MULTI-OMICS STUDY OF CHRONIC OBSTRUCTIVE PULMONARY DISEASE AND RELATED DISORDERS

IVANA PROKIĆ

Multi-omics Study of Chronic Obstructive Pulmonary Disease and Related Disorders

Ivana Prokić

Acknowledgements
The work described in this thesis was conducted at the Genetic Epidemiology Unit, Department of Epidemiology, Erasmus Medical Center, Rotterdam, The Netherlands.

The work presented in this thesis was supported by grant number 4.1.13.007 of Lung Foundation Netherlands (Longfonds), Biobanking and Biomolecular Resources Research Infrastracture (BBMRI)-NL (184.021.007), Corbell, and by ERAWEB scholarship.

The Erasmus Rucphen Family study as a part of EUROSPAN (European Special Populations Research Network) was supported by European Commission FP6 STRP grant number 018947 (LSHG-CT-2006-01947) and also received funding from the European Community's Seventh Framework Programme (FP7/2007-2013)/ grant agreement HEALTH-F4-2007-201413 by the European Commission under the programme "Quality of Life and Management of the Living Resources" of 5th Framework Programme (no. QLG2-CT-2002-01254). 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. All study participants, general practitioners, specialists, researchers, institutions and funders of all other studies from this thesis are appreciatively acknowledged.

The publication and printing of this thesis was financially supported by: Department of Epidemiology, Erasmus Medical Center, Rotterdam and Erasmus University Rotterdam.
Cover design: Slobodanka Nišić and Ivana Prokić
Layout: Optima Grafische Communicatie
Printing: Optima Grafische Communicatie
ISBN: 978-94-6361-290-6

Multi-omics Study of Chronic Obstructive Pulmonary Disease and Related Disorders

Multi-omics studie van Chronische Obstructieve Longziekte en verwante stoornissen

Proefschrift

ter verkrijging van de graad van doctor aan de
Erasmus Universiteit Rotterdam
op gezag van de

rector magnificus
Prof. dr. R.C.M.E. Engels

en volgens besluit van het College voor Promoties.

De openbare verdediging zal plaatsvinden op
woensdag, 4 september 2019 om 11:30 uur

door
Ivana Prokić (geboren Nedeljković)
geboren te Belgrado, Servië

Erasmus University Rotterdam

Promotiecommissie

Promotor:	Prof. dr. ir. C.M. van Duijn
Overige leden:	Prof. dr. H.M. Boezen
	Prof. dr. G.G. Brusselle
	Prof. dr. J. van Meurs
Copromotor:	Dr. N. Amin
Paranimphen:	Desana Kocevska
	Ashley van der Spek

Za moju večitu pordršku,
Jovanu, Marka, Anđu i Miška;
I moju novu snagu,
Petra i Lenku.

TABLE OF CONTENTS

PUBLICATIONS AND MANUSCRIPTS IN THIS THESIS

Chapter 2.1
Ivana Nedeljković, Natalie Terzikhan, Judith M. Vonk, Diana A. van der Plaat, Lies Lahousse, Cleo C. van Diemen, Brian D. Hobbs, Dandi Qiao, Michael H. Cho, Guy G. Brusselle, Dirkje S. Postma, H. MarikeBoezen, Cornelia M. van Duijn, Najaf Amin. *A genome-wide linkage study for chronic obstructive pulmonary disease in a Dutch genetic isolate identifies novel rare candidate variants.* Frontiers in Genetics, 2018 Apr 19; 9:133; doi: 10.3389/fgene.2018.00133.

Chapter 2.2
Ivana Nedeljković, Elena Carnero-Montoro, Lies Lahousse, Diana A. van der Plaat, Kim de Jong, Judith M. Vonk, Cleo C. van Diemen, AlenFaiz, Maarten van den Berge, Ma'enObeidat, Yohan Bossé, David C. Nickle, BIOS consortium, Andre G. Uitterlinden, Joyce B.J. van Meurs, Bruno H.C. Stricker, Guy G. Brusselle, Dirkje S. Postma, H. MarikeBoezen, Cornelia M. van Duijn, Najaf Amin. *Understanding the role of the chromosome 15q25.1 in COPD through epigenetics and transcriptomics.* European Journal of Human Genetics, 2018 Feb 8; 26:709–722; doi: 10.1038/s41431-017-0089-8.

Chapter 2.3
Ivana Nedeljković, Lies Lahousse, Elena Carnero-Montoro, AlenFaiz, Judith M. Vonk, Kim de Jong, Diana A. van der Plaat, Cleo C. van Diemen, Maarten van den Berge, Ma'enObeidat, Yohan Bossé, David C. Nickle, BIOS Consortium, Andre G. Uitterlinden, Joyce B.J. van Meurs, Bruno H.C. Stricker, Guy G. Brusselle, Dirkje S. Postma, H. MarikeBoezen, Cornelia M. van Duijn, Najaf Amin. *COPD GWAS variant at 19q13.2 in relation with DNA methylation and gene expression.* Human Molecular Genetics, 2018 Jan 15; 27(2):396-405; doi: 10.1093/hmg/ddx390.

Chapter 2.4
Maaike de Vries, **Ivana Nedeljković**, Diana A. van der Plaat, Lies Lahouse, Guy G. Brusselle, Alexandra Zhernakova, Cornelia M. van Duijn, Judith M. Vonk, H. Marike-Boezen. *DNA methylation is associated with lung function levels in never-smokers.* Manuscript submitted for publication.

Chapter 3.1
Herman T. den Dekker*, Kimberley Burrows*, Janine F. Felix*, Lucas A. Salas, **Ivana Nedeljković**, Jin Yao, Sheryl L. Rifas-Shiman, Carlos Ruiz-Arenas, N. Amin, Mariona Bustamante, Dawn L. DeMeo, A. John Henderson, Caitlin G. Howe, Marie-France Hivert, M. Arfan Ikram, Johan C. de Jongste, Lies Lahousse, Pooja R. Mandaviya, Joyce B. van Meurs, MarionaPinart, Gemma C. Sharp, Lisette Stolk, André G. Uitterlinden, Josep

M. Anto, Augusto A. Litonjua, Carrie V. Breton, Guy G. Brusselle, Jordi Sunyer, George Davey Smith, Caroline L. Relton[#], Vincent W.V. Jaddoe[#], LiesbethDuijts[#]. *New-born DNA-methylation, childhood lung function, and the risks of asthma and COPD across the life course.* Eur Respir J. February 2019:1801795. doi:10.1183/13993003.01795-2018

**first authors contributed equally*
[#]last authors contributed equally

Chapter 3.2
Ivana Nedeljković, Lies Lahousse, Maaike de Vries, Jun Liu, Marita Kalaoja, Judith M. Vonk, Diana A. van der Plaat, Cleo C. van Diemen, Ashley van der Spek, Alexandra Zhernakova, Jingyuan Fu, Mohsen Ghanbari, Mika Ala-Korpela, Johannes Kettunen, Aki S. Havulinna, Markus Perola, VeikkoSalomaa, Lars Lind, Johan Ärnlöv, Bruno H.C. Stricker, Guy G. Brusselle, H. MarikeBoezen, Cornelia M. van Duijn, Najaf Amin. *A cross-omics integrative study of metabolomic signature of Chronic obstructive pulmonary disease.* Manuscript submitted for publication.

Chapter 3.3
Ivana Nedeljković, Maaike de Vries, Judith M. Vonk, Lies Lahousse, Bruno H.C. Stricker, Guy G. Brusselle, Najaf Amin, Marike Boezen, Cornelia M. van Duijn. *Genetic correlation of Chronic Obstructive Pulmonary Disease and non-pulmonary comorbidity*. Manuscript in preparation for submission.

Chapter 3.4
Olivera Jovanova*, **Ivana Nedeljković***, Derek Spieler *, Rosie M. Walker*, Chunyu Liu*, Michelle Luciano*, Jan Bressler, Jennifer Brody, Amanda J. Drake, Kathryn L. Evans, Rahul Gondalia, Sonja Kunze, Brigitte Kuhnel, Jari Lahti, Rozenn N. Lemaitre, Riccardo E. Marioni, Brenton Swenson, Jayandra Jung Himali, Hongsheng Wu, Yun Li, Allan F. McRae, Tom C. Russ, James Stewart, Zhiying Wang, Guosheng Zhang, Karl-Heinz Ladwig, Andre G. Uitterlinden, Xiuqing Guo, Annette Peters, KatriRäikkönen, John M. Starr, Melanie Waldenberger, Naomi R. Wray, Whitsel A. Eric, Nona Sotoodehnia, Sudha Seshadri, David J. Porteous, Joyce van Meurs, Thomas H. Mosley, Andrew M. McIntosh, Michael M. Mendelson, Daniel Levy, Lifang Hou, Johan G. Eriksson, Myriam Fornage, Ian J. Deary, Andrea Baccarelli, Henning Tiemeier, Najaf Amin. *DNA methylation signatures of depressive symptoms in middle-aged and elderly persons: meta-analysis of multiethnic epigenome-wide studies.* JAMA Psychiatry. 2018;75(9):949-959. doi:10.1001/jamapsychiatry.2018.1725

**first authors contributed equally*

Chapter 1

Introduction

COPD

Chronic obstructive pulmonary disease (COPD) is the most common respiratory disease, characterized by chronic and progressive course.[1] It's pathology involves chronic inflammatory response of the airways, overproduction of mucus (resulting in chronic bronchitis), parenchymal tissue destruction (resulting in emphysema) and abnormal repair defence mechanisms (resulting in small airway fibrosis).[2] This leads to air trapping in the lungs, sputum production, obstructed exhalation, dyspnoea and cough, common symptoms associated with COPD.[2] Although COPD can be stable over time, exacerbations, defined as an acute worsening of respiratory symptoms resulting in additional therapy, often occur.[3]

Epidemiology and risk factors

Chronic obstructive pulmonary disease is a major public health burden.[3,4] COPD is currently the third leading cause of death worldwide with more than 3 million deaths per year.[5,6] Although it is difficult to estimate the prevalence due to the variability in diagnostic criteria, recent standardized meta-analyses show a significant increase in both global and regional prevalence in 2010, compared with 1990.[7] In 2010, the global prevalence based on spirometry was estimated to be 11.7% with 384 million cases.[7] Prevalence is higher in current smokers and ex-smokers, in males compared with females and increases with age and air pollution.[3]

The COPD prevalence and annual deaths are predicted to increase, due to the increased prevalence of smoking and air pollution in some regions and aging of the population.[4] Exacerbations are an important reason for hospitalization and are responsible for about 10% of all acute medical admissions, adding to the mortality and morbidity rates and overall burden of the disease.[8] Survival rates of COPD patients with three or more exacerbations in 5 years follow-up are markedly reduced compared with those without exacerbations (30% versus 80%).[9]

Although smoking is a predominant risk factor, 25-45% of never-smokers also develop COPD.[10,11] It has been hypothesized that COPD is the result of a more complex interaction of cumulative exposures to noxious gases and particles (smoking, air pollution and/or occupational exposure) and a range of host factors, including (epi) genetic factors, poor lung growth, age and airway hyper-responsiveness.[3] From a genetic perspective, an important question to answer is to what extent the genetic determinants of COPD are overlapping in smokers and non-smokers or whether there are specific gene-environment interactions that change the genetic architecture in these two groups.

Diagnosis

According to the Global initiative for chronic Obstructive Lung Disease (GOLD) the COPD diagnosis is based on the airflow limitation, as measured by the lung function tests.[2] Spirometry is the most objective lung function test and the post-bronchodilator ratio of the forced expiratory volume in 1 second (FEV_1) over the forced vital capacity of the lungs (FVC) resulting in <0.7 is a standard definition of the airflow limitation.[3] However, using this fixed ratio results in more frequent over-diagnosis in the elderly (the lung function normally lowers with age), and more frequent under-diagnosis in younger adults (<45 years).[12] Thus, the American Thoracic Society (ATS) and the European Respiratory Society (ERS) guidelines recommend the lower limit of normal (LLN) as a cut-off value (FEV_1/FVC<LLN). LLN represents the lower 5% of the healthy population, evaluated by comparison with the reference values based on age, height, sex and race.[13] However, this value is highly dependent on the reference population. Since simplicity and consistency of a diagnostic tool are highly valued in clinical practice and research, GOLD still prefers the use of the fixed ratio[3] and is therefore widely used in genetic and epidemiological studies as well as in the studies described in this thesis. In the new assessment tool proposed by GOLD 2017,[3] COPD is classified in stages of severity based on the combination of severity of airflow limitation (FEV_1 % predicted), exacerbation history and symptoms burden (**Figure 1**).[2,3] GOLD is confident that this tool will result in a decrease of misclassification and better diagnosis and treatment of COPD.

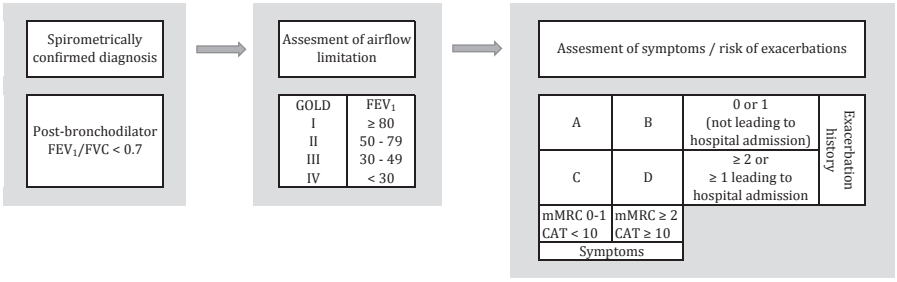

Figure 1. Combined COPD assessment tool proposed by GOLD 2017 (Adapted with permission from GOLD from "GOLD Management and Prevention of COPD 2017", Copyright © 2016 GOLD),[2] mMRC- Modified British Medical Research Council Questionnaire used for symptom assessment; CAT – COPD Assessment Test™.

Comorbidities

Various other pulmonary conditions are known to coexist with COPD and increase the severity of the disease. Those include asthma, pneumonia, pulmonary hypertension, pulmonary embolism, obstructive sleep apnoea, idiopathic pulmonary fibrosis and lung cancer.[14,15] Most are considered to be part of the COPD spectrum or a consequence of COPD pathology.[14] Asthma is considered to be a major risk factor for COPD, where people with asthma have 12-fold increased risk of COPD compared with those without asthma.[16] However, it is difficult to clinically differentiate asthma and COPD in adults as in 40% of the elderly it coexists with COPD.[17]

Furthermore, COPD is a systemic disorder that is associated with multiple extrapulmonary comorbid diseases.[18,19] Most common are cardiovascular diseases, metabolic diseases, cancer and depression, among many others.[15] The comorbidities may in part be explained by common factors such as smoking, alcohol, diet, ageing and polypharmacy or may share pathophysiological mechanisms and be consequence of the systemic inflammation.[15,18] Comorbidities have impact on the severity of the exacerbations and consequently on hospitalization rates and prognosis and are thus relevant for clinical care and management.[20] Depression is proposed to be one of the most underestimated, yet prevalent comorbidities of COPD[15] for which the common mechanisms are far from understood.[21] A total of 26% of COPD cases have depression, which has been associated with female gender, younger age, poor prognosis, smoking and severity of COPD with higher exacerbation risk.[22,23] Depression may be the result of (preclinical) pathology, which impacts quality of life. On the other hand, it has been speculated that there may be shared risk factors with effects on brain, such as smoking, ageing, hypoxaemia and systemic inflammation.[15,24] Alternatively, there may be shared genomics determinants.[15] In the present study, I studied the common genetic and epigenetic determinants of COPD, depression and other COPD related comorbidity.

OMICS OF COPD

The suffix *-omics* (from Greek word "ὁμοῖος" - common, general, one that concerns all parts) added to a molecular term denotes a comprehensive or global assessment of a set of molecules, which are collectively denoted with the suffix *–ome.*[25] Accordingly, *genomics, epigenomics, transcriptomics* and *metabolomics* represent a comprehensive study of a *genome, epigenome, transcriptome* and *metabolome,* respectively, the complete sets of different genes, transcripts of genes, proteins or active molecules (metabolites) of an organism (**Figure 2**).

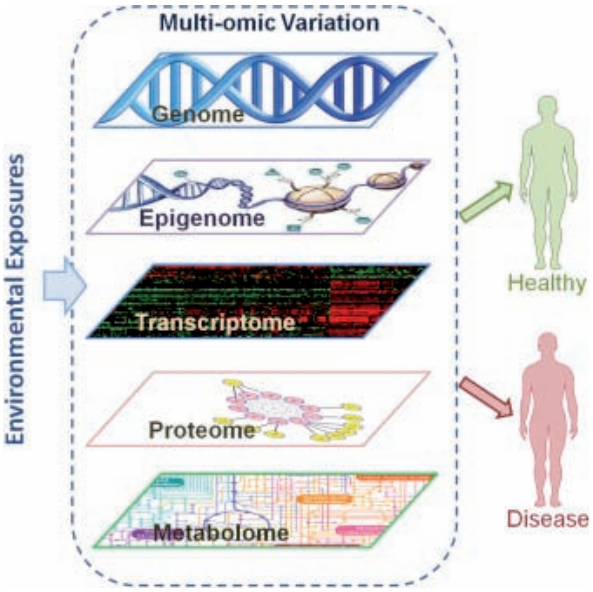

Figure 2. Multi-omics approach to studying a disease. Reprinted with permission from Elsevier. Sun YV, Hu YJ. Chapter Three-Integrative Analysis of Multi-omics Data for Discovery and Functional Studies of Complex Human Diseases. Advances in genetics. 2016 Dec 31;93:147-90. Copyright © 2016 Elsevier Inc.

Analyses that integrate these layers are powerful tools for understanding the pathogenesis and pathology of complex diseases.[25] Such integrative studies may improve our understanding of how specific genetic variations contribute to the disease.[26] The integration of data across multi-omics layers allows us to:

- gain understanding of the functional consequences and relevant interactions between different layers;[27]
- build pathways and networks based on a prior published or bioinformatic knowledge in order to understand the pathophysiology of a disease.[26,28]

There has been significant progress in understanding pulmonary diseases in recent years based on the development of omics research.[29] COPD is a complex disease with overlapping endophenotypes, which may be the result of interactions of many factors, both external and internal.[30] In this thesis I aim to disentangle the pathogenesis of COPD and its co-morbidity, using various omics approaches discussed below.

Genetics

Genetics focuses on identification of a DNA (Deoxyribonucleic acid) sequence changes, such as single nucleotide variations (SNVs). These may be associated with the risk and development of pathology, treatment response or prognosis.[26,31] The human genome is an important driver of the risk of COPD. The heritability of COPD is estimated to be 20-60%.[32,33] COPD as a complex disease is likely the result of the interplay of rare variants with moderate to large effects and common variants with small effects. Genetic studies identified several genetic risk factors for COPD. The first and most well-known genetic variant causing emphysema at young age is the rare variant in *SERPINA1* gene at chromosome 14q, resulting in Alpha-1-antitrypsin (AAT) deficiency.[34,35] Candidate-gene studies, focusing on genes encoding protein implicated in the pathogenesis of COPD, highlighted broad areas of the genome potentially involved in COPD, but did not yield informative reproducible results.[36] Genome-wide association studies (GWAS), using hypothesis-free and genome wide approach, have successfully identified common variants associated with COPD[37–43] and related outcomes, such as lung function measurements (FEV_1, FEV_1/FVC),[37,44–47] emphysema,[48] chronic bronchitis.[41,49] Findings are not only replicable within an endophenotype, but also show a substantial overlap accross.[43] The loci identified in COPD GWASs that were replicated include Hedgehog-interacting protein (*HHIP*), Family with sequence similarity 13 member A (*FAM13A*), Nicotinic cholinergic receptors (*CHRNA3/5*), Ion-responsive element binding protein 2 (*IREB2*), Cytochrome P450 family gene (*CYP2A6*), Member RAS oncogene family gene (*RAB4B*) and Egl-9 family hypoxic-inducible factor 2 (*EGLN2*).[37,43]

As has been the case in many other disorders, the use of endophenotypes, i.e., continuous heritable traits that are associated with the disease (diagnosis), has been even more successful in identifying genetic loci.[50] The major advantage of this approach is that it overcomes the problems of diagnostic classification, which for many disorders including COPD is arbitrary and may introduce misclassification. The use of endophenotypes results in loss of specificity as there is no 1:1 relationship between the endophenotype and the disease and endophenotype may be related to multiple disorders.[51] Yet, there is a gain in efficiency because the endophenotypes often have a higher heritability than the disease and are usually available in large number of persons, covering a full range of disease severity: from healthy, pre-clinic, moderate to severe. Based on a genome-wide association discovery in 48,943 individuals and follow-up in 95,375 individuals, Wain et al. reported 97 loci relevant for lung function, of which 43 were novel.[37] **Figure 3** gives an overview of the 97 loci, underlying those relevant for COPD.

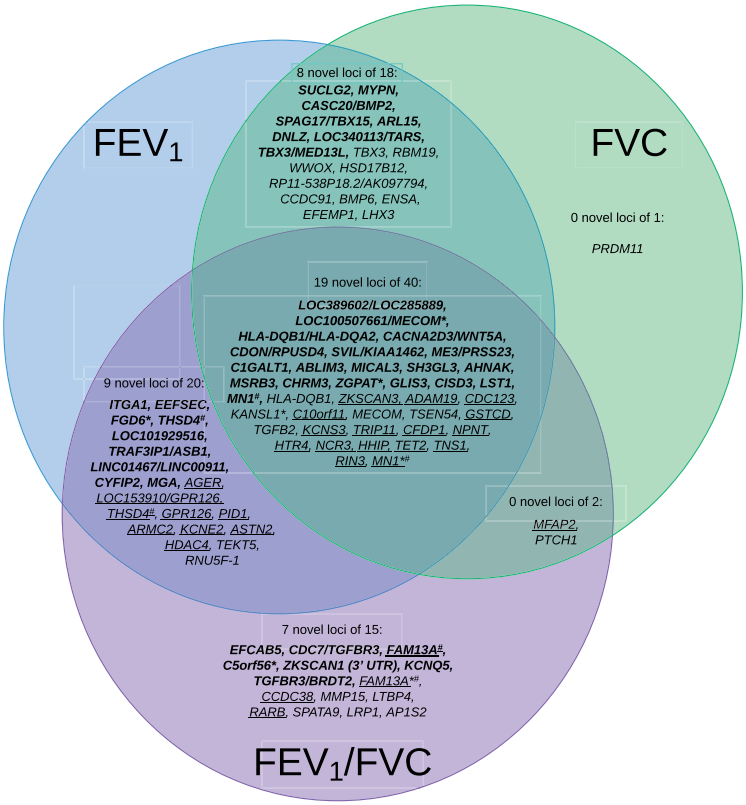

Figure 3. Loci associated with lung function related to COPD.[50] In bold - novel findings. Underlined – loci associated with COPD ($P<5.26\times10^{-4}$). *Loci associated with smoking. #Same gene has 2 variants one novel, one already known.

The genetic risk score derived from these is associated with COPD susceptibility results in 3.7-fold difference in COPD risk between highest and lowest genetic risk score deciles (**Figure 4**).[37] The odds ratios per standard deviation of the risk score (~6 alleles) (95% confidence interval) is 1.24 (1.20-1.27), $P=5.05\times10^{-49}$ show a consistent increase over the full distribution.

When interpreting the biological and physiological pathways the 97 genetic variants are implicated in those involved in development, elastic fibres and epigenetic regulation pathways. These pathways point to targets for drugs and compounds in development for COPD and asthma.

Despite the successes, a large part of the estimated heritability is still missing. This may be explained by:

- Rare variants that are not well covered to date by GWASs;
- Gene interactions;
- Epigenetic modifications that are in part driven by genetic variants.

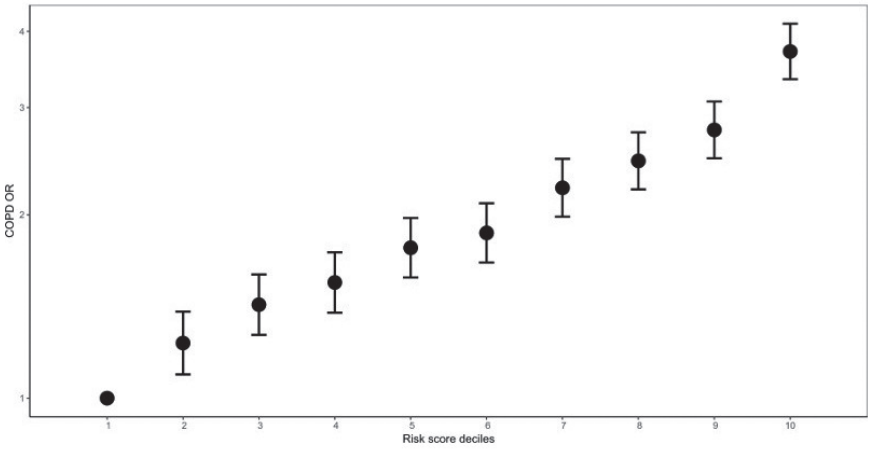

Figure 4. Odds ratios for spirometrically-defined COPD for weighted genetic risk score deciles in UK Biobank (10,547 cases, pre-bronchodilator % predicted FEV_1<80% and FEV_1/FVC<0.7, and 53,948 controls, FEV_1/FVC>0.7 and % predicted FEV_1>80%, weights derived from non-discovery populations). For each decile, odds ratios were obtained using a logistic regression adjusted for age, age2, sex, height, smoking status, pack-years and the first 10 ancestry principal components. Source: Wain et al.[37]

So far GWAS has brought to surface common variants. Rare variants are not covered by the arrays used for GWAS, but, more importantly, are also not well imputed using common reference panels (e.g. HapMap and 1000 Genomes).[52,53] Of note is that imputation is improving with larger reference panels, such as the Haplotype Reference Consortium panel (HRC) combining several widely used panels (with total of 64,976 haplotypes) and data from exome sequencing. Using HRC, rare variants can be imputed more reliably in GWAS.[52] An alternative route to discover rare variants is family based studies. While a variant is rare in the general population, within a family of first- and second-degree relatives such variant will be transmitted with a 50% probability. Thus, within a family, the variant is common. To find rare variants Qiao et al conducted a whole exome sequencing analysis in 2,543 subjects from two family-based studies.[54] Applying a gene-based segregation test in the family-based data, they identified significant segrega-

tion of rare loss of function variants in *TBC1D10A* and *RFPL1* ($P<2\times10^{-6}$) but were unable to find similar variants in the case–control study. Further, they identified individuals with putative high-risk variants, including patients harbouring homozygous mutations in genes associated with cutis laxa and Niemann–Pick Disease Type C.[54] Also a recent whole genome sequencing study in severe COPD identified a large number of potentially important functional variants, with the strongest associations being in known COPD risk loci, including *HHIP* and *SERPINA1*.[55] Encouraged by these findings, in this thesis I also used a family-based approach to identify rare variants implicated in COPD.

Epigenomics

Epigenomics investigates epigenome which is a set of chemical modifications of the chromatin and DNA molecule that regulate gene expression, without changing the DNA sequence.[26,28] These changes are usually reversible, and may be driven by genetic (heritable) and environmental factors. Of note is that in some instances the modifications may be permanent,[56,57] and cell-type (tissue) specific.[58]

The most commonly studied epigenetic mechanisms are DNA methylation and histone modifications.[29] DNA methylation is addition of a methyl group ($-CH_3$) to any cytosine (C) that is next to guanine (G) in the DNA sequence, converting it to 5-methylcytosine. These sites are called CpGs (short for 5'-C-phosphate-G-3'), and in humans around 70-80% of CpGs are methylated.[59] Epigenome-wide association studies (EWAS) have shown that differential DNA methylation patterns have a role in the disease development.[60] It has also been shown that smoking affects DNA methylation,[56,61,62] which in turn may lead to the disease. Furthermore, genetic variants may modulate regulatory mechanisms such as DNA methylation (methylation quantitative trait loci - meQTLs).[63] Epigenetic studies of COPD have identified differential DNA methylation associated with COPD severity, poor lung function and use of systemic corticosteroids.[64–66] It has been postulated that early exposure to risk factors, such as maternal tobacco smoking during pregnancy, are associated with risk of asthma and lower lung function, through changes in DNA methylation.[67] This may also affect the risk of COPD at old age. When combining epigenome and transcriptome data from lung tissues of COPD patients and controls, *EPAS1* gene has been proposed as a key regulator of COPD pathogenesis and has been confirmed by functional studies, highlighting the need for integrative studies.[68] This gene has not emerged in the list of genes implicated in COPD or endophenotypes to date.

In this thesis, I addressed the specific question whether the GWAS variants change the epigenome landscape and subsequently alter the transcription of the gene, integrating genetic, epigenetic and transcriptomic data. GWAS has been extremely successful, but the functional effects of the identified genes in COPD pathogenesis

were largely not investigated. Another poorly understood issue is the interaction of the genetic drivers of pathology with the environment. Integrating genetic research with other -omics may improve our understanding of functional effects and gene interactions, since at the omics level such effects are expected to be larger than at the level of a complex disease such as COPD, which involves a large range of phenotypes and comorbidities driven by both external and internal factors.[30] In this thesis, I aimed to understand the functional changes driving the association of GWAS hits to COPD at the level of epigenomics and transcriptomics. I further use genetics to address the question whether a common genetic background explains the comorbidity in COPD occurring in patients.

Transcriptomics

Transcriptomics explores genome-wide levels of RNA transcripts (gene expression) both qualitatively and quantitatively, which are directly influenced by the genome (expression quantitative trait loci – eQTLs) and epigenome (expression quantitative trait methylation – eQTMs),[26] besides environmental factors. It is known that gene expression can be tissue specific and in order to investigate a disease one should focus on the tissue of interest. One study showed that environmental risk factors such as smoking influences the transcriptome of the small airway epithelium,[69] even after smoking cessation.[70] However, some genes are expressed globally over tissues. An important issue to consider is that multiple tissues may be involved in a disease. Smoking, the major determinant of COPD, may affect the expression in blood, lung tissue or other tissues. Indeed, a study investigating blood of smokers with and without COPD, could discriminate the cases from the controls based on the expression profile of 26 genes involved in immune and inflammatory response and sphingolipid metabolism.[71] Although transcriptomic studies were useful in identifying specific gene expression pattern associated with COPD[72,73] and with drug response,[74,75] a global expression profile unique for COPD has not been found.[29] In this thesis, I chose to integrate genomics with gene expression to explore the functional effects of genetic and epigenetic changes.

Metabolomics

Metabolomic studies all metabolites present in a tissue, which are small molecules (<1 kDa) of endogenous or exogenous etiology.[29] These include peptides, amino acids, nucleic acids, carbohydrates, vitamins, polyphenols, and alkaloids, among other compounds that are involved in cellular metabolic functions. In pulmonary research of metabolomics, studied samples include blood, sputum, exhaled breath condensate, bronchoalveolar lavage fluid and lung tissue.[76] The identification

of changes in biomarkers that can identify or differentiate various disease phenotypes even in the early stages is of high importance in COPD.[76] Several studies used metabolomics methods to investigate biochemical effects induced by COPD, exacerbations and its related outcomes as well as external effects of smoking and drugs, using different samples.[76] Most of the studies identified metabolites involved in systemic inflammation, protein degradation and oxidative stress.[77–79] Consistent with the transcriptomics studies in blood, mentioned above, another study of lipids in sputum reported that sphingolipids were highly expressed in sputum of smokers with COPD compared with smoking controls.[80,81] However, these studies were very limited in sample size, therefore the results should be further confirmed in larger samples. In this thesis I have combined the data of two large population-based studies to understand the metabolomics changes in COPD. As a person's metabolism may change causing the disease or change as a result of the disease process, I used a genomic method, explained below, to disentangle these effects.

Mendelian Randomization

A major problem in observational epidemiological studies and the translation of findings to the clinics is the problem of causal inferences due to the possible reverse causation: e.g. to distinguish whether the metabolic or other omics changes are causing a disease or are the consequence of the pathology. One of the most important approaches developed in the omics era is the method referred to as Mendelian Randomization (MR). MR is a cross-omics approach, which uses genetic data as an instrumental variable (IV) to examine the evidence for causal effects between modifiable exposures (risk factors) and an outcome (disease).[82] The rationale is that similar to randomized controlled trials, the genotypes are assigned randomly and the disease starts after meiosis.[83] Randomisation is based on Mendel's second law that the inheritance of one trait is independent of the inheritance of other traits.[83] The IV (usually based on a combination of genotypes that are associated to the disease) has to comply with three assumptions: (1) to be associated with the exposure; (2) to be independent of any confounders of the exposure-outcome association and (3) to be related to the outcome only through the exposure.[83] MR analysis can be conducted unilateral, testing a specific hypothesis, e.g. if alcohol consumption is casually related to the risk of cardiovascular mortality.[84] In the setting of multi- or cross-omics research as in the metabolomics-COPD study I performed, the MR is often bi-directional, testing the hypothesis that: 1) the metabolite is causally related to COPD and therefore the genetic determinants of metabolite (used as instrumental variable) are also associated to COPD and 2) (pre)clinical COPD pathology affect the metabolite levels, which translates into the model where genes determining COPD are also associated to metabolite.

SCOPE OF THIS THESIS

The overall aim of this thesis is to identify novel molecular determinants of COPD, lower lung function and related pathology such as depression and to perform integrative studies to investigate the functional role and interaction of multiple omics layers.

In **Chapter 2** I investigate COPD applying different omics approaches. In **Chapter 2.1,** I describe a genome-wide linkage scan performed in a search for rare genetic variants which have a role in familial COPD, utilizing family-based settings of the Erasmus Rucphen Family (ERF) study and integrating the data from the Rotterdam Study (RS), the LifeLines study (LLS), Hobbs et al.[85] and the Vlagtwedde/ Vlaardingen study. **Chapter 2.2** and **Chapter 2.3** investigate the functional role of two established COPD GWAS loci by exploring a multi-omics approach linking the genetic loci to the epigenomic and transcriptomic effects in the Rotterdam study and the Lung expression quantitative loci mapping study. **Chapter 2.2** examines the chromosome 15q25 locus and its meQTL effects in blood and eQTL effects in lung tissue, to understand the functional effects of this locus in relation to COPD. Similarly, **Chapter 2.3** investigates a top variant from a novel locus on 19q13, identified in COPD GWAS, and mediation of its genetic risk on gene expression, through DNA methylation signatures. In **Chapter 2.4,** I present an EWAS meta-analysis of lung function levels in never-smokers only, to identify factors other than smoking which affect lung function through DNA methylation in RS and LLS.

In **Chapter 3**, the thesis focuses on comorbidities of COPD, including early and late metabolic effects. **Chapter 3.1** describes a large meta-analysis in Pregnancy And Childhood Epigenetics (PACE) consortium studying DNA methylation in relation to lung function at birth and the effects on lung function, asthma and COPD throughout life course. In **Chapter 3.2,** I study circulating metabolites in relation with COPD in ERF, RS and several replication cohorts and apply multi-omics Mendelian Randomization approach to investigate causal relations of the metabolite-COPD associations. In **Chapter 3.3**, I use an integrative genetic approach to overlap genetic drivers of COPD and its non-pulmonary comorbidity. In **Chapter 3.4,** I investigate DNA methylation patterns specific for depression in a largest to date EWAS study in Cohorts for Heart and Aging in Genomic Epidemiology (CHARGE) consortium with the view to determine the overlap with that seen in COPD.

The main findings and implications described in my thesis I discuss in the **Chapter 4,** which I summarize in English and in Dutch in **Chapter 5.**

REFERENCES

1. European Respiratory Society. European lung white book. http://www.erswhitebook.org/chapters/chronic-obstructive-pulmonary-disease/. Published 2016.
2. Vestbo J, Hurd SS, Agustí AG, et al. Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease GOLD executive summary. *Am J Respir Crit Care Med*. 2013;187(4):347-365. doi:10.1164/rccm.201204-0596PP
3. GOLD. From the Global Strategy for the Diagnosis, Management and Prevention of COPD, Global Initiative for Chronic Obstructive Lung Disease. *Available from http// goldcopd org*. 2017. http://goldcopd.org.
4. World Health Organization. Projections of global mortality and burden of disease from 2015 to 2030. *Heal Stat Inf Syst*. 2013;3(11):2015-2030. doi:978 92 4 156422 9
5. Naghavi M, Wang H, Lozano R, et al. Global, regional, and national age-sex specific all-cause and cause-specific mortality for 240 causes of death, 1990-2013: A systematic analysis for the Global Burden of Disease Study 2013. *Lancet*. 2015;385(9963):117-171. doi:10.1016/S0140-6736(14)61682-2
6. Lozano R, Naghavi M FK. Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet*. 2012;380(380):2095–2128. doi:https://doi.org/10.1016/S0140-6736(12)61728-0
7. Adeloye D, Chua S, Lee C, et al. Global and regional estimates of COPD prevalence: Systematic review and meta–analysis. *J Glob Health*. 2015;5(2). doi:10.7189/jogh.05.020415
8. Roberts CM, Stone RA, Lowe D, Pursey NA, Buckingham RJ. Co-morbidities and 90-day outcomes in hospitalized COPD exacerbations. *COPD J Chronic Obstr Pulm Dis*. 2011;8(5):354-361. doi:10.3109/15412555.2011.600362
9. Soler-Cataluna JJ. Severe acute exacerbations and mortality in patients with chronic obstructive pulmonary disease. *Thorax*. 2005;60(11):925-931. doi:10.1136/thx.2005.040527
10. Lamprecht B, McBurnie MA, Vollmer WM, et al. COPD in Never Smokers\nResults From the Population-Based Burden of Obstructive Lung Disease Study. *Chest*. 2011;139(April):752-763. doi:10.1378/chest.10-1253
11. Pavord ID, Yousaf N, Birring SS, Salvi SS, Barnes PJ. Chronic obstructive pulmonary disease in non-smokers. *Lancet*. 2009;374(9691):1964; author reply 1965--1966. doi:10.1016/S0140-6736(09)62114-0
12. Celli BR, Macnee W. Standards for the diagnosis and treatment of patients with COPD : a summary of the ATS / ERS position paper. *Eur Respir J*. 2004;23(6):932-946. doi:10.1183/09031936.04.00014304
13. van Dijk W, Tan W, Li P, Best G, Li S. Clinical Relevance of Fixed Ratio vs Lower Limit of Normal of FEV_1:FVC in COPD: Patient-Reported Out.pdf. *Ann Fam Med*. 2015;13(1):41-48.
14. Chatila WM, Thomashow BM, Minai OA, Criner GJ, Make BJ. Comorbidities in Chronic Obstructive Pulmonary Disease. *Proc Am Thorac Soc*. 2008;5(4):549-555. doi:10.1513/pats.200709-148ET
15. Barnes PJ, Celli BR. Systemic manifestations and comorbidities. *Eur Respir J*. 2009;33(5):1165-1185. doi:10.1183/09031936.00128008
16. Silva GE. Asthma as a Risk Factor for COPD in a Longitudinal Study. *Chest*. 2004;126(1):59-65. doi:10.1378/chest.126.1.59

17. Aryal S, Diaz-guzman E, Mannino DM. Asthma Treatment Options in Asthma and Chronic Obstructive Pulmonary Diseases Overlap Syndrome. *Touch Briefings, Eur Respir Dis*. 2011; 7(2):101-105.
18. Divo M, Cote CG, De Torres JP, et al. Comorbidities and risk of mortality in patients with chronic obstructive pulmonary disease. *Am J Respir Crit Care Med*. 2012;186(2):155-161. doi: 10.1164/rccm.201201-0034OC
19. Almagro P, Cabrera FJ, Diez-Manglano J, et al. Comorbidome and short-term prognosis in hospitalised COPD patients: The ESMI study. *Eur Respir J*. 2015;46(3):850-853. doi: 10.1183/09031936.00008015
20. Mannino DM, Thorn D, Swensen A, Holguin F. Prevalence and outcomes of diabetes, hypertension and cardiovascular disease in COPD. *Eur Respir J*. 2008;32(4):962-969. doi: 10.1183/09031936.00012408
21. Norwood RJ. A review of etiologies of depression in COPD. *Int J COPD*. 2007;2(4):485-491.
22. Hanania NA, Müllerova H, Locantore NW, et al. Determinants of depression in the ECLIPSE chronic obstructive pulmonary disease cohort. *Am J Respir Crit Care Med*. 2011;183(5):604-611. doi:10.1164/rccm.201003-0472OC
23. Maurer J, Rebbapragada V, Borson S, et al. Anxiety and depression in COPD: Current understanding, unanswered questions, and research needs. *Chest J*. 2008;134(4 SUPPL.):43S-56S. doi:10.1378/chest.08-0342
24. Anisman H, Merali Z, Hayley S. Neurotransmitter, peptide and cytokine processes in relation to depressive disorder: Comorbidity between depression and neurodegenerative disorders. *Prog Neurobiol*. 2008;85(1):1-74. doi:10.1016/j.pneurobio.2008.01.004
25. Wijmenga C, Zhernakova A. The importance of cohort studies in the post-GWAS era. *Nat Genet*. 2018;50(3):322-328. doi:10.1038/s41588-018-0066-3
26. Hasin Y, Seldin M, Lusis A. Multi-omics approaches to disease. *Genome Biol*. 2017;18(1):83. doi:10.1186/s13059-017-1215-1
27. Valdes AM, Glass D, Spector TD. Omics technologies and the study of human ageing. *Nat Rev Genet*. 2013;14(9):601-607. doi:10.1038/nrg3553
28. Sun Y V., Hu Y-J. Integrative Analysis of Multi-omics Data for Discovery and Functional Studies of Complex Human Diseases. *Adv Genet*. 2016;93:147-190. doi:10.1016/bs.adgen.2015.11.004
29. Kan M, Shumyatcher M, Himes BE. Using omics approaches to understand pulmonary diseases. *Respir Res*. 2017;18(1):149. doi:10.1186/s12931-017-0631-9
30. Huang Q. Genetic Study of Complex Diseases in the Post-GWAS Era. *J Genet Genomics*. 2015; 42(3):87-98. doi:10.1016/j.jgg.2015.02.001
31. Manichaikul A, Nguyen JN. Genetic studies as a tool for identifying novel potential targets for treatment of COPD. *Eur Respir J*. 2017;50(5):1702042. doi:10.1183/13993003.02042-2017
32. Ingebrigtsen T, Thomsen SF, Vestbo J, et al. Genetic influences on Chronic Obstructive Pulmonary Disease. *a Twin Study*. 2010;104(12):1890-1895. doi:10.1016/j.rmed.2010.05.004
33. Zhou JJ, Cho MH, Castaldi PJ, Hersh CP, Silverman EK, Laird NM. Heritability of Chronic Obstructive Pulmonary Disease and Related Phenotypes in Smokers. *Am J Respir Crit Care Med*. 2013;188(8):941-947. doi:10.1164/rccm.201302-0263OC [doi]
34. Bashir, A., Shah, N.N., Hazari, Y.M., Habib, M., Bashir, S., Hilal, N., Banday, M., Asrafuzzaman, S., and Fazili KM. Novel variants of SERPIN1A gene: Interplay between alpha1- antitrypsin deficiency and chronic obstructive pulmonary disease. *Respir Med*. 2016;117:139-149. doi: https://doi.org/10.1016/j.rmed.2016.06.005

35. Laurell CB ES, Laurell CB ES. The electrophoretic alpha 1-globulin pattern of serum in alpha 1-antitrypsin deficiency. *Scand J Clin Lab Invest*. 1963;15(2):132-140. doi:https://doi.org/10.1080/00365516309051324
36. Hobbs BD, Hersh CP. Integrative genomics of chronic obstructive pulmonary disease. *Biochem Biophys Res Commun*. 2014;452(2):276-286. doi:10.1016/J.BBRC.2014.07.086
37. Wain L V., Shrine N, Artigas MS, et al. Supplementary: Genome-wide association analyses for lung function and chronic obstructive pulmonary disease identify new loci and potential druggable targets. *Nat Genet*. 2017;49(3):416-425. doi:10.1038/ng.3787
38. Kim DK, Cho MH, Hersh CP, et al. Genome-wide association analysis of blood biomarkers in chronic obstructive pulmonary disease. *Am J Respir Crit Care Med*. 2012;186(12):1238-1247. doi:10.1164/rccm.201206-1013OC
39. Boezen HM. Genome-wide association studies: what do they teach us about asthma and chronic obstructive pulmonary disease? *Proc Am Thorac Soc*. 2009;6(8):701-703. doi:10.1513/pats.200907-058DP
40. Pillai SG, Ge D, Zhu G, et al. A Genome-Wide association study in chronic obstructive pulmonary disease (COPD): Identification of two major susceptibility loci. *PLoS Genet*. 2009;5(3):e1000421. doi:10.1371/journal.pgen.1000421
41. Silverman EK, Mosley JD, Palmer LJ, et al. Genome-wide linkage analysis of severe , early-onset chronic obstructive pulmonary disease : airflow obstruction and chronic bronchitis phenotypes. *Hum Mol Genet*. 2002;11(6):623-632. http://dx.doi.org/10.1093/hmg/11.6.623.
42. Cho MH, McDonald M-LN, Zhou X, et al. Risk loci for chronic obstructive pulmonary disease: a genome-wide association study and meta-analysis. *Lancet Respir Med*. 2014;2(3):214-225. doi:10.1016/S2213-2600(14)70002-5
43. Hobbs BD, De Jong K, Lamontagne M, et al. Genetic loci associated with chronic obstructive pulmonary disease overlap with loci for lung function and pulmonary fibrosis. *Nat Genet*. 2017;49(3):426-432. doi:10.1038/ng.3752
44. van der Plaat DA, de Jong K, Lahousse L, et al. Genome-wide association study on the FEV_1/FVC ratio in never-smokers identifies HHIP and FAM13A. *J Allergy Clin Immunol*. 2017;139(2):533-540. doi:10.1016/j.jaci.2016.06.062
45. Hansel NN, Ruczinski I, Rafaels N, et al. Genome-wide study identifies two loci associated with lung function decline in mild to moderate COPD. *Hum Genet*. 2013;132(1):79-90. doi:10.1007/s00439-012-1219-6
46. Repapi E, Sayers I, Wain L V., et al. Genome-wide association study identifies five loci associated with lung function. *Nat Genet*. 2010;42(1):36-44. doi:10.1038/ng.501
47. Artigas MS, Loth DW, Wain L V., et al. Genome-wide association and large-scale follow up identifies 16 new loci influencing lung function. *Nat Genet*. 2011;43(11):1082-1090. doi:10.1038/ng.941
48. Castaldi PJ, Cho MH, Estépar RSJ, et al. Genome-wide association identifies regulatory loci associated with distinct local histogram emphysema patterns. *Am J Respir Crit Care Med*. 2014;190(4):399-409. doi:10.1164/rccm.201403-0569OC
49. Lee JH, Cho MH, Hersh CP, et al. Genetic susceptibility for chronic bronchitis in chronic obstructive pulmonary disease. *Respir Res*. 2014;15(1):113. doi:10.1186/s12931-014-0113-2
50. Beauchaine TP, Constantino JN. Redefining the endophenotype concept to accommodate transdiagnostic vulnerabilities and etiological complexity. *Biomark Med*. 2017;11(9):769. doi:10.2217/BMM-2017-0002

51. Hasler G, Drevets WC, Manji HK, Charney DS. Discovering Endophenotypes for Major Depression. *Neuropsychopharmacology*. 2004;29(10):1765-1781. doi:10.1038/sj.npp.1300506
52. Iglesias AI, van der Lee SJ, Bonnemaijer PWM, et al. Haplotype reference consortium panel: Practical implications of imputations with large reference panels. *Hum Mutat*. 2017;38(8): 1025-1032. doi:10.1002/humu.23247
53. Huang J, Howie B, McCarthy S, et al. Improved imputation of low-frequency and rare variants using the UK10K haplotype reference panel. *Nat Commun*. 2015;6:8111. doi:10.1038/ncomms9111
54. Qiao D, Ameli A, Prokopenko D, et al. Whole exome sequencing analysis in severe chronic obstructive pulmonary disease. *Hum Mol Genet*. 2018;27(21):3801-3812. doi:10.1093/hmg/ddy269
55. Prokopenko D, Sakornsakolpat P, Fier HL, et al. Whole-Genome Sequencing in Severe Chronic Obstructive Pulmonary Disease. *Am J Respir Cell Mol Biol*. 2018;59(5):614-622. doi: 10.1165/rcmb.2018-0088OC
56. Joehanes R, Just AC, Marioni RE, et al. Epigenetic Signatures of Cigarette Smoking. *Circ Cardiovasc Genet*. 2016;9(5):436-447. doi:10.1161/CIRCGENETICS.116.001506
57. Ladd-Acosta C, Fallin MD. The role of epigenetics in genetic and environmental epidemiology. *Epigenomics*. 2016;8(2):271-283. doi:10.2217/epi.15.102
58. Taudt A, Colomé-Tatché M, Johannes F. Genetic sources of population epigenomic variation. *Nat Rev Genet*. 2016;17(6):319-332. doi:10.1038/nrg.2016.45
59. Jabbari K, Bernardi G. Cytosine methylation and CpG, TpG (CpA) and TpA frequencies. *Gene*. 2004;333(SUPPL.):143-149. doi:10.1016/j.gene.2004.02.043
60. M.B. T, L. D-C, N. V-R, H.C. W. DNA methylation in white blood cells: Association with risk factors in epidemiologic studies. *Epigenetics*. 2011;6(7):828-837. doi:http://dx.doi.org/10.4161/epi.6.7.16500
61. Breitling LP, Yang R, Korn B, Burwinkel B, Brenner H. Tobacco-smoking-related differential DNA methylation: 27 K discovery and replication. *Am J Hum Genet*. 2011;88(4):450-457. doi: 10.1016/j.ajhg.2011.03.003
62. Wan ES, Qiu W, Carey VJ, et al. Smoking-associated site-specific differential methylation in buccal mucosa in the COPDGene study. *Am J Respir Cell Mol Biol*. 2015;53(2):246-254. doi: 10.1165/rcmb.2014-0103OC
63. Shi J, Marconett C, Duan J, Hyland P, … PL-N, 2014 undefined. Characterizing the genetic basis of methylome diversity in histologically normal human lung tissue. *NatureCom*. 2014. https://www.nature.com/articles/ncomms4365.
64. Qiu W, Baccarelli A, Carey VJ, et al. Variable DNA methylation is associated with chronic obstructive pulmonary disease and lung function. *Am J Respir Crit Care Med*. 2012;185(4): 373-381. doi:10.1164/rccm.201108-1382OC
65. Wan ES, Qiu W, Baccarelli A, et al. Systemic steroid exposure is associated with differential methylation in chronic obstructive pulmonary disease. *Am J Respir Crit Care Med*. 2012; 186(12):1248-1255. doi:10.1164/rccm.201207-1280OC
66. Morrow JD, Cho MH, Hersh CP, et al. DNA methylation profiling in human lung tissue identifies genes associated with COPD. *Epigenetics*. 2016;11(10):730-739. doi: 10.1080/15592294.2016.1226451
67. Holloway JW, Meta-analysis GC, Joubert BR, et al. DNA Methylation in Newborns and Maternal Smoking in Pregnancy : Genome-wide Consortium Meta-analysis ARTICLE DNA

Methylation in Newborns and Maternal Smoking in Pregnancy : *Am J Hum Genet*. 2016; 98(April):680-696. doi:10.1016/j.ajhg.2016.02.019

68. Yoo S, Takikawa S, Geraghty P, et al. Integrative Analysis of DNA Methylation and Gene Expression Data Identifies EPAS1 as a Key Regulator of COPD. *PLoS Genet*. 2015;11(1):e1004898. doi:10.1371/journal.pgen.1004898
69. Hackett NR, Butler MW, Shaykhiev R, et al. RNA-Seq quantification of the human small airway epithelium transcriptome. *BMC Genomics*. 2012;13(1):82. doi:10.1186/1471-2164-13-82
70. Beane J, Sebastiani P, Liu G, Brody JS, Lenburg ME, Spira A. Reversible and permanent effects of tobacco smoke exposure on airway epithelial gene expression. *Genome Biol*. 2007;8(9): R201. doi:10.1186/gb-2007-8-9-r201
71. Bahr TM, Hughes GJ, Armstrong M, et al. Peripheral blood mononuclear cell gene expression in chronic obstructive pulmonary disease. *Am J Respir Cell Mol Biol*. 2013;49(2):316-323. doi: 10.1165/rcmb.2012-0230OC
72. Singh D, Fox SM, Tal-Singer R, et al. Induced sputum genes associated with spirometric and radiological disease severity in COPD ex-smokers. *Thorax*. 2011;66(6):489-495. doi:10.1136/thx.2010.153767
73. Almansa R, Socias L, Sanchez-Garcia M, et al. Critical COPD respiratory illness is linked to increased transcriptomic activity of neutrophil proteases genes. *BMC Res Notes*. 2012;5:401. doi:10.1186/1756-0500-5-401
74. Masuno K, Haldar SM, Jeyaraj D, et al. Expression profiling identifies klf15 as a glucocorticoid target that regulates airway hyperresponsiveness. *Am J Respir Cell Mol Biol*. 2011;45(3):642-649. doi:10.1165/rcmb.2010-0369OC
75. Misior AM, Deshpande DA, Loza MJ, Pascual RM, Hipp JD, Penn RB. Glucocorticoid- and protein kinase A-dependent transcriptome regulation in airway smooth muscle. *Am J Respir Cell Mol Biol*. 2009;41(1):24-39. doi:10.1165/rcmb.2008-0266OC
76. Nobakht M Gh BF, Aliannejad R, Rezaei-Tavirani M, Taheri S, Oskouie AA. The metabolomics of airway diseases, including COPD, asthma and cystic fibrosis. *Biomarkers Biochem Indic Expo response, susceptibility to Chem*. 2015;20(1):5-16. doi:10.3109/1354750X.2014.983167
77. Chen Q, Deeb RS, Ma Y, Staudt MR, Crystal RG, Gross SS. Serum metabolite biomarkers discriminate healthy smokers from COPD smokers. *PLoS One*. 2015;10(12):e0143937. doi:10.1371/journal.pone.0143937
78. Adamko DJ, Nair P, Mayers I, Tsuyuki RT, Regush S, Rowe BH. Metabolomic profiling of asthma and chronic obstructive pulmonary disease: A pilot study differentiating diseases. *J Allergy Clin Immunol*. 2015;136(3):571-580.e3. doi:10.1016/j.jaci.2015.05.022
79. Ubhi BK, Riley JH, Shaw PA, et al. Metabolic profiling detects biomarkers of protein degradation in COPD patients. *Eur Respir J*. 2012;40(2):345-355. doi:10.1183/09031936.00112411
80. Bowler RP, Jacobson S, Cruickshank C, et al. Plasma Sphingolipids Associated with Chronic Obstructive Pulmonary Disease Phenotypes. *Am J Respir Crit Care Med*. 2015;191(3):275-284. doi:10.1164/rccm.201410-1771OC
81. Telenga ED, Hoffmann RF, T'Kindt R, et al. Untargeted lipidomic analysis in chronic obstructive pulmonary disease uncovering sphingolipids. *Am J Respir Crit Care Med*. 2014;190(2): 155-164. doi:10.1164/rccm.201312-2210OC
82. Davey Smith G, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease?*. *Int J Epidemiol*. 2003;32(1):1-22. doi:10.1093/ije/dyg070

83. Lawlor DA, Harbord RM, Sterne JAC, Timpson N, Davey Smith G. Mendelian randomization: Using genes as instruments for making causal inferences in epidemiology. *Stat Med*. 2008; 27(8):1133-1163. doi:10.1002/sim.3034

84. Millwood IY, Walters RG, Mei XW, et al. Conventional and genetic evidence on alcohol and vascular disease aetiology: a prospective study of 500 000 men and women in China. *Lancet (London, England)*. 2019;393(10183):1831-1842. doi:10.1016/S0140-6736(18)31772-0

85. Hobbs BD, Parker MM, Chen H, et al. Exome array analysis identifies a common Variant in IL27 associated with chronic obstructive pulmonary disease. *Am J Respir Crit Care Med*. 2016; 194(1):48-57. doi:10.1164/rccm.201510-2053OC

Chapter 2

Omics studies of COPD and lung function

Chapter 2.1

A genome-wide linkage study for chronic obstructive pulmonary disease in a Dutch genetic isolate identifies novel rare candidate variants

Ivana Nedeljković, Natalie Terzikhan, Judith M. Vonk, Diana A. van der Plaat, Lies Lahousse, Cleo C. van Diemen, Brian D. Hobbs, Dandi Qiao, Michael H. Cho, Guy G. Brusselle, Dirkje S. Postma, H. Marike Boezen, Cornelia M. van Duijn, Najaf Amin

This chapter is published in Frontiers in Genetics, April 2018.
The supplemental material for this paper is available at:
https://doi.org/10.3389/fgene.2018.00133

ABSTRACT

Chronic obstructive pulmonary disease (COPD) is a complex and heritable disease, associated with multiple genetic variants. Specific familial types of COPD may be explained by rare variants, which have not been widely studied. We aimed to discover rare genetic variants underlying COPD through a genome-wide linkage scan. Affected-only analysis was performed using the 6K Illumina Linkage IV Panel in 142 cases clustered in 27 families from a genetic isolate, the Erasmus Rucphen Family (ERF) study. Potential causal variants were identified by searching for shared rare variants in the exome-sequence data of the affected members of the families contributing most to the linkage peak. The identified rare variants were then tested for association with COPD in a large meta-analysis of several cohorts.

Significant evidence for linkage was observed on chromosomes 15q14-15q25 (log of odds (LOD) score=5.52), 11p15.4-11q14.1 (LOD=3.71) and 5q14.3-5q33.2 (LOD=3.49). In the chromosome 15 peak, that harbors the known COPD locus for nicotinic receptors, and in the chromosome 5 peak we could not identify shared variants. In the chromosome 11 locus, we identified four rare (minor allele frequency (MAF) <0.02), predicted pathogenic, missense variants. These were shared among the affected family members. The identified variants localize to genes including neuroblast differentiation-associated protein (*AHNAK*), previously associated with blood biomarkers in COPD, phospholipase C Beta 3 (*PLCB3*), shown to increase airway hyperresponsiveness, solute carrier family 22-A11 (*SLC22A11*), involved in amino acid metabolism and ion transport, and metallothionein-like protein 5 (*MTL5*), involved in nicotinate and nicotinamide metabolism. Association of *SLC22A11* and *MTL5* variants were confirmed in the meta-analysis of 9,888 cases and 27,060 controls.

In conclusion, we have identified novel rare variants in plausible genes related to COPD. Further studies utilizing large sample whole-genome sequencing should further confirm the associations at chromosome 11 and investigate the chromosome 15 and 5 linked regions.

INTRODUCTION

Chronic obstructive pulmonary disease (COPD) is a common and complex disease, and one of the leading causes of death worldwide.[1] Previous studies provided heritability estimates for COPD of 20% to even 60%.[2,3] Both rare variants with a large impact and common variants with a modest impact on the risk to develop COPD have been identified. The *SERPINA1* gene at chromosome 14q32.13, encoding Alpha-1-antitrypsin (AAT), was in fact the first gene identified to be associated with COPD.[4,5] Rare variants in *SERPINA1* are known to contribute to COPD risk in AAT deficiency in homozygous and heterozygous carriers of the low-frequency Z allele.[6] In an exome study of severe, early-onset families, Qiao et al identified several genes with rare variants segregating in at least two pedigrees.[7] In extended families, genetic linkage studies have found evidence of linkage to chromosomes 2q, 6q, 8p, 12p and 19q, among others.[8,9] However, many initially promising findings from linkage or exome sequencing in candidate-gene studies could not be replicated in subsequent analyses.[10]

Common variants in several genes have been identified in multiple genome-wide association studies (GWAS), to be associated with COPD or obstructive lung function impairment. Among consistently replicated loci in GWAS are genes on chromosome 4 – Hedgehog-interacting protein (*HHIP*) and Family with sequence similarity 13 member A (*FAM13A*), chromosome 5 – 5-hydroxytryptamine receptor 4 (*HTR4*), chromosome 15 - Nicotinic cholinergic receptors (*CHRNA3/5*) and Ion-responsive element binding protein 2 (*IREB2*) and chromosome 19 – Cytochrome P450 family gene (*CYP2A6*), member RAS oncogene family gene (*RAB4B*) and Egl-9 family hypoxic-inducible factor 2 (*EGLN2*).[11,12] However, only few loci identified in GWAS could be functionally explained.

Despite the undeniable progress in understanding the genetic origins of COPD, a major part of its heritability remains unexplained. A complicating factor in studies on the genetics of COPD is that COPD is considered a complex genetic trait, i.e. multiple, possibly interacting, genetic and environmental factors are involved. Therefore, there is a need for fine mapping techniques that can identify functional, rare variants with large effects explaining specific types of COPD. Rare variant association studies can be carried out in relatively small sample sizes when using family-based settings.[13] In a genetically isolated population, alleles that are found at low or very-low (rare) frequencies in control samples may reach much higher proportions due to a limited number of founder individuals, genetic drift, minimal immigration and high inbreeding.[14] Therefore, attempting to identify risk genes for

COPD in populations that are relatively genetically and environmentally homogeneous could be beneficial.[15]

This study uses the Erasmus Rucphen Family (ERF) study, a Dutch genetically isolated population, to localize and identify rare genetic variants and subsequently shows the relevance of these variants in the general population by performing an association analysis in a large sample.

METHODS

Study populations

Linkage study

The linkage study was performed in 142 related participants from the ERF study. ERF is a family-based cohort study, studied as part of the Genetic Research in Isolated Population (GRIP) program. It is based in a genetically isolated community from the south-west area of the Netherlands, set up to investigate genes underlying different quantitative traits and common diseases.[14] The participants of ERF are living descendants of 22 couples from the religious isolate in the 19th century, who had at least six children baptized in the community church. The baseline data collection for over 3,000 people was conducted between June 2002 and February 2005. These individuals are related to each other through multiple lines of descent in a single large pedigree spanning 23 generations and connecting over 23,000 individuals. In 2015 a follow-up data collection for 1,500 participants was performed by reviewing general practitioner's records, including letters from the specialists and spirometry reports and medication use. In total 192 probable COPD cases were identified in the follow-up. The COPD diagnosis was confirmed by respiratory specialists based on an obstructive lung function, i.e. the ratio of Forced Expiratory Volume in one second over the Forced Vital Capacity (FEV_1/FVC) <0.7, with or without medication use (n=116). If the information on FVC was missing (n=14), the following criteria for COPD were used: FEV_1 <80%, use of respiratory medication and a COPD diagnosis in the report of the respiratory specialist to the general practitioner. If no lung function measurement was available (n=15), COPD diagnosis was based on: medication use with CT-scan of the lungs indicating COPD and/or a history of frequent COPD exacerbations mentioned in the medical documents. Thus, the COPD diagnosis could be confirmed for 145 participants, of which three did not have genotyping data, resulting in the final sample size for the linkage study of 142 COPD cases.

