Genetic and Non-genetic Determinants and Clinical Consequences of Impaired Lung Function

Natalie Terzikhan

Financial support by the Fund for Scientific Research Flanders (Grant number G035014 N)

The Rotterdam Study is funded by Erasmus Medical Center and Erasmus University, Rotterdam, Netherlands Organization for the Health Research and Development (ZonMw), the Research Institute for Diseases in the Elderly (RIDE), the Ministry of Education, Culture and Science, the Ministry for Health, Welfare and Sports, the European Commission (DG XII), and the Municipality of Rotterdam. The author is grateful to the study participants, the staff from the Rotterdam Study and the participating general practitioners and pharmacists.

Layout and printed by: Optima Grafische Communicatie, Rotterdam, the Netherlands (www.ogc.nl)

ISBN: 978-94-6361-191-6

Genetic and Non-genetic Determinants and Clinical Consequences of Impaired Lung Function

Genetisch en niet-genetische determinanten en klinische consequenties van verminderde longfunctie

Thesis

to obtain the degree of Doctor from the
Erasmus University Rotterdam
by command of the
rector magnificus
Prof. dr. R.C.M.E. Engels

also thesis
to obtain the degree of Doctor from the in medical sciences from the
University of Ghent
by command of the
rector magnificus
Prof. dr. R. Van de Walle

and in accordance with the decision of the Doctorate Board.

The public defence shall be held on
Wednesday 19 December 2018 at 15:30 hrs
by
Natalie Terzikhan
born in Aleppo, Syria

GHENT UNIVERSITY

Erasmus University Rotterdam

PROMOTIECOMMISSIE

Promotoren:	Prof. dr. G.G. Brusselle
	Prof. dr. B.H. Stricker
Co-promotoren:	dr. L. Lahousse
	dr. K.M.C. Verhamme
Overige leden:	Prof. dr. H. M. Boezen
	Prof. dr. ir. C. M. van Duijn
	Prof. dr. M.A. Ikram

إذا الشعب يوما أراد الحياة
فلا بد أن يستجيب القدر
ولا بد لليل أن ينجلي
ولابد للقيد أن ينكسر
ومن لم يعانقه شوق الحياة
تبخر في جوها واندثر
كذلك قالت لي الكائنات
وحدثني روحها المستتر
ومن لا يحب صعود الجبال ...
يعش ابد الدهر بين الحفر

ابو القاسم الشابي

If, one day, a people desires to live, then fate will answer their call
And their night will then begin to fade, and their chains break and fall
For they who are not embraced by a passion for life will dissipate into thin air
At least that is what all creation has told me, and what its hidden spirits declare
For they who shrink from scaling the mountain will live forever in patholes

translation by Elliott Colla

PUBLICATIONS AND MANUSCRIPTS IN THIS THESIS

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;31(8):785-92.*

Wyss AB*, Sofer T*, Lee MK*, **Terzikhan N***, Nguyen JN*, Lahousse L*, et al. Multiethnic meta-analysis identifies ancestry-specific and cross-ancestry loci for pulmonary function. *Nat Commun. 2018;9(1):2976.*

Terzikhan N*, Sun F*, Verhamme FM*, Adams HHH, Loth D, Bracke KR, et al. Heritability and genome-wide association study of diffusing capacity of the lung. *Eur Respir J. 2018.*

van der Plaat D, Vonk JM*, **Terzikhan N***, de Jong K, de Vries M, La Bastide-van Gemert S, et al. Occupational exposures affect lung function via methylation of genes regulating expression. *Submitted*

Terzikhan N*, Xu H*, Bracke KR, Verhamme FM, Stricker BH, Dupuis J, Lahousse L, O'Connor GT, Brusselle GG. Epigenome-wide association study on diffusing capacity of the lung (DLCO): replication and meta-analysis. *Manuscript in preparation*

Nedeljkovic I, **Terzikhan N**, Vonk JM, van der Plaat DA, Lahousse L, van Diemen CC, et al. A Genome-Wide Linkage Study for Chronic Obstructive Pulmonary Disease in a Dutch Genetic Isolate Identifies Novel Rare Candidate Variants. *Front Genet. 2018;9:133.*

Terzikhan N, Bos D, Lahousse L, Wolff L, Verhamme KMC, Leening MJG, et al. Pulmonary artery to aorta ratio and risk of all-cause mortality in the general population: the Rotterdam Study. *Eur Respir J. 2017;49(6).*

Terzikhan N, Lahousse L, Verhamme KMC, Franco OH, Ikram MA, Stricker BH, Brusselle GG. Chronic Obstructive Pulmonary Disease is associated with an increased risk of peripheral artery disease and mortality. *accepted in the Eur Resp J Open Research 2018.*

*These authors have equal contribution to the manuscript.

TABLE OF CONTENTS OF THIS THESIS

Chapter 1 General introduction and outline of thesis 9

Chapter 2 Epidemiology of COPD 25

Chapter 3 Genetic epidemiology

Chapter 3.1 Heritability and GWAS of lung function 43

Chapter 3.2 Heritability and GWAS of pulmonary diffusion 77

Chapter 3.3 EWAS of the occupational exposures 97

Chapter 3.4 EWAS of pulmonary diffusion 113

Chapter 3.5 Genome-wide linkage study for COPD 129

Chapter 4 Clinical epidemiology

Chapter 4.1 Pulmonary artery: aorta (PA:A) ratio and mortality 149

Chapter 4.2 COPD and peripheral artery disease development and effects on mortality

Chapter 5 General discussion 183

Chapter 6 Summary 203

Appendices 207

List of publications 208

PhD portfolio 210

Words of gratitude 212

About the author 215

General introduction and outline of thesis

1

The lungs have an important function to provide the body with oxygen and eliminate carbon dioxide from the blood. Breathing keeps us alive, and an impaired lung function adversely affects this process. Research into the pathophysiology of lung diseases aims to determine the factors which contribute to an impaired lung function, in order to reach the ultimate goal of preventing and treating respiratory diseases. Lung function tests and the genotype arrays are important tools that help us to understand lung health and to investigate pathways that lead to an impaired lung function and respiratory disease.

1. CHRONIC OBSTRUCTIVE PULMONARY DISEASE (COPD)

COPD is the third leading cause of death worldwide (1). According to the Global Initiative for Chronic Obstructive Lung Disease (GOLD), *"COPD is a common, preventable and treatable disease that is characterized by persistent respiratory symptoms and airflow limitation that is due to airway and/or alveolar abnormalities usually caused by significant exposure to noxious particles or gases"* (2). COPD has two features; small airways disease (**chronic bronchitis**), and/or parenchymal destruction (**emphysema**), of which the relative contributions might vary from one individual to another.

1.1. Risk factors

Tobacco smoking is considered as the most important risk factor, in addition to **occupational**, outdoor and indoor air pollution (e.g. biomass fuel, dust, fumes and agricultural pesticides (3, 4)). Other factors also contribute to the development of COPD, including genetics, chronic bronchitis, asthma and airway hyper-responsiveness, poor lung growth during childhood, but also aging, female sex and low socioeconomic status (5-8).

1.2. Comorbidities and consequences

Comorbidities are abundant in COPD (9-11), as COPD often coexists with pulmonary diseases such as asthma (12), pulmonary fibrosis (13) and lung cancer (14, 15) and with extrapulmonary disease conditions (16) such as cerebrovascular disease (including stroke (17-19)), cognitive impairments (20, 21), obstructive sleep apnoea, cardiovascular disease (e.g. sudden cardiac death (22), **peripheral artery disease** (23, 24)), osteoporosis (25-29), metabolic disorders, depression and anxiety, skeletal muscle dysfunction (30, 31) and gastrointestinal reflux (32, 33). It is unclear whether those co-morbidities are the consequence of COPD or simply co-occur with COPD resulting from shared risk factors (10).

Individuals with COPD have distinct features which define the disease progression (34, 35). Late stage COPD can lead to serious complications, including COPD exacerbation with hospitalization, **pulmonary hypertension** and an increased risk of **mortality**.

2. LUNG FUNCTION

Lung function tests are important to detect and diagnose chronic lung diseases, assess disease severity, and monitor disease progression and treatment effects. The two most widely used lung function tests are **spirometry** and **diffusing capacity of carbon dioxide**, each of them assessing a distinct physiological function of the airways and lungs, respectively. Below, these tests are described in short.

2.1. Spirometry

Spirometry is a technique that measures the ability of the lungs to transport the air in and out of the lungs. The two most important measures in spirometry are the Forced Expiratory Volume in one second (FEV_1) and the Forced Vital Capacity (FVC). FEV_1 provides information about the expiratory airflow, and the FVC about the maximal amount of air that is exhaled. The ratio of these two measurements FEV_1/FVC, the so-called Tiffeneau index, is used for the diagnosis of chronic obstructive lung diseases. A reduced FEV_1/FVC ratio is indeed indicative of the presence of airflow limitation, which is a key diagnostic criterion for the diagnosis of COPD, but can also be reduced in patients with uncontrolled or severe asthma.

Lung function parameters decline with age, and the rate of decline depends on the exposure to environmental factors, such as smoking and air pollution, and genetic predisposition (36). Impaired lung function can be determined by many factors, including impaired lung growth during foetal life, childhood and adolescence. Importantly, impaired lung function also serves as an independent determinant for adverse health outcomes, including subsequent COPD (37, 38), cardiovascular disease (39), hospitalizations (40) and all-cause mortality (38, 40-42).

2.2. Diffusing capacity

The diffusing capacity of the lung for carbon monoxide (DLCO; also called transfer factor, TLCO) measures the total capacity of the body to exchange gases by the single-breath technique (SBT). Herein, small concentrations of carbon monoxide (CO) and a tracer gas (helium or another inert gas) are added to the inspiratory air in order to determine the levels of gas exchange. Patients are asked to inhale small amounts of CO, hold their breath for 10 seconds and exhale. Exhaled air is then sampled and the difference

between the added and the exhaled CO concentration yields the diffusing capacity (**Figure 1**) (43-46). DLCO is the product of two components that can be obtained by the SBT; firstly, the alveolar volume (VA) and secondly, the rate constant of CO removal from alveolar gas, the so called KCO, after taking the barometric pressure into account. Although it is generally accepted that DLCO/VA and KCO are physiologically equal, both are not interchangeable according to Hughes et al. (47). It is important to interpret DLCO/VA as the DLCO expressed per unit alveolar volume and not as the DLCO corrected for the alveolar volume, since DLCO/VA is not constant when VA changes. Therefore, DLCO/VA reflects the physiology more accurately than DLCO (47), since the very same DLCO measure can reflect different combinations of KCO and VA, each of these combinations reflect different pathophysiology (47-49).

Diffusing capacity is decreased by diseases such as emphysema, pulmonary fibrosis or pulmonary hypertension, and can be slightly increased in asthma or obesity (50). Since DLCO is also decreased in patients with anaemia, it is important to correct the measured DLCO values for blood haemoglobin levels.

Like impaired lung function measured by spirometry, a decreased diffusing capacity measured by the SBT has been associated with adverse health outcomes (51). In patients with severe emphysema, low diffusing capacity was associated with poor long-term survival (52) and it may even be a better predictor of mortality in COPD outpatients than spirometry measures (53).

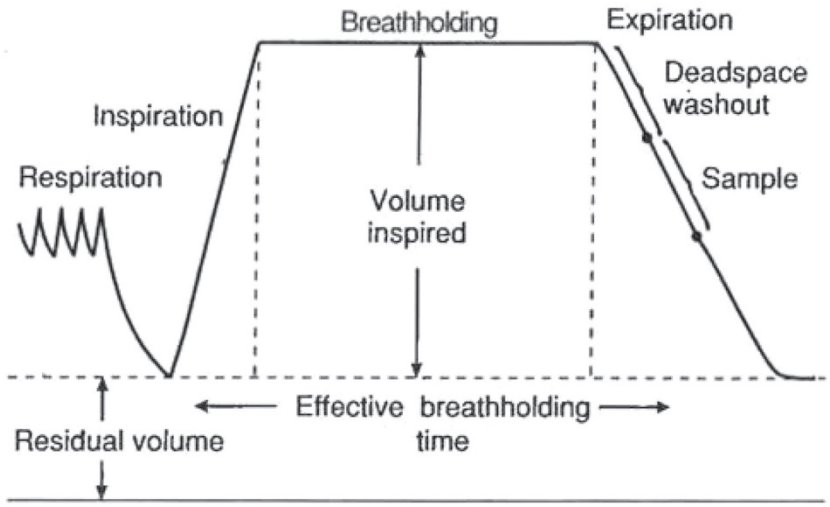

Figure 1 schematic representation of the single-breath DLCO breathing manoeuvre.
DLCO: Diffusing capacity of the Lung for Carbon Monoxide (CO)

3. DETERMINANTS OF IMPAIRED LUNG FUNCTION

There are many determinants of impaired lung function which can be distinguished into environmental, life style and genetic factors .

3.1. Environmental factors

Tobacco smoke is considered as the strongest risk factor for impaired lung function and COPD. However, a significant proportion of individuals who never smoked develop COPD, suggesting that other factors than smoking should also be considered.

Half of worlds households and the majority of the rural households use coal and biomass fuel as their primary source of domestic energy (54). It is suggested that exposure to biomass smoke might be the most important risk factor for COPD, since more people are exposed to biomass smoke worldwide than to tobacco smoke. In addition, **occupational exposures** to noxious gases, mineral dusts, biological dusts and fumes are also associated with airway obstruction (55, 56). Finally, raised concentrations of outdoor air pollutants, especially in fast growing industrial countries, are strongly associated with COPD exacerbations and worsening of pre-existing COPD (57).

3.2. Smoking

Cigarette smoking is considered as the major preventable risk factor of respiratory morbidity and mortality in developed countries (58). The effect of smoking on lung development and function is pronounced in early life and during adulthood. Maternal smoking during pregnancy, for example, is associated with increased risk of wheezing and asthma in the offspring at adolescence (59).

It is believed that smoking exaggerates chronic inflammation in diseased lungs. Oxidative stress, genetic susceptibility and epigenetic modifications play an important role in this, by amplifying the inflammation that is initiated by smoking (60).

Smoking has a modifying role in lung function. Hunninghake and colleagues demonstrated that genetic variation in *MMP12* (in smoking individuals) is associated with lung function. This effect was not found in non-smokers (61, 62). Another study by Xu et. al., also showed effect modification by smoking in the association between serum vitamin D and lung function (63). Furthermore, many interesting genetic and epigenetic loci were found for nicotine dependence (smoking behaviour), these pathways were also found in genome-wide and epigenome-wide association studies of lung function (64, 65). Whether there is genetic predisposition of nicotine dependence that influences lung function or whether we are dealing with pleiotropic effects remains to be investigated.

4. GENETIC EPIDEMIOLOGY

In genetic epidemiology, we study genetic and epigenetic determinants of health and disease in populations of unrelated individuals or in families. In the following paragraphs, we will summarize some important techniques used in genetic epidemiology.

4.1. Genome-wide association studies

A Genome-Wide Association Study (GWAS) is a hypothesis generating epidemiological study that aims to observe associations between Single Nucleotide Polymorphisms (SNPs) and traits of interest to understand the underlying biology and to ultimately improve treatment and prevention strategies (66). In the last decade, genetic studies such as GWAS and exome array studies (67), have been instrumental to identify genetic determinants of impaired lung function and COPD (68, 69). Those studies have also been successful in unravelling pathways and genes through which disease evolves in the body, such as the already consistently replicated loci *hedgehog-interacting protein (HHIP), Family with sequence similarity 13 member A (FAM13A)* and the smoking related *nicotinic acetylcholine receptor CHRNA5* (67, 70-72) (**Figure 2**).

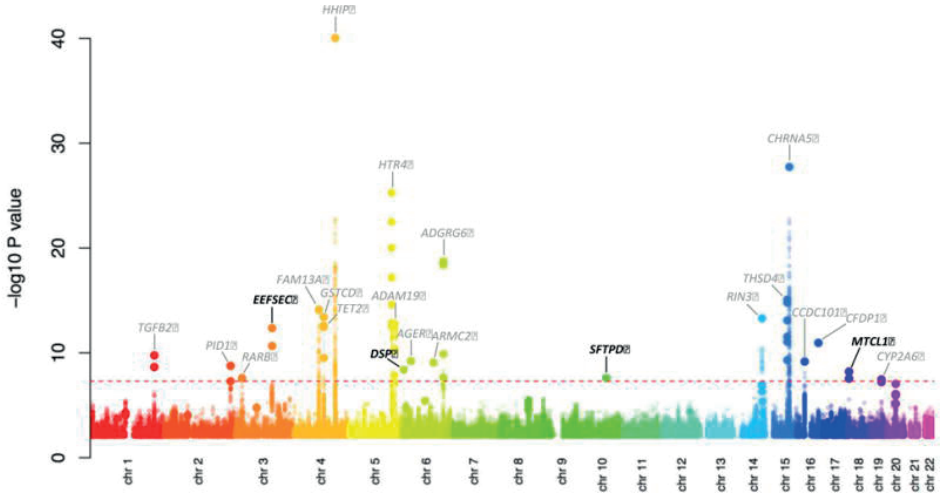

Figure 2 Manhattan plot from the publication of Hobbs and colleagues (70) Gene names in gray are previously known COPD or lung function (FEV1 or FEV1/FVC) loci; novel gene loci are indicated in black.

4.2. Epigenome-wide association studies

Similar to GWAS, an epigenome-wide association study (EWAS) is a hypothesis generating study investigating how epigenetic variations such as DNA methylation, histone modifications and non-coding RNA activities can influence phenotypes and

diseases (73). Among all forms of DNA methylation, methylation of cytosines in the context of cytosine–guanine dinucleotides (CpG) is the most prominent one. DNA methylation of cystosine is mapped through treatment of the genomic DNA with sodium bisulfite (74). The unmethylated cytosines will then convert to uracils, followed by thymidines that can be easily detected in a PCR, while the methylated cytosines remain unchanged (**Figure 3**). DNA methylation plays a key role in normal cellular function and homeostasis, whereas abnormal DNA methylation is associated with disease. So far, the role of epigenetic variation, particularly DNA methylation, has been mainly studied in the context of cancer (73, 75-77). EWAS on other complex diseases are also emerging and correlation between genotype and epigenotype (methQTL) favours the integrated approach between GWAS and EWAS in the future (73).

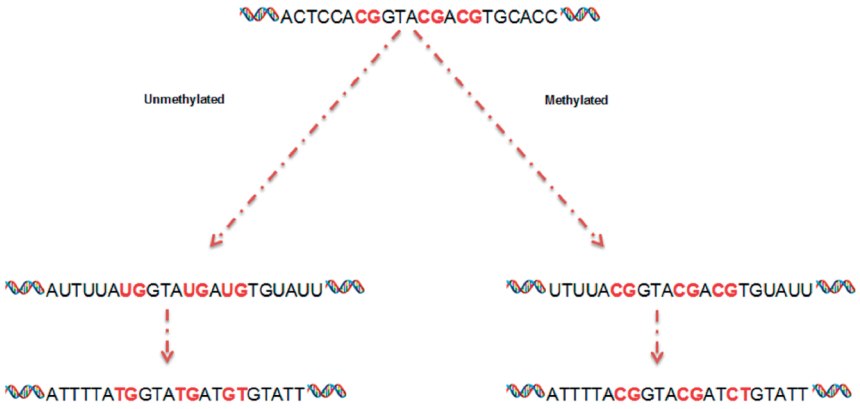

Figure 3 Detection of methylated regions after treatment with sodium bisulfite. DNA methylation of cystosine is mapped through treatment of the genomic DNA with sodium bisulfite (74). The unmethylated cytosines will then convert to uracils (UG), followed by thymidines (TG) that can be easily detected in a PCR, while the methylated cytosines remain unchanged (CG)

4.3. Genome-wide linkage analysis

GWAS studies have been functional to identify common genetic variants of disease. However, an emerging viewpoint suggests that rare variants -which are not well studied using GWAS- could explain a substantial part of the missing heritability (78, 79). Before, linkage analysis was the main tool to investigate rare genetic variance in complex traits with familial aggregation, until the focus was shifted to GWAS. Genome-wide linkage analysis is a powerful tool to localize disease genes on specific chromosomes . It relies on the theory that genes that are nearby each other are more linked and segregated together during meiosis compared to genes further away, and thereby links chromosomal fragments together among related individuals with trait similarity. In their most typical form, linkage studies have three steps. The first step is the identification of

the disease locus, next, sequence analysis is performed to define the rare deleterious variant, and finally, functional analysis follows to unravel the pathophysiology (80). With the increased use of whole genome sequencing (WGS), linkage studies have now more potential again to be used in identifying rare variants that cannot be captured in GWAS due to power issues (78-81).

