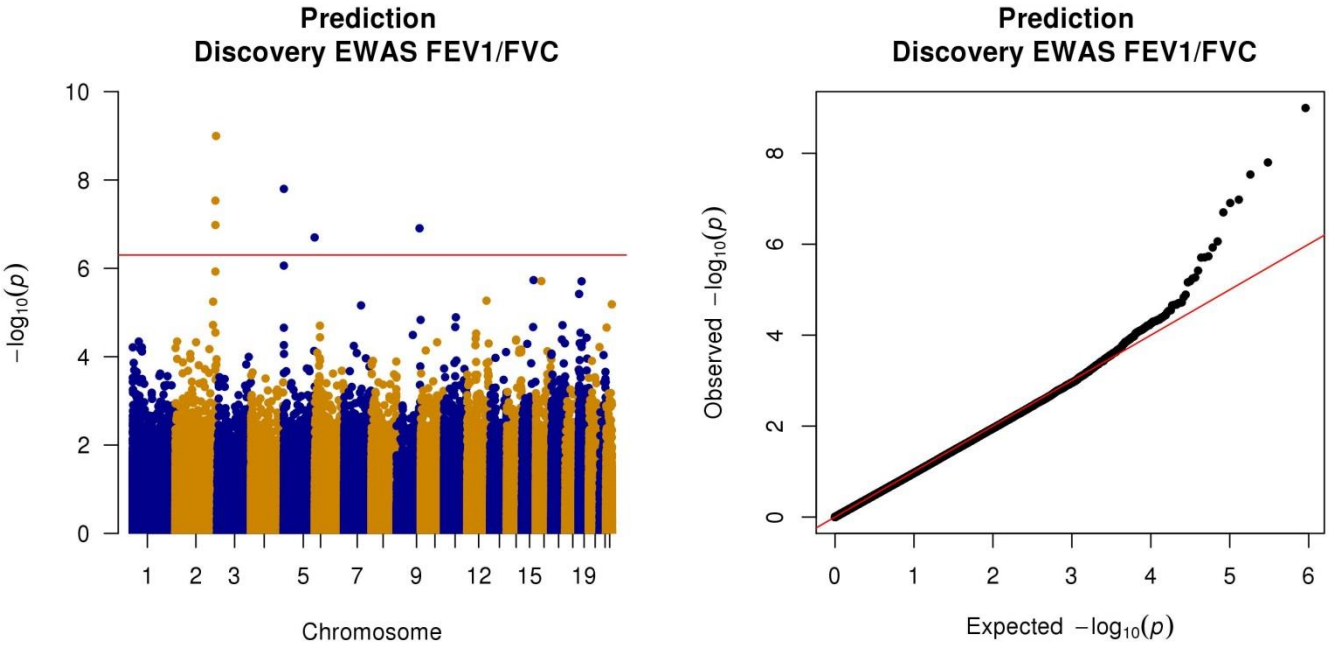




Footnote to figure 2:

*Base model (M_{base}): EWAS were covariate adjusted for age, age squared, height, squared deviation from the mean of height, sex and interaction terms of age, age squared, height and squared deviation of height with sex, education (low, medium, high), body mass index, spirometer type, study center, and cell composition. Smoking adjusted model (M_{smok}): Covariates applied for M_{base} and additionally smoking status and packyears smoked

figure 3: Manhattan and Quantile-Quantile plots of covariate-adjusted prediction* EWAS (M_{base}†) on FEV_1/FVC in all participants. Meta-analysis of the prediction association was performed without genomic control ($\lambda = 0.95$). Analogous figures for association with change in FEV_1 and with change in FVC, see online supplement figure S4.