Association Study

The association analysis was performed using data from the Rotterdam Study (RS; 1,588 cases and 9,784 controls), the Lifelines study (LLS; 1,647 cases and 9,530 controls), the Vlagtwedde/Vlaardingen-study (VlaVla; 375 cases and 1,019 controls) and the data from the study of Hobbs et al.[16] (6,161 cases and 6,004 controls), in addition to the ERF study (117 cases and 1,091 controls).

RS is a prospective, population-based study,[17] focusing on the diseases in the participants aged 45 or older. The COPD diagnosis in the RS was defined as having pre-bronchodilator obstructive spirometry ($FEV_1/FVC<0.7$), assessed either by spirometry in the research center or by reviewing medical histories of the participants. Spirometry was performed by trained paramedical personnel, according to the guidelines of the American Thoracic Society/European Respiratory Society (ATS/ERS). In absence of interpretable spirometry measures, all medical information of subjects regularly using respiratory medication was reviewed, including files from specialists and general practitioners, to confirm a diagnosis of COPD. Both ERF and RS have been approved by the Medical Ethics Committee of the Erasmus Medical Center. All participants provided written informed consent to participate in the study and to obtain information from their treating physicians.

LLS is a multi-disciplinary prospective population-based cohort of the Northern provinces of the Netherlands with a three generation design, focusing on the onset of common complex diseases.[18] COPD was defined as having pre-bronchodilator $FEV_1/FVC<0.7$, assessed by spirometry using a Welch Allyn Version 1.6.0.489, PC-based SpiroPerfect with Ca Workstation software. All subjects provided written informed consent and the study was approved by the Medical Ethics Committee of the University Medical Center Groningen, Groningen, the Netherlands.

The Vlagtwedde/Vlaardingen study is a prospective, Dutch population-based cohort including individuals from Vlagtwedde (a rural area) and Vlaardingen (an urban area), aimed to gain insight into the risk factors for chronic airway diseases and lung function.[19] COPD was defined as having pre-bronchodilator $FEV_1/FVC<0.7$. Data of the last survey in 1989/1990 were used and spirometry data were collected by performing a slow inspiratory maneuver, using a water-sealed spirometer (Lode instruments, Groningen, the Netherlands). The Committee on Human Subjects in Research of the University of Groningen reviewed the study and affirmed the safety of the protocol and study design and all participants gave their written informed consent.

In the study by Hobbs at al.[16] COPD cases were defined as having $FEV_1/FVC\leq0.7$ and $FEV_1\leq 80\%$ of the predicted value. It was multi-ethnic study with Asian, African, and European ancestry individuals. Institutional review board approval and written informed consent were obtained for all these cohorts. For more details please refer to their publication.[16]

Genotyping

DNA isolation

For all participants, DNA was extracted from venous blood using the salting out method.[20]

Linkage array

For the linkage analysis genotyping was performed using the 6K Illumina Linkage IV panel (Illumina, San Diego, CA, USA). Further, quality control (QC) was performed involving exclusion of the variants with call rate <98%, those diverging from Hardy-Weinberg equilibrium ($P<10^{-8}$) and X-chromosome variants and participants with an overall call rate <96%. Mendelian inconsistencies were designated as missing genotypes. The final dataset comprised 5,250 autosomal single nucleotide variants (SNVs) in 3,018 participants.

Exome-sequencing and genotyping

The sequencing and genotyping in the ERF study have been described elsewhere.[21] In short, for 1,336 ERF participants whole exome sequencing was performed at a mean depth of 74x (Agilent, v4 capture). After QC, 543,954 SNVs in 1,327 participants were retained. For 1,527 individuals whose exomes were not sequenced, the Illumina Infinium HumanExome BeadChip v1.1 was used for genotyping and variant calling was done using Genome Studio. After QC 70,000 polymorphic SNVs in 1,515 participants were retrieved. Of these, the overlap with COPD status information, was available for 636 participants (59 cases and 577 controls) with exome-sequence and 572 participants (58 cases and 514 controls) with exome-chip data. The cases overlap with the sample used in the linkage analysis. The ERF data is available in the EGA public repository (https://www.ebi.ac.uk/ega/home) with ID number: EGAS00001001134.

The Rotterdam Study was genotyped using Illumina 550K and Illumina 610K and 660K arrays, and genotyping QC was done as described elsewhere.[22] Haplotype Reference Consortium imputation panel (HRC)[23] was used for imputation. File preparation and imputation was done as described elsewhere.[22] In the final dataset we included 11,372 participants of RS (cases and controls) with HRC imputed genotype data available.

In LLS and VlaVla the genotyping was done using Illumina CytoSNP-12 arrays and QC was done as described elsewhere.[24] The Genome of the Netherlands (GoNL) panel was used for imputation of LLS and VlaVla and was done as described elsewhere.[18] The final dataset included 11,177 participants of LLS and 1,394 of VlaVla.

In Hobbs et al. work all individuals were genotyped using the Illumina HumanExome arrays (v1.1 and v1.2; Illumina, San Diego, CA). For more information please refer to their publication.[16]

Statistical analyses

Genome-wide linkage analysis

For the genome-wide linkage analysis, 142 related COPD cases from ERF were used. The cases were linked in a single large pedigree of 23 generations. However, due to the linkage software restraints, the cases were clustered into 27 smaller (≤24 bits) families using PEDCUT software.[25] We used Haplopainter[26] to illustrate all 27 pedigrees (Supplementary figure 1). We then performed affected-only parametric linkage analysis in MERLIN software[27] using incomplete penetrance and no phenocopies for both dominant (0, 0.5, 0.5) and recessive models (0, 0, 0.5).[28] The measure of the likelihood of linkage is the logarithm of the odds (LOD) score and we considered LOD≥3.3 to be statistically significant. Further we performed per-family analysis for significant regions to identify the families with COPD cases contributing the most to the LOD score.

Identification of variants in the identified regions

Next, we used exome-sequence data in ERF to identify rare variants that may explain the identified linkage peaks. For this, among all variants in this region we selected only variants with predicted damaging effects on protein (missense and stop-coding) based on the FunctionGVS column of the SeattleSeq Annotation database (http://snp.gs.washington.edu/SeattleSeqAnnotation138/) from the National Heart, Lung and Blood Institute (NHLBI) and with minor allele frequency (MAF) <0.05 in the general population (1000Genomes). As frequencies in a genetically isolated population may be inflated or deflated due to genetic drift,[14] we used the MAF from the general population for filtering. We selected variants shared among most (>50%) of the affected family members as candidate variants.

A formal test of association was performed for the identified candidate variants in each study - ERF, in samples with exome-sequence (N=636) and in exome-chip (N=572) data, in three RS cohorts (RS-I, RS-II and RS-III), using the HRC imputed data (N=11,372), the LLS (N=11,177), the VlaVla cohort (N=1,394) and the Hobbs et al results (N=11,797). For this analysis, in ERF we used "*seqMeta*" package in R[29] to perform single-variant analysis, adjusted for age, sex and smoking status (current/past/never smoking). Logistic regression analysis was used to associate the variants in the RS and the VlaVla cohort, using SPSS software[30] and in LLS, using PLINK,[31]

applying the same models as used in ERF. Variants were excluded from the analysis if the minor allele count was less than five in either the case or the control category. Summary statistics for identified the variants were extracted from the results of Hobbs et al.[16] A fixed-effects meta-analysis was performed with the summary statistics from all studies using the "*rmeta*" package in R.[32]

Functional look-up of the genes

We investigated the Ingenuity Knowledge Base for functional annotation and look up of the genes, harboring the identified variants (IPA, Qiagen bioinformatics).[33] Furthermore, we consulted the Gene network tool,[34] a bioinformatics database containing co-expression data, functional predictions from gene ontology, Biocarta and the Kyoto Encyclopedia of Genes and Genomes (KEGG) to investigate our findings.

RESULTS

The general characteristics of the study samples are presented in **Table 1**.

Table 1. General characteristics of the populations used in this study.

	ERF			RS	Hobbs et al.	Life Lines	Vlagtwedde/ Vlaardingen
	Linkage*	Exome-chip	Exome-sequence	HRC imputed	Exome-chip**	GoNL imputed	GoNL imputed
Number	142	572	636	11,372	12,165	11,177	1,394
Age, mean(sd)	59.7(10.9)	51.7(14.2)	48.5(14.0)	65.1(9.8)	58.4(10.3)	48.2(11.0)	52.7(10.2)
Female gender, %(n)	59.9(85)	56.8(325)	61.8(393)	58.0(6,592)	44.5(5,410)	58.6(6,547)	46.3(646)
COPD cases, %(n)	100(142)	10.1(58)	9.3(59)	14.0(1,588)	50.6(6,161)	14.7(1,647)	26.9(375)
Never smokers, %(n)	1.4(2)	27.1(155)	29.4(187)	35.3(4,011)	1.7(212)	40.7(4,549)	30.2(421)
Ex-smokers, %(n)	23.2(33)	27.8(159)	28.8(183)	48.8(5,546)	49.6(6,037)	36.7(4,104)	33.1(462)
Current smokers, %(n)	58.5(83)	45.1(258)	41.8(266)	16.0(1,815)	45.0(5,473)	22.6(2,524)	36.7 (511)

*Information on smoking was missing for 16.9% (24) participants; ** Full dataset reported in the Hobbs et al meta-analysis. Information on smoking was missing for 3.6% (443) participants.

All 27 families included in the linkage analyses in ERF are depicted in supplementary **Figure S1**. The affected relatives were mainly smokers: 81.7% of the cases included in the linkage analyses were current or ex-smokers. As shown in **Table 2** and **Figure 1**, we identified significant evidence for linkage of COPD to chromosomes 15q14-15q25 (Heterogeneity LOD score - HLOD=5.52), 11p15.4-11q14.1 (HLOD=3.71) and 5q14.3-5q33.2 (HLOD=3.49).

Table 2. Genome wide significant (HLOD>3.3) results of linkage analysis in the ERF study

Cytogenetic location*	Start SNP	End SNP	SNP with highest HLOD	Start position#	End position#	Dominant model HLOD	Recessive model HLOD
15q14-15q25	rs2004175	rs1402760	rs383902	39039593	79146817	4.24	5.52
11p15.4-11q14.1	rs1609812	rs7102569	rs626333	5247141	79184899	2.61	3.71
5q14.3-5q33.2	rs1366133	rs1432812	rs1154308	91114584	155274700	2.65	3.49

* Region under the linkage peak; Start SNP – single nucleotide polymorphism (SNP) at the beginning of the corresponding region; End SNP – SNP at the end of the corresponding region; HLOD - Heterogeneity log of odds score; #Corresponding to the region from base to base of the linkage peak, based on the hg19 assembly.

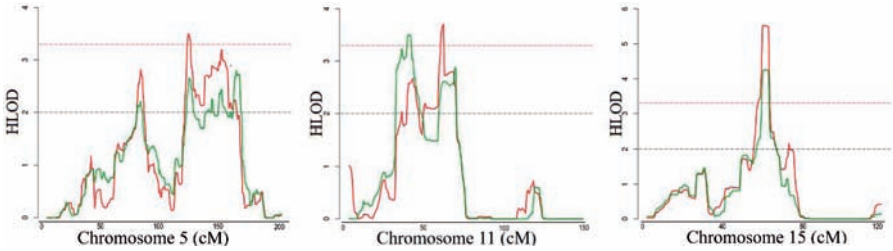

Figure 1. LOD score plot for the regions at (A) chromosomes 5, (B) 11 and (C) 15. X-axis shows the chromosomal position in cM and the Y-axis shows the HLOD score. Red line represents HLOD scores for recessive and green line for dominant model. Dashed red line represents the level of significance (HLOD=3.3), while dashed black line represents the suggestive level (HLOD=2).

We next searched for rare, deleterious and shared variants by most (>50%) of the affected family members in the three identified regions mentioned above. In the linked regions of chromosomes 5 and 15 we could not identify any variants that passed mentioned filtering criteria. For the linked region on chromosome 11, we identified two families that were contributing most (LOD>1) to the linkage score (**Figure 2**). Exome-sequence data were available for 8 of 17 COPD cases from these two families.

We identified four missense variants including rs116243978 (*AHNAK*), rs35169799 (*PLCB3*), rs141159367 (*SLC22A11*) and rs146043252 (*MTL5*), shared among five of the eight affected family members (**Table 3**).

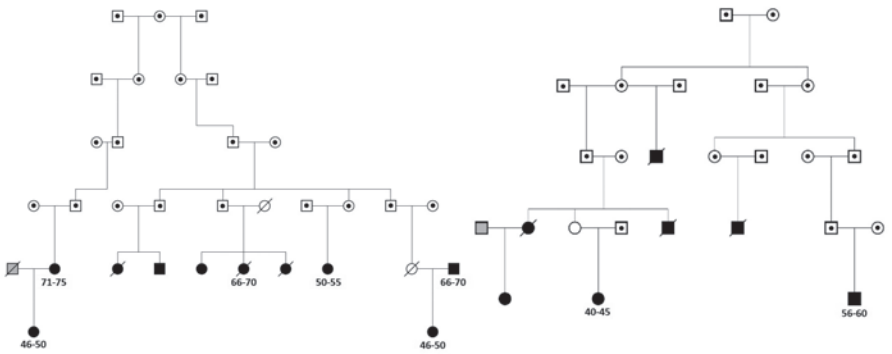

Figure 2. The two sub-families contributing most to the linkage peak on chromosome 11. Squares represent males and circles females. Cases are denoted in black, known controls are denoted in grey and the family members for which we do not have COPD information are denoted in white. Family members with dot in the middle are not included in ERF study and for them only pedigree information was available. Deceased family members are crossed. For cases with exome-sequence data used in the sharing analysis information on 5-year age range (in years) is provided.

Table 3. Deleterious variants from chromosome 11q (missense, stop codon or CADD > 15) with a frequency in the 1000 genomes <0.05 that are shared by at least 5 cases.

Gene	Variant	1KG MAF	ERF MAF	Cytogenetic band	Position (hg19)	A1	A2	Carrier-HET	Carrier-HOM	Function	CADD	PolyPhen
AHNAK	rs116243978	0.005	0.04	11q12.3	62286165	G	C	5/8	0/8	missense	15.55	1
PLCB3	rs35169799	0.023	0.08	11q13.1	64031241	T	C	6/8	1/8	missense	15.73	0.982
SLC22A11	rs141159367	0.0006	0.04	11q13.1	64323476	T	C	5/8	1/8	missense	18.25	1
MTL5	rs146043252	0.0002	0.04	11q13.3	68478487	G	A	5/8	0/8	missense	21	1

1KG MAF – minor allele (A1) frequency (MAF) in 1000 Genomes - EUR; A1: effect allele; A2: alternative allele; Carrier-HET: number of heterozygote carriers within the 8 COPD cases of the two top contributing families of the chromosome 11 region; Carrier-HOM: number of homozygote carriers within the 8 COPD cases; Function: predicted function of the variant; CADD: Combined Annotation Dependant Depletion score (>15 considered deleterious); PolyPhen: probability that variant is damaging.

Each of these variants was predicted to be highly pathogenic (Combined Annotation Dependent Depletion score, CADD>15, PolyPhen>0.98) which suggests their relevance for the disease development. Of these four variants, one (rs141159367 in *SLC22A11)* showed a significant association (OR=1.87, *P*=0.002) with COPD in the meta-analysis (**Table 4**). The variant rs146043252 in *MTL5* showed a nominal association signal (OR=1.66, *P*=0.04).

Table 4. Results of association analysis with COPD

Gene	Variant	β	OR	SE	P	N
AHNAK	rs116243978	0.14	1.15	0.18	0.422	13,402
PLCB3	rs35169799	0.05	1.05	0.04	0.247	36,948
SLC22A11	rs141159367	0.63	1.87	0.20	**0.002**	18,562
MTL5	rs146043252	0.51	1.66	0.25	**0.044**	12,050

Meta-analysis of the ERF exome-sequence, ERF exome-chip, Rotterdam Study – HRC, Hobbs et al. exome-chip, LifeLines – GoNL and Vlagtwedde/Vlaardingen cohort - GoNL results. In bold: significant results. β: Regression coefficient estimates from logistic regression model regressing COPD status on indicated variants, corrected for sex, age and smoking; OR: odds ratio; SE: standard error of the effect; P: p-value of the significance; N-sample size used in the analysis.

DISCUSSION

In this study, we found significant evidence for extensive linkage of COPD to the chromosomes 15q14-15q25 (40.1Mb), 11p15.4-11q14.1 (73.9Mb) and 5q14.3-5q33.2 (64.1Mb). We were able to identify four rare and predicted pathogenic variants under the chromosome 11 peak, in plausible genes (*AHNAK, PLCB3, SLC22A11* and *MTL5*), shared by at least five family members. One of these four variants, i.e. rs141159367 in *SLC22A11,* was significantly associated with COPD in 9,888 cases and 27,428 controls (*P*=0.002) while another variant (rs146043252 in *MTL5*) showed nominal association with COPD (*P*=0.04).

The finding of our family-based linkage analysis aligns with that of large scale GWASs implicating the *CHRNA3/5-CHRNB4* and *IREB2* region on chromosome 15q25 in COPD development. This region is also associated with lung cancer, peripheral arterial disease, nicotine addiction and smoking quantity.[35] The evidence in the literature on the role of smoking in the genetic risk of COPD thus far is controversial. On one hand, there is evidence to support that the variants in this region, although implicated in both lung disease and smoking behavior, are associated with COPD susceptibility, independently of cigarette smoke exposure.[36] On the other hand, in a previous study we show that two variants, previously associated with COPD in the *CHRNA3/5* locus, were associated with lung function measurements in ever-smokers, but not in never-smokers,[37] which is in line with the only longitudinal study on the relation between the nicotine receptor variant and annual lung function decline.[38] That study shows that carriers of the nicotinic receptors variants are significantly less able to quit smoking, leading to the lung function decline and, subsequently to COPD. Similarly, for the chromosome 5 linked region, we could not observe any shared rare variant. This region, known for its associations with pulmonary func-

tion and airflow obstruction[39,40] was recently associated with COPD by the largest GWAS to date.[11] The *HTR4* gene in 5q32 encodes a serotonin receptor involved in depression and is strongly expressed in respiratory complex neurons.[41]

However, the functional variants in these regions have still not been confirmed. In our families, we could not identify rare damaging variants shared between the cases in this region. This may be explained if rare intronic regulatory variants play a key role, which we could not investigate using the exome data. It is unlikely that these linkage peaks are attributed to the common variants which have small effects identified in GWASs, given the very strong evidence for linkage of this region to COPD. Future studies using whole-genome sequencing should investigate this region further, ideally in never smokers. This emphasizes the need for integration of available genomic information into more focused, candidate-gene based efforts to disentangle the functional role of the chromosome 5 and 15 regions.

In the identified region of chromosome 11 we were able to pinpoint four strong candidate genes for the association with COPD, i.e. *SLC22A11, AHNAK, PLCB3,* and *MTL5*. The most interesting finding is the rare variant in *SLC22A11* (solute carrier family 22 member 11), which encodes an integral membrane protein and part of the family of organic anion transporters (OATs), known to mediate the absorption and elimination of endogenous and exogenous organic anions and as such, are involved in the pharmacokinetic, pharmacodynamic and safety profiles in a wide range of drugs.[42] *SLC22A11* (OAT4) is mainly expressed in kidney and placenta. However, it is also shown to be expressed in lung tissue, fibroblasts and T-lymphocytes ($P<5\times10^{-7}$), among other tissues/cells reported in the Gene network.[34] In addition, *in vitro SLC22A11* mRNA was absent in normal human bronchial epithelial cells, but highly expressed in other bronchial cells models comprising transformed cells.[43] *SLC22A11* in particular is known to be a drug target for probenecid, a *SLC22A11* inhibitor, used in the gout prevention and to increase antibiotic blood levels, yet its direct role in lung disease treatment is still unknown.[42]

Our linkage analysis yielded different regions compared with those identified earlier. However, the fact that both *SLC22A11* and *MTL5* variants were associated with COPD in our meta-analysis confirms their role in COPD and makes them even more interesting candidates. *MTL5* (metallothionein-like protein 5) encodes testis expressed metallothionein like proteins (TESMIN*).* They are highly conserved, low-molecular-weight cysteine-rich proteins induced by and binding to heavy metal ions, and they do not have enzymatic activity. They play a central role in the regulation of cell growth and differentiation, and are involved in spermatogenesis, differentially regulating meiosis in male and female cells.[44] *MTL5* was shown to be involved in nicotinate and nicotinamide metabolism and is also expressed in fibroblasts and

lung tissue ($P<7\times10^{-29}$), based on the Gene network.[34] Metallothioneins were additionally shown to protect cells against oxidative stress damage and participate in differentiation, proliferation and/or apoptosis of normal and lung cancer cells.[45]

The main strength of our study is the genetically isolated family-based population, which can display increased frequencies of some variants found at very low proportions in panmictic populations. This allowed us to perform a genome-wide linkage scan and identify rare coding variants. However, even though we identified linkage of three regions to COPD, a limitation of our study is the low power to explain the peaks at chromosomes 5 and 15, possibly due to the use of exome data. As intronic regulatory variants may play a significant role, in the future, faster and cheaper whole-genome sequencing will allow us to improve identification of rare variants and our understanding of their involvement in COPD. As our sample consists of high percentage of current or ex-smokers, it is possible that we are demonstrating genetic effects on smoking which further affects the development of COPD. Nevertheless, we were able to demonstrate a positive association, independent of smoking, of two variants in the association meta-analysis comprising 9,888 cases and 27,060 controls. Yet, studies with very large sample sizes utilizing mediation or mendelian randomization techniques are needed to disentangle these relationships and confirm our results in the general population.

To conclude, using the powerful genome-wide linkage scan in a Dutch genetic isolate, we have confirmed the implication of the 15q25 region in COPD and identified regions at chromosomes 5 and 11. Within the region on chromosome 11 we identified four deleterious rare variants shared between most of the affected family members in *AHNAK, PLCB3, SLC22A11* and *MTL5.* The variants in *SLC22A11* and *MTL5* were significantly associated with COPD in our meta-analysis. Further studies pooling large sample sizes could confirm the role of the identified rare variants at chromosome 11 in the general population. Similarly, large studies utilizing whole-genome sequencing should further investigate the role of linked regions in chromosomes 5 and 15 in COPD.

REFERENCES

1. Lozano R, Naghavi M FK. Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet*. 2012;380(380):2095–2128. doi:https://doi.org/10.1016/S0140-6736(12)61728-0
2. Ingebrigtsen T, Thomsen SF, Vestbo J, et al. Genetic influences on Chronic Obstructive Pulmonary Disease. *a Twin Study*. 2010;104(12):1890-1895. doi:10.1016/j.rmed.2010.05.004
3. Zhou JJ, Cho MH, Castaldi PJ, Hersh CP, Silverman EK, Laird NM. Heritability of Chronic Obstructive Pulmonary Disease and Related Phenotypes in Smokers. *Am J Respir Crit Care Med*. 2013;188(8):941-947. doi:10.1164/rccm.201302-0263OC [doi]
4. Laurell CB ES, Laurell CB ES. The electrophoretic alpha 1-globulin pattern of serum in alpha 1-antitrypsin deficiency. *Scand J Clin Lab Invest*. 1963;15(2):132-140. doi:https://doi.org/10.1080/00365516309051324
5. Bashir, A., Shah, N.N., Hazari, Y.M., Habib, M., Bashir, S., Hilal, N., Banday, M., Asrafuzzaman, S., and Fazili KM. Novel variants of SERPIN1A gene: Interplay between alpha1- antitrypsin deficiency and chronic obstructive pulmonary disease. *Respir Med*. 2016;117:139-149. doi:https://doi.org/10.1016/j.rmed.2016.06.005
6. Foreman MG, Wilson C, DeMeo DL, et al. Alpha-1 antitrypsin PiMZ genotype is associated with chronic obstructive pulmonary disease in two racial groups. *Ann Am Thorac Soc*. 2017;14(8):1280-1287. doi:10.1513/AnnalsATS.201611-838OC
7. Qiao D, Lange C, Beaty TH, Crapo JD, Barnes KC, Bamshad M. Exome Sequencing Analysis in Severe , Early-Onset Chronic Obstructive Pulmonary Disease. *Am J Respir Crit Care Med*. 2016;193(12):1-61. doi:https://doi.org/10.1164/rccm.201506-1223OC
8. Silverman EK, Mosley JD, Palmer LJ, et al. Genome-wide linkage analysis of severe , early-onset chronic obstructive pulmonary disease : airflow obstruction and chronic bronchitis phenotypes. *Hum Mol Genet*. 2002;11(6):623-632. http://dx.doi.org/10.1093/hmg/11.6.623.
9. Palmer LJ, Celedón JC, Chapman HA, Speizer E. FE, Weiss ST, Silverman EK. Genome-wide linkage analysis of bronchodilator responsiveness and post-bronchodilator spirometric phenotypes in chronic obstructive pulmonary disease. *Hum Mol Genet*. 2003;12(10):1199-1210. doi:10.1093/hmg/ddg125
10. Hersh CP, DeMeo DL, Lange C, et al. Attempted replication of reported chronic obstructive pulmonary disease candidate gene associations. *Am J Respir Cell Mol Biol*. 2005;33(1):71-78. doi:10.1165/rcmb.2005-0073OC
11. Hobbs BD, De Jong K, Lamontagne M, et al. Genetic loci associated with chronic obstructive pulmonary disease overlap with loci for lung function and pulmonary fibrosis. *Nat Genet*. 2017;49(3):426-432. doi:10.1038/ng.3752
12. Wain L V., Shrine N, Artigas MS, et al. Supplementary: Genome-wide association analyses for lung function and chronic obstructive pulmonary disease identify new loci and potential druggable targets. *Nat Genet*. 2017;49(3):416-425. doi:10.1038/ng.3787
13. Auer PL, Lettre G. Rare variant association studies: Considerations, challenges and opportunities. *Genome Med*. 2015;7(1):16. doi:10.1186/s13073-015-0138-2
14. Pardo LM, MacKay I, Oostra B, van Duijn CM, Aulchenko YS. The effect of genetic drift in a young genetically isolated population. *Ann Hum Genet*. 2005;69(3):288-295. doi:10.1046/j.1529-8817.2005.00162.x

15. Van Diemen CC, Postma DS, Aulchenko YS, et al. Novel strategy to identify genetic risk factors for COPD severity: A genetic isolate. *Eur Respir J*. 2010;35(4):768-775. doi: 10.1183/09031936.00054408
16. Hobbs BD, Parker MM, Chen H, et al. Exome array analysis identifies a common Variant in IL27 associated with chronic obstructive pulmonary disease. *Am J Respir Crit Care Med*. 2016; 194(1):48-57. doi:10.1164/rccm.201510-2053OC
17. Ikram MA, Brusselle GGO, Murad SD, et al. The Rotterdam Study: 2018 update on objectives, design and main results. *Eur J Epidemiol*. 2017;32(9):807-850. doi:10.1007/s10654-017-0321-4
18. Scholtens S, Smidt N, Swertz MA, et al. Cohort Profile: LifeLines, a three-generation cohort study and biobank. *Int J Epidemiol*. 2015;44(4):1172-1180. doi:10.1093/ije/dyu229
19. Van Diemen CC, Postma DS, Vonk JM, Bruinenberg M, Scheuten JP, Boezen HM. A disintegrin and metalloprotease 33 polymorphisms and lung function decline in the general population. *Am J Respir Crit Care Med*. 2005;172(3):329-333. doi:10.1164/rccm.200411-1486OC
20. S.A.Miller DDD and HFP. A simple salting out procedure for extractin DNA from humam nnucleated cells. *Nucleic Acids Res*. 1988;15(3):1215.
21. Amin N, de Vrij FMS, Baghdadi M, et al. A rare missense variant in RCL1 segregates with depression in extended families. *Mol Psychiatry*. 2017. doi:10.1038/mp.2017.49
22. Iglesias AI, van der Lee SJ, Bonnemaijer PWM, et al. Haplotype reference consortium panel: Practical implications of imputations with large reference panels. *Hum Mutat*. 2017;38(8): 1025-1032. doi:10.1002/humu.23247
23. McCarthy S, Das S, Kretzschmar W, Durbin R, Abecasis G, Marchini J. A reference panel of 64,976 haplotypes for genotype imputation. *bioRxiv*. 2015;48(10):035170. doi:10.1101/035170
24. De Jong K, Vonk JM, Timens W, et al. Genome-wide interaction study of gene-by-occupational exposure and effects on FEV,levels. *J Allergy Clin Immunol*. 2015;136(6):1664-1672e14. doi: 10.1016/j.jaci.2015.03.042
25. Liu F, Kirichenko A, Axenovich TI, van Duijn CM, Aulchenko YS. An approach for cutting large and complex pedigrees for linkage analysis. *Eur J Hum Genet*. 2008;16(7):854-860. doi: 10.1038/ejhg.2008.24
26. Thiele H, Nürnberg P. HaploPainter: A tool for drawing pedigrees with complex haplotypes. *Bioinformatics*. 2005;21(8):1730-1732. doi:10.1093/bioinformatics/bth488
27. Abecasis GR, Cherny SS, Cookson WO, Cardon LR. Merlin—rapid analysis of dense genetic maps using sparse gene flow trees. *Nat Genet*. 2002;30(1):97-101. doi:10.1038/ng786
28. Durner M, Vieland VJ, Greenberg DA. Further evidence for the increased power of LOD scores compared with nonparametric methods. *Am J Hum Genet*. 1999;64(1):281-289. doi: 10.1086/302181
29. Voorman A, Brody J, Lumley T. skatMeta: an R Package for meta analyzing region-based tests of rare DNA variants June. *R Packag version*. 2013;1:1-17.
30. Norušis MJ. *SPSS/PC+ Professional Statistics Version 5.0*. SPSS, Inc.; 1992.
31. Purcell S, Neale B, Todd-Brown K, et al. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *Am J Hum Genet*. 2007;81(3):559-575. doi:10.1086/519795
32. Lumley T. rmeta: Meta-analysis. R package version 2.16. 2009. *Cited on*. 2011.
33. Ipa Q. Redwood City, CA. *CA Available online http//www.ingenuity.com/products/ipa*. 2015.
34. Fehrmann RSN, Karjalainen JM, Westra H, et al. Gene expression analysis identifies global gene dosage sensitivity in cancer. *Nat Genet*. 2015;47(2):115-126. doi:10.1038/ng.3173

35. Thorgeirsson TE, Geller F, Sulem1 P, et al. A variant associated with nicotine dependence, lung cancer and peripheral arterial disease. *Nat*. 2008;452(7187):638-642. doi:10.1038/nature06846
36. Hardin M, Silverman EK. Chronic Obstructive Pulmonary Disease Genetics : A Review of the Past and a Look Into the Future. *J COPD Found*. 2014;1(1):33-46. doi:10.15326/jcopdf.1.1.2014.0120
37. van der Plaat DA, de Jong K, Lahousse L, et al. Genome-wide association study on the FEV_1/FVC ratio in never-smokers identifies HHIP and FAM13A. *J Allergy Clin Immunol*. 2017;139(2):533-540. doi:10.1016/j.jaci.2016.06.062
38. Budulac SE, Vonk JM, Postma DS, Siedlinski M, Timens W, Boezen MH. Nicotinic acetylcholine receptor variants are related to smoking habits, but not directly to COPD. *PLoS One*. 2012;7(3):e33386. doi:10.1371/journal.pone.0033386
39. Hancock DB, Eijgelsheim M, Wilk JB, et al. Meta-analyses of genome-wide association studies identify multiple loci associated with pulmonary function. *Nat Genet*. 2010;42(1):45-52. doi:10.1038/ng.500
40. Wilk JB, Shrine NRG, Loehr LR, et al. Genome-wide association studies identify CHRNA5/3 and HTR4 in the development of airflow obstruction. *Am J Respir Crit Care Med*. 2012;186(7):622-632. doi:10.1164/rccm.201202-0366OC
41. Manzke T, Guenther U, Ponimaskin EG, et al. 5-HT4(a) receptors avert opioid-induced breathing depression without loss of analgesia. *Science (80-)*. 2003;301(5630):226-229. doi:10.1126/science.1084674
42. Bosquillon C. Drug Transporters in the Lung—Do They Play a Role in the Biopharmaceutics of Inhaled Drugs? *J Pharm Sci*. 2010;99(5):4215-4227. doi:10.1002/jps
43. Endter S, Francombe D, Gumbleton M, Ehrhardt C. RT-PCR analysis of ABC, SLC and SLCO drug transporters in human lung epithelial cell models. *J Pharm Pharmacol*. 2009;61(5):583-591. doi:10.1211/jpp/61.05.0006
44. Olesen C, Moller M, Byskov AG. Tesmin Transcription is Regulated Differently during Male and Female Meiosis. *Mol Reprod Dev*. 2004;67(1):116-126. doi:10.1002/mrd.20007
45. Werynska B, Pula B, Kobierzycki C, Dziegiel P, Podhorska-Okolow M. Metallothioneins in the lung cancer. *Folia Histochem Cytobiol*. 2015;53(1):1-10. doi:10.5603/FHC.a2015.0009

Chapter 2.2

Understanding the role of the chromosome 15q25.1 in COPD through epigenetics and transcriptomics

Ivana Nedeljković, Elena Carnero-Montoro, Lies Lahousse, Diana A. van der Plaat, Kim de Jong, Judith M. Vonk, Cleo C. van Diemen, Alen Faiz, Maarten van den Berge, Ma'en Obeidat, Yohan Bossé, David C. Nickle, BIOS consortium, Andre G. Uitterlinden, Joyce B.J. van Meurs, Bruno H.C. Stricker, Guy G. Brusselle, Dirkje S. Postma, H. Marike Boezen, Cornelia M. van Duijn, Najaf Amin

This chapter is published in European Journal of Human Genetics, February 2018.
The supplemental material for this paper is available at:
https://www.nature.com/articles/s41431-017-0089-8

ABSTRACT

Chronic obstructive pulmonary disease (COPD) is a major health burden in adults and cigarette smoking is considered the most important environmental risk factor of COPD. Chromosome 15q25.1 locus is associated with both COPD and smoking. Our study aims at understanding the mechanism underlying the association of chromosome 15q25.1 with COPD through epigenetic and transcriptional variation in a population-based setting.

To assess if COPD-associated variants in 15q25.1 are methylation quantitative trait loci, epigenome-wide association analysis of four genetic variants, previously associated with COPD ($P<5\times10^{-8}$) in the 15q25.1 locus (rs12914385:C>T-*CHRNA3*, rs8034191:T>C-*HYKK*, rs13180:C>T-*IREB2*, rs8042238:C>T-*IREB2*), was performed in the Rotterdam study (n=1 489). All four variants were significantly associated ($P<1.4\times10^{-6}$) with blood DNA methylation of *IREB2,* C*HRNA3* and *PSMA4*, of which two, including *IREB2* and *PSMA4* were also differentially methylated in COPD cases and controls ($P<0.04$). Further additive and multiplicative effects of smoking were evaluated, and no significant effect was observed. To evaluate if these four genetic variants are expression quantitative trait loci, transcriptome-wide association analysis was performed in 1 087 lung samples. All four variants were also significantly associated with differential expression of the *IREB2* 3'UTR in lung tissues ($P<5.4\times10^{-95}$).

We conclude that regulatory mechanisms affecting the expression of *IREB2* gene, such as DNA methylation, may explain the association between genetic variants in chromosome 15q25.1 and COPD, largely independent of smoking.

INTRODUCTION

Chronic obstructive pulmonary disease (COPD) affects over 300 million people and is the third leading cause of death worldwide, which makes it a major public health burden.[1] COPD is characterized by airflow limitation and chronic, inflammatory response of the airways to cigarette smoke, occupational exposures, air pollution etc.[2] Systemic inflammation and complications, together with comorbid conditions add to its complexity.[3]

COPD is determined by both genetic and environmental factors. Genetic factors explain 20-40% of the variance in the disease,[4] while the most important environmental risk factor is smoking. Smokers have 10 fold increased incidence of COPD compared to never-smokers.[5] Genome-wide association studies (GWAS) revealed genetic variants associated with COPD and lung function.[6,7] From a genetic-epidemiological perspective, the chromosome 15q25.1 locus is of interest, harbouring three nicotinic receptors (*CHRNA3, CHRNA5* and *CHRNB4*) and other genes that could have potential impact on COPD, including *IREB2, PSMA4* and *HYKK*.[8-11] The 15q25.1 region has also been associated with smoking[12,13] and lung cancer[14] in large GWAS. Because smoking is a risk factor for both COPD and lung cancer, the association of this locus with COPD and lung cancer might be mediated through smoking,[15] which is in line with the only longitudinal study investigating this hypothesis thus far.[16]

The mechanism through which these SNPs and smoking are involved in COPD and related outcomes remains obscure. SNPs in the 15q25.1 locus are shown to be *cis*- expression quantitative trait loci (*cis*-eQTLs) in blood,[17] brain,[18] sputum[19] and lungs.[20,21] This raises the question whether the variants are involved in COPD through the regulatory mechanisms. DNA methylation is a heritable, dynamic, epigenetic mark that plays a critical role in the regulation of gene expression.[22] Despite having a strong genetic component, DNA methylation is known to respond to changes in environmental factors,[23] and its role in mediating genetic risk effect and the interaction with environmental exposure has been widely proposed.[24]

Recent studies have shown the association between genome-wide patterns of DNA methylation variation with smoking behaviour[25,26] and COPD.[22,27] Though differential DNA methylation sites (CpG) in *CHRNA3* (15q25.1) were associated with COPD status and lung cancer,[28] the role of DNA methylation as a mechanism through which the genetic variants may be involved in COPD and related outcomes, remains unexplored.

In this study we selected all single nucleotide polymorphisms (SNP) in the 15q25.1 region, associated with COPD in GWAS: rs12914385:C>T (*CHRNA3*), rs8034191:T>C (*HYKK, CHRNA3, CHRNA5*), rs13180:C>T (*IREB2*) and rs8042238:C>T (*IREB2*).[8-11] Rs12914385:C>T is a top hit in the largest GWAS of COPD (OR=1.39, $P=2.7\times10^{-16}$).[8]

SNPs rs12914385:C>T and rs8034191:T>C are in moderate linkage disequilibrium (LD; $r^2 = 0.723$), while rs13180:C>T and rs8042238:C>T are in perfect LD ($r^2 = 0.997$). Rs8034191:T>C is estimated to explain 12.2% of COPD risk in the general population and 14.3% in current smokers.[11] Rs13180:C>T, is associated with COPD, independently of smoking and of rs8034191:T>C.[10,15,21] Evidence suggests that the association of COPD with rs8034191:T>C may be mediated by smoking while the association with rs13180:C>T is independent of smoking.[21,29] We examined whether SNPs in the 15q25.1 locus are also associated with differential DNA methylation in the population-based Rotterdam study (RS). Further, we tested if methylation patterns associated with these variants are associated with COPD and FEV_1/FVC (the ratio of the forced expiratory volume in first second (FEV_1) over forced vital capacity of the lungs (FVC)). We also tested whether the variants are associated with a differential expression in non-tumour lung tissue from the Lung eQTL study (LES).

METHODS

Study populations

Participants of the discovery and replication cohorts were part of the RS, a prospective, population-based study, designed to investigate the occurrence and determinants of diseases in the elderly, as described elsewhere.[30] The discovery cohort of our epigenetic analysis is a random sample of 723 participants from RS with complete phenotype, genome-wide genotype and methylation data available. An independent sample of 766 participants from RS, were included as a replication cohort. RS is part of the Biobanking and Biomolecular Resources Research Infrastructure for The Netherlands (BBMRI-NL), BIOS (Biobank-based Integrative Omics Studies) project.[31] The EWAS data of RS was made publically available as a Rainbow Project (RP3; BIOS) of the BBMRI-NL (data access link: http://wiki.bbmri.nl/wiki/BIOS_bios). Results of this study are available through dbGaP (accession number *phs000930,* https://www.ncbi.nlm.nih.gov/gap).

Detailed information on spirometry measures, COPD diagnosis, COPD SNPs selection, genotyping and DNA methylation assessment in RS and RNA array in LES is provided in the Supplementary information.

Statistical analyses

First, we tested the association of the four selected SNPs with COPD and FEV_1/FVC in our discovery and replication cohorts, using logistic and linear regression models, respectively, adjusted for age and sex in Model 1 and additionally adjusting

for current smoking and pack-years in Model 2. Results from the two cohorts were then meta-analysed using fixed effects models with "rmeta" package in R. Further, in the RS discovery cohort (n=723), we performed four epigenome-wide association studies (EWAS), to assess the relationship between dosages of each SNP, as independent variable, and epigenome-wide DNA methylation in blood as dependent variable. We applied linear regression methods using two models. One adjusted for age, sex, technical covariates to correct for batch effects (array number and position on array), and white blood cell types to correct for the cellular heterogeneity of blood (number of lymphocytes, monocytes and granulocytes) (Model 1). The other was adjusted additionally for current smoking and pack-years, the number of cigarette packs smoked in one year (Model 2). The False discovery rate (FDR) <0.05 was used to declare epigenome-wide significance. Significant sites were tested in the replication cohort (n=766) using the same models. Since the 15q25.1 region is also associated with smoking behaviour, significant CpG sites were also tested in a third model including 'SNP × current smoking' and 'SNP × pack-years' interaction terms to assess possible genetic-environment interaction between the tested SNPs and smoking per cohort. Per cohort results were meta-analysed using fixed effects models with "rmeta" package in R. Associations of the identified CpG sites with COPD and FEV_1/FVC, were further performed using logistic and linear regression, respectively, adjusted for age, sex, technical covariates and white blood cell counts in both the discovery and replication cohorts, and meta-analysed as mentioned above.

Finally, we assessed whether the identified SNPs were acting as eQTLs in lung tissue. We performed genome-wide eQTL analysis in 1 087 samples from GRN, UBC and Laval, compared to Nguyen et al. who used Laval (N=420) as the discovery and UBC and GRN samples for replication. First, cohort specific (GRN, Laval and UBC) principal components explaining at least one percent were calculated based on residuals from linear regression models on genome-wide 2-log transformed gene expression levels (of each probe separately) adjusted for COPD status, age, sex and smoking status. Second, in each cohort separately, linear regression analysis was used to test for association between the SNPs and genome-wide 2-log transformed gene expression levels. SNPs were tested in an additive genetic model and the models were adjusted for disease status (COPD, alpha-1 antitrypsin deficiency, idiopathic pulmonary fibrosis, pulmonary hypertension, cystic fibrosis, and other disease), age, sex, smoking status and the cohort specific number of PCs (14 PCs for GRN and Laval, and 16 for UBC). Finally, SNP effect estimates of the three cohorts were meta-analysed using fixed effects models with effect estimates weighted by the reciprocal of the estimated variance. We used FDR<0.05 to correct for multiple testing.

RESULTS

The discovery set comprised 723 participants of RS, with genotype and DNA methylation data from whole blood, including 114 COPD cases and 541 controls (68 excluded due to possible asthma). The replication set comprised 766 independent participants of Rotterdam study, with genotype and DNA methylation data, including 93 COPD cases and 591 controls (82 excluded due to possible asthma). The characteristics of the discovery and replication cohorts are shown in **Table 1**. COPD cases were more often male and smokers and had smoked on average more pack-years, compared to controls.

Table 1. Characteristics of the discovery and replication cohorts and per COPD status.

	Discovery cohort			Replication cohort			
	COPD	**Controls**	**All**	**COPD**	**Controls**	**All**	**P-value**
N (% of all)	114 (15.8)	541 (74.8)	723	93 (12.1)	591 (77.2)	766	0.054*
Age (years)	61.9±8.6	59.3±7.9	59.9±8.2	68.2±5.7	67.6±5.9	67.7±5.9	$<5\times10^{-6}$
Gender (% males)	59.6	43.1	45.8	58.1	42.1	42.3	0.176
FEV_1/FVC	0.63±0.07	0.78±0.04	0.75±0.08	0.63±0.07	0.78±0.05	0.76±0.08	0.08
Smoking status:							$<5\times10^{-6}$
Current smokers (%)	37.7	19.8	23.2	21.5	8.8	10.4	
Ex- smokers (%)	48.2	44.2	44.4	54.8	55.2	55.7	
Never smokers (%)	14	36	32.4	23.7	36	33.8	
Pack-years †	34.3±26.9	19.9±19.6	23.2±22.0	33.7±18.7	19.6±20.1	21.9±20.6	0.261

Data presented as % or mean±SD; COPD: Chronic Obstructive Pulmonary Disease cases; All: all participants included in EWAS; For COPD status, which was not available for all participants, the valid percentage is denoted in brackets (% of all); In the Discovery cohort 68 patients and in the Replication cohort 82 patients were excluded from the association analyses with COPD, due to possible asthma; P-value: P-value of the difference of Discovery and Replication cohorts; * P-value of the difference of COPD status in Discovery and Replication cohorts; † Pack-years calculated in current and ex-smokers only.

Three of the four selected SNPs (rs12914385:C>T, rs13180:C>T and rs8042238:C>T) on chromosome 15q25.1 were nominally associated with COPD in RS (n=1 339) while only rs12914385:C>T was nominally associated with COPD in a considerably smaller dataset from LES (n=512) (**Table S1**). None of the SNPs were associated with FEV_1/FVC (**Table S1**).

To determine whether the four SNPs (rs12914385:C>T, rs8034191:T>C, rs13180:C>T and rs8042238:C>T) in chromosome 15q25.1 are methylation quantitative trait loci (meQTLs), we performed EWAS in RS for each SNP. Significant associations (FDR<0.05) were detected at 14 unique CpG sites (**Figure 1**), 12 sites in *cis* (within the window of 400kb) (**Figure 2**), and two in different chromosomes, 6 and 12 (**Table 2**, Model 1).

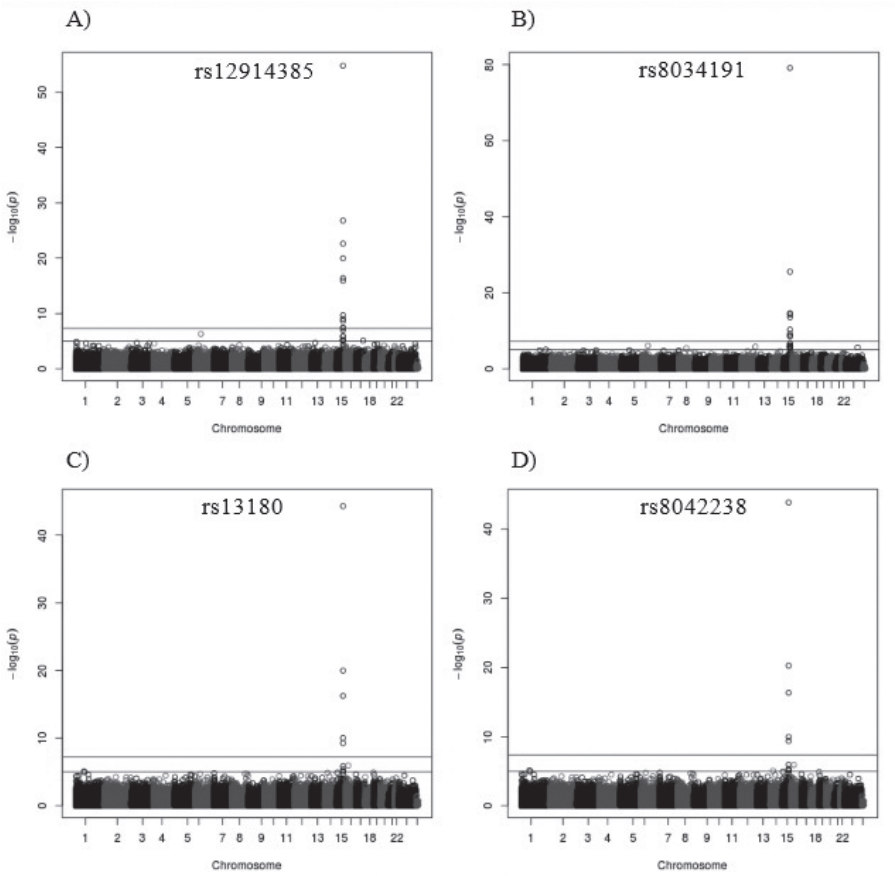

Figure 1. Plotted results of the discovery EWAS of the four SNPs at chromosome 15. In circles are represented all CpGs throughout the genome. X-axis shows all chromosomes. Y-axis shows negative logarithm of the p-value of the associations of (A) rs12914385:C>T, (B) rs8034191:T>C, (C) rs13180:C>T and (D) rs8042238:C>T with epigenome-wide DNA methylation levels. Upper line represents the significance threshold (FDR < 0.05). N=723.

Of these 14 CpG sites, 10 were significantly replicated in the independent sample from RS at a significance level of P<0.0019, corresponding to the Bonferroni correction for number of tests performed in the replication sample (n=26) (**Table 2**, Model 1). All four SNPs were significantly associated with differential methylation at three CpG sites (cg18825076, cg04882995 and cg04140906) in *IREB2,* C*HRNA3* and *PSMA4,* respectively (**Table 2**, **Figure 2**). Addition of smoking as a confounder did not change the results (**Table 2**; Model 2), suggesting that the effects of the SNPs on DNA methylation are independent of smoking. However, significant genetic-

Table 2. Association of SNPs with epigenome-wide DNA methylation in two models in discovery and replication cohorts.

			Discovery (N=723)						Replication (N=766)					
			Model 1			Model 2			Model 1			Model 2		
SNP	CpG	Annotation of the CpG (hg19)	ß	SE	P	ß	SE	P	ß	SE	P	ß	SE	P
rs12914385	cg18825076*	chr15:g.78729989, *IREB2*	-0.0225	0.0013	**1.54×10^{-55}**	-0.0225	0.0013	**1.19×10^{-55}**	-0.0309	0.0016	**1.39×10^{-68}**	-0.0307	0.0016	**2.05×10^{-67}**
CHRNA3	cg04882995*	chr15:g.78912658, *CHRNA3*	0.0124	0.0011	**1.59×10^{-27}**	0.0123	0.0011	**6.34×10^{-27}**	0.0148	0.0014	**4.61×10^{-25}**	0.0148	0.0014	**7.46×10^{-25}**
intronic variant	cg06752398	chr15:g.79053858, *ADAMTS7*	0.0165	0.0016	**2.34×10^{-23}**	0.0164	0.0016	**6.62×10^{-23}**	NA	NA	NA	NA	NA	NA
chr15:g.78606381C>T	cg21242079	chr15:g.79101063, *ADAMTS7*	0.0195	0.0022	**3.90×10^{-17}**	0.0194	0.0022	**4.64×10^{-17}**	0.0224	0.0034	**8.36×10^{-11}**	0.022	0.0034	**1.95×10^{-10}**
	cg00540400	chr15:g.79124168, *ADAMTS7-MORF4L1*	-0.0235	0.0028	**1.19×10^{-16}**	-0.0239	0.0027	**3.07×10^{-17}**	-0.0334	0.0039	**1.51×10^{-16}**	-0.0334	0.004	**2.13×10^{-16}**
	cg04140906*	chr15:g.78833505, *PSMA4*	-0.0061	0.0011	**3.25×10^{-8}**	-0.0062	0.0011	**2.08×10^{-8}**	-0.0125	0.0021	**3.49×10^{-9}**	-0.0125	0.0021	**4.35×10^{-9}**
	cg05786009	chr15:g.79152474, *ADAMTS7-MORF4L1*	0.0037	0.0007	**3.96×10^{-8}**	0.0037	0.0007	**5.15×10^{-8}**	0.0048	0.0011	**3.28×10^{-5}**	0.0048	0.0011	**3.16×10^{-5}**
	cg05012158	chr15:g.79051863, *ADAMTS7*	-0.0058	0.0011	**1.04×10^{-7}**	-0.0058	0.0011	**1.37×10^{-7}**	NA	NA	NA	NA	NA	NA
	cg21436520	chr6:g.53413189, *GCLC-LOC101929136*	0.006	0.0012	**4.74×10^{-7}**	0.0061	0.0012	**2.75×10^{-7}**	-0.0007	0.0021	0.75139	-0.0004	0.0021	0.830983
rs8034191	cg18825076*	chr15:g.78729989, *IREB2*	-0.0277	0.0013	**7.41×10^{-80}**	-0.0279	0.0013	**3.11×10^{-80}**	-0.0353	0.0015	**2.39×10^{-87}**	-0.0352	0.0015	**4.51×10^{-86}**
HYKK	cg04882995*	chr15:g.78912658, *CHRNA3*	0.0129	0.0012	**2.72×10^{-26}**	0.0128	0.0012	**1.24×10^{-25}**	0.016	0.0014	**1.96×10^{-27}**	0.0161	0.0014	**1.89×10^{-27}**
intronic variant	cg06752398	chr15:g.79053858, *ADAMTS7*	0.014	0.0017	**5.08×10^{-15}**	0.0139	0.0018	**1.32×10^{-14}**	NA	NA	NA	NA	NA	NA
chr15:g.78513681T>C	cg21242079	chr15:g.79101063, *ADAMTS7*	0.0189	0.0024	**2.49×10^{-14}**	0.0188	0.0024	**2.92×10^{-14}**	0.022	0.0035	**8.71×10^{-10}**	0.0217	0.0036	**1.85×10^{-9}**
	cg20117256	chr15:g.78726576, 4kb 5' *IREB2*	0.024	0.0039	**1.98×10^{-9}**	0.0235	0.004	**4.31×10^{-9}**	0.0268	0.0034	**1.44×10^{-14}**	0.0266	0.0034	**3.47×10^{-14}**
	cg00540400	chr15:g.79124168, *ADAMTS7-MORF4L1*	-0.0181	0.003	**3.02×10^{-9}**	-0.0185	0.003	**1.29×10^{-9}**	-0.0304	0.0041	**6.34×10^{-13}**	-0.0304	0.0042	**9.67×10^{-13}**
	cg16751781	chr15:g.78858589, *CHRNA5*	-0.0037	0.0007	**1.89×10^{-7}**	-0.0036	0.0007	**2.40×10^{-7}**	-0.0055	0.0016	**0.000467**	-0.0057	0.0016	**0.000312**
	cg21436520	chr6:g.53413189, *GCLC-LOC101929136*	0.0063	0.0013	**7.75×10^{-7}**	0.0065	0.0013	**4.12×10^{-7}**	-0.0002	0.0021	0.909182	-0.0001	0.0021	0.970673
	cg08701566	chr15:g.78911099, *CHRNA3*	-0.0037	0.0007	**8.90×10^{-7}**	-0.0038	0.0007	**5.35×10^{-7}**	-0.0077	0.0013	**3.56×10^{-9}**	-0.0078	0.0013	**2.55×10^{-9}**
	cg05786009	chr15:g.79152474, *ADAMTS7-MORF4L1*	0.0035	0.0007	**1.09×10^{-6}**	0.0035	0.0007	**1.44×10^{-6}**	0.0045	0.0012	**0.000151**	0.0045	0.0012	**0.000167**
	cg04140906*	chr15:g.78833505, *PSMA4*	-0.0057	0.0012	**1.37×10^{-6}**	-0.0059	0.0012	**8.65×10^{-7}**	-0.011	0.0022	**5.75×10^{-7}**	-0.0111	0.0022	**5.67×10^{-7}**
	cg23556238	chr12:g.125298876, *SCARB1*	-0.0071	0.0015	**1.53×10^{-6}**	-0.007	0.0015	**2.14×10^{-6}**	-0.0034	0.0012	0.005374	-0.0037	0.0012	0.002745
rs13180	cg18825076*	chr15:g.78729989, *IREB2*	-0.0201	0.0013	**4.84×10^{-45}**	-0.0201	0.0013	**1.32×10^{-44}**	-0.0257	0.0017	**8.23×10^{-47}**	-0.0256	0.0016	**1.2×10^{-46}**
IREB2	cg04140906*	chr15:g.78833505, *PSMA4*	-0.01	0.001	**1.02×10^{-20}**	-0.0101	0.001	**4.65×10^{-21}**	-0.0126	0.002	**1.23×10^{-9}**	-0.0125	0.002	**1.28×10^{-9}**
synonymous variant	cg13561554	chr15:g.78795944, *IREB2-HYKK*	-0.0245	0.0028	**5.59×10^{-17}**	-0.0244	0.0029	**8.73×10^{-17}**	-0.0207	0.0021	**1.26×10^{-21}**	-0.0206	0.0021	**2×10^{-21}**
chr15:g.78497146C>T	cg04882995*	chr15:g.78912658, *CHRNA3*	0.0072	0.0011	**5.44×10^{-10}**	0.0071	0.0011	**9.84×10^{-10}**	0.0095	0.0014	**2.79×10^{-11}**	0.0095	0.0014	**3.16×10^{-11}**
rs8042238	cg18825076*	chr15:g.78729989, *IREB2*	-0.0201	0.0013	**1.35×10^{-44}**	-0.02	0.0013	**3.64×10^{-44}**	-0.0258	0.0016	**1.47×10^{-47}**	-0.0256	0.0016	**2.12×10^{-47}**
IREB2	cg04140906*	chr15:g.78833505, *PSMA4*	-0.01	0.001	**5.47×10^{-21}**	-0.0102	0.001	**2.43×10^{-21}**	-0.0129	0.002	**4.09×10^{-10}**	-0.0129	0.002	**4×10^{-10}**
intronic variant	cg13561554	chr15:g.78795944, *IREB2-HYKK*	-0.0246	0.0028	**4.64×10^{-17}**	-0.0245	0.0029	**7.24×10^{-17}**	-0.0203	0.0021	**4.77×10^{-21}**	-0.0202	0.0021	**8.11×10^{-21}**
chr15:g.78481929C>T	cg04882995*	chr15:g.78912658, *CHRNA3*	0.0072	0.0011	**4.37×10^{-10}**	0.0071	0.0011	**8.02×10^{-10}**	0.0095	0.0014	**2.56×10^{-11}**	0.0095	0.0014	**2.73×10^{-11}**

β: Regression coefficient estimates from linear regression model regressing DNA methylation levels on indicated SNPs; In Model 1 coefficients are corrected for sex, age, technical covariates and different white blood cellular proportions; In Model 2 coefficients are additionally adjusted for current smoking and pack-years smoked; SE: standard error of the effect; P: p-value of the significance; N: number of participants in the analysis; in bold: significant results; *Methylation sites significantly associated with all four SNPs.

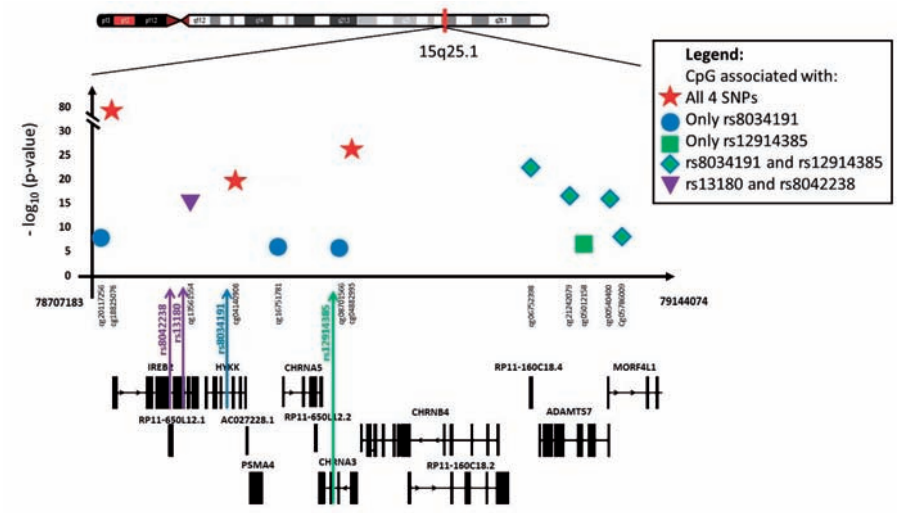

Figure 2. Regional P-value plot of chromosome 15q25.1 with significant SNP-CpG associations. Every point in the plot represents one SNP-CpG association (n=12). X-axis shows all genes in the region as well as the position of SNPs of interest; Y-axis shows negative logarithm of the p-values of the associations of CpGs with the corresponding SNP (distinguished by different colours). In the case a CpG associates with more than one SNP, the smaller p-value is taken into account.

environment interaction between pack-years and rs12914385:C>T (*CHRNA3*) was observed for DNA methylation levels at two CpG sites including the top hit, cg18825076 (*IREB2*; $P_{interaction}=5.0\times10^{-4}$), and cg00540400 (*ADAMTS7-MORF4L1*; $P_{interaction}=1.7\times10^{-3}$) (**Table S2**). The association between rs12914385:C>T (*CHRNA3*) and cg18825076 and cg00540400 remained significant, albeit slightly decreased (**Table S2**). However, the direction of the effect of both significant interactions was opposite in the discovery and replication cohorts (**Table S3**), suggesting this is likely a false positive finding.

Interestingly, these two CpG sites i.e. cg18825076 and cg00540400 and a third one, cg04140906 in *PSMA4,* were nominally associated (P<0.04) with COPD (**Table 3, Table S4**, Model 1). When correcting for smoking (**Table 3, Table S4**, Model 2) the association between cg18825076 and COPD disappeared (P=0.16), while the association of the other two CpG sites (cg00540400 and cg04140906) became stronger (P<0.03). None of the CpGs were associated with FEV_1/FVC (**Table S5**).

Table 3. Association of COPD with DNA methylation at 10 replicated sites from the meQTL analysis; meta-analyzed results (N=1339).