5. AIMS AND OUTLINE OF THIS THESIS

The general aim of this thesis was to study the epidemiology of impaired lung function and COPD, to unravel its genetic and non-genetic determinants, and finally to study the impact of impaired lung function on adverse health outcomes, such as mortality.

Chapter 2: Epidemiology of COPD

In the first chapter of this thesis we studied the epidemiology of COPD. Here we focused on the prevalence and incidence of COPD in middle-aged and older subjects in the Rotterdam Study, with approximately 15,000 participants and a follow-up period of up to 25 years.

Chapter 3: Genetic epidemiology

In the **chapters 3.1** and **3.2**, we studied the heritability and the genetic determinants of pulmonary function tests; spirometry and diffusing capacity, respectively. In **chapter 3.3**, we investigated the association between occupational exposures on DNA methylation, and assessed the effect of occupational exposures on lung function in never smoking individuals via epigenetic mechanisms. In **chapter 3.4**, we also studied the epigenetic signature of diffusing capacity. Finally, in **chapter 3.5**, we studied rare variants that possibly affect COPD through linkage and exome sequence analyses. For the research work in this chapter, we collaborated with other large cohorts from The Cohorts for Heart and Aging Research in Genomic Epidemiology (**CHARGE**) consortium (**chapter 3.1**), with colleagues from the **Framingham Heart Study** (**chapter 3.2** and **3.4**), **Lifelines** (**chapter 3.3**) and the Erasmus Rucphen Family (**ERF**) study (**chapter 3.5**).

Chapter 4: Clinical relevance

In the third chapter, we aimed to investigate the adverse effects of an impaired lung function on health outcomes. In **chapter 4.1**, we investigated the effect of an increased pulmonary artery to aorta ratio, as measured on CT scans of the thorax, on mortality in individuals with COPD, and explored the modifying effect of diffusing capacity in this association. Finally, in **chapter 4.2**, we investigated the association between COPD and the risk to develop peripheral artery disease (PAD) and mortality.

REFERENCES

1. Lozano R, Naghavi M, Foreman K, Lim S, Shibuya K, Aboyans V, 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. 2012;380(9859):2095-128.
2. GOLD. GOLD reports 2018 [Available from: https://goldcopd.org/.
3. Hopkinson NS, Polkey MI. Chronic obstructive pulmonary disease in non-smokers. Lancet. 2009;374(9706):1964; author reply 5-6.
4. Forbes LJ, Kapetanakis V, Rudnicka AR, Cook DG, Bush T, Stedman JR, et al. Chronic exposure to outdoor air pollution and lung function in adults. Thorax. 2009;64(8):657-63.
5. Rosenberg SR, Kalhan R, Mannino DM. Epidemiology of Chronic Obstructive Pulmonary Disease: Prevalence, Morbidity, Mortality, and Risk Factors. Semin Resp Crit Care. 2015;36(4):457-69.
6. Brusselle GG, Lahousse L. Sex-Specific Genetic Risk Factors for Chronic Obstructive Pulmonary Disease. Am J Respir Cell Mol Biol. 2017;56(3):281-2.
7. Barnes PJ. Sex Differences in Chronic Obstructive Pulmonary Disease Mechanisms. Am J Respir Crit Care Med. 2016;193(8):813-4.
8. Antuni JD, Barnes PJ. Evaluation of Individuals at Risk for COPD: Beyond the Scope of the Global Initiative for Chronic Obstructive Lung Disease. Chronic Obstr Pulm Dis. 2016;3(3):653-67.
9. Barr RG, Celli BR, Mannino DM, Petty T, Rennard SI, Sciurba FC, et al. Comorbidities, patient knowledge, and disease management in a national sample of patients with COPD. Am J Med. 2009;122(4):348-55.
10. Negewo NA, Gibson PG, McDonald VM. COPD and its comorbidities: Impact, measurement and mechanisms. Respirology. 2015;20(8):1160-71.
11. Divo M, Cote C, de Torres JP, Casanova C, Marin JM, Pinto-Plata V, et al. Comorbidities and risk of mortality in patients with chronic obstructive pulmonary disease. Am J Respir Crit Care Med. 2012;186(2):155-61.
12. McDonald VM, Higgins I, Gibson PG. Managing older patients with coexistent asthma and chronic obstructive pulmonary disease: diagnostic and therapeutic challenges. Drugs Aging. 2013;30(1):1-17.
13. Sanchez MEV, Steiner R. Images in COPD: Combined Pulmonary Fibrosis and Emphysema. Chronic Obstr Pulm Dis. 2015;2(4):367-9.
14. Schroedl C, Kalhan R. Incidence, treatment options, and outcomes of lung cancer in patients with chronic obstructive pulmonary disease. Curr Opin Pulm Med. 2012;18(2):131-7.
15. de Torres JP, Marin JM, Casanova C, Cote C, Carrizo S, Cordoba-Lanus E, et al. Lung cancer in patients with chronic obstructive pulmonary disease-- incidence and predicting factors. Am J Respir Crit Care Med. 2011;184(8):913-9.
16. Patel AR, Hurst JR. Extrapulmonary comorbidities in chronic obstructive pulmonary disease: state of the art. Expert Rev Respir Med. 2011;5(5):647-62.
17. Portegies ML, Lahousse L, Joos GF, Hofman A, Koudstaal PJ, Stricker BH, et al. Chronic Obstructive Pulmonary Disease and the Risk of Stroke. The Rotterdam Study. Am J Respir Crit Care Med. 2016;193(3):251-8.
18. Soderholm M, Inghammar M, Hedblad B, Egesten A, Engstrom G. Incidence of stroke and stroke subtypes in chronic obstructive pulmonary disease. Eur J Epidemiol. 2016;31(2):159-68.

19. Lin CS, Shih CC, Yeh CC, Hu CJ, Chung CL, Chen TL, et al. Risk of Stroke and Post-Stroke Adverse Events in Patients with Exacerbations of Chronic Obstructive Pulmonary Disease. PLoS One. 2017;12(1):e0169429.
20. Kakkera K, Padala KP, Kodali M, Padala PR. Association of chronic obstructive pulmonary disease with mild cognitive impairment and dementia. Curr Opin Pulm Med. 2018;24(2):173-8.
21. Cleutjens F, Spruit MA, Ponds R, Vanfleteren L, Franssen FME, Gijsen C, et al. Cognitive impairment and clinical characteristics in patients with chronic obstructive pulmonary disease. Chron Respir Dis. 2018;15(2):91-102.
22. Lahousse L, Niemeijer MN, van den Berg ME, Rijnbeek PR, Joos GF, Hofman A, et al. Chronic obstructive pulmonary disease and sudden cardiac death: the Rotterdam study. Eur Heart J. 2015;36(27):1754-61.
23. Houben-Wilke S, Jorres RA, Bals R, Franssen FM, Glaser S, Holle R, et al. Peripheral Artery Disease and Its Clinical Relevance in Patients with Chronic Obstructive Pulmonary Disease in the COPD and Systemic Consequences-Comorbidities Network Study. Am J Respir Crit Care Med. 2017;195(2):189-97.
24. Lin MS, Hsu KY, Chen YJ, Chen CR, Chen CM, Chen W. Prevalence and risk factors of asymptomatic peripheral arterial disease in patients with COPD in Taiwan. PLoS One. 2013;8(5):e64714.
25. Maggi S, Siviero P, Gonnelli S, Schiraldi C, Malavolta N, Nuti R, et al. Osteoporosis risk in patients with chronic obstructive pulmonary disease: the EOLO study. J Clin Densitom. 2009;12(3):345-52.
26. Bhattacharyya P, Paul R, Ghosh M, Dey R, Dey R, Barooah N, et al. Prevalence of osteoporosis and osteopenia in advanced chronic obstructive pulmonary disease patients. Lung India. 2011;28(3):184-6.
27. Jorgensen NR, Schwarz P, Holme I, Henriksen BM, Petersen LJ, Backer V. The prevalence of osteoporosis in patients with chronic obstructive pulmonary disease: a cross sectional study. Respir Med. 2007;101(1):177-85.
28. Jorgensen NR, Schwarz P. Osteoporosis in chronic obstructive pulmonary disease patients. Curr Opin Pulm Med. 2008;14(2):122-7.
29. Pobeha P, Lazurova I, Tkacova R. [Osteoporosis in chronic obstructive pulmonary disease] Osteoporoza pri chronickej obstrukcnej chorobe pl'uc. Vnitr Lek. 2010;56(11):1142-9.
30. Jaitovich A, Barreiro E. Skeletal Muscle Dysfunction in Chronic Obstructive Pulmonary Disease. What We Know and Can Do for Our Patients. Am J Respir Crit Care Med. 2018;198(2):175-86.
31. Kim HC, Mofarrahi M, Hussain SN. Skeletal muscle dysfunction in patients with chronic obstructive pulmonary disease. Int J Chron Obstruct Pulmon Dis. 2008;3(4):637-58.
32. Kim SW, Lee JH, Sim YS, Ryu YJ, Chang JH. Prevalence and risk factors for reflux esophagitis in patients with chronic obstructive pulmonary disease. Korean J Intern Med. 2014;29(4):466-73.
33. Jung KS. Reflux esophagitis is one of highly prevalent comorbidities among patients with chronic obstructive pulmonary disease. Korean J Intern Med. 2014;29(4):428-9.
34. Segal LN, Martinez FJ. Chronic obstructive pulmonary disease subpopulations and phenotyping. J Allergy Clin Immunol. 2018;141(6):1961-71.
35. Vestbo J, Lange P. Natural history of COPD: Focusing on change in FEV1. Respirology. 2016;21(1):34-43.
36. Molfino NA. Genetic predisposition to accelerated decline of lung function in COPD. Int J Chron Obstruct Pulmon Dis. 2007;2(2):117-9.

37. Krzyzanowski M, Jedrychowski W, Wysocki M. Factors associated with the change in ventilatory function and the development of chronic obstructive pulmonary disease in a 13-year follow-up of the Cracow Study. Risk of chronic obstructive pulmonary disease. Am Rev Respir Dis. 1986;134(5):1011-9.
38. Beeckman LAF, Wang ML, Petsonk EL, Wagner GR. Rapid declines in FEV1 and subsequent respiratory symptoms, illnesses, and mortality in coal miners in the United States. Am J Resp Crit Care. 2001;163(3):633-9.
39. Persson C, Bengtsson C, Lapidus L, Rybo E, Thiringer G, Wedel H. Peak expiratory flow and risk of cardiovascular disease and death. A 12-year follow-up of participants in the population study of women in Gothenburg, Sweden. Am J Epidemiol. 1986;124(6):942-8.
40. Lash TL, Johansen MB, Christensen S, Baron JA, Rothman KJ, Hansen JG, et al. Hospitalization rates and survival associated with COPD: a nationwide Danish cohort study. Lung. 2011;189(1):27-35.
41. Mannino DM, Buist AS, Petty TL, Enright PL, Redd SC. Lung function and mortality in the United States: data from the First National Health and Nutrition Examination Survey follow up study. Thorax. 2003;58(5):388-93.
42. Schunemann HJ, Dorn J, Grant BJB, Winkelstein W, Trevisan M. Pulmonary function is a long-term predictor of mortality in the general population - 29-year follow-up of the Buffalo Health Study. Chest. 2000;118(3):656-64.
43. Crapo RO, Forster RE, 2nd. Carbon monoxide diffusing capacity. Clin Chest Med. 1989;10(2):187-98.
44. Crapo RO, Jensen RL, Wanger JS. Single-breath carbon monoxide diffusing capacity. Clin Chest Med. 2001;22(4):637-49.
45. Crapo RO, Morris AH. Standardized single breath normal values for carbon monoxide diffusing capacity. Am Rev Respir Dis. 1981;123(2):185-9.
46. Hegewald MJ. Diffusing capacity. Clin Rev Allergy Immunol. 2009;37(3):159-66.
47. Hughes JM, Pride NB. Examination of the carbon monoxide diffusing capacity (DL(CO)) in relation to its KCO and VA components. Am J Respir Crit Care Med. 2012;186(2):132-9.
48. Hughes JM, Pride NB. In defence of the carbon monoxide transfer coefficient Kco (TL/VA). Eur Respir J. 2001;17(2):168-74.
49. Hughes JM, Pride NB. Carbon monoxide transfer coefficient (transfer factor/alveolar volume) in females versus males. Eur Respir J. 2003;22(1):186-7.
50. Ferguson MK, Lehman AG, Bolliger CT, Brunelli A. The role of diffusing capacity and exercise tests. Thorac Surg Clin. 2008;18(1):9-17, v.
51. Ferguson MK, Little L, Rizzo L, Popovich KJ, Glonek GF, Leff A, et al. Diffusing capacity predicts morbidity and mortality after pulmonary resection. J Thorac Cardiovasc Surg. 1988;96(6):894-900.
52. Bates DV, Knott JM, Christie RV. Respiratory function in emphysema in relation to prognosis. Q J Med. 1956;25(97):137-57.
53. Boutou AK, Shrikrishna D, Tanner RJ, Smith C, Kelly JL, Ward SP, et al. Lung function indices for predicting mortality in COPD. Eur Respir J. 2013;42(3):616-25.
54. Salvi SS, Barnes PJ. Chronic obstructive pulmonary disease in non-smokers. Lancet. 2009;374(9691):733-43.
55. Matheson MC, Benke G, Raven J, Sim MR, Kromhout H, Vermeulen R, et al. Biological dust exposure in the workplace is a risk factor for chronic obstructive pulmonary disease. Thorax. 2005;60(8):645-51.

56. de Jong K, Boezen HM, Kromhout H, Vermeulen R, Postma DS, Vonk JM, et al. Pesticides and other occupational exposures are associated with airway obstruction: the LifeLines cohort study. Occup Environ Med. 2014;71(2):88-96.
57. Sunyer J. Urban air pollution and chronic obstructive pulmonary disease: a review. Eur Respir J. 2001;17(5):1024-33.
58. Bergen AW, Caporaso N. Cigarette smoking. J Natl Cancer Inst. 1999;91(16):1365-75.
59. Hollams EM, de Klerk NH, Holt PG, Sly PD. Persistent effects of maternal smoking during pregnancy on lung function and asthma in adolescents. Am J Respir Crit Care Med. 2014;189(4):401-7.
60. Brusselle GG, Joos GF, Bracke KR. Chronic Obstructive Pulmonary Disease 1 New insights into the immunology of chronic obstructive pulmonary disease. Lancet. 2011;378(9795):1015-26.
61. Hunninghake GM, Cho MH, Tesfaigzi Y, Soto-Quiros ME, Avila L, Lasky-Su J, et al. MMP12, lung function, and COPD in high-risk populations. N Engl J Med. 2009;361(27):2599-608.
62. Brusselle GG. Matrix metalloproteinase 12, asthma, and COPD. N Engl J Med. 2009;361(27):2664-5.
63. Jiayi Xu TMB, Geetha Chittoor, Gudny Eiriksdottir, Ani W. Manichaikul, Fangui Sun, Natalie Terzikhan, Xia Zhou, Sarah L. Booth, Guy G. Brusselle, Ian H. de Boer, Myriam Fornage, Alexis C. Frazier-Wood, Mariaelisa Graff, Vilmundur Gudnason, Tamara B. Harris, Albert Hofman, Ruixue Hou, Denise K. Houston, David R. Jacobs Jr., Stephen B. Kritchevsky, Jeanne Latourelle, Rozenn N. Lemaitre, Pamela L. Lutsey, George O'Connor, Elizabeth C. Oelsner, James S. Pankow, Bruce M. Psaty, Rebecca R. Rohde, Stephen S. Rich, Jerome I. Rotter, Lewis J. Smith, Bruno H. Stricker, V. Saroja Voruganti, Thomas J. Wang, M. Carola Zillikens, R. Graham Barr, Josée Dupuis, Sina A. Gharib, Lies Lahousse, Stephanie J. London, Kari E. North, Albert V. Smith, Lyn M. Steffen, Dana B. Hancock, Patricia A. Cassano. Large Meta-Analysis in the CHARGE Consortium Provides Evidence For an Association of Serum Vitamin D With Pulmonary Function. bioRxiv. 2018.
64. Joehanes R, Just AC, Marioni RE, Pilling LC, Reynolds LM, Mandaviya PR, et al. Epigenetic Signatures of Cigarette Smoking. Circ Cardiovasc Genet. 2016;9(5):436-47.
65. Tobacco, Genetics C. Genome-wide meta-analyses identify multiple loci associated with smoking behavior. Nat Genet. 2010;42(5):441-7.
66. Visscher PM, Wray NR, Zhang Q, Sklar P, McCarthy MI, Brown MA, et al. 10 Years of GWAS Discovery: Biology, Function, and Translation. Am J Hum Genet. 2017;101(1):5-22.
67. Jackson VE, Ntalla I, Sayers I, Morris R, Whincup P, Casas JP, et al. Exome-wide analysis of rare coding variation identifies novel associations with COPD and airflow limitation in MOCS3, IFIT3 and SERPINA12. Thorax. 2016;71(6):501-9.
68. Li Y, Cho MH, Zhou XB. What do polymorphisms tell us about the mechanisms of COPD? Clin Sci. 2017;131(24):2847-63.
69. Melen E, Guerra S. Recent advances in understanding lung function development. F1000Res. 2017;6:726.
70. Hobbs BD, de Jong K, Lamontagne M, Bosse Y, Shrine N, Artigas MS, et al. Genetic loci associated with chronic obstructive pulmonary disease overlap with loci for lung function and pulmonary fibrosis. Nat Genet. 2017;49(3):426-32.
71. 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.

72. Wain LV, Shrine N, Artigas MS, Erzurumluoglu AM, Noyvert B, Bossini-Castillo L, et al. Genome-wide association analyses for lung function and chronic obstructive pulmonary disease identify new loci and potential druggable targets. Nat Genet. 2017;49(3):416-25.
73. Rakyan VK, Down TA, Balding DJ, Beck S. Epigenome-wide association studies for common human diseases. Nat Rev Genet. 2011;12(8):529-41.
74. Clark SJ, Harrison J, Paul CL, Frommer M. High sensitivity mapping of methylated cytosines. Nucleic Acids Res. 1994;22(15):2990-7.
75. Akhavan-Niaki H, Samadani AA. DNA methylation and cancer development: molecular mechanism. Cell Biochem Biophys. 2013;67(2):501-13.
76. Jones PA, Baylin SB. The epigenomics of cancer. Cell. 2007;128(4):683-92.
77. Yang IV, Schwartz DA. Epigenetic control of gene expression in the lung. Am J Respir Crit Care Med. 2011;183(10):1295-301.
78. Ott J, Wang J, Leal SM. Genetic linkage analysis in the age of whole-genome sequencing. Nat Rev Genet. 2015;16(5):275-84.
79. McClellan J, King MC. Genetic heterogeneity in human disease. Cell. 2010;141(2):210-7.
80. Altshuler D, Daly MJ, Lander ES. Genetic mapping in human disease. Science. 2008;322(5903):881-8.
81. Pulst SM. Genetic linkage analysis. Arch Neurol. 1999;56(6):667-72.

