

Early View

Original article

Epigenome-wide association study on diffusing capacity of the lung (DLCO)

Natalie Terzikhan, Hanfei Xu, Ahmed Edris, Ken R. Bracke, Fien M. Verhamme, Bruno H. C. Stricker, Josée Dupuis, Lies Lahousse, George T. O'Connor, Guy G. Brusselle

Please cite this article as: Terzikhan N, Xu H, Edris A, *et al.* Epigenome-wide association study on diffusing capacity of the lung (DLCO). *ERJ Open Res* 2020; in press (<https://doi.org/10.1183/23120541.00567-2020>).

This manuscript has recently been accepted for publication in the *ERJ Open Research*. It is published here in its accepted form prior to copyediting and typesetting by our production team. After these production processes are complete and the authors have approved the resulting proofs, the article will move to the latest issue of the ERJOR online.

Copyright ©ERS 2020. This article is open access and distributed under the terms of the Creative Commons Attribution Non-Commercial Licence 4.0.

Epigenome-wide association study on diffusing capacity of the lung (DLCO)

Natalie Terzikhan^{*1,2}, Hanfei Xu^{*3}, Ahmed Edris^{2,5}, Ken R. Bracke¹, Fien M. Verhamme¹, Bruno H. C. Stricker², Josée Dupuis^{*3}, Lies Lahousse^{*2,5}, George T. O'Connor^{*6}, Guy G. Brusselle^{*1,2,7}

¹Department of Respiratory Medicine, Ghent University Hospital, Ghent, Belgium.

²Department of Epidemiology, Erasmus MC – University Medical Center Rotterdam, Rotterdam, the Netherlands.

³Department of Biostatistics, Boston University School of Public Health, Boston, Massachusetts, USA.

⁵Department of Bioanalysis, Faculty of Pharmaceutical Sciences, Ghent University, Ghent, Belgium.

⁶Pulmonary Center, Boston University Schools of Medicine and Public Health, Boston, Massachusetts, USA.

⁷Department of Respiratory Medicine, Erasmus MC – University Medical Centre Rotterdam, Rotterdam, The Netherlands.

Correspondence:

Guy G. O. Brusselle, MD, PhD, FERS

Department of Respiratory Medicine, Ghent University Hospital, B-9000 Ghent, Belgium

Email: guy.brusselle@ugent.be

* These authors contributed equally to this work.

Funding

This work was supported by a grant from the Fund for Scientific Research Flanders (FWO) project [G035014N]. FMV is a post-doctoral fellow of the Fund for Scientific Research–Flanders (FWO). The Rotterdam Study is funded by Erasmus MC and Erasmus University Rotterdam; the Netherlands Organisation for the Health Research and Development (ZonMw); the Research Institute for Diseases in the Elderly (RIDE); the Ministry of Education, Culture and Science; the Ministry for Health, Welfare and Sports; the European Commission (DG XII); and the Municipality of Rotterdam.

The generation and management of the Illumina 450K methylation array data (EWAS data) for the Rotterdam Study was executed by the Human Genotyping Facility of the Genetic Laboratory of the Department of Internal Medicine, Erasmus MC, the Netherlands. The EWAS data was funded by the Genetic Laboratory of the Department of Internal Medicine, Erasmus MC, and by the Netherlands Organization for Scientific Research (NWO; project number 184021007) and made available as a Rainbow Project (RP3; BIOS) of the Biobanking and Biomolecular Research Infrastructure Netherlands (BBMRI-NL). We thank Mr. Michael Verbiest, Ms. Mila Jhamai, Ms. Sarah Higgins, Mr. Marijn Verkerk, and Lisette Stolk PhD for their help in creating the methylation database.

The Framingham Heart Study is funded by National Institutes of Health contract N01-HC-25195 and HHSN268201500001I. The laboratory work for this investigation was funded by the Division of Intramural Research, National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, MD, and the NIH Director's Challenge Award (PI: D. Levy).

Ethical statement

The Rotterdam Study has been approved by the Medical Ethics Committee of the Erasmus MC (registration number MEC 02.1015) and by the Dutch Ministry of Health, Welfare and Sport (Population Screening Act WBO, license number 1071272-159521-PG). The Rotterdam Study has been entered into the Netherlands National Trial Register (NTR; www.trialregister.nl) and into the WHO International Clinical Trials Registry Platform (ICTRP; www.who.int/ictip/network/primary/en/) under shared catalogue number NTR6831. All participants provided written informed consent to participate in the study and to have their information obtained from treating physicians.

Abstract

Background: Epigenetics may play an important role in pathogenesis of lung diseases. However, little is known about the epigenetic factors that influence impaired gas exchange at the lungs.

Aim: To identify the epigenetic signatures of the diffusing capacity of the lung measured by carbon monoxide uptake.

Methods: Epigenome-Wide Association Study (EWAS) was performed on diffusing capacity, measured by carbon monoxide uptake (D_{LCO}) and per alveolar volume (D_{LCO} / V_A) using the single-breath technique in 2,674 individuals from two population-based cohort studies, the Rotterdam Study (the discovery panel) and the Framingham Heart Study (the replication panel). We assessed the clinical relevance of our findings by investigating the identified sites in whole blood and lung tissue specific gene expression.