			Model 1			Model 2		
CpG	**Position (hg19)**	**Gene**	**β**	**SE**	**P**	**β**	**SE**	**P**
cg18825076	78729989	*IREB2*	-5.165	2.514	**0.04**	-3.65	2.58	0.158
cg04882995	78912658	*CHRNA3*	1.6	3.098	0.606	-0.57	3.225	0.86
cg21242079	79101063	*ADAMTS7*	2.321	1.354	0.086	2.454	1.389	0.077
cg00540400	79124168	*ADAMTS7-MORF4L1*	2.577	1.27	**0.042**	3.228	1.313	**0.014**
cg04140906	78833505	*PSMA4*	-5.674	2.714	**0.037**	-6.002	2.772	**0.03**
cg05786009	79152474	*ADAMTS7-MORF4L1*	2.043	4.307	0.635	1.723	4.469	0.7
cg20117256	78726576	*IREB2*	-1.227	1.188	0.302	-1.54	1.23	0.212
cg16751781	78858589	*CHRNA5*	0.891	3.57	0.803	2.092	3.707	0.573
cg08701566	78911099	*CHRNA3*	6.785	3.971	0.087	6.333	4.136	0.126
cg13561554	78795944	*IREB2-HYKK*	-1.185	1.521	0.436	-0.62	1.559	0.689

β: coefficient estimates from the logistic regression models: Model 1 adjusted for age, sex, technical covariates and different white blood cellular proportions; Model 2 additionally adjusted for current smoking and pack-years; SE: standard error of the effect; P: p-value of the significance; in bold: nominally significant results.

Results involving differential DNA methylation at *IREB2* are of special interest, as we have detected association with COPD risk allele (rs12914385:C>T) and with COPD status. These relationships are illustrated in Figure 3. We found a higher frequency of T allele among the COPD cases, compared to the controls in our dataset (**Figure 3a**). Further, the T allele was associated with lower DNA methylation of the cg18825076 (**Figure 3b**). Lastly, COPD cases showed lower DNA methylation at cg18825076 compared to controls (**Figure 3c**).

Finally, in the eQTL analysis in 1087 lung tissue samples from LES significant association (FDR<0.05) was observed at 15 expression probe sets both in *cis* and *trans* (**Table 4**). All four SNPs were significantly associated with probe set 100154936_TGI_at (3'UTR of the *IREB2* gene, $P<3.2\times10^{-98}$) (**Figure 4**). The T allele of rs12914385:C>T was associated with higher expression of this probe. *Trans*-eQTL effects were observed in the chromosomes 2, 3, 4, 13 and 14 (**Table 4**). Rs8034191:T>C was associated with gene expression of *FAM13A* intron in chromosome 4 (100158626_TGI_at), a very well-known gene involved in COPD.[10]

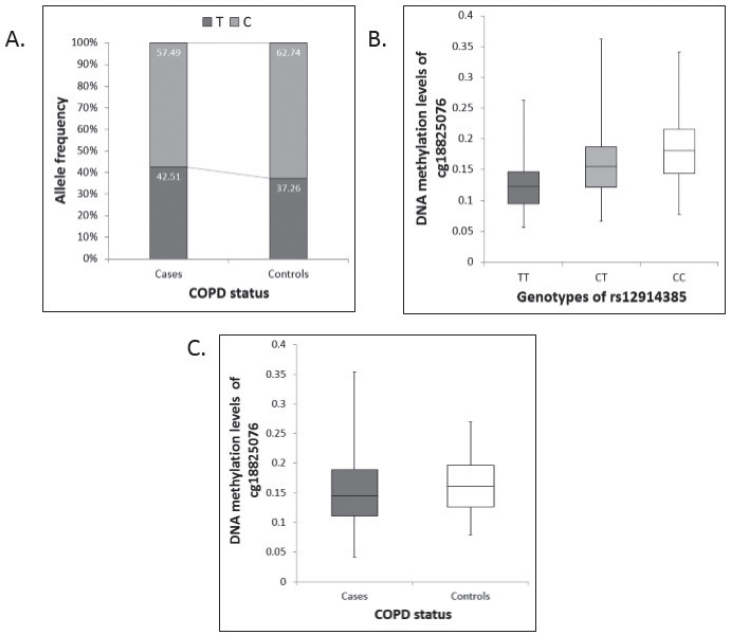

Figure 3. Interplay between genetics, DNA methylation and COPD status at IREB2 gene. (A) The frequency of T and C alleles of rs12914385:C>T (Y-axis) among COPD cases and controls (X-axis) (N=1339, P=0.043). Homozygotes were counted as carrying two and heterozygotes as carrying one copy of a given allele. (B) Differences in cg18825076 DNA methylation levels (Y-axis) between rs12914385:C>T genotypes (X-axis) (N=1489, $P=1.05\times10^{-125}$). (C) Differences in cg18825076 DNA methylation levels (Y-axis) in COPD cases and controls (X-axis) (N=1339, P=0.04).

Table 4. Association of COPD-SNPs with transcriptome-wide gene expression in lungs (N=1087).

SNP	Tested allele	Probe	Chr	Position (hg19)	Annotation	β	SE	P
rs12914385	T	100154936_TGI_at*	15	78792408-78792890	3'UTR *IREB2* (AF147302)	0.23	0.009	**2.13×10^{-142}**
CHRNA3		100148404_TGI_at	15	78832747-78841563	*PSMA4* (NM_002789)	-0.06	0.006	**1.62×10^{-28}**
intronic variant		100156434_TGI_at	15	78857862-78887611	*CHRNA5* (NM_000745)	0.14	0.014	**1.00×10^{-23}**
chr15:g.78606381C>T		100313804_TGI_at	15	78828532-78828986	*mRNA intron HYKK* (AA532820)	-0.10	0.015	**2.38×10^{-10}**
		100132425_TGI_at	15	78799906-78829715	*HYKK* (NM_001083612)	0.08	0.015	**2.56×10^{-8}**
		100123649_TGI_at	15	78887647-78913637	*CHRNA3* (NM_000743)	0.11	0.026	**7.25×10^{-6}**
rs8034191	C	100154936_TGI_at*	15	78792408-78792890	3'UTR *IREB2* (AF147302)	-0.24	0.008	**3.27×10^{-184}**
HYKK		100148404_TGI_at	15	78832747-78841563	*PSMA4* (NM_002789)	0.07	0.006	**3.17×10^{-37}**
intronic variant		100156434_TGI_at	15	78857862-78887611	*CHRNA5* (NM_000745)	-0.16	0.014	**9.11×10^{-32}**
chr15:g.78513681T>C		100313804_TGI_at	15	78828532-78828986	*mRNA intron HYKK* (AA532820)	0.11	0.015	**9.15×10^{-14}**
		100132425_TGI_at	15	78799906-78829715	*HYKK* (NM_001083612)	-0.10	0.015	**2.90×10^{-12}**
		100146326_TGI_at	3	31745705-31763038	*OSBPL10* (BX103488)	0.05	0.011	**7.57×10^{-6}**
		100312267_TGI_at	15	78887350-78889074	*CHRNA3* (AJ584709)	-0.09	0.020	**9.42×10^{-6}**
		100123649_TGI_at	15	78887647-78913637	*CHRNA3* (NM_000743)	-0.11	0.025	**2.02×10^{-5}**
		100158626_TGI_at	4	89658713-89659241	*mRNA intron FAM13A* (AI922966)	-0.06	0.015	**2.37×10^{-5}**
		100143446_TGI_at	15	78730518-78793798	*IREB2* (NM_004136)	-0.03	0.006	**2.76×10^{-5}**
rs13180	T	100154936_TGI_at*	15	78792408-78792890	3'UTR *IREB2* (AF147302)	-0.20	0.010	**3.17×10^{-98}**
IREB2		100129239_TGI_at	15	78780991-78786607	*IREB2* (AW118658)	0.06	0.011	**2.46×10^{-7}**
synonymous variant		100143446_TGI_at	15	78730518-78793798	*IREB2* (NM_004136)	-0.03	0.006	**1.41×10^{-6}**
chr15:g.78497146C>T		100161812_TGI_at	2	242943360-242948160	(AK126530)	-0.04	0.009	**8.78×10^{-6}**
		100140544_TGI_at	14	104145786-104168220	*KLC1* (CO247613)	0.04	0.010	**9.38×10^{-6}**
		100144578_TGI_at	2	37869087-37871279	UTR *CDC42EP3* (AK055915)	-0.04	0.010	**1.44×10^{-5}**
		100141776_TGI_at	13	50571143-50592603	*TRIM13* (NM_052811)	0.04	0.010	**2.83×10^{-5}**
rs8042238	T	100154936_TGI_at*	15	78792408-78792890	3'UTR *IREB2* (AF147302)	-0.20	0.010	**1.10×10^{-98}**
IREB2		100129239_TGI_at	15	78780991-78786607	*IREB2* (AW118658)	0.06	0.011	**2.84×10^{-7}**
intronic variant		100143446_TGI_at	15	78730518-78793798	*IREB2* (NM_004136)	-0.03	0.006	**1.09×10^{-6}**
chr15:g.78481929C>T		100161812_TGI_at	2	242943360-242948160	(AK126530)	-0.04	0.009	**5.73×10^{-6}**
		100140544_TGI_at	14	104145786-104168220	*KLC1* (CO247613)	0.04	0.010	**8.79×10^{-6}**
		100144578_TGI_at	2	37869087-37871279	UTR *CDC42EP3* (AK055915)	-0.04	0.010	**1.36×10^{-5}**
		100141776_TGI_at	13	50571143-50592603	*TRIM13* (NM_052811)	0.04	0.010	**2.47×10^{-5}**

β: Regression coefficient estimates from linear regression models regressing gene expression on indicated SNPs (see model in Methods section); SE: standard error; P: p-value of the significance; in bold: significant results; *Probe set associated with all four SNPs.

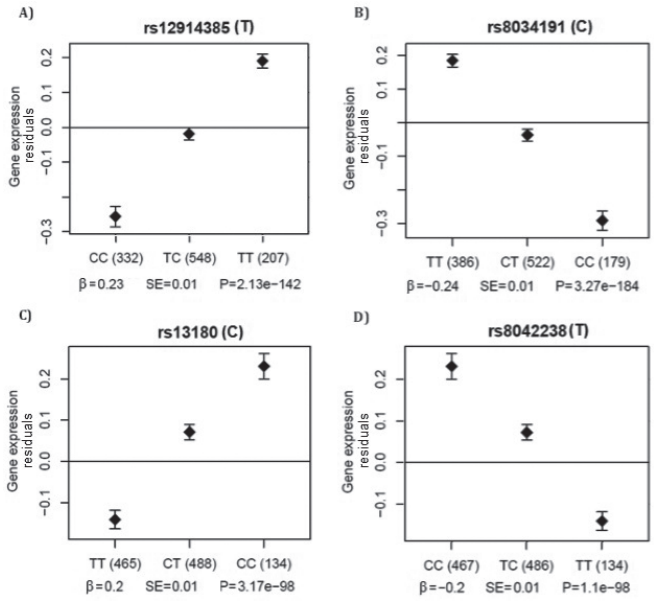

Figure 4. 3'UTR IREB2 expression plot in regard to the genotypes of four SNPs. Plots of genotype specific mean residuals (with 95% confidence intervals) from the linear regression models on gene expression levels adjusted for disease status, age, sex, smoking status and the cohort specific number of PCs. The association of residuals of the expression levels (probe set 100154936_TGI_at, 3'UTR *IREB2*) on the Y-axis with genotypes of (A) rs12914385:C>T, (B) rs8034191:T>C, (C) rs13180:C>T and (D) rs8042238:C>T on the X-axis, risk/tested allele is given in brackets; β: Regression coefficient estimate, SE: standard error, P: p-value of the significance**,** TT/TC/CC are different genotypes of the given SNP and number of carriers is given in brackets.

DISCUSSION

In the current study we show that four COPD-associated SNPs (rs12914385:C>T, *CHRNA3*; rs8034191:T>C, *HYKK;* rs13180:C>T and rs8042238:C>T, *IREB2*) in the 15q25.1 locus are also blood meQTLs that regulate mostly nearby DNA methylation levels. All variants are associated with differential DNA methylation of *IREB2, CHRNA3* and *PSMA4* genes, independently of smoking. We further show that DNA methylation at two sites in genes *IREB2* and *PSMA4,* together with the site between *ADAMTS7* and *MORF4L1,* are associated with COPD. Finally, we show that all four SNPs are also lung *cis*- and *trans*-eQTLs, affecting the expression of several genes including *IREB2, PSMA4, CHRNA3, CHRNA5, HYKK, FAM13A, KLC1* and *TRIM13.* Our results demonstrate that COPD-SNPs shape the epigenetic regulatory landscape in the 15q25.1 locus in blood and lung tissues, and suggest that the genetic risk of these SNPs on COPD

might be mediated and/or modified by DNA methylation levels in this region. Overall, our findings put forward the role of DNA methylation in COPD as an important mechanism in the complex regulation of the 15q25.1 locus.

Rs12914385:C>T and rs8034191:T>C are in moderate linkage disequilibrium (LD) (**Table S6**), so as expected they yielded multiple overlapping results. Rs13180:C>T and rs8042238:C>T are in perfect LD (**Table S6**) and thus showed almost the same effects. We show that rs12914385:C>T, rs8034191:T>C and rs13180:C>T are associated with COPD in RS, but only rs12914385:C>T in LES, which is expected given the lower sample size. While rs12914385:C>T was not vastly studied, rs8034191:T>C is a well-studied SNP with regards to COPD and lung function in different populations.[32-35] It has been shown that rs8034191:T>C is associated with an increased risk for COPD independent of rs13180:C>T and pack-years of smoking.[33] On the contrary, the same study shows that the association of rs13180:C>T with severe COPD is possibly driven by moderate LD with rs8034191:T>C (r^2=0.21). However, a study in a Chinese population shows association of both SNP with lung function but not with COPD, and the association of rs8034191:T>C with pack-years in COPD cases.[35] A study including 3,424 COPD cases and 1,872 controls showed that the association of rs8034191:T>C with COPD is 30% mediated by pack-years, and this mediation increases to 42% when adjusted for rs13180:C>T.[15] In our study we show that the rs8034191:T>C yielded more epigenome-wide significant results and that some overlapped with rs13180:C>T (**Figure 2**).

The most interesting finding of this study is that all four SNPs influence the three CpG sites in the *IREB2, CHRNA3* and *PSMA4* genes. Furthermore, the same site in *PSMA4* is also associated with COPD, independent of smoking. Site in *IREB2*, our top hit, is also associated with COPD, but this association drops after correcting for smoking. This suggests that the four genetic variants may influence COPD susceptibility through changes in DNA methylation of *IREB2*, *PSMA4* and *CHRNA3*.

We focused on chromosome 15q25.1 region hits from COPD GWASs. However, this region has also been reported in the association with smoking by several large smoking genetics consortia.[12,13] They showed that locus in 15q25.1, represented by rs16969968:G>A, and other SNPs, are mostly associated with smoking quantity. Saccone et al.[36] showed no significant association between rs16969968:G>A and COPD in smokers, adjusted for cigarettes per day. This SNP was in high LD with our SNPs, rs12914385:C>T (r^2=0.84) and rs8034191:T>C (r^2=0.93), but in this study we show that the proposed pathway, in which SNPs act as meQTLs, is mainly independent of smoking. Additionally, none of our four SNPs, were found to be significant meQTLs in the brain in a study of nicotine dependence.[37] Despite the well-described role of smoking behaviour in shaping the methylome at multiple tissues,[25,26] our SNP-DNA

methylation associations could not be explained by exposure to smoking, making a mediating effect of smoking improbable. In line with our observations, the studies of the role of smoking on DNA methylation have failed to show an effect in this region.[25,38-41] In a recent large smoking EWAS meta-analysis,[41] none of our 10 replicated sites were associated with smoking, strengthening our claims. Yet the failure to detect an association in these analyses does not necessarily represent absence of a true effect, it may rather reflect the lack of the statistical power to detect a true interaction. Further larger studies are needed to elucidate this question.

In blood, rs12914385:C>T, rs13180:C>T and rs8042238:C>T have been previously associated with differential expression of *IREB2* and *PSMA4,* while rs8034191:T>C was associated with the expression of *IREB2, PSMA4* and *CHRNA5*.[17] However, in the present study we examined the gene expression in the lung tissue and found that all four SNPs were also lung eQTLs for multiple genes in *cis,* including *IREB2, PSMA4, CHRNA3, CHRNA5, HYKK*, as well as with other genes in *trans* involved in COPD pathogenesis, such as *FAM13A* (chromosome 4).[10] Compared to Nguyen et al.,[20] who only used a part of this dataset as discovery, we showed that in addition to *CHRNA3*, *CHRNA5* and *PSMA4*, the four SNPs were also associated with differential expression of *IREB2* and *HYKK*. Again, all SNPs were associated with differential expression of *IREB2* gene, suggesting that possibly the genetic variants are involved in the pathogenesis of the disease through differential DNA methylation and regulation of expression.

Based on our results in relation to *IREB2*, we propose a disease model in which the COPD-risk allele (rs12914385:C>T, *CHRNA3*), exert its risk by lowering the DNA methylation level at *IREB2* gene and subsequently increasing its expression in COPD patients. The lower level of *IREB2* DNA methylation in blood from COPD cases and the positive effect of the risk allele on gene expression in lungs support this scenario. As we do not have the methylation and expression data in the same tissue, we were not able to validate this hypothesis directly, but future integrative studies in lung tissue should elucidate this further. *IREB2* gene is coding the RNA-binding protein that binds to iron-responsive elements and can regulate the expression of transferrin receptor and ferritin by changing its own protein expression and thereby regulate iron metabolism, important in pathogenesis of lung diseases.[42] It is shown that *IREB2* gene interacts with *MYC* and *MAX* genes involved in the regulation of the gene transcription through epigenetic changes.[43]

The strength of this study is that our findings are based on a large and unique sample of patients in-depth characterized genomically and epigenetically and our findings on the role of the genetic variants in blood corroborate with the changes in the transcriptome in lung. However, there are some limitations to our study. The

first and main limitation is the use of blood in the DNA methylation analysis as a proxy for clinically and biologically relevant changes that develop in the lungs. In absence of lung tissue DNA methylation measurements, blood is the most reasonable surrogate for examination of methylation changes related to COPD and smoking. This is because the disease, apart from affecting lung tissue, also induces systemic changes and has been associated with elevation in markers of systemic inflammation.[44] Furthermore, studies comparing the DNA methylation patterns in multiple tissues confirm that there is a great overlap in patterns, encouraging us to believe that blood is a good surrogate to study differences that occur in lungs.[45,46] Second limitation is the use of COPD definition based on the pre-bronchodilator spirometry. This measure demonstrates the variability of the smooth muscle contraction, while by using post-bronchodilator spirometry we can observe the irreversibility of the airflow limitation, the main characteristic of COPD.[47] However, in the attempt to minimize potential misclassification, we have identified and excluded all possible asthmatic patients. In addition, some epidemiological studies show that both pre- and post-bronchodilator spirometry predicted mortality related to COPD, with a similar degree of accuracy.[48] Third limitation to this study is the use of the whole lung tissue for gene expression analysis, which is very heterogeneous, instead of identifying the source cells for our eQTL signals. Finally, we formulated hypothesis based on results obtained from a cross-sectional study, in which inferences on directionality of effects are complicated. Future results based on longitudinal studies will help to support the role of DNA methylation and gene expression as regulatory mediators for COPD genetic risk. In addition, further replication of our results in other ethnicities and more diverse studies may corroborate our results. Since genetic variants identified through GWASs usually map to non-coding intergenic and intronic regulatory regions, the functional role is often unclear.[49,50] They are more likely to modulate gene expression through regulatory mechanisms and epigenetic modifications (e.g. DNA methylation), as we show in this study. However, we do not show significant association of all SNPs with the disease, but this is expected since our sample size is considerably smaller than that used in the original GWAS. Our study did not aim to prioritize between the four variants in terms of relevance to COPD, but to investigate if the variants in 15q25.1 region, associated to COPD, are involved in regulatory mechanisms. The changes in DNA methylation levels that we observed in blood are small, most likely because we assess differential methylation in a variety of cell types. More substantial changes may be found in lung tissue in future integrative studies.

In summary, we found evidence suggesting that genetic variations underlying *IREB2, HYKK* and *CHRNA3* act as *cis* meQTLs and eQTLs. They all affect the DNA

methylation and expression of *IREB2*, which also contributes to the risk of having COPD. We did not find evidence that smoking mediates these relationships, although this should be corroborated in larger sample sizes. This finding is compatible with the hypothesis that the genetic variants are involved in the pathogenesis of COPD through differential methylation and regulation of expression. Future integrative studies quantifying both DNA methylation and gene expression in lung tissue, as well as functional studies, are needed to confirm suggested hypothesis.

REFERENCES

1. Lozano R: Bin Abdulhak A, Birbeck G, Blyth F, Bolliger I, Boufous S, Bucello C, Burch M, et al: Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet, The* 2012; **380:** 2095.
2. de Jong K, Vonk JM, Timens W *et al*: Genome-wide interaction study of gene-by-occupational exposure and effects on FEV_1 levels. *Journal of Allergy and Clinical Immunology* 2015; **136:** 1664-1672.e1614.
3. Vestbo J, Hurd SS, Agusti AG *et al*: Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: GOLD executive summary. *Am J Respir Crit Care Med* 2013; **187:** 347-365.
4. Zhou JJ, Cho MH, Castaldi PJ, Hersh CP, Silverman EK, Laird NM: Heritability of chronic obstructive pulmonary disease and related phenotypes in smokers. *American journal of respiratory and critical care medicine* 2013; **188:** 941-947.
5. Society ER: European lung white book, 2016, Vol 2016.
6. Soler Artigas M, Loth DW, Wain LV *et al*: Genome-wide association and large-scale follow up identifies 16 new loci influencing lung function. *Nat Genet* 2011; **43:** 1082-1090.
7. Wilk JB, Shrine NR, Loehr LR *et al*: Genome-wide association studies identify CHRNA5/3 and HTR4 in the development of airflow obstruction. *Am J Respir Crit Care Med* 2012; **186:** 622-632.
8. Cho MH, McDonald M-LN, Zhou X *et al*: Risk loci for chronic obstructive pulmonary disease: a genome-wide association study and meta-analysis. *The lancet Respiratory medicine* 2014; **2:** 214-225.
9. Lee JH, Cho MH, Hersh CP *et al*: Genetic susceptibility for chronic bronchitis in chronic obstructive pulmonary disease. *Respiratory research* 2014; **15:** 1.
10. Cho MH, Boutaoui N, Klanderman BJ *et al*: Variants in FAM13A are associated with chronic obstructive pulmonary disease. *Nat Genet* 2010; **42:** 200-202.
11. Pillai SG, Ge D, Zhu G *et al*: A genome-wide association study in chronic obstructive pulmonary disease (COPD): identification of two major susceptibility loci. *PLoS Genet* 2009; **5:** e1000421.
12. Tobacco, Genetics C: Genome-wide meta-analyses identify multiple loci associated with smoking behavior. *Nature genetics* 2010; **42:** 441-447.
13. Liu JZ, Tozzi F, Waterworth DM *et al*: Meta-analysis and imputation refines the association of 15q25 with smoking quantity. *Nature genetics* 2010; **42:** 436-440.
14. Hung RJ, McKay JD, Gaborieau V *et al*: A susceptibility locus for lung cancer maps to nicotinic acetylcholine receptor subunit genes on 15q25. *Nature* 2008; **452:** 633-637.
15. Siedlinski M, Tingley D, Lipman PJ *et al*: Dissecting direct and indirect genetic effects on chronic obstructive pulmonary disease (COPD) susceptibility. *Human genetics* 2013; **132:** 431-441.
16. Budulac SE, Vonk JM, Postma DS, Siedlinski M, Timens W, Boezen MH: Nicotinic acetylcholine receptor variants are related to smoking habits, but not directly to COPD. *PloS one* 2012; **7:** e33386.
17. Westra H-J, Peters MJ, Esko T *et al*: Systematic identification of trans eQTLs as putative drivers of known disease associations 2013; **45:** 1238-1243.

18. Wang JC, Cruchaga C, Saccone NL *et al*: Risk for nicotine dependence and lung cancer is conferred by mRNA expression levels and amino acid change in CHRNA5. *Hum Mol Genet* 2009; **18:** 3125-3135.
19. Qiu W, Cho MH, Riley JH *et al*: Genetics of sputum gene expression in chronic obstructive pulmonary disease. *PLoS One* 2011; **6:** e24395.
20. Nguyen JD, Lamontagne M, Couture C *et al*: Susceptibility loci for lung cancer are associated with mRNA levels of nearby genes in the lung. *Carcinogenesis* 2014; **35:** 2653-2659.
21. DeMeo DL, Mariani T, Bhattacharya S *et al*: Integration of genomic and genetic approaches implicates IREB2 as a COPD susceptibility gene. *The American Journal of Human Genetics* 2009; **85:** 493-502.
22. Vucic EA, Chari R, Thu KL *et al*: DNA methylation is globally disrupted and associated with expression changes in chronic obstructive pulmonary disease small airways. *Am J Respir Cell Mol Biol* 2014; **50:** 912-922.
23. Yet I, Tsai PC, Castillo-Fernandez JE, Carnero-Montoro E, Bell JT: Genetic and environmental impacts on DNA methylation levels in twins. *Epigenomics* 2016; **8:** 105-117.
24. Ladd-Acosta C, Fallin MD: The role of epigenetics in genetic and environmental epidemiology. *Epigenomics* 2016; **8:** 271-283.
25. Zeilinger S, Kuhnel B, Klopp N *et al*: Tobacco smoking leads to extensive genome-wide changes in DNA methylation. *PLoS One* 2013; **8:** e63812.
26. Hillemacher T, Frieling H, Moskau S *et al*: Global DNA methylation is influenced by smoking behaviour. *European Neuropsychopharmacology* 2008; **18:** 295-298.
27. Qiu W, Baccarelli A, Carey VJ *et al*: Variable DNA methylation is associated with chronic obstructive pulmonary disease and lung function. *Am J Respir Crit Care Med* 2012; **185:** 373-381.
28. Paliwal A, Vaissière T, Krais A *et al*: Aberrant DNA methylation links cancer susceptibility locus 15q25. 1 to apoptotic regulation and lung cancer. *Cancer research* 2010; **70:** 2779-2788.
29. Bossé Y: Updates on the COPD gene list. *Int J Chron Obstruct Pulmon Dis* 2012; **7:** 607-631.
30. Ligthart S, Steenaard RV, Peters MJ *et al*: Tobacco smoking is associated with DNA methylation of diabetes susceptibility genes. *Diabetologia* 2016; **59:** 998-1006.
31. Netherlands BaBrriot: BIOS 2016, Vol 2016.
32. Cui K, Ge X, Ma H: Four SNPs in the CHRNA3/5 alpha-neuronal nicotinic acetylcholine receptor subunit locus are associated with COPD risk based on meta-analyses. *PLoS One* 2014; **9:** e102324.
33. Hardin M, Zielinski J, Wan ES *et al*: CHRNA3/5, IREB2, and ADCY2 are associated with severe chronic obstructive pulmonary disease in Poland. *Am J Respir Cell Mol Biol* 2012; **47:** 203-208.
34. Kim WJ, Wood AM, Barker AF *et al*: Association of IREB2 and CHRNA3 polymorphisms with airflow obstruction in severe alpha-1 antitrypsin deficiency. *Respir Res* 2012; **13:** 16.
35. Zhou H, Yang J, Li D *et al*: Association of IREB2 and CHRNA3/5 polymorphisms with COPD and COPD-related phenotypes in a Chinese Han population. *J Hum Genet* 2012; **57:** 738-746.
36. Saccone NL, Culverhouse RC, Schwantes-An TH *et al*: Multiple independent loci at chromosome 15q25.1 affect smoking quantity: a meta-analysis and comparison with lung cancer and COPD. *PLoS Genet* 2010; **6**.
37. Hancock DB, Wang J-C, Gaddis NC *et al*: A multiancestry study identifies novel genetic associations with CHRNA5 methylation in human brain and risk of nicotine dependence. *Human molecular genetics* 2015**:** ddv303.

38. Wan ES, Qiu W, Carey VJ *et al*: Smoking-Associated Site-Specific Differential Methylation in Buccal Mucosa in the COPDGene Study. *American journal of respiratory cell and molecular biology* 2015; **53:** 246-254.
39. Tsaprouni LG, Yang T-P, Bell J *et al*: Cigarette smoking reduces DNA methylation levels at multiple genomic loci but the effect is partially reversible upon cessation. *Epigenetics* 2014; **9:** 1382-1396.
40. Shenker NS, Ueland PM, Polidoro S *et al*: DNA methylation as a long-term biomarker of exposure to tobacco smoke. *Epidemiology* 2013; **24:** 712-716.
41. Joehanes R, Just AC, Marioni RE *et al*: Epigenetic Signatures of Cigarette Smoking. *Circ Cardiovasc Genet* 2016; **9:** 436-447.
42. Cheng Z, Dai LL, Song YN *et al*: Regulatory effect of iron regulatory protein-2 on iron metabolism in lung cancer. *Genet Mol Res* 2014; **13:** 5514-5522.
43. Ning B, Liu G, Liu Y *et al*: 5-aza-2'-deoxycytidine activates iron uptake and heme biosynthesis by increasing c-Myc nuclear localization and binding to the E-boxes of transferrin receptor 1 (TfR1) and ferrochelatase (Fech) genes. *J Biol Chem* 2011; **286:** 37196-37206.
44. Agustí A: Systemic effects of chronic obstructive pulmonary disease: what we know and what we don't know (but should). *Proceedings of the American Thoracic Society* 2007; **4:** 522-525.
45. Lokk K, Modhukur V, Rajashekar B *et al*: DNA methylome profiling of human tissues identifies global and tissue-specific methylation patterns. *Genome Biol* 2014; **15:** r54.
46. Yang X, Shao X, Gao L, Zhang S: Systematic DNA methylation analysis of multiple cell lines reveals common and specific patterns within and across tissues of origin. *Hum Mol Genet* 2015; **24:** 4374-4384.
47. Sterk PJ: Let's not forget: the GOLD criteria for COPD are based on post-bronchodilator FEV_1. *Eur Respir J* 2004; **23:** 497-498.
48. Mannino DM, Diaz-Guzman E, Buist S: Pre-and post-bronchodilator lung function as predictors of mortality in the Lung Health Study. *Respiratory research* 2011; **12:** 1.
49. Hindorff LA, Sethupathy P, Junkins HA *et al*: Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci U S A* 2009; **106:** 9362-9367.
50. Maurano MT, Humbert R, Rynes E *et al*: Systematic localization of common disease-associated variation in regulatory DNA. *Science* 2012; **337:** 1190-1195.

Chapter 2.3

COPD GWAS variant at 19q13.2 in relation with DNA methylation and gene expression

Ivana Nedeljković, Lies Lahousse, Elena Carnero-Montoro, Alen Faiz, Judith M Vonk, Kim de Jong, Diana A van der Plaat, Cleo C van Diemen, Maarten van den Berge, Ma'en Obeidat, Yohan Bossé, David C Nickle, BIOS Consortium, Andre G Uitterlinden, Joyce B J van Meurs, Bruno H C Stricker, Guy G Brusselle, Dirkje S Postma, H Marike Boezen, Cornelia M van Duijn, Najaf Amin

This chapter is published in Human Molecular Genetics, October 2017.
The supplemental material for this paper is available at:
https://doi.org/10.1093/hmg/ddx390

ABSTRACT

Chronic obstructive pulmonary disease (COPD) is among the major health burdens in adults. While cigarette smoking is the leading risk factor, a growing number of genetic variations have been discovered to influence disease susceptibility. Epigenetic modifications may mediate the response of the genome to smoking and regulate gene expression. Chromosome 19q13.2 region is associated with both smoking and COPD, yet its functional role is unclear. Our study aimed to determine whether rs7937 (*RAB4B, EGLN2*), a top genetic variant in 19q13.2 region identified in genome-wide association studies of COPD, is associated with differential DNA methylation in blood (N=1490) and gene expression in blood (N=721) and lungs (N=1087).

We combined genetic and epigenetic data from the Rotterdam Study (RS) to perform the epigenome-wide association analysis of rs7937. Further, we used genetic and transcriptomic data from blood (RS) and from lung tissue (Lung expression quantitative trait loci mapping study), to perform the transcriptome-wide association study of rs7937. Rs7937 was significantly (FDR<0.05) and consistently associated with differential DNA methylation in blood at 4 CpG sites in cis, independent of smoking. One methylation site (cg11298343-*EGLN2)* was also associated with COPD (*P*=0.001). Additionally, rs7937 was associated with gene expression levels in blood in cis (*EGLN2*), 42% mediated through cg11298343, and in lung tissue, in cis and trans (*NUMBL*, *EGLN2, DNMT3A, LOC101929709* and *PAK2*).

Our results suggest that changes of DNA methylation and gene expression may be intermediate steps between genetic variants and COPD, but further causal studies in lung tissue should confirm this hypothesis.

INTRODUCTION

Chronic obstructive pulmonary disease (COPD) is a common, systemic, lung disease, mainly characterized by airway obstruction and inflammation.[1] COPD often develops as a response to chronic exposure to cigarette smoke, fumes and gases.[2, 3] There is significant inter-individual variability in the response to these environmental exposures[4, 5] that has been attributed to genetic factors.[6, 7] Genome-wide association studies (GWAS) have identified genetic variants associated with COPD susceptibility on chromosomes 4q31, 4q22, 15q25 and 19q13.[8-11] However, the mechanism explaining how these variants are involved in the pathogenesis of COPD remains elusive.[12]

As is the case for many complex diseases, many single nucleotide polymorphisms (SNPs) associated with COPD and lung function by GWAS are located in non-protein coding intergenic and intronic regulatory regions.[13, 14] It has been hypothesized that these SNPs may modulate regulatory mechanisms, such as RNA expression, splicing, transcription factor binding and epigenetic modifications (e.g. DNA methylation). Changes in RNA expression as well as in DNA methylation regulating expression have recently been associated with COPD, suggesting that genetic and epigenetic factors are working in concert in the pathogenesis of COPD.[15, 16] Emerging evidence suggests that differential methylation sites (CpGs) are potentially important for COPD susceptibility,[15-17] but their location was not linked to the GWAS loci. However, important associations in COPD genomic regions may have been missed as arrays with limited coverage (27K) were used in the studies conducted to date.

The 19q13.2 region is associated with COPD and cigarette smoking,[18, 19] lung function[20] and emphysema patterns.[21] Genes in this region include *RAB4B* (member RAS oncogene family), *EGLN2* (Egl-nine homolog 2), *MIA* (melanoma inhibitory activity) and *CYP2A6* (cytochrome P450 family 2 subfamily A member 6). The top variant in the region, rs7937:C>T, has been identified by Cho et al..[9] This SNP (*RAB4B, EGLN2*) was associated with COPD (OR=1.37, P=2.9×10^{-9}), but not with smoking. Nevertheless, in a study of 10 healthy non-smokers and 7 healthy smokers, *EGLN2* was found to be expressed at a higher level in airway epithelium of smokers compared to non-smokers.[22] In this small, underpowered study of airway epithelial DNA, there was no significant evidence for differential DNA methylation of *EGLN2* between smokers and non-smokers.

In this study we set out to determine whether rs7937 is involved in regulatory mechanisms like DNA methylation and gene expression and whether these mechanisms are also associated with COPD. We further evaluated the role of smoking in these regulatory mechanisms. For that purpose, we performed an epigenome-wide association study (EWAS) of rs7937 in blood using an array with high coverage (450K) and a transcriptome-wide association study of rs7937 in blood and lung tissues.

RESULTS

Our discovery cohort comprised 724 participants with genotype and DNA methylation data, while the replication cohort comprised 766 participants from the Rotterdam Study (RS).[23] The summary statistics of the discovery and replication cohorts are shown in **Table 1.**

Table 1. Characteristics of the discovery and replication cohorts and per COPD status

	Discovery cohort			Replication cohort		
	COPD	**Controls**	**All**	**COPD**	**Controls**	**All**
N (% of all)	114 (15.7)*	541 (74.7)	724	93 (12.1)*	591 (77.2)	766
Age (years)*	61.9±8.6	59.3±7.9	59.9±8.2	68.2±5.7	67.6±5.9	67.7±5.9
Males (%)	68 (59.6)	233 (43.1)	331 (45.7)	54 (58.1)	249 (42.1)	324 (42.3)
FEV_1/FVC (% of all)	0.63±0.07 (71.9)	0.78±0.04 (66.0)	0.76±0.08 (67.0)	0.63±0.07 (95.7)	0.79±0.05 (91.0)	0.76±0.08 (91.8)
Current smokers, n (%)*	43 (37.7)	107 (19.8)	168 (23.2)	20 (21.5)	52 (8.8)	80 (10.4)
Ex-smokers, n (%)	55 (48.2)	239 (44.2)	322 (44.5)	51 (54.8)	326 (55.2)	427 (55.7)
Never smokers, n (%)	16 (14.0)	195 (36.0)	234 (32.3)	22 (23.7)	213 (36.0)	259 (33.8)
Pack-years#	34.3±26.9	19.9±19.6	23.2±22.0	33.7±18.7	19.6±20.1	21.9±20.6

Data for quantitative measures presented as mean±SD. COPD: Chronic Obstructive Pulmonary Disease cases, All: all participants included in EWAS. For traits that were not available for all participants (COPD status and FEV_1/FVC), the valid percentage is denoted in brackets (% of all). * Significantly different between the discovery and replication cohort. #Pack-years data were available for all participants (mean and SD calculated in current and ex-smokers only).

As expected, the prevalence of males, smokers and the average pack-years of smoking were higher in cases, compared to controls. Compared to the replication cohort, participants in the discovery cohort were on average 8 years younger and included significantly more COPD cases and current smokers, although the pack-years of smoking were comparable. The overview of the analysis pipeline and sample sizes used is presented in **Figure 1**.

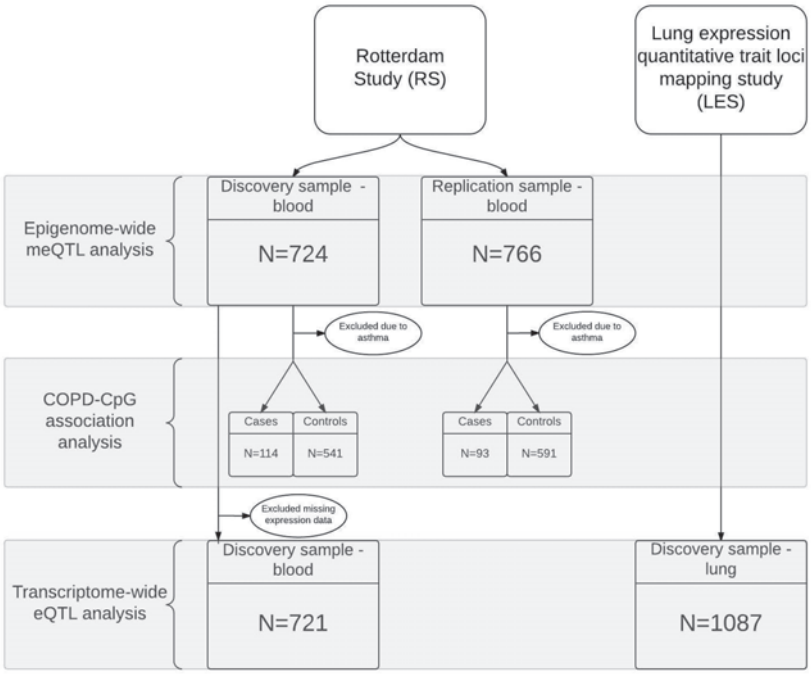

Figure 1. Analysis pipeline and datasets overview.

Methylation quantitative trait locus (meQTL) analysis

In the genome-wide blood meQTL analysis of rs7937 in the discovery cohort, rs7937 was significantly (False Discovery Rate (FDR) <0.05) associated with differential DNA methylation at 6 CpG sites in the genes *ITPKC* and *EGLN2*, located within the same 19q13.2 region (Model 1, **Table 2, Figure 2A**). Five of the six methylation sites were available in the replication dataset and four were significantly replicated with the same direction as found in the discovery cohort (**Table 2, Figures 2B** and **3**). Adding smoking as a confounder (Model 2, **Table 2**) and testing interaction with smoking (Model 3, data not shown) did not change the results, suggesting that the association between rs7937 and DNA methylation at these sites is independent of smoking. In an additional Model 4, we show that adding COPD to the model did not change the effect of rs7937 on DNA methylation (**Supplementary Table 1**).

Table 2. Association of rs7937-T with epigenome-wide DNA methylation in discovery (N=724) and replication (N=766) cohorts

		Discovery				Replication			
CpG	**Model**	**β**	**SE**	**P**	**FDR**	**β**	**SE**	**P**	**FDR**
cg21653913	1	-0.08925	0.00496	9.17×10^{-59}	4.35×10^{-53}	-0.10030	0.00491	1.13×10^{-72}	4.74×10^{-67}
19:41307778 *EGLN2*	2	-0.08956	0.00497	5.70×10^{-59}	2.71×10^{-53}	-0.10059	0.00492	1.17×10^{-72}	4.92×10^{-67}
cg11298343	1	-0.02036	0.00201	1.94×10^{-22}	3.53×10^{-17}	-0.01410	0.00141	5.42×10^{-22}	6.36×10^{-17}
19:41306150 *EGLN2*	2	-0.02070	0.00199	1.37×10^{-23}	3.25×10^{-18}	-0.01402	0.00140	5.30×10^{-22}	7.42×10^{-17}
cg10585486	1	-0.00901	0.00089	2.23×10^{-22}	3.53×10^{-17}	-0.01349	0.00130	1.47×10^{-23}	3.09×10^{-18}
19:41304133 *EGLN2*	2	-0.00902	0.00089	2.19×10^{-22}	3.47×10^{-17}	-0.01347	0.00130	2.25×10^{-23}	4.72×10^{-18}
cg24958765	1	0.00235	0.00031	1.17×10^{-13}	9.27×10^{-9}	-0.00427	0.00075	2.02×10^{-8}	1.06×10^{-3}
19:41283667 *RAB4B*	2	0.00237	0.00031	1.02×10^{-13}	8.06×10^{-9}	-0.00433	0.00075	1.26×10^{-8}	6.62×10^{-4}
cg13791183	1	0.01244	0.00209	4.23×10^{-9}	2.51×10^{-4}	NA			
19:41316697 *CYP2A6-AK097370*	2	0.01255	0.00209	3.28×10^{-9}	1.94×10^{-4}				
cg25923056	1	-0.00952	0.00177	1.00×10^{-7}	5.30×10^{-3}	-0.01153	0.00141	1.50×10^{-15}	1.05×10^{-10}
19:41306455 *EGLN2*	2	-0.00985	0.00175	2.54×10^{-8}	1.34×10^{-3}	-0.01140	0.00141	2.73×10^{-15}	1.91×10^{-10}

β: Regression coefficient estimates from linear regression model regressing DNA methylation levels on indicated SNPs. In Model 1 coefficients are corrected for sex, age, technical covariates and different white blood cellular proportions. In Model 2 coefficients are additionally adjusted for current smoking and pack-years smoked. SE: standard error of the effect, P: p-value of the significance, FDR: False discovery rate value. NA: Not available in the replication dataset.

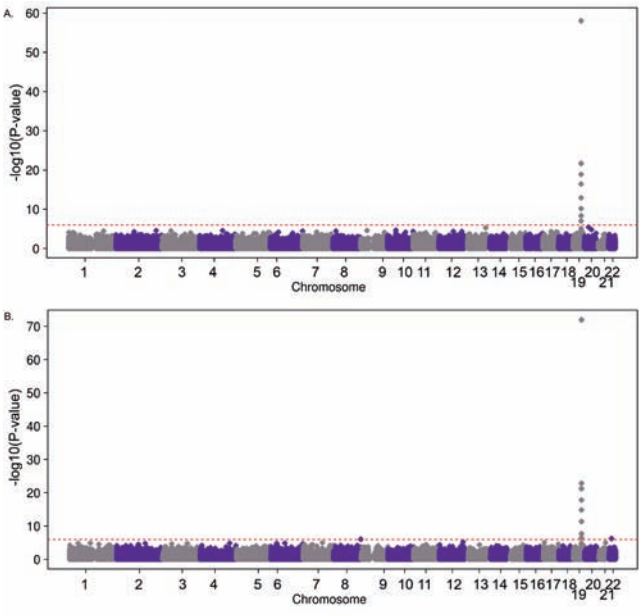

Figure 2. Association of the rs7937 with DNA methylation across the genome. In circles are represented all CpGs throughout the genome. X-axis shows chromosome locations; Y-axis shows negative logarithm of the p-value of the associations of the SNP with each CpG site. Red line represents the significance threshold (FDR<0.05). A. Discovery analysis; B. Replication analysis.

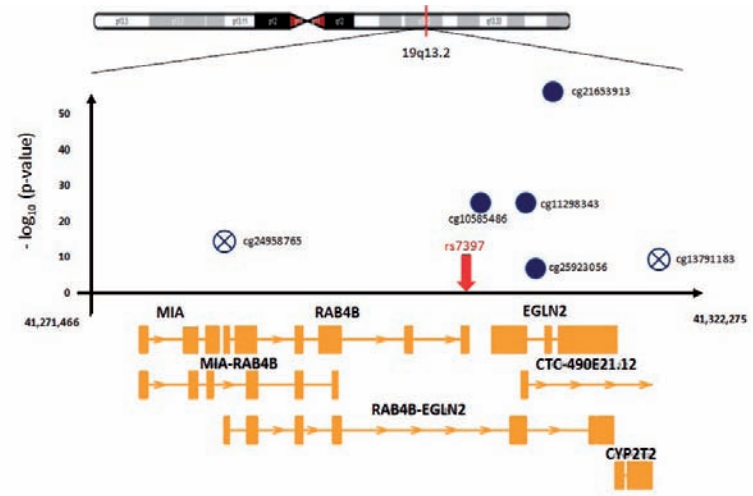

Figure 3. Region plot of chromosome 19q13.2 with significant SNP-CpG associations. The circles represent SNP-CpG associations; X-axis shows all genes in the region; Y-axis shows negative logarithm of the p-values of the associations of CpGs with the SNP. Crossed circle represents the non-replicated associations.

COPD and FEV_1/FVC analyses

When testing for association of DNA methylation at the four replicated differentially methylated CpG sites with COPD, we observed a significant association with cg11298343 (*EGLN2*) in Model 1 (β (SE)=-7.080 (2.16), *P*=0.001) (**Table 3, Supplementary Table 2**), which remained nominally significant with diminished but still strong and concordant negative effect, after adjusting for smoking (Model 2; β (SE)=-4.924 (2.25), *P*=0.029). Further we show that additionally adjusting for rs7937 slightly deteriorated the effect of cg11298343 on COPD (Model 4; **Supplementary Table 1**).

In the association with the quantitative determinant of COPD (**Table 3, Supplementary Table 2**), the ratio of forced expiratory volume in one second (FEV_1) over the forced vital capacity (FVC), we observed nominal significance for the same site (Model 1; β (SE)=0.138 (0.07), *P*=0.04), which deteriorated with adjusting for smoking (Model 2; β (SE)=0.047 (0.06), *P*=0.46).

Table 3. Association of DNA methylation at significant CpG sites in 19q13.2, with COPD and FEV_1/FVC ratio – meta-analysis results

Trait	CpG	N	Model	β	SE	P
COPD	cg21653913	1339	1	-0.084	0.711	0.906
			2	0.175	0.738	0.812
	cg11298343	1339	1	-7.080	2.163	**0.001**
			2	-4.924	2.254	**0.029**
	cg10585486	1339	1	-2.215	3.657	0.545
			2	-3.110	3.811	0.415
	cg25923056	1339	1	-0.368	2.153	0.864
			2	1.551	2.255	0.492
FEV_1/FVC	cg21653913	1188	1	-0.008	0.020	0.676
			2	-0.017	0.020	0.382
	cg11298343	1188	1	0.138	0.065	**0.035**
			2	0.047	0.064	0.464
	cg10585486	1188	1	0.096	0.093	0.304
			2	0.094	0.090	0.299
	cg25923056	1188	1	-0.021	0.064	0.743
			2	-0.098	0.062	0.114

N: number of participants in the meta-analysis; Model 1 is adjusted for age, sex, technical covariates and different white blood cellular proportions, Model 2 is additionally adjusted for smoking; β: Regression coefficient estimates from logistic/linear regression models; SE: standard error of the effect; P: p-value of the significance. In bold: nominally significant associations.

Blood and lung expression quantitative trait loci (eQTL) analysis

In the genome-wide blood eQTL analysis in RS, rs7937 was significantly associated with differential expression of the ILMN_2354391 probe in the *EGLN2* gene (β (SE)=0.064 (0.01), P=9.3×10^{-9}, **Table 4**). The risk allele (T) was associated with increased expression of *EGLN2* (**Supplementary figure 1**). The association signal dropped albeit remained significant (β (SE)=0.058 (0.01), P=1.9×10^{-6}) after adjusting for the top CpG site cg11298343 in the same gene (**Table 4**).

Table 4. Association of rs7937 with transcriptome-wide gene expression in blood (N=721)

SNP	A1	A2	Probe	Chr	Position (GRCh37/ hg19)	Gene	Model	β	SE	P	FDR
rs7937	T	C	NM_080732.1	19	46006086-46006135	*EGLN2*	1	0.0635	0.0109	9.29×10^{-9}	0.000197
							2	0.0577	0.0120	1.88×10^{-6}	0.039942

A1: effect allele (tested allele), A2: alternative allele, Chr: chromosome of the probe, Model 1 is adjusted for age, sex, current smoking, technical covariates and different white blood cellular proportions, Model 2 is additionally adjusted for cg11298343; β: Regression coefficient estimates from linear regression models regressing gene expression on indicated SNP, SE: standard error, P: p-value of the significance, FDR: False discovery rate value.

Suggesting that differential DNA methylation at cg11298343 is partly responsible for the differential expression of ILMN_2354391 in *EGLN2* in blood. In further investigation we performed the formal mediation analysis where we show (**Table 5**) that 42% of the association between rs7937 and *EGLN2* expression is indeed mediated through cg11298343 (P=0.04).

Table 5. Mediation of the rs7937-*EGLN2* expression association through cg11298343 (N=721)

	Estimate	95%CI Lower	95%CI Upper	P-value
ACME	0.02	0.005	0.032	**0.01**
ADE	0.02	-0.014	0.068	0.22
Total Effect	0.04	0.005	0.080	**0.03**
Proportion Mediated	**0.42**	0.058	1.995	**0.04**

ACME: average causal mediation effect by DNA methylation at cg11298343; ADE: average direct effect of rs7937 on EGLN2 expression; Total: total effect rs7937 on EGLN2 expression; Proportion Mediated: proportion of the association between rs7937 and EGLN2 expression, explained by methylation at cg11298343; Regression models adjusted for sex, age, current smoking, pack-years, technical variance and estimated blood cell composition; p-value < 0.05.

Moreover, genome-wide eQTL analysis of rs7937 in lung tissue of 1087 participants from the Lung expression quantitative loci mapping study (LES), showed significant associations ($P<1.36\times10^{-6}$) with 5 probes in the same region (*in cis; AK097370(EGLN2)*, *NUMBL*) and other chromosomes (*in trans;* LOC101929709, *DNMT3A* and *PAK2*) (**Table 6**). In all cases, except one (LOC101929709 at chromosome 8), the T allele of rs7937 was consistently associated with decreased expression of the genes in lung tissue (**Table 6, Supplementary figure 2**).

Table 6. Association of rs7937 with transcriptome-wide gene expression in lung tissue (N=1087)

SNP	A1	A2	Probe	Chr	Position (GRCh37/ hg19)	Annotation	β	SE	P	FDR
rs7937 (T/C)	T	C	NM_004756	19	40665906-40690658	*NUMBL*	-0.099	0.013	4.39×10^{-15}	1.17×10^{-11}
			BC037804	19	40808443-40810818	*AK097370* (*EGLN2*)	-0.077	0.012	3.77×10^{-10}	1.01×10^{-6}
			BX330016	8	89720919-89724906	*LOC101929709*	0.049	0.011	6.15×10^{-6}	0.016
			AK025230	2	25233434-25246179	*DNMT3A*	-0.028	0.006	8.00×10^{-6}	0.021
			BQ445924	3	196829093-196830253	*PAK2*	-0.028	0.006	1.36×10^{-5}	0.036

β: Regression coefficient estimates from linear regression models regressing gene expression on indicated SNP, SE: standard error, P: p-value of the significance, FDR: False discovery rate value.

DISCUSSION

Our study shows that the rs7937 in 19q13.2 is associated with differential blood DNA methylation of 4 CpG sites located in the *EGLN2* in the discovery and replication cohort. The COPD risk allele (T) is associated with lower DNA methylation at these sites. These relationships are independent of smoking and of COPD. We further show that DNA methylation in blood at cg11298343 (*EGLN2*) is associated with COPD and remains nominally significant after adjusting for smoking. Finally, rs7937 is associated with differential expression in blood of *EGLN2*, 42% explained by *EGLN2* DNA methylation at site cg11298343, and in lung tissue of *NUMBL, AK097370 (EGLN2), LOC101929709, DNMT3A* and *PAK2*.

EGLN2 is coding prolyl hydroxylase domain-containing protein 1 (PHD1) which regulates posttranscriptional modifications of hypoxia induced factor (HIF), a transcriptional complex involved in oxygen homeostasis. At normal oxygen levels, the alpha

subunit of HIF is targeted for degradation by PHD1, which is an essential component of the pathway through which cells sense oxygen,[24] is also known to be involved in activation of inflammatory and immune genes, including those implicated in COPD.[25] Furthermore, read-through transcription exists between this gene and the upstream *RAB4B*, and together they were shown to be involved in invasive lung cancer.[26]

Our study of DNA methylation in blood replicates the findings of a study on meQTLs in blood across the human life course: during pregnancy, at birth, childhood, adolescence and middle age[27] which reported three (cg10585486, cg11298343, cg25923056) out of our four replicated CpGs. We additionally report a novel finding in association with rs7937, our top hit, cg21653913. Interestingly, they report differential DNA methylation at cg11298343 to be associated with rs7937 at all five time points. In the present study, we now show that rs7937 is also associated with cg11298343 in our elderly sample, the age category at highest risk for developing COPD. This finding goes in line with our hypothesis that the life-long change in the DNA methylation is involved in the pathogenesis and onset of COPD in older age, rather than the other way around. However, further longitudinal studies are needed, testing this hypothesis in lung tissue.

In line with our findings, rs7937 has previously been associated with differential expression of *EGLN2* in blood.[28] Having both DNA methylation and transcription data available, we could further test whether the relation of rs7937 and *EGLN2* expression could be explained by DNA methylation levels of *EGLN2* at cg11298343 site. We show for the first time that in blood there is indeed a mediation of 42% through the DNA methylation at cg11298343, confirming our hypothesis. We report two novel findings in this region in the lung tissue. We found that rs7937 is involved in expression of *AK097370* in lung tissue, a DNA clone in the proximity of *EGLN2*, as well as with *NUMBL*. *NUMBL* is known as a negative regulator of NF-kappa-B signaling pathway in neurons[29] and was also found to be expressed in the lungs in the GTEx database.[30] The association between rs7937 and gene expression in the same dataset has been tested earlier by Lamontagne et al.[31] but no significant results were reported. In the present study, using a more powerful meta-analysis approach, rs7937 was associated with two loci in the region (*NUMBL* and *AK097370,* close to *EGLN2*) and three more loci in other chromosomes (*LOC101929709, DNMT3A* and *PAK2*). Taken together, our findings raise the hypothesis that the genetic effect of rs7937 on COPD might be mediated by DNA methylation at cg11298343 and subsequent alteration of expression of *EGLN2* and other genes in this region, such as *NUMBL*. We show this in blood but further formal mediation analyses in lung tissue are needed to confirm this hypothesis, requiring the assembly of a large dataset of lung tissue characterized for genetic, epigenetic and transcriptomic data.

Furthermore, we have found significant associations in lungs of rs7937 in *trans,* i.e. with the expression of genes on other chromosomes. These effects include differential expression of *PAK2* (chromosome 3), *DNMT3A* (chromosome 2) and long non-coding RNA on chromosome 8. The protein encoded by *PAK2* gene is activated by proteolytic cleavage during caspase-mediated apoptosis, and may play a role in regulating the apoptotic events in the dying cell.[32] *DNMT3A* is the gene encoding the DNA methyltransferase which plays a key role in *de novo* methylation. This may imply that rs7937 is involved in the pathogenesis of COPD through differential DNA methylation and regulation of expression throughout the genome, again asking for further research of DNA methylation.

The strength of our analysis is the use of large and unique samples of patients whose genetic, epigenetic and transcriptomic characteristics were assessed in detail. However, a limitation of our study is the use of blood tissue for the assessment of DNA methylation and gene expression. Nevertheless, our findings regarding the role of the genetic variants in blood corroborate with the changes in the transcriptome in lung tissue. It has been shown that blood can be used to evaluate methylation changes related to COPD and smoking, as the disease induces systemic changes associated with elevated markers of systemic inflammation in blood.[25] A second limitation of our study is that we cannot distinguish the expression in lung parenchymal tissue, which comprises multiple cell types. It may be speculated that expression of only distinct cells is affected by rs7939. If this is the case, the most likely effect is that the power of our study is reduced but has not biased our findings in the sense of generating false positives. However, although eQTLs are frequently cell- and tissue-specific,[33] many eQTLs are also shared across tissues.[34] Finally, the number of patients with COPD and spirometry measures were limited, which may have compromised the power of the study. Nevertheless, we observed significant findings that are relevant for COPD.

In conclusion, our findings suggest that genetic variations underlying *EGLN2* methylation contribute to the risk of developing COPD. This finding adds insight into how genetic variants are involved in th e pathogenesis of COPD, through differential DNA methylation and regulation of expression, irrespective of smoking. Future integrative studies involving genetics, epigenetics and transcriptomics in lung tissue are crucial to elucidate the molecular mechanisms behind COPD genetic susceptibility and to translate the findings to clinical care and prevention. This may lead to an increased specificity and sensitivity of diagnostic and prognostic tools. In addition, novel DNA methylation loci may be used as a target for future drug design in COPD. While smoking cessation is shown to be a useful prevention tool for disease risk and mortality reduction, DNA methylation loci independent of smoking may be used as a target for a more personalized and focused treatment approach.

MATERIALS AND METHODS

Study population

For our analyses in blood we used two independent subsets of participants from the RS.[23] The full discovery set for meQTL analysis was comprised of 724 participants with full genomic and epigenetic data, while replication set included 766 participants. The replication subset is part of the Biobanking and Biomolecular Resources Research Infrastructure for The Netherlands (BBMRI-NL), BIOS (Biobank-based Integrative Omics Studies) project.[35] RS has been approved by the Medical Ethics Committee of the Erasmus MC and by the Ministry of Health, Welfare and Sport of the Netherlands, implementing the Population Studies Act: Rotterdam Study. All participants provided written informed consent to participate in the study and to obtain information from their treating physicians. The detailed information on our samples can be found in Online supplementary material.

Spirometry measures and COPD diagnosis

From the initial full datasets for meQTL analysis ($n_{discovery}$=724, $n_{replication}$=766), after excluding participants with asthma, we used data from 655 participants in the discovery and 684 participants in the replication cohort for the association analyses with COPD. COPD diagnosis was defined as pre-bronchodilator $FEV_1/FVC<0.7$. More detailed information can be found in Online supplementary material.

COPD SNPs selection

Using the GWAS catalog[36] on 15th January 2017, we performed a search with the term "19q13.2", additionally applying filters for the p-value$\leq 5\times10^{-8}$ and for the trait to include term "Chronic obstructive pulmonary disease". One GWAS study passed these filtering criteria, and reported the top SNP, rs7937-T (NC_000019.10:g.40796801C>T), to be associated with COPD (OR=1.37).[9] We used this SNP to perform all analyses in our study.

Genotyping in RS

Genotyping was performed using 610K and 660K Illumina arrays for which whole blood genomic DNA was used. Detailed information can be found in Online supplementary material. Imputation was done using 1000 Genomes (1KG) phase I v3 reference panel, with measured genotypes that had minor allele frequencies (MAF)>1%, performed with MacH software and Minimac implementation. We extracted dosages of rs7937 (risk allele T, allele frequency (AF) =0.54), from RS imputed data using DatABEL library of R-package.[37]

DNA methylation array in RS

We used Illumina Infinium Human Methylation 450K array to quantify DNA methylation levels across the genome from whole blood in RS. The detailed QC and normalization criteria can be found in the Online supplementary material. Ultimately, after the QC and normalization steps our discovery set included 724 Caucasian participants and 463,456 probes, while the replication set included 766 Caucasian participants and 419,936 probes.

RNA array in blood in RS

In the discovery sample we used the same blood samples at baseline to isolate RNA, which we hybridized to Illumina Whole-Genome Expression Beadchips Human HT-12 v4 array. Raw probe intensities were quantile-normalized, and 2-log transformed and controlled for quality as described elsewhere.[28] After all normalization and QC steps the sample consisted of 21.238 probes in 721 participants with available full data on SNP and RNA arrays and all covariates.

RNA array in lung tissue

Gene expression was quantified using lung tissue samples obtained from patients that underwent lung resection surgery at three facilities participating in the LES: University of Groningen (GRN), Laval University (Laval) and University of British Columbia (UBC).[38] Illumina Human1M-Duo BeadChip arrays were used for genotyping, and a custom Affymetrix microarray (GPL10379) for gene expression profiling. The final dataset for the eQTL analysis consisted of 1087 subjects. More detailed information can be found in Online supplementary material.

Statistical analyses

meQTL analysis

We performed EWAS in the discovery cohort using linear regression analysis with rs7937-T as independent variable and DNA methylation sites as dependent variable. We fitted two models; first adjusted for age, sex, technical covariates to correct for batch effects (array number and position on array) and the estimated white blood cell counts[39] (including monocytes, T-lymphocytes: CD4 and CD8, B-lymphocytes, natural killer cells, neutrophils and eosinophils) (Model 1); and second, for significant sites additionally adjusted for current smoking and pack-years smoked (Model 2). We used the FDR<0.05 as an epigenome-wide significance threshold.[40] Significant sites from Model 1 were then tested for association in the replication cohort using the same models as in the discovery. Since the 19q13.2 region was also

implicated in smoking behavior, we also tested significant CpG sites in a third model including 'rs7937×smoking' interaction term to assess possible interaction between rs7937 and smoking (for both current smoking and pack-years smoked), in the discovery and replication cohorts. For the significant sites we used another model additionally adjusted for COPD status, which we compared to the Model 1 in attempt to further elucidate the direction of the effect between DNA methylation and COPD.