2

Prevalence and incidence of COPD in smokers and non-smokers: the Rotterdam Study

ABSTRACT

Background: COPD is the third leading cause of death in the world and its global burden is predicted to increase further. Even though the prevalence of COPD is well studied, only few studies examined the incidence of COPD in a prospective and standardized manner.

Methods: In a prospective population-based cohort study (Rotterdam Study) enrolling subjects aged ≥45, COPD was diagnosed based on a pre-bronchodilator obstructive spirometry ($FEV_1/FVC < 0.70$). In absence of an interpretable spirometry within the Rotterdam Study, cases were defined as having COPD diagnosed by a physician on the basis of clinical presentation and obstructive lung function measured by the general practitioner or respiratory physician. Incidence rates were calculated by dividing the number of incident cases by the total number of person years of subjects at risk.

Results: In this cohort of 14,619 participants, 1,993 subjects with COPD were identified of whom 689 as prevalent ones and 1,304 cases as incident ones. The overall incidence rate (IR) of COPD was 8.9/1,000 person-years (PY); 95% Confidence Interval (CI) 8.4 to 9.4. The IR was higher in males and in smokers. The proportion of female COPD participants without a history of smoking was 27.2%, while this proportion was 7.3% in males.

Conclusion: The prevalence of COPD in the Rotterdam Study is 4.7% and the overall incidence is approximately 9/1,000 PY, with a higher incidence in males and in smokers. The proportion of never-smokers among female COPD cases is substantial.

Key words: COPD, GOLD, LLN, prevalence, incidence, the Rotterdam Study.

INTRODUCTION

Worldwide, Chronic Obstructive Pulmonary Disease (COPD) is the third leading cause of death [1]. COPD is characterized by persistent airflow limitation that is typically progressive and associated with an enhanced chronic inflammatory response in the airways and lung tissue to harmful particles or gases [2]. The chronic airflow limitation in COPD is caused by the combination of parenchymal destruction (emphysema) and small airways disease (obstructive bronchiolitis), of which the relative presence varies from person to person [2]

According to estimates from the Global Burden of Disease Study, COPD was prevalent in more than 300 million people in 2013 [3]. The disease burden and its financial impact is predicted to increase, mainly due to population aging [4-6]. Several studies reported on the prevalence of COPD. In European adult populations over 40 years, the prevalence of COPD ranges between 15-20% and is higher in men than in women [7-9]. Even though the prevalence of COPD is well known, only few studies examined its incidence rate in a prospective and standardized manner (please see supplementary **Table 1S** in Online Resource 1 for an overview of studies which investigated the incidence of COPD).

While tobacco smoking is a major risk factor for COPD, only approximately 20% of smokers develop the disease . More evidence is rising to suggest that other risk factors such as air pollution, respiratory infections, poor nutritional status, chronic asthma, impaired lung growth, poor socio-economic status and genetic factors are also important for disease development [10-12]. About 15-20% of COPD cases are due to occupational exposures to pollutants at the workplace [9], and about 50% of subjects who died from COPD in developing countries have been exposed to biomass smoke during lifetime [10]. These facts emphasize the need for action in order to reduce the impact of those risk factors on disease development. To this end, investigating the incidence of COPD is important, since it might shed light on new trends in the development and course of the disease, which in turn can lead to new insights and guidance for prevention and treatment. Therefore, the objective of this study is to investigate the prevalence and incidence of COPD by age, sex and smoking status in the participants of the Rotterdam Study, a large ongoing prospective population-based cohort study with 25 years of follow-up.

MATERIALS AND METHODS

The present study was embedded within the Rotterdam Study, an ongoing prospective population-based cohort study that investigates the occurrence of chronic diseases and

risk factors in elderly. The objective and methods of this cohort have been published previously [13, 14]. Briefly, The Rotterdam Study (RS) includes approximately 15,000 participants aged ≥ 45 years, living in Ommoord, a well-defined suburb of the city of Rotterdam, The Netherlands, and encompasses three cohorts: RS I, RS II and RS III. Baseline data were collected between 1989 and 1992 in RS I (n = 7,983), between 2000 and 2003 in RS II (n = 3,011) and between 2006 and 2009 in RS III (n = 3,932); thereafter cross-sectional surveys and examinations have been conducted every 4 to 5 years. Participants were initially interviewed at home for information on their health status. This was followed by an extensive set of examinations performed at a specially built research facility in the study district. Trained research assistants collected information from medical records of the general practitioners (GPs), nursing homes and hospitals. The study was approved by the medical ethics committee of Erasmus Medical Center, Rotterdam. All participants gave their written informed consent and permission to retrieve information from treating physicians.

COPD diagnosis

COPD was diagnosed based on an obstructive pre-bronchodilator spirometry (FEV_1/FVC < 0.70) according to the GOLD guidelines [2, 15, 16]. We also diagnosed COPD according to the lower level of normal (LLN) instead of GOLD as a sensitivity analysis as proposed by Hankinson *et al.* [17]. Spirometry was performed by trained paramedical personnel according to the ATS/ERS guidelines, using a portable spirometer (SpiroPro; Erich Jaeger GmbH; Hoechberg, Germany) from 2002-2008, and using a Master Screen® PFT Pro (Care Fusion, the Netherlands) since 2009. Spirometry results which did not meet ATS/ERS criteria for acceptability were classified as not interpretable [18, 19]. Reversibility tests were not performed.

Within the Rotterdam study, pre-bronchodilator spirometry was performed in 8,411 participants. In 7,188 subjects, the spirometry met ATS/ERS criteria and was thus interpretable. In absence of an interpretable study-acquired spirometry, the medical records including letters from specialists and the electronic GP files were reviewed of all patients who regularly used medication for obstructive lung disease (Anatomical Therapeutic Chemical Classification codes: R03). Drug use was exclusively used for potential case finding; each such potential case was subsequently validated through careful evaluation of all medical records, hospitalizations and specialist letters and only included if a clear and well-founded diagnosis of COPD was retained. Cases were then defined as having a physician diagnosed COPD based on clinical presentation and obstructive lung function measured by the GP or respiratory physician. Physician diagnosed asthma patients (n=460) were excluded from the COPD cases, but were controls, as they were at risk to develop COPD. In addition, 54 asthma cases were added to the COPD group since they developed COPD during follow-up. All asthma

cases had a validated physician diagnosis of asthma identified by thorough examination of medical charts information, as spirometry data provide us only with information about airflow limitation. A total of 311 of the 460 (68%) asthma cases also performed an interpretable (pre-bronchodilator) spirometry within the Rotterdam Study. Within the 311 subjects with physician-diagnosed asthma and an interpretable spirometry, 60 had an obstructive lung function examination (FEV1/FVC < 0.7). Those subjects would have been classified as COPD if we had solely based our COPD classification on spirometry data. This demonstrates the strength of our approach of combining medical chart information with spirometry data.

The incident date was defined as the date of the first obstructive lung function examination, the date of COPD diagnosis in the medical records or the date of first prescription of the COPD medication (in those with established COPD) whichever came first.

Prevalent cases were defined as having COPD at inclusion. Incident cases were defined as participants who acquired COPD during follow-up. For incident COPD cases, follow-up time was defined as the time period between the start of the study and the diagnosis of COPD, lost to follow up, death, or the last visit to the study centre (December, 2014).

STATISTICAL ANALYSIS

Statistical analysis was performed using SPSS statistical software (SPSS for Windows, version 21; SPSS; Chicago, IL), R (Foundation for Statistical Computing, Vienna, Austria) and Microsoft Excel (version 2010). For the statistical analyses, patients without informed consent for follow-up were excluded. The prevalence of COPD (%) was calculated by dividing the total number of COPD cases at baseline (prevalent cases) by the total number of participants included. The prevalence at the end of follow-up (%) was calculated as the total number of COPD cases at the end of this study divided by the total number of participants included. For the analysis of the incidence rate of COPD, patients with prevalent COPD were excluded. Median follow up time was estimated using the reverse Kaplan Meier method (also called Kaplan-Meier estimate of potential follow-up method). Incidence rates were calculated by dividing the number of incident cases by the total number of person years of subjects at risk and are presented per 1,000 person years. The 95% Confidence Intervals (CI) were calculated using a Poisson distribution. Incidence rates were stratified for sex, age, and self-reported smoking behaviour at baseline. Age was categorized into eight groups (45-<50, 50-<55, 55-<60, 60-<65, 65-<70, 70-<75, 75-<80, >80 years) as described before [18]. Smoking behaviour was categorised as current, former and never.

To study age-specific incidence rates, follow-up time was divided by five-year age intervals. Subjects contributed to the subsequent age intervals until they developed COPD, were lost to follow up, died, or reached the end of study (December, 2014).

RESULTS

In this cohort of 14,619 participants with informed consent for follow-up, a total of 1,993 individuals (56.5% males) were identified as having COPD and 12,626 participants (38.8% males) did not have COPD. A total of 689 COPD subjects were identified as having prevalent COPD at baseline and 1,304 COPD cases were incident **(Figure 1)**. The median follow-up time was 10.7 years (with a maximum follow-up time up to 25 years) and mean age at baseline was 65.8 ± 10.4 years.

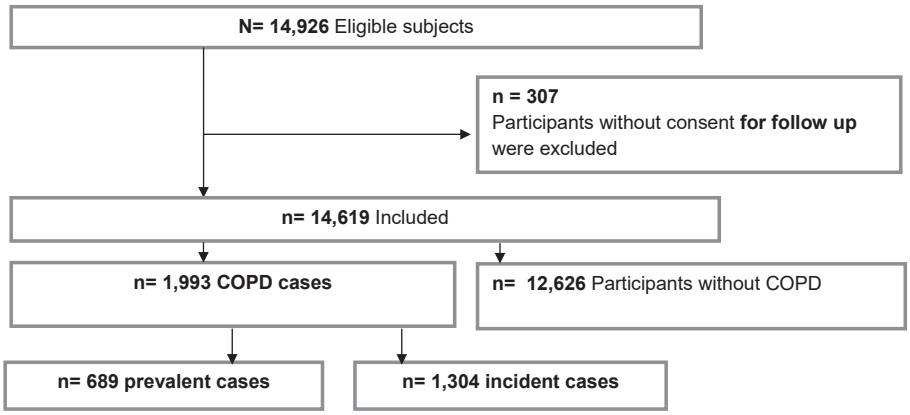

Fig. 1 Flow chart of participants in the study.

Regarding the smoking status 21.7% of the study participants were current smokers, 41.7% former smokers and 34.2% never smokers **(Table 1)**. In ever smokers, 17.8% (1,663/9,169) had COPD (including incident and prevalent cases), whereas in never smokers the prevalence of COPD was 6.4% (318/4,997). In men, 17.3% (n=1,042/6,024) were never smokers, compared to 46.0% (n=3,955/8,595) never smoking women. The proportion of COPD female cases without a smoking history was 27.2% (236/867), while the proportion of never smokers among COPD male cases was 7.3% (82/1,126). Amongst the incident COPD patients who never smoked, questionnaire information on passive smoking was available in 170 patients. The proportion of passive smoking in these patients was 51.2% (n=87) and amongst these passive smokers, the majority were female (67/87,77%).

Table 1 Baseline characteristics of the study population (n= 14,619)[a]

Characteristics	Total *n* = 14,619	COPD cases *n* = 1993	Non-cases *n* = 12,626
Age (years) at baseline	65.8 (10.4)	64.8 (8.5)	65.9 (10.6)
Gender *n* (%)			
Males	6024 (41.2)	1126 (56.5)	4898 (38.8)
Females	8595 (58.8)	867 (43.5)	7728 (61.2)
Genetic ethnicity *n* (%)			
Central European	11,617 (98.0)	1638 (98.9)	9979 (97.9)
Asian	145 (1.2)	14 (0.8)	131 (1.3)
African	69 (0.6)	2 (0.1)	67 (0.7)
Admixed	21 (0.2)	3 (0.2)	18 (0.2)
Missing genetic data *n*	2767	336	2431
Smoking at baseline *n* (%)			
Current smoker	3078 (21.7)	800 (41.0)	2278 (18.6)
Former smoker	6091 (43.0)	831 (42.6)	5260 (43.1)
Never smoker	4997 (35.3)	318 (16.3)	4679 (38.3)
Missing *n*	453	44	409
Pack years of smoking mean (SD)			
Current smoker	30.3 (21.3)	34.6 (19.8)	28.7 (21.5)
Former smoker	22.0 (23.8)	33.6 (28.5)	20.2 (22.5)
Missing *n*	770	77	693
Anthropometry mean (SD)			
Weight (Kg)	76.0 (13.9)	76.2 (13.4)	76.0 (14.0)
Height (cm)	168.0 (9.6)	170.5 (9.4)	167.6 (9.5)
BMI	26.9 (4.1)	26.1 (3.9)	27.0 (4.1)
Missing *n*	1559	143	1416
Blood pressure mean (SD)			
Systolic blood pressure (mmHg)	139.6 (36.3)	138.1 (32.4)	139.8 (36.9)
Diastolic blood pressure (mmHg)	78.5 (33.7)	77.7 (29.1)	78.6 (34.4)
Missing *n*	1392	131	1261

a Data are presented as n (%) or Mean ± Standard deviation (SD).

The prevalence of COPD at baseline in the Rotterdam Study was 4.7% (689/14,619) and the prevalence at the end of follow-up was 13.6% (1993/14,619). The overall incidence rate (IR) of COPD was 8.9/1,000 person-years (PY) (95% CI 8.4 to 9.4/1,000 PY). For the sensitivity analysis using LLN instead of GOLD classification method, the overall incidence rate was 5.5/1,000 PY (95% CI 5.2 to 5.9) (See **Table 2S** in online resource 2 for detailed information on the prevalence and incidence data according to different classification methods; GOLD and LLN).

Subgroup analysis of the spirometry data based on GOLD (n=7,153) *versus* medical record data (n=7,466) was also performed. The prevalence of COPD was 5.3% *versus*

4.2%, respectively. The incidence rate of COPD was 11.7/1,000PY (95% CI 10.9 to 12.4) *versus* 5.8/1,000PY (95% CI 5.3 to 6.4), respectively (**Table 2S** in online resource 2). Additional information is provided on severity and respiratory complaints in the spirometry group in **Table 3S** in online resource 3.

The overall IR was higher in men (13.3/1,000 PY, 95% CI 12.4-14.3) than in women (6.1/1,000 PY, 95% CI 5.6-6.6); age specific IR ranged between 8.7 and 17.6/1,000 PY in males and 3.0 to 7.9/1,000 PY in females. The incidence of COPD increased from the age of 45 in both sexes to the age of 80 in men and 75 in women **(Figure 2)**. The IR was higher in current and former smokers than in never-smokers (19.7/1,000 PY, 95% CI 18.1-21.4 in current smokers, 8.3/1,000 PY, 95% CI 7.6-9.1 in former smokers and 4.1/1,000 PY, 95% CI 3.6-4.7, in never smokers). The IR of COPD in smoking men was 15.0/1,000 PY (95% CI 13.9-16.2) compared to 8.6/1,000 PY (95% CI 7.8-9.5) in smoking women. The age-specific IR of COPD in ever smokers ranged between 7.3 and 15.3/1,000 PY. The IR was 6.0/1,000 PY (95% CI 4.6-7.8) in never smoking men and 3.7/1,000 PY (95% CI 3.1-4.3) in never smoking women. The age-specific incidence of COPD in never smokers increased by age, but to a lesser extent than the incidence of COPD in ever smokers **(Figure 3)**. After stratification by sex and smoking history, the age-specific incidence in never smoking women showed the same pattern **(Figure 4).**

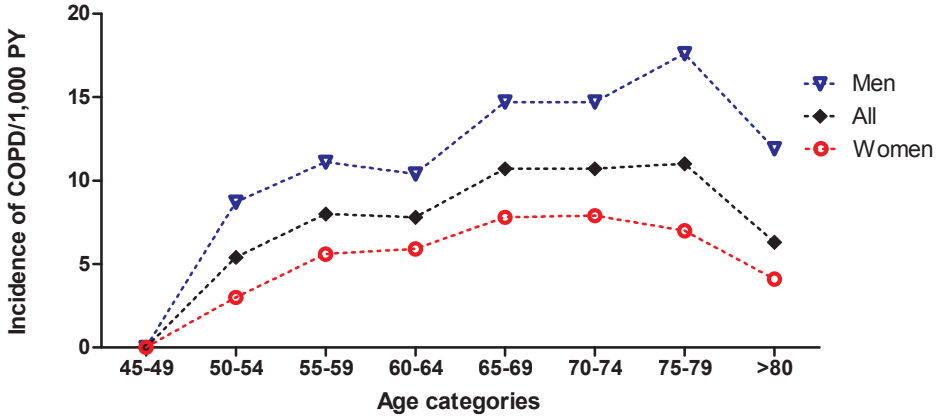

Fig. 2 Age-specific incidence of COPD by sex

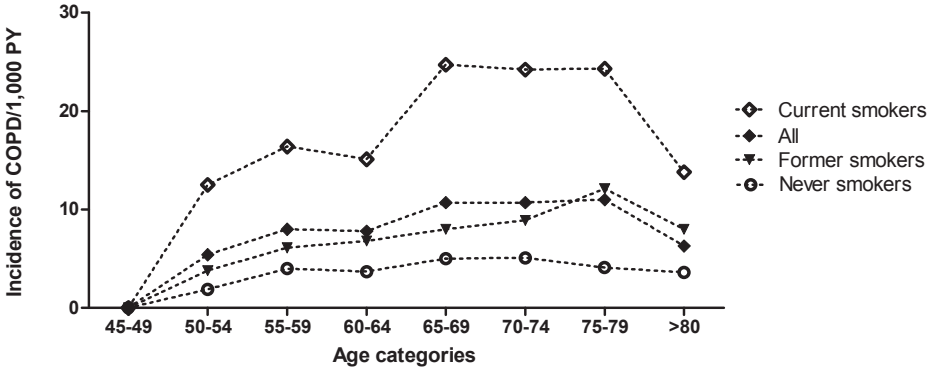

Fig. 3 Age specific incidence of COPD by smoking behaviour

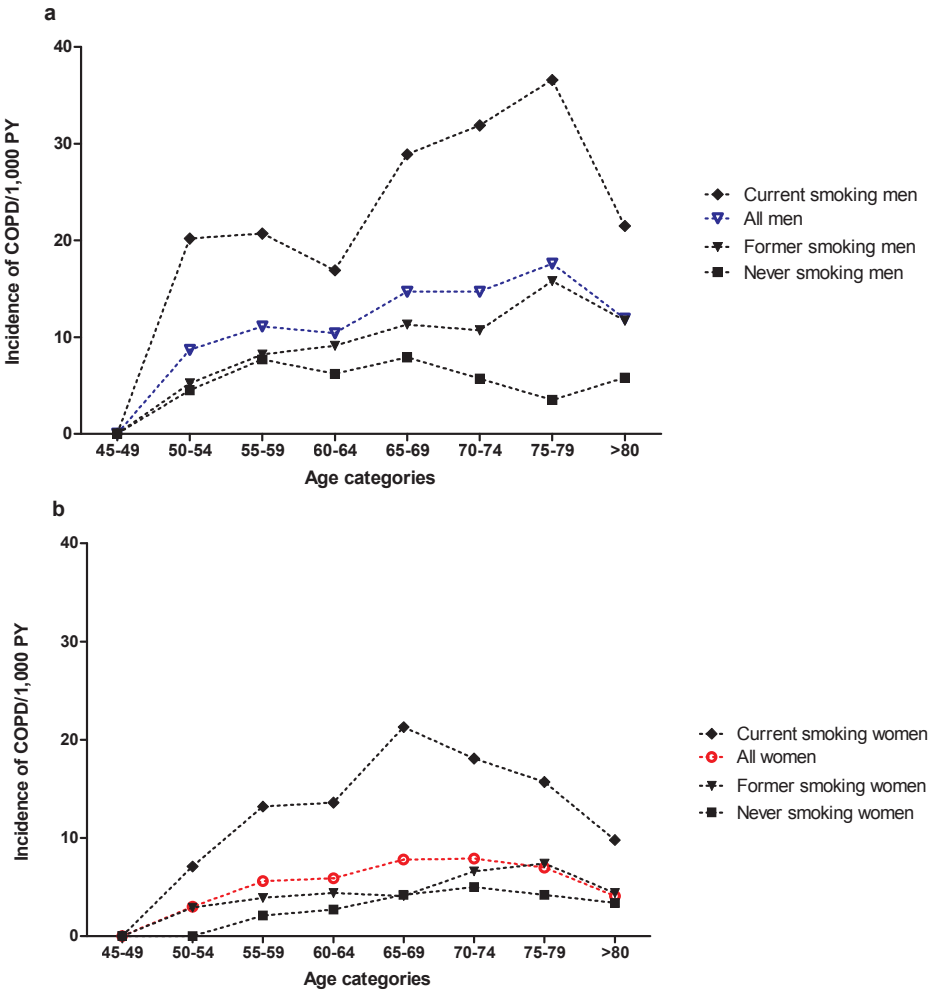

Fig. 4 Age-specific incidence of COPD by sex (a: men and b: women) and smoking behaviour

DISCUSSION

In this large prospective cohort study, the baseline prevalence of COPD was 4.7% and the prevalence at the end of follow-up was 13.6%. The overall incidence rate of COPD was approximately 9/1,000 PY. This rate increased progressively with age, was higher in men than in women and higher in ever smokers compared to never smokers. Importantly, more than one in four female COPD subjects was a never smoker.