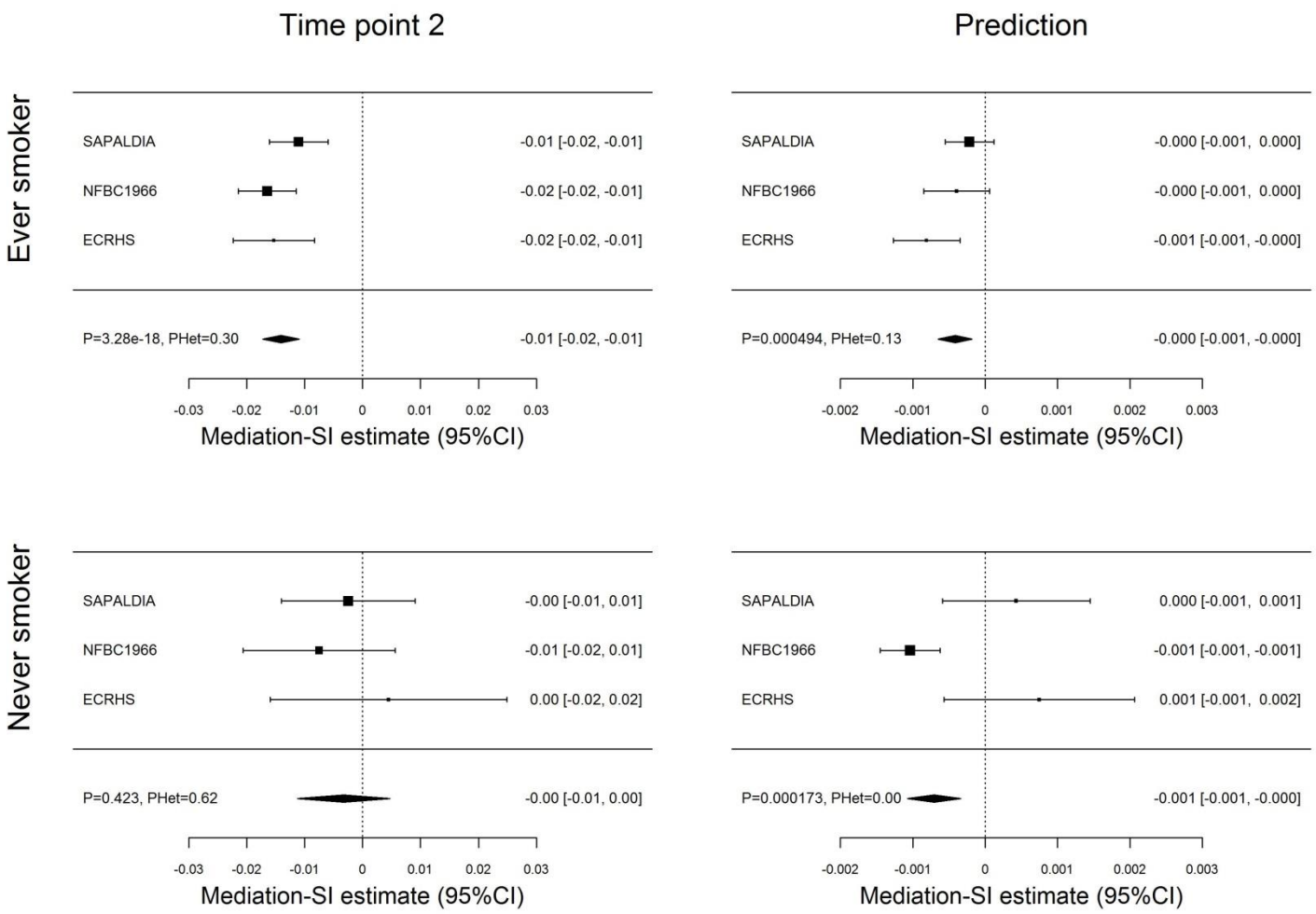


Footnote to figure 3:

* Predictive associations of DNA methylation at first time point (DNAme1) with change in lung function during follow-up

†Base model (M_{base}): EWAS were covariate adjusted for age, age squared, height, FEV_1/FVC at time point 1, squared deviation from the mean of height, sex and interaction terms of age, age squared, height and squared deviation of height with sex, education (low, medium, high), body mass index, spirometer type, study center, and cell composition.