COPD and FEV_1/FVC analysis

To test if the significantly associated methylation sites are also associated with the lung phenotypes, we performed logistic and linear regression analyses with COPD and FEV_1/FVC ratio, respectively as dependent variables and DNA methylation as independent variable. In the first model we adjusted for age, sex, technical covariates and estimated white blood cell counts and additionally for current smoking and pack-years smoked in the second model, in both the discovery and the replication cohort. Results from the two cohorts were meta-analyzed using fixed effect models with "rmeta" package in R.[41] Bonferroni correction was applied to adjust for multiple testing. For the significant sites we used another model additionally adjusted for rs7937, which we compared to the first model in attempt to further elucidate the direction of the effect between DNA methylation and COPD.

Blood eQTL analysis

In the discovery cohort we tested whether rs7937 is associated with differential expression in the whole blood. We used linear regression analysis with rs7937 as independent variable and genome-wide normalized gene expression as dependent variable. For this analysis we used model adjusted for age, sex, current smoking, technical batch effects (plate ID and RNA quality) and white blood cell counts (lymphocytes, monocytes and granulocytes). For significant (FDR<0.05) probes, the second model was additionally adjusted for significant DNA methylation levels.

Mediation analysis

We have performed formal mediation analysis using the bootstrapping method in the "mediation" package in R,[42] to assess the potential mediator role of significant DNA methylation in the SNP-expression association. One thousand bootstraps were run to estimate the confidence intervals.[43] We used models adjusted for age, sex, current smoking, pack-years, expression technical batch effects (plate ID and RNA quality), methylation technical batch effects (position on array and array number) and estimated blood cell composition.

Lung eQTL analysis

To test if the SNP rs7937 is associated (FDR<0.05) with differential expression in lung tissue in LES, we performed a genome-wide linear regression analysis with the SNP as the independent variable and 2-log transformed gene expression levels as dependent variable. This analysis was performed for each of the three participating cohorts (GRN, Laval and UBC) separately, adjusted for lung disease status, age, sex, smoking status and cohort specific principal components (PCs). The inverse-variance weighted fixed effect meta-analysis of the results obtained from the three cohorts was performed with "*rmeta*" package in R software. The detailed overview of the fitted models can be found in the Online supplementary material.

REFERENCES

1 Lozano, R. (2012) Bin Abdulhak A, Birbeck G, Blyth F, Bolliger I, Boufous S, Bucello C, Burch M, et al: Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet, The*, **380**, 2095.

2 de Jong, K., Vonk, J.M., Timens, W., Bossé, Y., Sin, D.D., Hao, K., Kromhout, H., Vermeulen, R., Postma, D.S. and Boezen, H.M. (2015) Genome-wide interaction study of gene-by-occupational exposure and effects on FEV_1 levels. *Journal of Allergy and Clinical Immunology*, **136**, 1664-1672.e1614.

3 Gibson, G.J., Loddenkemper, R., Lundbäck, B. and Sibille, Y. (2013). European Respiratory Society, in press.

4 Burrows, B., Knudson, R.J., Cline, M.G. and Lebowitz, M.D. (1977) Quantitative Relationships between Cigarette Smoking and Ventilatory Function 1, 2. *American Review of Respiratory Disease*, **115**, 195-205.

5 Higgins, M.W., Keller, J.B., Landis, J.R., Beaty, T.H., Burrows, B., Demets, D., Diem, J.E., Higgins, I.T., Lakatos, E., Lebowitz, M.D. *et al.* (1984) Risk of chronic obstructive pulmonary disease. Collaborative assessment of the validity of the Tecumseh index of risk. *Am Rev Respir Dis*, **130**, 380-385.

6 McCloskey, S.C., Patel, B.D., Hinchliffe, S.J., Reid, E.D., Wareham, N.J. and Lomas, D.A. (2001) Siblings of patients with severe chronic obstructive pulmonary disease have a significant risk of airflow obstruction. *Am J Respir Crit Care Med*, **164**, 1419-1424.

7 Silverman, E.K., Chapman, H.A., Drazen, J.M., Weiss, S.T., Rosner, B., Campbell, E.J., O'Donnell, W.J., Reilly, J.J., Ginns, L., Mentzer, S. *et al.* (1998) Genetic epidemiology of severe, early-onset chronic obstructive pulmonary disease. Risk to relatives for airflow obstruction and chronic bronchitis. *Am J Respir Crit Care Med*, **157**, 1770-1778.

8 Boezen, H.M. (2009) Genome-wide association studies: what do they teach us about asthma and chronic obstructive pulmonary disease? *Proc Am Thorac Soc*, **6**, 701-703.

9 Cho, M.H., Castaldi, P.J., Wan, E.S., Siedlinski, M., Hersh, C.P., Demeo, D.L., Himes, B.E., Sylvia, J.S., Klanderman, B.J., Ziniti, J.P. *et al.* (2012) A genome-wide association study of COPD identifies a susceptibility locus on chromosome 19q13. *Hum Mol Genet*, **21**, 947-957.

10 Cho, M.H., McDonald, M.-L.N., Zhou, X., Mattheisen, M., Castaldi, P.J., Hersh, C.P., DeMeo, D.L., Sylvia, J.S., Ziniti, J. and Laird, N.M. (2014) Risk loci for chronic obstructive pulmonary disease: a genome-wide association study and meta-analysis. *The lancet Respiratory medicine*, **2**, 214-225.

11 Pillai, S.G., Ge, D., Zhu, G., Kong, X., Shianna, K.V., Need, A.C., Feng, S., Hersh, C.P., Bakke, P., Gulsvik, A. *et al.* (2009) A genome-wide association study in chronic obstructive pulmonary disease (COPD): identification of two major susceptibility loci. *PLoS Genet*, **5**, e1000421.

12 Huang, Q. (2015) Genetic study of complex diseases in the post-GWAS era. *J Genet Genomics*, **42**, 87-98.

13 Hindorff, L.A., Sethupathy, P., Junkins, H.A., Ramos, E.M., Mehta, J.P., Collins, F.S. and Manolio, T.A. (2009) Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci U S A*, **106**, 9362-9367.

14 Maurano, M.T., Humbert, R., Rynes, E., Thurman, R.E., Haugen, E., Wang, H., Reynolds, A.P., Sandstrom, R., Qu, H., Brody, J. *et al.* (2012) Systematic localization of common disease-associated variation in regulatory DNA. *Science*, **337**, 1190-1195.

15 Qiu, W., Baccarelli, A., Carey, V.J., Boutaoui, N., Bacherman, H., Klanderman, B., Rennard, S., Agusti, A., Anderson, W., Lomas, D.A. *et al.* (2012) Variable DNA methylation is associated with chronic obstructive pulmonary disease and lung function. *Am J Respir Crit Care Med*, **185**, 373-381.

16 Vucic, E.A., Chari, R., Thu, K.L., Wilson, I.M., Cotton, A.M., Kennett, J.Y., Zhang, M., Lonergan, K.M., Steiling, K., Brown, C.J. *et al.* (2014) DNA methylation is globally disrupted and associated with expression changes in chronic obstructive pulmonary disease small airways. *Am J Respir Cell Mol Biol*, **50**, 912-922.

17 Qiu, W., Wan, E., Morrow, J., Cho, M.H., Crapo, J.D., Silverman, E.K. and DeMeo, D.L. (2015) The impact of genetic variation and cigarette smoke on DNA methylation in current and former smokers from the COPDGene study. *Epigenetics*, **10**, 1064-1073.

18 Thorgeirsson, T.E., Gudbjartsson, D.F., Surakka, I., Vink, J.M., Amin, N., Geller, F., Sulem, P., Rafnar, T., Esko, T., Walter, S. *et al.* (2010) Sequence variants at CHRNB3-CHRNA6 and CYP2A6 affect smoking behavior. *Nat Genet*, **42**, 448-453.

19 Tobacco and Genetics, C. (2010) Genome-wide meta-analyses identify multiple loci associated with smoking behavior. *Nature genetics*, **42**, 441-447.

20 Artigas, M.S., Wain, L.V., Miller, S., Kheirallah, A.K., Huffman, J.E., Ntalla, I., Shrine, N., Trochet, H., McArdle, W.L. and Alves, A.C. (2015) Sixteen new lung function signals identified through 1000 Genomes Project reference panel imputation. *Nature communications*, **6**.

21 Castaldi, P.J., Cho, M.H., San José Estépar, R., McDonald, M.-L.N., Laird, N., Beaty, T.H., Washko, G., Crapo, J.D. and Silverman, E.K. (2014) Genome-wide association identifies regulatory Loci associated with distinct local histogram emphysema patterns. *American journal of respiratory and critical care medicine*, **190**, 399-409.

22 Ryan, D.M., Vincent, T.L., Salit, J., Walters, M.S., Agosto-Perez, F., Shaykhiev, R., Strulovici-Barel, Y., Downey, R.J., Buro-Auriemma, L.J., Staudt, M.R. *et al.* (2014) Smoking dysregulates the human airway basal cell transcriptome at COPD risk locus 19q13.2. *PLoS One*, **9**, e88051.

23 Hofman, A., Brusselle, G.G., Darwish Murad, S., van Duijn, C.M., Franco, O.H., Goedegebure, A., Ikram, M.A., Klaver, C.C., Nijsten, T.E., Peeters, R.P. *et al.* (2015) The Rotterdam Study: 2016 objectives and design update. *Eur J Epidemiol*, **30**, 661-708.

24 Bruick, R.K. and McKnight, S.L. (2001) A conserved family of prolyl-4-hydroxylases that modify HIF. *Science*, **294**, 1337-1340.

25 Cummins, E.P., Berra, E., Comerford, K.M., Ginouves, A., Fitzgerald, K.T., Seeballuck, F., Godson, C., Nielsen, J.E., Moynagh, P., Pouyssegur, J. *et al.* (2006) Prolyl hydroxylase-1 negatively regulates IkappaB kinase-beta, giving insight into hypoxia-induced NFkappaB activity. *Proc Natl Acad Sci U S A*, **103**, 18154-18159.

26 Hsu, Y.C., Yuan, S., Chen, H.Y., Yu, S.L., Liu, C.H., Hsu, P.Y., Wu, G., Lin, C.H., Chang, G.C., Li, K.C. *et al.* (2009) A four-gene signature from NCI-60 cell line for survival prediction in non-small cell lung cancer. *Clin Cancer Res*, **15**, 7309-7315.

27 Gaunt, T.R., Shihab, H.A., Hemani, G., Min, J.L., Woodward, G., Lyttleton, O., Zheng, J., Duggirala, A., McArdle, W.L., Ho, K. *et al.* (2016) Systematic identification of genetic influences on methylation across the human life course. *Genome Biol*, **17**, 61.

28 Westra, H.-J., Peters, M.J., Esko, T., Yaghootkar, H., Schurmann, C., Kettunen, J., Christiansen, M.W., Fairfax, B.P., Schramm, K., Powell, J.E. *et al.* (2013) Systematic identification of trans eQTLs as putative drivers of known disease associations. **45**, 1238-1243.

29 Ma, Q., Zhou, L., Shi, H. and Huo, K. (2008) NUMBL interacts with TAB2 and inhibits TNFalpha and IL-1beta-induced NF-kappaB activation. *Cell Signal*, **20**, 1044-1051.

30 Lonsdale, J., Thomas, J., Salvatore, M., Phillips, R., Lo, E., Shad, S., Hasz, R., Walters, G., Garcia, F., Young, N. *et al.* (2013) The Genotype-Tissue Expression (GTEx) project. **45**, 580-585.

31 Lamontagne, M., Couture, C., Postma, D.S., Timens, W., Sin, D.D., Pare, P.D., Hogg, J.C., Nickle, D., Laviolette, M. and Bosse, Y. (2013) Refining susceptibility loci of chronic obstructive pulmonary disease with lung eqtls. *PLoS One*, **8**, e70220.

32 Marlin, J.W., Eaton, A., Montano, G.T., Chang, Y.W. and Jakobi, R. (2009) Elevated p21-activated kinase 2 activity results in anchorage-independent growth and resistance to anticancer drug-induced cell death. *Neoplasia*, **11**, 286-297.

33 Fu, J., Wolfs, M.G., Deelen, P., Westra, H.J., Fehrmann, R.S., Te Meerman, G.J., Buurman, W.A., Rensen, S.S., Groen, H.J., Weersma, R.K. *et al.* (2012) Unraveling the regulatory mechanisms underlying tissue-dependent genetic variation of gene expression. *PLoS Genet*, **8**, e1002431.

34 Grundberg, E., Small, K.S., Hedman, A.K., Nica, A.C., Buil, A., Keildson, S., Bell, J.T., Yang, T.P., Meduri, E., Barrett, A. *et al.* (2012) Mapping cis- and trans-regulatory effects across multiple tissues in twins. *Nat Genet*, **44**, 1084-1089.

35 Biobanking and BioMolecular Resources research Infrastructure, B. (2016), Vol. 2016.

36 MacArthur, J., Bowler, E., Cerezo, M., Gil, L., Hall, P., Hastings, E., Junkins, H., McMahon, A., Milano, A., Morales, J. *et al.* (2017) The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucleic Acids Research*, **45**, D896-D901.

37 Aulchenko, Y., Struchalin, M., Ripke, S. and Johnson, T. (2010) GenABEL: genome-wide SNP association analysis. *R package version*, in press., 1.6-4.

38 Hao, K., Bossé, Y., Nickle, D.C., Paré, P.D., Postma, D.S., Laviolette, M., Sandford, A., Hackett, T.L., Daley, D. and Hogg, J.C. (2012) Lung eQTLs to help reveal the molecular underpinnings of asthma. *PLoS Genet*, **8**, e1003029.

39 Houseman, E.A., Accomando, W.P., Koestler, D.C., Christensen, B.C., Marsit, C.J., Nelson, H.H., Wiencke, J.K. and Kelsey, K.T. (2012) DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics*, **13**, 86.

40 Hochberg, Y. and Benjamini, Y. (1990) More powerful procedures for multiple significance testing. *Stat Med*, **9**, 811-818.

41 Lumley, T. and Lumley, M.T. (2006) The rmeta Package. in press.

42 Tingley, D., Yamamoto, T., Hirose, K., Keele, L. and Imai, K. (2014) Mediation: R package for causal mediation analysis. in press.

43 Mayer, A., Thoemmes, F., Rose, N., Steyer, R. and West, S.G. (2014) Theory and Analysis of Total, Direct, and Indirect Causal Effects. *Multivariate Behav Res*, **49**, 425-442.

Chapter 2.4

DNA methylation is associated with lung function in never-smokers

Maaike de Vries, Ivana Nedeljković, Diana A. van der Plaat, Alexandra Zhernakova, Lies Lahousse, Guy G. Brusselle, BIOS Consortium, Najaf Amin, Cornelia M. van Duijn, Judith M. Vonk, H. Marike Boezen

This chapter is submitted for publication.
The supplemental material for this paper is available at:
https://drive.google.com/drive/folders/1jDZzf55AvGEuODVGxzs0Qj7pIU5zKu7R?usp=sharing

ABSTRACT

Active smoking is the main risk factor for COPD. Here, epigenetic mechanisms may play a role, since cigarette smoking is associated with differential DNA methylation in whole blood. So far, it is unclear whether epigenetics also play a role in subjects with COPD who never smoked. Therefore, we aimed to identify differential DNA methylation associated with lung function in never-smokers.

We determined genome-wide DNA methylation levels of 396,243 CpG-sites (Illumina 450K) in blood of never smokers in four independent cohorts, LifeLines COPD&C (N=903), LifeLines DEEP (N=166), Rotterdam Study (RS)-III (N=150) and RS-BIOS (N=206). We meta-analysed the cohort-specific methylation results to identify differentially methylated CpG-sites with FEV_1/FVC. Expression Quantitative Trait Methylation (eQTM) analysis was performed in the Biobank-based Integrative Omics Studies (BIOS). A total of 36 CpG-sites were associated with FEV_1/FVC in never-smokers at p-value<0.0001, but the meta-analysis did not reveal any epi-genome wide significant CpG-sites. Of interest, 35 of these 36 CpG-sites have not been associated with lung function before in studies including subjects irrespective of smoking history. Among the top hits were cg10012512, cg02885771, annotated to the gene LTV1 Ribosome Biogenesis factor (*LTV1*), and, cg25105536, annotated to Kelch Like Family Member 32 (*KLHL32*). Moreover, a total of 11 eQTMS were identified.

With the identification of 35 CpG-sites that are unique for never smokers, our study shows that DNA methylation is also associated with FEV_1/FVC in subjects that never smoked and therefore not merely related to smoking.

INTRODUCTION

Chronic Obstructive Pulmonary Disease (COPD) is a progressive inflammatory lung disease characterized by persistent airway obstruction that causes severe respiratory symptoms and poor quality of life.[1] Although smoking is generally considered the main environmental risk factor, estimations are that 25-45% of patients with COPD have never smoked.[2] Despite extensive research, the etiology of COPD remains incompletely understood. It is known that the development of this complex heterogeneous disease is influenced by both genetic and environmental factors, as well as their interactions.[3,4,5,6] As interface between the inherited genome and environmental exposure, an important role has been postulated for the epigenome.[7] The epigenome includes multiple epigenetic mechanisms that affect gene expression without modifying the DNA sequence. These epigenetic mechanisms are highly dynamic and respond to environmental exposures, ageing and diseases.[8] One such epigenetic mechanism is DNA methylation, which involves the binding of a methyl group to a cytosine base located adjacent to a guanine base. Methylation of these so called CpG-sites in regulatory regions of the DNA generally result in decreased expression of a particular gene.[9]

So far, only a few studies have investigated the association between DNA methylation in peripheral blood and COPD or lung function using an epigenome-wide hypothesis free approach.[10,11,12,13,14,15,16] Although findings across the studies are not consistent, there is suggestive evidence that alterations in DNA methylation might play a role in the etiology of COPD. However, in previous studies, subjects were included irrespective of smoking status, thus including current smokers, ex-smokers, and never smokers. As a consequence, it is currently not known if there are differences in DNA methylation between healthy individuals and patients with COPD who have never smoked. Recently, we studied the association between epigenome-wide DNA methylation and COPD in both current smokers and never smokers.[16] Although we did not find any epigenome-wide significant association in current smokers nor in never smokers, the associations between DNA methylation and COPD were different between both groups. Hence, by further exploring the role of DNA methylation in a much larger set of never smokers together with a continuous measurement of lung function, we might be able to reveal important novel insights in the etiology of COPD. In this study, we aim to assess the association between DNA methylation and lung function in never smokers, meta-analyzing four independent population-based cohorts.

RESULTS

Subject characteristics

An overview of the characteristics of the subjects included in the study is shown in **Table 1**.

Table 1: Subject characteristics of the subjects from the four different DNA methylation datasets.

	LL COPD&C	LLDEEP	RS-III-1	RS-BIOS
Number of subjects, N (%)	903	166	150	206
Male, N (%)	508 (56.3)	71 (42.8)	74 (49.3)	80 (38.8)
Age (years), median (min-max)	46 (18-80)	42 (20-78)	63 (53-93)	68 (52-79)
Airway obstruction (FEV_1/FVC<70%), N (%)	316 (35.0)	15 (9.0)	13 (8.7)	19 (9.0)
FEV_1 (L), mean (SD)	3.5 (0.9)	3.6 (0.9)	3.2 (0.8)	2.7 (0.7)
FEV_1/FVC, mean (SD)	84.5 (8.2)	78.6 (6.2)	77.8 (5.9)	77.9 (5.9)

LL: Lifelines; RS: Rotterdam study; FEV_1: Forced expiratory volume in one second; FVC: Forced Vital Capacity; L: Liter; SD: standard deviation

LL COPD&C was the largest cohort included in this meta-analysis. Notably, since this cohort is a non-random selection from the LifeLines cohort study with COPD (defined as FEV_1/FVC < 0.70) as one of the selection criteria, the percentages of COPD cases should not be interpreted as prevalence.

Meta-analysis of the four epigenome-wide association studies

An epigenome-wide association study (EWAS) on FEV_1/FVC was performed in all four cohorts separately and combined with a meta-analysis. The meta-analysis did not reveal CpG-sites that were epigenome-wide significantly associated with FEV_1/FVC. We identified 36 CpG-sites as our top associations (**Table 2**).

The Manhattan plot of the meta-analysis is shown in **Figure 1A**.

Forest plots of the three most significant CpG-sites cg10012512, located in the intergenic region of chromosome 7q36.3 ($p=5.94x10^{-7}$), cg02285771, annotated to LTV1 Ribosome Biogenesis Factor (*LTV1)* ($p=4.10x10^{-6}$) and, cg25105536, annotated to Kelch Like Family Member 32 (*KLHL32)* ($p=9.09x10^{-6}$) are shown in **Figure 1B-D**. An overview of all CpG-sites associated with FEV_1/FVC at nominal p-value of 0.05 can be found in Supplementary Table 1. Complete summary statistics can be obtained upon request by the corresponding author.

The direction of the effect of the 36 top CpG-sites did not change in a sensitivity analysis in the LL COPD&C cohort excluding the subjects that were exposed to environmental tobacco smoke (ETS)(N=659 subjects)(**Supplementary Table 2**).

Table 2: Results of the meta-analysis and individual EWA studies on FEV_1/FCV in never smokers

		Meta-analysis			LL COPD&C			LLDEEP			RS-III-1			RS-BIOS		
CpG site	**Gene**	**Beta**	**SE**	**P-value**	**Beta**	**SE**	**P-value**	**Beta**	**SE**	**P-value**	**Beta**	**SE**	**P-value**	**Beta**	**SE**	**P-value**
cg10012512	*Intergenic*	-38.27	7.67	5.94×10^{-7}	-45.54	12.14	1.76×10^{-4}	-16.71	26.68	5.31×10^{-1}	-33.86	15.33	2.72×10^{-2}	-38.23	14.78	9.71×10^{-3}
cg02885771	*LTV1*	20.66	4.48	4.10×10^{-6}	21.53	8.76	1.40×10^{-2}	27.73	15.33	7.05×10^{-2}	21.95	6.05	2.86×10^{-4}	5.67	13.95	6.84×10^{-1}
cg25105536	*KLHL32*	-59.71	13.46	9.09×10^{-6}	-76.36	44.35	8.51×10^{-2}	-97.80	235.46	6.78×10^{-1}	-54.41	14.81	2.38×10^{-4}	-94.28	47.91	4.91×10^{-2}
cg20102034	*RTKN*	36.14	8.28	1.28×10^{-5}	42.57	15.29	5.35×10^{-3}	29.70	15.94	6.25×10^{-2}	40.85	14.65	5.29×10^{-3}	22.02	24.20	3.63×10^{-1}
cg03703840	*KIAA1731*	84.04	19.38	1.45×10^{-5}	100.48	42.84	1.90×10^{-2}	-43.70	187.80	8.16×10^{-1}	88.13	23.36	1.61×10^{-4}	33.87	62.55	5.88×10^{-1}
cg21614201	*SYNPO2*	-22.66	5.23	1.45×10^{-5}	-28.17	13.55	3.76×10^{-2}	-25.53	28.56	3.71×10^{-1}	-21.10	6.11	5.58×10^{-4}	-25.22	17.72	1.55×10^{-1}
cg07957088	*PRIC285*	35.48	8.33	2.06×10^{-5}	49.48	15.72	1.64×10^{-3}	31.33	16.68	6.03×10^{-2}	38.68	13.97	5.62×10^{-3}	-.10	24.74	9.97×10^{-1}
cg05304461	*C1orf127*	-80.31	19.00	2.37×10^{-5}	-95.35	36.04	8.16×10^{-3}	152.12	153.04	3.20×10^{-1}	-82.63	25.66	1.28×10^{-3}	-68.52	47.73	1.51×10^{-1}
cg11749902	*Intergenic*	-22.32	5.30	2.55×10^{-5}	-26.22	7.75	7.17×10^{-4}	-16.37	12.44	1.88×10^{-1}	-12.69	14.61	3.85×10^{-1}	-24.69	11.32	2.91×10^{-2}
cg02207312	*PRPF19*	75.53	18.05	2.87×10^{-5}	79.32	53.44	1.38×10^{-1}	-177.08	222.75	4.27×10^{-1}	77.18	20.22	1.35×10^{-4}	74.46	63.10	2.38×10^{-1}
cg19734370	*NPTx1*	12.65	3.04	3.19×10^{-5}	12.29	4.11	2.76×10^{-3}	12.09	6.95	8.21×10^{-2}	9.23	8.85	2.97×10^{-1}	17.64	8.07	2.88×10^{-2}
cg03077331	*FN3K*	14.19	3.45	3.99×10^{-5}	16.08	4.94	1.14×10^{-3}	9.62	8.41	2.52×10^{-1}	29.01	16.49	7.85×10^{-2}	11.51	6.31	6.84×10^{-2}
cg18387671	*ANKRD13B*	-88.73	21.86	4.92×10^{-5}	-110.71	69.61	1.12×10^{-1}	4.44	272.02	9.87×10^{-1}	-87.37	24.33	3.30×10^{-4}	-83.43	73.78	2.58×10^{-1}
cg03224276	*ZFHx3*	37.55	9.26	5.00×10^{-5}	52.17	19.25	6.73×10^{-3}	16.06	44.59	7.19×10^{-1}	28.97	11.60	1.25×10^{-2}	71.59	31.14	2.15×10^{-2}
cg02137691	*FGFR3*	28.80	7.11	5.11×10^{-5}	13.24	13.60	3.30×10^{-1}	40.83	15.87	1.01×10^{-2}	35.10	10.64	9.74×10^{-4}	16.63	25.22	5.10×10^{-1}
cg25884324	*UNC45A*	-36.97	9.16	5.45×10^{-5}	-42.03	19.42	3.05×10^{-2}	-32.96	50.06	5.10×10^{-1}	-35.47	11.31	1.71×10^{-3}	-36.84	30.86	2.32×10^{-1}
cg27158523	*PPIL4*	-49.97	12.40	5.54×10^{-5}	-62.31	22.65	5.94×10^{-3}	-241.34	161.10	1.34×10^{-1}	-37.48	14.71	1.09×10^{-2}	-83.47	40.23	3.80×10^{-2}
cg01157143	*NAV2*	-23.11	5.74	5.63×10^{-5}	-31.05	15.70	4.80×10^{-2}	-10.87	23.51	6.44×10^{-1}	-24.64	6.82	3.03×10^{-4}	-8.89	18.20	6.25×10^{-1}
cg07160694	*DCAF5*	77.84	19.34	5.69×10^{-5}	63.24	40.81	1.21×10^{-1}	54.41	155.03	7.26×10^{-1}	73.37	27.79	8.29×10^{-3}	98.91	36.83	7.24×10^{-3}
cg22127773	*KDM6B*	-48.39	12.03	5.75×10^{-5}	-58.63	19.17	2.22×10^{-3}	3.55	81.11	9.65×10^{-1}	-56.26	21.72	9.60×10^{-3}	-29.26	22.85	2.00×10^{-1}
cg20939319	*TEx15*	-14.90	3.71	5.84×10^{-5}	-17.12	8.37	4.07×10^{-2}	-26.90	17.30	1.20×10^{-1}	-13.61	4.55	2.80×10^{-3}	-13.49	12.02	2.62×10^{-1}

Table 2: Results of the meta-analysis and individual EWA studies on FEV_1/FCV in never smokers (continued)

		Meta-analysis			LL COPD&C			LLDEEP			RS-III-1			RS-BIOS		
CpG site	**Gene**	**Beta**	**SE**	**P-value**	**Beta**	**SE**	**P-value**	**Beta**	**SE**	**P-value**	**Beta**	**SE**	**P-value**	**Beta**	**SE**	**P-value**
cg02206852	*PROCA1*	23.87	5.97	6.39×10^{-5}	28.18	16.23	8.24×10^{-2}	26.98	20.97	1.98×10^{-1}	22.38	7.02	1.45×10^{-3}	27.78	24.10	2.49×10^{-1}
cg17075019	*Intergenic*	35.53	8.90	6.56×10^{-5}	49.59	13.38	2.12×10^{-4}	26.62	17.55	1.29×10^{-1}	13.65	25.97	5.99×10^{-1}	28.14	20.81	1.76×10^{-1}
cg25556432	*Intergenic*	23.02	5.78	6.75×10^{-5}	25.96	8.69	2.82×10^{-3}	21.69	13.17	9.95×10^{-2}	32.14	17.96	7.36×10^{-2}	15.46	11.29	1.71×10^{-1}
cg22742965	*TMEFF2*	-17.79	4.47	6.76×10^{-5}	-24.96	11.10	2.45×10^{-2}	0.42	20.86	9.84×10^{-1}	-17.82	5.43	1.03×10^{-3}	-14.83	13.14	2.59×10^{-1}
cg16734845	*CTDSPL2*	-33.94	8.52	6.82×10^{-5}	-54.67	21.90	1.26×10^{-2}	-38.26	26.03	1.42×10^{-1}	-31.88	10.86	3.32×10^{-3}	-15.33	24.10	5.25×10^{-1}
cg09108394	*PRKCB*	-14.93	3.76	7.11×10^{-5}	-16.43	8.33	4.84×10^{-2}	-27.78	14.95	6.31×10^{-2}	-14.34	4.92	3.55×10^{-3}	-9.74	9.71	3.16×10^{-1}
cg10034572	*Intergenic*	-20.08	5.08	7.77×10^{-5}	-19.86	13.39	1.38×10^{-1}	-56.52	27.77	4.18×10^{-2}	-19.29	5.90	1.09×10^{-3}	-12.71	17.73	4.73×10^{-1}
cg20066227	*C1QL3*	32.20	8.16	7.92×10^{-5}	26.51	18.29	1.47×10^{-1}	24.42	30.70	4.26×10^{-1}	40.00	10.35	1.12×10^{-4}	3.19	24.73	8.97×10^{-1}
cg07148038	*TN×B*	44.32	11.26	8.23×10^{-5}	51.79	16.72	1.95×10^{-3}	41.06	24.11	8.85×10^{-2}	55.29	30.47	6.96×10^{-2}	22.61	25.67	3.78×10^{-1}
cg23396786	*SF×N5*	20.16	5.12	8.26×10^{-5}	22.48	7.68	3.43×10^{-3}	13.97	10.89	2.00×10^{-1}	45.93	18.48	1.30×10^{-2}	13.79	10.08	1.71×10^{-1}
cg06218079	*TBCD*	8.18	2.08	8.34×10^{-5}	5.68	3.00	5.79×10^{-2}	12.74	3.45	2.26×10^{-4}	3.33	8.96	7.10×10^{-1}	6.35	6.52	3.30×10^{-1}
cg06982745	*ADAMTS14*	-40.80	10.44	9.37×10^{-5}	-36.77	18.57	4.77×10^{-2}	13.29	44.30	7.64×10^{-1}	-48.83	14.67	8.71×10^{-4}	-42.55	30.04	1.57×10^{-1}
cg05946118	*Intergenic*	-20.27	5.19	9.38×10^{-5}	-17.24	6.98	1.35×10^{-2}	-23.39	14.23	1.00×10^{-1}	-25.24	13.56	6.28×10^{-2}	-23.41	12.66	6.46×10^{-2}
cg08065963	*Intergenic*	-16.72	4.28	9.56×10^{-5}	-18.12	5.84	1.93×10^{-3}	-9.56	11.07	3.88×10^{-1}	-29.63	11.66	1.10×10^{-2}	-8.68	10.18	3.94×10^{-1}
cg12064372	*Intergenic*	32.85	8.43	9.75×10^{-5}	48.15	18.52	9.33×10^{-3}	26.64	92.88	7.74×10^{-1}	31.50	10.10	1.81×10^{-3}	7.96	28.48	7.80×10^{-1}

LL: Lifelines; RS: Rotterdam study; FEV_1: Forced expiratory volume in one second; FVC: Forced Vital Capacity; Beta: effect estimate; SE: standard error.

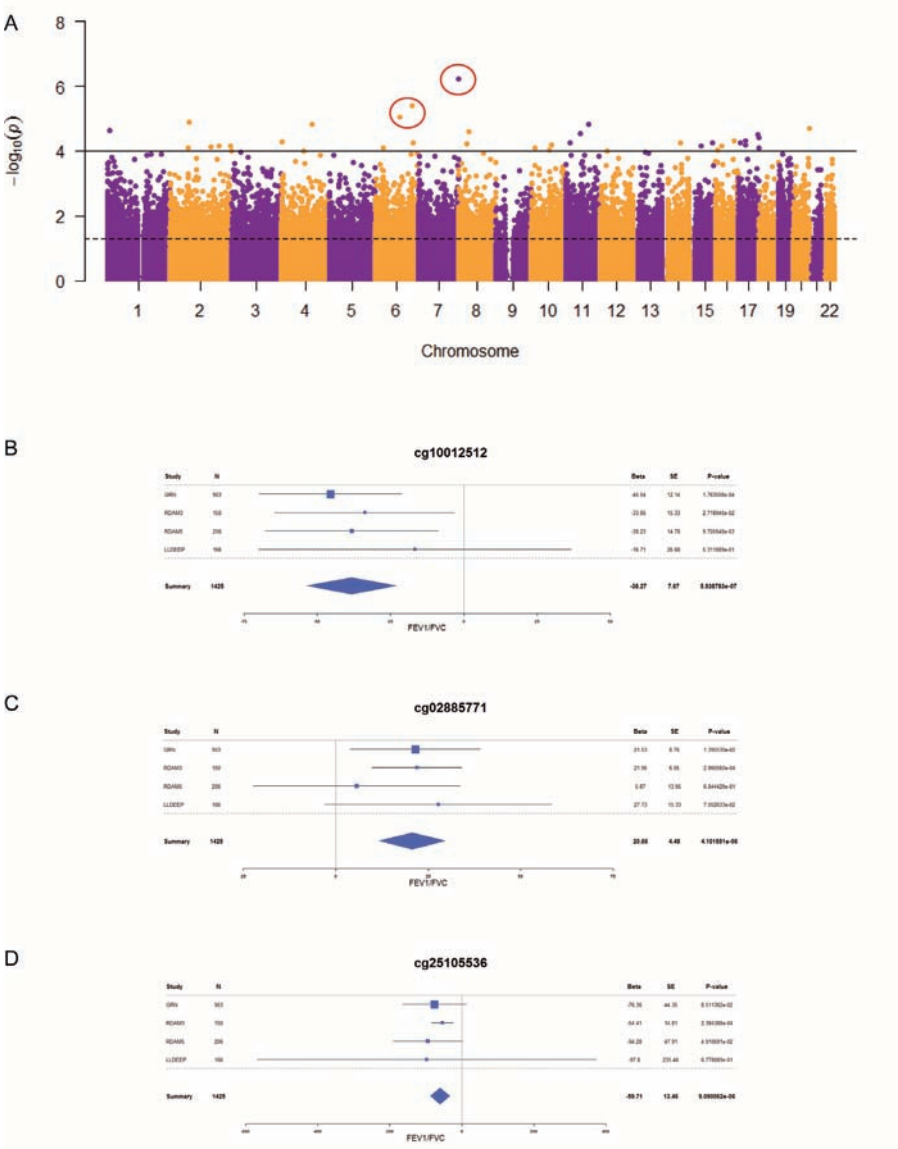

Figure 1: Manhattan and forest plots of the meta-analysis on four independent epigenome-wide association studies on FEV_1/FVC in never smokers. A) Manhattan plot in which every dot represents an individual CpG-site. Location on the X-axis indicated the chromosomal position and location on the Y-axis indicates the inversed log [10] p-value of the meta-analysis. Dotted horizontal line indicates p-value of 0.0001, horizontal fixed line indicates genome wide significance (p-value $< 0.05/396{,}243 = 1.26\times10^{-7}$). B-D) Forest plots showing the effect estimates and standard errors of the 4 independent EWA studies and meta-analysis for the top hits cg10012512 (B), cg028885771 (C) and cg25105536 (D).

Expression Quantitative trait Methylation (eQTM) analysis

To test if the top CpG-sites were associated with gene expression levels, we performed eQTM analysis. In total, 803 genes were located within 2 MB of the 36 CpG-sites. The expression of 11 genes was significantly associated with DNA methylation levels at the 9 different CpG-sites (**Table 3**).

Table 3: Overview of the results of the meta-analysis of the eQTM analysis.

CpG-site	Gene annotation CpG-site	Genes located within 1MB (N)	Gene (expression)	Beta	SE	P-value	Adjusted P-value
cg02137691	*FGFR3*	31	*SLC26A1*	0.0156	0.0038	3.53E-05	0.0011
cg02206852	*PROCA1*	52	*NUFIP2*	0.0084	0.0022	1.06E-04	0.0055
cg02206852	*PROCA1*	52	*GIT1*	0.0080	0.0023	6.11E-04	0.0318
cg02885771	*LTV1*	11	*VDAC1P8*	0.0096	0.0033	3.51E-03	0.0386
cg07148038	*TNXB*	89	*ATP6V1G2*	0.0074	0.0021	3.79E-04	0.0337
cg07148038	*TNXB*	89	*STK19B*	0.0035	0.0010	3.77E-04	0.0335
cg08065963		12	*ABAT*	0.0127	0.0034	1.85E-04	0.0022
cg20939319	*TEX15*	10	*SARAF*	-0.0029	0.0010	3.36E-03	0.0336
cg22127773	*KDM6B*	80	*TMEM88*	0.0011	0.0003	1.82E-04	0.0146
cg23396786	*SFXN5*	18	*CYP26B1*	0.0024	0.0008	1.78E-03	0.0321
cg25105536	*KLHL32*	4	*KLHL32*	-0.0004	0.0002	5.52E-03	0.0221

eQTM: Expression Quantitative Trait Methylation; Beta: effect estimate; SE: standard error.

DNA methylation at cg25105536, annotated to *KLHL32*, was significantly associated with gene expression levels of *KLHL32*. DNA methylation levels at cg08065963, located in the intergenic region on chromosome 16 and not yet annotated to a gene, showed a significant association with gene expression levels of 4-Aminobutyrate Aminotransferase *(ABAT)*. For the other 7 CpG-sites, DNA methylation levels were associated with gene expression levels of one or two genes other than the previously annotated genes. An overview of the association between DNA methylation and gene expression levels of all genes can be found in **Supplementary Table 3**.

DISCUSSION

This study is the first large general population-based EWA study on lung function in never smokers. So far, virtually all EWA studies on the origin of COPD included subjects with a history of cigarette smoking. As a consequence, these studies mainly

addressed the origins of COPD in response to smoking. It is unclear if the results of these studies help to explain the etiology of COPD or rather explain the contribution of cigarette smoke towards the disease. Therefore, our study importantly contributes to the current understanding of COPD in never smokers.

We identified 36 CpG-sites that were significantly associated with FEV_1/FVC at p-value below 0.0001. The top hit of our meta-analysis, cg10012512, is located in the intergenic region of chromosome 7q36.3. It is therefore not possible to speculate on the functional effect of differences in DNA methylation at this specific CpG-site and how these differences may affect FEV_1/FVC. While associations found with an eQTM analysis may help to get more insight in the function of a CpG-site, our eQTM analysis did not reveal any nominal significant associations for cg10012512. However, this CpG-site was differentially methylated between never smokers and current smokers.[17] Presumably, this CpG-site does also respond to other inhaled deleterious substances, which in turn affects lung function. The second top hit, cg02885771 located on chromosome 6q24.2 is annotated *LTV1*. Previously, this CpG-site has been associated with asthma in airway epithelial cells[18] and *LTV1* was shown to be expressed in lung tissue in the Genotype Tissue Expression (GTEx) project. Although studies in yeast describe *LTV1* as a conserved 40S-associated biogenesis factor that functions in small subunit nuclear export, a specific role for *LTV1* in respiratory diseases is not known.[19] The third top hit, cg25105536, is annotated to *KLHL32* on chromosome 6q16.1 and we found a significant association between DNA methylation levels of cg25105536 and gene expression levels of *KLHL32*. The function of *KLHL32* is poorly understood, however, four genetic variants in the *KLHL32* gene have been associated with FEV_1 and FEV_1/FVC in African American subjects with COPD and a history of smoking.[20] Notwithstanding the fact that these associations were only identified in a specific group, it might suggest a role for *KLHL32* in the respiratory system. Next to *KLHL32*, we found that gene expression levels of 10 additional genes were significantly associated with DNA methylation levels at one of the 36 CpG-sites. cg08065963, which was not yet annotated to a gene, was significantly associated with 4-Aminobutyrate Aminotransferase (*ABAT*). Interestingly, a role for *ABAT* in COPD has not been described before. The remaining nine genes were other genes than the annotated genes of the particular CpG-sites. This suggest that the CpG-sites may also regulate distant genes within a region of 2 MB, which complicates the functional assessment of differences in DNA methylation even further. To the best of our knowledge, there are seven studies in literature describing the association between DNA methylation and lung function (**Table 4**).

Table 4: Overview of studies reporting results of differential DNA methylation with lung function or COPD in whole blood

Study	Study population	Trait	Adjustment included in model	DNA methylation platform	Number of CpG-sites available for comparison
No association between DNA methylation and COPD in never and current smokers *De Vries* et al, 2018 [16]	Non-random selection from LifeLines cohort (N=1561 subjects) - Smoking status: Stratified for smoking (658 smokers and 903 never smokers)	- COPD (defined as $FEV_1/FVC \leq 0.7$)	Sex, Age, Pack years (in smoking stratified analysis), Batch effects, Blood cell composition	Illumina Infinium Human Methylation450 BeadChips array - Number of included probes: 420,938	Smokers: 19492 † Never smokers: 19393 †
Lung function discordance in monozygotic twins and associated differences in blood DNA methylation *Bolund* et al, 2017 [11]	Sub-population of twins from the Middle-Aged Danish Twin (MADT) study (N=169 twin pairs) - Smoking status: subjects with and without smoking history	Intra-pair difference in *z*-score calculated as "superior" minus "inferior" twin at baseline and during follow-up period for: - FEV_1 - FVC - FEV_1/FVC	Sex, Age, BMI, Pack years, Smoking status at follow-up, Blood cell composition *Intra-pair difference was calculated for all the variables	Illumina Infinium Human Methylation450 BeadChips array - Number of included probes: 453,014	37 *
Epigenome-wide association study of chronic obstructive pulmonary disease and lung function in Koreans *Lee* et al, 2017 [12]	Sample of Korean COPD cohort (N=100 subjects) - Smoking status: subjects with and without smoking history	- COPD status (defined as $FEV_1/FVC < 0.7$) - FEV_1 - FVC - FEV_1/FVC	Sex, Age, Height, Smoking status, Pack years, Blood cell composition	Illumina Infinium Human Methylation450 BeadChips array - Number of included probes: 402,508	16 *
Differential DNA methylation marks and gene comethylation of COPD in African-Americans with COPD exacerbations *Busch* et al, 2016 [13]	Sample of PA-SCOPE AA study population (N=362 subjects) - Smoking status: smokers >20 pack years	- COPD (defined as $FEV_1/FVC \leq 0.7$ and $FEV_1 \leq 80\%$)	Sex, Age, Pack years, Batch number, Blood cell composition	Illumina Infinium Human Methylation27 BeadChips array - Number of included probes: 19,302	12 *
The epigenetic clock is correlated with physical and cognitive fitness in the Lothian Birth Cohort *Marioni* et al, 2015 [15]	The Lothian Birth Cohort of 1936 (N=1,091) - Smoking status: self-reported, subjects with and without smoking history	- FEV_1	Sex, Age, Heigth, Smoking status, Blood cell composition	Illumina Infinium Human Methylation450 BeadChips array - Number of included probes: 450,726	2 *

Table 4: Overview of studies reporting results of differential DNA methylation with lung function or COPD in whole blood (continued)

Study	Study population	Trait	Adjustment included in model	DNA methylation platform	Number of CpG-sites available for comparison
Variable DNA methylation is associated with chronic obstructive pulmonary disease and lung function *Qiu* et al, 2012 [10]	Test-replication approach in 2 family-based cohorts (N=1,085 and 369 subjects) - Smoking status: subjects with and without smoking history	- COPD status (FEV_1/FVC ≤0.7 and FEV_1 ≤70%) - FEV_1/FVC - FEV_1	Random family effect	Illumina Infinium Human Methylation27 BeadChips array - Number of included probes: 26,485	349 *
Epigenome-wide scans identify differentially methylated regions for age and age-related phenotypes in a healthy ageing population *Bell* et al, 2012 [14]	Sample of the TwinsUK cohort (N=172 female twin pairs) - Smoking status: unknown	- FEV_1 - FVC	Age, Batch effects	Illumina Infinium Human Methylation27 BeadChips array - Number of included probes: 24,641	1 *

[†] CpG-sites obtained from the online available data; [$] CpG-sites selected at nominal p-value <0.05 available from self-performed analyses; COPD: Chronic Obstructive Pulmonary Disease; FEV_1: Forced Expiratory Volume in 1 sec; FVC: Forced Expiratory Capacity.

Six of these studies included both subjects with and without a history of cigarette smoking and, except for the study by *Qui* et al, adjusted for smoking status in the statistical analysis. Altogether, these studies identified 406 unique CpG-sites. Interestingly, none of the 36 CpG-sites from our meta-analysis in never smokers were among these 406 previously identified CpG-sites (**Table 5**). Apparently these 36 CpG-sites are only associated with lung function level in never smokers. The fact that 17 CpG-sites (47%) were associated at nominal p-value <0.05 with COPD (dichotomously defined as the ratio of FEV_1/FVC below 70%) in our previously EWAS stratified for never smoking, further underscores this assumption.[16] There is, however, one exception, since cg22742965, annotated to Transmembrane Protein With EGF Like And Two Follistatin Like Domains 2 (*TMEFF2)*, was also significantly associated with COPD in smokers. Most likely, this CpG-site shows a general response to inhaled deleterious substances such as cigarette smoke and other yet unknown substances.

Table 5: Overview of CpG location, gene annotation, gene function and literature comparison of the top 36 CpG-sites of the meta-analysis

CpG-site	CpG location	Gene annotation	Gene function	Previously associated with lung function
cg10012512	7:157224041	*Intergenic*	*NA*	Yes [1]
cg02885771	6:144163654	LTV1	Involved in ribosome biogenesis	No
cg25105536	6:97372436	KLHL32	Only described as protein coding gene	No
cg20102034	2:74653166	RTKN	Negative regulator of GTPase activity of Rho proteins	Yes [1]
cg03703840	11:93394809	KIAA1731	Mediating of centriole-to-centrosome conversion at late mitosis	No
cg21614201	4:119888794	SYNPO2	Only described as protein coding gene	No
cg07957088	20:62196387	PRIC285	Nuclear transcriptional co-activator for peroxisome proliferator activated receptor alpha	Yes [1]
cg05304461	1:11019377	C1orf127	Only described as protein coding gene	No
cg11749902	8:41093619	*Intergenic*	*NA*	Yes [1]
cg02207312	11:60674164	PRPF19	Involved in cell survival and DNA repair	No
cg19734370	17:78444348	NPTX1	Exclusively localized to the nervous system as binding protein for taipoxin	Yes [1]
cg03077331	17:80693076	FN3K	Catalyzes the phosphorylation of fructosamines	Yes [1]
cg18387671	17:27920246	ANKRD13B	Only described as protein coding gene	Yes [1]
cg03224276	16:72829831	ZFHX3	Regulates myogenic and neuronal differentiation	No
cg02137691	4:1805671	FGFR3	Involved in bone development and maintenance	No
cg25884324	15:91482502	UNC45A	Regulator of the progesterone receptor chaperoning pathway	No
cg27158523	6:149867355	PPIL4	Involved in protein folding, immunosuppression and infection of HIV-1 virions	Yes [1]
cg01157143	11:19478542	NAV2	Plays a role in cellular growth and migration	No
cg07160694	14:69619856	DCAF5	Only described as protein coding gene	No
cg22127773	17:7754785	KDM6B	Demethylation of di- or tri-methylated lysine 27 of histone H3	Yes [1]
cg20939319	8:30707701	TEX15	Involved in cell cycle processes of spermatocytes	No
cg02206852	17:27030540	PROCA1	Only described as protein coding gene	No
cg17075019	10:79541650	*Intergenic*	*NA*	Yes [1]
cg25556432	2:239628926	*Intergenic*	*NA*	Yes [1]

Table 5: Overview of CpG location, gene annotation, gene function and literature comparison of the top 36 CpG-sites of the meta-analysis (continued)

CpG-site	CpG location	Gene annotation	Gene function	Previously associated with lung function
cg22742965	2:192891657	TMEFF2	Cellular context-dependent oncogene or tumor suppressor	Yes
cg16734845	15:44781962	CTDSPL2	Only described as protein coding gene	No
cg09108394	16:23850106	PRKCB	As kinase involved in diverse cellular signaling pathways	No
cg10034572	2:160921789	*Intergenic*	*NA*	No
cg20066227	10:16564552	C1QL3	Only described as protein coding gene	No
cg07148038	6:32061160	TNXB	Anti-adhesive protein involved in matrix maturation during wound healing	Yes [1]
cg23396786	2:73299151	SFXN5	Only described as protein coding gene	Yes [1]
cg06218079	17:80834228	TBCD	As co-factor D involved in the correct folding of beta-tubulin	No
cg06982745	10:72454006	ADAMTS14	The matured enzyme is involved in the formation of collagen fibers	No
cg05946118	16:8985638	*Intergenic*	*NA*	Yes [1]
cg08065963	16:8985593	*Intergenic*	*NA*	Yes [1]
cg12064372	12:30948792	*Intergenic*	*NA*	Yes [1]

[1] Only observed in study by *de Vries* et al in never smokers; Gene function obtained by www.genecards.org

Assuming that the observed differential DNA methylation at the majority of the CpG-sites in our study occurs without exposure to smoking, the question arises why this differential DNA methylation is observed. One possible explanation may be that other factors within the environment such as air pollution and job-related exposures are responsible for the observed differences in DNA methylation. Recently, we studied the epigenome-wide association between DNA methylation and exposure to air pollution and job-related exposures in a selection of the LifeLines population cohort including both never and current smokers.[21,22] While we did find significant associations, none of them were replicated in independent cohorts. Additional analyses in never smokers for this paper did not reveal novel associations between DNA methylation and environmental exposures (Online supplement Table 4 and Online supplement Figure 1). This might potentially be due to lack of power, since only a small percentage of the subjects that have never smoked in the LL COPD&C cohort have been exposed to environmental exposures. Moreover, exposure levels

to air pollution in the LL COPD&C are relatively low compared to the average Dutch levels determined within the 2012 Dutch national health survey as described by Strak *et al.*[23] Next to environmental exposures, another explanation may be that a reduced lung function level precedes the differences in DNA methylation. However, with the cross-sectional design of this study, we cannot derive conclusions on the direction of the association and causality. Large longitudinal studies are required to investigate causality between DNA methylation and FEV_1/FVC.

In conclusion, with this study we show that epigenetics indeed may be associated with FEV_1/FVC in subjects who never smoked. Moreover, since 35 out of the 36 identified CpG-sites are unique for never smokers, our data suggest that factors other than smoking affect FEV_1/FVC via DNA methylation.

METHODS

Study population

To study the association between epigenome-wide DNA methylation and lung function, defined as the ratio between the Forced Expiratory Volume in one second (FEV_1) and Forced Vital Capacity (FVC), in never smokers, we performed a meta-analysis in four different cohorts. Two cohorts originated from the LifeLines population-based cohort study[24]: the LifeLines COPD & Controls DNA methylation study[16,22] (LL COPD&C, n=903) and the LifeLines DEEP study[25] (LLDEEP, n=166). The two other cohorts originated from the population-based Rotterdam study (RS)[26] : The first visit of the third RS cohort (RS-III-1, n=150) and a cohort selected for the Biobank-based Integrative Omics Studies (BIOS) project (RS-BIOS, n=206). Both population-based cohort studies were approved by the local university medical hospital ethical committees and all participants signed written informed consent. In all cohorts, never smoking was defined based on self-reported never-smoking history and zero pack years included in the standardized questionnaires.

Measurements

Lung function

Within the LifeLines population-based cohort study, pre-bronchodilator spirometry was performed with a Welch Allyn Version 1.6.0.489, PC-based Spiroperfect with CA Workstation software according to ATS/ERS guidelines. Technical quality and results were evaluated by well-trained assistants and difficult to interpret results were re-evaluated by a lung physician. Within the population-based Rotterdam

study, pre-bronchodilator spirometry was performed during the research center visit using a SpiroPro portable spirometer (RS-III-1) or a Master Screen® PFT Pro (RS-BIOS) by trained paramedical staff according to the ERS/ATS Guidelines. Spirometry results were analyzed by two researchers and verified by a specialist in pulmonary medicine.

DNA methylation

In all four cohorts, DNA methylation levels in whole blood were determined with the Illumina Infinium Methylation 450K array. Data was presented as beta values (ratio of methylated probe intensity and the overall intensity) ranging from 0 to 1. Quality control has been performed for all datasets separately as described before.[22,27] After quality control, data was available on 396,243 CpG-sites in all four datasets.

Statistical analysis

Epigenome-wide association study and meta-analysis

We performed EWAS on lung function defined as FEV_1/FVC in all four cohorts separately using robust linear regression analysis in R. The analysis was adjusted for the potential confounders: age and sex. To adjust for the cellular heterogeneity of the whole blood samples, we included proportional white blood cell counts of mononuclear cells, lymphocytes, neutrophils, and eosinophils, obtained by standard laboratory techniques. For LL COPD&C, we adjusted for technical variation by performing a principal components analysis using the 220 control probes incorporated in the Illumina 450k Chip. The 7 principal components that explained >1% of the technical variation were included in the analysis. For LLDEEP, data on technical variance was not accessible. For the two RS cohorts, we included the position on the array and array number to adjust for technical variation. Regression estimates from all four individual EWA studies were combined by a random-effect meta-analysis using the effect estimates and standard errors in "rmeta" package in R. CpG-sites with a p-value below 1.26×10^{-7} (Bonferroni corrected p-value by number of CpG-sites 0.05/396,243) were considered epigenome-wide significant. CpG-sites with a p-value below 0.0001 in the meta-analysis were defined as top associations in our study.

Expression Quantitative Trait Methylation (eQTM) analysis

To assess whether top associations were also associated with gene expression levels, we used the never smokers included in the Biobank-based Integrative Omics Studies (BIOS). For all cohorts separately, reads were normalized to counts per

million. To adjust for technical variation for gene expression and DNA methylation, principal component analysis was conducted on the residual normalized counts and beta-values excluding the potential confounders age and gender. Principal components that explained more than 5% of the technical variation in gene expression or DNA methylation were included in the analysis. Subsequently, robust linear regression analysis was performed on the CpG-sites and the genes within 1 MB around the CpG-sites. The analyses were adjusted for the potential confounders: age, sex, and technical variation by principal components as stated before. The individual eQTM analysis were combined by a random-effect meta-analysis using the effect estimates and standard errors in rmeta. An eQTM was considered significant when the Bonferroni-adjusted p-value for the number of genes within 1 MB around the CpG-sites was below 0.05.

REFERENCES

1. From the *global strategy for the diagnosis, management and prevention of COPD*, global initiative for chronic obstructive lung disease (GOLD) 2015. available from: Http://www.goldcopd.org/. .
2. Salvi S.S. and Barnes P.J. (2009) Chronic obstructive pulmonary disease in non-smokers. *Lancet,* **374**, 733-743.
3. van der Plaat D.A., de Jong K., Lahousse L., Faiz A., Vonk J.M., van Diemen C.C., Nedeljkovic I., Amin N., Brusselle G.G., Hofman A. *et al.* (2016) Genome-wide association study on the FEV_1/FVC ratio in never-smokers identifies HHIP and FAM13A. *J. Allergy Clin. Immunol.,* .
4. de Jong K., Boezen H.M., Kromhout H., Vermeulen R., Postma D.S. and Vonk J.M. (2014) Association of occupational pesticide exposure with accelerated longitudinal decline in lung function. *Am. J. Epidemiol.,* **179**, 1323-1330.
5. de Jong K., Vonk J.M., Timens W., Bosse Y., Sin D.D., Hao K., Kromhout H., Vermeulen R., Postma D.S. and Boezen H.M. (2015) Genome-wide interaction study of gene-by-occupational exposure and effects on FEV_1 levels. *J. Allergy Clin. Immunol.,* **136**, 1664-72.e1-14.
6. Hobbs B.D., de Jong K., Lamontagne M., Bosse Y., Shrine N., Artigas M.S., Wain L.V., Hall I.P., Jackson V.E., Wyss A.B. *et al.* (2017) Genetic loci associated with chronic obstructive pulmonary disease overlap with loci for lung function and pulmonary fibrosis. *Nat. Genet.,* **49**, 426-432.
7. Jirtle R.L. and Skinner M.K. (2007) Environmental epigenomics and disease susceptibility. *Nat. Rev. Genet.,* **8**, 253-262.
8. Yang I.V., Lozupone C.A. and Schwartz D.A. (2017) The environment, epigenome, and asthma. *J. Allergy Clin. Immunol.,* **140**, 14-23.
9. Jones P.A. (2012) Functions of DNA methylation: Islands, start sites, gene bodies and beyond. *Nat. Rev. Genet.,* **13**, 484-492.
10. Qiu W., Baccarelli A., Carey V.J., Boutaoui N., Bacherman H., Klanderman B., Rennard S., Agusti A., Anderson W., Lomas D.A. *et al.* (2012) Variable DNA methylation is associated with chronic obstructive pulmonary disease and lung function. *Am. J. Respir. Crit. Care Med.,* **185**, 373-381.
11. Bolund A.C.S., Starnawska A., Miller M.R., Schlunssen V., Backer V., Borglum A.D., Christensen K., Tan Q., Christiansen L. and Sigsgaard T. (2017) Lung function discordance in monozygotic twins and associated differences in blood DNA methylation. *Clin. Epigenetics,* **9**, 132-017-0427-2. eCollection 2017.
12. Lee M.K., Hong Y., Kim S.Y., Kim W.J. and London S.J. (2017) Epigenome-wide association study of chronic obstructive pulmonary disease and lung function in koreans. *Epigenomics,* **9**, 971-984.
13. Busch R., Qiu W., Lasky-Su J., Morrow J., Criner G. and DeMeo D. (2016) Differential DNA methylation marks and gene comethylation of COPD in african-americans with COPD exacerbations. *Respir. Res.,* **17**, 143-016-0459-8.
14. Bell J.T., Tsai P.C., Yang T.P., Pidsley R., Nisbet J., Glass D., Mangino M., Zhai G., Zhang F., Valdes A. *et al.* (2012) Epigenome-wide scans identify differentially methylated regions for age and age-related phenotypes in a healthy ageing population. *PLoS Genet.,* **8**, e1002629.
15. Marioni R.E., Shah S., McRae A.F., Ritchie S.J., Muniz-Terrera G., Harris S.E., Gibson J., Redmond P., Cox S.R., Pattie A. *et al.* (2015) The epigenetic clock is correlated with physical and cognitive fitness in the lothian birth cohort 1936. *Int. J. Epidemiol.,* **44**, 1388-1396.

16. de Vries M., van der Plaat D.A., Vonk J.M. and Boezen H.M. (2018) No association between DNA methylation and COPD in never and current smokers. *BMJ Open Respir. Res.,* **5**, e000282-2018-000282. eCollection 2018.
17. Joehanes R., Just A.C., Marioni R.E., Pilling L.C., Reynolds L.M., Mandaviya P.R., Guan W., Xu T., Elks C.E., Aslibekyan S. *et al.* (2016) Epigenetic signatures of cigarette smoking. *Circ. Cardiovasc. Genet.,* **9**, 436-447.
18. Nicodemus-Johnson J., Myers R.A., Sakabe N.J., Sobreira D.R., Hogarth D.K., Naureckas E.T., Sperling A.I., Solway J., White S.R., Nobrega M.A. *et al.* (2016) DNA methylation in lung cells is associated with asthma endotypes and genetic risk. *JCI Insight,* **1**, e90151.
19. Merwin J.R., Bogar L.B., Poggi S.B., Fitch R.M., Johnson A.W. and Lycan D.E. (2014) Genetic analysis of the ribosome biogenesis factor Ltv1 of saccharomyces cerevisiae. *Genetics,* **198**, 1071-1085.
20. Lutz S.M., Cho M.H., Young K., Hersh C.P., Castaldi P.J., McDonald M.L., Regan E., Mattheisen M., DeMeo D.L., Parker M. *et al.* (2015) A genome-wide association study identifies risk loci for spirometric measures among smokers of european and african ancestry. *BMC Genet.,* **16**, 138-015-0299-4.
21. de F C Lichtenfels A.J., van der Plaat D.A., de Jong K., van Diemen C.C., Postma D.S., Nedeljkovic I., van Duijn C.M., Amin N., la Bastide-van Gemert S., de Vries M. *et al.* (2018) Long-term air pollution exposure, genome-wide DNA methylation and lung function in the LifeLines cohort study. *Environ. Health Perspect.,* **126**, 027004.
22. van der Plaat D.A., de Jong K., de Vries M., van Diemen C.C., Nedeljkovic I., Amin N., Kromhout H., Biobank-based Integrative Omics Study Consortium, Vermeulen R., Postma D.S. *et al.* (2018) Occupational exposure to pesticides is associated with differential DNA methylation. *Occup. Environ. Med.,* **75**, 427-435.
23. Strak M., Janssen N., Beelen R., Schmitz O., Vaartjes I., Karssenberg D., van den Brink C., Bots M.L., Dijst M., Brunekreef B. *et al.* (2017) Long-term exposure to particulate matter, NO2 and the oxidative potential of particulates and diabetes prevalence in a large national health survey. *Environ. Int.,* **108**, 228-236.
24. Scholtens S., Smidt N., Swertz M.A., Bakker S.J., Dotinga A., Vonk J.M., van Dijk F., van Zon S.K., Wijmenga C., Wolffenbuttel B.H. *et al.* (2015) Cohort profile: LifeLines, a three-generation cohort study and biobank. *Int. J. Epidemiol.,* **44**, 1172-1180.
25. Tigchelaar E.F., Zhernakova A., Dekens J.A., Hermes G., Baranska A., Mujagic Z., Swertz M.A., Munoz A.M., Deelen P., Cenit M.C. *et al.* (2015) Cohort profile: LifeLines DEEP, a prospective, general population cohort study in the northern netherlands: Study design and baseline characteristics. *BMJ Open,* **5**, e006772-2014-006772.
26. Ikram M.A., Brusselle G.G.O., Murad S.D., van Duijn C.M., Franco O.H., Goedegebure A., Klaver C.C.W., Nijsten T.E.C., Peeters R.P., Stricker B.H. *et al.* (2017) The rotterdam study: 2018 update on objectives, design and main results. *Eur. J. Epidemiol.,* **32**, 807-850.
27. Ligthart S., Steenaard R.V., Peters M.J., van Meurs J.B., Sijbrands E.J., Uitterlinden A.G., Bonder M.J., BIOS consortium, Hofman A., Franco O.H. *et al.* (2016) Tobacco smoking is associated with DNA methylation of diabetes susceptibility genes. *Diabetologia,* **59**, 998-1006.