Results: We identified and replicated two CpG sites (cg05575921 and cg05951221) that were significantly associated with D_{LCO} / V_A and one (cg05575921) suggestively associated with D_{LCO} . Furthermore, we found a positive association between *AHRR* (cg05575921) hypomethylation and gene expression of *EXOC3* in whole blood. We confirmed that the expression of *EXOC3* in lung tissue is positively associated with D_{LCO} / V_A and D_{LCO} .

Conclusions: We report on epigenome wide associations with diffusing capacity in the general population. Our results suggest *EXOC3* to be an excellent candidate through which smoking induced hypomethylation of *AHRR* might affect pulmonary gas exchange.

Introduction

Tests for diffusing capacity of the lung for carbon monoxide (D_{LCO} and D_{LCO} per alveolar volume (D_{LCO} / V_A)) provide a quantitative measure of gas exchange in the lung. In addition to its utility in the diagnosis and monitoring of lung diseases such as emphysema and pulmonary fibrosis, measure of gas exchange is an independent predictor of mortality in COPD patients.(1) D_{LCO} and D_{LCO} / V_A measurements are influenced by environmental factors.(2) Exposure to environmental factors such as smoking, occupation related compounds and air pollution, decreases the level of gas exchange in the lungs.(3)

While D_{LCO} and D_{LCO} / V_A are heritable traits, the fraction of variation in these lung function tests unexplained by genetic variation suggests an important role for environmental interactions.(4) Although genetics play an important role, epigenetic mechanisms such as DNA methylation are important for regulation of gene expression but also essential to understand the interplay between genes, environment and disease.(5, 6) Like D_{LCO} and D_{LCO} / V_A , DNA methylation is also influenced by smoking (7, 8); however, the question remains whether DNA methylation is associated with gas exchange independently of smoking.

Little is known about the effects of DNA methylation on lung function (9-11) and epigenome wide association studies (EWAS) of pulmonary gas exchange are lacking. Therefore, we aimed to perform an EWAS to investigate which epigenetic variants are related to the phenotypic variation in D_{LCO} and D_{LCO} / V_A . In addition, we examined the impact of the discovered epigenetic variants on gene expression in the lungs.

Methods

Study populations

This EWAS study encompassed a discovery study embedded within the Rotterdam Study and a replication study embedded within the Framingham Heart Study. The Rotterdam Study is an ongoing prospective population-based cohort study in Rotterdam, the Netherlands. The design has been previously described.^(12, 13) Briefly, the Rotterdam Study includes four sub-cohorts. For this study, data from the third visit of the second sub-cohort (RSII-3) and the second visit of the third sub-cohort (RSIII-2) was used. The discovery panel consisted of 659 participants from a random subset of 747 individuals of European descent with methylation and diffusing capacity data available.

Replication of the identified CpG sites was performed in 2,114 individuals from the Framingham Heart Study Offspring cohort with methylation and diffusing capacity data was available. The design of the Framingham Heart Offspring Study has been described extensively before.⁽¹⁴⁾ The replication analyses were focused on Offspring cohort participants of European descent who attended the eighth exam (2005-2008). In both cohorts, we only included participants over 45 years old with blood cell counts and batch effects data available.

DNA methylation

DNA was extracted from whole peripheral blood (stored in EDTA tubes) by standardised salting out methods. Genome-wide DNA methylation levels were measured using the Illumina Human Methylation 450K array (Illumina, San Diego, CA, USA). In summary, samples (500 ng of DNA per sample) were first treated with bisulfite using the Zymo EZ-96 DNA-methylation kit (Zymo Research, Irvine, CA, USA). Subsequently, samples were hybridised to the arrays according to the

standardized protocols. The methylation percentage of a CpG site was presented as a β value ranging between 0 (no methylation) to 1 (full methylation). Data processing was performed in the Genetic Laboratory of Internal Medicine, Erasmus University Medical Centre, Rotterdam. Quality control was performed using Genome Studio (v2011.1, methylation module version 1.9.0; Illumina, USA). For each probe, individuals with methylation levels higher than three times the inter-quartile range (IQR) were excluded. In both studies, cross reactive and polymorphic probes, or probes that have an underlying SNP at the CpG site were excluded, in addition to probes within 10 bp of the single base extension (minor allele frequency >1% in European ancestry (EUR) 1000 genomes project data).(15) Additionally, probes with a missing rate >20% at $p < 0.01$ were excluded in the Framingham cohort. In total, 363,387 CpGs were included.

Diffusing capacity of the lung: D_{LCO} and D_{LCO} / V_A

D_{LCO} (mmol/min/kPa) and alveolar volume (V_A) in litre were measured by the single breath technique **using a Master Screen® PFT Pro (CareFusion, San Diego, CA)** in accordance with ERS / ATS guidelines.(16) The D_{LCO} per alveolar volume (D_{LCO} / V_A ; mmol/min/kPa/litre) was calculated by dividing the D_{LCO} by V_A . Analyses were restricted to participants with two interpretable and reproducible measurements of D_{LCO} and D_{LCO} / V_A . Two measurements were considered reproducible if the difference between the first and second D_{LCO} measurement ((highest-lowest value)/highest value) was $\leq 10\%$ and the difference between the first and second D_{LCO} / V_A measurement was $\leq 15\%$.