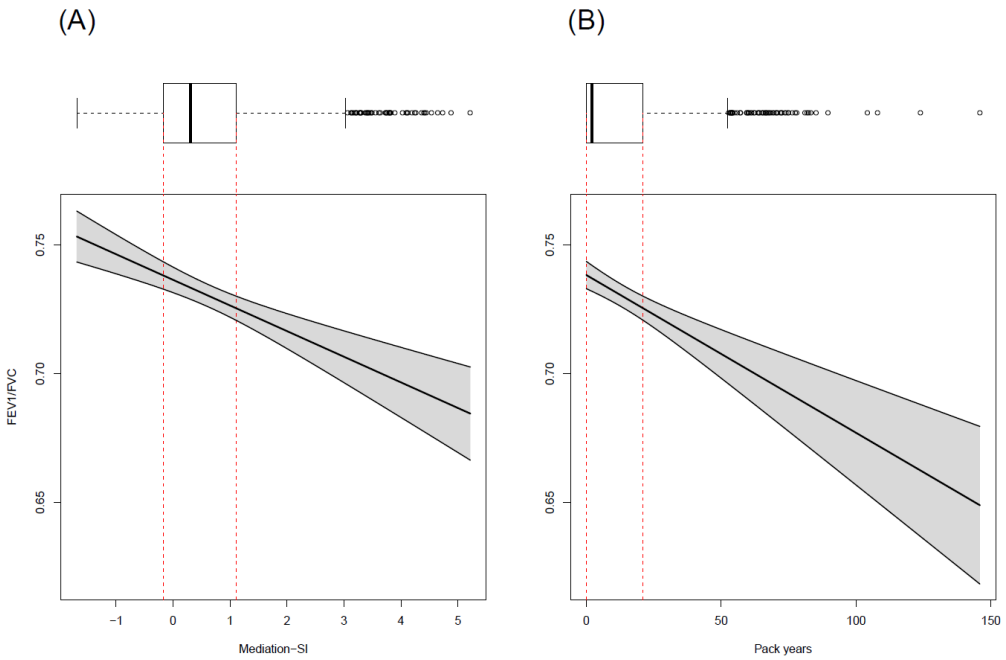
figure 4: Forest plots of cohort-specific results and meta-analyses of the association of the Mediation Smoking Index with FEV_1/FVC and change in FEV_1/FVC in ever - and never smokers in the discovery cohorts. Associations run applying base model adjustment (M_{base}*).



Footnote to figure 4:

*Base model (M_{base}): EWAS were covariate adjusted for age, age squared, height, squared deviation from the mean of height, sex and interaction terms of age, age squared, height and squared deviation of height with sex, education (low, medium, high), body mass index, spirometer type, study center, and cell composition. Prediction models were additionally adjusted for FEV_1/FVC at time point 1.

figure 5: Distribution and association* of Mediation smoking index (SI)† and self-reported smoking history (packyears) with FEV1/FVC, with 95% confidence interval. Dotted lines mark boxplot interquartile range borders. (A) Boxplot of Mediation-SI (median: 0.3 and range: -1.7 to 5.2) in all participants of SAPALDIA. (B) Boxplot of packyears (median: 2.0 and range: 0 to 145.9) in all participants of SAPALDIA. Analogous figures for association of Mediation-SI with FEV1 and with FVC, see online supplement figure S6 and figure S7.



Footnote to figure 5:

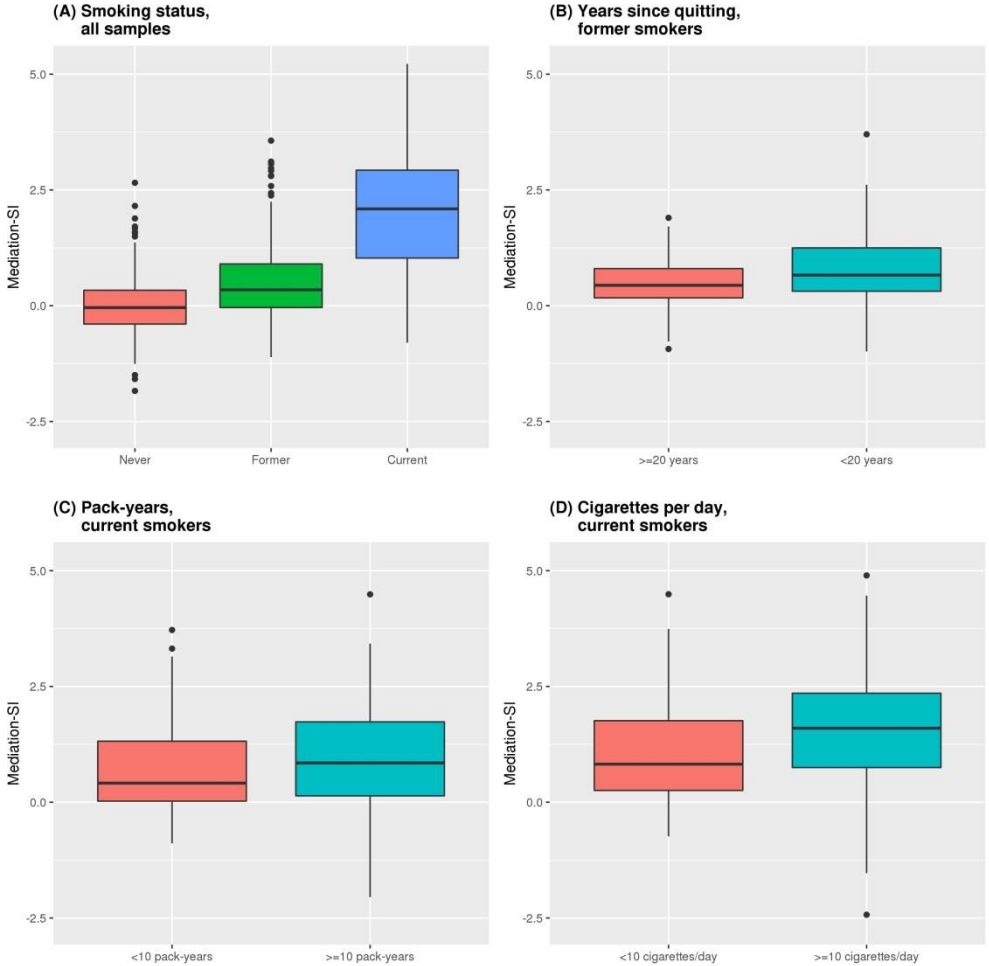
*Associations were adjusted for the base model (M_{base}): age, age squared, height, squared deviation from the mean of height, sex and interaction terms of age, age squared, height and squared deviation of height with sex, education (low, medium, high), body mass index, spirometer type, study center, and cell composition.

†The Mediation-SI can be constructed for all participants irrespective of their smoking status.

The M_{base}-adjusted model explained 17.5% of the variance in the outcome. The M_{base}-adjusted model additionally adjusted for the Mediation-SI explained 19.6% of the FEV1/FVC variance (total adjusted R^2 = 0.196) of which 2.8% of the variance was specifically explained by the SI variable. This was comparable to the variance explained by the M_{base}-adjusted model additionally

adjusted for packyears and smoking status corresponding to the M_{smok} model (R^2 = 0.198, and with 1.6% of the variance specifically explained by the packyears variable). Model including both smoking adjustments (M_{smok} and additionally Mediation-SI) explained 20.1% of the FEV1/FVC variance.

figure 6: Distribution of adjusted* Mediation smoking index (Mediation-SI) in SAPALDIA at time point 2.



Footnote to figure 6:

* Mediation-SI were adjusted as follows: A) adjusted for age, sex, and education; never smokers (n=395); former smokers (n=356); current smokers (n=211); B) adjusted for age, sex, education, packyears, and cigday; former smokers (n=356); C) adjusted for age, sex, education, and cigday; current smokers n=211); D) adjusted for age, sex, education, and packyears; current smokers n=211).