Chapter 3

Omics studies of COPD and comorbidities: asthma and depression

Chapter 3.1

New-born DNA-methylation, childhood lung function, and the risks of asthma and COPD across the life course

Herman T. den Dekker*, Kimberley Burrows*, Janine F. Felix*, Lucas A. Salas, Ivana Nedeljković, Jin Yao, Sheryl L. Rifas-Shiman, Carlos Ruiz-Arenas, N. Amin, Mariona Bustamante, Dawn L. DeMeo, A. John Henderson, Caitlin G. Howe, Marie-France Hivert, M. Arfan Ikram, Johan C. de Jongste, Lies Lahousse, Pooja R. Mandaviya, Joyce B. van Meurs, Mariona Pinart, Gemma C. Sharp, Lisette Stolk, André G. Uitterlinden, Josep M. Anto, Augusto A. Litonjua, Carrie V. Breton, Guy G. Brusselle, Jordi Sunyer, George Davey Smith, Caroline L. Relton#, Vincent W.V. Jaddoe#, Liesbeth Duijts#

**first authors contributed equally*

#last authors contributed equally

This chapter is published in European Respiratory Journal, February 2019.

The supplemental material for this paper is available at:

https://doi.org/10.1183/13993003.01795-2018

ABSTRACT

We aimed to identify differentially methylated regions (DMRs) in cord blood DNA associated with childhood lung function, asthma and chronic obstructive pulmonary disease (COPD) across the life course.

We meta-analysed epigenome-wide data of 1688 children from five cohorts to identify cord blood DMRs and their annotated genes, in relation to Forced Expiratory Volume in 1 second (FEV_1), FEV_1/Forced Vital Capacity (FVC), and Forced Expiratory Flow at 75% of FVC (FEF_{75}) at ages 7 to 13 years. Identified DMRs were explored for associations with childhood asthma, adult lung function and COPD, gene expression and involvement in biological processes. We identified 59 DMRs associated with childhood lung function, of which 18 were associated with childhood asthma and 9 with COPD in adulthood. Genes annotated to the top ten identified DMRs were HOXA5, PAOX, LINC00602, ABCA7, PER3, CLCA1, VENTX, NUDT12, PTPRN2 and TCL1A. Differential gene expression in blood was observed for 32 DMRs in childhood and 18 in adulthood. Genes related with 16 identified DMRs were associated with respiratory developmental or pathogenic pathways.

Our findings suggest that the epigenetic status of the new-born affects respiratory health and disease across the life course.

INTRODUCTION

Asthma and chronic obstructive pulmonary disease (COPD) have become major global health problems in the last decades.[1] Both diseases are characterized by airway obstruction, as indicated by a reduced Forced Expiratory Volume in 1 second (FEV_1), FEV_1 to Forced Vital Capacity (FVC) ratio, and Forced Expiratory Flow at 75% of FVC (FEF_{75}).[2] Childhood lung function predicts lung function and risks of asthma and COPD in later life.[3] An accumulating body of evidence suggests that asthma and COPD have at least part of their origins in fetal life.[4, 5] Genetics alone fail to explain the quickly altering prevalence of allergies and chronic respiratory diseases in the past decades, because any mutation would require multiple generations to occur on a population level.[6] Furthermore, adverse fetal exposures, such as maternal smoking and suboptimal diet, increase the risk of asthma and COPD.[5] The pathways linking genetic predisposition and environmental exposures in fetal life with life course respiratory disease may include epigenetic changes, including DNA-methylation.[5] Epigenetic changes are influenced by environmental exposures and could exert population effects much more rapidly than genetic mutations.[6] DNA-methylation is currently the best understood epigenetic mechanism, and techniques have been developed to assess epigenome-wide DNA-methylation patterns in large population- based studies. Fetal development is characterized by high rates of DNA-methylation changes and rapid organ development.[5] DNA-methylation may affect fetal development through effects on gene transcription and expression.[7] Recent studies assessing the associations between DNA-methylation and childhood respiratory health are mainly limited to candidate genes and small sample sizes.[3, 8] We focused on identification of differential DNA-methylated regions (DMRs) because regional methylation of CpGs controls cell-type-specific transcription. Also, the use of DMRs increases statistical power and minimizes the effects of genetic variants in the methylation analyses.[9] Identification of genomic regions with altered DNA-methylation levels related to lung function and respiratory diseases across the life course is important to understand mechanisms underlying associations of environmental and genetic factors with the development of lower lung function and risk of chronic respiratory diseases. We hypothesized that fetal differential DNA-methylation reflected in cord blood DNA of newborns affect gene expression and subsequent respiratory tract development, and predispose individuals for obstructive airway diseases in later life.[10, 11]

We meta-analyzed five epigenome-wide association studies using data from 1,688 children participating in prospective cohort studies to identify differential DNA-methylated regions (DMRs) of newborns associated with childhood FEV_1, FEV_1/FVC

and FEF_{75}. Identified top DMRs were subsequently explored for their associations with childhood asthma, lung function in adolescence and adulthood, and COPD in adulthood, and explored for association with gene expression and involvement in biological processes.

METHODS

Study design and data sources

We included population-based cohort studies participating in PACE consortium with data on epigenome-wide DNA-methylation at birth and lung function in childhood.[12] We used data from 1,688 Caucasian children aged 7 to 13 years participating in the Avon Longitudinal Study of Parents and Children (ALSPAC, United Kingdom), Generation R (Netherlands), INfancia y Medio Ambiente Study (INMA) (Spain), Children's Health Study (CHS) and Project Viva (both from the U.S.A.). These data were used for the primary discovery epigenome-wide meta-analysis to identify DMRs of newborns related to childhood lung function. We aimed to identify DMRs instead of single CpGs while differences at any individual CpG may be small, and the use of DMRS might minimize the effects of genetic variants in the methylation analysis.[13, 14]

We used several resources for the secondary analyses. For clinical outcomes, we used childhood asthma data (Generation R, mean age 6 years), lung function data from adolescents (ALSPAC, mean age 15 years) and adults (Rotterdam Study, mean age 66 years, The Netherlands), and COPD data in adults (Rotterdam Study) (**Figure 1**). For gene expression, we used blood samples from children (INMA, at birth and mean age 4 years) and adults (Rotterdam Study). Last, we used publicly available resources to relate identified DMRs with biological processes.[15-17] Parents, legal representatives or participants provided informed consent in accordance with local ethics policies. Detailed information about the study design and cohorts is provided in the **Supplementary Appendix**.

DNA-methylation

All cohorts extracted DNA from blood samples and used the EZ-96 DNA Methylation kit (Zymo Research Corporation, Irvine, USA) for bisulfite conversion. Samples were processed with the Infinium HumanMethylation450 BeadChip (Illumina Inc., San Diego, USA) followed by cohort-specific quality control, probe exclusion and data normalization. Detailed information on cohort-specific data acquisition and quality control is provided in the **Supplementary Appendix**.

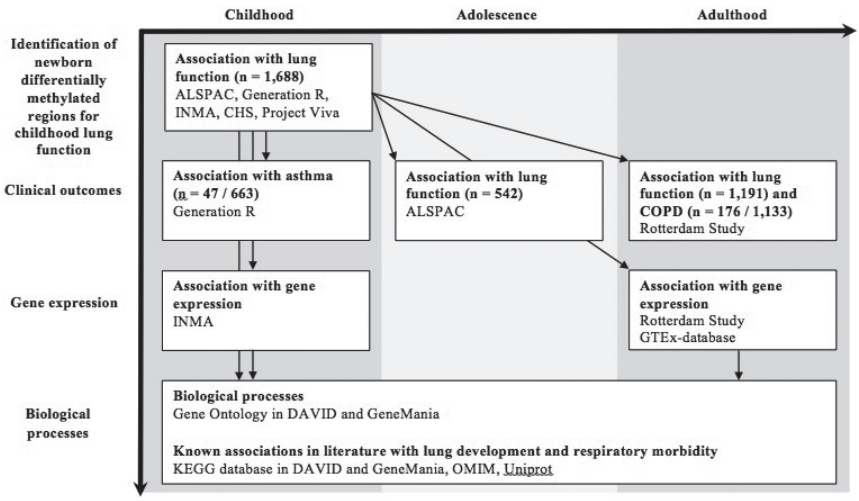

Figure 1. Overall Study Design. Epigenome-wide meta-analyses were performed to identify methylated CpGs associated with lung function in children using data from 1,689 children participating in ALSPAC, Generation R, INMA, CHS and Project Viva. Identified differentially methylated regions (DMRs) were annotated to their nearest gene using PAVIS. Next, we examined if identified DMRs were associated with asthma in children participating in Generation R, lung function in adolescents and adults participating in the ALSPAC or Rotterdam Study, or COPD in adults participating in Rotterdam Study, and with gene expression levels in children participating in INMA, adults in Rotterdam Study, and the GTEx-database. We further explored biological processes and associations with lung development and respiratory morbidity using publicly available resources (DAVID, GeneMania, OMIM and UniProt). N = x: number of participants included for the analysis. N = x/x: number of cases / total number of participants included in the analysis.

Respiratory outcome assessment

Lung function measures comprised pre-bronchodilator FEV_1, FEV_1/FVC and FEF_{75}, which were converted into sex-, age-, height- and ethnicity-adjusted z-scores.[18] Physician-diagnosed asthma was obtained by questions adapted from the International Study on Asthma and Allergy in Childhood.[19] COPD was defined as pre-bronchodilator FEV_1/FVC <0.70 in the absence of asthma, or a doctor diagnosis.[20]

Statistical analyses

Primary meta-analysis on childhood lung function

A detailed description of applied methods is presented in the **Supplementary Appendix**. Individual cohorts used robust linear regression models to examine the associations of DNA-methylation levels of CpGs with childhood FEV_1, FEV_1/FVC and

FEF_{75}. Analyses were adjusted for maternal age, socio-economic status, smoking during pregnancy, parity, asthma or atopy, technical covariates and estimated cell counts.[21] Results were combined using inverse variance-weighted fixed-effect meta-analyses. Results from unadjusted models were similar to fully adjusted models (**Supplementary Appendix Table 1**). Using p-values obtained from the meta-analyses, we identified DMRs using the software-tool Comb-p, which is the most robust tool to identify DMRs with small effect sizes.[9, 22] Regions were defined as a minimum of 2 probes within a window size of 500 bases with an FDR-threshold <0.05.[9] Comb-p uses unadjusted p-values for each probe as input, and calculates adjusted p-values for each probe that account for the correlation with nearby CpGs.[23] Next, the SLK p-values were adjusted for multiple testing and adjusted into q-values. Comb-p finds DMRs based on these q-values and calculates p-values for these DMRs based on the original p-values. Finally, the DMR p-values were adjusted for multiple testing using the Šidák-correction based on the size of the region and number of possible regions of that size. A sliding window identifies a DMR without any predefined regional borders, and therefore (theoretically) does not have a maximum number of windows and DMRs. A more extensive description of the identification of DMRS is provided in the Supplementary Material. Annotation of the genes located nearest to the DMRs was performed using Peak Annotation and Visualization (PAVIS).[24] We limited annotation to a region of 500kb (250kb upstream, 250kb downstream of the beginning and end of the region, respectively). All annotations were based on human GRCh37/hg19 assembly. Because genetic variants in Infinium probes could result in spurious methylation measurements, we performed a sensitivity analysis in a subset of high-quality probes (n=294,834) without SNPs, insertions or deletions, repeats, polymorphic probes and bisulfite induced reduced genomic complexity.[25]

Secondary analyses on later life lung function and respiratory diseases

We used linear and logistic regression models to examine the associations of CpGs within identified DMRs with asthma in childhood, FEV_1, FEV_1/FVC and FEF_{75} in adolescence and adulthood, and COPD in adulthood. Single CpG p-values were used to reconstruct the identified DMRs with Comb-p, applying identical parameter settings as in the discovery meta-analyses including false discovery rate (FDR)-correction.[9, 26] We did not apply Šidák-correction because analyses were limited to the identified DMRs.

Gene expression analyses

We assessed the associations of CpGs within identified DMRs with gene expression in a region of ±250kb in blood samples from children and adults. P-values of CpGs

associated with gene expression were combined for each DMR using a modified generalized Fisher method and FDR-correction.[26, 27] Additionally, we explored whether the annotated and differentially expressed genes were expressed in human lung specimens of the Genotype-Tissue Expression (GTEx) database.[15]

Exploration biological processes

The Gene Ontology database implemented in DAVID and Genemania was used to examine gene function in biological processes.[16, 17] We examined pathways for all genes annotated to the DMRs and for genes with differential expression in association with the identified DMRs. We used the Kyoto Encyclopedia of Genes and Genomes (KEGG) database in DAVID and Genemania, the OMIM database and the Universal Protein Resource (UniProt) to explore whether annotated or expressed genes have been related to respiratory development or diseases.[28] We used the Ensemble Genome browser to visualize the genomic structure of the identified DMRs.

RESULTS

Meta-analysis of epigenome-wide association studies on childhood lung function

Characteristics of the participating cohorts are given in **Table 1** and **Supplementary Appendix Table 2**.

We identified 22, 15 and 22 DMRs associated with FEV_1, FEV_1/FVC and FEF_{75}, respectively (**Figure 2**, **Supplementary Appendix Tables 3** and **4**).

A higher mean methylation of CpGs located within 37 (63%) of the identified DMRs was associated with higher lung function measures, and within 22 (37%) of identified DMRs with lower lung function measures. We observed a high homogeneity across the included studies (CpGs with I^2 <50: FEV_1 140/163 (86%), FEV_1/FVC 82/89 (92%) and FEF_{75} 139/148 (87%)) (**Supplemental Tables 4a-c**).

Of the top ten significant DMRs associated with childhood lung function, the 5 DMRs and their annotated genes for FEV_1 were located at chr7:27,183,133-27,184,854 (*HOXA5),* chr10:135,202,522-135,203,201 (*PAOX),* chr6:166,418,799-166,419,139 *(LINC00602)*, chr19:1,063,624-1,064,219 (*ABCA7)* and chr1:7,887,199-7,887,561 (*PER3*). Three DMRs and their annotated genes for FEV_1/FVC were located at chr1:86,968,087-86,968,544 *(CLCA1)*, chr10:135,051,149-135,051,582 (*VENTX),* and chr5:102,898,223-102,898,734 (*NUDT12).* Two DMRs for FEF_{75} and their annotated genes were located at chr7:158,045,980-158,046,359 *(PTPRN2)* and chr14:96,180,406-96,181,045 (*TCL1A)*. After exclusion of potentially problematic probes containing genomic variants, 41

Table 1. Characteristics of Cohorts and Their Participants.

	No. of participants	Type of blood sample for DNA-methylation	No. of available CpGs	No. of subjects with expression data	Age at lung function measurement	Asthma		COPD	
					Years (SD)	Cases	Controls	Cases	Controls
Primary analyses									
ALSPAC (UK)	654	Cord blood	471,193	NA	8·6 (0·2)	NA	NA	NA	NA
Generation R (NL)	643	Cord blood	436,013	NA	9·8 (0·3)	47	663	NA	NA
INMA (Spain)	140	Cord blood	439,306	107	6·9 (0·3)	NA	NA	NA	NA
CHS (USA)	75	Cord blood	383,857	NA	13·3 (0·6)	NA	NA	NA	NA
Project Viva (USA)	176	Cord blood	470,870	NA	7·9 (0·7)	NA	NA	NA	NA
Secondary analyses									
ALSPAC (UK)	542	Cord blood	NA	NA	15·4 (0·2)	NA	NA	NA	NA
Rotterdam Study – I (NL)	488	Peripheral blood	NA	488	64·0 (6·3)	NA	NA	63	425
Rotterdam Study – II (NL)	703	Peripheral blood	NA	703	67·5 (5·9)	NA	NA	92	611

Lung function was obtained by spirometry and sex-, age-, height- and ethnicity-adjusted Z-scores were calculated. FEV_1: Forced Expiratory Volume in 1 second; FVC: Forced Vital Capacity; FEF75: Forced Expiratory Flow at 75% of FVC; NA: not applicable. UK: United Kingdom. NL: the Netherlands. USA: United States of America.

of the 59 previously identified DMRs still contained ≥2 CpGs (**supplementary appendix Table 4**). Of these 41 DMRs, 54% (n=22) remained to be associated with childhood lung function (**Supplementary appendix Table 5**).

Identified DMRs and lung function and respiratory diseases across the life course

Of all 59 identified DMRs related with childhood lung function, 18 (31%) were associated with childhood asthma (**Figure 3**, **Supplementary appendix Table 6**).

Furthermore, 11 (19%) and 9 (15%) DMRs were associated with lung function in adolescence and adulthood, respectively, and 9 (15%) were associated with COPD. The DMRs annotated to *HOXA5* and *PAOX* were associated with childhood and adolescence FEV_1 and COPD, but not with childhood asthma or adult lung function. The DMRs annotated to *PER3* and *VENTX* were associated with childhood and adolescence FEV_1 and FEV_1/FVC, respectively. The DMR annotated to *NUDT12* was associated with childhood FEV_1/FVC and COPD. The DMRs annotated to *PTPRN2* and *TCL1A* were associated with childhood FEF_{75} and asthma. The DMRs annotated to *LINC00602*, *ABCA7* and *CLCA1* were associated with childhood lung function but not with other outcomes.

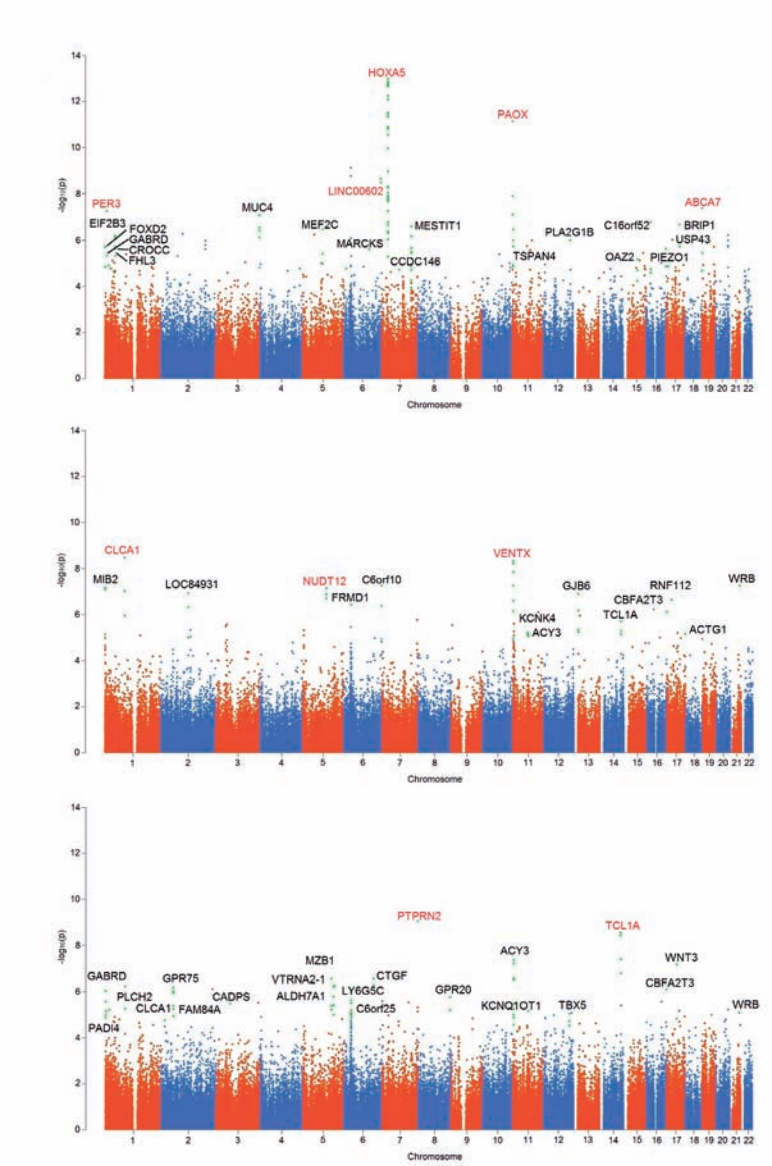

Figure 2. Manhattan Plots of Associations of CpGs located in Differentially Methylated Regions with Childhood Lung Function Outcomes. Green dots represent p-values from associations of CpGs located in differentially methylated regions (DMRs) at birth with childhood (A) Forced Expiratory Volume in 1 second (FEV_1), (B) FEV_1/ Forced Vital Capacity (FVC) and (C) Forced Expiratory Flow at 75% of FVC (FEF_{75}). P-values of DMRs ranged from 3·05E-14 to 0·031, and details are provided in Supplementary Appendix Table 3. Nearest annotated genes of DMRs are provided. The genes annotated to the top ten significant DMRs associated with childhood lung function are written in red. Single CpGs are presented as red and blue dots, corrected for correlations with neigboring CpGs.

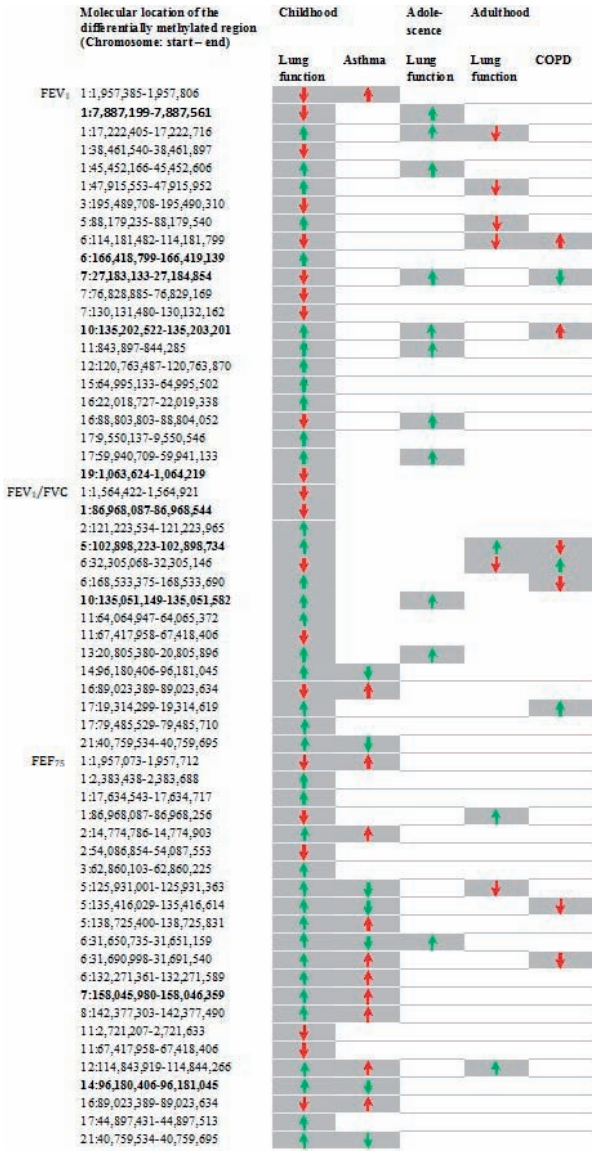

Figure 3. Identified Differentially Methylated Regions and Their Location, and Direction of Associations with Childhood Lung Function, Childhood Asthma, Adolescent Lung Function, and Adult Lung Function and COPD. Results present identified differentially methylated regions (DMRs) from association-analyses of DNA-methylation at birth with childhood Forced Expiratory Volume in 1 second (FEV_1), FEV_1/ Forced Vital Capacity (FVC) and Forced Expiratory Flow at 75% of FVC (FEF_{75}), their location, and their direction of association with childhood lung function, childhood asthma, adolescent lung function, and adult lung function and COPD. Molecular locations of the top ten significant DMRs are presented in bold. Identified DMRs associated with childhood lung function and other respiratory outcomes are marked in grey, and if not associated with respiratory outcomes in white. Directions of associations are marked ↓ if a higher mean methylation of the DMRs was associated with a lower z-score for lung function or lower risk of asthma or COPD and marked ↑ if a higher mean methylation of the DMRs was associated with a higher z-score of lung function or higher risk of asthma or COPD. Red colored arrows represent disadvantageous effect estimates (lower lung function, increased risk of asthma or COPD), and green colored arrows beneficial effect estimates (higher lung function, lower risk of asthma or COPD).

Identified DMRs and Gene Expression

Of the 59 identified DMRs, 32 (54%) DMRs at birth were associated with gene expression at age 4 years, and 18 (31%) DMRs with gene expression in adulthood (**Supplementary Appendix Table 7**). The DMR annotated to *HOXA5* was associated with differential expression of several genes of the *HOX*-family (**Table 2**). The DMRs annotated to *PER3*, *VENTX*, *NUDT12* and *TCL1A* were associated with differential expression of their respective genes. The DMRs annotated to *PAOX*, *LINC00602*, *ABCA7*, *CLCA1* and *PTPRN2* were not associated with expression of their corresponding genes. Genes annotated to 28 (47%) of all identified DMRs were expressed in adult lung tissue, including the top significant DMRs annotated to *PAOX, ABCA7, CLCA1*, *VENTX* and *NUDT12* (**Supplementary Appendix Table 8**).

Identified DMRs and related biological processes

Of all 59 identified DMRs, 43 were annotated to genes not previously associated with lung function or respiratory morbidity (**Supplementary Appendix Table 7)**. Of the genes annotated to the top ten significant DMRs, *HOXA5*, *CLCA1*, *TCL1A* and *NUDT12* were previously associated with respiratory development including alveogenesis, respiratory diseases and cellular immunity (**Table 2**). Genes related to the identified DMRs, including *HOXA5*, *PER3*, *CLCA1*, *NUDT12* and *PTPRN2*, were located in pathways related to regionalization, DNA- and RNA-regulation and embryonic development (**Supplementary Appendix Tables 9**). The genes *HLA-DRB4* and *HLA-DRB5* were enriched in processes including asthma. These genes were associated with the DMR located at chr6:32,305,068-32,305,146, which was related with childhood and adulthood FEV_1/FVC and COPD.

Of the top ten significant DMRs, the DMRs annotated to *HOXA5*, *CLCA1* and *TCL1A* contained *CTCF*-binding sites (**Supplementary Figure 2**). The DMRs annotated to *HOXA5*, *PAOX*, *PER3* and *NUDT12* were located in promotor regions of their respective genes. The DMR annotated to *ABCA7* was located in a CpG-island.

DISCUSSION

We identified 59 DMRs in neonatal cord blood associated with childhood lung function. Eighteen (31%) of all identified DMRs were also associated with childhood asthma, 11 (19%) and 9 (15%) with adolescent and adult lung function, respectively, and 9 (15%) with COPD. Differential gene expression was observed for 32 (54%) DMRs in childhood and 18 (31%) DMRs in adulthood. Multiple genes related to the identified DMRs have previously been associated with respiratory development

Table 2. Associations of the top ten significant identified DMRs with gene expression and related respiratory outcomes.

Molecular location of the DMR (Chromosome: start – end)	Lung function	Annotated gene*	Expressed gene†	Gene expression in children‡	Gene expression in adults§	Previously associated with lung development or respiratory morbidity‖
chr1: 7,887,199 - 7,887,561	FEV_1	*PER3*	*PER3, RP3-467I1.4, RNA5SP23, RP4-726F1.1*	↓	-	
			RP11-431K24.1	-	↑	
chr1: 86,968,087 – 86,968,544	FEV_1/FVC	*CLCA1*	no expression	-	-	Lung development, asthma, COPD
chr5: 102,898,223 - 102,898,734	FEV_1/FVC	*NUDT12*	*NUDT12*	↓	-	Smoking behavior in COPD
			CMBL	↓	-	
chr6: 166,418,799 – 166,419,139	FEV_1	*LINC00602*	no expression	-	-	
chr7: 27,183,133 - 27,184,854	FEV_1	*HOXA5*	*HOXA1, HOTTIP*	↓	↓	Lung development, FEV_1, FEV_1/FVC
			EVX1, HOXA4, HOXA7	↓	-	Lung development, asthma, COPD
chr7: 158,045,980 – 158,046,359	FEF_{75}	*PTPRN2*	no expression	-	-	
chr10: 135,202,522 – 135,203,201	FEV_1	*PAOX*	no expression	-	-	
chr10: 135,051,149 – 135,051,582	FEV_1/FVC	*VENTX*	*TUBGCP2, RP11-122K13.12*	↓	-	
			VENTX, ECHS1	↑	-	
			SPRN	↑	↓	
			ZNF511	-	↑	
chr14: 96,180,406 – 96,181,045	FEF_{75}	*TCL1A*	*TCL1A, CCDC85C*	↓	-	Asthma
chr19: 1,063,624 – 1,064,219	FEV_1	*ABCA7*	no expression	-	-	

Results present identified differentially methylated regions (DMRs) from association-analyses of DNA-methylation at birth with childhood Forced Expiratory Volume in 1 second (FEV_1), FEV_1/ Forced Vital Capacity (FVC) and Forced Expiratory Flow at 75% of FVC (FEF_{75}). *DMRs were annotated to their nearest gene. †: Identified DMRs at birth were associated with gene expression in: ‡: childhood (INMA; mean age 4 years) and §: adulthood (the Rotterdam Study, mean age 66 years). Gene expressions levels were assessed limited to 250kb up- and downstream of the outer border of the DMR. Directions of associations are marked ↓ if a higher methylation of the DMR was associated with a decreased expression of the specific gene, ↑ if a higher methylation of the DMR was associated with an increased expression of the specific gene, and - if no direction of associations were observed. ‖: Associations of expressed genes with lung development and respiratory morbidity were explored in previous published studies the OMIM database and UniProt.

and morbidity, and many identified DMRs were located within known regulatory elements for gene expression.

Reduced lung function in childhood is associated with reduced lung function and increased risks of asthma and COPD many decades later.[10, 29] Pathways of environmental exposures in early life, such as tobacco smoke exposure or lack of breastfeeding, that affect lung development and risk of chronic obstructive respiratory diseases in later life might be modified by genetic susceptibility. Vice versa, genetic susceptibility could partly explain the difference in adverse effects of early environmental exposures on the risk of chronic obstructive respiratory diseases in later life. Identified genetic variants associated with childhood asthma in large-scale GWA studies only account for up to 7.5% of the explained variance.[30] Epigenetic mechanisms could link environmental exposures with the unexplained heritability for childhood asthma.[31, 32] Studies that examined associations of DNA-methylation with lung function, asthma or COPD are scarce, limited to candidate genes or high-risk population and lack replication. An epigenome-wide study among 97 asthmatics and 97 healthy children aged 6-12 years identified 81 DMRs associated with asthma, of which 16 DMRs were also associated with FEV_1.[8] Of these 81 DMRs, 19 were located within 500kb of our identified DMRs and may affect the same genes. Another epigenome-wide study in 1,454 adults identified 349 CpGs associated with COPD.[33] Four annotated genes in this adult study (*CBFA2T3*, *PADI4*, *LST1*, *KCNQ1*) were replicated in our study of children. Multiple genes associated with the identified DMRs have previously been related with asthma and COPD in genome-wide association studies. *TCL1A* has been identified as asthma-susceptibility gene.[34] Nine (15%) of the 59 DMRs we identified were associated with adult lung function, and annotated to, or associated with differential expression of 11 genes. Nine of these genes were previously linked with pulmonary structures (*CROCC*, *CLCA1*), immunity (*MARCKS*, *FOXD2*, *MEF2C*, *CMBL*, *CLCA1*), asthma (*MARCKS*, *HCG23*, *CLCA1*), COPD (*MARCKS*, *TBX5*, *CLCA1*), and smoking behavior in COPD (*NUDT12*).[28] This suggests that genes associated with respiratory diseases could be influenced by differential DNA-methylation from early life onwards.

We explored the biological processes of the top significant DMRs for development of respiratory morbidity.[28] The DMR annotated to *HOXA5* was associated with childhood and adolescent FEV_1, COPD and differential expression of *HOXA1*, *HOXA4* and *HOXA7*. One DMR associated with childhood FEV_1/FVC was annotated to *VENTX*, which is a member of the *HOX*-gene family. The DMR annotated to *LINC00602* (Long Intergenic Non-Protein Coding RNA (lncRNA) 602) was linked to childhood FEV_1. LncRNAs influence gene-specific epigenetic regulation and interact amongst others with the transcription of *HOX*-genes. *HOX*-genes are critical for segmental fetal

development, and especially *HOXA5* is required for embryonic respiratory tract morphogenesis.[35,36] The DMR annotated to *PAOX* was linked to childhood and adolescence FEV_1 and COPD. *PAOX* is involved in the regulation of intracellular polyamine, which is essential for protein synthesis. The DMR linked to *ABCA7* was associated with FEV_1 in childhood and adolescence. *ABCA7* is involved in the lipid homeostasis in the cellular immune system and is essential for phagocytosis of apoptotic cells by alveolar macrophages.[37] *PER3*, annotated to a DMR associated with FEV_1 in children and adolescents, is a key element in the endogenous circadian rhythm. The DMR linked to *CLCA1* was associated with childhood FEV_1/FVC and FEF_{75} and expressed in adult lung tissue. *CLCA1* affects IL-13 driven mucus production in human airway epithelial cells and is associated with asthma and COPD.[38-40] *NUDT12*, annotated to a DMR associated with childhood FEV_1/FVC and COPD, is involved in intracellular biochemical reactions. *NUDT12* is associated with smoking behavior in COPD.[41]

PTPRN2, annotated to a DMR associated with childhood FEF_{75} and asthma is member of a gene family regulating cell growth and differentiation, and is involved in vesicle-mediated secretory processes. DNA-methylation of *PTPRN2* differentiates between lung cancer, pulmonary fibrosis and COPD.[42] *TCL1A*, annotated to a DMR associated with childhood FEV_1/FVC, FEF_{75} and asthma, is specific to developing lymphocytes when expressed and is associated with asthma.[34] Thus, many of the genes annotated to the top significant DMRs are involved in respiratory development, cellular immunity and respiratory morbidity, which warrant further studies.

This is the largest study to date evaluating the associations of newborn epigenome-wide DNA-methylation with lung function and respiratory disease in children and adults, and it provides new insights into the epigenetic changes in fetal life that increase the risk of life-time respiratory morbidity. To the best of our knowledge, no other cohort studies with data on cord blood DNA-methylation and childhood lung function are available. We aimed to strengthen our results using public databases on gene expression and biological pathways, which added additional support for the observed associations. Ideally, the presence of identified DMRs would be replicated in lung cells. However, in cohort studies, this is ethically not done. It is unknown whether nasal cells, which are easier to acquire, have a high enough correlation in DNA-methylation with lung tissue. Therefore, further studies should aim to examine whether DNA-methylation in nasal cells is a good proxy for lung tissue and data on DNA-methylation and phenotypes should be shared in consortia to increase the statistical power to identify DNA-methylation patterns affecting respiratory health across the life course. These results cannot currently be used as predictors of disease in individuals but are important from an etiological perspective. Genes associated with 29 of the identified DMRs, including *HOXA5, PAOX, VENTX, PTPRN2*

and *TCL1A*, have been reported to be differentially methylated in relation with maternal smoking during pregnancy.[12] Genes related with four identified DMRs associated with childhood lung function were differentially methylated in association with maternal folate levels during pregnancy.[43] This supports the hypothesis that adverse exposures in fetal life may impact DNA-methylation at birth, gene expression and subsequent respiratory development in the child, predisposing individuals for obstructive airway diseases. Further experimental or Mendelian randomization studies on the identified DMRs and associated genes might inform strategies in early life to improve lung function and lower the lifetime risk of obstructive respiratory diseases.

Some limitations should be discussed. We measured DNA-methylation in blood because it is easily accessible in large cohort studies. Blood DNA-methylation does not necessarily reflect lung epithelial DNA-methylation. However, asthma and COPD have systemic manifestations, characterized by increased inflammatory blood markers.[44, 45] Although the analyses were adjusted for estimated cell counts, we cannot rule out residual confounding due to alterations in cell type distribution. Recently, two new reference sets for cell type adjustment in cord blood samples were published.[46, 47] These reference sets are currently being validated, and future studies will shed light on the differences between reference panels. In our secondary analyses, we assessed whether the identified DMRs were associated with lung function measured in adolescence and adulthood, similar to the associations identified between cord blood DNA methylation and childhood lung function. DNA methylation patterns and expression of genes vary depending on the developmental stage, and these changes could be non-linear.[48] We were not able to assess the stability of DNA methylation in the identified DMRs in the same individuals from birth to adulthood. In a recent study addressing DNA methylation changes in early life, significantly reduced or increased methylation of single CpGs between ages 0 to 4 years and 4 to 8 years occurred in <4% of all CpGs, suggesting only a minor global change in DNA methylation in childhood.[49] Longitudinal changes in DNA methylation from early life until adulthood in relation to respiratory morbidity have not been studied yet. We observed similar associations between DNA methylation of a specific genetic region with lung function, asthma or COPD observed in early life and adulthood, and this strengthens our hypothesis that specific DNA methylation patterns affect respiratory health across the life course. Further studies in longitudinal cohort studies with repeated measures of DNA methylation from birth into adulthood in the same individuals are needed to confirm this.

We presented our primary results including all probes, and provided results of analyses excluding potentially problematic probes, namely those with SNPs, inser-

tions or deletions, repeats and bisulfite induced reduced genomic complexity. These underlying variants may affect probe binding and as such, affect the identified associations. In our stringent sensitivity analyses, we observed similar size and direction of the effect estimates in 54% of the identified DMRs associated with childhood lung function. However, the true exact impact of potentially problematic probes on the measurement of DNA methylation in our analyses remains unknown.[25,50] Discarding probes a priori may discard information. Therefore, we present all results of the main and sensitivity analysis.

Genetic variation as opposed to environmental variation might be influencing the DMRs associated with respiratory health. A recent study in two ethnic diverse adult cohorts in 557 subjects showed that DNA methylation of airway epithelium plays a central role in mediating the effects of SNPs and gene expression on asthma risk and its clinical course.[51] Another study in 115 subjects participating in an adult cohort study reported a potential mediating effect of DNA methylation of single CpGs on the associations between SNPs located at chromosomal locus 17q21 and asthma.[52] The study identified 6 CpGs associated with gene expression of ORMDL3 and GSDMB. The authors did not assess the associations of DNA methylation with asthma or lung function. We did not identify any DMR located near the 17q21 locus. This could be explained by the young age of our study subjects or differences in main respiratory outcomes measurements. The previous studies stepwise assessed the effect of DNA methylation with the gene expression, and associations of SNPs with asthma, whereas our study focused on the direct associations between DNA methylation and respiratory outcomes. Further research is needed to assess this potential biological pathway.

Several identified DMRs were associated with gene expression other than the nearest and therefore annotated gene, which limits the potential biological effect of the annotated genes. The genomic inflation factor for the primary analyses ranged from 1.07 to 1.21 (**Supplementary Figure 1**). Recently, it was shown that the genomic inflation factor provides an invalid estimate of test-statistic inflation when the outcome of interest is associated with many, small genetic effects.[53] Furthermore, estimating the inflation factor using the genomic inflation factor results both in an overestimation of the actual inflation and in imprecise estimates contributing to the previously unexplained, high variability across studies. This might explain the genomic inflation in our analyses. The statistical steps in Comb-P limit the final number of DMRs identified, and genomic inflation in the identification of DMRs could not be tested. Further studies are needed to develop statistical tools dealing with genomic inflation in epigenome-wide studies.

There were no cohort studies available for replication analyses. We included all available cohorts in the meta-analysis to obtain the largest possible power to detect new associations. A previously published study has shown that in (epi)genome-wide association analyses a meta-analysis of all participating cohorts rather than a split sample analysis with a properly selected level of (epi)genome-wide significance is the most powerful approach to identify new associations.[54] The high between-study homogeneity observed for the vast majority of CpGs in our meta-analysis (**Supplementary Tables 4a-c**) also provides support for the stability of the reported associations. Nevertheless, confirmatory studies are needed.

In conclusion, we identified 59 DMRs in cord blood that were associated with childhood lung function. Multiple DMRs were additionally related with childhood asthma, adolescent and adult lung function, or adult COPD. Also, multiple DMRs were associated with differential gene expression of genes involved in embryonic and respiratory tract development or were located in regulatory elements for gene expression. These findings suggest that epigenetic changes during fetal life might modify the risk of respiratory diseases across the life course.

REFERENCES

1. Chronic obstructive pulmonary disease (COPD): World Health Organisation; 2015 [Available from: http://www.who.int/mediacentre/factsheets/fs315/en/.
2. McDonough JE, Yuan R, Suzuki M, Seyednejad N, Elliott WM, Sanchez PG, et al. Small-airway obstruction and emphysema in chronic obstructive pulmonary disease. N Engl J Med. 2011; 365(17):1567-75.
3. McGeachie MJ, Yates KP, Zhou X, Guo F, Sternberg AL, Van Natta ML, et al. Patterns of Growth and Decline in Lung Function in Persistent Childhood Asthma. New Engl J Med. 2016;374(19): 1842-52.
4. Martinez FD. Early-Life Origins of Chronic Obstructive Pulmonary Disease. N Engl J Med. 2016;375(9):871-8.
5. Duijts L, Reiss IK, Brusselle G, de Jongste JC. Early origins of chronic obstructive lung diseases across the life course. Eur J Epidemiol. 2014;29(12):871-85.
6. Begin P, Nadeau KC. Epigenetic regulation of asthma and allergic disease. Allergy Asthma Clin Immunol. 2014;10(1):27.
7. Krauss-Etschmann S, Meyer KF, Dehmel S, Hylkema MN. Inter- and transgenerational epigenetic inheritance: evidence in asthma and COPD? Clin Epigenetics. 2015;7:53.
8. Yang IV, Pedersen BS, Liu A, O'Connor GT, Teach SJ, Kattan M, et al. DNA methylation and childhood asthma in the inner city. J Allergy Clin Immunol. 2015;136(1):69-80.
9. Pedersen BS, Schwartz DA, Yang IV, Kechris KJ. Comb-p: software for combining, analyzing, grouping and correcting spatially correlated P-values. Bioinformatics. 2012;28(22):2986-8.
10. Postma DS, Rabe KF. The Asthma-COPD Overlap Syndrome. N Engl J Med. 2015;373(13):1241-9.
11. Fu JJ, McDonald VM, Baines KJ, Gibson PG. Airway IL-1 beta and Systemic Inflammation as Predictors of Future Exacerbation Risk in Asthma and COPD. Chest. 2015;148(3):618-29.
12. Joubert BR, Felix JF, Yousefi P, Bakulski KM, Just AC, Breton C, et al. DNA Methylation in Newborns and Maternal Smoking in Pregnancy: Genome-wide Consortium Meta-analysis. Am J Hum Genet. 2016;98(4):680-96.
13. Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, Tonti-Filippini J, et al. Human DNA methylomes at base resolution show widespread epigenomic differences. Nature. 2009; 462(7271):315-22.
14. Bock C. Analysing and interpreting DNA methylation data. Nat Rev Genet. 2012;13(10):705-19.
15. Carithers LJ, Ardlie K, Barcus M, Branton PA, Britton A, Buia SA, et al. A Novel Approach to High-Quality Postmortem Tissue Procurement: The GTEx Project. Biopreserv Biobank. 2015; 13(5):311-9.
16. Huang da W, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic acids research. 2009;37(1):1-13.
17. Warde-Farley D, Donaldson SL, Comes O, Zuberi K, Badrawi R, Chao P, et al. The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. Nucleic acids research. 2010;38(Web Server issue):W214-20.
18. Quanjer PH, Stanojevic S, Cole TJ, Baur X, Hall GL, Culver BH, et al. Multi-ethnic reference values for spirometry for the 3-95-yr age range: the global lung function 2012 equations. Eur Respir J. 2012;40(6):1324-43.
19. Asher MI, Keil U, Anderson HR, Beasley R, Crane J, Martinez F, et al. International Study of Asthma and Allergies in Childhood (ISAAC): rationale and methods. Eur Respir J. 1995;8(3): 483-91.

20. Terzikhan N, Verhamme KM, Hofman A, Stricker BH, Brusselle GG, Lahousse L. Prevalence and incidence of COPD in smokers and non-smokers: the Rotterdam Study. Eur J Epidemiol. 2016.
21. Houseman EA, Accomando WP, Koestler DC, Christensen BC, Marsit CJ, Nelson HH, et al. DNA methylation arrays as surrogate measures of cell mixture distribution. BMC bioinformatics. 2012;13:86.
22. Peters TJ, Buckley MJ, Statham AL, Pidsley R, Samaras K, Lord RV, et al. De novo identification of differentially methylated regions in the human genome. Epigenet Chromatin. 2015;8.
23. Eckhardt F, Lewin J, Cortese R, Rakyan VK, Attwood J, Burger M, et al. DNA methylation profiling of human chromosomes 6, 20 and 22. Nat Genet. 2006;38(12):1378-85.
24. Huang W, Loganantharaj R, Schroeder B, Fargo D, Li L. PAVIS: a tool for Peak Annotation and Visualization. Bioinformatics. 2013;29(23):3097-9.
25. Naeem H, Wong NC, Chatterton Z, Hong MK, Pedersen JS, Corcoran NM, et al. Reducing the risk of false discovery enabling identification of biologically significant genome-wide methylation status using the HumanMethylation450 array. BMC Genomics. 2014;15:51.
26. Benjamini Y, Hochberg Y. Controlling for False Discovery Rate: a Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society, Series B. 1995;57:289-300.
27. Dai H, Leeder JS, Cui Y. A modified generalized Fisher method for combining probabilities from dependent tests. Front Genet. 2014;5:32.
28. UniProt C. UniProt: a hub for protein information. Nucleic acids research. 2015;43(Database issue):D204-12.
29. McGeachie MJ, Yates KP, Zhou X, Guo F, Sternberg AL, Van Natta ML, et al. Patterns of Growth and Decline in Lung Function in Persistent Childhood Asthma. N Engl J Med. 2016;374(19): 1842-52.
30. Soler Artigas M, Loth DW, Wain LV, Gharib SA, Obeidat M, Tang W, et al. Genome-wide association and large-scale follow up identifies 16 new loci influencing lung function. Nat Genet. 2011;43(11):1082-90.
31. Cookson W, Moffatt M, Strachan DP. Genetic risks and childhood-onset asthma. J Allergy Clin Immunol. 2011;128(2):266-70; quiz 71-2.
32. English S, Pen I, Shea N, Uller T. The information value of non-genetic inheritance in plants and animals. PLoS One. 2015;10(1):e0116996.
33. 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.
34. George BJ, Reif DM, Gallagher JE, Williams-DeVane CR, Heidenfelder BL, Hudgens EE, et al. Data-driven asthma endotypes defined from blood biomarker and gene expression data. PLoS One. 2015;10(2):e0117445.
35. Golpon HA, Geraci MW, Moore MD, Miller HL, Miller GJ, Tuder RM, et al. HOX genes in human lung: altered expression in primary pulmonary hypertension and emphysema. Am J Pathol. 2001;158(3):955-66.
36. Mandeville I, Aubin J, LeBlanc M, Lalancette-Hebert M, Janelle MF, Tremblay GM, et al. Impact of the loss of Hoxa5 function on lung alveogenesis. American Journal of Pathology. 2006; 169(4):1312-27.

37. Jehle AW, Gardai SJ, Li S, Linsel-Nitschke P, Morimoto K, Janssen WJ, et al. ATP-binding cassette transporter A7 enhances phagocytosis of apoptotic cells and associated ERK signaling in macrophages. J Cell Biol. 2006;174(4):547-56.
38. Poole A, Urbanek C, Eng C, Schageman J, Jacobson S, O'Connor BP, et al. Dissecting childhood asthma with nasal transcriptomics distinguishes subphenotypes of disease. J Allergy Clin Immunol. 2014;133(3):670-8 e12.
39. Hegab AE, Sakamoto T, Uchida Y, Nomura A, Ishii Y, Morishima Y, et al. CLCA1 gene polymorphisms in chronic obstructive pulmonary disease. J Med Genet. 2004;41(3):e27.
40. Alevy YG, Patel AC, Romero AG, Patel DA, Tucker J, Roswit WT, et al. IL-13-induced airway mucus production is attenuated by MAPK13 inhibition. J Clin Invest. 2012;122(12):4555-68.
41. Siedlinski M, Cho MH, Bakke P, Gulsvik A, Lomas DA, Anderson W, et al. Genome-wide association study of smoking behaviours in patients with COPD. Thorax. 2011;66(10):894-902.
42. Wielscher M, Vierlinger K, Kegler U, Ziesche R, Gsur A, Weinhausel A. Diagnostic Performance of Plasma DNA Methylation Profiles in Lung Cancer, Pulmonary Fibrosis and COPD. EBioMedicine. 2015;2(8):929-36.
43. Joubert BR, den Dekker HT, Felix JF, Bohlin J, Ligthart S, Beckett E, et al. Maternal plasma folate impacts differential DNA methylation in an epigenome-wide meta-analysis of newborns. Nat Commun. 2016;7:10577.
44. Barnes PJ. Inflammatory mechanisms in patients with chronic obstructive pulmonary disease. J Allergy Clin Immunol. 2016;138(1):16-27.
45. Nadif R, Siroux V, Boudier A, le Moual N, Just J, Gormand F, et al. Blood granulocyte patterns as predictors of asthma phenotypes in adults from the EGEA study. Eur Respir J. 2016.
46. Bakulski KM, Feinberg JI, Andrews SV, Yang J, Brown S, S LM, et al. DNA methylation of cord blood cell types: Applications for mixed cell birth studies. Epigenetics. 2016;11(5):354-62.
47. Gervin K, Page CM, Aass HC, Jansen MA, Fjeldstad HE, Andreassen BK, et al. Cell type specific DNA methylation in cord blood: a 450K-reference data set and cell count-based validation of estimated cell type composition. Epigenetics. 2016:0.
48. Feng L, Wang J, Cao B, Zhang Y, Wu B, Di X, et al. Gene expression profiling in human lung development: an abundant resource for lung adenocarcinoma prognosis. PLoS One. 2014; 9(8):e105639.
49. Xu CJ, Bonder MJ, Soderhall C, Bustamante M, Baiz N, Gehring U, et al. The emerging landscape of dynamic DNA methylation in early childhood. BMC Genomics. 2017;18(1):25.
50. Lehne B, Drong AW, Loh M, Zhang WH, Scott WR, Tan ST, et al. A coherent approach for analysis of the Illumina HumanMethylation450 BeadChip improves data quality and performance in epigenome-wide association studies (vol 16, 37, 2015). Genome Biol. 2016;17.
51. Kothari PH, Qiu W, Croteau-Chonka DC, Martinez FD, Liu AH, Lemanske RF, Jr., et al. Role of local CpG DNA methylation in mediating the 17q21 asthma susceptibility gasdermin B (GSDMB)/ORMDL sphingolipid biosynthesis regulator 3 (ORMDL3) expression quantitative trait locus. J Allergy Clin Immunol. 2018;141(6):2282-6 e6.
52. Nicodemus-Johnson J, Myers RA, Sakabe NJ, Sobreira DR, Hogarth DK, Naureckas ET, et al. DNA methylation in lung cells is associated with asthma endotypes and genetic risk. JCI Insight. 2016;1(20):e90151.
53. van Iterson M, van Zwet EW, Consortium B, Heijmans BT. Controlling bias and inflation in epigenome- and transcriptome-wide association studies using the empirical null distribution. Genome Biol. 2017;18(1):19.

54. Skol AD, Scott LJ, Abecasis GR, Boehnke M. Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. Nat Genet. 2006;38(2):209-13.

Chapter 3.2

A cross-omics integrative study of metabolic signatures of Chronic obstructive pulmonary disease

Ivana Nedeljković, Lies Lahousse, Maaike de Vries, Jun Liu, Marita Kalaoja, Emmi Tikkanen, Peter Würtz, Judith M. Vonk, Diana A. van der Plaat, Cleo C. van Diemen, Ashley van der Spek, Alexandra Zhernakova, Jingyuan Fu, Mohsen Ghanbari, Mika Ala-Korpela, Johannes Kettunen, Aki S. Havulinna, Markus Perola, Veikko Salomaa, Lars Lind, Johan Ärnlöv, Bruno H.C. Stricker, Guy G. Brusselle, H. Marike Boezen, Cornelia M. van Duijn*, Najaf Amin*.

**Both authors contributed equally.*

This chapter is submitted for publication.

The supplemental material for this paper is available at:
https://drive.google.com/file/d/1O9mbkPsUclOuOdR7IqxWxPqS3rzf2U8I/view?usp=sharing

ABSTRACT

Importance: Chronic obstructive pulmonary disease (COPD) is a disorder characterized by persistent and progressive airflow limitation. Beyond lung function impairments, metabolic changes in the circulation have been reported but their relation to the risk factors and prognosis of COPD has not been addressed. *Objective:* To identify metabolic signatures for COPD. *Design:* A comprehensive metabolic study of COPD and lung function was conducted in two large population-based studies in the Netherlands, the Rotterdam Study and the Erasmus Rucphen Family study. Significant findings were replicated in Lifelines-DEEP study, FINRISK and Prospective Investigation of the Vasculature in Uppsala Seniors (PIVUS) studies. The data were integrated with publicly available data sets. *Setting:* Multicenter, population-based setting. *Participants:* A random sample of 5,557 individuals was included in the discovery cohort, whose lung function was characterized by spirometry. *Exposure:* Circulating levels of metabolites as measured by proton Nuclear Magnetic Resonance Spectroscopy. *Main outcomes and measures:* The primary outcome was COPD, defined as the ratio of Forced Expiratory Volume in 1 second (FEV_1) to Forced Vital Capacity (FVC) <0.7. Secondary outcomes were FEV_1/FVC (continuous variable), smoking status and pack-years of smoking. Tertiary outcome was all cause mortality. *Results:* There were 602 cases of COPD and 4955 controls used in the discovery meta-analysis. Our logistic regression results showed that higher levels of plasma Glycoprotein acetyls (GlycA) were significantly associated with COPD (OR=1.16, P=5.6×10^{-4} in the discovery and OR=1.30, P =1.8×10^{-6} in the replication sample). Smoking status (P=1.3×10^{-22}) and pack-years of smoking (P= 2.5×10^{-16}) were significantly associated with levels of GlycA. A bi-directional two-sample Mendelian randomization analysis has suggested that circulating blood GlycA is not causally related to COPD, but that COPD is causally associated with GlycA. Using the prospective data of the same sample of Rotterdam Study in Cox-regression, we show that circulating GlycA levels are predictive biomarker of COPD risk (HR=1.42, 95%CI 1.24-1.63, P=7.61×10^{-7}, comparing those in the highest and lowest quartile of GlycA) but are not significantly associated with mortality in patients (HR=1.06, SE=0.06, P=0.31). *Conclusions and Relevance:* Our study shows that circulating blood GlycA is a biomarker of preclinical COPD pathology.

INTRODUCTION

Chronic obstructive pulmonary disease (COPD) is a progressive inflammatory lung disease and currently the third leading cause of death worldwide.[1,2] COPD is characterised by chronic airway inflammation, airway remodelling and airflow limitation.[3] A reduced ratio of the Forced Expiratory Volume in 1 second (FEV_1) to Forced Vital Capacity (FVC) is a measure of obstruction and is used to diagnose COPD but also as an endophenotype for preclinical lung function.[3,4] Smoking is the most important risk factor for COPD and related impaired lung function.[2] COPD is a complex heterogeneous disease in which systemic features beyond airflow obstruction, including systemic inflammation, oxidative stress, muscle dysfunction, cachexia and vascular pathology occur.[5,6] Understanding these systemic effects may give new insights in the pathogenesis and progression of COPD but may alternatively yield important clues for preventive research.

Recent developments in metabolomics have made it possible to investigate the associations between circulating metabolites and the systemic effects in COPD. Glycoprotein acetyls (GlycA) were found to be predictive for several chronic diseases, among which COPD.[7] In a previous metabolomics study using proton Nuclear Magnetic Resonance (^{1}H-NMR), lower levels of lipoproteins, N,N-dimethylglycine and higher levels of glutamine, phenylalanine, 3-methylhistidine and ketone bodies were found in the circulation of ex-smoking COPD patients compared with ex-smoking controls.[8] In severe COPD patients, branched chain amino acids (BCAAs) were found to be lower, compared with controls.[8] Interestingly, BCAAs, 3-methylhistidine, ketone bodies, and triglycerides were negatively correlated with cachexia and positively correlated with systemic inflammation,[8] but these findings have not been replicated. Another question that remains to be answered is whether the metabolic changes are a cause or a consequence of COPD. If the latter is true, the metabolites may be relevant for the disease progression and prognosis.

To answer these questions, we performed a comprehensive integrative metabolic analysis to identify plasma metabolic measures associated with COPD and lung function levels, defined as FEV_1/FVC, using the NMR approach in a set of large epidemiological studies, in depth characterized for genetic and environmental risk factors. The discovery phase of the study was conducted in two population-based studies in the Netherlands, the Rotterdam Study (RS)[9] and the Erasmus Rucphen Family study (ERF).[10,11] A replication meta-analysis was conducted in Lifelines-DEEP study (LLD),[12] two cohorts of FINRISK study[13,14] and Prospective Investigation of the Vasculature in Uppsala Seniors (PIVUS) study.[15,16]

METHODS

Study population

Studies included in the discovery sample

The RS is a population-based study of 14,926 people older than 45 years, from the Ommoord area of Rotterdam, incorporating three cohorts: RS-I (established in 1989), RS-II (2000) and RS-III (2006), with multiple subsequent visits.[9] Participants filled in questionnaires, underwent physical examination and provided fasting blood samples at each visit. For this analysis, three independent samples from different RS cohorts were enrolled: Sample 1) visit 4 of RS-I (RS-I-4); sample 2) a combined sample, which we collectively call RS-E5 in this manuscript, comprising of visit 5 of RS-I (RS-I-5), visit 3 of RS-II (RS-II-3), and visit 2 of RS-III (RS-III-2); and sample 3) another independent set from RS-III-2.

ERF is a population-based study from the south-west of the Netherlands. It is a genetically isolated population comprising 3,465 living descendants of 22 couples from the 19th century and their spouses.[10] The baseline data collection was performed in 2002-2005 when participants underwent physical examinations, provided blood samples and completed questionnaires. A follow-up of the participants was performed in 2015-2018, reviewing the medical records at the general practitioner's office.

Both RS and ERF were approved by the Medical Ethics committee of the Erasmus Medical Center and all participants gave informed consent for participation in the study and for evaluation of the available information from their physicians.

Studies included in the replication sample

LLD is a sub-cohort of the large general population-based cohort study LifeLines, which was initiated to study genes, exposures and their interactions in the etiology of complex multifactorial diseases and healthy ageing.[17,18] LLD consists of 1,500 LifeLines participants who registered at the LifeLines research site in Groningen between April and August 2013. These subjects gave additional biological materials, including blood samples for metabolite and inflammation profiling, and extensive phenotype information.[12] Metabolic and lung function data were available for 717 LLD individuals and these subjects are included in the current study. LLD was approved by the ethics committee of the University Medical Center Groningen and all participants signed an informed consent prior to enrolment.

The FINRISK cohorts comprise cross-sectional population surveys that are carried out every 5 years since 1972, to assess the risk factors of chronic diseases (e.g. CVD,

diabetes, obesity, cancer) and health behaviour in the working age population (25-74 years of age), in 3-5 large study areas of Finland. The FINRISK surveys are conducted by the National Institute for Health and Welfare, THL (previously National Public Health Institute, KTL). Extensive information from each participant was collected at baseline via questionnaire and health examination with blood collection. The cohorts were followed up by linking them to national health registers. The cohorts FINRISK 1997 (total of 6898 participants) and an extension of FINRISK 2007, known as DIetary, Lifestyle and Genetic determinants of Obesity and Metabolic syndrome (DILGOM) study[19] (total of 4600 participants) are included in our replication sample for COPD analysis. The FINRISK 1997 study was approved by the Ethical Committee of the National Public Health Institute, while the DILGOM study was approved by the Coordinating Ethical Committee of the Hospital District of Helsinki and Uusimaa. All participants have signed an informed consent, allowing the use of their data and samples for studying environmental and genetic risk factors of chronic diseases.

The PIVUS study started in 2001 with the aim to investigate endothelial function as a prospective cardiovascular risk factor in elderly subjects. A random sample of Uppsala city residents were invited from the register of inhabitants within one month following their 70th birthday. No exclusion criteria were applied except that participants were required to have a Swedish identification number. In PIVUS, 1,016 agreed to participate, resulting in a participation rate of 50.1% of all invited, whereof 51.5 % were female. The participants have undergone a range of physical measurements, and given information about their medical history, lifestyle habits and regular medication. In addition, blood samples were drawn. The Ethics Committee of the University of Uppsala approved the study and the participants gave informed consent (approval number 00-419).

Assessment of COPD status and lung function measurements

COPD in the RS was defined as pre-bronchodilator $FEV_1/FVC<0.7$, assessed either by spirometry at the RS research center or by reviewing medical histories of the participants. Spirometry was performed in the RS by trained paramedical personnel, according to the guidelines of the American Thoracic Society/European Respiratory Society (ATS/ERS). When spirometry measurements were absent or uninterpretable, all files from specialists and general practitioners were reviewed to set a diagnosis of COPD. In total, this analysis included 541 COPD subjects and 4,407 subjects without COPD from all three RS cohorts.