Covariates

Covariates were selected based on known association with DNA methylation and diffusing capacity and included age, sex, smoking status, weight, height and batch

effects; array number and position. **Smoking status was obtained using a questionnaire. Cigarette pack years were computed as duration of smoking (years) multiplied by the number of smoked cigarettes, divided by 20.**

Functional analysis

We used data from the BBMRI atlas (17) to identify methylation-gene expression associations; the so called expression quantitative trait methylation (eQTM). **The BBMRI atlas includes all association studies performed by the BBMRI-NL consortium. We queried the database (<http://bbmri.researchlumc.nl/atlas/#query>) for association between the significant CpGs and differential gene expression and downloaded the original data tables (<http://bbmri.researchlumc.nl/atlas/#data>) to learn more about the direction of the estimated effects.**

Gene Expression in lung tissue

Lung resection specimens were obtained from 92 patients, of which 78 from surgery for solitary pulmonary tumours (Ghent University Hospital, Ghent, Belgium) and 14 from explant lungs of end-stage COPD patients undergoing lung transplantation (University Hospital Gasthuisberg, Leuven, Belgium). Lung tissue at maximum distance from the pulmonary lesions and without signs of retro-obstructive pneumonia or tumour invasion, was collected by a pathologist. None of the patients operated for malignancy were treated with neo-adjuvant chemotherapy. Written informed consent was obtained from all subjects. This study was approved by the medical ethical committees of the Ghent University Hospital (2011/14) and the University Hospital Gasthuisberg Leuven (S51577).

RNA was extracted with the miRNeasy Mini kit (Qiagen) from total lung tissue blocks submersed in RNA-later. cDNA was obtained by the miScript II RT kit (Qiagen), following manufacturer's instructions. Expression of target genes ADGRG6 (GPR126) and reference genes Glyceraldehyde-3-phosphate dehydrogenase (GAPDH), Hypoxanthine phosphoribosyltransferase-1 (HPRT-1) and Succinate Dehydrogenase Complex Flavoprotein Subunit A (SDHA) were analyzed using Taqman Gene Expression Assays (Applied Biosystems, Forster City, CA, USA). Real-time PCR reactions were set up in duplicate using diluted cDNA using identical amplification conditions for each of the target and reference genes. A standard curve derived from serial dilutions of a mixture of all samples were included in each run. The amplification conditions consisted of: 10 minutes at 95°C and 60 cycles of 95°C for 10 seconds and 60°C for 15 seconds. Amplifications were performed using a LightCycler 96 detection system (Roche). Data were processed using the standard curve method. Expression of target genes was corrected by a normalization factor that was calculated based on the expression of three reference genes, using the geNorm applet according to the guidelines and theoretical framework previously described (18).

Statistical analysis

Methylation probes were tested for association with D_{LCO} or D_{LCO} / V_A using a linear regression model in the Rotterdam Study and linear mixed model in the Framingham Heart Study (with a random effect to account for familial relationships in the Framingham cohort). Surrogate variable analysis was used to adjust for inflation in the effect estimates by batch effect in the Framingham Heart Study. The first model (Model 1) included age, sex, **current smoking, former smoking**, white blood cells, array number and position. In the second model (Model 2), we additionally adjusted

for weight and height. **Model 2 was used for all sensitivity analyses.** False discovery rate of 5% was used to correct for multiple testing (FDR < 0.05 was considered statistically significant). We tested the association between mRNA expression of *AHRR* and *EXOC3* genes and gas exchange in a linear regression adjusting for all covariates used in Model 2. For these analyses, outliers (mean ± 3 SD) were excluded and mRNA levels were log transformed to make sure the normality assumption was fulfilled. A p-value lower than 0.05 was considered statistically significant for the gene expression analysis.

Results

General characteristics

The general characteristics of the study populations are shown in **Table 1**. The mean age (SD) was 67.4 (5.9) in the Rotterdam Study and 65.6 (8.4) in the Framingham Heart Study. Mean levels of D_{LCO} and D_{LCO} / V_A were similar in both study populations.

Discovery phase

Data from the Rotterdam Study was used in the discovery phase of both Epigenome-wide association studies (EWAS) of D_{LCO} and D_{LCO} / V_A .

EWAS of D_{LCO}

No statistically significant associations were found between DNA methylation and D_{LCO} in analyses adjusted for age, sex, smoking status, white blood cell counts and batch effects (model 1), nor in analyses additionally adjusted for weight and height (model 2). However, cg05575921 (Gene: *AHRR*, chromosome: 5) was suggestively associated with D_{LCO} in both models ($FDR_{model1}=0.07$, $FDR_{model2}= 0.09$).