References

1. Klimentidis YC, Vazquez AI, de Los Campos G, Allison DB, Dransfield MT, Thannickal VJ. Heritability of pulmonary function estimated from pedigree and whole-genome markers. *Front Genet* 2013: 4: 174.
2. Joehanes R, Just AC, Marioni RE, Pilling LC, Reynolds LM, Mandaviya PR, Guan W, Xu T, Elks CE, Aslibekyan S, Moreno-Macias H, Smith JA, Brody JA, Dhingra R, Yousefi P, Pankow JS, Kunze S, Shah SH, McRae AF, Lohman K, Sha J, Absher DM, Ferrucci L, Zhao W, Demerath EW, Bressler J, Grove ML, Huan T, Liu C, Mendelson MM, Yao C, Kiel DP, Peters A, Wang-Sattler R, Visscher PM, Wray NR, Starr JM, Ding J, Rodriguez CJ, Wareham NJ, Irvin MR, Zhi D, Barrdahl M, Vineis P, Ambatipudi S, Uitterlinden AG, Hofman A, Schwartz J, Colicino E, Hou L, Vokonas PS, Hernandez DG, Singleton AB, Bandinelli S, Turner ST, Ware EB, Smith AK, Klengel T, Binder EB, Psaty BM, Taylor KD, Gharib SA, Swenson BR, Liang L, DeMeo DL, O'Connor GT, Herceg Z, Ressler KJ, Conneely KN, Sotoodehnia N, Kardia SL, Melzer D, Baccarelli AA, van Meurs JB, Romieu I, Arnett DK, Ong KK, Liu Y, Waldenberger M, Deary IJ, Fornage M, Levy D, London SJ. Epigenetic Signatures of Cigarette Smoking. *Circ Cardiovasc Genet* 2016: 9(5): 436-447.
3. Gao X, Jia M, Zhang Y, Breitling LP, Brenner H. DNA methylation changes of whole blood cells in response to active smoking exposure in adults: a systematic review of DNA methylation studies. *Clin Epigenetics* 2015: 7: 113.
4. de Vries M, van der Plaat DA, Nedeljkovic I, Verkaik-Schakel RN, Kooistra W, Amin N, van Duijn CM, Brandsma CA, van Diemen CC, Vonk JM, Marike Boezen H. From blood to lung tissue: effect of cigarette smoke on DNA methylation and lung function. *Respiratory research* 2018: 19(1): 212.
5. Machin M, Amaral AF, Wielscher M, Rezwan FI, Imboden M, Jarvelin MR, Adcock IM, Probst-Hensch N, Holloway JW, Jarvis DL, study A. Systematic review of lung function and COPD with peripheral blood DNA methylation in population based studies. *BMC Pulm Med* 2017: 17(1): 54.
6. Carmona JJ, Barfield RT, Panni T, Nwanaji-Enwerem JC, Just AC, Hutchinson JN, Colicino E, Karrasch S, Wahl S, Kunze S, Jafari N, Zheng Y, Hou L, DeMeo DL, Litonjua AA, Vokonas PS, Peters A, Lin X, Schwartz J, Schulz H, Baccarelli AA. Metastable DNA methylation sites associated with longitudinal lung function decline and aging in humans: an epigenome-wide study in the NAS and KORA cohorts. *Epigenetics* 2018: 1-17.
7. Kodal JB, Kobylecki CJ, Vedel-Krogh S, Nordestgaard BG, Bojesen SE. AHRR hypomethylation, lung function, lung function decline and respiratory symptoms. *The European respiratory journal : official journal of the European Society for Clinical Respiratory Physiology* 2018: 51(3).
8. 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.
9. Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD, Irizarry RA. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics* 2014: 30(10): 1363-1369.
10. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* 2010: 26(17): 2190-2191.
11. Lee MK, Hong Y, Kim SY, Kim WJ, London SJ. Epigenome-wide association study of chronic obstructive pulmonary disease and lung function in Koreans. *Epigenomics* 2017: 9(7): 971-984.
12. Morrow JD, Cho MH, Hersh CP, Pinto-Plata V, Celli B, Marchetti N, Criner G, Bueno R, Washko G, Glass K, Choi AM, Quackenbush J, Silverman EK, DeMeo DL. DNA methylation profiling in human lung tissue identifies genes associated with COPD. *Epigenetics* 2016: 1-10.
13. Zinellu A, Sotgiu E, Fois AG, Zinellu E, Sotgia S, Ena S, Mangoni AA, Carru C, Pirina P. Blood global DNA methylation is decreased in non-severe chronic obstructive pulmonary disease (COPD) patients. *Pulm Pharmacol Ther* 2017: 46: 11-15.