We previously published on the prevalence and incidence rates of COPD in the first Rotterdam Study (RS I; encompassing 7,983 participants) over a follow-up period of 15.5 years (from 1989 till 2004) [18]. Here we report on the prevalence and incidence rates of COPD in all three RS cohorts (encompassing 14,926 participants) with an extended follow-up period of up to 25 years (from baseline till 2014). Comparing the results, the overall incidence rate in both studies showed high consistency (IR: 9,2/1,000 PY [95% CI; 8,5-10] in RS I versus 8,9/1,000 PY [95% CI; 8.4-9.4] in RS I, II and III combined). Given that both cohort studies used overlapping but different data sources (RS I versus RS I, II and III) and had different lengths of follow-up (15,5 versus 25 years), the consistency found highlights the reliability of the epidemiologic data.

Reviewing the literature, several studies reported on the prevalence of COPD. These prevalences varied widely and ranged from 0.2% in Japan to 37% in the USA and between 2.1% and 26.1% in Europe [7, 8]. Bischoff et al. presented data on the prevalence of COPD in a dynamic general practice population aged 40 and older in the Netherlands using data from the Continuous Morbidity Registration [20]. Their prevalence (5.4%) was in line with the one found in our study (4.7%).

Numerous studies reported on the incidence of COPD but only few studies reported the incidence rate taking the individual contribution to follow-up into account, in large cohorts with long follow-up time (see supplementary **Table 1S**).The reported measures on the incidence varied widely when reported in terms of rates per 1000 persons, ranging from 8.2 to 81.6 [21-23], while the incidence rates reported per person time units ranged from 2.6 to 9.2 per 1,000 PY [18, 24-27]. This variation in incidence rates can be explained by variability in terms of the definition of COPD, research methods, source population and calendar time [7].

Overall, our study confirms that the incidence of COPD is higher in men than in women and in elderly (>75y) than in younger subjects. At our study centre, COPD was diagnosed based on an obstructive (pre-bronchodilator) spirometry. If an interpretable spirometry was not available, COPD was defined as a validated diagnosis made by the GP or the respiratory physician on the basis of clinical presentation and obstructive lung function.Therefore, not only symptomatic but (in contrast to the patients seen by the physician in the clinic) also asymptomatic or oligosymptomatic subjects with COPD were diagnosed in the RS. Since mild COPD cases rarely seek medical attention, the

true incidence of COPD is frequently underestimated in clinical settings [23]. Studies that reported on physician diagnosed COPD showed lower IRs compared to the rate found in our epidemiologic study [24, 25, 28] **(Table 1S)**. In our study, subgroup analysis of the spirometry data *versus* medical record data were also calculated. The incidence rates showed a similar pattern as compared to the literature and were 11.7/1,000 PY using spirometry data *versus* 5.8/1,000 PY using medical records data.

In this study, we classified COPD cases according to GOLD guidelines. Since ATS and ERS recently advocate the use of the Lower Limit of Normal (LLN), we also recalculated the incidence rate using LLN classification in the spirometry group instead of GOLD. The overall incidence rate after reclassifying the spirometry based COPD cases according to LLN was lower than the initial incidence rate using GOLD (5.5/1,000PY versus 8.9/1,000 PY) which is in line with the literature [29] This difference is ascribed to the fact that mild COPD cases as classified according to GOLD were considered as controls when LLN was used as a cut-off. Whether mild (asymptomatic) COPD should be classified as COPD is sometimes debated. However, Mannino *et al*. demonstrated that subjects classified as "cases" using GOLD but as "normal" using LLN have a significantly higher risk of COPD-related hospitalization and mortality [30].

In our study, the age-specific incidence of COPD in never smokers increased by age, but to a lesser extent than the incidence of COPD in current and former smokers **(Figure 3)**. The detection of COPD cases in never smokers indicates that, besides tobacco smoking, other factors such as genetic susceptibility, impaired lung growth, respiratory infections and environmental exposures including occupational exposures and (outdoor and indoor) air pollution might contribute to the development of COPD [10-12]. Interestingly, in our study, approximately 27.2% of all female COPD cases were never smokers, whereas this prevalence was much lower in men (7.3%). This suggests that the contribution of environmental exposures other than active smoking leading to COPD seems more substantial in females than in males. Indeed, our data confirm that one of these environmental exposures namely; passive smoking, is higher in females than in males [31-33].

Likewise, more evidence is emerging on the increasing occurrence of COPD in non-smoking individuals, especially in females. Worldwide an estimated 25-45% of patients with COPD never smoked [10]. Nevertheless, most randomized clinical trials (RCT) that examine the efficacy and safety of pharmacologic treatments for COPD, only include COPD patients with a history of cigarette smoking of at least 10 pack years [10].

The burden of COPD in never smokers is higher than previously believed [10, 11, 31, 34], therefore more research is needed to unravel the characteristics of non-smoking COPD in order to address the true burden, prognosis, clinical, radiographic

and physiological features and treatment possibilities in this specific and neglected group.

The strengths of the Rotterdam study are the prospective, population-based design with a maximum follow-up time of 25 years. In addition, measurements of the variables in this prospective cohort is done independently of the research question, which makes it less prone to information and selection bias.

A limitation is that spirometry measurements were introduced in the Rotterdam study in January 2002 and therefore measured in only 8,411 participants (out of 9,950 still alive). This could lead to an underestimation of asymptomatic COPD in the Rotterdam Study in participants without spirometry measurement. A second limitation is that within the Rotterdam Study, as in most population-based cohort studies, reversibility tests were not performed, because the use of inhaled bronchodilators could interfere with other tests during the study visit. This could inflate the prevalence of COPD considerably [35, 36]. While some researchers state that the use of a bronchodilator is necessary to eliminate the variable airflow limitation in order to diagnose COPD [37], others suggest that bronchodilator responsiveness is anyway greatly variable and that more than 50% of the patients who initially were classified as reversible would be reclassified, had they attended on a different occasion [38, 39]. The use of pre-bronchodilator spirometry implies that we cannot exclude the possibility of misclassification of some asthma patients as COPD patients. To minimize the risk of misclassification, we additionally identified and validated patients with physician-diagnosed asthma. However, we still acknowledge the use of pre-bronchodilatory test results as weakness because some unknown degree of inflation of COPD diagnoses might still be present.

In conclusion, the overall incidence rate of COPD in the Rotterdam Study was approximately 9/1,000 PY, with a higher incidence in males and in smokers. The proportion of never smokers among COPD cases is substantial and higher in females than in males.

REFERENCES

1. Lozano R, Naghavi M, Foreman K, 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*. 2012; **380**: 2095-128.
2. 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 Resp Crit Care*. 2013; **187**: 347-65.
3. Vos Tea. Global, regional, and national incidence, prevalence, and years lived with disability for 301 acute and chronic diseases and injuries in 188 countries, 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013. *The Lancet*. 2015.
4. Prince MJ, Wu F, Guo Y, et al. The burden of disease in older people and implications for health policy and practice. *Lancet*. 2014.
5. Muka T, Imo D, Jaspers L, et al. The global impact of non-communicable diseases on healthcare spending and national income: a systematic review. *Eur J Epidemiol*. 2015; **30**: 251-77.
6. Herse F, Kiljander T, Lehtimaki L. Annual costs of chronic obstructive pulmonary disease in Finland during 1996-2006 and a prediction model for 2007-2030. *NPJ Prim Care Respir Med*. 2015; **25**: 15015.
7. Atsou K, Chouaid C, Hejblum G. Variability of the chronic obstructive pulmonary disease key epidemiological data in Europe: systematic review. *Bmc Med*. 2011; **9**: 7.
8. Rycroft CE, Heyes A, Lanza L, Becker K. Epidemiology of chronic obstructive pulmonary disease: a literature review. *Int J Chron Obstruct Pulmon Dis*. 2012; **7**: 457-94.
9. Gibson L, Sibille, Lundbäck, Fletcher,,. Lung health in Europe, facts and figures. European lung foundation 2013.
10. Salvi SS, Barnes PJ. Chronic obstructive pulmonary disease in non-smokers. *Lancet*. 2009; **374**: 733-43.
11. Zeng G, Sun B, Zhong N. Non-smoking-related chronic obstructive pulmonary disease: a neglected entity? *Respirology*. 2012; **17**: 908-12.
12. Lange P, Celli B, Agusti A, et al. Lung-Function Trajectories Leading to Chronic Obstructive Pulmonary Disease. *N Engl J Med*. 2015; **373**: 111-22.
13. Hofman A, Darwish Murad S, van Duijn CM, et al. The Rotterdam Study: 2014 objectives and design update. *Eur J Epidemiol*. 2013; **28**: 889-926.
14. Hofman A, Brusselle GG, Darwish Murad S, et al. The Rotterdam Study: 2016 objectives and design update. *Eur J Epidemiol*. 2015; **30**: 661-708.
15. Lahousse L, Maes B, Ziere G, et al. Adverse outcomes of frailty in the elderly: the Rotterdam Study. *Eur J Epidemiol*. 2014; **29**: 419-27.
16. Lahousse L, Niemeijer MN, van den Berg ME, et al. Chronic obstructive pulmonary disease and sudden cardiac death: the Rotterdam study. *Eur Heart J*. 2015.
17. Hankinson JL, Odencrantz JR, Fedan KB. Spirometric reference values from a sample of the general U.S. population. *Am J Respir Crit Care Med*. 1999; **159**: 179-87.
18. van Durme YM, Verhamme KM, Stijnen T, et al. Prevalence, incidence, and lifetime risk for the development of COPD in the elderly: the Rotterdam study. *Chest*. 2009; **135**: 368-77.
19. Huitema MP, Reesink HJ, Post MC. Incremental value of pulmonary artery diameter indexed for BSA in pulmonary hypertension. *Thorax*. 2015.
20. Bischoff EW, Schermer TR, Bor H, Brown P, van Weel C, van den Bosch WJ. Trends in COPD prevalence and exacerbation rates in Dutch primary care. *Br J Gen Pract*. 2009; **59**: 927-33.

21. Crighton EJ, Ragetlie R, Luo J, To T, Gershon A. A spatial analysis of COPD prevalence, incidence, mortality and health service use in Ontario. *Health Rep*. 2015; **26**: 10-8.
22. Gershon AS, Wang C, Wilton AS, Raut R, To T. Trends in chronic obstructive pulmonary disease prevalence, incidence, and mortality in ontario, Canada, 1996 to 2007: a population-based study. *Arch Intern Med*. 2010; **170**: 560-5.
23. Mannino DM, Homa DM, Akinbami LJ, Ford ES, Redd SC. Chronic obstructive pulmonary disease surveillance--United States, 1971-2000. *Respir Care*. 2002; **47**: 1184-99.
24. Afonso AS, Verhamme KM, Sturkenboom MC, Brusselle GG. COPD in the general population: prevalence, incidence and survival. *Respir Med*. 2011; **105**: 1872-84.
25. Garcia Rodriguez LA, Wallander MA, Tolosa LB, Johansson S. Chronic obstructive pulmonary disease in UK primary care: incidence and risk factors. *Copd*. 2009; **6**: 369-79.
26. Kojima S, Sakakibara H, Motani S, et al. Incidence of chronic obstructive pulmonary disease, and the relationship between age and smoking in a Japanese population. *J Epidemiol*. 2007; **17**: 54-60.
27. de Marco R, Accordini S, Cerveri I, et al. Incidence of chronic obstructive pulmonary disease in a cohort of young adults according to the presence of chronic cough and phlegm. *Am J Respir Crit Care Med*. 2007; **175**: 32-9.
28. Nihlen U, Nyberg P, Montnemery P, Lofdahl CG. Influence of family history and smoking habits on the incidence of self-reported physician's diagnosis of COPD. *Respir Med*. 2004; **98**: 263-70.
29. Guder G, Brenner S, Angermann CE, et al. "GOLD or lower limit of normal definition? A comparison with expert-based diagnosis of chronic obstructive pulmonary disease in a prospective cohort-study". *Respir Res*. 2012; **13**: 13.
30. Mannino DM, Sonia Buist A, Vollmer WM. Chronic obstructive pulmonary disease in the older adult: what defines abnormal lung function? *Thorax*. 2007; **62**: 237-41.
31. Hagstad S, Bjerg A, Ekerljung L, et al. Passive smoking exposure is associated with increased risk of COPD in never smokers. *Chest*. 2014; **145**: 1298-304.
32. Yin P, Jiang CQ, Cheng KK, et al. Passive smoking exposure and risk of COPD among adults in China: the Guangzhou Biobank Cohort Study. *Lancet*. 2007; **370**: 751-7.
33. Tan WC, Sin DD, Bourbeau J, et al. Characteristics of COPD in never-smokers and ever-smokers in the general population: results from the CanCOLD study. *Thorax*. 2015; **70**: 822-9.
34. Lamprecht B, McBurnie MA, Vollmer WM, et al. COPD in Never Smokers Results From the Population-Based Burden of Obstructive Lung Disease Study. *Chest*. 2011; **139**: 752-63.
35. Johannessen A, Omenaas ER, Bakke PS, Gulsvik A. Implications of reversibility testing on prevalence and risk factors for chronic obstructive pulmonary disease: a community study. *Thorax*. 2005; **60**: 842-7.
36. Tilert T, Dillon C, Paulose-Ram R, Hnizdo E, Doney B. Estimating the U.S. prevalence of chronic obstructive pulmonary disease using pre- and post-bronchodilator spirometry: the National Health and Nutrition Examination Survey (NHANES) 2007-2010. *Respir Res*. 2013; **14**: 103.
37. Sterk PJ. Let's not forget: the GOLD criteria for COPD are based on post-bronchodilator FEV1. *Eur Respir J*. 2004; **23**: 497-8.
38. Albert P, Agusti A, Edwards L, et al. Bronchodilator responsiveness as a phenotypic characteristic of established chronic obstructive pulmonary disease. *Thorax*. 2012; **67**: 701-8.
39. Calverley PM, Burge PS, Spencer S, Anderson JA, Jones PW. Bronchodilator reversibility testing in chronic obstructive pulmonary disease. *Thorax*. 2003; **58**: 659-64.

Genetic Epidemiology

3

3.1 Multiethnic meta-analysis identifies ancestry-specific and cross-ancestry loci for pulmonary function

ABSTRACT

Background: Near 100 loci have been identified for pulmonary function, almost exclusively in studies of European ancestry populations.

Methods: We extend previous research by meta-analyzing genome-wide association studies of 1000 Genomes imputed variants in relation to pulmonary function in a multiethnic population of 90,715 individuals of European (*N*=60,552), African (*N*=8,429), Asian (*N*=9,959), and Hispanic/Latino (*N*=11,775) ethnicities.

Results: We identify over 50 additional loci at genome-wide significance in ancestry-specific or multiethnic meta-analyses. Using recent fine-mapping methods incorporating functional annotation, gene expression, and differences in linkage disequilibrium between ethnicities, we further shed light on potential causal variants and genes at known and newly identified loci. Several of the novel genes encode proteins with predicted or established drug targets, including *KCNK2* and *CDK12*.

Conclusion: Our study highlights the utility of multiethnic and integrative genomics approaches to extend existing knowledge of the genetics of lung function and clinical relevance of implicated loci.

INTRODUCTION

Pulmonary function traits (PFTs), including forced expiratory volume in the first second (FEV_1) and forced vital capacity (FVC), and their ratio FEV_1/FVC, are important clinical measures for assessing respiratory health, diagnosing chronic obstructive pulmonary disease (COPD), and monitoring the progression and severity of various other lung conditions. Further, even when within the normal range, these parameters are related to mortality, independently of standard risk factors[1-3].

In addition to lifestyle and environmental factors, such as smoking and air pollution, genetics influences pulmonary function[4-6]. Previous genome-wide association studies (GWAS) have identified nearly 100 loci associated with PFTs[7-15]. These analyses have been primarily conducted using HapMap imputed data among European ancestry populations[7-12]. Recently, the UK BiLEVE Study (*N*=48,943) and SpiroMeta Consortium (*N*=38,199) have also examined associations between 1,000 Genomes imputed variants and PFTs, but only among Europeans[13-15].

The present Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) meta-analysis builds upon previous studies by examining PFTs in relation to the more comprehensive 1000 Genomes panel in a larger study population (90,715 individuals from 22 studies, Table 1) comprised of multiple ancestral populations: European (60,552 individuals from 18 studies), African (8,429 individuals from 7 studies), Asian (9,959 individuals from 2 studies), and Hispanic/Latino (11,775 individuals from six ethnic background groups in 1 study). Along with look-up of our top findings in existing analyses of lung function traits and COPD, we additionally investigate correlation with gene expression in datasets from blood and lung tissue, identify known or potential drug targets for newly identified lung function associated loci, and assess the potential biological importance of our findings using recently developed methods integrating linkage disequilibrium (LD), functional annotation, gene expression and the multiethnic nature of our data. By conducting a GWAS meta-analysis in a large multiethnic population and employing recently developed integrative genomic methods, we identify over 50 additional loci associated with pulmonary function, including some with functional or clinical relevance

RESULTS

Ancestry-Specific Meta-Analyses

Each study used linear regression to model the additive effect of variants on PFTs, adjusting for age, sex, height, cigarette smoking, weight (for FVC only), and center, ancestral principal components, and a random familial effect to account for family

Table 1. Sample size and location of studies included in the CHARGE consortium 1000 Genomes and pulmonary function meta-analysis

		Sample Size by Ancestry			
Study[a]	**Country**	**European**	**African**	**Hispanic/Latino**	**Asian**
AGES[b]	Iceland	1620			
ALHS	United States	2844			
ARIC[b]	United States	8878	1837		
CARDIA[b]	United States	1580	883		
CHS[b]	United States	3135	566		
FamHS	United States	1679			
FHS[b]	United States	7689			
GOYA	Denmark	1456			
HCHS/SOL	United States			11775	
HCS[b]	Australia	1822			
Health ABC[b]	United States	1472	943		
Healthy Twin	South Korea				2098
JHS	United States		2015		
KARE3	South Korea				7861
LifeLines[b]	Netherlands	11851			
LLFS[b]	United States and Denmark	3787			
MESA[b]	United States	1339	863		
NEO	Netherlands	5460			
1982 Pelotas	Brazil	1357	1322		
RS I[b]	Netherlands	1232			
RS II[b]	Netherlands	1135			
RS III[b]	Netherlands	2216			
	Total	**60552**	**8429**	**11775**	**9959**

[a]AGES Age Gene Environment Susceptibility Study; ALHS Agricultural Lung Health Study (1180 asthma cases and 1664 controls); ARIC Atherosclerosis Risk in Communities Study; CARDIA Coronary Artery Risk Development in Young Adults; CHS Cardiovascular Health Study; FamHS Family Heart Study; FHS Framingham Heart Study; GOYA Genetics of Overweight Young Adults Study (670 obese cases and 786 controls); HCHS/SOL Hispanic Community Health Study/Study of Latinos; HCS Hunter Community Study; JHS Jackson Heart Study; KARE3 Korean Association Resource Phase 3 Study; LLFS Long Life Family Study; MESA Multi-Ethnic Study of Atherosclerosis; NEO Netherlands Epidemiology of Obesity Study; RS Rotterdam Study

[b]Studies included in one or more previous CHARGE papers: Hancock et al 2010 included ARIC, CHS, FHS, RSI and RSII; Soler-Artigas et al 2011 included AGES, ARIC, CHS, FHS, Health ABC, RSI and RSII in stage 1 and HCS, CARDIA, LifeLines, MESA, and RSIII in stage 2; and Loth et al 2014 included AGES, ARIC, CARDIA, CHS, FHS, Health ABC, HCS, MESA, RSI, RSII, and RSIII in stage 1 and LifeLines and LLFS in stage 2.

relatedness when appropriate. Ancestry-specific fixed-effects inverse-variance weighted meta-analyses of study-specific results, with genomic control correction, were conducted in METAL (http://www.sph.umich.edu/csg/abecasis/metal/). Meta-analyses included approximately 11.1 million variants for European ancestry, 18.1 million for African ancestry, 4.2 million variants for Asian ancestry, and 13.8 million for Hispanic/Latino ethnicity (see Methods).