EWAS of D_{LCO} / V_A

In model 1, we identified one CpG site associated with D_{LCO} / V_A ; cg05575921 (Gene: *AHRR*, chromosome: 5, FDR: 0.017). In model 2, we identified two CpG sites associated with D_{LCO} / V_A ; cg05575921 (Gene: *AHRR*, chromosome: 5, FDR: 0.047) and cg05951221 (Closest gene: *ALPG*, chromosome: 2, FDR: 0.047). Sensitivity analyses of the top associations by additionally adjusting for smoking related CpGs, FEV_1/FVC or in never smokers are presented in the supplement.

Replication phase

We replicated our findings using data from the Framingham Heart Study. The statistically significant associations between DNA methylation and DLCO/ V_A are summarized in **Table 2**.

Functional analysis

We used BBMRI-atlas to identify eQTM using the identified D_{LCO} / V_A associated CpG sites (cg05575921 and cg05951221). **Cg05575921 (AHRR gene region) was significantly associated with the expression of the EXOC3 gene ($\beta= 0.15$ (SE=0.039), FDR= 3.9E-4) and no eQTM was retrieved between AHRR and AHRR expression. No eQTM was retrieved for cg05951221.**

Sensitivity analyses

Although we adjusted for smoking in our analyses, residual confounding might still play a role. To further disentangle the potential role of smoking, we performed two sensitivity analyses:

First, we adjusted the analyses of the lead associations for well-known smoking CpGs with the highest explained variance according to the publication by Zeilinger et al.(8) We ended up adjusting the analyses of the lead associations for each other and for another CpG (cg06126421 - chr 6).

Second, we performed the analysis in a subset of 207 non-smokers. The results of the sensitivity analyses are presented in supplementary tables 6, 7 and 8. The results suggest that smoking partially explains the observed associations, supporting both scenarios; 1) the effect of smoking on diffusing

capacity is mediated by DNA methylation, and 2) the potential for additional smoking-independent effects.

Gene Expression in lung tissue

Based on the results of the previous section, we were interested in the association between EXOC3 and D_{LCO} or D_{LCO} / V_A independent of smoking. Because AHRR is strongly associated with smoking, we also investigated the association between AHRR and D_{LCO} or D_{LCO} / V_A after adjustment for smoking. Therefore, mRNA was extracted from lung resection specimens of 92 patients who underwent surgery for lung transplantation or solitary pulmonary tumours, including 44 patients without COPD and 48 patients with COPD; and mRNA expression of AHRR and EXOC3 in lung tissue was examined using quantitative RT-PCR. See Table 3 for the general characteristics of this study population. Two outliers were excluded for AHRR, and four for EXOC3.

Figure 1 shows the expression levels of AHRR and EXOC3 genes by smoking and COPD status. mRNA levels of AHRR were significantly higher in current smokers compared to never or former smokers, while mRNA levels of EXOC3 were not significantly different between groups.

Regression analysis revealed no significant association between mRNA expression of *AHRR* and D_{LCO} or D_{LCO} / V_A after adjustment for age, sex and smoking status in Model 1, or after additional adjustment for weight and height in Model 2.

On the other hand, mRNA expression of *EXOC3* and D_{LCO} or D_{LCO} / V_A showed no statistically significant association in Model 1; however, the association became significant after additional adjustment for weight and height in Model 2 (see results in supplementary **Table S1**).

Discussion

In this EWAS study, we investigated the epigenetic signature of D_{LCO} and D_{LCO} / V_A as a measure of gas exchange in the lung. We observed two CpG sites with D_{LCO} / V_A epigenome-wide associations in the Rotterdam Study, and those were replicated in the Framingham Heart Study. Additional analysis revealed a potential pathway through which methylation of *AHRR* might influence gas exchange.

This is the first EWAS on diffusing capacity of carbon monoxide uptake in the lungs. Few studies investigated the effect of DNA methylation on lung function, and were restricted to spirometry measures (FEV_1 and FVC). Bell and colleagues (9) investigated age-related phenotypes including FEV_1 and FVC in 172 twin females. Authors reported one association at chromosome 11 in the *WT1* gene region. Another study by Marioni *et al.* (10) performed an EWAS on FEV_1 in 920 individuals and did not find epigenome-wide associations. Qui et al. performed EWAS on FEV_1 , FVC and COPD and found many associations, but the analyses were not adjusted for differences in white blood cell count.(11)

In this EWAS, we discovered and confirmed two epigenome-wide associations, one CpG site in the aryl hydrocarbon receptor repressor gene (*AHRR*) which encodes a repressor protein of the aryl hydrocarbon receptor (AHR), and the second CpG site on chromosome 2 (2q37.1). These CpG sites are well described as being strongly associated with smoking behaviour. (7, 8, 19) Cg05575921 is located in the *AHRR* gene region and is involved in the metabolism of smoking-released chemicals, where the *AHRR* gene suppresses the function of the AHR gene –which is responsible for the regulation of smoking related substances- through a negative feedback loop (20). Although the interaction between the genetics and DNA methylation is complex, it is believed that *AHRR* hypomethylation inhibits the translation of the gene by

preventing transcription factors from binding to the promotor regions.(21, 22) We hypothesise that *AHRR* hypomethylation by smoking might intervene with the elimination of the smoking related substances from the body.