14. Gaunt TR, Shihab HA, Hemani G, Min JL, Woodward G, Lyttleton O, Zheng J, Duggirala A, McArdle WL, Ho K, Ring SM, Evans DM, Davey Smith G, Relton CL. Systematic identification of genetic influences on methylation across the human life course. *Genome Biol* 2016: 17: 61.
15. Hemani G, Zheng J, Elsworth B, Wade KH, Haberland V, Baird D, Laurin C, Burgess S, Bowden J, Langdon R, Tan VY, Yarmolinsky J, Shihab HA, Timpson NJ, Evans DM, Relton C, Martin RM, Davey Smith G, Gaunt TR, Haycock PC. The MR-Base platform supports systematic causal inference across the human phenome. *Elife* 2018: 7.
16. Stover PJ, James WPT, Krook A, Garza C. Emerging concepts on the role of epigenetics in the relationships between nutrition and health. *J Intern Med* 2018.
17. Franceschi C, Garagnani P, Morsiani C, Conte M, Santoro A, Grignolio A, Monti D, Capri M, Salvioli S. The Continuum of Aging and Age-Related Diseases: Common Mechanisms but Different Rates. *Front Med (Lausanne)* 2018: 5: 61.
18. Bojesen SE, Timpson N, Relton C, Davey Smith G, Nordestgaard BG. AHRR (cg05575921) hypomethylation marks smoking behaviour, morbidity and mortality. *Thorax* 2017: 72(7): 646-653.
19. Andersen AM, Philibert RA, Gibbons FX, Simons RL, Long J. Accuracy and utility of an epigenetic biomarker for smoking in populations with varying rates of false self-report. *American journal of medical genetics Part B, Neuropsychiatric genetics : the official publication of the International Society of Psychiatric Genetics* 2017: 174(6): 641-650.
20. Zaghlool SB, Mook-Kanamori DO, Kader S, Stephan N, Halama A, Engelke R, Sarwath H, Al-Dous EK, Mohamoud YA, Roemisch-Margl W, Adamski J, Kastenmuller G, Friedrich N, Visconti A, Tsai PC, Spector T, Bell J, Falchi M, Wahl A, Waldenberger M, Peters A, Gieger C, Pezer M, Lauc G, Graumann J, Malek JA, Suhre K. Deep molecular phenotypes link complex disorders and physiological insult to CpG methylation. *Human molecular genetics* 2018.
21. Teschendorff AE, Yang Z, Wong A, Pipinikas CP, Jiao Y, Jones A, Anjum S, Hardy R, Salvesen HB, Thirlwell C, Janes SM, Kuh D, Widschwendter M. Correlation of Smoking-Associated DNA Methylation Changes in Buccal Cells With DNA Methylation Changes in Epithelial Cancer. *JAMA Oncol* 2015: 1(4): 476-485.
22. Gao X, Zhang Y, Breitling LP, Brenner H. Relationship of tobacco smoking and smoking-related DNA methylation with epigenetic age acceleration. *Oncotarget* 2016: 7(30): 46878-46889.
23. Monick MM, Beach SR, Plume J, Sears R, Gerrard M, Brody GH, Philibert RA. Coordinated changes in AHRR methylation in lymphoblasts and pulmonary macrophages from smokers. *American journal of medical genetics Part B, Neuropsychiatric genetics : the official publication of the International Society of Psychiatric Genetics* 2012: 159B(2): 141-151.
24. Knopik VS, Marceau K, Bidwell LC, Rolan E. Prenatal substance exposure and offspring development: Does DNA methylation play a role? *Neurotoxicol Teratol* 2018.
25. Nielsen CH, Larsen A, Nielsen AL. DNA methylation alterations in response to prenatal exposure of maternal cigarette smoking: A persistent epigenetic impact on health from maternal lifestyle? *Arch Toxicol* 2016: 90(2): 231-245.
26. Joubert BR, Felix JF, Yousefi P, Bakulski KM, Just AC, Breton C, Reese SE, Markunas CA, Richmond RC, Xu CJ, Kupers LK, Oh SS, Hoyo C, Gruzieva O, Soderhall C, Salas LA, Baiz N, Zhang H, Lepeule J, Ruiz C, Ligthart S, Wang T, Taylor JA, Duijts L, Sharp GC, Jankipersadsing SA, Nilsen RM, Vaez A, Fallin MD, Hu D, Litonjua AA, Fuemmeler BF, Huen K, Kere J, Kull I, Munthe-Kaas MC, Gehring U, Bustamante M, Saurel-Coubizolles MJ, Quraishi BM, Ren J, Tost J, Gonzalez JR, Peters MJ, Haberg SE, Xu Z, van Meurs JB, Gaunt TR, Kerkhof M, Corpeleijn E, Feinberg AP, Eng C, Baccarelli AA, Benjamin Neelon SE, Bradman A, Merid SK, Bergstrom A, Herceg Z, Hernandez-Vargas H, Brunekreef B, Pinart M, Heude B, Ewart S, Yao J, Lemonnier N, Franco OH, Wu MC, Hofman A, McArdle W, Van der Vlies P, Falahi F, Gillman MW, Barcellos LF, Kumar A, Wickman M, Guerra S, Charles MA, Holloway J, Auffray C, Tiemeier HW, Smith GD, Postma D, Hivert MF, Eskenazi B, Vrijheid M, Arshad H, Anto JM, Dehghan A, Karmaus W, Annesi-Maesano I, Sunyer J, Ghantous A, Pershagen G, Holland N, Murphy SK, DeMeo DL, Burchard EG, Ladd-Acosta C, Snieder H, Nystad W, Koppelman GH, Relton