For the ERF study, the doctor's diagnosis of COPD was confirmed by reviewing medical records based on $FEV_1/FVC<0.7$, with or without medication use. If the information on FVC was missing, the following criteria for COPD were used: $FEV_1<80\%$

of predicted, use of respiratory medication and a COPD diagnosis mentioned in the report of the respiratory specialist to the general practitioner. In total, 61 COPD subjects and 548 subjects without COPD were included from ERF. For ERF participants, we did not have lung function measurements at the time of the metabolomics measurements, so we did not include this cohort in the FEV_1/FVC analysis.

For LLD, COPD was defined as a FEV_1/FVC below 70%. Pre-bronchodilator spirometry was performed according to the ATS/ERS guidelines using a Welch Allyn Version 1.6.0.489, PC-based Spiroperfect with CA Workstation. Technical quality and results were assessed by well-trained assistants and abnormal results were re-evaluated by lung physicians.

In the FINRISK study the COPD information was extracted based on diagnoses and reimbursement information from the National health register, which include the Drug Reimbursement Register, the Care Register for Health Care, the Register for Prescribed Drug Purchases, the Causes-of-Death Register, and the Cancer Register. The maximum retrospective time period available for obtaining prevalent disease events was 20 years for DILGOM and 10 years for FINRISK97.

In the PIVUS study FEV_1 and FVC were assessed with spirometry using a Vitalograph Alpha spirometer (Vitalograph Ltd. Buckingham, UK) according to the American Thoracic Society recommendations.[20,21] The best value of three acceptable recordings was used. FEV_1 and FVC expressed as percent of predicted values, were adjusted for age, sex and height according to Hedenström's formula.[22,23]

Assessment of blood metabolites

Metabolic profiling in RS, ERF and LLD was done as part of the 4th Rainbow Project of the BioBanking for Medical Research Infrastructure of the Netherlands (BBMRI-NL) (https://www.bbmri.nl/omics-metabolomics/). To quantify the metabolite biomarkers from all samples fasting EDTA plasma samples were used for quantitative high-throughput ^{1}H-NMR metabolomics platform (Nightingale Ltd, Helsinki, Finland). Details and advantages of the NMR-based metabolomics analyses using plasma were described elsewhere.[24,25] Using this method, we were able to quantify a wide range of blood metabolite biomarkers such as lipoprotein fractions, amino-acids, cholesterol levels, glycerides, phospholipids, fatty acids, ketone bodies and metabolites related to inflammation and glycolysis. In total, 161 metabolites, overlapping between RS and ERF, were used in the discovery analysis.

Statistical analyses

Association of COPD and FEV_1/FVC with metabolites

The distributions of all metabolites were inspected for normality and natural logarithm or rank transformations were applied. Per cohort, we used transformed metabolite levels as independent variable and COPD status or FEV_1/FVC as dependent variables in logistic and linear regression models, respectively. The models were adjusted for age, sex, BMI (kg/m^2), lipid lowering medication use and smoking status (current, ex- or never smokers). For the discovery sample, the results from ERF, RS-I-4, RS-E5 and RS-III-2 were meta-analysed using fixed effect models in "*METAL*" software.[26] As the metabolites are known to be highly correlated, we applied the method by Li and Ji[27] to assess the number of independent metabolites. Using this method, we calculated that for the 161 metabolites, the number of independent tests was 45, which resulted in the Bonferroni significance threshold of P=0.001 (0.05/45). Significant metabolites were further tested for replication in the meta-analysis of LLD, FINRISK1997 and DILGOM studies for the COPD analysis and of LLD and PIVUS studies for the FEV_1/FVC analysis. Again, the same regression models were used for the fixed effect meta-analysis in "*METAL*" software.

For significant COPD metabolites, we investigated the odds ratios per quartile of the metabolite distribution in the discovery sample. To investigate the effects of smoking on this association, we used two logistic regression models, one adjusted for age, sex, BMI and lipid lowering medication use, and a second model additionally adjusted for smoking status (current, ex- and never smokers). Results from each cohort were combined using inverse-variance weighted fixed effects meta-analysis in "*rmeta*" package in R.

Association of smoking with metabolites

We further tested if the replicated metabolites from the COPD and FEV_1/FVC analyses (as dependent variables) were associated with smoking status (current, ex- and never smokers) and pack-years of smoking. We used models adjusted for age, sex, BMI and lipid lowering medication use in the discovery sample. Associations with smoking status were further tested for replication in the FINRISK1997 and DILGOM studies, using same models. All analyses were performed in R (version 3.2.1.). Replication studies did not have pack-years data to investigate further. Next, for comparison, we tested the same models in the discovery cohort after excluding the COPD cases.

Association of genetic variants with metabolites

We have used a bi-directional approach in which we examined whether: 1) the genetic determinants of the significant metabolic measures are associated with COPD and lung function, which would lead to the conclusion that the metabolites are most likely driving the disease; 2) the genetic determinants of COPD are associated with significant metabolites when the metabolites would most likely be altered as an integral part of the disease pathophysiology and may be biomarkers. In these analyses we use the genes as instrumental variable and a method which is referred to in genetics as a bi-directional Mendelian Randomization (MR) approach.[28] MR was conducted using "*gtx*" package in R.[29] To maximize the statistical power of the study[28] we used the genetic information from previously published genome-wide association studies (GWAS) on metabolites (Model 1)[25] and COPD (Model 2).[30] Genetic risk score (GRS), summarizing the effect of the SNPs genome-widely associated with either the significant metabolites or COPD, were used as instrumental variable. In GRS we included unique SNPs (mapped to human genome build hg19) in low linkage disequilibrium based on the data in ERF study (within 500Kb and R^2<0.05). MRs were performed with GRS explaining >1% of the variance, because the power of the MR using GRS that explains a lower proportion of the variance is too low to yield trustable results. To control for pleiotropic effects, we checked the heterogeneity of the SNPs included in the GRS and excluded the SNPs which were also genome-wide significantly associated with the outcome.

Association of metabolites with mortality

To investigate whether metabolites have a clinical utility in predicting COPD, we constructed classical receiver operating curves (ROC) and compared area's under the curve (AUC).[31] To further investigate whether the identified metabolites may act as biomarker of the disease prognosis, we performed a survival analysis in SPSS, similar to the previous study by Fischer and colleagues for all-cause mortality, ignoring any underlying morbidity.[32] To check whether the metabolites associated with mortality in COPD patients, we performed the Cox proportional hazards model in three RS cohorts. Analyses were adjusted for age at sampling, sex and smoking. We further performed a similar analysis using four quartiles of metabolite, testing in COPD cases and controls.

RESULTS

Descriptive characteristics of the samples

Descriptive characteristics of all cohorts used in the analysis are presented in **Table 1**.

Table 1. Discovery population characteristics per cohort

	Discovery cohorts				Replication cohorts			
Study	**ERF**	**RS-I-4**	**RS-E5**	**RS-III-2**	**LLD**	**FINRISK97**	**DILGOM**	**PIVUS**
N	609	2777	686	1485	717	6898	4600	854
Age, mean (sd)	49.0 (13.3)	74.8 (6.5)	68.4 (5.7)	62.8 (5.8)	46.0 (14.3)	48.0 (13.1)	52.3 (13.5)	70 (0)
Women, % (n)	55.8 (340)	58.2 (1615)	57.6 (395)	57.8 (859)	56.3 (404)	51.6 (3561)	53.4 (2458)	48.2 (412)
COPD cases, % (n)	10.0 (61)	12.1 (336)	10.3 (71)	9.0 (134)	13.8 (99)	0.6 (43)	0.8 (35)	NA
FEV_1/FVC, mean (sd), % of all	NA	0.73 (0.08), 48.8	0.76 (0.07), 91.3	0.77 (0.07), 91.9	0.77 (0.08), 100	NA	NA	0.76 (0.11), 100
BMI, mean (sd)	27.2 (4.85)	27.4 (4.1)	27.8 (4.3)	27.4 (4.5)	25.4 (4.1)	26.6 (4.5)	27.2 (4.8)	27.1 (4.26)
Current smokers, % (n)	43.3 (264)	12.6 (349)	9.5 (65)	13.7 (203)	20.5 (147)	23.9 (1648)	17.6 (810)	10.2 (87)
Ex-smokers, % (n)	30.0 (183)	56.1 (1559)	57.0 (391)	50.2 (746)	NA	22.9 (1577)	26.3 (1210)	41.5 (354)
Never smokers, % (n)	26.6 (162)	31.3 (869)	33.5 (230)	36.1 (536)	79.4 (570)	53.2 (3673)	56.1 (2580)	48.2 (412)
Pack-years of smoking, mean (sd), % of all[a]	24.9 (20.4) 72.7	24.2 (23.4), 64.7	22.0 (20.8) 66.3	19.5 (20.3) 63.8	NA	NA	NA	NA
Lipid lowering medication users, % (n)	12.3 (75)	22.4 (621)	32.5 (223)	22.2 (329)	3.9 (28)	3.4 (237)	15.7 (721)	16.5 (141)

RS-E5: consists of RS-I-5, RS-II-3 and RS-III-2; [a] Pack-years calculated in current and ex-smokers only, so "% of all" excludes never smokers; NA - not applicable;

Comparing the discovery cohorts, ERF participants were younger (mean age 49.0±13.3) and had a higher percentage of current smokers compared to the participants of the three RS cohorts (RS-I-4 mean age 74.8±6.5; RS-E5 mean age 68.4±5.7; RS-III-2 mean age 62.8±5.8). The RS cohorts had a higher percentage of lipid lowering medication users, compared to ERF (**Table 1**). The mean FEV_1/FVC and BMI were comparable across the studies. Descriptive characteristics for COPD cases and

subjects without COPD separately in the discovery cohorts are provided in **eTable 1** in the Supplement**.** In general, COPD subjects were older and more often smokers compared with subjects without COPD.

Association of COPD and FEV_1/FVC with metabolites

In the discovery sample, six plasma metabolites were associated with COPD at a significance level of 5% (**Table 2, Figure 1**).

Table 2. Metabolites associated with COPD in the discovery and replication studies

	Discovery meta-analysis						Replication meta-analysis					
Metabolite	**B**	**SE**	**OR**	**P-value**	**Direction[a]**	**N**	**β**	**SE**	**OR**	**P-value**	**Direction[b]**	**N**
GlycA	0.152	0.044	1.16	**5.6×10^{-4}**	++++	5557	0.266	0.053	1.30	**1.8×10^{-6}**	+++	12205
3-hydroxybutyrate	0.122	0.041	1.13	0.003	++++	5002	-0.031	0.057	0.97	0.662	+--	12173
Histidine	-0.097	0.047	0.91	0.037	----	5534	-0.153	0.063	0.86	0.020	---	12200
Free cholesterol in med. HDL	0.099	0.049	1.10	0.045	+-++	5557	0.004	0.063	1.00	0.867	+--	12208
Acetoacetate	0.084	0.042	1.09	0.047	++-+	5551	-0.061	0.059	0.94	0.360	---	12204
18:2, linoleic acid	-0.095	0.048	0.91	0.049	+---	5546	-0.036	0.057	0.96	0.238	+-+	12167

Model adjusted for age, sex, BMI, lipid lowering medication and smoking status; GlycA – Glycoprotein acetyls; HDL - high density lipoprotein; β - effect size; SE - standard error; OR - odds ratio; Direction - direction of the effect in individual studies; N - meta-analysis sample size; [a] Direction of the effect in the discovery studies in order: ERF, RS-III-2, RS-E5, RS-I-4; [b] Direction of the effect in the replication studies in order: DILGOM, FINRISK 1997, LLD; In bold: significant results (P<0.001).

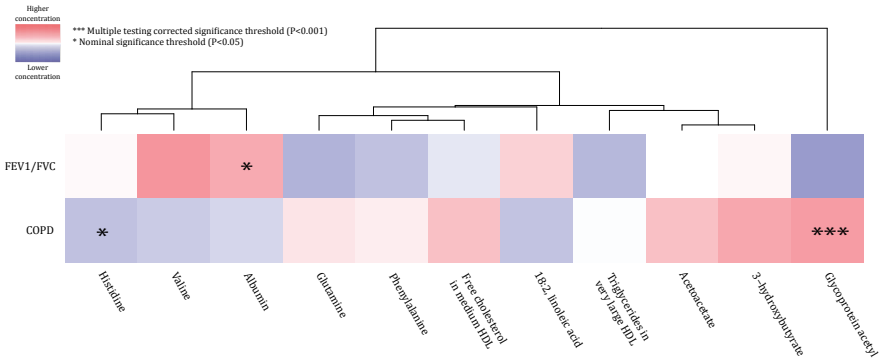

Figure 1. Top metabolites associated with COPD and/or FEV_1/FVC. Colors represent standardized effect estimates of the metabolite association with corresponding trait (COPD, FEV_1/FVC). Red color represents the trait associated with an increase in metabolite concentration, while blue represents a decrease. For replicated metabolites, replication significance threshold is shown with stars: *P<0.05 and ***P<0.001. HDL – high-density lipoprotein.

At nominal significance, higher levels of GlycA (OR=1.16; P=5.6×10^{-4}), 3-hydroxybutyrate (OR=1.13; P=0.003), free cholesterol in medium HDL (OR=1.10; P=0.045) and acetoacetate (OR=1.09; P=0.047) were associated with a higher prevalence of COPD. Higher levels of histidine and 18:2 linoleic acid (OR=0.91 for both, P=0.04 and P=0.05 respectively) were associated with a lower prevalence of COPD. When taking into account the multiple testing correction threshold, only GlycA was significantly associated with COPD (P= 5.6×10^{-4}).

We tested all six metabolites for replication in the independent samples. The association of higher levels of GlycA with COPD was significantly replicated (OR=1.30, P=1.75×10^{-4}) in the 12,205 participants of the replication sample, after multiple testing correction.

Findings for the FEV_1/FVC ratio were not consistent over the discovery and replication studies. Adjusting for multiple testing, we found in the discovery cohorts that lower levels of valine (β=0.005, P=2.5×10^{-4}) and higher levels of GlycA (β=-0.005, P=4.5×10^{-4}) were associated with a lower FEV_1/FVC ratio (**Table 3, Figure 1**). Other metabolites that reached nominal significance in the discovery included albumin which was positively associated with FEV_1/FVC, and glutamine, triglycerides in very large HDL and phenylalanine which were negatively associated with FEV_1/FVC (**Table 3, Figure 1**).

Table 3. Top metabolites associated with FEV_1/FVC - Results of the discovery and replication studies

	Discovery meta-analysis					Replication meta-analysis				
Metabolite	**B**	**SE**	**P-value**	**Direction[a]**	**N**	**β**	**SE**	**P-value**	**Direction[b]**	**N**
Valine	0.005	0.001	**2.5×10-4**	+++	3324	-0.0015	0.0023	0.5314	-+	1460
GlycA	-0.005	0.001	**4.5×10-4**	---	3324	-0.0010	0.0022	0.6438	--	1463
Albumin	0.004	0.001	0.0047	+++	3324	0.0045	0.0021	0.0353	++	1463
Glutamine	-0.003	0.001	0.0097	---	3323	0.0029	0.0023	0.1923	++	1393
Triglycerides in very large HDL	-0.003	0.001	0.0160	---	3324	0.0031	0.0022	0.1491	++	1469
Phenylalanine	-0.003	0.001	0.0334	---	3324	-0.0012	0.0023	0.5899	+-	1450

Model adjusted for age, sex, BMI, lipid lowering medication and smoking status; HDL - high density lipoprotein; β - effect size; SE - standard error; Direction - direction of the effect in individual studies; N - meta-analysis sample size; [a] Direction of the effect in the discovery studies in order: RS-III-2, RS-E5, RS-I-4; [b] Direction of the effect in the replication studies in order: LLD, PIVUS; In bold: significant results (P<0.001).

Only the association of FEV_1/FVC to albumin showed nominal significance in the replication samples (β=0.005, P=0.03), but none were significantly associated when considering multiple testing correction.

Association of smoking with metabolites

Although the above analyses were adjusted for smoking, metabolite levels may have changed as a consequence of smoking and may be an intermediary in the relation of smoking to COPD. We tested the association of smoking status and pack-years with the metabolite levels that were significantly associated with COPD. The results are presented in the **eTable 2** in the Supplement. GlycA was significantly positively associated with both smoking status (current, ex- or never smoker) and pack-years of smoking (β=0.15, P=1.31×10^{-22} and β=0.006, P=2.52×10^{-16}, respectively). **eTable 2** in the Supplement also shows the replication of the association between the metabolites and smoking status in the FINRISK1997 and DILGOM studies. Data on pack-years was not available in these replication studies. When excluding COPD cases from the discovery sample, the identified associations of GlycA with smoking status and pack-years attenuated, both in the effect size and p-value, yet remained significant (**eTable 3** in the Supplement).

Association of genetic variants of circulating GlycA and COPD with metabolites

Next, we performed a Mendelian Randomisation experiment investigating the hypothesis that: 1) GlycA is increasing the risk of COPD and therefore the genetic determinants of GlycA (used as instrumental variables) are also associated with COPD and 2) the opposite scenario is true in which (pre)clinical COPD pathology increases GlycA levels. The results of both models are presented in **Table 4.**

Table 4. Results of the bidirectional Mendelian randomization approach on GlycA and COPD.

Model	Exposure	Outcome	R^2	nSNP	β	SE	P-value	Ors	Phet
1	GlycA	COPD	0.023	9	-0.001	0.047	**0.988**	13.6	0.09
2	COPD	GlycA	0.03	7	0.165	0.053	**0.0018**	9.6	0.14

R^2 - the explained variance in the exposure by applied genetic risk score; nSNP - number of SNPs used to construct the genetic risk score; β - the weighted effect of the genetic risk score of exposure on outcome; SE - standard error; Significance threshold = P-value < 0.05; Ors: heterogeneity test statistic; Phet: heterogeneity test P-value

The GRS for Model 1 included nine independent SNPs (R^2=0.023, **eTable 4** in the Supplement) and yielded no significant evidence for association (P=0.99). In Model 2, we found that genes associated with a higher risk of COPD are also associated with higher levels of GlycA (**Table 4**, P=0.002), suggesting that COPD pathology increased GlycA levels. This analysis is based on seven independent SNPs in the GRS (R^2=0.03, **eTable 5** in the Supplement). No heterogeneity effect or potential pleiotropic SNPs were found in either model.

Is circulating GlycA predictive biomarker for COPD?

The question to answer next is whether GlycA in the circulation is a biomarker of early pathology thus can be used as a predictive or diagnostic biomarker or rather a prognostic biomarker for mortality in COPD patients. To this end, we performed an analysis in the Rotterdam Study in which we associated GlycA to the future risk of COPD. We determined the relative risk by quartile of the GlycA concentrations in the circulation, using the lowest quartile as a reference (**eTable 6** in the Supplement). Only incident patients are included in these analyses; prevalent COPD patients are excluded. Compared to the lowest quartile, those subjects in the highest quartile of GlycA had a 1.99-fold (95% Confidence interval: 1.52-2.60) higher risk of COPD, after adjustment for age, sex, BMI and lipid lowering medication. Smoking accounted for a part of the observed association between plasma GlycA and COPD attenuating the OR for those in the highest quartile of GlycA to 1.74, while the association remained significant (95% Confidence interval: 1.32-2.28). To test whether circulating GlycA adds to the predictive value, we compared the AUC curves for the models including: 1) age and sex (AUC=0.601); 2) age, sex and smoking (AUC=0.670) and 3) age, sex, smoking and circulating GlycA levels in blood (AUC=0.675). The AUC comparing model 2 and 1 shows that smoking is associated with an increase in AUC by 0.069. Adding circulating GlycA increased the AUC further by only 0.005 (**eFigure 1**).

Is circulating GlycA a prognostic biomarker for mortality in COPD?

Previous study has shown that GlycA is a predictor of mortality.[32] We confirm this in current study, after adjustment for age, sex and smoking (HR=1.16, $P=4.93\times10^{-9}$) (**eTable 7**). We therefore tested the hypothesis that GlycA is a marker of COPD related to future mortality. We first compared mortality across the quartiles of GlycA and found that those in the highest quartile have 1.42-fold (95% Confidence interval: 1.24-1.63, $P=7.61\times10^{-7}$) increased risk of mortality during follow-up compared to those in the lowest quartile (**eTable 7**). However, when stratifying these analyses by COPD status, we observed that this association is driven by controls (**eTable 7, eFigure 2**). In COPD patients, circulating GlycA levels were not significantly associated with mortality when studying GlycA as a continuous variable (HR=1.06, $P=0.31$) nor for those in the highest quartile (HR=1.02, $P=0.93$ in COPD cases). In those without COPD, the association of GlycA to mortality was stronger and significant (HR=1.18, $P=1.32\times10^{-9}$).

DISCUSSION

In our metabolome-wide discovery analysis we identified 11 plasma metabolites associated with COPD or lung function levels (FEV_1/FVC) at marginal significance. Of the 11, only higher levels of GlycA were significantly associated with COPD when adjusting for multiple testing and this is the only metabolite we could replicate in an independent sample. The association of GlycA with COPD remained significant when adjusting for smoking. GlycA levels in the circulation were significantly associated with smoking. Our MR analysis showed that the genetic predisposition to COPD associates with GlycA. Although GlycA was found to be a predictor of mortality in the general population,[33] the metabolite did not predict mortality in COPD patients.

The most convincing and interesting finding of our study is that of GlycA. We recently associated this metabolite with the incidence of a variety of disorders, including COPD in our study based on record linkage.[7] The record linkage study focussed specifically on the relation of GlycA with a wide variety of disorders. Using two population-based cohorts, we identified new associations with GlycA including incident COPD, alcoholic liver disease, chronic renal failure, glomerular diseases and inflammatory polyarthropathies. The GlycA associations were for a large part independent of that of high-sensitivity C-reactive protein (hsCRP), but GlycA and hsCRP also share contributions to mortality risk, suggesting chronic inflammation as the common pathway. GlycA is shown to be a biomarker for chronic inflammation, neutrophil activity and risk of future severe infection, even superior compared with CRP.[34,35]

The present study extends our findings published previously in that we have increased the number of NMR metabolites studied and found that GlycA is the only metabolite significantly associated with COPD when adjusting for multiple testing. In the present study we also have studied effects of GlycA beyond COPD, and found GlycA is consistently associated with smoking status and quantity (pack-years of smoking). Smoking is related to GlycA levels in the circulation but it does not explain the association between GlycA and COPD. This is compatible with the view that smoking, the major driver of COPD risk in the population, is associated with GlycA which in turn is associated with COPD risk. In the present paper we used data integration approach (MR) to test the hypothesis that GlycA increases the risk of COPD causally or rather is a biomarker that is part of the disease pathogenesis. The findings of the present paper suggest that the latter is more likely, as the genes associated with COPD also associate with GlycA levels. No marginally significant support was found for the hypothesis that GlycA is a determinant of COPD: the genes that are known to determine GlycA levels are not associated with the risk of COPD.

In the present paper we do not find evidence that GlycA is associated with COPD mortality. Such a relationship was seen in our findings for cardiovascular disease. GlycA not only increased the risk of incident cardiovascular disease[7,36] but was also associated with a 5-fold increased 12-year risk of mortality in those with the highest GlycA levels.[7]

GlycA, also called orosomucoid,[37] is a positive acute phase protein, and its concentration increases in response to systemic tissue injury, inflammation or infection.[38] GlycA is mainly produced by the liver, but it is also synthesized in myelocytes and released by activated neutrophils.[39] Being a type I acute phase protein, GlycA is induced by cytokines, interleukins and tumor necrosis factor alpha (TNFα),[40,41] which among others stimulate a systemic inflammatory response in COPD patients who lose weight.[42] GlycA is one of the main drug binding proteins, carrying basic and neutral lipophilic drugs such as steroid hormones or medications in blood.[43]

A strength of our study is that it is the largest and most comprehensive metabolic study of COPD and lung function. Another strength is the use of the NMR platform, which is valued for being non-invasive, non-destructive, fast and for providing highly reproducible results.[44] Our MR approach allowed us to gain more insight into the direction of the effects, yielding a new interpretation of our data suggesting that GlycA is an independent risk factor of COPD. Yet we have to acknowledge that a limitation of MR is that our knowledge of the genetic determinants of both COPD and GlycA is very limited. In addition, we acknowledge possible limitations of MR due to pleiotropy, the lack of trans-ethnic studies and remaining bias due to canalization.

Altogether, combining the epidemiological data with our MR analyses suggests that GlycA is a predictor of COPD and may be a mediator in the causal pathway linking smoking to COPD. Further functional studies investigating the role of GlycA in COPD will provide more insight into the pathogenesis, prognosis and treatment response of COPD and lung function decline. Our study highlights the power of cross-omics and epidemiological data integration.

REFERENCES

1. World Health Organizaon - WHO. *World Health Statistics 2008*. World Health Organization; 2008.
2. Lozano R, Naghavi M FK. Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet*. 2012;380(380):2095–2128. doi:https://doi.org/10.1016/S0140-6736(12)61728-0
3. Vestbo J, Hurd SS, Agustí AG, et al. Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease GOLD executive summary. *Am J Respir Crit Care Med*. 2013;187(4):347-365. doi:10.1164/rccm.201204-0596PP
4. Wain L V., Shrine N, Artigas MS, et al. Supplementary: Genome-wide association analyses for lung function and chronic obstructive pulmonary disease identify new loci and potential druggable targets. *Nat Genet*. 2017;49(3):416-425. doi:10.1038/ng.3787
5. Vestbo J, Anderson W, Coxson HO, et al. Evaluation of COPD Longitudinally to Identify Predictive Surrogate End-points (ECLIPSE). *Eur Respir J*. 2008;31(4):869-873. doi: 10.1183/09031936.00111707
6. Agusti A, Calverley PMA, Celli B, et al. Characterisation of COPD heterogeneity in the ECLIPSE cohort. *Respir Res*. 2010;11(1):122.
7. Kettunen J, Ritchie SC, Anufrieva O, et al. Biomarker Glycoprotein Acetyls Is Associated With the Risk of a Wide Spectrum of Incident Diseases and Stratifies Mortality Risk in Angiography Patients. *Circ Genomic Precis Med*. 2018;11(11):e002234. doi:10.1161/CIRCGEN.118.002234
8. Ubhi BK, Riley JH, Shaw PA, et al. Metabolic profiling detects biomarkers of protein degradation in COPD patients. *Eur Respir J*. 2012;40(2):345-355. doi:10.1183/09031936.00112411
9. Ikram MA, Brusselle GGO, Murad SD, et al. The Rotterdam Study: 2018 update on objectives, design and main results. *Eur J Epidemiol*. 2017;32(9):807-850. doi:10.1007/s10654-017-0321-4
10. Pardo LM, MacKay I, Oostra B, van Duijn CM, Aulchenko YS. The effect of genetic drift in a young genetically isolated population. *Ann Hum Genet*. 2005;69(3):288-295. doi: 10.1046/j.1529-8817.2005.00162.x
11. Henneman P, Aulchenko YS, Frants RR, van Dijk KW, Oostra BA, van Duijn CM. Prevalence and heritability of the metabolic syndrome and its individual components in a Dutch isolate: The Erasmus Rucphen Family study. *J Med Genet*. 2008;45(9):572-577. doi:10.1136/jmg.2008.058388
12. Tigchelaar EF, Zhernakova A, Dekens JAM, et al. Cohort profile: LifeLines DEEP, a prospective, general population cohort study in the northern Netherlands: study design and baseline characteristics. *BMJ Open*. 2015;5(8):e006772. doi:10.1136/bmjopen-2014-006772
13. Borodulin K, Tolonen H, Jousilahti P, et al. Cohort Profile: The National FINRISK Study. *Int J Epidemiol*. 2018;47(3):696-696i. doi:10.1093/ije/dyx239
14. Borodulin K, Vartiainen E, Peltonen M, et al. Forty-year trends in cardiovascular risk factors in Finland. *Eur J Public Health*. 2015;25(3):539-546. doi:10.1093/eurpub/cku174
15. Lind L, Fors N, Hall J, Marttala K, Stenborg A. A comparison of three different methods to evaluate endothelium-dependent vasodilation in the elderly: the Prospective Investigation of the Vasculature in Uppsala Seniors (PIVUS) study. *Arterioscler Thromb Vasc Biol*. 2005; 25(11):2368-2375. doi:10.1161/01.ATV.0000184769.22061.da
16. Salihovic S, Ganna A, Fall T, et al. The metabolic fi ngerprint of p , p ′-DDE and HCB exposure in humans. *Environ Int*. 2016. doi:10.1016/j.envint.2015.12.015

17. Scholtens S, Smidt N, Swertz MA, et al. Cohort Profile: LifeLines, a three-generation cohort study and biobank. *Int J Epidemiol*. 2015;44(4):1172-1180. doi:10.1093/ije/dyu229
18. Stolk RP, Rosmalen JGM, Postma DS, et al. Universal risk factors for multifactorial diseases. *Eur J Epidemiol*. 2008;23(1):67-74. doi:10.1007/s10654-007-9204-4
19. Konttinen H, Llewellyn C, Silventoinen K, et al. Genetic predisposition to obesity, restrained eating and changes in body weight: a population-based prospective study. *Int J Obes (Lond)*. 2018;42(4):858-865. doi:10.1038/ijo.2017.278
20. Engström G, Gerhardsson de Verdier M, Dahlbäck M, Janson C, Lind L. BP variability and cardiovascular autonomic function in relation to forced expiratory volume: a population-based study. *Chest*. 2009;136(1):177-183. doi:10.1378/chest.08-2529
21. Standardization of Spirometry, 1994 Update. American Thoracic Society. *Am J Respir Crit Care Med*. 1995;152(3):1107-1136. doi:10.1164/ajrccm.152.3.7663792
22. Hedenström H, Malmberg P, Fridriksson H V. Reference values for lung function tests in men: regression equations with smoking variables. *Ups J Med Sci*. 1986;91(3):299-310. http://www.ncbi.nlm.nih.gov/pubmed/3811032. Accessed November 15, 2018.
23. Hedenström H, Malmberg P, Agarwal K. Reference values for lung function tests in females. Regression equations with smoking variables. *Bull Eur Physiopathol Respir*. 21(6):551-557. http://www.ncbi.nlm.nih.gov/pubmed/4074961. Accessed November 15, 2018.
24. Soininen P, Kangas AJ, Wurtz P, Suna T, Ala-Korpela M. Quantitative Serum Nuclear Magnetic Resonance Metabolomics in Cardiovascular Epidemiology and Genetics. *Circ Cardiovasc Genet*. 2015;8(1):192-206. doi:10.1161/Circgenetics.114.000216
25. Kettunen J, Demirkan AA, Würtz P, et al. Genome-wide study for circulating metabolites identifies 62 loci and reveals novel systemic effects of LPA. *Nat Commun*. 2016;7:11122. doi:10.1038/ncomms11122
26. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics*. 2010;26(17):2190-2191. doi:10.1093/bioinformatics/btq340
27. Li J, Ji L. Adjusting multiple testing in multilocus analysis using the eigenvalues of a correlaton matrix. *Heredity (Edinb)*. 2005;95(3):1-7.
28. Liu J, van Klinken JB, Semiz S, et al. A Mendelian Randomization Study of Metabolite Profiles, Fasting Glucose, and Type 2 Diabetes. *Diabetes*. 2017;66(11):2915-2926. doi:10.2337/db17-0199
29. Dastani Z, Hivert MF, Timpson NJ, et al. Novel loci for adiponectin levels and their influence on type 2 diabetes and metabolic traits: a multi-ethnic meta-analysis of 45,891 individuals. *PLoS Genet*. 2012;8(3):e1002607. doi:10.1371/journal.pgen.1002607PGENETICS-D-11-02099 [pii]
30. Hobbs BD, De Jong K, Lamontagne M, et al. Genetic loci associated with chronic obstructive pulmonary disease overlap with loci for lung function and pulmonary fibrosis. *Nat Genet*. 2017;49(3):426-432. doi:10.1038/ng.3752
31. Akobeng AK. Understanding diagnostic tests 3: receiver operating characteristic curves. *Acta Paediatr*. 2007;96(5):644-647. doi:10.1111/j.1651-2227.2006.00178.x
32. Fischer K, Kettunen J, Würtz P, et al. Biomarker Profiling by Nuclear Magnetic Resonance Spectroscopy for the Prediction of All-Cause Mortality: An Observational Study of 17,345 Persons. Minelli C, ed. *PLoS Med*. 2014;11(2):e1001606. doi:10.1371/journal.pmed.1001606
33. Lawler PR, Akinkuolie AO, Chandler PD, et al. Circulating N-Linked Glycoprotein Acetyls and Longitudinal Mortality Risk. *Circ Res*. 2016;118(7):1106-1115. doi:10.1161/CIRCRESAHA.115.308078

34. Ritchie SC, Würtz P, Nath AP, et al. The Biomarker GlycA Is Associated with Chronic Inflammation and Predicts Long-Term Risk of Severe Infection. *Cell Syst*. 2015;1(4):293-301. doi:10.1016/j.cels.2015.09.007
35. Otvos JD, Shalaurova I, Wolak-Dinsmore J, et al. GlycA: A Composite Nuclear Magnetic Resonance Biomarker of Systemic Inflammation. *Clin Chem*. 2015;61(5):714-723. doi:10.1373/clinchem.2014.232918
36. Gruppen EG, Riphagen IJ, Connelly MA, Otvos JD, Bakker SJL, Dullaart RPF. GlycA, a Pro-Inflammatory Glycoprotein Biomarker, and Incident Cardiovascular Disease: Relationship with C-Reactive Protein and Renal Function. Shimosawa T, ed. *PLoS One*. 2015;10(9):e0139057. doi:10.1371/journal.pone.0139057
37. Fournier T, Medjoubi-N N, Porquet D. Alpha-1-acid glycoprotein. *Biochim Biophys Acta - Protein Struct Mol Enzymol*. 2000;1482(1):157-171. doi:https://doi.org/10.1016/S0167-4838(00)00153-9
38. Gabay C, Kushner I. Acute-Phase Proteins and Other Systemic Responses to Inflammation. *N Engl J Med*. 1999;340(6):448-454. doi:10.1056/NEJM199902113400607
39. Theilgaard-Mönch K, Jacobsen LC, Rasmussen T, et al. Highly glycosylated alpha1-acid glycoprotein is synthesized in myelocytes, stored in secondary granules, and released by activated neutrophils. *J Leukoc Biol*. 2005;78(2):462-470. doi:10.1189/jlb.0105042
40. Moshage H. REVIEW ARTICLECytokines and the Hepatic Acute Phase Response. *J Pathol*. 1997;266(August 1996):257-266.
41. Bistrian BR. Acute phase proteins and the systemic inflammatory response. *Crit Care Med*. 1999;27(3):452-453. doi:10.1097/00003246-199903000-00004
42. de Godoy I, Donahoe M, Calhoun WJ, Mancino J, Rogers RM. Elevated TNF-alpha production by peripheral blood monocytes of weight-losing COPD patients. *Am J Respir Crit Care Med*. 1996;153(2):633-637. doi:10.1164/ajrccm.153.2.8564110
43. Israili ZH, Dayton PG. Human alpha 1-acid glycoprotein and its interactions with drugs. *Drug Metab Rev*. 2001;33(2):161-235.
44. Pan Z, Raftery D. Comparing and combining NMR spectroscopy and mass spectrometry in metabolomics. *Anal Bioanal Chem*. 2007;387(2):525-527. doi:10.1007/s00216-006-0687-8

Chapter 3.3

Genetic correlation of Chronic Obstructive Pulmonary Disease and non-pulmonary comorbidity

Ivana Nedeljković, Maaike de Vries, Judith M. Vonk, Lies Lahousse, Bruno H.C. Stricker, Guy G. Brusselle, Najaf Amin, Marike Boezen, Cornelia M. van Duijn.

This chapter is in preparation.

The supplemental material for this paper is available at: https://drive.google.com/file/d/1xJqj60sAejh6dkrwyLzBblYbtXzZ-qyh/view?usp=sharing

ABSTRACT

COPD is a complex disease which co-occurs with a range of pulmonary and systemic pathologies. The co-occurrence may be explained by common determinants or a joint pathogenesis that may in part be driven by genetic factors. We explored the genetic overlap between COPD and comorbid conditions. For this study, genome wide association study (GWAS) summary statistics for COPD were obtained from the International COPD Genetics Consortium (ICGC). Linkage Disequilibrium (LD) regression was used to determine the genetic correlation of COPD with phenotypes for which GWAS summary statistics data are publicly available at the LD Hub database on 16th February 2019 (http://ldsc.broadinstitute.org/ldhub/).

As expected, we find marginal significant evidence for genetic correlation of COPD with a variety of comorbidities including cardio-metabolic traits (acute myocardial infarction, coronary artery disease, angina pectoris, hypertension, diabetes, chronic kidney disease). The strongest association were seen for diabetes based on significance and essential hypertension in terms of strength of correlation. An unexpected but intriguing finding is the correlation of COPD with family history of depression in siblings (most significant finding) and attention deficit hyperactivity disorder (strongest correlation). Finally, we find marginal evidence for significance of genetic correlation of COPD to female reproductive traits, autoimmune diseases of the bowel and aging disorders such as cataract.

This study pinpoints diseases that are genetically correlated with COPD and highlights the significance of studying comorbidities of COPD.

INTRODUCTION

Chronic obstructive pulmonary disease (COPD) is a common cause of morbidity and mortality worldwide, characterized by persistent and progressive limitation in lung function.[1,2] COPD often co-occurs with other pulmonary pathology. Co-occurrence may be explained by a shared pathogenesis which is part of the COPD pathology spectrum or a consequence of COPD.[3] The disorders associated with COPD include a range of pulmonary pathologies such as asthma, pneumonia, pulmonary hypertension, pulmonary embolism, obstructive sleep apnoea, idiopathic pulmonary fibrosis and lung cancer. However, COPD is also a systemic disorder that is associated with non-pulmonary comorbid diseases.[4,5] These involve cardio-metabolic pathology, loss of bone mass density, depression, among many others.[6]

Both the pulmonary and non-pulmonary comorbidities may in part be explained by common factors such as smoking, sedentary behavior, lcohol, diet, ageing and polypharmacy or by shared pathophysiological mechanisms such as the systemic inflammation.[4,6] There is evidence for genetic correlation between COPD and asthma.[7] Although only one of the COPD loci identified to date was also genome wide significantly implicated in asthma,[7,8] an integrative genomic approach called Linkage Disequilibrium (LD) regression brought to surface a strong and significant genetic correlation between COPD, asthma and lung function parameters.[7] In addition, there was also a significant evidence for genetic correlation with height and smoking.[7] To our knowledge, up until now, only these specific traits were tested using a "candidate-disease" approach based on prior knowledge of COPD comorbidity.

The aim of the present analyses was to explore the genetic overlap between COPD and other diseases using LD regression and the wealth of GWAS data available for data mining.

METHODS

For this study, GWAS results for COPD were obtained from the largest study performed to date by Hobbs et al.[7] This study was conducted by the International COPD Genetics Consortium (ICGC), in which 22 studies with GWAS and COPD data (case-control or population based) were included. For the COPD GWAS, the consortium performed a genetic association analyses in 15,256 cases and 47,936 controls. To benchmark the finding the consortium replicated top results ($P < 5\times10^{-6}$) in 9,498 cases and 9,748 controls. In the combined meta-analysis, 22 loci were associated to

COPD at genome-wide significance. All SNPs, independent of the significance, are included in the genetic correlation analysis of the present study.

LD regression exploits data of the GWAS, available in the public domain. For many different comorbidities, summary-level GWAS results are publicly available at LD Hub for LD regression.[9] We evaluated the genetic correlations of COPD with the 126 diseases on LD Hub (access date: 16th February 2019). For a given pair of traits, LD score regression estimates the expected population correlation between the best possible linear SNP-based predictor for each trait. The analysis is restricted to common SNPs. As pulmonary comorbidities are already investigated and shown to have genetic overlap with COPD, we were interested in extra-pulmonary comorbid conditions. As we focus on the question whether there is a genetic overlap between COPD and extra-pulmonary disorders, we focus on disorders and other determinants that are positively correlated to COPD.

RESULTS

Our LD score regression showed marginal significant positive correlations of COPD and 20 different traits. All significant (P<0.05) extra-pulmonary correlations are presented in **Table 1.** These and other diseases tested in the analysis are presented in **Supplementary table 1**, which also shows confirmation of the already established genetic overlap with pulmonary comorbidities**.** The most statistically significant genetic correlation of COPD was found for depression of a sibling (r_g=0.29; P=0.0007) while attention deficit hyperactivity disorder (ADHD) showed a remarkably strong correlation (r_g=0.51; P=0.0031). Further among neuro-psychiatric diseases, we found evidence for correlation of COPD and schizophrenia (r_g=0.09; P=0.049).

Various cardio-metabolic diseases were found to be genetically correlated to COPD. The strongest association in terms of significance is seen for diabetes diagnosed by doctor (r_g=0.17; P=0.002). Also diabetes of a sibling and mother (r_g=0.16; P=0.02 and r_g=0.17; P=0.02, respectively) is found to be positively correlated as well as high blood pressure diagnosed by doctor and essential hypertension (r_g=0.11; P=0.01 and r_g=0.36; P=0.04, respectively). We also see marginal evidence for genetic correlation with acute myocardial infarction (AMI, r_g=0.21; P=0.02), syncope and collapse (r_g=0.23; P=0.03), angina pectoris (r_g=0.15; P=0.03) and heart attack diagnosed by doctor (r_g=0.14; P=0.05). Finally, in this paper, we show significant genetic correlation to coronary artery disease (CAD, r_g=0.11; P=0.05) which was reported earlier, but was not significant.[7] We further found evidence for correlation to typical aging disorders including, senile cataract (r_g=0.32; P=0.03), and chronic kidney disease (r_g=0.26; P=0.03), which is strongly associated to hypertension and cardiovascular disease.

Table 1. Significant (P<0.05) genetic correlation results of COPD with comorbidities

Comorbidity	Study PMID	Ethnicity	r_g	SE	h^2	P
Illnesses of siblings: Severe depression	UKBB	European	0.289	0.085	0.014	**0.0007**
Diabetes diagnosed by doctor	UKBB	European	0.171	0.056	0.043	**0.0021**
Attention deficit hyperactivity disorder	27663945	European	0.510	0.173	0.075	**0.0031**
Inflammatory Bowel Disease	26192919	European	0.137	0.055	0.321	**0.013**
High blood pressure diagnosed by doctor	UKBB	European	0.105	0.043	0.116	**0.014**
Crohns disease	26192919	European	0.133	0.056	0.493	**0.018**
Diagnoses - main ICD10: I21 Acute myocardial infarction	UKBB	European	0.207	0.089	0.010	**0.020**
Illnesses of siblings: Diabetes	UKBB	European	0.162	0.071	0.021	**0.022**
Illnesses of mother: Diabetes	UKBB	European	0.165	0.072	0.019	**0.022**
Diagnoses - main ICD10: H25 Senile cataract	UKBB	European	0.317	0.141	0.004	**0.025**
Chronic Kidney Disease	26831199	Mixed	0.255	0.114	0.019	**0.025**
Diagnoses - main ICD10: R55 Syncope and collapse	UKBB	European	0.230	0.108	0.006	**0.033**
Angina pectoris diagnosed by doctor	UKBB	European	0.153	0.072	0.022	**0.034**
Diagnoses - main ICD10: N92 Excessive frequent and irregular menstruation	UKBB	European	0.227	0.107	0.008	**0.035**
Diagnoses - main ICD10: N81 Female genital prolapse	UKBB	European	0.248	0.121	0.006	**0.041**
Diagnoses - main ICD10: I10 Essential (primary) hypertension	UKBB	European	0.361	0.179	0.003	**0.044**
Coronary artery disease	26343387	Mixed	0.110	0.056	0.079	**0.047**
Heart attack diagnosed by doctor	UKBB	European	0.144	0.072	0.019	**0.047**
Schizophrenia	25056061	Mixed	0.091	0.046	0.458	**0.049**
Illnesses of mother: Breast cancer	UKBB	European	0.177	0.090	0.010	**0.050**

Study PMID: PubMed ID for a given GWAS used in the analysis; UKBB: United Kingdom BioBank - unpublished GWAS results; r_g: genetic correlation coefficient; SE: standard error of r_g; h^2: SNP heritability; P: P-value.

Among the other diseases found to show marginal genetic correlation with COPD were inflammatory bowel disease (r_g=0.14; P=0.01), Crohn's disease (r_g=0.13; P=0.02), and female reproductive conditions including excessive frequent and irregular menstruation and genital prolapse (r_g=0.23; P=0.04 and r_g=0.25; P=0.04, respectively). Finally, we found that a breast cancer of mother was genetically correlated to COPD (r_g=0.18; P=0.05).

DISCUSSION

In this study focussing on the question whether there is a genetic overlap between COPD and extra pulmonary disorders, we find marginally significant evidence for genetic correlation of COPD with psychiatric, cardiovascular, inflammatory disorders of the bowel, age-related disorders such as cataract and female reproductive system disorders. Of note is that for three disorders (diabetes, severe depression and breast cancer) we find evidence for genetic correlation of COPD with family history rather than the co-occurrence of the disease in an individual.

COPD is genetically correlated to wide range of cardiovascular disorders (AMI, CAD and angina pectoris) and their risk factors (hypertension, diabetes, chronic kidney disease). The genetic correlation of COPD to cardiovascular disease is expected based on the findings that COPD and cardiovascular comorbidities may have common pathogenic mechanisms and that cardiovascular mortality accounts for 20-30% of deaths in COPD.[10,11]

Lung function has been found to be a better predictor of cardiovascular mortality than cholesterol.[12] Yet, previous study failed to show significant genetic overlap of COPD with cardiovascular disease.[7] Using larger genome wide association studies, we confirm the genetic overlap with cardiovascular disease and its risk factors. Of note is that diabetes shows the strongest genetic correlation in term of statistical significance. Not only diabetes of the person but also the family history of diabetes in siblings and the mother are found to be genetically correlated to COPD. Such correlations with family history strongly suggests a joint genetic aetiology. Although not the most significantly correlated, the strongest genetic correlation with COPD and cardiovascular pathology is seen for essential hypertension.

COPD is more strongly genetically correlated to neuropsychiatric pathology than to cardiovascular pathology, both in terms of statistical significance (family history of depression in a sibling) and strength of correlation (ADHD). These findings are both puzzling and intriguing, in particular for ADHD, which is not only the third most significantly associated disorder but shows an extremely strong association to COPD, a disorders with a typically late onset and a strong link to smoking. It is tempting to speculate that, as both disorders show a strong association to smoking, ADHD is also significantly genetically correlated to lung cancer. However, ADHD is also genetically correlated to diabetes and related traits, which may be another logical pathway.

The strong genetic correlation between COPD and family history of depression in a sibling is of interest since the prevalence of depression in COPD is high.[13,14] Also it has been suggested that relationship between depression and COPD is likely to

be bidirectional, in that COPD may increase the risk of depression and vice versa, depression may increase the risk of COPD.[15] However, we found no significant genetic overlap between depression and COPD in the same person which is difficult to explain. This is why we further tested the correlation of COPD and major depressive disorder using results of the largest to date GWAS on depression[17] and showed borderline significant correlation (r_g=0.25, P=0.058). This can mean that increasing sample sizes further we can reach enough power to detect true risk variants which play a role in the overlapping pathways. However, at this point we cannot exclude a false positive finding. The same holds for the genetic correlation of COPD to schizophrenia, which is a disease associated with many loci in the genome, which may generate false positive findings.

There are two other disorders for which we also only find a positive family history is genetically associated to COPD: diabetes and breast cancer in the mother. The genetic overlap with diabetes is very consistent for the person her/himself and the family history. For breast cancer the same problem occurs as with depression: no genetic correlation is found with the disease itself. These finding should be followed up in the near future. Finally, to our knowledge, we found for the first time evidence for a genetic correlation between COPD and cataract and inflammatory gastrointestinal disorders. Also this finding raise interesting hypothesis: the genetic correlation of COPD and cataract may be related to diabetes while the genetic correlation of COPD and inflammatory gastrointestinal disorders may be explained by common immune pathways.

Although we both confirmed long expected genetic correlations with cardiovascular disease and found new intriguing ones, this study is by no means an endpoint. Our finding provide new leads into the comorbidity research. There may be various explanations for the genetic correlation including the genetic localization on the same chromosome, pleiotropy, i.e., the protein encoded by a gene has a biological function in the lung but also in other tissues, shared pathogenesis which may be partial, and/or that one trait is in the causal pathway of the other.

REFERENCES

1. Lozano R, Naghavi M FK. Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet*. 2012;380(380):2095–2128. doi:https://doi.org/10.1016/S0140-6736(12)61728-0
2. GOLD. From the Global Strategy for the Diagnosis, Management and Prevention of COPD, Global Initiative for Chronic Obstructive Lung Disease. *Available from http// goldcopd org*. 2017. http://goldcopd.org.
3. Chatila WM, Thomashow BM, Minai OA, Criner GJ, Make BJ. Comorbidities in Chronic Obstructive Pulmonary Disease. *Proc Am Thorac Soc*. 2008;5(4):549-555. doi:10.1513/pats.200709-148ET
4. Divo M, Cote CG, De Torres JP, et al. Comorbidities and risk of mortality in patients with chronic obstructive pulmonary disease. *Am J Respir Crit Care Med*. 2012;186(2):155-161. doi:10.1164/rccm.201201-0034OC
5. Almagro P, Cabrera FJ, Diez-Manglano J, et al. Comorbidome and short-term prognosis in hospitalised COPD patients: The ESMI study. *Eur Respir J*. 2015;46(3):850-853. doi:10.1183/09031936.00008015
6. Barnes PJ, Celli BR. Systemic manifestations and comorbidities. *Eur Respir J*. 2009;33(5):1165-1185. doi:10.1183/09031936.00128008
7. Hobbs BD, De Jong K, Lamontagne M, et al. Genetic loci associated with chronic obstructive pulmonary disease overlap with loci for lung function and pulmonary fibrosis. *Nat Genet*. 2017;49(3):426-432. doi:10.1038/ng.3752
8. González JR, Cáceres A, Esko T, et al. A common 16p11.2 inversion underlies the joint susceptibility to asthma and obesity. *Am J Hum Genet*. 2014;94(3):361-372. doi:10.1016/j.ajhg.2014.01.015
9. Zheng J, Erzurumluoglu AM, Elsworth BL, et al. LD Hub: a centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. *Bioinformatics*. 2017;33(2):272-279. doi:10.1093/bioinformatics/btw613
10. Camilli AE, Robbins DR, Lebowitz MD. Death Certificate Reporting of Confirmed Airways Obstructive Disease. *Am J Epidemiol*. 1991;133(8):795-800. doi:10.1093/oxfordjournals.aje.a115958
11. Huiart L, Ernst P, Suissa S. Cardiovascular Morbidity and Mortality in COPD. *Chest*. 2005;128(4):2640-2646. doi:10.1378/chest.128.4.2640
12. Hole DJ, Watt GC, Davey-Smith G, Hart CL, Gillis CR, Hawthorne VM. Impaired lung function and mortality risk in men and women: findings from the Renfrew and Paisley prospective population study. *BMJ*. 1996;313(7059):711-715; discussion 715-6. http://www.ncbi.nlm.nih.gov/pubmed/8819439. Accessed March 4, 2019.
13. Hanania NA, Müllerova H, Locantore NW, et al. Determinants of depression in the ECLIPSE chronic obstructive pulmonary disease cohort. *Am J Respir Crit Care Med*. 2011;183(5):604-611. doi:10.1164/rccm.201003-0472OC
14. Dodd JW, Getov S V., Jones PW. Cognitive function in COPD. *Eur Respir J*. 2010;35(4):913-922. doi:10.1183/09031936.00125109
15. Atlantis E, Fahey P, Cochrane B, Smith S. Bidirectional Associations Between Clinically Relevant Depression or Anxiety and COPD. *Chest*. 2013;144(3):766-777. doi:10.1378/chest.12-1911

16. Nobakht M Gh BF, Aliannejad R, Rezaei-Tavirani M, Taheri S, Oskouie AA. The metabolomics of airway diseases, including COPD, asthma and cystic fibrosis. *Biomarkers Biochem Indic Expo response, susceptibility to Chem*. 2015;20(1):5-16. doi:10.3109/1354750X.2014.983167
17. Wray NR, Ripke S, Mattheisen M, et al. Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nat Genet*. 2018;50(5):668-681. doi:10.1038/s41588-018-0090-3

Chapter 3.4

DNA methylation signatures of depressive symptoms in middle-aged and elderly persons: meta-analysis of multi-ethnic epigenome-wide studies

Olivera Story Jovanova*, Ivana Nedeljković*, Derek Spieler*, Rosie M. Walker*, Chunyu Liu*, Michelle Luciano*, Jan Bressler, Jennifer Brody, Amanda J. Drake, Kathryn L. Evans, Rahul Gondalia, Sonja Kunze, Brigitte Kuhnel, Jari Lahti, Rozenn N. Lemaitre, Riccardo E. Marioni, Brenton Swenson, Jayandra Jung Himali, Hongsheng Wu, Yun Li, Allan F. McRae, Tom C. Russ, James Stewart, Zhiying Wang, Guosheng Zhang, Karl-Heinz Ladwig, Andre G. Uitterlinden, Xiuqing Guo, Annette Peters, Katri Räikkönen, John M. Starr, Melanie Waldenberger, Naomi R. Wray, Eric A. Whitsel, Nona Sotoodehnia, Sudha Seshadri, David J. Porteous, Joyce van Meurs, Thomas H. Mosley, Andrew M. McIntosh, Michael M. Mendelson, Daniel Levy, Lifang Hou, Johan G. Eriksson, Myriam Fornage, Ian J. Deary, Andrea Baccarelli, Henning Tiemeier, Najaf Amin.

**first authors contributed equally*

This chapter is published in the Journal of the American Medical Association (JAMA) Psychiatry, July 2018.

The supplemental material for this paper is available at: archpsyc.jamanetwork.com/article.aspx?doi=10.1001/jamapsychiatry.2018.1725

ABSTRACT

Importance: Depressive disorders arise from a combination of genetic and environmental risk factors, however, the pathophysiology and underlying molecular events leading to depression remain elusive. Epigenetic disruption provides a plausible mechanism through which gene-environment interactions lead to depression. Large-scale epigenome-wide studies on depression are missing, hampering the identification of potentially modifiable biomarkers. *Objective*: To identify robust epigenetic mechanisms underlying depression in middle-aged and elderly persons using DNA methylation in blood. *Design*: We performed the first cross-ethnic meta-analysis of epigenome-wide association studies (EWAS) within the framework of the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium. *Setting*: Discovery EWAS was performed in nine population-based cohorts. Results of the EWAS from all cohorts were pooled using sample-size weighted meta-analysis. Replication of the top epigenetic sites from the discovery stage was performed in two independent population-based cohorts. *Participants*: The discovery sample included 7,948 individuals of European origin and the replication sample included 3,308 individuals of African-American and European origin. Only participants that were assessed for both depressive symptoms and whole blood DNA methylation were included in the study. *Outcome*: Whole blood DNA methylation levels were assayed with Illumina-Infinium Human Methylation 450K BeadChip and depressive symptoms were assessed by questionnaire. *Results*: The discovery cohorts consisted of 7,948 individuals (48% female) with a mean age of 65.4 (SD=5.8) years. The replication cohort consisted of 3,308 individuals (74% female) with a mean age of 60.3 (SD=6.4) years. The EWAS identified methylation of three CpG sites including cg04987734 (*p-value*=1.57×10^{-8}, *n*=11256, *CDC42BPB* gene), cg12325605 (*p-value*=5.24×10^{-9}, *n*=11256, *ARHGEF3* gene) and an intergenic CpG site cg14023999 (*p-value*=5.99×10^{-8}, *n*=11256, chromosome=15q26.1) significantly associated with increased depressive symptoms. The predicted expression of *CDC42BPB* and *ARHGEF3* was significantly associated with major depression in brain and fibroblasts, respectively. *Conclusion*: We report the first robustly associated methylated sites for depressive symptoms. All three findings point towards axon guidance as the common disrupted pathway in depression. Our findings provide new insights into the molecular mechanisms underlying the complex pathophysiology of depression. Further research is warranted to determine the utility of these findings as biomarkers of depression and evaluate any potential role in the pathophysiology of depression and their downstream clinical effects.

INTRODUCTION

Depression is one of the most common mental health disorders that is projected to play a leading role in disease burden by the year 2030.[1] In later life, depression is associated with disability, increased mortality, dementia and poor outcomes from physical illness.[2] Further, more people aged over 65 years commit suicide than in any other age group, and most have major depression.[3] Limited understanding of the molecular mechanisms underlying depression is a major bottleneck in the development of innovative treatment, prognostic markers, and prevention strategies.

Studying depression is challenging, as it is a heterogeneous disorder with a multifactorial etiology.[4] This heterogeneity increases with age as the incidence of chronic diseases and disability rises. The contribution of genetics to the risk of depression is moderate with heritability estimates ranging from 40 to 50%[5] and modest (18%) in the elderly.[6] Genome-wide association studies (GWAS) have recently identified numerous rare and common genetic variants associated with depression and related traits.[7-10] However, genetic variation alone does not completely explain an individual's risk for developing depression. Among environmental factors, adverse life-events and stress are major risk factors for depression.[11] Converging evidence from animal and human studies suggest that psychosocial stressors trigger depression onset by inducing elevations in pro-inflammatory cytokines.[12] These psychosocial stressors are also known to influence epigenetic mechanisms, such as DNA methylation[13] that can drive sustained changes in gene expression.[14] The high contribution of environmental factors to depression in the elderly makes DNA methylation an interesting candidate mechanism for studies of the molecular basis of late-life depression.

DNA methylation may be global or tissue-specific.[15] Tissues likely to be involved in complex psychiatric disorders, such as brain, are not directly accessible from living patients. The, use of post-mortem brain tissue to study DNA methylation is a possible solution, although obtaining a sufficient sample size is challenging.[16] To study differential DNA methylation associated with mental health symptoms on a large scale, peripheral tissues such as blood constitutes a useful proxy for detecting trans-tissue changes and the most appropriate tissue for biomarkers.[16,17] Moderate correlation has been demonstrated between blood and brain tissues at non-tissue specific regulatory regions across the methylome.[18] To date, several studies have assessed the correlation between depression and blood DNA methylation.[19,20] However, these studies have been limited to a small number of DNA methylation sites (CpGs) and/or small samples. For instance, the largest published epigenome-wide association study (EWAS) assessed brain DNA methylation in 76 cases persons who

died during a depressive episode and 45 controls.[21] Moreover, these studies compared depressed cases with healthy controls. Depression represents an arbitrarily selected extreme of the continuum of varying severity and duration,[22] whereas a broad phenotype approach can be more representative for the general population. In a large study consisting of 252,503 individuals from 68 countries showed that sub-threshold depressive disorders produce significant decrements in health and do not qualitatively differ from full-blown episodes of depression.[23] A meta-analysis in individuals aged over 55 found two to three times higher prevalence of sub-threshold depressive symptomology than major depression.[24] Use of rating scales have therefore been recommended for the assessment of depressive problems in the elderly.[2]

In the current study, we performed EWAS of depressive symptoms using whole blood samples of 7,948 individuals of European ethnicity from the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) consortium. We replicated our findings in 3,308 individuals of African-American and European ancestry. Finally, we used publicly available expression quantitative methylation (eQTM) loci and expression quantitative loci (eQTL) databases to identify the downstream effects of the associated methylation loci.

MATERIALS AND METHODS

Study population

The study sample for the discovery analysis included a total of 7,948 participants of European ancestry from nine population-based cohorts of the CHARGE consortium (**Table 1**): Cardiovascular Health Study (CHS),[25] Framingham Heart Study (FHS),[26] Helsinki Birth Cohort Study (HBCS),[27] Cooperative Health Research in the Augsburg Region (KORA) study,[28] two sub-cohorts from Lothian Birth-Cohort born in 1921 (LBC1921)[29] and 1936 (LBC1936),[30] two sub-cohorts from Rotterdam Study (RS-III and RS-BIOS)[31] and Generation Scotland: Scottish Family Health Study (GS) study.[32] These cohorts included community dwelling individuals, who were not selected based on disease status. Informed consent was obtained from all participants. The same cohorts have been successfully used to identify differentially methylated sites associated with cognitive traits,[33] inflammation[34] and smoking.[35] The protocol for each study was approved by the institutional review board of each institution.

The replication sample included 3,308 participants, largely of African American origin from the Atherosclerosis Risk in Communities Study (ARIC)[36] and European origin from the Women's Health Initiative - Epigenetic Mechanisms of PM-Mediated

Cardiovascular disease (WHI-EMPC) that joined the consortium later for the replication phase of the study.[37] Detailed information for each cohort is provided in the **Supplementary Text**.

Depressive symptoms assessment

Depressive symptoms were measured using self-reported questionnaires or structured interview performed by a trained researcher, psychologist, or psychiatrist at the same time point when blood samples were obtained for DNA methylation quantification (**Table 1**). Four cohorts (FHS, HBCS, RS-III, and RS-BIOS) assessed depressive symptoms using the 20-item Centre for Epidemiologic Studies Depression (CES-D) scale,[38] while CHS used the 10-item CES-D scale. Participants could score from zero to 60 (or 30 for CHS) points, where higher scores suggest more depressive symptoms. WHI-EMPC used a cohort specific CES-D/DIS screening instrument, which is described in detail in the **Supplementary Text**. The LBC1921 and LBC1936 assessed self-reported depressive symptoms using the Hospital Anxiety and Depression Scale-depression subscale (HADS-D),[39] which consists of seven items. Participants could score from zero to 21. The KORA study used the self-administered Patient Health Questionnaire (PHQ-9)[40] representing a depression module that scores each of the nine Diagnostic and Statistical Manual of Mental Disorders, 4th Edition (DSM-IV) criteria for depression from zero to three. The GS study assessed life-time history of depression using the Structured Clinical Interview for DSM-IV Disorders (SCID).[41] The ARIC study assessed depressive symptoms using the 21-item Maastricht Questionnaire (21-MQ). In all cohorts, depressive symptoms were analyzed as continuous variable except for GS, which studied depression status as binary trait.

DNA methylation sample and measurement

In all cohorts, DNA was extracted from whole blood and methylation levels were assessed using the Illumina-Infinium Human Methylation 450K BeadChip (Illumina Inc., San Diego, CA, USA) using standard manufacturer's protocols. The 450K array includes more than 450,000 CpGs and is enriched for genic regions, covering 99% of all genes. DNA methylation data pre-processing, including quality control (QC) and normalization, was conducted per cohort using study-specific methods. In all cohorts, DNA methylation levels were quantified as β-values, which range from zero to one, and indicate the proportion of DNA strands in a sample methylated at a specific CpG. Detailed information about cohort specific DNA extraction, bisulfite conversion, DNA methylation profiling, normalization and QC is described in detail in the **Supplementary Text**.