In addition to the association with smoking, *AHRR* has been also recently associated with impaired lung function. Bojesen et al. (19) showed that hypomethylation of cg05575921 was associated with smoking related phenotypes such as COPD, COPD exacerbations and lung cancer. Similarly, Kodal et al. (23) found that hypomethylation of cg05575921 was associated with low lung function, steeper lung function decline and respiratory symptoms. Finally, **Imboden et al.(24) and Morrow et al.(25) confirmed the CpG association with smoking, and suggested its potential use as a biomarker for predicting smoking associated morbidity and mortality.**

A recent study by Li and colleagues, proved the causal effect of smoking on DNA methylation.(20) This information eliminates the possibility for smoking to be an intermediate in the association between DNA methylation of smoking related probes and lung function. Alternatively, smoking can be a confounding factor for which we can adjust in our models. However, residual confounding is still possible as assessment of smoking status is self-reported.(26) Therefore the question remains, does residual confounding by smoking explain the whole effect in the association between smoking related probes and lung function? Or is there any significant effect of DNA methylation of those smoking related probes independent of smoking? Both scenarios lead to very interesting hypotheses and might have important clinical implications. In case smoking explains the entire association, the hypothesis might be, as proposed by Li et al. (20) and Kodal et al. (23), that the association between

hypomethylation of smoking related probes and a decreased lung function might reflect the smoking induced damage to the lungs through methylation of the smoking related probes. In other words, DNA methylation might act as an intermediate in the effect of smoking on smoking related diseases. The second scenario would be, that smoking affects DNA methylation, but the effect of the hypomethylation of the smoking related probes on lung function is (partly) independent from smoking. In that case, hypomethylation of the smoking related probes might provide clinically relevant information beyond the effect of smoking on lung function. Unfortunately, it is difficult to investigate these hypotheses in cross-sectional observational studies. **Nevertheless, our sensitivity analyses by either additionally adjusting for smoking probes in the analysis, or performing the main analysis in never-smokers revealed interesting results; smoking explained only part of the association. We hypothesize that smoking as well as independent effects might both play a role at the same time in the epigenetic signature of diffusing capacity of the lungs.** However, finding any association in never smokers is still no proof that smoking related probes are independently associated with lung function, as passive smoking or the unmeasured effects of air pollution still might bias these associations.

Regardless of the role of smoking in the association between hypomethylation of *AHRR* and a decreased pulmonary gas exchange, we propose in this study a pathway through which *AHRR* might affect lung function. We observed that the hypomethylation of the cg05575921 site in *AHRR* is associated with decreased expression of *EXOC3* gene in whole blood. Subsequently, we were able to link a decreased expression of *EXOC3* in human lung tissue to a decreased gas exchange in the lungs.

Exocyst complex component 3 (EXOC3), previously known as SEC6, is located on chromosome 5 upstream of AHRR gene and downstream of SLC9A3 gene. EXOC3 is part of the exocyst protein complex and the protein encoded by it, is involved in post Golgi trafficking and essential for biogenesis of epithelial cell surface polarity.(27-29) Variation in EXOC3 is linked to variability in the clinical phenotype of subjects with cystic fibrosis.(30, 31) This makes EXOC3 an interesting candidate to elucidate its role in membrane pathology and gas exchange.

The strength of this study is the population-based setting with standardized data collection. Also, this study provides unique data, since we are not aware of other population-based studies with data on gas exchange. Finally, we provided additional results of gene expression analysis on lung tissue. However, this study also has some limitations. **First, we are aware that our discovery panel is relatively small, and that bigger sample sizes or joint analyses may help in the identification of more CpG sites that might give us additional insight in pathology of gas exchange. To ensure our study is sufficiently powered to replicate our findings, we opted to use the smaller cohort in the discovery phase, as advocated by Held and colleagues.(32)** Second, although we were able to replicate findings of the lead association, we cannot exclude the possibility of unmeasured confounding by smoking and air pollution. **Third, due to the difficulty of obtaining lung samples from healthy subjects, our gene expression analysis included samples from lung cancer and end-stage COPD patients, potentially affecting the gene expression profile.** Fourth, the current study analyses the data in a cross-sectional manner. As the epigenome changes over time, longitudinal data-analyses might be more informative about the epigenetic modifications over time, as it may play a crucial role in the aetiology of a decreased gas exchange. **Finally, our**

methylation analyses used blood samples, and therefore may not be fully representative of differential methylation in lungs as the target tissue. However, previous EWAS using lung samples demonstrated that smoking-induced differential methylation in the *AHRR* locus (cg05575921) was implicated in lung carcinogenesis, and suggested that blood samples could be used as surrogates for lung samples in some methylation analyses.(33)

In conclusion, impaired gas exchange is associated with smoking-related epigenetic changes. We propose a pathway through which *AHRR* hypomethylation might affect gas exchange.