European ancestry meta-analyses identified 17 novel loci (defined as more than 500kb in either direction from the top variant of a known locus as has been used in previous multiethnic GWAS[16]) which were significantly (defined as $P<5.0x10^{-8}$) [14,17] associated with pulmonary function: 2 loci for FEV_1 only, 6 loci for FVC only, 7 loci for FEV_1/FVC only, and 2 loci for both FEV_1 and FVC (Table 2, Figure 1, Supplementary Figures 2-3). The African ancestry meta-analysis identified 8 novel loci significantly associated with pulmonary function: 2 loci for FEV_1, 1 locus for FVC, and 5 loci for FEV_1/FVC (Table 3, Supplementary Figures 1-3). Five of these loci were also significant at a stricter $P<2.5x10^{-8}$ threshold as has been suggested for populations of African ancestry[17]. Six of the African ancestry loci were identified based on variants with low allele frequencies (0.01-0.02) in African ancestry and which were monomorphic or nearly monomorphic (allele frequency < 0.004) in other ancestries (European, Asian, and Hispanic; Supplementary Table 2). In the Hispanic/Latino ethnicity meta-analysis, we identified one novel locus for FVC (Table 3, Supplementary Figures 1-3). Another locus was significantly associated with FEV_1, but this region was recently reported by HCHS/SOL[18]. For FEV_1/FVC among Hispanics/Latinos, all significant variants were in loci identified in previous studies of European ancestry populations. In the Asian ancestry meta-analysis, all variants significantly associated with PFTs were also in loci previously identified among European ancestry populations (Supplementary Figure 1). Within each ancestry, variants discovered for one PFT were also looked-up for associations with the other two PFTs (Supplementary Table 1).

Multiethnic Meta-analysis

In multiethnic fixed effects meta-analyses of 10.9 million variants, we identified 47 novel loci significantly associated with pulmonary function. Thirteen of these loci were also identified in the ancestry-specific meta-analyses, while 34 were uniquely identified in the multiethnic meta-analysis: 11 loci for FEV_1 only, 14 loci for FVC only, 7 loci for FEV_1/FVC only, 1 locus for FEV_1 and FEV_1/FVC, and 1 locus for all three phenotypes (Table 4, Figure 1, Supplementary Figures 2-3). Although many of the 34 loci uniquely identified in the multiethnic meta-analysis were just shy of significance in the European ancestry meta-analysis, and therefore benefited from the additional sample size of the multiethnic meta-analysis, some multiethnic loci contained variants near genome-wide significance in at least one other ancestry-specific meta-analysis

Table 2. Top Variants from Novel Loci Discovered in European Ancestry Meta-analysis of Pulmonary Function in the CHARGE Consortium

Nearest Gene(s)[a]	Trait[b]	Top Variant	Chr:Pos	Coded Allele	Allele Freq	N	Beta[c]	SE	P-value
LOC728989	FVC	rs12724426	1:146494027	a	0.21	31315	-36.75	6.63	2.95E-08
CENPF, KCNK2	FVC	rs512597	1:215095003	t	0.81	60507	-24.26	4.12	3.92E-09
C1orf140, DUSP10	FVC	rs6657854	1:221630555	a	0.72	60508	-19.89	3.49	1.18E-08
RBMS3	FEV_1/FVC	rs17666332	3:29469675	t	0.72	60531	0.003	0.0005	4.76E-08
AFAP1	FEV_1/FVC	rs28520091	4:7846240	t	0.48	60527	0.003	0.0004	2.17E-09
AP3B1	FEV_1	rs252746	5:77392117	a	0.78	60551	20.05	3.45	6.19E-09
	FVC	rs12513481	5:77450828	c	0.23	60507	-25.01	3.74	2.15E-11
LINC00340	FEV_1/FVC	rs1928168	6:22017738	t	0.51	60522	0.003	0.0004	6.74E-14
SLC25A51P1,BAI3	FEV_1/FVC	rs9351637	6:67863782	t	0.61	60528	0.002	0.0004	2.89E-08
CNTNAP2	FEV_1/FVC	rs1404154	7:146651409	t	0.99	23748	-0.03	0.006	2.80E-08
DMRT2, SMARCA2	FVC	rs771924	9:1555835	a	0.42	60507	-18.4	3.18	7.16E-09
ALX1, RASSF9	FEV_1	rs10779158	12:85724096	a	0.34	60550	15.89	2.9	4.36E-08
	FVC	rs10779158	12:85724096	a	0.34	60506	18.72	3.31	1.52E-08
LOC644172, CRHR1	FEV_1	rs143246821	17:43685698	a	0.79	39416	30.58	4.99	9.06E-10
WNT3	FEV_1	rs916888	17:44863133	t	0.75	60551	20.53	3.48	3.76E-09
DCC	FVC	rs8089865	18:50957922	a	0.59	60509	20.57	3.23	1.95E-10
TSHZ3	FEV_1/FVC	rs1353531	19:31846907	t	0.14	60530	-0.003	0.0006	4.53E-08
EYA2	FVC	rs2236519	20:45529571	a	0.38	60508	-18.06	3.28	3.51E-08
KLHL22, MED15	FEV_1/FVC	rs4820216	22:20854161	t	0.15	60528	-0.004	0.0006	1.53E-09

[a]Nearest gene: indicates gene either harboring the variant or nearest to it. *C1orf140/DUSP10* locus also includes *HLX. CRHR1/LOC644172* locus also includes *ARHGAP27, MGC57346, CRHR1-IT1, LRRC37A4P. KLHL22/MED15* locus also includes *ZNF74, SCARF2*.

[b]Phenotypes: FEV_1 forced expiratory volume in 1 second (in ml), FVC forced vital capacity (in ml), Ratio FEV_1/FVC (as a proportion)

[c]Additive effect of variant on pulmonary function, adjusting for age, age^2, sex, height, height^2, smoking status, pack-years of smoking, weight (for FVC only), and center, ancestral principal components, and a random familial effect to account for family relatedness when appropriate.

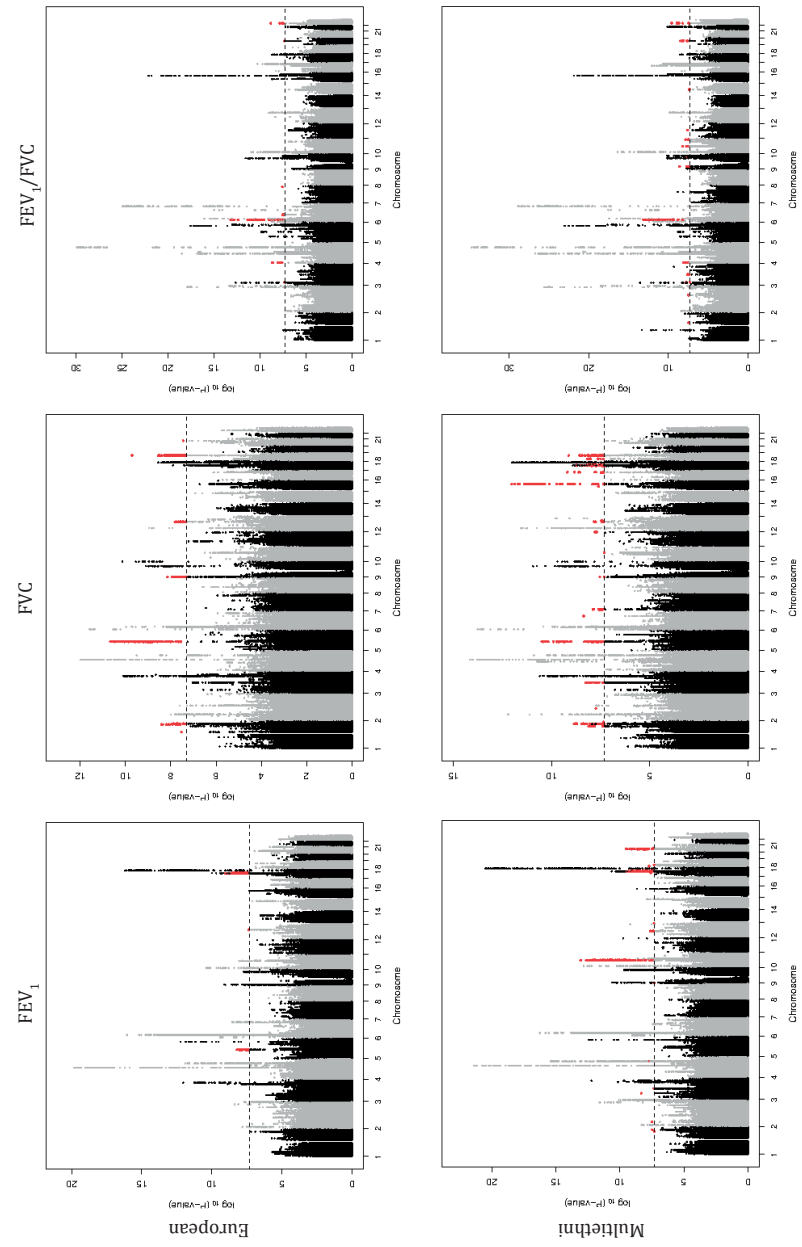

Figure 1
Manhattan plots for genome-wide association results for pulmonary function (FEV1, FVC, and FEV1/FVC) from European and multiethnic meta-analyses in CHARGE. Novel loci indicated by red. Significance level (5x10-8) indicated by dashed line.

Table 3. Top Variants from Novel Loci Discovered in African Ancestry and Hispanic Ethnicity Meta-analyses of Pulmonary Function in the CHARGE Consortium

Nearest Gene(s)[a]	Trait[b]	Top Variant	Chr:Pos	Coded Allele	Allele Freq	N	Beta[c]	SE	P-value
			African Ancestry						
RYR2	FEV_1	rs3766889	1:237941781	t	0.82	8428	52.21	9.52	4.12E-08
C2orf48, HPCAL1	FEV_1/FVC	rs139215025	2:10418806	a	0.01	5653	-0.07	0.01	9.03E-11
EN1, MARCO	FVC	rs114962105	2:119660943	a	0.98	7099	178.48	32.44	3.77E-08
CADPS	FEV_1/FVC	rs111793843	3:62386350	t	0.01	7857	-0.05	0.008	1.97E-08
ANKRD55, MAP3K1	FEV_1	rs11748173	5:55922145	t	0.21	8429	67.07	10.72	3.91E-10
HDC	FEV_1/FVC	rs180930492	15:50555681	t	0.01	3852	-0.07	0.01	2.59E-09
LOC283867, CDH5	FEV_1/FVC	rs144296676	16:66060569	t	0.99	6536	-0.03	0.006	5.35E-09
CPT1C	FEV_1/FVC	rs147472287	19:50213396	t	0.01	5653	-0.05	0.009	3.25E-08
			Hispanic Ethnicity						
DKFZp686O1327, PABPC1P2	FVC	rs6746679	2:147046592	a	0.56	11759	-37.36	6.67	2.17E-08

[a]Nearest gene: indicates gene either harboring the variant or nearest to it.

[b]Phenotypes: FEV_1 forced expiratory volume in 1 second (in ml), FVC forced vital capacity (in ml), Ratio FEV_1/FVC (as a proportion)

[c]Additive effect of variant on pulmonary function, adjusting for age, age^2, sex, height, height^2, smoking status, pack-years of smoking, weight (for FVC only), and center, ancestral principal components, and a random familial effect to account for family relatedness when appropriate.

with some nominal significance (P<0.05) in the remaining ancestry-specific meta-analyses (Supplementary Table 3). For example, rs7899503 in *JMJD1C* was significantly associated with FEV_1 in the multiethnic meta-analysis (β=21.16ml, P=8.70x10^{-14}) and had the following ancestry-specific results: Asian β=28.29ml, P=4.56x10^{-7}; European β=17.35ml, P=1.35x10^{-5}; Hispanic β=19.86ml, P=0.002; African β=29.14ml, P=0.03; I^2=0 and $P_{heterogeneity}$=0.40 across the four ancestry-specific results.

In addition to the fixed-effects multiethnic meta-analysis, we conducted a random-effects meta-analysis using the Han and Eskin method[19] in METASOFT (http://genetics.cs.ucla.edu/meta/) as a sensitivity analysis. In instances where significant heterogeneity is present, the Han-Eskin method mitigates power loss[19]. In the Han-Eskin random-effects model, 37 of the 47 loci identified in the fixed-effects model at P<5x10^{-8} had a p-value below the same threshold (Supplementary Table 4). Among the 10 loci that did not, eight loci still gave a P<5x10^{-7} in the Han-Eskin random-effects model (*PIK3C2B, SUZ12P1, NCOR2/SCARB1, CTAGE1/RBBP8, C20orf112, COMTD1/ZNF503-AS1, EDAR,* and *RBMS3*) while only two did not (*CRADD* and *CCDC41*) (Supplementary Table 4). In addition, there were six loci for FEV_1/FVC that were genome-wide significant in the Han-Eskin random-effects model that had not quite achieved genome-wide significance in the fixed-effects model: G*STO1/GSTO2* (chr10, rs10883990), *FRMD4A* (chr10, rs1418884), *ETFA/SCAPER* (chr15, rs12440815), *APP* (chr21, rs2830155), *A4GNT* (chr3, rs9864090), *UBASH3B* (chr11, rs4935813) (Supplementary Table 4).

X-Chromosome Meta-Analysis

Imputed data for X-chromosome variants was available in 12 studies (ARIC, FHS, CHS, MESA, AGES, ALHS, NEO, RS1, RS2, RS3, JHS, Pelotas; N=43,153). Among these studies, fixed-effects inverse-variance weighted meta-analyses were conducted separately in males and females using METAL and the resulting sex-specific results were combined using a weighted sums approach. No X-chromosome variants were associated with PFTs at genome-wide significance in ancestry-specific or multiethnic meta-analyses. Although the absence of associations between X-chromosome variants and PFTs could reflect the reduced sample size, previous GWAS of pulmonary function have only identified one variant[13].

Look-up Replication of European and Multiethnic Novel Loci

Our primary look-up replication was conducted in the UK BiLEVE study (N=48,943)[14]. Since this study only included individuals of European ancestry, we sought replication only for the 52 novel loci (excluding the major histocompatibility complex, MHC) identified in either the European ancestry or multiethnic discovery meta-analyses. Data for the lead variant was available in the UK BiLEVE study for 51 loci, including 49 loci with a consistent direction of effect between our results and those from UK

Table 4. Top Variants from Novel Loci Discovered in Multiethnic Meta-analysis of Pulmonary Function in the CHARGE Consortium

Nearest Gene(s)[a]	Trait[b]	Top Variant	Chr:Pos	Coded Allele[c]	Allele Freq	N	Beta[d]	SE	P-value
DCAF8	FEV_1/FVC	rs11591179	1:160206067	t	0.45	90624	-0.002	0.0003	3.48E-08
NR5A2	FVC	rs2821332	1:200085714	a	0.47	90642	14.5	2.51	7.65E-09
PIK3C2B	FEV_1	rs12092943	1:204434927	t	0.74	90703	-14.57	2.67	4.83E-08
C1orf140, DUSP10	FVC	rs12046746	1:221635207	c	0.71	90427	-16.99	2.81	1.41E-09
	FEV_1	1:221765779: C_CA	1:221765779	i	0.12	55548	-36.25	6.57	3.38E-08
RYR2	FVC	1:237929787: T_TCA	1:237929787	i	0.11	48215	-37.17	6.79	4.46E-08
PKDCC, EML4	FEV_1	rs963406	2:42355947	a	0.12	80755	-23.13	4.18	3.17E-08
EDAR	FVC	rs17034666	2:109571508	a	0.23	82747	-27.93	4.96	1.81E-08
KCNJ3, NR4A2	FEV_1/FVC	rs72904209	2:157046432	t	0.88	90453	0.003	0.0005	3.09E-08
RBMS3	FEV_1/FVC	rs28723417	3:29431565	a	0.74	90358	0.002	0.0004	1.77E-08
DNAH12	FEV_1	rs79294353	3:57494433	a	0.92	79170	-29.56	5.05	4.82E-09
DCBLD2, MIR548G	FEV_1	rs6778584	3:98815640	t	0.7	90393	12.98	2.37	4.51E-08
	FVC	rs1404098	3:98806782	a	0.71	90334	15.93	2.73	5.45E-09
	FEV_1/FVC	rs80217917	3:99359368	t	0.88	90617	-0.003	0.0005	2.58E-08
AFAP1	FEV_1/FVC	rs28520091	4:7846240	t	0.44	80715	0.002	0.0004	8.40E-09
OTUD4, SMAD1	FEV_1	rs111898810	4:146174040	a	0.20	80752	-20.24	3.61	2.14E-08
AP3B1	FVC	rs72776440	5:77440196	c	0.21	90631	-21.3	3.21	3.20E-11
LINC00340	FEV_1/FVC	rs9350408	6:22021373	t	0.51	82761	-0.003	0.0003	7.45E-14
CENPW, RSPO3	FVC	rs11759026	6:126792095	a	0.72	80687	-20.2	3.44	4.35E-09
AGMO	FVC	rs55905169	7:15506529	c	0.31	90511	-17.57	3.09	1.28E-08
DMRT2, SMARCA2	FVC	rs9407640	9:1574877	c	0.42	80686	-16.82	3.03	2.87E-08
	FEV_1	rs9407640	9:1574877	c	0.41	80754	-14.48	2.65	4.77E-08

Table 4. Top Variants from Novel Loci Discovered in Multiethnic Meta-analysis of Pulmonary Function in the CHARGE Consortium (*continued*)