Statistical analysis

Epigenome-wide association analysis

In all cohorts, the association between depressive symptoms and CpG sites was assessed using linear regression analysis in the R software. In the regression analysis, DNA methylation β-value at each CpG site was specified as the dependent variable and the depressive symptoms/depression as the predictor of interest. Association analysis was adjusted for age,[42] sex,[43] smoking[35] (assessed at the time of blood sampling for methylation), methylation batch effects, white blood cell composition (imputed or directly measured), principal components estimated using genome-wide genotype data to control for population stratification and familial relationships when required. Cohort specific details of these analyses are provided in the **Supplementary Text**. Further, sensitivity analysis was performed by adjusting the initial model for antidepressant medication use at the time of sample collection.

To pool the results from independent studies we performed sample-size weighted meta-analysis in METAL.[44] We chose the 'sample-size weighted' method because of the differences in the measurement scales of depressive symptoms across studies. A drawback of using sample-size weighted method is that no pooled effect estimates are generated. To obtain pooled effect estimates we additionally performed inverse-variance weighted meta-analysis for the top sites in cohorts that used CES-D 20 item scale for the assessment of depressive symptoms. CpG sites missing in more than three of the participating cohorts were removed. In total, 484,516 probes were tested for association, giving a Bonferroni-corrected genome-wide significance threshold of $0.05/484{,}516 = 1.03\times10^{-7}$. All CpG sites suggestive of association ($p\text{-}value \leq 10^{-5}$) were tested for association in the independent replication cohorts using the same model as used in the discovery EWAS. Finally, a sample size weighted meta-analysis was performed for all cohorts included in the discovery and replication phases in METAL. To evaluate the contribution of each study to the association results we generated posterior probabilities of the effects in each study (M-values) using the METASOFT package.[45] M-value and Forest plots for z-scores were generated using custom-made scripts in R. For annotating CpG sites we used the annotation provided by Illumina and the UCSC database (GRCh37/hg19).

Gene expression analyses

To evaluate the downstream effects of the three identified CpG sites in blood we used the BIOS database to search for eQTM.[46] To evaluate whether the expression of the genes associated with or harbored the significant methylation site is associated with major depression (also smoking and inflammation to check specificity) we

used the MetaXcan package.[47,48] MetaXcan associates the expression of the genes with the phenotype by integrating functional data generated by large-scale efforts, e.g., Genotype-Tissue Expression (GTEx) with that of the GWAS. MetaXcan is trained on transcriptome models in 44 human tissues from GTEx and is able to estimate their tissue-specific effect on phenotypes from GWAS. We used the GTEx-V6p-HapMap-2016-09-08 database and the publicly available GWAS datasets of major depression,[49] and C-reactive protein[50] and smoking,[51] which represent important potential confounders in the present study.

Causal inference analysis

To help infer causal relationships, we studied the *cis*-SNPs identified by the BIOS consortium[46] as instrumental variables for the CpG sites as proposed by Smith et al.[52] We checked the association of these cis-SNPs with depression, smoking and inflammation in the published GWAS of these traits. Similarly, we checked whether the single nucleotide polymorphisms (SNPs) associated with inflammation (CRP levels),[50] smoking[51] and depression[7] were associated with the identified CPG sites using the BIOS consortium database. We chose smoking and inflammation as these are highly correlated with both depression and DNA methylation and thus could influence the relationship between depression and DNA methylation.

RESULTS

The mean age in the discovery cohorts ranged from 52.4 years (SD=8.1) in GS to 79.1 years (SD=0.57) in LBC1921. Forty-eight percent of the total discovery sample were female. The replication cohort comprised 74% women and had an average age of 60.3 years (SD=6.4) (**Table 1**).

Epigenome-wide association analysis

In the meta-analysis of depressive symptoms of European ancestry, we identified one CpG site on chromosome 14q32.32 (cg04987734, *CDC42BPB*, *p-value*=4.93×10^{-8}, *n*=7948) that passed the Bonferroni threshold for significance (**Table 2, eFigure 1**). Further, suggestive association was observed at 19 additional CpG sites (**Table 2**).

Adjusting for anti-depressive medication use did not meaningfully change the results (**eTable 1**). No inflation in the test statistic was observed (Lambda=1.03, **eFigure 2**). We tested all 20 CpG sites for association in the replication sample. The top CpG site from the discovery (cg04987734) showed nominal association (*p-value*<0.05, *n*=3308) with depressive symptoms in the validation data set (**Table 2**).

Table 1. Descriptive statistics of the discovery and replication cohorts.

Study	Ethnicity	N	Female (%)	Mean Age (SD)	Current smokers (%)	Depressive symptoms	Antidepressant medication use (%)
Discovery (N = 7948)							
CHS	European	323	194 (60.1)	75.6 (5.2)	173 (53.6)	CESD[1] (10 item)	19 (5.9)
FHS	European	2722	1508 (53.6)	58.5 (11.6)	948 (34.8)	CESD[2] (20 item)	251 (16.1)
HBCS	European	122	0 (0)	65.2 (2.7)	24 (19.7)	CESD[2] (20 item)	11 (9.0)
KORA	European	1727	882 (51.1)	61.0 (8.9)	250 (14.5)	PHQ-9[3]	82 (4.7)
LBC 1921	European	432	261 (60.4)	79.1 (0.6)	194 (44.9)	HADS[4]	15 (3.5)
LBC 1936	European	916	452 (49.3)	69.6 (0.8)	504 (55)	HADS[4]	30 (3.3)
RS III	European	722	391 (54.2)	59.8 (8.1)	167 (23.1)	CESD[2] (20 item)	38 (5.3)
RS BIOS	European	757	319 (42.1)	67.6 (5.9)	78 (10.3)	CESD[2] (20 item)	51 (6.7)
GS[a]	European	227	151 (64.5)	52.4 (8.1)	46 (19.7)	SCID[5]	44 (18.8)
Total		7948	4158 (48.4)	65.4 (5.8)	2384 (30.6)	-	541 (8.1)
Replication (N = 3308)							
ARIC	African	2297	1445 (63)	56.1 (5.7)	584 (25.4)	21-MQ[6]	74 (3.3)
WHI-EMPC	European	1011	1011 (100)	64.6 (7.1)	509 (50.3)	CES-D/DIS[7]	61 (6.0)
Total		3308	2456 (74.2)	60.3 (6.4)	1093 (37.9)	-	135 (4.7)

Characteristics are depicted as mean (SD), unless otherwise specified. *CHC* Cardiovascular health cohort, *FHS* Framingham Heart Study, *HBCS* Helsinki Birth Cohort Study, *KORA* Cooperative Health Research in the Augsburg Region, *LBC* Lothian Birth Cohort, *RS* Rotterdam Study, *GS* Generation Scotland Study, and [a] CASE-CONTROL STUDY, *ARIC* Atherosclerosis Risk in Communities Study and *WHI-EMPC* the Women's Health Initiative - Epigenetic Mechanisms of PM-Mediated Cardiovascular disease; "(in brackets we state number of item of the questionnaires)". 1 Irwin, M. et al. 1999. [2] Radloff, LS. et al. 1977. [3] Kroenke, K. et al. 2001. [4] Zigmond, AS. et al. 1983. [5] First, MB. et al. 1996. [6] Wattanakit,K. et al. 2005. [7] Burnam, MA. et al. 1988.

In addition, significant association of a CpG site (cg12325605; *p-value*=9.17×10^{-05}, *n*=3308, **Table 2**) annotated to the *ARHGEF3* gene with depressive symptoms was observed in the replication sample.

Meta-analysis of discovery and replication cohorts showed a significant association of both cg04987734 (*p-value*=1.57×10$^{-08,}$ *n*=11256) and cg12325605 (*p-value*=5.24×10^{-09}, *n*=11256) with depressive symptoms (**Table 2; Figures 1 and 2**).

Also, an additional intergenic CpG site (cg14023999; *p-value*=5.99×10^{-08}, *n*=11256) at chromosome 15q26.1 locus showed genome-wide significant association with depressive symptoms (**eTable 2, eFigures 3 and 4**). The independent contributions of each cohort to the association signals of the three CpG are depicted in **eFigure 5** and also provided in **eTable 3**. For all three CpG sites the association signals were not driven by a single cohort but appeared to be linearly related to the sample size, sug-

Table 2. Top DNA methylation sites associated with depressive symptoms in the discovery EWAS.

				Discovery (N=7948)	Replication (N=3308)	Meta-analysis (N=11256)
CpG site ID	**Chr**	**Location**	**Gene symbol**	**P-value**	**P-value**	**P-value**
cg04987734	14	103415873	*CDC42BPB*	**4.93×10^{-8}**	4.82×10^{-02}	**1.57×10^{-08}**
cg07012687	17	80195180	*SLC16A3*	3.47×10^{-7}	1.58×10^{-01}	4.45×10^{-06}
cg08796240	16	70733832	*VAC14*	7.43×10^{-7}	2.56×10^{-01}	1.80×10^{-06}
cg06096336	2	231989800	*PSMD1; HTR2B*	8.06×10^{-7}	3.01×10^{-01}	2.51×10^{-06}
cg16745930	10	100220809	*HPSE2*	1.34×10^{-6}	4.01×10^{-01}	6.26×10^{-06}
cg09849319	5	1494983	*LPCATI1*	1.81×10^{-6}	4.64×10^{-01}	1.04×10^{-04}
cg17237086	22	40814966	*MKL1*	3.44×10^{-6}	2.51×10^{-01}	6.10×10^{-06}
cg03985718	2	105924245	*TGFBRAP1*	3.61×10^{-6}	8.54×10^{-01}	6.53×10^{-05}
cg21098005	20	44538605	*PLTP*	4.36×10^{-6}	9.60×10^{-01}	1.01×10^{-04}
cg16466652	19	6271960	*MLLT1*	4.39×10^{-6}	3.97×10^{-01}	1.57×10^{-05}
cg07884764	11	64107517	*CCDC88B*	5.03×10^{-6}	9.99×10^{-01}	1.25×10^{-04}
cg01541347	7	4729920	*FOXK1*	5.64×10^{-6}	3.77×10^{-01}	8.46×10^{-04}
cg02341197	21	34185927	*C21or62*	5.84×10^{-6}	2.02×10^{-01}	6.80×10^{-06}
cg01947751	3	196728969	-	6.23×10^{-6}	6.63×10^{-01}	3.68×10^{-04}
cg13747876	17	80195402	*SLC16A3*	6.32×10^{-6}	1.04×10^{-01}	2.93×10^{-06}
cg12764201	1	105101123	*CORT; APITD1*	7.15×10^{-6}	7.20×10^{-01}	7.29×10^{-05}
cg08295111	5	133866097	*PHF15*	7.87×10^{-6}	5.76×10^{-01}	5.64×10^{-04}
cg18030453	3	45506216	*LARS2*	9.16×10^{-6}	3.87×10^{-03}	1.20×10^{-07}
cg12325605	3	56810151	*ARHGEF3*	9.62×10^{-6}	**9.17×10^{-05}**	**5.24×10^{-09}**
cg23282441	10	73533927	*C10orf54; CDH23*	9.69×10^{-6}	1.77×10^{-01}	8.63×10^{-06}

gesting stronger association in larger studies (**eFigure 5**). Pooled effect estimates in cohorts that used CES-D scale suggest that a 1-unit increase in CES-D score increases methylation by 0.05% at cg04987734, 0.04% at cg12325605, and 0.03% at cg14023999.

Gene expression analyses

Cg04987734 was significantly associated with increased expression of CDC42BPB gene (FDR *p-value*=7.7×10^{-04}, *n*=2101) and cg14023999 was significantly associated with decreased expression of SEMA4B (FDR *p-value*=4.7×10^{-03}, *n*=2101) in blood (**eTable 4**). No significantly associated gene expression probes were identified for cg12325605 in blood. Further, the predicted expression of *CDC42BPB* gene in the brain (basal ganglia) (effect=0.14, *p-value*=2.7×10^{-03}) and of *ARHGEF3* in fibroblasts (effect=-0.48, *p-value*=9.8×10^{-04}) was associated with major depression (**eTable 5**). No association was observed with either smoking or inflammation.

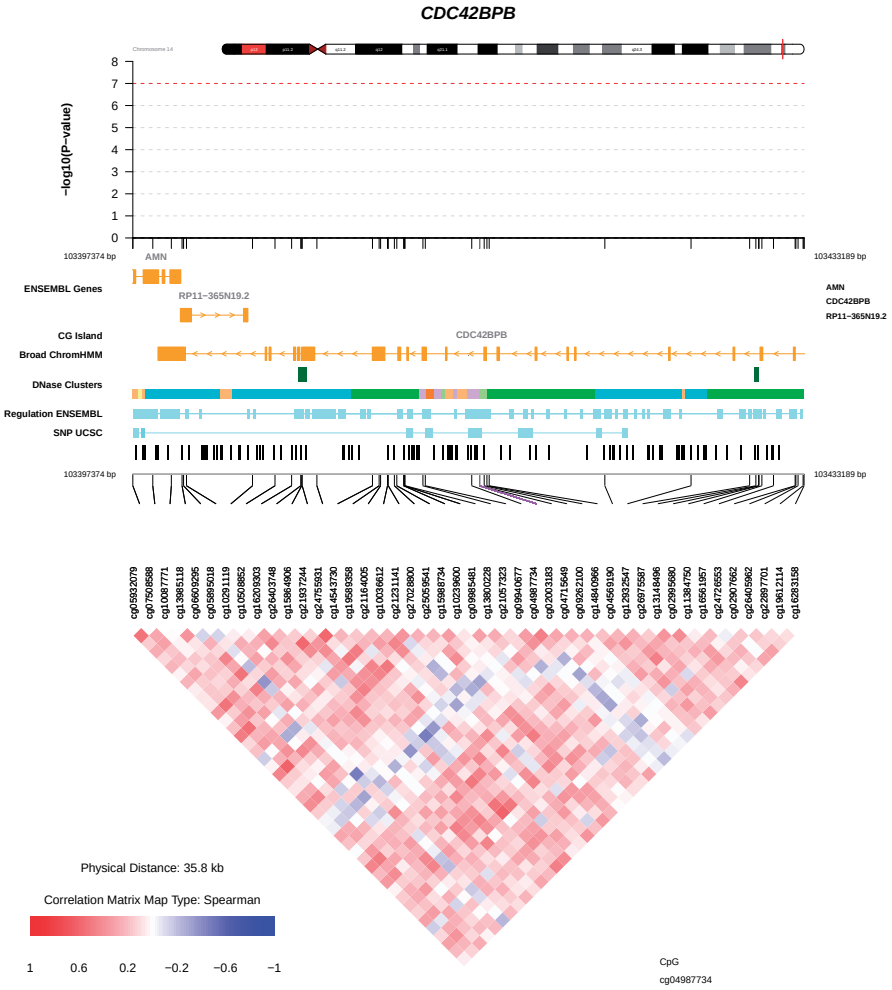

Figure 1: Regional association plot for the top CpG site cg04987734. The horizontal axis depicts the position in base pair (hg19) for the entire *CDC42BPB* gene region. The Vertical axis indicates the strength of association in terms of negative logarithm of the association *p*-value. Each circle represents CpG site. Red dashed line indicates the genome-wide significance threshold. Below the horizontal axis the figure shows the regulatory information and correlation matrix of other CpG sites in the region with the top hit. Color intensity marks the strength of the correlation and color the direction of the correlation. Figure is made using web-based plotting tool and R-based package "CoMET" (http://comet.epigen.kcl.ac.uk:3838/coMET1/).

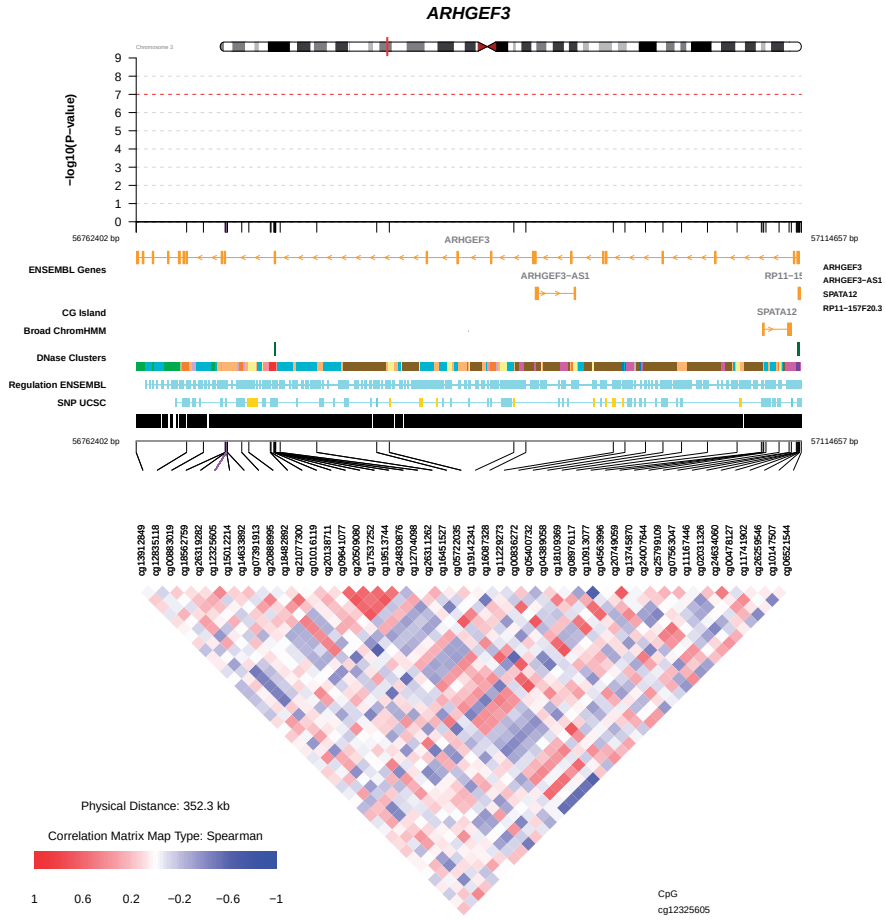

Figure 2: Regional association plot for the top CpG site cg12325605. The horizontal axis depicts the position in base pair (hg19) for the entire *ARHGEF3* gene region. The Vertical axis indicates the strength of association in terms of negative logarithm of the association *p*-value. Each circle represents CpG site. Red dashed line indicates the genome-wide significance threshold. Below the horizontal axis the figure shows the regulatory information and correlation matrix of other CpG sites in the region with the top hit. Color intensity marks the strength of the correlation and color the direction of the correlation. Figure is made using web-based plotting tool and R-based package "CoMET" (http://comet.epigen.kcl.ac.uk:3838/coMET1/).

Blood and brain correlation

We checked the correlation between methylation in blood and various brain regions at the three identified sites using a web-based tools, BECon[18] and a blood brain DNA methylation comparison tool (http://epigenetics.essex.ac.uk/bloodbrain/). BECon showed strong correlation between blood and brain DNA methylation, e.g. methylation at cg04987734 in the CDC42BPB gene was highly correlated (r=0.81) between blood and the Brodmann area 7 that spans the medial and lateral walls of the parietal cortex (**eFigure 6**). Methylation at the other two sites was negatively correlated with methylation in the Brodmann area 10 than spans anterior prefrontal cortex (cg12325605, r=-0.39; cg14023999, r=-0.42) suggesting strong but reverse methylation patterns in blood and brain (**eFigures 7 and 8**). However, the blood brain DNA methylation comparison tool that compares DNA methylation between blood and prefrontal cortex, entorhinal cortex, superior temporal gyrus and cerebellum, showed only modest correlations. For instance, methylation in blood at cg04987734 showed the strongest correlation with methylation in superior temporal gyrus (r = 0.18; http://epigenetics.essex.ac.uk/bloodbrain/?probe nameg=cg04987734), while methylation in blood at cg12325605 (http://epigenetics .essex.ac.uk/bloodbrain/?probenameg=cg12325605) and cg14023999 (http:// epigenetics.essex.ac.uk/bloodbrain/?probenameg=cg14023999) showed strongest correlation with methylation in cerebellum (r = 0.16 and 0.19 respectively). Nevertheless, the findings from the two databases suggest some degree of correlation between methylation in blood and methylation in brain for the three identified CpG sites.

Causal inference

In the BIOS database we identified two cis-SNPs for cg04987734 and 4 cis-SNPs for cg12325605 (**eTable 6**) and none for cg14023999. We took the most significant cis-SNP as the proxy for the CpG sites if available. For cg04987734 we used rs751837 as a proxy and for cg12325605 we used rs3821412 as a proxy (top cis-SNP rs9880418 was not available in the GWAS of depression, smoking or inflammation). Rs751837 was suggestively associated with major depression (p-value=0.07; albeit in opposite direction) (**eTable 7**). Rs3821412 was not associated with any of the three tested phenotypes. None of the SNPs associated with depression, inflammation or smoking was associated with any of the three CpG sites.

DISCUSSION

In this large-scale EWAS of depressive symptoms, we identified methylation at three CpG sites (cg04987734, cg12325605 and cg14023999) associated with depressive symptoms in the middle-aged and elderly persons. Cg04987734 is annotated to the *CDC42BPB* gene, cg12325605 to the *ARHGEF3* gene, and cg14023999 lies in an intergenic region on chromosome 15q26.1 locus. The predicted expression of *CDC42BPB* and *ARHGEF3* genes associate with major depression in brain and fibroblasts respectively.

CDC42BPB (CDC42 Binding Protein Kinase Beta) encodes a member of the serine/threonine protein kinase family, which is an important downstream effector of CDC42 and plays a role in the regulation of cytoskeleton reorganization, cell migration and regulation of neurite outgrowth.[53] CDC42BPB is highly expressed in the brain https://www.proteinatlas.org/ENSG00000198752-CDC42BPB/tissue. Hyper-methylation of cg04987734 has been associated with increased expression of CDC42BPB in blood.[46] Interestingly, methylation levels at this CpG site (cg04987734) in *CDC42BPB* gene were also previously associated with C-reactive protein (CRP) levels in blood;[34] and smoking.[35] In our study, however, we adjusted for smoking in the regression model; therefore, the association between depression and DNA methylation of this CpG site may be independent of smoking habits. Also, our causal inference analyses provide no support for the possibility that smoking, or inflammation explained the observed association with depressive symptoms nor the predicted expression of the gene showed an association with smoking or inflammation.

ARHGEF3 encodes for Rho Guanine Nucleotide Exchange Factor 3 protein. The gene is highly expressed (https://www.proteinatlas.org/ENSG00000163947-ARHGEF3/tissue) in adrenal glands, brain and uterus. Both *ARHGEF3* and *CDC42BPB* are co-expressed with several members of the Rho subfamily (RHOA, RHOB and RHOC; **eFigures 9 and 10**) of the Rho GTPase family that also includes CDC42.[54] The Rho family of GTPases is a family of small signaling G proteins involved in p75 neurotrophin receptor (p75NTR)-mediated signaling[55] and semaphorin signaling pathways.[56] P75NTR is a transmembrane receptor for neurotrophic factors of the neurotrophin family, which includes the brain-derived neurotrophic factor (BDNF).[57] P75NTR is widely expressed in the developing central and peripheral nervous system during the period of synaptogenesis and developmental cell death.[58] Both p75NTR and semaphorins are implicated in axon guidance.[59,60] In this context, the third associated CpG site cg14023999 that lies in an intergenic region on chromosome 15q26.1 is also interesting. Cg14023999 is associated with decreased expression of *SEMA4B* gene in blood. *SEMA4B* encodes for Semaphorin 4B protein. Sema4B is

believed to function through a direct interaction with post-synaptic density protein PSD-95[61] to promote synapse maturation.[61-63] The knock-down of Sema4B causes a decrease in GABAergic synapse number[62] suggesting a role in the assembly of excitatory and inhibitory postsynaptic specializations.[63] Previously cg14023999 was found to be significantly correlated with Parkinson's disease[64] and significant association of a CpG site in *SEMA4B* was observed in individuals with schizophrenia carrying the 22q11.2 deletion.[65] These findings point towards a functional of *SEMA4B* in neuro-psychiatric disorders. When comparing our findings with that of the previous EWAS of depression, we did not find an overlap. These studies were small (<100 individuals) and did not report reproducible results.[20]

To summarize, we report the first EWAS of depressive symptoms. We identified and replicated association of two methylation sites in the genome with depressive symptoms. A third site was identified in the meta-analysis of discovery and replication cohorts, which requires further replication. All three findings point towards axon guidance as the common disrupted pathway in depression (http://www.genome.jp/kegg/pathway/hsa/hsa 04360.html). Our findings provide new insights into the molecular mechanisms underlying the complex pathophysiology of depression.

Strengths and limitations of the study

This is the largest epigenome-wide study of depressive symptoms reported to date. Our major strength is the sample size that enabled detection of a replicable epigenome-wide significant locus, which suggests that in blood, DNA methylation signatures associated with depression may be subtle and will require large samples to be detected. Using peripheral blood tissue for DNA methylation profiling is a limitation of this study, as DNA methylation is known to be tissue specific.[66] While peripheral blood is not considered to be the most relevant tissue for the pathophysiology of depression, some sites show correlated methylation profiles between-tissues.[15,66] The three sites identified in our study show some degree of correlation between methylation in blood and various brain regions. Second, while replication in African-American samples suggests that some depressive symptoms related differences in DNA methylation may be similar across ethnicities,[67-69] it may also have resulted in false negatives due to different genetic background. Third, in these analyses we mostly used quantitative measures of depressive symptoms. Quantitative endo-phenotypes provide powerful alternatives for several complex outcomes, for example, hypertension.[70] This is likely to be especially true for a trait such as depressive symptoms, for which the severity and duration of illness can be highly heterogeneous.[22] Genome-wide studies of depressive traits, using quantita-

tive endo-phenotypes, have been suggested to improve statistical power.[22] However, the use of different phenotypic measures by different cohorts means that there may be some loss of statistical power due to the heterogeneity in the phenotype assessment. Nevertheless, the top three sites in our study were robustly associated with depressive symptoms independent of the depressive symptom measure used. Fourth, although we adjusted for potential confounders, the possibility of residual confounding cannot be excluded. Antidepressant medication indicates treated depression but itself may result in epigenetic modifications involved in depression pathophysiology.[71] Antidepressants can thus mediate or confound the relation between DNA methylation and depression. However, in sensitivity analysis additionally adjusted for antidepressant medication, our results did not change. Fifth, most cohorts included in this EWAS are cohorts of elderly persons. The aetiology of depression is more heterogeneous in elderly people than in younger and often hidden behind somatic symptoms, either because of somatization of the disorder or because of accentuation of symptoms of concomitant illness.[72] This may affect the generalizability of the results to younger populations. Finally, we made an attempt to disentangle cause and consequence using SNPs associated with the identified CpG sites and depression, inflammation and smoking as instrumental variables. The results did not support a causative role yet the association of the predicted gene expression of CDC42BPB in brain and ARHGEF3 in fibroblasts with major depression does suggest a possible causal role of the regulatory effects of these genes.[47]

REFERENCES

1. Lepine JP, Briley M. The increasing burden of depression. *Neuropsychiatr Dis Treat.* 2011; 7(Suppl 1):3-7.
2. Rodda J, Walker Z, Carter J. Depression in older adults. *BMJ.* 2011;343:d5219.
3. Conwell Y, Duberstein PR, Cox C, Herrmann JH, Forbes NT, Caine ED. Relationships of age and axis I diagnoses in victims of completed suicide: a psychological autopsy study. *Am J Psychiatry.* 1996;153(8):1001-1008.
4. Hidaka BH. Depression as a disease of modernity: explanations for increasing prevalence. *J Affect Disord.* 2012;140(3):205-214.
5. Lohoff FW. Overview of the genetics of major depressive disorder. *Curr Psychiatry Rep.* 2010; 12(6):539-546.
6. Gatz M, Pedersen NL, Plomin R, Nesselroade JR, McClearn GE. Importance of shared genes and shared environments for symptoms of depression in older adults. *J Abnorm Psychol.* 1992;101(4):701-708.
7. Wray NR, Sullivan PF. Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *bioRxiv.* 2017.
8. CONVERGE. consortium. Sparse whole-genome sequencing identifies two loci for major depressive disorder. *Nature.* 2015;523:588–591.
9. Hek K, Demirkan A, Lahti J, et al. A genome-wide association study of depressive symptoms. *Biol Psychiatry.* 2013;73(7):667-678.
10. Direk N, Williams S, Smith JA, et al. An Analysis of Two Genome-wide Association Meta-analyses Identifies a New Locus for Broad Depression Phenotype. *Biol Psychiatry.* 2017; 82(5):322-329.
11. Agid O, Kohn Y, Lerer B. Environmental stress and psychiatric illness. *Biomed Pharmacother.* 2000;54(3):135-141.
12. Berk M, Williams LJ, Jacka FN, et al. So depression is an inflammatory disease, but where does the inflammation come from? *BMC Med.* 2013;11:200.
13. Dahl C, Guldberg P. DNA methylation analysis techniques. *Biogerontology.* 2003;4(4):233-250.
14. Rakyan VK, Blewitt ME, Druker R, Preis JI, Whitelaw E. Metastable epialleles in mammals. *Trends Genet.* 2002;18(7):348-351.
15. Lokk K, Modhukur V, Rajashekar B, et al. DNA methylome profiling of human tissues identifies global and tissue-specific methylation patterns. *Genome Biol.* 2014;15(4):r54.
16. Byrne EM, Carrillo-Roa T, Henders AK, et al. Monozygotic twins affected with major depressive disorder have greater variance in methylation than their unaffected co-twin. *Transl Psychiatry.* 2013;3:e269.
17. Latham KE. Stage-specific and cell type-specific aspects of genomic imprinting effects in mammals. *Differentiation.* 1995;59(5):269-282.
18. Edgar RD, Jones MJ, Meaney MJ, Turecki G, Kobor MS. BECon: A tool for interpreting DNA methylation findings from blood in the context of brain. *bioRxiv.* 2017.
19. Bakusic J, Schaufeli W, Claes S, Godderis L. Stress, burnout and depression: A systematic review on DNA methylation mechanisms. *J Psychosom Res.* 2017;92:34-44.
20. Januar V, Saffery R, Ryan J. Epigenetics and depressive disorders: a review of current progress and future directions. *Int J Epidemiol.* 2015;44(4):1364-1387.
21. Nagy C, Suderman M, Yang J, et al. Astrocytic abnormalities and global DNA methylation patterns in depression and suicide. *Mol Psychiatry.* 2015;20(3):320-328.

22. van der Sluis S, Posthuma D, Nivard MG, Verhage M, Dolan CV. Power in GWAS: lifting the curse of the clinical cut-off. *Mol Psychiatry.* 2013;18(1):2-3.
23. Ayuso-Mateos JL, Nuevo R, Verdes E, Naidoo N, Chatterji S. From depressive symptoms to depressive disorders: the relevance of thresholds. *Br J Psychiatry.* 2010;196(5):365-371.
24. Meeks TW, Vahia IV, Lavretsky H, Kulkarni G, Jeste DV. A tune in "a minor" can "b major": a review of epidemiology, illness course, and public health implications of subthreshold depression in older adults. *J Affect Disord.* 2011;129(1-3):126-142.
25. Fried LP, Borhani NO, Enright P, et al. The Cardiovascular Health Study: design and rationale. *Ann Epidemiol.* 1991;1(3):263-276.
26. 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):281-290.
27. Eriksson JG, Forsen T, Tuomilehto J, Osmond C, Barker DJ. Early growth and coronary heart disease in later life: longitudinal study. *BMJ.* 2001;322(7292):949-953.
28. Wichmann HE, Gieger C, Illig T, Group MKS. KORA-gen--resource for population genetics, controls and a broad spectrum of disease phenotypes. *Gesundheitswesen.* 2005;67 Suppl 1: S26-30.
29. Deary IJ, Gow AJ, Pattie A, Starr JM. Cohort profile: the Lothian Birth Cohorts of 1921 and 1936. *Int J Epidemiol.* 2012;41(6):1576-1584.
30. Deary IJ, Gow AJ, Taylor MD, et al. The Lothian Birth Cohort 1936: a study to examine influences on cognitive ageing from age 11 to age 70 and beyond. *BMC Geriatr.* 2007;7:28.
31. Ikram MA, Brusselle GGO, Murad SD, et al. The Rotterdam Study: 2018 update on objectives, design and main results. *Eur J Epidemiol.* 2017;32(9):807-850.
32. Smith BH, Campbell A, Linksted P, et al. Cohort Profile: Generation Scotland: Scottish Family Health Study (GS:SFHS). The study, its participants and their potential for genetic research on health and illness. *Int J Epidemiol.* 2013;42(3):689-700.
33. Marioni RE, McRae AF, Bressler J, et al. Meta-analysis of epigenome-wide association studies of cognitive abilities. *Mol Psychiatry.* 2018.
34. Ligthart S, Marzi C, Aslibekyan S, et al. DNA methylation signatures of chronic low-grade inflammation are associated with complex diseases. *Genome Biol.* 2016;17(1):255.
35. Joehanes R, Just AC, Marioni RE, et al. Epigenetic Signatures of Cigarette Smoking. *Circ Cardiovasc Genet.* 2016;9(5):436-447.
36. The Atherosclerosis Risk in Communities (ARIC) Study: design and objectives. The ARIC investigators. *Am J Epidemiol.* 1989;129(4):687-702.
37. Design of the Women's Health Initiative clinical trial and observational study. The Women's Health Initiative Study Group. *Control Clin Trials.* 1998;19(1):61-109.
38. Lewinsohn PM, Seeley JR, Roberts RE, Allen NB. Center for Epidemiologic Studies Depression Scale (CES-D) as a screening instrument for depression among community-residing older adults. *Psychol Aging.* 1997;12(2):277-287.
39. Zigmond AS, Snaith RP. The hospital anxiety and depression scale. *Acta Psychiatr Scand.* 1983;67(6):361-370.
40. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. *J Gen Intern Med.* 2001;16(9):606-613.

41. First MB, Spitzer, Robert L, Gibbon Miriam, and Williams, Janet B.W. Structured Clinical Interview for DSM-IV-TR Axis I Disorders, Research Version, Patient Edition. (SCID-I/P) November 2002.
42. Horvath S. DNA methylation age of human tissues and cell types. *Genome Biol.* 2013;14(10): R115.
43. Singmann P, Shem-Tov D, Wahl S, et al. Characterization of whole-genome autosomal differences of DNA methylation between men and women. *Epigenetics Chromatin.* 2015;8:43.
44. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics.* 2010;26(17):2190-2191.
45. Han B, Eskin E. Interpreting meta-analyses of genome-wide association studies. *PLoS Genet.* 2012;8(3):e1002555.
46. Bonder MJ, Luijk R, Zhernakova DV, et al. Disease variants alter transcription factor levels and methylation of their binding sites. *Nat Genet.* 2017;49(1):131-138.
47. Barbeira A, Dickinson SP, Torres JM, et al. Integrating tissue specific mechanisms into GWAS summary results. *bioRxiv.* 2017.
48. Gamazon ER, Wheeler HE, Shah KP, et al. A gene-based association method for mapping traits using reference transcriptome data. *Nat Genet.* 2015;47(9):1091-1098.
49. Major Depressive Disorder Working Group of the Psychiatric GC, Ripke S, Wray NR, et al. A mega-analysis of genome-wide association studies for major depressive disorder. *Mol Psychiatry.* 2013;18(4):497-511.
50. Dehghan A, Dupuis J, Barbalic M, et al. Meta-analysis of genome-wide association studies in >80 000 subjects identifies multiple loci for C-reactive protein levels. *Circulation.* 2011;123(7): 731-738.
51. Tobacco, Genetics C. Genome-wide meta-analyses identify multiple loci associated with smoking behavior. *Nat Genet.* 2010;42(5):441-447.
52. Relton CL, Davey Smith G. Two-step epigenetic Mendelian randomization: a strategy for establishing the causal role of epigenetic processes in pathways to disease. *Int J Epidemiol.* 2012;41(1):161-176.
53. Chen XQ, Tan I, Leung T, Lim L. The myotonic dystrophy kinase-related Cdc42-binding kinase is involved in the regulation of neurite outgrowth in PC12 cells. *J Biol Chem.* 1999;274(28): 19901-19905.
54. Etienne-Manneville S, Hall A. Rho GTPases in cell biology. *Nature.* 2002;420(6916):629-635.
55. DeGeer J, Lamarche-Vane N. Rho GTPases in neurodegeneration diseases. *Exp Cell Res.* 2013; 319(15):2384-2394.
56. Liu BP, Strittmatter SM. Semaphorin-mediated axonal guidance via Rho-related G proteins. *Curr Opin Cell Biol.* 2001;13(5):619-626.
57. Woo NH, Teng HK, Siao CJ, et al. Activation of p75NTR by proBDNF facilitates hippocampal long-term depression. *Nat Neurosci.* 2005;8(8):1069-1077.
58. Davies AM. Nerve growth factor synthesis and nerve growth factor receptor expression in neural development. *Int Rev Cytol.* 1991;128:109-138.
59. Fiore R, Puschel AW. The function of semaphorins during nervous system development. *Front Biosci.* 2003;8:s484-499.
60. Schecterson L, Bothwell M. An all-purpose tool for axon guidance. *Neuroscience.* 2008;1(47): Pe50.

61. Burkhardt C, Muller M, Badde A, Garner CC, Gundelfinger ED, Puschel AW. Semaphorin 4B interacts with the post-synaptic density protein PSD-95/SAP90 and is recruited to synapses through a C-terminal PDZ-binding motif. *FEBS Lett.* 2005;579(17):3821-3828.
62. Paradis S, Harrar DB, Lin Y, et al. An RNAi-based approach identifies molecules required for glutamatergic and GABAergic synapse development. *Neuron.* 2007;53(2):217-232.
63. Pasterkamp RJ, Giger RJ. Semaphorin function in neural plasticity and disease. *Curr Opin Neurobiol.* 2009;19(3):263-274.
64. Chuang YH, Paul KC, Bronstein JM, Bordelon Y, Horvath S, Ritz B. Parkinson's disease is associated with DNA methylation levels in human blood and saliva. *Genome Med.* 2017;9(1):76.
65. Starnawska A, Hansen CS, Sparso T, et al. Differential DNA methylation at birth associated with mental disorder in individuals with 22q11.2 deletion syndrome. *Transl Psychiatry.* 2017;7(8):e1221.
66. Bagot RC, Labonte B, Pena CJ, Nestler EJ. Epigenetic signaling in psychiatric disorders: stress and depression. *Dialogues Clin Neurosci.* 2014;16(3):281-295.
67. Klengel T, Pape J, Binder EB, Mehta D. The role of DNA methylation in stress-related psychiatric disorders. *Neuropharmacology.* 2014;80:115-132.
68. Galanter JM, Gignoux CR, Oh SS, et al. Differential methylation between ethnic sub-groups reflects the effect of genetic ancestry and environmental exposures. *Elife.* 2017;6.
69. Kader F, Ghai M. DNA methylation-based variation between human populations. *Mol Genet Genomics.* 2017;292(1):5-35.
70. International Consortium for Blood Pressure Genome-Wide Association S, Ehret GB, Munroe PB, et al. Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. *Nature.* 2011;478(7367):103-109.
71. Lisoway AJ, Zai CC, Tiwari AK, Kennedy JL. DNA Methylation and Clinical Response to Antidepressant Medication in Major Depressive Disorder: A Review and Recommendations. *Neurosci Lett.* 2017.
72. Gottfries CG. Is there a difference between elderly and younger patients with regard to the symptomatology and aetiology of depression? *Int Clin Psychopharm.* 1998;13:S13-S18.

Chapter 4

General discussion

This thesis aimed to identify novel risk factors for COPD, lower lung function, asthma and depression and to perform integrative studies to investigate the functional role and interaction of multiple omics layers. This chapter will emphasize the main findings, discuss challenges, possible clinical implications and future directions of COPD research.

FINDINGS OF THIS THESIS

COPD as a common and complex disease is a major public health burden.[1,2] The aim of this thesis has been to identify novel molecular determinants of COPD, lower lung function and related pathology such as depression and to perform integrative studies to investigate the functional role and interaction of multiple omics layers. GWAS identified common genetic variants to be associated with COPD, however, they exert small effects and their functional role is unknown. In **Chapter 2.1** of this thesis, I performed a genome-wide linkage scan to identify rare genetic variants associated with COPD. Genetic linkage analysis is a powerful tool to identify genomic regions shared among the affected family members. This was done in the large genetically isolated Erasmus Rucphen Family study. Using a genetic isolate, characterized with shared lifestyle and environment effects as well as less genetic variance, increases the power of the analysis. I found significant evidence for extensive linkage of COPD to the known COPD GWAS region, chromosome 15q14-15q25 and to two novel regions, 11p15.4-11q14.1 and 5q14.3-5q33.2. More importantly, I was able to identify four pathogenic, rare variants, shared by family members with COPD using exome-sequence data. They belong to the chromosome 11 peak, genes *AHNAK, PLCB3, SLC22A11* and *MTL5*. The variants in *SLC22A11* and *MTL5* were confirmed in association with COPD in our meta-analysis of 9,888 cases and 27,428 controls. These two genes are both expressed in lung tissue and are interesting candidates. Further functional studies should confirm their role in respiratory pathology in general population. Although I confirm the linkage of chromosome 15 region, I could not identify any shared rare variants by means of exome-sequence data. This may be explained by the fact that only common variants are relevant, or, alternatively that common or rare regulatory variants outside of exons may play a key role in COPD. Future candidate-gene studies using whole-genome sequencing should further investigate the regions on chromosomes 5 and 15. Also, possible gene interaction studies are promising avenue for future of genetic COPD research.

It has been postulated that epigenetic modifications are in part driven by genetic variants.[3] In attempt to integrate multiple -omics layers, I focused on the effects of

known COPD genetic variants from chromosomes 15q25.1 (**Chapter 2.2**) and 19q13.2 (**Chapter 2.3**) on genome-wide DNA methylation in blood and gene expression in blood and lungs. Overall, findings of these two chapters highlight the need for integration of multi-omics data to further understand the role of GWAS loci.

In **Chapter 2.2** I performed meQTL and eQTL study and showed evidence suggesting that genetic variations underlying genes in the chromosome 15q25.1 region (in *IREB2, HYKK* and *CHRNA3*) influence COPD susceptibility through changes in blood DNA methylation of *IREB2, PSMA4* and *CHRNA3*. Furthermore, I observed the association of the same variants with several *cis* and *trans* gene expression changes in the lung tissue. These results suggest a disease model in which COPD risk allele of *CHRNA3* variant lowers DNA methylation and increases expression of *IREB2* in COPD patients. We suggest mediation by DNA methylation levels in this region, but future studies using the data from the same tissue should confirm this hypothesis. Even though genetic variants in chromosome 15q25.1 (encompassing the nicotinic receptor genes - *CHRNA3, CHRNA5* and *CHRNB4*) were also previously associated with smoking and smoking has a known effect on DNA methylation, our findings were independent of smoking, making the mediation by smoking not a necessary mechanism driving the relation between DNA methylation and COPD. Both smoking and genetic determinant may lead independently to differential DNA methylation. However, targeted studies in lung tissue should verify these findings.

Similarly, **Chapter 2.3** shows that the top COPD GWAS variant in chromosome 19q13.2 region (rs7937 in *RAB4B, EGLN2*) is associated with lower blood DNA methylation of *EGLN2*, independent of smoking and of COPD. I further showed that this DNA methylation site in *EGLN2* is associated with COPD, again independent of smoking. Having both DNA methylation and expression data in blood, I performed a mediation analysis and showed that differential DNA methylation mediates 42% of the association between the genetic variant and differential expression of *EGLN2*. I also showed the effect of the variant on *cis* and *trans* gene expression changes in lung tissue of *NUMBL, AK097370 (EGLN2), LOC101929709, DNMT3A* and *PAK2*. Our findings are in line with our hypothesis that the life-long change in the DNA methylation is involved in the pathogenesis and onset of COPD in older age, yet further longitudinal studies are needed, testing this hypothesis in large set of lung tissue characterized for multi-omics data.

As smoking plays an important role in the disease development and has a known effect on epigenetic landscape, it is crucial to take this determinant into account in epidemiological studies of COPD.[3,4] In lack of better assessment, self-reported current smoking status (current-, ex- and never-smokers) is widely used variable. To truly remove smoking effects from the equation, the analyses should be performed

in never-smokers only. However, such stratification would significantly reduce the statistical power and the analysis would require large sample sizes to show the real effects. With the aim to investigate DNA methylation signature of airflow obstruction, independent of smoking effects, I performed the largest to date EWAS on lung function levels in never smokers only, presented in **Chapter 2.4**. I identified 36 DNA methylation sites that were highly significantly associated with FEV_1/FVC. This chapter importantly contributes to the current understanding of epigenetic changes in COPD, when smoking effect is excluded. Moreover, this chapter presents current literature on blood DNA methylation in COPD and smoking and shows that the majority of the identified 36 DNA methylation sites are unique for never smokers. DNA methylation of *KLHL32* and *LTV1* genes, among others, indeed may play a role in the disease development in subjects with COPD that never smoked. Although we see associations in blood, many of the methylated genes are expressed in the lung tissue. Genetic variants in *KLHL32* have been associated with post-bronchodilator FEV_1 and FEV_1/FVC in COPD.[5] *LTV1* is shown to be an enhancer of two genes identified in GWAS of FEV_1/FVC response to bronchodilators (*PLAGL1*)[5] and lung cancer (*PHACTR2*)[6] in GeneHancer database.[7] These findings propose a possible new regulatory pathway, independent of smoking, through which DNA methylation of *LTV1* influences genetic susceptibility of COPD. Of course, this is just a hypothesis based on our findings and future functional studies of causality should investigate it further.

Chapter 3 of this thesis is investigating comorbidities of COPD which are known to influence the severity of the disease.[8] Asthma is considered the most common comorbidity, being both a risk factor for COPD and a co-existing disease in the elderly.[9,10] There is a significant overlap between genetic risk factors of asthma and COPD, but the regulatory mechanisms are unknown.[11] It is shown that lower lung function in childhood with subsequent normal or accelerated decline had increased risk of COPD.[12] Early life allergic diseases, lung infections, parental asthma, and maternal smoking predicted worse lung function, while personal smoking amplified the effect of maternal smoking.[12] In **Chapter 3.1** we hypothesized that umbilical cord blood DNA methylation influences childhood lung function, lung development and gene expression, and increases the risk of asthma and COPD in later life. We performed large EWAS meta-analyses of FEV_1, FEV_1/FVC and Forced Expiratory Flow at 75% of FVC ($FEF_{75)}$. EWAS results were pooled into differentially methylated regions (DMRs) and 59 such DMRs in neonatal cord blood were associated with childhood lung function. Multiple DMRs were additionally associated with childhood asthma, adolescent and adult lung function, COPD and with differential gene expression. These findings suggest that epigenetic changes during foetal life might modify the risk of respiratory diseases across the life course. As epigenetic landscape of a foetus is highly

influenced by environment and maternal behaviour,[13] identifying and reducing risk factors in pregnancy, may become a future strategy for prevention of lung diseases.

As COPD is a systemic disease, there are manifestations beyond the airflow obstruction, such as systemic inflammation, muscle degeneration and oxidative stress, which can result in a specific plasma biomarkers profile.[14–16] The identification of these specific biomarker changes can identify or differentiate disease phenotypes even in the early stages of COPD.[17] Therefore, in **Chapter 3.2** I performed a hypothesis-free analysis and identified and replicated the association of Glycoprotein acetyls (GlycA) with COPD, the only association that passed the multiple testing correction. The Rotterdam study and other epidemiological follow-up studies show that GlycA, as measured before the disease onset, increased the risk of COPD and may thus be a risk factor of the disease rather than a consequence. To validate the causal pathway, I further performed Mendelian Randomization (MR) analysis. This analysis clearly showed that the genes driving the risk of COPD, are also associated with GlycA. No evidence was found for the risk factor model in which the genes driving GlycA associate significantly to COPD. This finding suggest that GlycA is an early marker of COPD pathology. It is of interest that **Chapter 3.2** shows that GlycA which is a marker of the acute phase response, also strongly associates to smoking and may thus be a part of an inflammation pathway linking smoking to COPD. However, further functional studies should investigate the specific role of GlycA in COPD pathogenesis, prognosis, severity and treatment response.

Aside from asthma and other pulmonary conditions, many different comorbid diseases add to the burden of COPD. In attempt to identify common pathophysiology and decipher the co-occurrence of those diseases, I performed the genetic correlation analysis of COPD with 126 diseases available on the LD hub database, presented in **Chapter 3.3**. I describe the significant correlation of neuro-psychiatric and cardio-metabolic pathology, as well as female reproductive conditions autoimmune diseases of the bowel and aging disorders such as cataract. Of note is, for the first time shown, significant correlation of coronary artery disease, acute myocardial infarction, angina pectoris, hypertension, diabetes, chronic kidney disease, attention deficit hyperactivity disorder, schizophrenia, family history of depression (depression in sibling) and suggestive correlation with major depressive disorder.

The understudied comorbidity of COPD is depression, investigated in **Chapter 3.4.** As genetic risk factors of depression and COPD do not seem to overlap strongly (**in Chapter 3.3**), I studied different omics layers to try to identify common mechanisms explaining this co-occurrence. Depression related DNA methylation changes have poorly been studied, hence, we have performed largest to date EWAS of depressive symptoms. This chapter presents the identified DNA methylation sites annotated

to the *CDC42BPB* gene, *ARHGEF3* gene, and one intergenic region on chromosome 15q26.1 locus. All three findings point towards axon guidance as the common disrupted pathway in depression. DNA methylation site in *CDC42BPB* gene was also associated with inflammation,[18] smoking status[3] and pack-years of smoking[19] in two independent studies. It has been speculated that *CDC42BPB* may be a future biomarker of COPD[20] and *CDC42BPB* is a downstream target of *CDC42*, whose expression is altered in obese children with astma.[21] At the genetic level, *CDC42BPB* is associated with gamma-glutamyl transferase (http://atlas.ctglab.nl/phewas), which have been found to be differentially expressed in COPD mouse models.[22] Identifying depression related DNA methylation sites was a first step towards unravelling the complex mechanisms underlying depression, but may also shed light on the co-occurrence of depression and COPD and the role of smoking. Future studies should investigate the role of identified genes in COPD and other way around. Using multi-omics approach for investigating comorbidities has a potential to disentangle these relations and provide better treatment options and prognosis for both diseases.

METHODOLOGICAL CONSIDERATIONS

All studies in this thesis were performed using participants' data from Rotterdam Study, a population-based cohort study consisting of 45 years or older people form Rotterdam. In addition, genetic linkage study (**Chapter 2.1**) and metabolic study (**Chapter 3.2**) used data from the Erasmus Rucphen Family study, a genetic isolate from southwest of the Netherlands. Aside from the two studies we also used several Dutch and international studies as part of collaborative efforts of big consortiums. Details and methodological issues of each study are discussed in each chapter of this thesis. Here I would like to mention several general issues important to be considered.

Smoking assessment

Tobacco smoking is one of the major risk factors for many chronic diseases and different types of cancers (**Figure 1**). Therefore, it is one of the most investigated risk factors in epidemiological studies, yet the assessment of smoking has not been standardized. The effects of smoking are commonly assessed using self-administered questionnaires and studied as different variables: smoking behaviour (current smoking status, ever smoking status), smoking quantity (pack-years of smoking, cigarettes per day), nicotine dependence (time from waking up until first cigarette),[23] age of smoking initiation, smoking cessation, second-hand smoke and other. Also, tobacco is used in different ways: in form of cigarettes (different brands

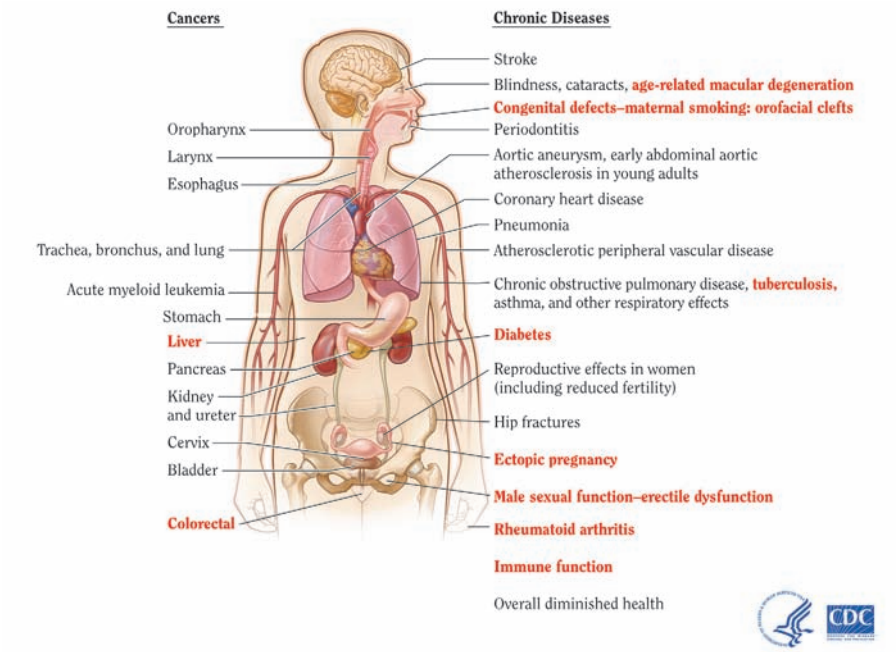

Figure 1. Diseases causally linked to smoking. Source: US Centers for Disease Control and Prevention (https://www.cdc.gov/). In red: a new disease causally linked to smoking in the 2014 Surgeon General's Report: The Health Consequences of Smoking—50 Years of Progress.

vary in the amount of nicotine and noxious particles), smoked in pipes and chewed. Using self-administered questionnaires to assess very heterogeneous exposure may introduce information bias, due to the erroneous classification of subjects. Specifically, patients with lung diseases are shown to likely falsely report their smoking behaviour.[24] The reason for this is often inaccurate memory of past events (recall bias) or deliberately understating as people tend to diminish the adverse behaviour.[25] This leads to frequent underestimation of smoking rates and, when used as a confounder in regression analysis, to over- or underestimation of effect estimates. This is especially important in epigenetic and transcriptomic studies of lung diseases as DNA methylation and gene expression is known to be affected by smoking.[3,26] Genetic studies have been identified risk loci associated with smoking, such as nicotinic receptors on chromosome 15, studied in **Chapter 2.2** of this thesis. In attempt to exclude the smoking effect, in **Chapter 2.4** I study DNA methylation in never smokers only. In **Chapter 3.2** I investigate the role of smoking in the identified COPD-GlycA associations. When studying multi-omics of lung diseases, where smoking plays an important role, it is imperative to take real smoking ef-

fects into account. However, in most of the cohorts used in this thesis, in absence of a better assessment, smoking behaviour is measured using self-administered questionnaires as current smoking status (current, ex or never smokers) and pack-years of any smoking (including cigarettes and pipes). Although self-administered questionnaire is quick, easy and cheap way of assessing smoking and thus widely used, the question remains whether it captures the real smoking exposure. Objective assessment of smoking biomarkers, such as nicotine, cotinine and the exhaled carbon monoxide concentrations, is available but is more expensive and less used in population-based studies.[27] These methods would undoubtedly report actual current smoking exposure, even in never-smokers exposed to second-hand smoke, but would not show difference between never-smokers and ex-smokers.[28] It is shown that there is a big discrepancy between self-reported and objectively measured current smoking status.[24,29] While the next generation technologies for assessment of omics data are developing fast and large datasets are readily available through national biobanks and large consortia, phenotyping of smoking exposure is still plummeted with misclassification of subjects. Therefore, it is of utmost importance for objective smoking assessment methods to become more widely used, especially in the next-generation multi-omics studies of lung diseases.

DNA methylation assessment

All cohorts included in the epigenetic studies of this thesis were quantifying DNA methylation using Infinium Human Methylation 450 Bead Chip from llumina, the array with more than 450 thousand CpG sites. It is known that 450K array measures only 1.7% of all CpGs mostly covering CpG-islands.[30] CpG island is a stretch of DNA with the highest frequency of CpGs, mostly located in the gene promotor regions and gene bodies. Therefore, when analysing data, limited genome coverage should be considered, as enhancer regions are almost completely missing. Furthermore, it has been discovered that some hybridization probes of 450K array co-hybridize to similar genomic sequences (cross-reactive CpGs) or target CpGs that overlap with genetic polymorphisms (polymorphic CpGs), so the measured methylation levels may reflect the underlying SNPs.[31] Potentially cross-reactive and polymorphic CpGs can cause a measurement bias and have been annotated previously.[31] This can confound the data, though unknown to which extent, and found associations should be interpreted with caution. Being a technical issue, adjusting for the underlying SNPs would not correct this error, while excluding the biased sites would possibly generate false-negative findings. To be stringent and sure of the validity of our findings, in most of our epigenetic studies we have excluded both cross-reactive and polymorphic CpGs. Future studies using the new Illumina 850K EPIC array, in which

problematic CpGs are corrected and many more added, covering the enhancer regions, this issue should be solved.[32]

Tissue of interest

For the quantification of the multi-omics data, blood is the most commonly used tissue as it is easily obtained and cheap. However, when using blood in studying epigenetics, transcriptomics and metabolomics of lung diseases the question arises whether changes identified in blood represent the processes in the tissue of interest – the lungs. On the other hand, COPD represents more of a syndrome than a single disease including different phenotypes and underlying processes which we are still trying to comprehend. Those are airway obstruction in bronchitis, loss of lung parenchyma in emphysema, as well as systemic effects such as the inflammation and oxidative stress as a response to the noxious particles, muscle wasting and changes in metabolism.[33] It has been shown that those systemic effects are detectable in plasma e.g. through the role of the macrophage and neutrophils in the pathogenesis of COPD (**Figure 2**),[34] justifying the use of blood as the tissue of interest.[35,36] In this

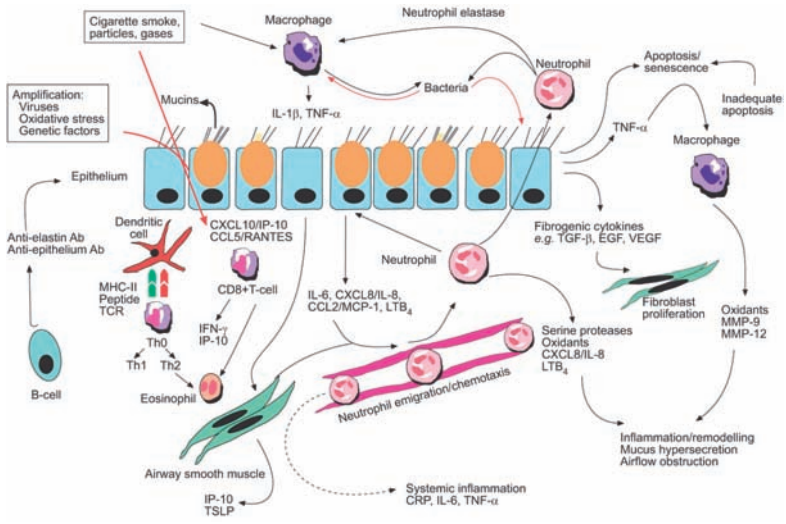

Figure 2. Summary of inflammatory and cellular interactions linking chronic cigarette exposure to the chronic inflammation of chronic obstructive pulmonary disease (COPD). Ab: antibody; Th: T-helper cell; MHC: major histocompatibility complex; TCR: T-cell receptor; CXCL: CXC chemokine ligand; IP: interferon (IFN)-γ-inducible protein; CCL: CC chemokine ligand; RANTES: regulated on activation, normal T-cell expressed and secreted; TSLP: thymic stromal lymphopoietin; IL: interleukin; TNF: tumour necrosis factor; MCP: monocyte chemotactic protein; LT: leukotriene; CRP: C-reactive protein; TGF: transforming growth factor; EGF: epidermal growth factor; VEGF: vascular endothelial growth factor; MMP: matrix metalloproteinase. Source: Chung KF, Adcock IM. *Eur Respir J*. 2008;31(6):1334-1356. doi:10.1183/09031936.00018908.

thesis blood was used for quantification of DNA methylation levels as well as gene expression and metabolic levels. However, for gene expression analyses described in **Chapters 2.2** and **2.3** we confirmed the regulation of gene expression by the COPD-associated genetic variants in lung tissue. When integrating multi-omics levels, it is very important to have all the data in lung tissue so that causal relationships involved in COPD pathophysiology can be investigated.

An important drawback to be discussed about the use of expression in blood is that it is quite heterogeneous tissue, consisting of many cell types, whose proportions may vary in different people and with disease severity. In whole blood DNA methylation studies, such as those described in this thesis, different types of leukocytes are used (lymphocytes, monocytes and granulocytes). As DNA methylation can be cell-specific, using a mix of cell types prone to inter-individual variability can give biased results.[37] Therefore, these analyses should be adjusted for cell proportions, either measured within the study or predicted based on the DNA methylation data using computational methods.[38] In the chapters of this thesis we have used both measured cell proportions and estimated proportions utilizing Houseman method.[39] However, it is important to consider that, while Housman estimation is good method to use for consistency in the big consortium efforts, it is based on half a dozen reference individuals and may not completely represent the proportions in the general population.

Causality

In this thesis we often performed cross-sectional association analyses, disabling us to claim any causal relationships between the studied exposure and the outcome. Even the variants identified in linkage analysis of **Chapter 2.1** may not be causal for COPD but simply in a linkage disequilibrium with the causal variant which is not genotyped. The fact that the identified genes are expressed in the lung tissue, may suggest that the real causal variant is within the gene. However, any claims of causality should be taken with caution and confirmed by functional studies using knockout models. On the other hand, DNA methylation is a dynamic process subjected to change and influenced by both external and internal effects. In this thesis I have reported associations of GWAS SNPs with blood DNA methylation, further associated with COPD (**Chapters 2.2** and **2.3**), yet it may be that having COPD changes DNA methylation at those sites. I have also showed association of new-born cord blood DNA methylation with lung function and asthma and COPD in later life in **Chapter 3.1.** In this chapter, using different time points for measuring the data could suggest that reverse causality may not be possible, but this should be confirmed in future longitudinal studies on the same people. In **Chapter 2.3,**

having both DNA methylation and gene expression data derived from blood, I could perform mediation analysis and I showed that DNA methylation of *EGLN2* mediates 42% of the association between rs7937 and expression of *EGLN2.* This may suggest a direction of the effect, but the reverse causality may also happen, as differential gene expression can affect DNA methylation levels of *EGLN2*.