Tables and Figures

Table 1 General characteristics of the discovery and validation study populations

	RS		FHS	
N total	659		2,114	
Age (yrs.), mean (SD)	67.4	(5.90)	65.6	(8.38)
Female, N (%)	369	(56%)	1159	(55%)
Current smokers, N (%)	70	(11%)	151	(7%)
Past smokers, N (%)	379	(58%)	1,186	(56%)
Weight (kg), mean (SD)	79.9	(13.94)	79.5	(17.83)
Height (cm), mean (SD)	169.8	(9.16)	167.3	(9.54)
FEV1 (L)	2.72	(0.71)	2.68	(0.75)
FEV1%predicted (%)	102.85	(22.82)	97.94	(16.94)
FVC (L)	3.56	(0.90)	3.68	(0.98)
FVC%predicted (%)	108.93	(16.92)	101.39	(15.09)
FEV1/FVC (%)	76.65	(7.33)	72.89	(6.86)
D _{LCO} (mmol/min/kPa), mean (SD)*	7.81	(1.64)	7.47	(2.13)
D _{LCO} / V _A (mmol/min/kPa/ V _A), mean (SD)*	1.52	(0.23)	1.42	(0.23)

DLCO: Diffusing capacity of the lung for carbon monoxide; D_{LCO} / V_A: Diffusing capacity of the lung for carbon monoxide by alveolar volume; FHS: The Framingham Heart Study; RS: The Rotterdam Study.

Values are means (standard deviation (SD)) for continuous variables or counts (percentages %) for dichotomous variables.*= **D_{LCO} was corrected for hemoglobin levels in the Rotterdam study. Interpretable spirometry was available in N=625 in the RS and N= 2114 in the FHS.**

Table 2 Epigenome-wide associations between genome-wide DNA-methylation and D_{LCO} / V_A

Model	CpG	Chr	Pos	Gene	Discovery cohort RS			Replication cohort FHS		
					B	SE	FDR	B	SE	FDR
M1	cg05575921	5	373378	AHRR	0.065	0.012	0.017	0.037	0.0055	8.72e-06
M2	cg05575921	5	373378	AHRR	0.064	0.012	0.047	0.040	0.0057	1.56e-06
	cg05951221	2	233284402	ALPG*	0.049	0.0093	0.047	0.030	.0051	3.60e-04

B: effect estimate; Chr: Chromosome; D_{LCO} / V_A : Diffusing capacity of the lung for carbon monoxide by alveolar volume; FDR: false discovery rate; FHS: The Framingham Heart Study; M1: Adjusted for age, sex, **current smoking, former smoking**, white blood cell count and batch effect; M2: adjusted for age, sex, current smoking, former smoking, weight, height, white blood cell count and batch effect; Pos: position; RS: Rotterdam Study; SE: Standard error.*= Closest gene(distance>8Kbp)

Table 3 Characteristics of study individuals for lung mRNA analysis (by RT-PCR)

(n=92)

	Never smokers	Smokers without COPD	COPD GOLD II	COPD GOLD III-IV
Number	18	26	34	14
Gender ratio (m/f)	6/12 §	19/7 §	31/3 §	8/6 §
Age (years)	65 (56-70)	63 (55-70)	66 (58-69) ‡	56 (54-60)* ‡ ‡
Current- / ex-smoker	NA	16/10	22/12	0/14
Pack years of smoking	NA	28 (15-45)*	45 (40-60)* ‡	30 (25-30)* ‡ ‡
D_{LCO}	21.58	23.30	17.22*	2.9* ‡ ‡
D_{LCO} %predicted	90.00	80.00	67.00*	35.00* ‡ ‡
D_{LCO} / V_A	4.64	3.89*	3.51*	0.87* ‡ ‡
D_{LCO} / V_A %predicted	102.85	91.00*	87.00*	58.50* ‡ ‡
FEV₁%predicted	102.00	95.00	68.00* ‡	25.50* ‡ ‡
Fev₁/FVC	78.00	75.00	56.00* ‡	32.00* ‡ ‡

Data are presented as median (IQR)

Mann-Whitney U test: * P < 0.05 versus never smokers; † P < 0.05 versus COPD GOLD II; ‡ P < 0.05