CL, Jaddoe VW, Wilcox A, Melen E, London SJ. DNA Methylation in Newborns and Maternal Smoking in Pregnancy: Genome-wide Consortium Meta-analysis. *American journal of human genetics* 2016: 98(4): 680-696.
27. Richmond RC, Joubert BR. Contrasting the effects of intra-uterine smoking and one-carbon micronutrient exposures on offspring DNA methylation. *Epigenomics* 2017: 9(3): 351-367.
28. Reynolds LM, Magid HS, Chi GC, Lohman K, Barr RG, Kaufman JD, Hoeschele I, Blaha MJ, Navas-Acien A, Liu Y. Secondhand Tobacco Smoke Exposure Associations With DNA Methylation of the Aryl Hydrocarbon Receptor Repressor. *Nicotine Tob Res* 2017: 19(4): 442-451.
29. de Lichtenfels AJ, van der Plaat DA, de Jong K, van Diemen CC, Postma DS, Nedeljkovic I, van Duijn CM, Amin N, la Bastide-van Gemert S, de Vries M, Ward-Caviness CK, Wolf K, Waldenberger M, Peters A, Stolk RP, Brunekreef B, Boezen HM, Vonk JM. Long-term Air Pollution Exposure, Genome-wide DNA Methylation and Lung Function in the LifeLines Cohort Study. *Environmental health perspectives* 2018: 126(2): 027004.
30. Hesselbach K, Kim GJ, Flemming S, Haupl T, Bonin M, Dornhof R, Gunther S, Merfort I, Humar M. Disease relevant modifications of the methylome and transcriptome by particulate matter (PM2.5) from biomass combustion. *Epigenetics* 2017: 12(9): 779-792.
31. Stueve TR, Li WQ, Shi J, Marconett CN, Zhang T, Yang C, Mullen D, Yan C, Wheeler W, Hua X, Zhou B, Borok Z, Caporaso NE, Pesatori AC, Duan J, Laird-Offringa IA, Landi MT. Epigenome-wide analysis of DNA methylation in lung tissue shows concordance with blood studies and identifies tobacco smoke-inducible enhancers. *Human molecular genetics* 2017: 26(15): 3014-3027.
32. Yao C, Joehanes R, Johnson AD, Huan T, Liu C, Freedman JE, Munson PJ, Hill DE, Vidal M, Levy D. Dynamic Role of trans Regulation of Gene Expression in Relation to Complex Traits. *American journal of human genetics* 2017: 100(4): 571-580.
33. Tingley D. YT, Hirose K., Keele L., Imai K. Mediation: R package for causal mediation analysis. *Journal of Statistical Software* 2014: 59(5): 1-38.