The Holy Grail for testing causality in observational epidemiological research is to conduct (nested) follow-up studies in which the exposure is measured before the disease. We conducted a nested follow-up study of various metabolites and COPD (**Chapter 3.2**) and found evidence that that Glycoprotein acetyls (GlycA) were associated with an increased risk of COPD. Recent development in genetics have made it possible to use genes as instrumental variable (IV) to test for causality using Mendelian Randomisation (MR; explained in **Chapter 1**). In **Chapter 3.2** I tested the most likely causal pathway underlying the association of COPD with GlycA. I showed in this chapter that COPD is causally related to the elevated levels of GlycA, rather than other way around. Although the conclusion is important, it is important to consider the possibility that the assumptions underlying MR are violated. These include:

1. The IV has to be associated with the exposure;
2. The IV has to be independent of any confounders of the exposure-outcome association and
3. The IV has to be related to the outcome only through the exposure.

However, as genetic variants used as IV may associate with other unknown confounders of the COPD-GlycA association, the condition two may have been violated. The finding that the genes predicting GlycA do not predict COPD makes it unlikely that GlycA is a risk factor. However, the hypothesis that GlycA is a biomarker of COPD pathology remains to be studied further prospectively.

External validity

External validity, the extent to which the results can be generalized to other situations and to other people, is an important factor to consider in epidemiological study. This issue should be taken into account when interpreting the results of **Chapter 2.1** of this thesis. Identified rare variants, identified in the genetic isolate, may be population-specific thus rare or non-existent in other populations.[40] This especially stands for the two variants which we could not replicate in the population-based setting. Furthermore, the replication cohorts were mainly of European ancestry, questioning the generalizability to other ancestries, since it is known that genetic variants have different effects in different ethnic populations. Utilizing large datas-

ets to replicate our findings is required before speculating on the external validity of these findings.

Differentially methylated regions

It has been postulated that differential DNA methylation at a single CpG site has a small effect on the risk for the disease and it should be evaluated considering the effects of the neighbouring sites or the whole region,[41] which makes sense from the biological point of view. These differentially methylated regions (DMR) are estimated using Comb-p method in **Chapter 3.1** of this thesis.[42] This method calculates auto-correlation, combines neighbouring P-values, corrects for false discovery rate, finds regions of enrichment and assigns a P-value to those regions.[42] The number of identified DMRs is much larger than identified single sites since this method increases power to identify region of interest. We identified 59 DMRs associated with childhood lung function, of which 31% were associated with childhood asthma, 19% with adolescent lung function, 15% with adult lung function, and 15% with COPD while 54% were influencing gene expression in childhood and 31% in adulthood. However, as Illumina 450K array is known to have unequable coverage of CpG sites throughout the genome, this method has its limitations. Therefore, simply pooling the CpG sites using computational methods, without any a-priori knowledge is questionable and the results should be interpreted with caution. The use of EPIC 850K array in the future, which has much better coverage of CpGs across the genome, may bridge this limitation.

POTENTIAL IMPLICATIONS AND FUTURE DIRECTIONS

Clinical implications and directions for future research were discusses in detail in every chapter of this thesis. Here I will highlight my most important ideas for the future.

Development of technical means which can withstand computationally demanding analyses opened new avenues for research of complex diseases. COPD, as a complex disease, already benefitted from the use of hypothesis-generating -omics approaches. GWAS identified over 40 loci associated with either lung function or COPD, yet the reproducibility of these findings is very low and the heritability is still largely missing.[43] GWAS variants are usually common with very small effect size. Using stringent, genome-wide significant thresholds we will need to increase sample sizes to reach enough statistical power to discover new common genetic variants. In the future, this can be done using large biobanks, collections of nation-wide data,

lately developing in several countries. It will be crucial to include information on different ancestries, as currently most of the findings are identified on Caucasians, and is has been speculated that this is leading to a disparity in future health care. Worldwide collaborations would also facilitate these efforts through large consortiums such as CHARGE and PACE consortiums, used in **Chapter 3** of this thesis. In order to confirm the role of rare variants on chromosome 11, described in **Chapter 2.1,** and to generalize findings to the general population, larger candidate-gene studies will be useful. Furthermore, large studies utilizing whole-exome and -genome sequencing should investigate the role of linked regions in chromosomes 5 and 15 in COPD. Such studies are now feasible and affordable at a large scale for instance the UK biobank. Gene-gene and gene-environment interactions should be investigated in further attempt to explain the missing heritability in the future. There has been little success up until now and this field awaits advances through upscaling the size of studies and the development of new computational approaches (e.g. deep learning). Functional studies are further needed to definitely confirm the role of novel variants in the disease pathogenesis.

Further, we show usefulness of investigating metabolic profiles of the disease to hopefully differentiate molecular processes in specific tissues. I found that GlycA is a biomarker of early COPD pathology that is elevated before the diagnosis of COPD and future studies widening the net of metabolites studied are likely to find other metabolites, improving the prediction of disease and yielding new information that in combination can empower future (preventive) trials.

In the future, multi-omics studies of COPD would benefit from longitudinal design, measuring multiple omics layers in the lung tissue of the same people in multiple time points. Basing the study in well characterized epidemiological cohorts will allow to remove the confounding effects of smoking and medication. This would lead to better inferences on causality and direction of the effects. MR is a useful method to infer causality and disentangle complex relations of multiple omics layers and should be used in the future research. Such studies would also allow to derive objective smoking assessment based on methylation, transcription and metabolomics profiles. Similarly, one can aim to capture the effect of air pollution and other risk factor of COPD in omics signatures.

Improved and more discrete phenotyping of COPD and its confounders, such as smoking behaviour, is required in order to understand the genetic architecture of the disease. This includes investigating comorbidities of COPD and their shared and overlapping risk factors. As comorbidities are influencing disease severity, prognosis and treatment response and complete well-being of the patient, it is imperative to determine to what extend these can be prevented.

Large scale genome-wide multi-omics studies in the lung tissue are the next phase in the respiratory research, integrating data in the context of a network medicine.[44] This may improve understanding of the disease heterogeneity, improve classification and identification of individuals in high risk and translate the findings to clinical care and prevention (**Figure 3**). This may open new avenues for precision medicine in the future.[45]

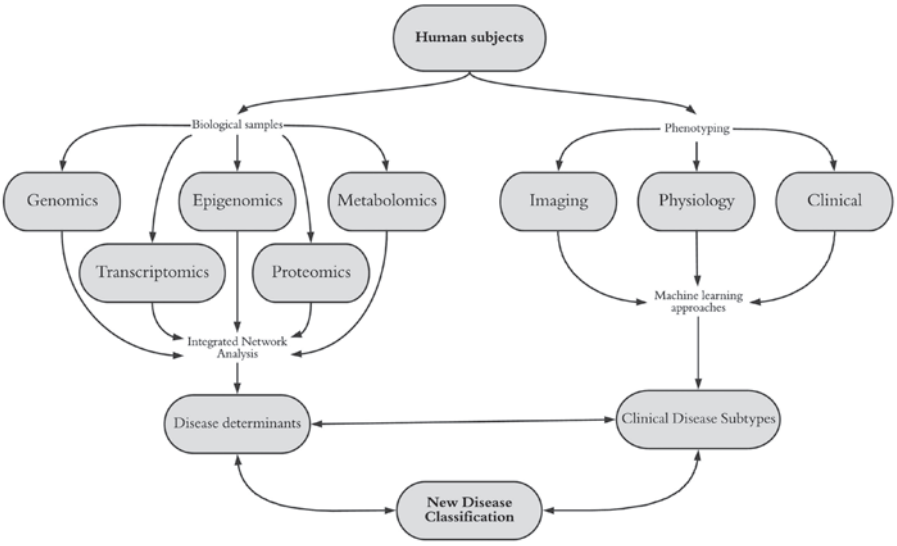

Figure 3. Potential Approaches to Reclassify Complex Diseases in Network Medicine. Adapted from Silverman and Loscalzo *Discov Med*. 2012;14(75):143-152.

The final aim of the network medicine, the identification of important disease determinants and reclassification of complex disease, such as COPD, is to have novel drug development strategies in the future and improve clinical care (**Figure 4**).[46,47]

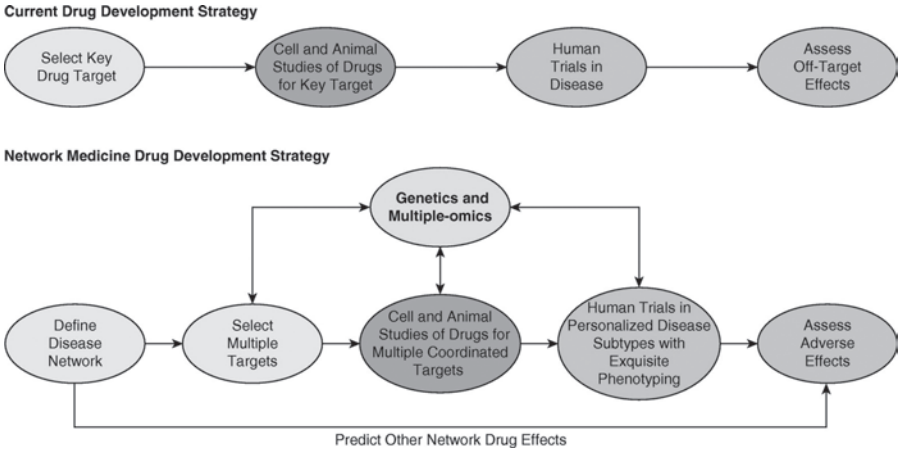

Figure 4. Current and Network Medicine Approaches to Drug Development for Complex Diseases. Adapted from Silverman EK, Loscalzo J. Clin Pharmacol Ther. 2012;93(1):26-8.

Conclusion

In this thesis I studied COPD using several omics layers in attempt to elucidate the molecular mechanisms underlying the disease. I have identified several novel risk factors for COPD and its comorbidities, proposed regulatory pathways and highlighted the need for integration of the multi-omics data. Furthermore, I have discussed methodological challenges and ideas from which the future respiratory research could benefit. Findings of this thesis require functional confirmation, but I am confident that it represents another important step on a path towards improved clinical care and prevention of COPD, based on precision medicine.

REFERENCES

1. Lozano R, Naghavi M FK. Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet*. 2012;380(380):2095–2128. doi:https://doi.org/10.1016/S0140-6736(12)61728-0
2. European Respiratory Society. European lung white book. http://www.erswhitebook.org/chapters/chronic-obstructive-pulmonary-disease/. Published 2016.
3. Yet I, Tsai P-C, Castillo-Fernandez JE, Carnero-Montoro E, Bell JT. Genetic and environmental impacts on DNA methylation levels in twins. *Epigenomics*. 2016;8(1):105-117. doi:10.2217/epi.15.90
4. Joehanes R, Just AC, Marioni RE, et al. Epigenetic Signatures of Cigarette Smoking. *Circ Cardiovasc Genet*. 2016;9(5):436-447. doi:10.1161/CIRCGENETICS.116.001506
5. Lokke A, Lange P, Scharling H, Fabricius P, Vestbo J. Developing COPD: a 25 year follow up study of the general population. *Thorax*. 2006;61(11):935-939. doi:10.1136/thx.2006.062802
6. Lutz SM, Cho MH, Young K, et al. A genome-wide association study identifies risk loci for spirometric measures among smokers of European and African ancestry. *BMC Genet*. 2015;16(1):138. doi:10.1186/s12863-015-0299-4
7. Wang L-E, Gorlova OY, Ying J, et al. Genome-Wide Association Study Reveals Novel Genetic Determinants of DNA Repair Capacity in Lung Cancer. *Cancer Res*. 2013;73(1):256-264. doi:10.1158/0008-5472.CAN-12-1915
8. Fishilevich S, Nudel R, Rappaport N, et al. GeneHancer: genome-wide integration of enhancers and target genes in GeneCards. *Database*. 2017;2017. doi:10.1093/database/bax028
9. Chatila WM, Thomashow BM, Minai OA, Criner GJ, Make BJ. Comorbidities in Chronic Obstructive Pulmonary Disease. *Proc Am Thorac Soc*. 2008;5(4):549-555. doi:10.1513/pats.200709-148ET
10. Silva GE. Asthma as a Risk Factor for COPD in a Longitudinal Study. *Chest*. 2004;126(1):59-65. doi:10.1378/chest.126.1.59
11. Aryal S, Diaz-guzman E, Mannino DM. Asthma Treatment Options in Asthma and Chronic Obstructive Pulmonary Diseases Overlap Syndrome. *Touch Briefings, Eur Respir Dis*. 2011;7(2):101-105.
12. Hobbs BD, De Jong K, Lamontagne M, et al. Genetic loci associated with chronic obstructive pulmonary disease overlap with loci for lung function and pulmonary fibrosis. *Nat Genet*. 2017;49(3):426-432. doi:10.1038/ng.3752
13. Bui DS, Lodge CJ, Burgess JA, et al. Childhood predictors of lung function trajectories and future COPD risk: a prospective cohort study from the first to the sixth decade of life. *Lancet Respir Med*. 2018;6(7):535-544. doi:10.1016/S2213-2600(18)30100-0
14. Eckhardt F, Lewin J, Cortese R, et al. DNA methylation profiling of human chromosomes 6, 20 and 22. *Nat Genet*. 2006;38(12):1378-1385. doi:10.1038/ng1909
15. Chen Q, Deeb RS, Ma Y, Staudt MR, Crystal RG, Gross SS. Serum metabolite biomarkers discriminate healthy smokers from COPD smokers. *PLoS One*. 2015;10(12):e0143937. doi:10.1371/journal.pone.0143937
16. Adamko DJ, Nair P, Mayers I, Tsuyuki RT, Regush S, Rowe BH. Metabolomic profiling of asthma and chronic obstructive pulmonary disease: A pilot study differentiating diseases. *J Allergy Clin Immunol*. 2015;136(3):571-580.e3. doi:10.1016/j.jaci.2015.05.022
17. Ubhi BK, Riley JH, Shaw PA, et al. Metabolic profiling detects biomarkers of protein degradation in COPD patients. *Eur Respir J*. 2012;40(2):345-355. doi:10.1183/09031936.00112411

18. Nobakht M Gh BF, Aliannejad R, Rezaei-Tavirani M, Taheri S, Oskouie AA. The metabolomics of airway diseases, including COPD, asthma and cystic fibrosis. *Biomarkers Biochem Indic Expo response, susceptibility to Chem*. 2015;20(1):5-16. doi:10.3109/1354750X.2014.983167
19. Ligthart S, Marzi C, Aslibekyan S, et al. DNA methylation signatures of chronic low-grade inflammation are associated with complex diseases. *Genome Biol*. 2016;17(1):255. doi:10.1186/s13059-016-1119-5
20. de Vries M, van der Plaat DA, Nedeljkovic I, et al. From blood to lung tissue: effect of cigarette smoke on DNA methylation and lung function. *Respir Res*. 2018;19(1):212. doi:10.1186/s12931-018-0904-y
21. Leidinger P, Keller A, Heisel S, et al. Novel autoantigens immunogenic in COPD patients. *Respir Res*. 2009;10(1):20. doi:10.1186/1465-9921-10-20
22. Rastogi D, Nico J, Johnston AD, et al. CDC42-related genes are upregulated in helper T cells from obese asthmatic children. *J Allergy Clin Immunol*. 2018;141(2):539-548.e7. doi:10.1016/j.jaci.2017.04.016
23. Zhang H, Sun D, Li D, et al. Long non-coding RNA expression patterns in lung tissues of chronic cigarette smoke induced COPD mouse model. *Sci Rep*. 2018;8(1):7609. doi:10.1038/s41598-018-25702-3
24. Transdisciplinary Tobacco Use Research Center (TTURC) Tobacco Dependence TB, Baker TB, Piper ME, et al. Time to first cigarette in the morning as an index of ability to quit smoking: implications for nicotine dependence. *Nicotine Tob Res*. 2007;9 Suppl 4(Suppl 4):S555-70. doi:10.1080/14622200701673480
25. Stelmach R, Fernandes FLA, Carvalho-Pinto RM, et al. Comparison between objective measures of smoking and self-reported smoking status in patients with asthma or COPD: are our patients telling us the truth? *J Bras Pneumol*. 2015;41(2):124-132. doi:10.1590/S1806-37132015000004526
26. Sponsiello-Wang Z, de La Bourdonnaye G, David M, Lüdicke F, Weitkunat R. Accuracy of the Smoking Questionnaire. *Beiträge zur Tab Int to Tob Res*. 2017;27(8):224-239. doi:10.1515/cttr-2017-0023
27. Hackett NR, Butler MW, Shaykhiev R, et al. RNA-Seq quantification of the human small airway epithelium transcriptome. *BMC Genomics*. 2012;13(1):82. doi:10.1186/1471-2164-13-82
28. SRNT Subcommittee on Biochemical Verification NL, Jacob P, Ahijevych K, et al. Biochemical verification of tobacco use and cessation. *Nicotine Tob Res*. 2002;4(2):149-159. doi:10.1080/14622200210123581
29. Baltar VT, Xun WW, Chuang S-C, et al. Smoking, secondhand smoke, and cotinine levels in a subset of EPIC cohort. *Cancer Epidemiol Biomarkers Prev*. 2011;20(5):869-875. doi:10.1158/1055-9965.EPI-10-1235
30. Hald J, Overgaard J, Grau C. Evaluation of Objective Measures of Smoking Status A Prospective Clinical Study in a Group of Head and Neck Cancer Patients Treated with Radiotherapy. *Acta Oncol (Madr)*. 2003;42(2):154-159. doi:10.1080/02841860310005020
31. Dedeurwaerder S, Defrance M, Calonne E, Denis H, Sotiriou C, Fuks F. Evaluation of the Infinium Methylation 450K technology. *Epigenomics*. 2011;3(6):771-784. doi:10.2217/epi.11.105
32. Chen YA, Lemire M, Choufani S, et al. Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray. *Epigenetics*. 2013;8(2):203-209. doi:10.4161/epi.23470

33. Moran S, Arribas C, Esteller M. Validation of a DNA methylation microarray for 850,000 CpG sites of the human genome enriched in enhancer sequences. *Epigenomics*. 2016;8(3):389-399. doi:10.2217/epi.15.114
34. Wouters EFM, Wouters. Chronic obstructive pulmonary disease * 5: Systemic effects of COPD. *Thorax*. 2002;57(12):1067-1070. doi:10.1136/thorax.57.12.1067
35. Chung KF, Adcock IM. Multifaceted mechanisms in COPD: inflammation, immunity, and tissue repair and destruction. *Eur Respir J*. 2008;31(6):1334-1356. doi:10.1183/09031936.00018908
36. Rahman I, Morrison D, Donaldson K, MacNee W. Systemic oxidative stress in asthma, COPD, and smokers. *Am J Respir Crit Care Med*. 1996;154(4):1055-1060. doi:10.1164/ajrccm.154.4.8887607
37. NOGUERA A, BUSQUETS X, SAULEDA J, VILLAVERDE JM, MacNEE W, AGUSTÍ AGN. Expression of Adhesion Molecules and G Proteins in Circulating Neutrophils in Chronic Obstructive Pulmonary Disease. *Am J Respir Crit Care Med*. 1998;158(5):1664-1668. doi:10.1164/ajrccm.158.5.9712092
38. Houseman EA, Kim S, Kelsey KT, Wiencke JK. DNA Methylation in Whole Blood: Uses and Challenges. *Curr Environ Heal Reports*. 2015;2(2):145-154. doi:10.1007/s40572-015-0050-3
39. Jaffe AE, Irizarry RA. Accounting for cellular heterogeneity is critical in epigenome-wide association studies. *Genome Biol*. 2014;15(2):R31. doi:10.1186/gb-2014-15-2-r31
40. Houseman EA, Accomando WP, Koestler DC, et al. DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics*. 2012;13(1):86. doi:10.1186/1471-2105-13-86
41. Tennessen JA, Bigham AW, O'Connor TD, et al. Evolution and Functional Impact of Rare Coding Variation from Deep Sequencing of Human Exomes. *Science (80-)*. 2012;337(6090):64-69. doi:10.1126/science.1219240
42. Jaffe AE, Murakami P, Lee H, et al. Bump hunting to identify differentially methylated regions in epigenetic epidemiology studies. *Int J Epidemiol*. 2012;41(1):200-209. doi:10.1093/ije/dyr238
43. Pedersen BS, Schwartz DA, Yang I V., Kechris KJ. Comb-p: software for combining, analyzing, grouping and correcting spatially correlated P-values. *Bioinformatics*. 2012;28(22):2986-2988. doi:10.1093/bioinformatics/bts545
44. Ortega VE. Picking the Right Fruit: Intersecting Chronic Obstructive Pulmonary Disease Genome-Wide Association Study Discoveries with Epigenetics. *Am J Respir Crit Care Med*. 2018;197(10):1237-1239. doi:10.1164/rccm.201801-0084ED
45. Silverman EK, Loscalzo J. Network medicine approaches to the genetics of complex diseases. *Discov Med*. 2012;14(75):143-152. http://www.ncbi.nlm.nih.gov/pubmed/22935211. Accessed November 29, 2018.
46. Gligorijević V, Malod-Dognin N, Pržulj N. Integrative methods for analyzing big data in precision medicine. *Proteomics*. 2016;16(5):741-758. doi:10.1002/pmic.201500396
47. Agustí A, Antó JM, Auffray C, et al. Personalized Respiratory Medicine: Exploring the Horizon, Addressing the Issues. Summary of a BRN-AJRCCM Workshop Held in Barcelona on June 12, 2014. *Am J Respir Crit Care Med*. 2015;191(4):391-401. doi:10.1164/rccm.201410-1935PP
48. Silverman EK, Loscalzo J. Developing new drug treatments in the era of network medicine. *Clin Pharmacol Ther*. 2013;93(1):26-28. doi:10.1038/clpt.2012.207

Chapter 5

Summary/Samenvatting

SUMMARY

COPD is the most common respiratory disease and a third leading cause of death worldwide. It is a complex disease with overlapping sub-phenotypes and epidemiological studies suggest that COPD is a consequence of the combination of cumulative exposure to external and internal factors and their interactions. Therefore, in order to understand the pathophysiology of COPD, I performed integrative studies to empower the development of preventive, diagnostic and therapeutic tools. The main aim of this thesis is to gain insight in the biological mechanisms by studying the functional role and interaction of multiple omics layers.

In **Chapter 2.1** I describe a genome-wide linkage scan, in search for rare genetic variants which have a role in familial COPD. I show that there is a significant linkage of genomic regions in chromosomes 5, 11 and 15 and identify novel rare variants in chromosome 11 region, shared among the COPD cases of the family (in *SLC22A11* and *MTL5*) which play a role in the disease. This study shows the importance of studying rare variants, which may have large effects, even if they are population specific. In **Chapter 2.2** I present a study, which investigates a functional role of chromosome 15q25 region variants, one of the most replicated COPD GWAS loci. I show that variants in *IREB2, HYKK* and *CHRNA3* genes exert effects on DNA methylation in blood, relevant to COPD, and gene expression in lung tissue. Similarly, **Chapter 2.3** I investigate a top variant from novel locus on 19q13 (*EGLN2* gene), identified in COPD GWAS, and show that genetic variations underlying *EGLN2* DNA methylation contribute to the risk of developing COPD, by mediating the genetic effects on *EGLN2* expression. The associations identified in **Chapters 2.2 and 2.3** are independent of smoking, which puts forward a genetic driven pathway of DNA methylation implicated in COPD which may be used as a target for a more personalized and focused treatment approach. In **Chapter 2.4** I describe a study of lung function levels in relation to DNA methylation in never-smokers and show a specific DNA methylation pattern, again independent of smoking.

In **Chapter 3.1** I describe a large meta-analysis of epigenome-wide DNA methylation studies of lung function at birth and show that 59 DMRs associate with lung function at birth of which some also affect lung function, asthma and COPD in later life. This study goes in line with the hypothesis that DNA methylation landscape at birth can predispose a person for COPD in later life. In **Chapter 3.2** I study circulating metabolites in relation with COPD and FEV_1/FVC and show importance of Glycoprotein acetyls (GlycA). The MR results show, for the first time, that higher levels of GlycA are a pre-diagnostic biomarker of COPD. The results suggest that COPD shows a specific metabolomics signature related to systemic inflammation. This

study underlines that metabolomics studies are important to provide a rationale for innovative personalized treatments in patients with COPD. In **Chapter 3.3** I briefly highlight significant genetic overlap between COPD and several comorbid diseases. COPD was not only positively correlated with pulmonary comorbidity, but also to various cardio-metabolic diseases (carotid artery disease, heart attack, hypertension, and diabetes), psychiatric diseases (depression, attention deficit hyperactivity disorder and schizophrenia) and autoimmune intestinal diseases. I further study the epigenetic mechanisms which play a role in the COPD and depression and may explain their co-occurrence. In **Chapter 3.4** I further investigate DNA methylation patterns specific for depression in a largest to date EWAS study in Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) consortium. I identify three DNA methylation sites associated with depressive symptoms, which point towards the axon guidance as the common disrupted pathway in depression. One of the genes differentially methylated in depression *(CDC42BPB)* has also been implicated in COPD.

Finally, in the **Chapter 4** I discuss the main findings of this thesis, challenges, possible clinical implications and future directions. In general, the findings of this thesis identify and improve the understanding of the role of specific risk factors for COPD and hopefully highlight the importance of combining the knowledge on multiple omics layers to investigate such a complex and important disease. I hope that such studies can lead to the development of better and more precise prevention of COPD.

SAMENVATTING

COPD (Chronische Obstructieve Longziekte) is de meest voorkomende longziekte en wereldwijd komt COPD op de derde plaats van ziekten die de hoogste sterfte veroorzaken. Het is een complexe ziekte met overlappende subfenotypen en epidemiologische studies suggereren dat COPD een gevolg is van de combinatie van cumulatieve blootstelling aan externe en interne factoren en hun interacties. Daarom heb ik, om de pathofysiologie van COPD te begrijpen, integratieve studies uitgevoerd om de ontwikkeling van preventieve, diagnostische en therapeutische hulpmiddelen mogelijk te maken. Het belangrijkste doel van dit proefschrift is om inzicht te krijgen in de biologische mechanismen door het bestuderen van de functionele rol en interactie van meerdere omics-lagen.

In **Hoofdstuk 2.1** beschrijf ik een genoomwijde linkage scan, waarbij ik op zoek ging naar zeldzame genetische varianten die een rol spelen bij familiaire COPD. Ik laat zien dat er een significante koppeling van genomische gebieden op chromosomen 5, 11 en 15 is. Daarnaast heb ik aangetoond dat nieuwe zeldzame varianten op chromosoom 11 (in *SLC22A11* en *MTL5*) een rol spelen in de ziekte. Deze zelzame varianten komen voor bij alle COPD patiënten van de familie die ik heb onderzocht. Deze studie toont hiermee het belang aan van het bestuderen van zeldzame varianten die grote effecten kunnen hebben, zelfs als ze populatiespecifiek zijn. In **Hoofdstuk 2.2** laat ik een studie zien waarin ik een functionele rol onderzoek van varianten op chromosoom 15q25, een van de meest gerepliceerde COPD genoomwijde associatie studies (GWAS) loci. Ik laat zien dat varianten in *IREB2*, *HYKK* en *CHRNA3* genen effect hebben op de DNA-methylatie in het bloed, wat relevant is voor COPD en genexpressie in longweefsel. Evenzo onderzoek ik in **Hoofdstuk 2.3** een topvariant van een nieuw gevonden locus op 19q13 (*EGLN2* gen) in een COPD GWAS, en laat zien dat genetische variaties die ten grondslag liggen aan DNA-methylatie in *EGLN2* bijdragen aan het risico op het ontwikkelen van COPD, door hun rol als mediators op de genetische effecten op EGLN2-expressie. Vervolgens laat ik in **Hoofdstukken 2.2** en **2.3** zien dat de gevonden associaties onafhankelijk zijn van roken, wat duidt op een pathway gedreven door genetische varianten die belangrijk zijn bij DNA methylatie in COPD. Deze bevinding kan mogelijk gebruikt worden voor een meer gepersonaliseerde en gerichte behandeling. In **Hoofdstuk 2.4** beschrijf ik een studie van longfunctieniveaus in relatie tot DNA-methylatie bij niet-rokers en laat ik een specifiek DNA-methylatiepatroon zien, wederom onafhankelijk van roken.

In **Hoofdstuk 3.1** beschrijf ik een grote meta-analyse van epigenoomwijde DNA-methylatiestudies van de longfunctie bij de geboorte. Ik laat zien dat er 59 gebieden zijn met verschillen in DNA methylatie die associëren met de longfunctie

bij de geboorte. Hiervan beïnvloeden sommige ook de longfunctie, astma en COPD op latere leeftijd. Deze studie sluit aan bij de hypothese dat het DNA-methylatie landschap bij de geboorte een persoon vatbaarder kan maken voor het ontwikkelen van COPD op latere leeftijd. In **Hoofdstuk 3.2** onderzoek ik de relatie tussen circulerende metabolieten in het bloed en COPD en FEV_1/FVC en toon ik het belang aan van Glycoproteïne acetyls (GlycA). De resultaten van de Mendeliaanse randomisatie studie laten voor het eerst zien dat hogere GlycA concentraties een pre-diagnostische biomarker van COPD zijn. De resultaten suggereren dat COPD een specifieke metabolomische signatuur vertoont die samenhangt met systemische ontsteking. Deze studie benadrukt dat metabolomische onderzoeken belangrijk zijn voor innovatieve gepersonaliseerde behandelingen bij patiënten met COPD. In **Hoofdstuk 3.3** belicht ik kort een significante genetische overlap tussen COPD en verschillende comorbide aandoeningen. COPD was niet alleen positief gecorreleerd met pulmonale comorbiditeit, maar ook met verschillende cardio-metabole ziekten (halsslagaderziekte, hartaanval, hypertensie en diabetes), psychiatrische ziekten (depressie, attention deficit hyperactivity disorder en schizophrenia) en auto-immuunziekten van het gastro-intestinale stelsel. Verder bestudeer ik in dit hoofdstuk de epigenetische mechanismen die een rol spelen bij COPD en depressie en die mogelijk hun gelijktijdig voorkomen verklaren. In **Hoofdstuk 3.4** onderzoek ik DNA-methylatiepatronen specifiek voor depressie in de tot nu toe grootste epigenoom-wijde associatie studie in het CHARGE (Cohorten voor Hart- en Verouderingsonderzoek in Genomische Epidemiologie) consortium. In deze studie identificeer ik drie DNA-methylatie posities die geassocieerd zijn met depressieve symptomen. Dit wijst erop dat de geleiding van axonen vaak verstoord is bij depressie. Daarnaast is één van de genen met verschillen in DNA methylatie status in depressie (*CDC42BPB*) ook betrokken bij COPD.

Tot slot bespreek ik in **Hoofdstuk 4** de belangrijkste bevindingen van dit proefschrift, de uitdagingen, de mogelijke klinische implicaties en de toekomstige richtingen. In het algemeen vergroten en verbeteren de bevindingen van dit proefschrift ons begrip in de rol van specifieke risicofactoren voor COPD. Hopelijk benadrukt dit het belang van het combineren van de kennis verkregen via meerdere omicslagen bij het onderzoeken van deze complexe en belangrijke ziekte. Ik hoop dat zulke studies tot de ontwikkeling van een betere en preciezere preventie van COPD kunnen leiden.

Chapter 6

Appendix

6.1 ACKNOWLEDGEMENTS

"The truth is rarely pure and never simple" - Oscar Wilde

Likewise, my quest for finding the truth was not simple (nor over), but was facilitated by the support of many people. With a big sense of accomplishment, gratitude and a bit of relief I am writing this chapter to thank all those people without whose help I would not be where I am today.

Firstly, I appreciate **Bojana Matejić** who introduced me to the scientific world by saying: „There is this program called SPSS". Thank you for your help and understanding when I needed it the most. I feel like you enabled me to accomplish everything that followed and I am very grateful for this.

My promotor **prof. Cornelia van Duijn**, your critical thinking and power of deduction are very inspiring. I am thankful for getting the opportunity to work with you in the Genepi group. Thank you for embarking upon an unknown path of respiratory diseases with me and for pushing me to the limits.

My co-promotor **Najaf Amin**, thank you for being my supervisor and always having the patience to answer my questions when I come to your office. And for loving my cakes, it's always a pleasure to make them!

The reading committee, thank you for accepting to be a part of it and for reading my thesis. Dear **prof. Guy Brusselle**, your passion for science is very inspiring and I really appreciate your input and discussions over the years. Dear **prof. Marike Boezen**, thank you for many fruitful meetings, input, help and support. I have learned a lot from you. Dear **prof. Joyce van Meurs**, thank you for initiating useful scientific discussions at the department and for your passion for teaching. To the rest of the committee, thank you for accepting the role and making time for the defense day.

My "partners in crime", moji sinovi **Desi** and **Jelena**, without you and our after-work brainstorming/complaining sessions (and caffetin) I would probably never survive the cold Dutch days, let alone finish the PhD. You are my "friends, but family", the world cannot separate us! Thank you for paranymph-ing and always being there for me. Also, thanks to **Jovan** and **Ray** for making you happier. Looking forward for a lifetime together with you!

To **Ashley**, my favorite Duchie, we've been next to each other since day one, battling similar battles. Thank you for your patience, translations, explanations and answers to all my questions about the "Dutch ways". Also a special thank you for being my paranymph! I hope to see you soon at this stage and for many more playdates in the neighborhood.

To the rest of the Genepi group – **Dina**, thank you for your help and patience over the years, you've been a true scientific role-model to me; **Dream**, my little Chinese sister, thank you for bringing joy and laughter to our office, I hope you will be very successful and satisfied with your thesis; **Adriana**, it has been a pleasure to hang out with you in and out of the office, I am hoping for many more garden chill-outs; **Shahzad**, thank you for fruitful discussions and for your patience with three loud girls in the office; **Hata**, you are very polite and diligent and great addition to the group; **Sven** we missed your unique ways since you left; **Ayse**, thank you for always being easy going and approachable, I hope you have a nice time in the UK; **Maaike**, thank you for your help and kind approach; **Elisa**, I wish you and your family all the best; and last but not least, to the colleagues from the lab, **Andy**, **Jeannette**, **Andrea** and **Bernadette**, we all owe our thanks to you.

To the rest of the Groningen group – dear **Diana, Kim, Cleo, Maaike, Judith, Ana Julia, prof. Dirkje Postma** thank you for nice collaboration, brainstorming and your input. Also, thank you for your company and support at the ERS congress.

Dear **Kate**, **Manuela**, **Gogo**, **Niko**, **Elena**, **Giorgia**, **Ruchika**, **Nemanja**, **Julijana**, I feel very lucky to have met you and consider you my friends. I miss you and hope to cross your path again soon.

To all other colleagues from the EMC Epidemiology department and all my co-authors - thank you – **Fjorda, Pooja, Eralda, Jinluan, Marten, Lies, Natalie, Emmely, Eliza, Mirjana, Gavro, Tamara, Nevena, Kozeta, Adela, Natalia, Kim, Leila, Martijn, Olivera, Rosa, Hieab, Mohsen, Peter, Tijana, Carolina, Vanja, Irma, Olja, Zorana, Milan T., Milan S., Strahinja, Ryan, Ana, Trudy, Jannine, Henning, Oscar, Arfan** – thank you for many discussions, meetings, collaborations, coffee breaks and chats at the department.

Nano, Solange, Lennie, Astrid and Frank, each of you played an important role in my PhD, thank you for your help.

To my Spanish family – **Jim, JM, Peter, Daniel, Josep, Effie** - thank you for the hospitality and for so much great food! I wish you good health, success and collaboration! ;) To **prof. Josep Roca**, thank you for the opportunity to join you at the department and for opening new opportunities for me.

To our group therapy squad throughout the faculty days and beyond - **Danka, Sanja, Sole, Boka, Goga, Nataša, Pavke, Sale, Stefi, Blazo –** thank you for your support, many visits in lonely days, for always being available and happy to meet for a coffee'n'update afternoons and for being able to share all our happy and sad news. **Danka**, thank you for our endless conversations over the phone and in person, for always having my back and supporting me and for my beautiful godson, Teo.

To **Mihailo, Goxy, Glišo, Marija, Čuka, Tina, Marijana, Marina and Bogdan -** some friendships cannot be broken by distance or time. Thank you for always making us feel like we never left. I hope to never lose you.

Kata and **Doda,** I am very pleased to have you in my life. You should know I will always be there for you!

Milisavi thank you for your love and support throughout my whole life.

To my **family in law**, thank you for accepting me as an equal member of the Prokić family.

To **my siblings,** thank you for enriching my life and for always being there for me, I am missing you every day. **Igor** and **Ana**, thank you for my beautiful nephew. **Joka**, you are my role-model and the voice in my head. I don't know a better person and I am proud to be your little sister. Thank you and **Marko** for being a great team and teaching us how to be great parents. **Makice**, it's so nice to have a little brother and see how he grows to be a good and strong person. You are my protection and sound mind, always giving the best advice. I wish you all the success in the world. **Boba**, thank you for making my brother a happier and better man. Also, thank you for the beautiful cover, one day you will be famous!

Petra, Jana and **Luka** thank you for being such wonderful, beautiful, smart kids, for the unconditional love and for making me a proud aunt.

To **my parents**, I can only say thank you and I love you. I know you are proud of what I have become, how I chose my path and what lays ahead for me. But, in truth, I can credit all this only to you, your rightful upbringing, unconditional love and selfless sacrifice. My wish is to always be there for you and never let you down.

To my husband, **Petar**, our Prokić team is the reason I feel I can do anything I set my mind to, knowing I have my rock and foundation to lean on. You've always been such a good influence and support throughout the years and I can't wait to share every next experience with you. Thank you for our beautiful daughter. I love you.

To my daughter, **Lenka**, you are such a calm and wonderful baby which allowed me to finish writing this book. So, thank you, this book would definitely not exist without you. Having you in my life gave me the final push I needed when it was the hardest. I hope to one day inspire you to never give up and push yourself to the limits for the things you care about. The immeasurable love and joy of having you shifted my perspectives and changed my life completely. I am excited about every new day and to experience every new milestone with you.

Ivana Prokić

Rotterdam

6.2 ABOUT THE AUTHOR

Ivana Prokić (born Nedeljković), was born in Belgrade, Serbia on 25th of June 1988, as a middle child of Anđelka and Miroslav Nedeljković. She completed her secondary education in the First Belgrade Gymnasium in 2007 and in the same year started her medical studies at the University of Belgrade, School of Medicine in Belgrade, Serbia. In 2013 she finished her medical studies and received the ERAWEB scholarship for continuing her studies towards the Master of Science degree. She came to Rotterdam in August 2013 and started the master in Genetic epidemiology at the Netherlands Institute for Health Sciences (NIHES), under the supervision of prof. Cornelia van Duijn at the Genetic epidemiology unit of Erasmus Medical Center. In August 2014 she received a second ERAWEB scholarship to continue pursuing her PhD and for her thesis she also finished the Doctor of Science program at NIHES. During her PhD studies Ivana received the Marie Skłodowska-Curie scholarship (MCSA-RISE) and worked in LinkCare Health Services in Barcelona, Spain. Ivana will continue her work on respiratory diseases aiming to connect the research, clinical practice and industry. Since May 2017 she is wife of Petar Prokić and since September 2018 mother of Lenka Prokić.

6.3 PHD PORTFOLIO

Name PhD student: Ivana Prokić
Erasmus MC department: Epidemiology
Research school: Netherlands Institute of Health Sciences (NIHES)
PhD period: August 2013 – December 2017
Promotor: Prof. C.M.van Duijn
Copromotor: Dr. N. Amin

Education	Year	ECTS*
Master of Science in Genetic epidemiology		
Core Curriculum		
Study Design	2013	4.3
Biostatistical Methods I: Basic Principles	2013	5.7
Biostatistical Methods II: Clasical Regression Models	2013	4.3
Genetic-Epidemiologic Research Methods	2013	5.1
Linux for Scientists	2013	0.6
Advanced Short Courses		
Bayesian Statistics	2014	1.4
Women's Health	2014	0.9
Advances in Genome-Wide Association Studies	2014	1.4
Family-based Genetic Analysis	2014	1.4
Mendelian Randomization	2014	0.9
A first encounter with next-generation sequencing data	2014	1.4
Skills Courses		
English Language	2013	1.4
Introduction to Medical Writing	2014	1.1
Courses for the Quantitative Researcher	2014	1.4
Erasmus Summer Programme		
Principles of Research in Medicine	2013	0.7
Clinical Decision Analysis	2013	0.7
Methods of Public Health Research	2013	0.7
Health Economics	2013	0.7
Genome Wide Association Analysis	2014	1.4
Conceptual Foundation of Epidemiologic Study Design	2013	0.7
Principles of Genetic Epidemiology	2013	0.7
Genomics in Molecular Medicine	2014	1.4
Markers and Prognostic Research	2013	0.7
Advances in Genomics Research	2014	0.4
The Practice of Epidemiologic Analysis	2013	0.7

Education	Year	ECTS*
Doctor of Science in Genetic epidemiology		
Causal Inference	2015	0.7
History of Epidemiologic Ideas	2015	0.7
Advances in Epidemilogic Analysis	2015	0.4
Causal Mediation Analysis	2015	0.7
Psychopharmacology	2015	1.4
Missing Values in Clinical Research	2015	0.7
Principles of Epidemiologic Data Analysis	2015	0.7
Qualitative Research	2015	1.4
Advanced Genetic & Omics Data Analysis 2	2014	0.7
Course Biomedical Research Techniques XIII	2014	1.5
NGS in DNA Diagnostics Course	2014	1.0
Research Integrity	2016	0.3
Attended Seminars and Meetings		
Seminars at the department of Epidemiology, Erasmus MC	2013-2017	4.5
Genetic epidemiology working group meetings, Erasmus MC	2013-2017	4.5
PhD students of Epidemiology department seminars (2020 meetings), ErasmusMC	2015-2017	2
Longfonds consortium meetings	2014-2017	3
BBMRI-NL consortium meetings	2016-2017	2
CHARGE consortium meetings	2014-2017	4
GIANT consortium meetings	2013-2014	2
Presentations at Conferences and Meetings		
Epigenetics of common diseases conference (Cambridge, UK) – oral	2015	1
BIOS symposium (Amsterdam, the Netherlands) – oral	2016	0.5
European Respiratory Society Congress (Milan, Italy) – oral	2017	1
CHARGE consortium meeting (New York, USA) – poster	2017	0.8
ERS Satellites: Advances in precision medicine in COPD & ILD	2019	1
Awards and scholarships		
ERAWEB I – scholarship for Master of Science program	2013	
ERAWEB II– scholarship for PhD	2014	
Travel award Epigenetics of common diseases conference	2015	
Marie Skłodowska-Curie Research and Innovation Staff Exchange	2017	
Teaching		
Teaching assistant for Principles of Genetic Epidemiology course	2016	0.5
Other		
Peer review of articles for Epigenetics journal	2018	

*ECTS (European Credit Transfer System) equals a workload of 28 hours.

LIST OF PUBLICATIONS

1. **DNA Methylation Signatures of Depressive Symptoms in Middle-aged and Elderly Persons**, *JAMA Psychiatry*, vol. 75, no. 9, p. 949, Sep. 2018
O. Story Jovanova*, **I. Nedeljković***, D. Spieler*, R. M. Walker*, C. Liu*, M. Luciano, J. Bressler*, J. Brody, A. J. Drake, K. L. Evans, R. Gondalia, S. Kunze, B. Kuhnel, J. Lahti, R. N. Lemaitre, R. E. Marioni, B. Swenson, J. J. Himali, H. Wu, Y. Li, A. F. McRae, T. C. Russ, J. Stewart, Z. Wang, G. Zhang, K.-H. Ladwig, A. G. Uitterlinden, X. Guo, A. Peters, K. Räikkönen, J. M. Starr, M. Waldenberger, N. R. Wray, E. A. Whitsel, N. Sotoodehnia, S. Seshadri, D. J. Porteous, J. van Meurs, T. H. Mosley, A. M. McIntosh, M. M. Mendelson, D. Levy, L. Hou, J. G. Eriksson, M. Fornage, I. J. Deary, A. Baccarelli, H. Tiemeier, and N. Amin.

2. **From blood to lung tissue: effect of cigarette smoke on DNA methylation and lung function**, *Respir. Res.*, vol. 19, no. 1, p. 212, Dec. 2018
M. de Vries, D. A. van der Plaat**, I. Nedeljković**, R. N. Verkaik-Schakel, W. Kooistra, N. Amin, C. M. van Duijn, C.-A. Brandsma, C. C. van Diemen, J. M. Vonk, and H. Marike Boezen.

3. **A genome-wide linkage study for chronic obstructive pulmonary disease in a dutch genetic isolate identifies novel rare candidate variants**, *Front. Genet.*, vol. 9, no. APR 2018
I. Nedeljković, N. Terzikhan, J. M. Vonk, D. A. van der Plaat, L. Lahousse, C. C. van Diemen, B. D. Hobbs, D. Qiao, M. H. Cho, G. G. Brusselle, D. S. Postma, H. M. Boezen, C. M. van Duijn, and N. Amin.

4. **Life-Course Genome-wide Association Study Meta-analysis of Total Body BMD and Assessment of Age-Specific Effects**, *Am. J. Hum. Genet.*, vol. 102, no. 1, pp. 88–102, 2018
C. Medina-Gomez, J. P. Kemp, K. Trajanoska, J. Luan, A. Chesi, T. S. Ahluwalia, D. O. Mook-Kanamori, A. Ham, F. P. Hartwig, D. S. Evans, R. Joro, **I. Nedeljković**, H. F. Zheng, K. Zhu, M. Atalay, C. T. Liu, M. Nethander, L. Broer, G. Porleifsson, B. H. Mullin, S. K. Handelman, M. A. Nalls, L. E. Jessen, D. H. M. Heppe, J. B. Richards, C. Wang, B. Chawes, K. E. Schraut, N. Amin, N. Wareham, D. Karasik, N. Van der Velde, M. A. Ikram, B. S. Zemel, Y. Zhou, C. J. Carlsson, Y. Liu, F. E. McGuigan, C. G. Boer, K. Bønnelykke, S. H. Ralston, J. A. Robbins, J. P. Walsh, M. C. Zillikens, C. Langenberg, R. Li-Gao, F. M. K. Williams, T. B. Harris, K. Akesson, R. D. Jackson, G. Sigurdsson, M. den Heijer, B. C. J. van der Eerden, J. van de Peppel, T. D. Spector, C. Pennell, B. L. Horta, J. F. Felix, J. H. Zhao, S. G. Wilson, R. de Mutsert, H. Bisgaard, U. Styrkársdóttir, V. W. Jaddoe, E. Orwoll, T. A. Lakka, R. Scott, S. F. A. Grant, M. Lorentzon, C. M. van Duijn, J. F. Wilson, K. Stefansson, B. M. Psaty, D. P. Kiel, C. Ohlsson, E. Ntzani, A. J. van Wijnen, V. Forgetta, M. Ghanbari, J. G. Logan, G. R. Williams, J. H. D. Bassett, P. I. Croucher, E. Evangelou, A. G. Uitterlinden, C. L. Ackert-Bicknell, J. H. Tobias, D. M. Evans, and F. Rivadeneira.

5. **Long-term Air Pollution Exposure, Genome-wide DNA Methylation and Lung Function in the LifeLines Cohort Study**, *Environ. Health Perspect.*, vol. 126, no. 02, Feb. 2018
A. J. de F.C. Lichtenfels, D. A. van der Plaat, K. de Jong, C. C. van Diemen, D. S. Postma, **I. Nedeljković**, C. M. van Duijn, N. Amin, S. la Bastide-van Gemert, M. de Vries, C. K. Ward-Caviness, K. Wolf, M. Waldenberger, A. Peters, R. P. Stolk, B. Brunekreef, H. M. Boezen, and J. M. Vonk.

6. **Occupational exposure to pesticides is associated with differential DNA methylation,** *Occup. Environ. Med.*, p. oemed-2017-104787, Feb. 2018
D. A. van der Plaat, K. de Jong, M. de Vries, C. C. van Diemen, **I. Nedeljković**, N. Amin, H. Kromhout, R. Vermeulen, D. S. Postma, C. M. van Duijn, H. M. Boezen, and J. M. Vonk.

7. **Understanding the role of the chromosome 15q25.1 in COPD through epigenetics and transcriptomics**, *Eur. J. Hum. Genet.*, pp. 1–14, Feb. 2018
I. Nedeljković, E. Carnero-Montoro, L. Lahousse, D. A. van der Plaat, K. de Jong, J. M. Vonk, C. C. van Diemen, A. Faiz, M. van Den Berge, M. Obeidat, Y. Bossé, D. C. Nickle, B. Consortium, A. G. Uitterlinden, J. J. B. van Meurs, B. C. H. Stricker, G. G. Brusselle, D. S. Postma, H. M. Boezen, C. M. van Duijn, and N. Amin.

8. **COPD GWAS variant at 19q13.2 in relation with DNA methylation and gene expression**, *Hum. Mol. Genet.*, vol. 27, no. 2, pp. 396–405, Jan. 2017
I. Nedeljković, L. Lahousse, E. Carnero-Montoro, A. Faiz, J. M. Vonk, K. de Jong, D. A. van der Plaat, C. C. van Diemen, M. van den Berge, M. Obeidat, Y. Bossé, D. C. Nickle, B. I. O. S. Consortium, A. G. Uitterlinden, J. B. J. van Meurs, B. H. C. Stricker, G. G. Brusselle, D. S. Postma, H. M. Boezen, C. M. van Duijn, and N. Amin.

9. **Genome-wide association study on the FEV_1/FVC ratio in never-smokers identifies HHIP and FAM13A**, *J. Allergy Clin. Immunol.*, vol. 139, no. 2, pp. 533–540, Feb. 2017
D. A. van der Plaat, K. de Jong, L. Lahousse, A. Faiz, J. M. Vonk, C. C. van Diemen, **I. Nedeljković**, N. Amin, G. G. Brusselle, A. Hofman, C. A. Brandsma, Y. Bossé, D. D. Sin, D. C. Nickle, C. M. van Duijn, D. S. Postma, and H. M. Boezen.

10. **The well-known gene HHIP and novel gene MECR are implicated in small airway obstruction**, *Am. J. Respir. Crit. Care Med.*, vol. 194, no. 10, pp. 1299–1302, Nov. 2016
D. A. Van Der Plaat, K. De Jong, L. Lahousse, A. Faiz, J. M. Vonk, C. C. Van Diemen, **I. Nedeljković**, N. Amin, M. Obeidat, C. M. Van Duijn, H. Marike Boezen, and D. S. Postma.

11. **Novel Genetic Susceptibility Loci for FEV_1 in the Context of Occupational Exposure in Never-Smokers**, *Am. J. Respir. Crit. Care Med.*, vol. 194, no. 6, pp. 769–772, Sep. 2016
K. de Jong, J. M. Vonk, A. Faiz, D. A. van der Plaat, W. Timens, Y. Bossé, H. Kromhout, **I. Nedeljković**, D. S. Postma, and H. M. Boezen.

12. **Novel DNA methylation sites of glucose and insulin homeostasis: an integrative cross-omics analysis.** *bioRxiv*. October 2018:432070. doi:10.1101/432070
J. Liu, E. Carnero-Montoro, J. van Dongen, S. Lent, **I. Nedeljković**, S. Ligthart, P.C. Tsai, T. C. Martin, P. R. Mandaviya, R. Jansen, M. J. Peters, L. Duijts, V. W.V. Jaddoe, H. Tiemeier, J. F. Felix, A. Y. Chu, D. Levy, S. J. Hwang, J. Bressler, R. Gondalia, E. L. Salfati, C. Herder, B. A. Hidalgo, T. Tanaka, A. Zenobia Moore, R. N. Lemaitre, M. A. Jhun, J. A. Smith, N. Sotoodehnia, S. Bandinelli, L. Ferrucci, D. K. Arnett, H. Grallert, T. L. Assimes, L. Hou, A. Baccarelli, E. A. Whitsel, K. Willems van Dijk, N. Amin, A. G. Uitterlinden, E. J.G. Sijbrands, O. H. Franco, A. Dehghan, T. D. Spector, J. Dupuis, M. F. Hivert, J. I. Rotter, J. B. Meigs, J. S. Pankow, J. B.J. van Meurs, A. Isaacs, D. I. Boomsma, J. T. Bell, A. Demirkan, C. M. van Duijn.

13. **Newborn DNA-methylation, childhood lung function, and the risks of asthma and COPD across the life course.** *Eur Respir J*. February 2019:1801795. doi:10.1183/13993003.01795-2018
H. T. den Dekker*, K. Burrows*, J. F. Felix*, L. A. Salas, **I. Nedeljković**, J. Yao, S. L. Rifas-Shiman, C. Ruiz-Arenas, N. Amin, M. Bustamante, D. L. DeMeo, A. John Henderson, C. G. Howe, M. F. Hivert, M. A. Ikram, J. C. de Jongste, L. Lahousse, P. R. Mandaviya, J. B. van Meurs, M. Pinart, G. C. Sharp, L. Stolk, A. G. Uitterlinden, J. M. A., Augusto A. Litonjua, C. V. Breton, G. G. Brusselle, J. Sunyer, G. Davey Smith, C. L. Relton[#], V. W.V. Jaddoe[**], L. Duijts.[**]

14. **A cross-omics integrative study of metabolomic signatures of Chronic obstructive pulmonary disease.** *Submitted for publication.*
I. Nedeljković, L. Lahousse, M. de Vries, J. Liu, M. Kalaoja, J. M. Vonk, D. A. van der Plaat, C. C. van Diemen, A. van der Spek, A. Zhernakova, J. Fu, M. Ghanbari, M. Ala-Korpela, J. Kettunen, A. S. Havulinna, M. Perola, V. Salomaa, L. Lind, J. Ärnlöv, B. H.C. Stricker, G. G. Brusselle, H. M. Boezen, C. M. van Duijn*, N. Amin*.

15. **DNA methylation is associated with lung function in never-smokers.** *Submitted for publication.*
M. de Vries, **I. Nedeljković**, D. van der Plaat, A. Zhernakova, L. Lahousse, G. G. Brusselle, BIOS Consortium, N. Amin, C. M. van Duijn, J. M. Vonk, H. M. Boezen.

16. **Limited overlap in significant hits between genome-wide association studies on two airflow obstruction definitions in the same population.** *Submitted for publication.*
D. A. van der Plaat, J. M. Vonk, L. Lahousse, K. de Jong, A. Faiz, **I. Nedeljković**, N. Amin, C. C. van Diemen, G. G. Brusselle, Y. Bossé, CA, K. Hao, P. D. Paré, C. M. van Duijn, D. S. Postma, H. M. Boezen.

17. **Occupational exposure to gases/fumes and mineral dust affect DNA methylation levels of genes regulating expression.** *Submitted for publication.*
D. A. van der Plaat; J. M. Vonk†; N. Terzikhan†; K. de Jong; M. de Vries; S. La Bastide-van Gemert; C. C. van Diemen; L. Lahousse; G. G. Brusselle; **I. Nedeljković**; N. Amin; BIOS consortium; H. Kromhout; R. C. H. Vermeulen; D. S. Postma; C. M. van Duijn; H. M. Boezen.

18. **Genetic correlation of Chronic Obstructive Pulmonary Disease and non-pulmonary comorbidity.** *In preparation.*
I. Nedeljković, M. de Vries, J. M. Vonk, L. Lahousse, B. H.C. Stricker, G. G. Brusselle, N. Amin, M. Boezen, C. M. van Duijn

19. **Novel adiponectin-associated variants implicate new obesity and lipid biology.** *Submitted for publication.*
C. N. Spracklen*, T. Karaderi*, H. Yaghootkar, C. Schurmann, R. S. Fine, Z. Kutalik, M. H. Preuss, Y. Lu, L. B.L. Wittemans, L. S. Adair, M. Allison, N. Amin, P. L. Auer, T. M. Bartz, M. Blüher, M. Boehnke, J. B. Borja, J. Bork-Jensen, L. Broer, D. I. Chasman, Y.-Der Ida Chen, P. Chirstofidou, A. Demirkan, K. W. van Dijk, C. M. van Duijn, M. F. Feitosa, M. E. Garcia, M. Graff, H. Grallert, N. Grarup, X. Guo, J. Haesser, T. Hansen, T. B. Harris, H. M. Highland, J. Hong, M. A. Ikram, E. Ingelsson, R. Jackson, P. Jousilahti, M. Kähönen, J. R. Kizer, P. Kovacs, J. Kriebel, M. Laakso, L. A. Lange, T. Lehtimäki, J. Li, R. Li-Gao, L. Lind, J. Luan, L.-Pekka Lyytikäinen, S.

MacGregor, D. Mackey, A. Mahajan, M. Mangino, S. Männistö, M. McCarthy, B. McKnight, C. Medina-Gomez, J. B. Meigs, S. Molnos, D. Mook-Kanamori, A. P. Morris, R. de Mutsert, M. A. Nalls, **I. Nedeljković**, K. E. North, C. E. Pennell, A. D. Pradhan, M. A. Province, O. T. Raitakari, C. K. Raulerson, A. P. Reiner, P. M. Ridker, S. Ripatti, N. Robertson, J. I. Rotter, V. Salomaa, A. A Sandoval-Zárate, C. M. Sitlani, T. D. Spector, K. Strauch, M. Stumvoll, K. D. Taylor, B. Thuesen, A. Tönjes, A. G. Uitterlinden, C. Venturini, M. Walker, C. A. Wang, S. Wang, N. J Wareham, S. M. Willems, J. G. Wilson, Y. Wu, J. Yao, K. L. Young, C. Langenberg, T. M. Frayling, T. O. Kilpeläinen, C. M. Lindgren**, R. J.F. Loos**, K. L. Mohlke**

** shared first authorship*
*** shared last authorship*

Abstracts

1. **Identification of novel rare genetic variants associated with COPD in the general population**
   M. de Vries, D. A. van der Plaat, **I. Nedeljković**, K. Joeri van der Velde, N. Amin, C. van Duijn, J. M. Vonk, H. Marike Boezen, C. C. van Diemen. European Respiratory Journal 52 (suppl 62) OA2193; DOI: 10.1183/13993003.congress-2018.OA2193 Published 19 November 2018

2. **A genome-wide SNP-by-NO2 interaction study on lung function in the LifeLines study**
   A. Julia De Faria Coimbra Lichtenfels, D. van der Plaat, K. de Jong, C. C. van Diemen, **I. Nedeljković**, C. M. van Duijn, N. Amin, R. P. Stolk, B. Brunekreef, M. Boezen, J. M. Vonk. European Respiratory Journal 50 (suppl 61) PA4475; DOI: 10.1183/1393003.congress-2017.PA4475 Published 6 December 2017

3. **DNA methylation mediates the association between occupational exposures and lung function**
   D. van der Plaat, M. de Vries, S. Bastide-van Gemert, K. de Jong, C. van Diemen**, I. Nedeljković**, N. Amin, H. Kromhout, R. Vermeulen, D. Postma, C. van Duijn, J. Vonk, M. Boezen. European Respiratory Journal 50 (suppl 61) OA2946; DOI: 10.1183/1393003.congress-2017.OA2946 Published 6 December 2017

4. **A genome-wide linkage study for COPD in a Dutch genetic isolate**
   I. Nedeljković, N. Terzikhan, J. M. Vonk, D. A. van der Plaat, L. Lahousse, C. C. van Diemen, G. G. Brusselle, H. Marike Boezen, C. M. van Duijn, N. Amin. European Respiratory Journal 50 (suppl 61) OA1789; DOI: 10.1183/1393003.congress-2017.OA1789 Published 6 December 2017

5. **LATE-BREAKING ABSTRACT: Occupational exposure to pesticides is associated with differential DNA methylation**
   D. van der Plaat, K. de Jong, J. Vonk, C. van Diemen, **I. Nedeljković**, N. Amin, H. Kromhout, R. Vermeulen, C. van Duijn, D. Postma, M. Boezen. European Respiratory Journal 48 (suppl 60) PA1211; DOI: 10.1183/13993003.congress-2016.PA1211 Published 8 November 2016

6. **Identification of Novel Genes Related to Two Definitions of Airway Obstruction in Never-Smokers**
   K. de Jong, A. Faiz, J. M. Vonk, C. C. van Diemen, **I. Nedeljković**, N. Amin, C. Anke Brandsma, Y. Bosse, D. D. Sin, D. C. Nickle, C. M. van Duijn, Dirkje S. Postma, H. Marike Boezen. MAY 2016 B105. COPD: EPIDEMIOLOGY AND GENETICS

7. **Identification of Novel Genes Related to Obstruction of the Small Airways in Adult Never-Smokers**
   K. de Jong, A. Faiz, J. M. Vonk, C. C. van Diemen**, I. Nedeljković**, N. Amin, C. Anke Brandsma, Y. Bosse, D. D. Sin, D. C. Nickle, C. M. van Duijn, H. Marike Boezen, D. S. Postma. MAY 2016 B105. COPD: EPIDEMIOLOGY AND GENETICS

8. **Variants in the Nicotine Acetylcholine Receptor (CHRNA 3/5) Locus May Play a Role in COPD Development Via Smoking Status**
   K. de Jong, J. M. Vonk, C. C. van Diemen, **I. Nedeljković**, N. Amin, C. M. van Duijn, D. S. Postma, H. Marike Boezen. MAY 2016 B105. COPD: EPIDEMIOLOGY AND GENETICS

9. **DNA Methylation Mediates the Effect of Cigarette Smoke Exposure on Lung Function Levels**
   D. A. van Der Plaat, A. J.F.C. Lichtenfels**, I. Nedeljković**, N. Amin, C. M. van Duijn, D. S. Postma, C. C. van Diemen, J. M. Vonk, H. M. Boezen. MAY 2017 D28. THE WEB OF OMICS: DNA, RNA, AND PROTEIN MODIFICATIONS