versus smokers without COPD; Fisher's exact test: § P < 0.001

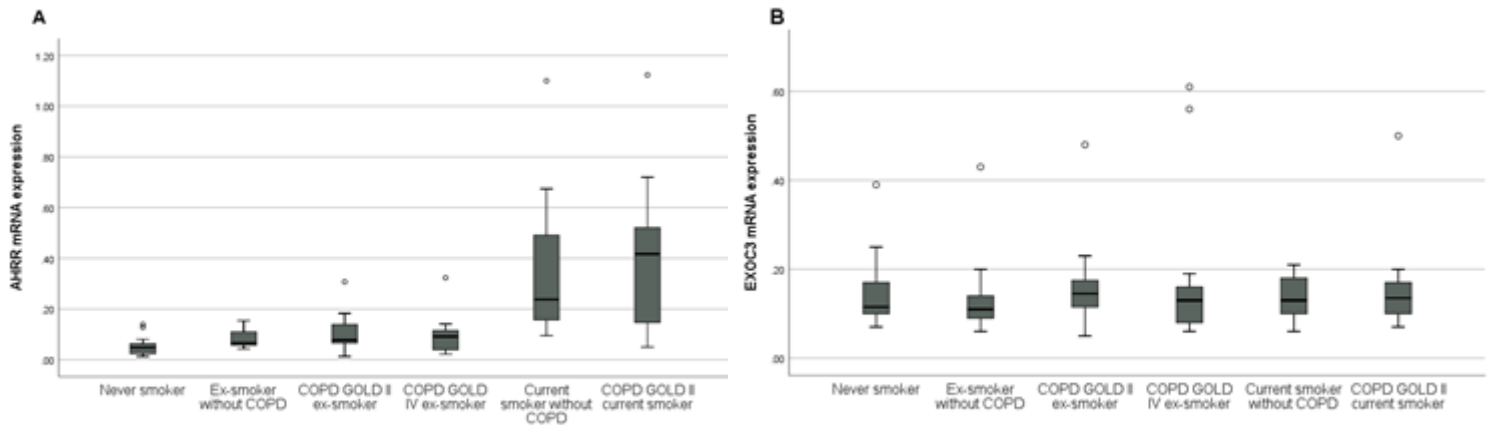


Figure 1 human lung tissue mRNA expression of the **A) AHRR** gene and **B) EXOC3** gene stratified by smoking and COPD status.

AHRR: Aryl Hydrocarbon Receptor Repressor gene; COPD: Chronic obstructive pulmonary disease; ex: ex-smoker; EXOC3: Exocyst Complex Component 3 gene. GOLD: Global initiative for chronic obstructive lung disease;

Acknowledgment

The Rotterdam Study is funded by Erasmus Medical Center and Erasmus University, Rotterdam, Netherlands Organization for the Health Research and Development (ZonMw), the Research Institute for Diseases in the Elderly (RIDE), the Ministry of Education, Culture and Science, the Ministry for Health, Welfare and Sports, the European Commission (DG XII), and the Municipality of Rotterdam. The authors are grateful to the study participants, the staff from the Rotterdam Study and the participating general practitioners and pharmacists. The Framingham Heart Study is conducted and supported by the National Heart, Lung, and Blood Institute (NHLBI) in collaboration with Boston University (Contract No. N01-HC-25195, HHSN268201500001I and 75N92019D00031).

References

1. Boutou AK, Shrikrishna D, Tanner RJ, Smith C, Kelly JL, Ward SP, et al. Lung function indices for predicting mortality in COPD. *Eur Respir J*. 2013;42(3):616-25.
2. Hegewald MJ. Diffusing capacity. *Clin Rev Allergy Immunol*. 2009;37(3):159-66.
3. Fricker M, Goggins BJ, Mateer S, Jones B, Kim RY, Gellatly SL, et al. Chronic cigarette smoke exposure induces systemic hypoxia that drives intestinal dysfunction. *JCI Insight*. 2018;3(3).
4. Terzikhan N, Sun F, Verhamme FM, Adams HHH, Loth D, Bracke KR, et al. Heritability and genome-wide association study of diffusing capacity of the lung. *Eur Respir J*. 2018;52(3).
5. Shah S, McRae AF, Marioni RE, Harris SE, Gibson J, Henders AK, et al. Genetic and environmental exposures constrain epigenetic drift over the human life course. *Genome Res*. 2014;24(11):1725-33.
6. Vercelli D. Genetics, epigenetics, and the environment: switching, buffering, releasing. *J Allergy Clin Immunol*. 2004;113(3):381-6; quiz 7.
7. Shenker NS, Polidoro S, van Veldhoven K, Sacerdote C, Ricceri F, Birrell MA, et al. Epigenome-wide association study in the European Prospective Investigation into Cancer and Nutrition (EPIC-Turin) identifies novel genetic loci associated with smoking. *Hum Mol Genet*. 2013;22(5):843-51.
8. Zeilinger S, Kuhnel B, Klopp N, Baurecht H, Kleinschmidt A, Gieger C, et al. Tobacco smoking leads to extensive genome-wide changes in DNA methylation. *PLoS One*. 2013;8(5):e63812.
9. Bell JT, Tsai PC, Yang TP, Pidsley R, Nisbet J, Glass D, et al. Epigenome-wide scans identify differentially methylated regions for age and age-related phenotypes in a healthy ageing population. *PLoS Genet*. 2012;8(4):e1002629.
10. Marioni RE, Shah S, McRae AF, Ritchie SJ, Muniz-Terrera G, Harris SE, et al. The epigenetic clock is correlated with physical and cognitive fitness in the Lothian Birth Cohort 1936. *Int J Epidemiol*. 2015;44(4):1388-96.
11. Qiu W, Baccarelli A, Carey VJ, Boutaoui N, Bacherman H, Klanderman B, et al. Variable DNA methylation is associated with chronic obstructive pulmonary disease and lung function. *Am J Respir Crit Care Med*. 2012;185(4):373-81.
12. Ikram MA, Brusselle GGO, Murad SD, van Duijn CM, Franco OH, Goedegebure A, et al. The Rotterdam Study: 2018 update on objectives, design and main results. *Eur J Epidemiol*. 2017;32(9):807-50.
13. Hofman A, Brusselle GG, Darwish Murad S, van Duijn CM, Franco OH, Goedegebure A, et al. The Rotterdam Study: 2016 objectives and design update. *Eur J Epidemiol*. 2015;30(8):661-708.
14. Kannel WB, Feinleib M, McNamara PM, Garrison RJ, Castelli WP. An investigation of coronary heart disease in families. The Framingham offspring study. *Am J Epidemiol*. 1979;110(3):281-90.
15. Chen YA, Lemire M, Choufani S, Butcher DT, Grafodatskaya D, Zanke BW, et al. Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray. *Epigenetics*. 2013;8(2):203-9.
16. Macintyre N, Crapo RO, Viegi G, Johnson DC, van der Grinten CP, Brusasco V, et al. Standardisation of the single-breath determination of carbon monoxide uptake in the lung. *Eur Respir J*. 2005;26(4):720-35.
17. Bonder MJ, Luijk R, Zhernakova DV, Moed M, Deelen P, Vermaat M, et al. Disease variants alter transcription factor levels and methylation of their binding sites. *Nat Genet*. 2017;49(1):131-8.

18. Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, De Paepe A, et al. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol.* 2002;3(7):RESEARCH0034.
19. Bojesen SE, Timpson N, Relton C, Davey Smith G, Nordestgaard BG. AHRR (cg05575921) hypomethylation marks smoking behaviour, morbidity and mortality. *Thorax.* 2017;72(7):646-53.
20. Li S, Wong EM, Bui M, Nguyen TL, Joo JE, Stone J, et al. Causal effect of smoking on DNA methylation in peripheral blood: a twin and family study. *Clin Epigenetics.* 2018;10:18.
21. Bird AP. CpG-rich islands and the function of DNA methylation. *Nature.* 1986;321(6067):209-13.
22. Machin M, Amaral AF, Wielscher M, Rezwan FI, Imboden M, Jarvelin MR, et al. Systematic review of lung function and COPD with peripheral blood DNA methylation in population based studies. *BMC Pulm Med.* 2017;17(1):54.
23. Kodal JB, Kobylecki CJ, Vedel-Krogh S, Nordestgaard BG, Bojesen SE. AHRR hypomethylation, lung function, lung function decline and respiratory symptoms. *Eur Respir J.* 2018;51(3).
24. Imboden M, Wielscher M, Rezwan FI, Amaral AFS, Schaffner E, Jeong A, et al. Epigenome-wide association study of lung function level and its change. *Eur Respir J.* 2019;54(1).
25. Morrow JD, Make B, Regan E, Han M, Hersh CP, Tal-Singer R, et al. DNA Methylation Is Predictive of Mortality in Current and Former Smokers. *Am J Respir Crit Care Med.* 2020;201(9):1099-109.
26. Shipton D, Tappin DM, Vadiveloo T, Crossley JA, Aitken DA, Chalmers J. Reliability of self reported smoking status by pregnant women for estimating smoking prevalence: a retrospective, cross sectional study. *BMJ.* 2009;339:b4347.
27. Grindstaff KK, Yeaman C, Anandasabapathy N, Hsu SC, Rodriguez-Boulan E, Scheller RH, et al. Sec6/8 complex is recruited to cell-cell contacts and specifies transport vesicle delivery to the basal-lateral membrane in epithelial cells. *Cell.* 1998;93(5):731-40.
28. Oztan A, Silvis M, Weisz OA, Bradbury NA, Hsu SC, Goldenring JR, et al. Exocyst requirement for endocytic traffic directed toward the apical and basolateral poles of polarized MDCK cells. *Mol Biol Cell.* 2007;18(10):3978-92.
29. Rodriguez-Boulan E, Macara IG. Organization and execution of the epithelial polarity programme. *Nat Rev Mol Cell Biol.* 2014;15(4):225-42.
30. Dorfman R. Modifier gene studies to identify new therapeutic targets in cystic fibrosis. *Curr Pharm Des.* 2012;18(5):674-82.
31. Corvol H, Blackman SM, Boelle PY, Gallins PJ, Pace RG, Stonebraker JR, et al. Genome-wide association meta-analysis identifies five modifier loci of lung disease severity in cystic fibrosis. *Nat Commun.* 2015;6:8382.
32. Held L. A new standard for the analysis and design of replication studies. *Journal of the Royal Statistical Society: Series A (Statistics in Society).* 2020;183(2):431-48.
33. Stueve TR, Li WQ, Shi J, Marconett CN, Zhang T, Yang C, et al. Epigenome-wide analysis of DNA methylation in lung tissue shows concordance with blood studies and identifies tobacco smoke-inducible enhancers. *Hum Mol Genet.* 2017;26(15):3014-27.