Nearest Gene(s)[a]	Trait[b]	Top Variant	Chr:Pos	Coded Allele[c]	Allele Freq	N	Beta[d]	SE	P-value
FLJ35282, ELAVL2	FEV_1/FVC	rs10965947	9:23588583	t	0.39	90475	0.002	0.0004	2.70E-09
TMEM38B, ZNF462	FEV_1/FVC	rs2451951	9:109496630	t	0.47	88436	0.002	0.0003	2.36E-08
JMJD1C	FEV_1	rs7899503	10:65087468	c	0.25	90712	21.16	2.84	8.70E-14
	FEV_1/FVC	rs75159994	10:64916064	t	0.77	86988	-0.003	0.0004	6.09E-09
COMTD1, ZNF503-AS1	FVC	10:77002679: TC_T	10:77002679	d	0.22	55498	22.36	4.1	4.89E-08
HTRA1	FEV_1/FVC	rs2293871	10:124273671	t	0.23	90481	0.002	0.0004	1.51E-08
FAM168A	FEV_1/FVC	11:73280955: GA_G	11:73280955	d	0.20	55521	0.004	0.0006	2.74E-08
KIRREL3-AS3, ETS1	FVC	rs73025192	11:127995904	t	0.12	90529	-24.18	4.28	1.63E-08
RAB5B	FEV_1	rs772920	12:56390364	c	0.72	90572	13.86	2.49	2.48E-08
ALX1, RASSF9	FVC	rs7971039	12:85724305	a	0.26	90639	16.36	2.88	1.44E-08
CRADD	FVC	rs11107184	12:94184082	t	0.34	88548	14.89	2.71	3.87E-08
CCDC41	FVC	rs10859698	12:94852628	a	0.21	88159	21.19	3.84	3.49E-08
NCOR2, SCARB1	FEV_1	rs11057793	12:125230287	t	0.75	78930	17.66	3.24	4.78E-08
DDHD1, MIR5580	FEV_1/FVC	rs4444235	14:54410919	t	0.54	80712	0.002	0.0004	4.03E-08
SQRDL, SEMA6D	FVC	rs4775429	15:46722435	t	0.17	79231	40.23	7.21	2.45E-08
SMAD3	FVC	rs8025774	15:67483276	t	0.29	88524	-20.87	2.92	9.34E-13
PDXDC2P	FVC	rs3973397	16:70040398	a	0.48	44921	-22.38	4.05	3.31E-08
PMFBP1, ZFHX3	FVC	rs55771535	16:72252097	a	0.13	80688	-29.88	4.83	6.38E-10
SUZ12P1	FEV_1	rs62070631	17:29087285	a	0.15	82835	20.26	3.64	2.57E-08
MED1, CDK12	FVC	rs8067511	17:37611352	t	0.8	90632	18.3	3.2	1.08E-08
LOC644172, CRHR1	FEV_1	rs186806998	17:43682323	t	0.82	43927	29.5	4.7	3.47E-10
	FVC	rs150741403	17:43682405	c	0.85	43896	35.83	5.97	1.94E-09

Table 4. Top Variants from Novel Loci Discovered in Multiethnic Meta-analysis of Pulmonary Function in the CHARGE Consortium (*continued*)

Nearest Gene(s)[a]	Trait[b]	Top Variant	Chr:Pos	Coded Allele[c]	Allele Freq	N	Beta[d]	SE	P-value
WNT3	FEV_1	rs199525	17:44847834	t	0.8	80753	18.85	3.08	9.59E-10
	FVC	rs199525	17:44847834	t	0.8	80686	20.32	3.52	7.52E-09
SOGA2	FEV_1	rs513953	18:8801351	a	0.29	82871	-14.5	2.58	1.96E-08
CTAGE1, RBBP8	FEV_1	rs7243351	18:20148531	t	0.45	90708	12.31	2.25	4.69E-08
CABLES1	FVC	rs7238093	18:20728158	a	0.22	90240	18.15	3.13	6.78E-09
DCC	FVC	rs8089865	18:50957922	a	0.53	90578	15.81	2.57	7.38E-10
TSHZ3	FEV_1/FVC	rs9636166	19:31829613	a	0.86	80714	0.003	0.0005	3.25E-09
ZNF337	FEV_1	rs6138639	20:25669052	c	0.79	90593	17.91	2.85	3.17E-10
C20orf112	FEV_1	rs1737889	20:31042176	t	0.22	80755	-16.82	3.07	4.17E-08
KLHL22, MED15	FEV_1/FVC	rs4820216	22:20854161	t	0.13	82714	-0.003	0.0005	2.61E-10

[a]Nearest gene: indicates gene either harboring the variant or nearest to it. *JMJD1C* locus also includes *EGR2, NRBF2, JMJD1C-AS1,REEP3*. *HTRA1* locus also includes *DMBT1*. *RAB5B* locus also includes *SOUX*. *SMAD3* locus also includes *AAGAB, IQCH*. *MED1/CDK12* locus also includes *FBXL20*. LOC644172/*CRHR1* locus also includes *ARHGAP27, MGC57346, CRHR1-IT1, LRRC37A4P*. *ZNF337* locus also includes *ABHD12, PYGB, GINS1, NINL, NANP, FAM182B, LOC100134868. KLHL22/MED15* locus also includes *ZNF74, SCARF2*. *C1orf140/DUSP10* locus also includes *HLX*.

[b]Phenotypes: FEV_1 forced expiratory volume in 1 second (in ml), FVC forced vital capacity (in ml), Ratio FEV_1/FVC (as a proportion)

[c]Alleles for INDELS: I Insertion, D Deletion

[d]Additive effect of variant on pulmonary function, adjusting for age, age^2, sex, height, height^2, smoking status, pack-years of smoking, weight (for FVC only), and center, ancestral principal components, and a random familial effect to account for family relatedness when appropriate.

BiLEVE (Supplementary Table 5). Based on a two-sided $P<9.6x10^{-4}$ (0.05/52), 15 loci replicated for the same trait based on the lead variant from our analysis: *DCBLD2/ MIR548G, SUZ12P1, CRHR1, WNT3, ZNF337, ALX1/RASSF9, MED1/CDK12, EYA2, RBMS3, LINC00340, FLJ35282/ELAVL2, DDHD1/MIR5580, TSHZ3, KLHL22/MED15, FAM168A* (Supplementary Table 5). It was recently demonstrated that using one-sided replication p-values in GWAS, guided by the direction of association in the discovery study, increases replication power while being protective against type 1 error compared to the two-sided p-values[20]; under this criterion, an additional four loci replicated for the same trait based on the lead variant: *RAB5B, JMJD1C, AGMO,* and *C20orf112* (Supplementary Table 5).

We also conducted a secondary look-up replication for European ancestry and multiethnic lead variants in the much larger UK BioBank study (*N*=255,492 with PFTs) from which the UK BiLEVE study is sampled. Unlike the UK BiLEVE results which were adjusted for age, age^2, sex, height, pack-years of smoking, and ancestral principal components[14], the publicly available UK BioBank results (https://sites.google.com/ broadinstitute.org/ ukbbgwasresults/home) are only adjusted for sex and ancestral principal components. In addition, only results for FEV_1 and FVC (not the ratio FEV_1/ FVC) were currently available. Nevertheless, this secondary look-up yielded evidence of replication for the same trait for an additional 9 loci with a two-sided $P<9.6x10^{-4}$: *NR5A2, PIK3C2B, OTUD4/SMAD1, AP3B1, CENPW/RSPO3, SMAD3, PDXDC2P, SOGA2, DCC* (Supplementary Table 5). Another locus also replicated for the same trait with a one-sided $P<9.6x10^{-4}$ (*DNAH12*) and another discovered for FEV_1/FVC also replicated for FEV_1 and FVC (*KCNJ3/NR4A2*) in the UK Biobank data. In summary, we found evidence of replication in UK BiLEVE or UK Biobank for 30 novel loci.

Look-up Replication of African and Hispanic Novel Loci

Look-up replication of lead variants for novel African ancestry loci was sought in three smaller studies of African Americans: COPDGene (*N*=3,219)[21,22], SAPPHIRE (*N*=1,707)[23,24], and SAGE (*N*=1,405; predominantly children)[25]. We did not find evidence of replication for most of the African ancestry loci identified in our study (Supplementary Table 6). This could possibly reflect low power given the smaller sample sizes of the external studies combined with the low minor allele frequencies (MAF) of most (six out of eight) of the African ancestry variants. We found the strongest evidence for replication for *RYR2* (rs3766889). This variant was common (MAF=0.18) and well imputed ($r^2>0.90$) in CHARGE. The effect size was similar across CHARGE (β=52.21ml, $P=4.12x10^{-8}$) and the two adult replication studies (COPDGene β=46.85ml, *P*=0.03 and SAPPHIRE β=22.00ml, *P*=0.32); meta-analysis of these adult studies resulted in a significant combined association (β=47.35ml, SE=8.00ml, $P=3.30x10^{-9}$). In SAGE, which includes mostly children and examined percent predicted values, the result was

in the opposite direction and not significant. In our Hispanic ethnicity meta-analysis, the lead variant from the single novel locus (rs6746679, *DKFZp686O1327/PABPC1P2*) did not replicate in two smaller external studies of Hispanics: MESA (*N*=806; MESA Hispanics not included in discovery) and GALA II (*N*=2,203; predominantly children)[26] (Supplementary Table 6).

Overlap of Newly Identified Loci with COPD

Pulmonary function measures are the basis for the diagnosis of COPD, an important clinical outcome; therefore, we also looked-up the 52 novel loci identified in the European ancestry or multiethnic meta-analyses in the International COPD Genetics Consortium (ICGC). This consortium recently published a meta-analysis of 1000 Genomes imputed variants and COPD primarily among individuals of European ancestry (*N*= 15,256 cases and 47,936 controls), including some of the same individuals included in the present lung function analysis[27]. Ten lead variants representing 8 novel loci were associated with COPD at $P<9.6\times10^{-4}$: *RBMS3, OTUD4/SMAD1, TMEM38B/ZNF462, NCOR2/SCARB1, SUZ12P1, WNT3, SOGA2, C20orf112* (Supplementary Table 7). Directions of effects were consistent between our results and the COPD findings for these variants (i.e. variants associated with increased pulmonary function values were associated with decreased odds of COPD and vice-versa). Our top variant in *SOGA2* (also known as *MTCL1*) is in LD (R^2=0.8) with the top variant for COPD as reported by the IGCG Consortium[27].

eQTL and mQT Signals

To query whether novel loci contained variants associated with gene expression (eQTLs), we looked-up variants from all 60 novel loci identified in any ancestry-specific or multiethnic meta-analyses in the following datasets: 1) lung samples from 278 individuals in GTEx (https://www.gtexportal.org/home/)[28]; 2) lung samples from 1,111 participants in studies from the Lung eQTL Consortium including Laval University, the University of Groningen and the University of British Columbia[29-31]; 3) whole blood samples from 5,257 Framingham Heart Study participants[32]; 4) peripheral blood samples from 5,311 participants in EGCUT, InCHIANTI, Rotterdam Study, Fehrmann, HVH, SHIP-TREND and DILGOM[33]; and 5) peripheral blood samples from 2,116 participants in 4 Dutch studies collectively known as BIOS[34,35]. We examined both whole blood and lung datasets to take advantage of the much larger size, and higher statistical power, of available blood eQTL datasets since we have previously found substantial overlap between lung and blood eQTLs for lung function GWAS loci, as well as complementary information from these two different tissues[29]. The Lung eQTL Consortium study used a 10% FDR cut-off, while all other studies used a 5% FDR cut-off (see Supplementary Note 1 for further study descriptions and methods).

A significant cis-eQTL in at least one dataset was found for 34 lead variants representing 27 novel loci (Supplementary Table 8). Of these, 13 loci had significant cis-eQTLs only in datasets with blood samples and three loci only in datasets with lung samples (*COMTD1/ZNF503-AS1, FAM168A, SOGA2*). Eleven loci had significant cis-eQTLs in both blood and lung samples based on lead variants, with one locus having a significant cis-eQTL across all five datasets (*SMAD3*) and another four loci having a significant cis-eQTL in four datasets (*RAB5B, CRHR1, WNT3, ZNF337*). Significant trans-eQTLs in at least one dataset were found for seven lead variants representing four novel loci (*TMEM38B/ZNF462*, *RAB5B, CRHR1*, and *WNT3*, Supplementary Table 8).

In addition, mQTL data were available from FHS and BIOS. Significant cis-mQTLs and trans-mQTLs in at least one dataset were found for 52 lead variants (43 novel loci) and one lead variant (one novel locus), respectively (Supplementary Table 8).

None of the novel variants discovered for African and Hispanic ancestry indicated significant cis-eQTLs in GTex which includes some slight multiethnic representation (12% African American and 3% other races/ethnicities). Although few ancestry-specific eQTL datasets exist, we identified two such studies with blood samples from African American participants: SAPPHIRE (*N*=597) and MESA (*N*=233)[36]. In SAPPHIRE, none of the eight African ancestry variants identified in the meta-analysis indicated significant cis-eQTLs (FDR<0.05), but rs180930492 was associated with a trans-eQTL among individuals without asthma and rs111793843 and rs139215025 were associated with trans-eQTLs among individuals with asthma at FDR<0.05 (Supplementary Table 9). In MESA, eQTL data were available for only three of the African ancestry variants (rs11748173, rs3766889, rs144296676) and none indicated significant (FDR<0.05) cis-eQTLs (Supplementary Table 9).

Heritability and Genetic Correlation

We used LD Score regression[37] to estimate the variance explained by genetic variants investigated in our European ancestry meta-analysis, also known as SNP heritability. Across the genome, the SNP heritability (narrow-sense) was estimated to be 20.7% (SE 1.5%) for FEV_1, 19.9% (SE 1.4%) for FVC and 17.5% (SE 1.4%) for FEV_1/FVC.

We also partitioned heritability by functional categories to investigate whether particular subsets of common variants were enriched[38]. We found significant enrichment in 6 functional categories for all three PFTs: conserved regions in mammals, DNase I hypersensitive sites (DHS), super enhancers, the histone methylation mark H3K4me1 and histone acetylation marks H3K9Ac and H3K27Ac (Supplementary Figure 4). Another seven categories showed enrichment for at least one PFT (Supplementary Figure 5). We observed the largest enrichment of heritability (14.5-15.3 fold) for conserved regions in mammals, which has ranked highest in previous partitioned heritability analyses for other GWAS traits (Supplementary Figure 5)[38].

Since both height and smoking are important determinants of pulmonary function, and have been associated with many common variants in previous GWAS, we also used LD score regression to investigate genetic overlap[39] between our FEV_1, FVC and FEV_1/FVC results and publicly available GWAS results of ever smoking[40] and height[41]. No significant genetic correlation was found between PFTs and smoking or height (Supplementary Table 10), indicating our findings are independent of those traits.

In addition, we used LD Score regression to investigate genetic overlap between each PFT and the other two PFTs, as well as with asthma. Based on the overall PFT results presented in this paper, we found significant genetic correlation between FEV_1 and FVC (P<0.001) and between FEV_1 and FEV_1/FVC (P<0.001), but not between FVC and FEV_1/FVC (P=0.23; Supplementary Table 10). Since measures of FEV_1 and FVC (independent of genetics) are highly correlated, and to lesser degree FEV1/FVC[10], these results are not surprising. Using publicly available GWAS results for asthma[42], we also found significant correlation between PFTs and asthma (P<0.003; Supplementary Table 10).

Functional Annotation

For functional annotation, we considered all novel variants with $P<5x10^{-8}$ from the 60 loci discovered in our ancestry-specific and multiethnic meta-analyses, plus significant variants from the MHC region, two loci previously discovered in the CHARGE exome chip study (*LY86/RREB1* and *SEC24C*)[43] and *DDX1*. Using Ensembl VEP[44], we found six missense variants in four loci outside of the MHC region and 22 missense variants in the MHC region (Supplementary Table 11). Of the 28 total missense variants, two (chr15:67528374 in *AAGAB* and chr6:30899524 in the MHC region) appear to be possibly damaging based on SIFT[45] and PolyPhen-2[46] scores (Supplementary Table 11). Using CADD[47], we found an additional 28 deleterious variants from 15 loci based on having a scaled C-score greater than 15 (Supplementary Data 1). In the MHC region, we found another 11 deleterious variants based on CADD. Based on RegulomeDB[48], which annotates regulatory elements especially for non-coding regions, there were 52 variants from 18 loci with predicted regulatory functions (Supplementary Data 1). In the MHC region, there were an additional 72 variants with predicted regulatory functions.

Pathway Enrichment Analysis

Gene set enrichment analyses conducted using DEPICT[49] on genes annotated to variants with $P<1x10^{-5}$ based on the European ancestry meta-analysis results revealed 218 significantly enriched pathways (FDR<0.05) (Supplementary Data 2). The enriched pathways were dominated by fundamental developmental processes, including many involved in morphogenesis of the heart, vasculature, and lung. Tissue and cell type analysis noted significant enrichment (FDR<0.05) of smooth muscle, an important

component of the lung (Supplementary Table 12, Supplementary Figure 6). We found 8, 1, and 82 significantly prioritized genes (FDR<0.05) for FEV_1, FVC, and FEV_1/FVC, respectively (Supplementary Data 3). Given the generally smaller p-values for variants associated with FEV_1/FVC, enriched pathways and tissue/cell types as well as prioritized genes were predominantly discovered from DEPICT analyses of FEV_1/FVC.

Due to extended LD across the MHC locus on chromosome 6 (positions 25000000 to 35000000), DEPICT excludes this region[49]. Standard Ingenuity Pathway Analysis (IPA) run without excluding the MHC highlighted 16 enriched networks based on the European ancestry meta-analysis results, including three involved in inflammatory diseases or immunity; 33 of the 84 genes in these three networks are in the MHC region (Supplementary Table 13). IPA based on the multiethnic results highlighted 21 enriched networks, including similar inflammatory and immunity related networks (Supplementary Table 14).

Identification of Potential Causal Variants using PAINTOR

Using a multiethnic fine-mapping analysis incorporating strength of association, variation in genetic background across major ethnic groups, and functional annotations in PAINTOR[50], we examined 40 loci that contained at least five genome-wide significant variants in the European ancestry and multiethnic meta-analyses or at least one significant variant in the African ancestry or Hispanic/Latino ethnicity meta-analyses. We identified 15 variants representing 13 loci as having high posterior probabilities of causality (>0.8): 3 for FEV_1, 3 for FVC, and 9 for FEV_1/FVC (Supplementary Table 15, Supplementary Figure 7). Of the 15 putative casual variants, 11 showed high posterior probabilities of causality (>0.8) before considering annotations and 4 were identified by adding functional annotations. Nine were the top SNPs at that locus from the fixed-effects meta-analysis (loci: *WNT3*, *PMFBP1/ZFHX3*, *EN1/MARCO*, *C2orf48/HPCAL1*, *CPT1C*, *CADPS*, *LOC283867/CDH5*, *HDC*, and *CDC7/TGFBR3*), while 6 were not (loci: *CDK2/RAB5B*, *BMS1P4*, *PMFBP1/ZFHX3*, *FLJ35282/ELAVL2*, *HDC*, and *COL8A1*).

Identification of Independent Signals using FINEMAP

We used FINEMAP[51] to identify variants with a high posterior probability of causality (>0.6) independent of 118 lead variants in pulmonary function loci identified in the current or previous studies[14]. We identified 37 independent variants for 23 previously identified loci and one independent variant at each of two novel loci (*LINC00340* and *SLC25A51P1/BAI3*; Supplementary Table 16).

Gene-based Analysis of GWAS Results using S-PrediXcan

Among the novel loci identified in the current GWAS of PFTs, we identified seven variants corresponding to nine genes demonstrating genome-wide significant evidence of association with lung or whole blood tissue-specific expression (Supplementary Table 17) based on the gene-based S-PrediXcan approach[52]. Bayesian colocalization analysis[53] indicated the following associations demonstrated at least 50% probability of shared SNPs underlying both gene expression and PFTs: *ARHGEF17* and *FAM168A* in analysis of multiethnic GWAS for FEV_1/FVC based on GTEx whole blood models, and *WNT3* in analysis of multiethnic GWAS for FVC based on GTEx lung models (Supplementary Table 18).

Druggable Targets

To investigate whether the genes identified in our study encode proteins with predicted drug targets, we queried the ChEMBL database (https://www.ebi.ac.uk/chembl/). In addition, we incorporated an IPA to identify potential upstream targets. Genes associated with pulmonary function, but not included in the drug target analysis performed by Wain et al[14], were evaluated, for a total of 139 genes outside of the MHC: 110 genes representing the 60 novel loci identified in our fixed-effects ancestry-specific and multiethnic meta-analysis, 13 genes representing the 6 novel loci identified in our random effects meta-analysis[19], 3 genes representing an additional 3 loci near significance in the African ancestry meta-analysis (*BAZ2B, NONE/PCDH10, and ADAMTS17*), 9 genes representing 2 loci identified in a recent CHARGE analysis of exome variants[43] which were also significant in our 1000 Genomes analysis (*LY86/RREB1* and *SEC24C*), and 4 genes representing one locus identified at genome-wide significance in a separate publication from one of our participating studies (HCHS/SOL)[18] but also significant in our analysis (*ADORA2B/ZSWIM7/TTC19/NCOR1*). In the ChEMBL database, 17 of these genes encode proteins with predicted or known drug targets: *NR5A2, KCNK2, EDAR, KCNJ3, NR4A2, BAZ2B, A4GNT, GSTO1, GSTO2, NCOR2, SMAD3, NCOR1, CDK12, WNT3, PYGB, NANP, EYA2* (Supplementary Table 19). Of these, two genes (*KCNK2 and CDK12)* have approved drug targets. Using IPA, four additional genes were predicted as drug targets (*ADORA2B, APP, CRHR1, and MAP3K1*; Supplementary Table 20) and 37 genes had drugs or chemicals as upstream regulators (Supplementary Table 21).

DISCUSSION

By conducting a GWAS meta-analysis in a large multiethnic population we increased the number of known loci associated with pulmonary function by over 50%. In total,

we identified 60 novel genetic regions (outside of the MHC region): 17 from European ancestry, 8 from African ancestry, 1 from Hispanic/Latino ethnicity, and 34 from multiethnic meta-analyses.

For 32 of the 52 loci novel loci identified in our European ancestry and multiethnic meta-analyses, we found evidence for look-up replication in the UK BiLEVE study, UK Biobank study, or ICGC COPD consortium. For an additional three loci, we found support for validation using new genomic methods such as PAINTOR, FINEMAP, or S-PrediXcan. Specifically, 19 novel variants replicated in look-up in a smaller independent sample of Europeans from the UK BiLEVE study[14]: *DCBLD2/MIR548G, SUZ12P1, CRHR1, WNT3, ZNF337, ALX1/RASSF9, MED1/CDK12, EYA2, RBMS3, LINC00340, FLJ35282/ELAVL2, DDHD1/MIR5580, TSHZ3, KLHL22/MED15, FAM168A*, *RAB5B, JMJD1C, AGMO,* and *C20orf112*. Based on a minimally adjusted publicly available analysis in a larger sample of Europeans from the UK Biobank, an additional 11 loci replicated: *NR5A2, PIK3C2B, OTUD4/SMAD1, AP3B1, CENPW/RSPO3, SMAD3, PDXDC2P, SOGA2, DCC*, *DNAH12*, and *KCNJ3/NR4A2*. Because UK BiLEVE is sampled from UK Biobank we are not able to perform a combined replication meta-analysis. Additionally, the studies adjusted for different covariates (UK BiLEVE results were adjusted for age, sex, height, pack-years and ancestral principal components while UK Biobank results were adjusted for only sex and ancestral components). Among those loci which did not directly replicate for PFTs in the UK BiLEVE or UK Biobank datasets, the lead variants in an additional two European or multiethnic loci were significantly associated in the ICGC Consortium with COPD, which was defined using PFT measures[27]: *TMEM38B/ZNF462* and *NCOR2/SCARB1*. FINEMAP and S-PrediXcan also produced evidence for causality for three European ancestry and multiethnic loci which had not replicated in UK BiLEVE, UK Biobank or ICGC: *DCAF8, AFAP1*, and *SLC25A51P1/BAI3*.

The few additional studies with 1000 Genomes imputed variants and pulmonary function in African ancestry individuals have smaller samples sizes making replication challenging for the eight novel loci identified in our African ancestry meta-analyses. Further, lead variants for six of the eight loci were low frequency in African-Ancestry (*C2orf48/HPCAL1, EN1/MARCO, CADPS, HDC, LOC283867/CDH5*, and *CPT1C*) (MAF<0.02), including three not well imputed (r^2<0.75), and monomorphic or nearly monomorphic in other ancestries (European, Asian, and Hispanic). For the two novel African ancestry variants with MAF>0.02 and well imputed (r^2>0.90), we found the strongest evidence for replication for *RYR2* (rs3766889). This variant had a similar effect estimate for FEV_1 in CHARGE, COPDGene, and SAPPHIRE with a significant combined association across these adult studies. Although this particular variant did not quite meet genome-wide significance in the multiethnic meta-analysis for FEV_1 (P=6.56x10^{-4}), another variant in this gene did for FVC (1:237929787:T_TCA, p=4.46x10^{-8}).

Our analysis also sheds light on additional potential causal genes at a complex locus (chromosome 17 near positions 43600000 to 44300000, hg19) previously discovered from GWAS of FEV_1 which identified *KANSL1* in European populations as the top finding for this region[14,15]. With the exception of a single INDEL in *KANSL1* in our European ancestry meta-analysis (17:44173680:T_TC, P=1.03x10^{-10}), we found *CRHR1* as the strongest gene associated with FEV_1 in this region. Although some variants in *CRHR1* identified in our study are within 500kb of *KANSL1* (e.g., rs16940672, 17:43908152, P=2.07x10^{-10}), a number of significant variants in this gene are more than 500kb away from previously identified hits [our definition of novel] (e.g., rs143246821, 17:43685698, P=9.06x10^{-10}). In our multiethnic meta-analysis, several variants in *CRHR1* were associated with FEV_1 at smaller p-values than variants in *KANSL1*. Definitive assessment of the causal variants at this locus, as well as other multigenic GWAS loci, will likely require additional data from ongoing large-scale sequencing studies to enable detailed fine mapping.

In both our European and multiethnic meta-analyses we also noted a significant association with *WNT3* on chromosome 17 near position 44800000 (hg19) which is more than 500kb from *KANSL1* or *CRHR1* [our definition of novel]. We found that the top variant in *WNT3* for FEV_1 among individuals of European ancestry (rs916888, 17:44863133, P=3.76x10^{-9}) had a high probability for causality based on PAINTOR, an analysis which integrates functional annotations along with association statistics and LD for each ethnicity[50]. We also found evidence that *WNT3* may be the causal gene at this locus using S-PrediXcan, a gene level association test that prioritizes potentially causal genes while filtering out LD-induced false positives[52,53]. Notably, S-PrediXcan implicated *WNT3* as a likely mediating gene for FVC based on the top variant in our multiethnic meta-analyses (rs199525, 17:44847834, P=7.52x10^{-9}), which is an eQTL SNP for *WNT3* in lung and other tissues. Further, the lead *WNT3* variants for both FEV_1 and FVC (rs916888 and rs199525) were significantly associated with COPD in a look-up of a large published meta-analysis dataset[27]. In addition, other genes in the *WNT* signaling pathway, a fundamental development pathway, have been implicated as influencing pulmonary function[54]. This pathway was also one of the significant pathways identified in our analysis. In a previous pathway analysis of asthma, *SMAD3* has been shown to interact with the *WNT* signaling pathway[55]. Finally, *WNT3* also emerged as having a potential druggable target, and incorporation of pathway analysis to identify upstream regulators found an additional four drugs in clinical use for which *WNT3* is a target molecule (chemotherapeutic agents doxorubicin and paclitaxel, the hormone beta-estradiol and LGK-974, a novel agent that targets a WNT-specific acyltransferase)[56]. Again, further evaluation of this interesting and complex locus which contains many significant variants in LD will benefit from data being generated in ongoing large-scale sequencing studies.

Some genes identified in our study play key roles in inflammation, immunity and pulmonary biology. For example, *MARCO* (macrophage receptor with collagenous structure) has been shown in murine models to be required for lung defense against pneumonia and inhaled particles[57,58]. *SMAD3* is part of the SMAD family of proteins which are signal transducers and transcriptional modulators that mediate multiple signaling pathways. *SMAD3* is activated by transforming growth factor beta (TGF-B) which plays a key role in airway remodeling. S*MAD3* has a predicted drug target and SNPs in *SMAD3* are significantly associated with asthma in GWAS[42,59].

Other genes identified in our study that are targeted by approved drugs include *CDK12* and *KCNK2*. *CDK12* drug targets include AT-7519, Roniciclib, AZD-5438, and PH.A-793887. Roniciclib has been used in clinical trials including lung cancer patients[60]. *KCNK2* (potassium channel subfamily K member 2) is targeted by five inhalational anesthetic agents. These agents have anti-inflammatory effects both systemically[61] and in the lungs[62] and meta-analysis of clinical studies shows protection against pulmonary complications after cardiac surgery[63]. A recent trial suggested that one of these inhalation agents, sevoflurane, offers promise for reducing epithelial injury and improving outcomes in patients with acute respiratory distress syndrome[64].

In addition to querying commonly used genome databases for functional annotation of variants, we sought to narrow down causal variants in implicated loci using recently developed methods that incorporate LD, functional data and/or the multiethnic analysis done in this paper. In particular, PAINTOR is a useful tool to identify potential causal variants in our novel loci as it leverages LD across ancestral groups along with association statistics and functional annotations[50]. PAINTOR identified 15 putative causal variants from 13 loci, including seven loci uniquely identified in the multiethnic meta-analyses such as *PMFBP1/ZFHX3* and *COL8A1* (part of the *DCBLD2* loci). Several of the putative causal variants from PAINTOR were the top SNPs from the fixed-effects meta-analysis (e.g., rs916888 *WNT3*). Similarly, FINEMAP has been shown to be an accurate and efficient tool for investigating whether lead SNPs for a given loci are driven by independent variants in the same region, especially when annotation information is not available[51]. Among previous and novel loci identified in individuals of European ancestry, we identified 37 independent variants for 23 previously identified loci and two lead variants for two novel loci (rs1928168 *LINC00340* and rs9351637 *SLC25A51P1/BAI3)* with a high probability of causality. Finally, we ran S-PrediXcan a gene level association test that prioritizes potentially causal genes[52]. Seven of our novel loci contained putative causal genes based on S-PrediXcan for lung or whole blood tissues, including *NRBF2* (part of the *JMJD1C* locus) and *WNT3*. S-PrediXcan also highlighted the region around chromosome 11 position 73280000 (hg19), noting strong evidence for both *FAM168A* and *ARHGEF17* which was further supported by the co-localization analysis. Interestingly, DEPICT also prioritized *ARHGEF17*, a member

of the guanine nucleotide exchange factor (GEF) family of genes which can mediate actin polymerization and contractile sensitization in airway smooth muscle[65,66].

Rather than conducting a standard gene-based pathway analysis, we performed a newer integrative method, DEPICT, that incorporates cell and tissue-specific functional data into a pathway analysis to prioritize genes within implicated loci[49]. In addition to identifying potential causal variants, this approach revealed a number of fundamental development processes, including pathways related to lung development, growth regulation, and organ morphogenesis. The *WNT* signaling pathway was also highlighted along with processes relevant to the pathogenesis of COPD including extracellular matrix structure and collagen networks. Tissue/cell type enrichment results highlighted smooth muscle which is highly relevant for lung function. DEPICT excludes the MHC due to extended LD in this region, which likely explains the relative paucity of inflammation-related pathways identified compared to previous pathway analyses in GWAS of PFTs[29,54]. Indeed, standard IPA analysis of our data including the MHC region, found that 33 of 84 genes (39%) in the 3 (out of 16) enriched networks involved in immune or inflammatory processes are in the MHC. The predominance of fundamental pathways related to lung growth, differentiation and structure is consistent with recent work[67] that has rekindled interest in the observation made 40 years ago[68] that individuals can cross the threshold for diagnosis of COPD either by rapid decline in adulthood or by starting from a lower baseline of maximal pulmonary function attained during growth. Within this context, understanding the genetic (and environmental) factors that influence the variability in maximal lung function attained during the first three decades of life is essential to reducing the public health burden of COPD[69].

In summary, our study extends existing knowledge of the genetic landscape of PFTs by utilizing the more comprehensive 1000 Genomes imputed variants, increasing the sample size, including multiple ancestries and ethnicities, and employing newly developed computational applications to interrogate implicated loci. We discovered 60 novel loci associated with pulmonary function and found evidence of replication in UK BiLEVE, UK Biobank, or ICGC for 32 novel loci and validation for another 3 loci. We found evidence that several variants in these loci were missense mutations and had possible deleterious or regulatory effects, and many had significant eQTLs. Further, using new genomic methods that incorporate LD, functional data and the multiethnic structure of our data, we shed light on potential causal genes and variants in implicated loci. Finally, several of the newly identified genes linked to lung function are druggable targets, highlighting the clinical relevance of our integrative genomics approach.

METHODS

Studies

Member and affiliate studies from The Cohorts for Heart and Aging Research in Genomic *Epidemiology* (CHARGE) consortium with pulmonary function and 1000 Genomes imputed genetic data were invited to participate in the present meta-analysis. Participating studies included: AGES, ALHS, ARIC, CARDIA, CHS, FamHS, FHS, GOYA, HCHS/SOL, HCS, Health ABC, Healthy Twin, JHS, KARE3, LifeLines, LLFS, MESA, NEO, 1982 PELOTAS, RSI, RSII, RIII. Characteristics of these studies are provided in Supplementary Table 22 and descriptions of study designs are provided in the Supplementary Note 1; informed consent was obtained from participants in each study. Although our meta-analysis included studies of asthma (ALHS) and obesity (GOYA and NEO), exclusion of these studies did not materially change results (Supplementary Note 2). Further, previous meta-analyses of GWAS of pulmonary function have demonstrated high correlation between results when including or excluding asthma and COPD cases[8].

Pulmonary Function

Spirometry measures of pulmonary function (FEV_1, FVC, and the ratio FEV_1/FVC) were collected by trained staff in each study according to American Thoracic Society (ATS) or European Respiratory Society guidelines. See cohort descriptions in Supplementary Note 1 for more details.

Variants

Studies used various genotyping platforms, including Affymetrix Human Array 6.0, Illumina Human Omni Chip 2.5, and others, as described in cohort descriptions in the Supplementary Note 1. Using MACH, MINIMAC, or IMPUTE2, studies then used genotyped data to impute variants based on the 1000 Genomes Integrated phase 1 reference panel. One study (Hunter Community) imputed to the 1000 Genomes European phase 1 reference panel; sensitivity analyses excluding this study from the European ancestry meta-analysis showed no material differences (see Supplementary Note 2). The two Asian studies (Healthy Twin and KARE3) imputed to the 1000 Genomes Asian phase 1 reference panel.

Statistical Analysis

Within each study, linear regression was used to model the additive effect of variants on PFTs. FEV_1 and FVC were modeled as milliliters and FEV_1/FVC as a proportion. Studies were asked to adjust analyses for age, age^2, sex, height, height^2, smoking status (never, former, current), pack-years of smoking, center (if multicenter study), and ancestral principal components, including a random familial effect to account for family

relatedness when appropriate[70]. Models of FVC were additionally adjusted for weight. Analyses were conducted using ProbAbel, PLINK, FAST, or the R kinship package as described in the cohort descriptions of the Supplementary Note 1.

Ancestry-specific and multiethnic fixed effects meta-analyses using inverse variance weighting of study-specific results with genomic control correction were conducted in Meta Analysis Helper (METAL, http://www.sph.umich.edu/csg/abecasis/metal/). Multiethnic random effects meta-analyses using the four ancestry-specific fixed effects meta-analysis results were conducted using the Han-Eskin model[19] in METASOFT (http://genetics.cs.ucla.edu/meta/). Only variants with p-values for association <0.05 or p-values for heterogeneity <0.1 from fixed-effects models were included in the random-effects models.

Variants with imputation quality scores (r^2) less than 0.3 and/or a minor allele count (MAC) less than 20 were excluded from each study prior to meta-analysis. Following meta-analysis, we also excluded variants with less than one-third the total sample size or less than the sample size of the largest study for a given meta-analysis to achieve the following minimal sample sizes: 20,184 for European ancestry; 2,810 for African ancestry; 7,862 for Asian ancestry; 4,435 for Hispanic/Latino ethnicity and 30,238 for Multiethnic.

Significance was defined as $P<5x10^{-8}$ [14,17]. Novel variants were defined as being more than +/-500kb from the top variant of a loci identified in a previous GWAS of pulmonary function; previous multiethnic GWAS have used this definition[16]. We used the list of 97 known variants as published in the recent UK BiLEVE paper[14] with the following modifications: added variants in *DDX1, DNER, CHRNA5* since listed in GWAS catalog; added variants in *LCT, FGF10, LY86/RREB1, SEC24C, RPAP1, CASC17*, and *UQCC1* since identified in exome chip paper[43]; added variant in *TMEM163* identified in Loth et al paper[10]; used 17:44339473 instead of 17:44192590 to represent *KANSL1* since 17:44339473 was the original variant listed for *KANSL1* in Wain et al 2015[15]; and used 12:28283187 instead of 12:28689514 to represent *PTHLH* since 12:28283187 was the original variants listed for *PTHLH* in Soler Artigas et al 2015[13].

Genomic inflation factors (lambda values) from quantile-quantile plots of observed and expected p-values for ancestry- and phenotype-specific meta-analyses are presented in Supplementary Table 23. Lambda values were slightly higher in European and multiethnic meta-analyses (range of lambda 1.12 to 1.16) than in other ancestry-specific meta-analyses (range of lambda 1.01 to 1.06) likely due to the much larger sample sizes of the European and multiethnic meta-analyses[71].

LD Score Regression

The SNP heritability, i.e. the variance explained by genetic variants, was calculated from the European ancestry GWAS summary statistics (with genomic control off) using

LD Score regression (https://github.com/bulik/ldsc)[37]. Partitioned heritability was also calculated using the method described by Finucane and colleagues[38]. In total, 28 functional annotation classes were used for this analysis, including: coding regions, regions conserved in mammals, CCCTC-binding factor (CTCF), DNase genomic foot printing (DGF), DNase I hypersensitive sites (DHS), fetal DHS, enhancer regions; including super-enhancers and active enhancers from the FANTOM5 panel of samples, histone marks including two versions of acetylation of histone H3 at lysine 27 (H3K27ac and H3K27ac2), histone marks monomethylation (H3K4me1), trimethylation of histone H3 at lysine 4 (H3K4me) and acetylation of histone H3 at lysine 9 (H3K9ac5). In addition to promotor and intronic regions, transcription factor binding site (TFBS), transcription start site (TSS) and untranslated regions (UTR3 and UTR5). A p-value of 0.05/28 classes $< 1.79 \times 10^{-3}$ was considered statistically significant. Genetic correlation between our pulmonary function (FEV_1, FVC and FEV_1/FVC) results and publicly available GWAS of ever smoking[40] and height[41] was also investigated using LD Score regression[39].

Functional Annotation

To find functional elements in novel genome-wide significant signals, we annotated SNPs using various databases. We used Ensembl Variant Effect Predictor (VEP)[44] (Accessed 17 Jan 2017) and obtained mapped genes, transcripts, consequence of variants on protein sequence, Sorting Intolerant from Tolerant (SIFT)[45] scores, and Polymorphism Phenotyping v2 (PolyPhen-2)[46] scores. We checked if there were deleterious variants using Combined Annotation Dependent Depletion (CADD) v1.3[47] which integrates multiple annotations, compares each variant with possible substitutions across the human genome, ranks variants, and generates raw and scaled C-scores. A variant having a scaled C-score of 10 or 20 indicates that it is predicted to be in the top 10% or 1% deleterious changes in human genome, respectively. We used a cutoff of 15 to provide deleterious variants since it is the median for all possible splice site changes and non-synonymous variants (http://cadd.gs.washington.edu/info, Accessed 18 Jan 2017). To find potential regulatory variants, we used RegulomeDB[48] (Accessed 17 Jan 2017) which integrates DNA features and regulatory information including DNAase hypersensitivity, transcription factor binding sites, promoter regions, chromatin states, eQTLs, and methylation signals based on multiple high-throughput datasets and assign a category to each variant. Variants having RegulomeDB categories 1 or 2, meaning 'likely to affect binding and linked to expression of a gene target' or 'likely to affect binding,' were considered as regulatory variants.

Pathway Analysis using DEPICT and IPA

For gene prioritization and identification of enriched pathways and tissues/cell types, we used Data-driven Expression Prioritized Integration for Complex Traits (DEPICT)[49]

with association results for FEV_1, FVC, and FEV_1/FVC. We used association results from our European ancestry meta-analysis and the LD structure from 1000 Genomes European (CEU, GBR, and TSI) reference panel. The software excludes the major histocompatibility complex (MHC) region on chromosome 6 due to extended LD structure in the region. We ran a version of DEPICT for 1000 Genomes imputed meta-analysis results using its default parameters with an input file containing chromosomal location and p-values for variants having unadjusted p-values $<1x10^{-5}$. For gene set enrichment analyses, DEPICT utilizes 14,461 reconstituted gene sets generated by genes' co-regulation patterns in 77,840 gene expression microarray data. For tissue/cell type enrichment analysis, mapped genes were tested if they are highly expressed in 209 medical subject headings (MeSH) annotations using 37,427 microarray data. Gene prioritization analysis using co-functionality of genes can provide candidate causal genes in associated loci even if the loci are poorly studied or a gene is not the closest gene to a genome-wide significant variant. We chose FDR<0.05 as a cutoff for statistical significance in these enrichment analyses and gene prioritization results. Because DEPICT excludes the MHC, we also ran a pathway analysis with Ingenuity Pathway Analysis (IPA) (Ingenuity Systems, Redwood City, CA, USA, http://www.ingenuity.com/) on genes to which variants with $P<1x10^{-5}$ annotated.

PAINTOR

To identify causal variants in novel genome-wide significant loci, we used a trans-ethnic functional fine mapping method[50] implemented in PAINTOR (https://github.com/gkichaev/PAINTOR_FineMapping, Accessed on May 2, 2016). This method utilizes functional annotations along with association statistics (Z-scores) and LD information for each locus for each ancestry. We included our ancestry-specific meta-analysis results and used the African, American, European, and East Asian individuals from 1000 Genomes to calculate LD[72]. For PAINTOR we focused on 23 novel loci identified in our European ancestry and multiethnic fixed effects meta-analyses which had at least five genome-wide significant variants as well as all nine African or Hispanic loci which had at least one genome-wide significant variant. In addition, we included six loci which overlapped with the UK BiLEVE 1000 Genomes paper[14] and one locus with the CHARGE exome paper[43], since we ran PAINTOR prior to those publications. To reduce computational burden, we limited flanking regions to ±100 kilobase (kb) from the top single nucleotide polymorphisms (SNPs) and included variants with absolute value of Z-score greater than 1.96.

We used 269 publicly available annotations relevant to 'lung', 'bronch', or 'pulmo' from the following: hypersensitivity sites (DHSs)[73], super enhancers[74], Fantom5 enhancer and transcription start site regions[75], immune cell enhancers[76], and methylation and acetylation marks ENCODE[77]. We ran PAINTOR for each phenotype separately to

prioritize annotations based on likelihood-ratio statistics[78,79]. We included minimally correlated top annotations (less than five for each phenotype) to identify causal variants.

For the 40 loci from the fixed-effects meta-analysis, we used PAINTOR to construct credible sets of causal variants using a Bayesian meta-analysis framework. To obtain 95% credible sets for each locus, we ranked SNPs based on posterior probabilities of causality (high to low) and then took the SNPs filling in 95% of the summed posterior probability. We computed the median number of SNPs in the credible sets for ancestry-specific and multiethnic analyses of each trait.

FINEMAP

We used FINEMAP[51] to identify signals independent of lead variants for pulmonary function loci identified in the current or previous studies[14]. The Rotterdam Study (*N*=6,291), one of the larger cohort studies included in our meta-analysis, was used as a reference population. SNPs with MAF of <1% were excluded, leaving 118 SNPs for analysis. 10 SNPs for FEV_1 and FVC and 20 SNPs for FEV_1/FVC were further excluded because the LD matrix of the reference file from the Rotterdam Study did not represent the correlation matrix of the total study population. We allowed up to 10 causal SNPs per loci in FINEMAP analyses. To reduce the chance of false positive findings, we also conducted sensitivity analyses allowing up to 15 causal SNPs for loci with more than four SNPs with posterior probabilities of >0.8.

S-PrediXcan

S-PrediXcan is a summary statistics based approach for gene-based analysis[52] that was derived as an extension of the PrediXcan method for integration of GWAS and reference transcriptome data[80]. We used the S-PrediXcan approach to prioritize potentially causal genes, coupled with a Bayesian colocalization procedure[53] used to filter out LD-induced false positives. S-PrediXcan was used to analyze both European ancestry and multiethnic GWAS summary data for PFTs from the current study.

S-PrediXcan analysis was performed using the following publicly available tissue-specific expression models (http://predictdb.org) from the Genotype-Tissue Expression (GTEx) project v6p[28]: (1) GTEx Lung (278 samples) and (2) GTEx Whole blood (338 samples). Approximately 85% of participants in GTEx are white, 12% African American, and 3% of other races/ethnicities. Gene-based S-PrediXcan results were filtered on the following: (1) Proportion of SNPs used = (n SNPs available in GWAS summary data)/(n SNPs in prediction model) > 0.6, and (2) prediction performance R-squared > 0.01. Following application of S-PrediXcan to each of the GWAS summary data sets, we computed Bonferroni-corrected p-values derived as the nominal p-value for each gene-based test divided by the number of genes passing specified filters in each analysis to test whether genetically regulated gene expression was associated with the trait of

interest. The genome-wide S-PrediXcan results were then merged with novel loci from the current GWAS study by identifying all matches in which the novel locus SNP was within 500kb of the start of the gene.

We further incorporated a Bayesian colocalization approach[53] to interpret the extent to which S-PrediXcan results may have been influenced by LD within the region of interest. The Bayesian colocalization procedure was run using the following priors: p1 = 1e-4; prior probability SNP associated to trait 1, p2 = 1e-4; prior probability SNP associated to trait 2, p12 = 1e-5; prior probability SNP associated to both traits. The procedure generated posterior probabilities that correspond to one of the following hypotheses: a region is (H0) has no association with neither trait, (H1) associated with PFT phenotype but not gene expression, (H2) associated with gene expression but not PFT phenotype, (H3) associated with both traits, due to two independent SNPs, (H4) associated with both traits, due to one shared SNP.

Druggable Targets

We searched annotated gene lists against the ChEMBL database (v22.1, updated on November 15, 2016) to identify genes as targets of approved drugs or drugs in development. In addition, we used the Ingenuity Pathway Analysis (IPA, www.ingenuity.com, content of 2017-06-22) to identify drug targets and upstream regulators of the gene lists. We reported the upstream regulators in the following categories, biologic drug, chemical - endogenous mammalian, chemical - kinase inhibitor, chemical – other, chemical drug, chemical reagent, and chemical toxicant.

Data and Code Availability

The complete meta-analysis results have been deposited in the database of Genotypes and Phenotypes (dbGaP) under the CHARGE acquisition number phs000930 [https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000930.v6.p1]. GWAS data for most US studies are already available in dbGAP. For all other studies, please send requests to the study PI or Stephanie London (london2@niehs.nih.gov) who will forward them to the relevant party. Requests for METAL code can be directed to Stephanie London.

REFERENCES

1. Hole, D.J. *et al.* Impaired lung function and mortality risk in men and women: findings from the Renfrew and Paisley prospective population study. *BMJ* **313**, 711-5; discussion 715-6 (1996).
2. Young, R.P., Hopkins, R. & Eaton, T.E. Forced expiratory volume in one second: not just a lung function test but a marker of premature death from all causes. *Eur Respir J* **30**, 616-22 (2007).
3. Schunemann, H.J., Dorn, J., Grant, B.J., Winkelstein, W., Jr. & Trevisan, M. Pulmonary function is a long-term predictor of mortality in the general population: 29-year follow-up of the Buffalo Health Study. *Chest* **118**, 656-64 (2000).
4. Wilk, J.B. *et al.* Evidence for major genes influencing pulmonary function in the NHLBI family heart study. *Genet Epidemiol* **19**, 81-94 (2000).
5. Givelber, R.J. *et al.* Segregation analysis of pulmonary function among families in the Framingham Study. *Am J Respir Crit Care Med* **157**, 1445-51 (1998).
6. Chen, Y., Horne, S.L., Rennie, D.C. & Dosman, J.A. Segregation analysis of two lung function indices in a random sample of young families: the Humboldt Family Study. *Genet Epidemiol* **13**, 35-47 (1996).
7. Repapi, E. *et al.* Genome-wide association study identifies five loci associated with lung function. *Nat Genet* **42**, 36-44 (2010).
8. Hancock, D.B. *et al.* Meta-analyses of genome-wide association studies identify multiple loci associated with pulmonary function. *Nat Genet* **42**, 45-52 (2010).
9. Hancock, D.B. *et al.* Genome-wide joint meta-analysis of SNP and SNP-by-smoking interaction identifies novel loci for pulmonary function. *PLoS Genet* **8**, e1003098 (2012).
10. Loth, D.W. *et al.* Genome-wide association analysis identifies six new loci associated with forced vital capacity. *Nat Genet* **46**, 669-77 (2014).
11. Soler Artigas, M. *et al.* Genome-wide association and large-scale follow up identifies 16 new loci influencing lung function. *Nat Genet* **43**, 1082-90 (2011).
12. Tang, W. *et al.* Large-scale genome-wide association studies and meta-analyses of longitudinal change in adult lung function. *PLoS One* **9**, e100776 (2014).
13. Soler Artigas, M. *et al.* Sixteen new lung function signals identified through 1000 Genomes Project reference panel imputation. *Nat Commun* **6**, 8658 (2015).
14. Wain, L.V. *et al.* Genome-wide association analyses for lung function and chronic obstructive pulmonary disease identify new loci and potential druggable targets. *Nat Genet* **49**, 416-425 (2017).
15. Wain, L.V. *et al.* Novel insights into the genetics of smoking behaviour, lung function, and chronic obstructive pulmonary disease (UK BiLEVE): a genetic association study in UK Biobank. *Lancet Respir Med* **3**, 769-81 (2015).
16. Mahajan, A. *et al.* Genome-wide trans-ancestry meta-analysis provides insight into the genetic architecture of type 2 diabetes susceptibility. *Nat Genet* **46**, 234-44 (2014).
17. Pe'er, I., Yelensky, R., Altshuler, D. & Daly, M.J. Estimation of the multiple testing burden for genomewide association studies of nearly all common variants. *Genet Epidemiol* **32**, 381-5 (2008).
18. Burkart, K.M. *et al.* A Genome-wide Association Study in Hispanics/Latinos Identifies Novel Signals for Lung Function. The Hispanic Community Health Study/Study of Latinos. *American Journal of Respiratory and Critical Care Medicine* doi: 10.1164/rccm.201707-1493OC (2018).
19. Han, B. & Eskin, E. Random-effects model aimed at discovering associations in meta-analysis of genome-wide association studies. *Am J Hum Genet* **88**, 586-98 (2011).

20. Sofer, T. *et al.* A powerful statistical framework for generalization testing in GWAS, with application to the HCHS/SOL. *Genet Epidemiol* **41**, 251-258 (2017).
21. Cho, M.H. *et al.* Risk loci for chronic obstructive pulmonary disease: a genome-wide association study and meta-analysis. *Lancet Respir Med* **2**, 214-25 (2014).
22. Regan, E.A. *et al.* Genetic epidemiology of COPD (COPDGene) study design. *COPD* **7**, 32-43 (2010).
23. Levin, A.M. *et al.* Nocturnal asthma and the importance of race/ethnicity and genetic ancestry. *Am J Respir Crit Care Med* **190**, 266-73 (2014).
24. Padhukasahasram, B. *et al.* Gene-based association identifies SPATA13-AS1 as a pharmacogenomic predictor of inhaled short-acting beta-agonist response in multiple population groups. *Pharmacogenomics J* **14**, 365-71 (2014).
25. Borrell, L.N. *et al.* Childhood obesity and asthma control in the GALA II and SAGE II studies. *Am J Respir Crit Care Med* **187**, 697-702 (2013).
26. Drake, K.A. *et al.* A genome-wide association study of bronchodilator response in Latinos implicates rare variants. *J Allergy Clin Immunol* **133**, 370-8 (2014).
27. Hobbs, B.D. *et al.* Genetic loci associated with chronic obstructive pulmonary disease overlap with loci for lung function and pulmonary fibrosis. *Nat Genet* **49**, 426-432 (2017).
28. The Genotype-Tissue Expression (GTEx) project. *Nat Genet* **45**, 580-5 (2013).
29. Obeidat, M. *et al.* Molecular mechanisms underlying variations in lung function: a systems genetics analysis. *Lancet Respir Med* **3**, 782-95 (2015).
30. Obeidat, M. *et al.* GSTCD and INTS12 regulation and expression in the human lung. *PLoS One* **8**, e74630 (2013).
31. Hao, K. *et al.* Lung eQTLs to help reveal the molecular underpinnings of asthma. *PLoS Genet* **8**, e1003029 (2012).
32. Joehanes, R. *et al.* Integrated genome-wide analysis of expression quantitative trait loci aids interpretation of genomic association studies. *Genome Biol* **18**, 16 (2017).
33. Westra, H.J. *et al.* Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nat Genet* **45**, 1238-43 (2013).
34. Bonder, M.J. *et al.* Disease variants alter transcription factor levels and methylation of their binding sites. *Nat Genet* **49**, 131-138 (2017).
35. Zhernakova, D.V. *et al.* Identification of context-dependent expression quantitative trait loci in whole blood. *Nat Genet* **49**, 139-145 (2017).
36. Mogil, L.S. *et al.* Genetic architecture of gene expression traits across diverse populations. *bioRxiv* (2018).
37. Bulik-Sullivan, B.K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet* **47**, 291-5 (2015).
38. Finucane, H.K. *et al.* Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat Genet* **47**, 1228-35 (2015).
39. Bulik-Sullivan, B. *et al.* An atlas of genetic correlations across human diseases and traits. *Nat Genet* **47**, 1236-41 (2015).
40. Genome-wide meta-analyses identify multiple loci associated with smoking behavior. *Nat Genet* **42**, 441-7 (2010).
41. Wood, A.R. *et al.* Defining the role of common variation in the genomic and biological architecture of adult human height. *Nat Genet* **46**, 1173-86 (2014).
42. Moffatt, M.F. *et al.* A large-scale, consortium-based genomewide association study of asthma. *N Engl J Med* **363**, 1211-1221 (2010).

43. Jackson, V.E. *et al.* Meta-analysis of exome array data identifies six novel genetic loci for lung function. Wellcome Open Research. Preprint [version 1; referees: 1 approved, 1 approved with reservations] at doi: 10.12688/wellcomeopenres.12583.1 (2018).
44. McLaren, W. *et al.* The Ensembl Variant Effect Predictor. *Genome Biol* **17**, 122 (2016).
45. Kumar, P., Henikoff, S. & Ng, P.C. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nature Protocols* **4**, 1073-1082 (2009).
46. Adzhubei, I.A. *et al.* A method and server for predicting damaging missense mutations. *Nature Methods* **7**, 248-249 (2010).
47. Kircher, M. *et al.* A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet* **46**, 310-5 (2014).
48. Boyle, A.P. *et al.* Annotation of functional variation in personal genomes using RegulomeDB. *Genome Research* **22**, 1790-1797 (2012).
49. Pers, T.H. *et al.* Biological interpretation of genome-wide association studies using predicted gene functions. *Nat Commun* **6**, 5890 (2015).
50. Kichaev, G. & Pasaniuc, B. Leveraging Functional-Annotation Data in Trans-ethnic Fine-Mapping Studies. *Am J Hum Genet* **97**, 260-71 (2015).
51. Benner, C. *et al.* FINEMAP: efficient variable selection using summary data from genome-wide association studies. *Bioinformatics* **32**, 1493-501 (2016).
52. Barbeira, A. *et al.* Exploring the phenotypic consequences of tissue specific gene expression variation inferred from GWAS summary statistics. *Nat Commun* 9, 1825 (2018).
53. Giambartolomei, C. *et al.* Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet* **10**, e1004383 (2014).
54. Gharib, S.A. *et al.* Integrative pathway genomics of lung function and airflow obstruction. *Hum Mol Genet* **24**, 6836-48 (2015).
55. Barreto-Luis, A. *et al.* A pathway-based association study reveals variants from Wnt signalling genes contributing to asthma susceptibility. *Clin Exp Allergy* **47**, 618-626 (2017).
56. Liu, J. *et al.* Targeting Wnt-driven cancer through the inhibition of Porcupine by LGK974. *Proc Natl Acad Sci U S A* **110**, 20224-9 (2013).
57. Arredouani, M. *et al.* The scavenger receptor MARCO is required for lung defense against pneumococcal pneumonia and inhaled particles. *J Exp Med* **200**, 267-72 (2004).
58. Dorrington, M.G. *et al.* MARCO is required for TLR2- and Nod2-mediated responses to Streptococcus pneumoniae and clearance of pneumococcal colonization in the murine nasopharynx. *J Immunol* **190**, 250-8 (2013).
59. Torgerson, D.G. *et al.* Meta-analysis of genome-wide association studies of asthma in ethnically diverse North American populations. *Nat Genet* **43**, 887-92 (2011).
60. Bahleda, R. *et al.* Phase I dose-escalation studies of roniciclib, a pan-cyclin-dependent kinase inhibitor, in advanced malignancies. *Br J Cancer* **116**, 1505-1512 (2017).
61. Yu, Q.B., Li, H.M., Li, L.L., Wang, S.Y. & Wu, Y.B. Sevoflurane downregulates interleukin-6 and interleukin-8 levels in patients after cardiopulmonary bypass surgery: a meta-analysis. *Genet Mol Res* **14**, 19016-27 (2015).
62. Cho, E.J., Yoon, J.H., Hong, S.J., Lee, S.H. & Sim, S.B. The effects of sevoflurane on systemic and pulmonary inflammatory responses after cardiopulmonary bypass. *J Cardiothorac Vasc Anesth* **23**, 639-45 (2009).
63. Uhlig, C. *et al.* Effects of Volatile Anesthetics on Mortality and Postoperative Pulmonary and Other Complications in Patients Undergoing Surgery: A Systematic Review and Meta-analysis. *Anesthesiology* **124**, 1230-45 (2016).

64. Jabaudon, M. *et al.* Sevoflurane for Sedation in Acute Respiratory Distress Syndrome. A Randomized Controlled Pilot Study. *Am J Respir Crit Care Med* **195**, 792-800 (2017).
65. Hirshman, C.A. & Emala, C.W. Actin reorganization in airway smooth muscle cells involves Gq and Gi-2 activation of Rho. *Am J Physiol* **277**, L653-61 (1999).
66. Billington, C.K. & Penn, R.B. Signaling and regulation of G protein-coupled receptors in airway smooth muscle. *Respir Res* **4**, 2 (2003).
67. Lange, P. *et al.* Lung-Function Trajectories Leading to Chronic Obstructive Pulmonary Disease. *N Engl J Med* **373**, 111-22 (2015).
68. Fletcher, C. & Peto, R. The natural history of chronic airflow obstruction. *Br Med J* **1**, 1645-8 (1977).
69. Martinez, F.D. Early-Life Origins of Chronic Obstructive Pulmonary Disease. *N Engl J Med* **375**, 871-8 (2016).
70. Hankinson, J.L., Odencrantz, J.R. & Fedan, K.B. Spirometric reference values from a sample of the general U.S. population. *Am J Respir Crit Care Med* **159**, 179-87 (1999).
71. Yang, J. *et al.* Genomic inflation factors under polygenic inheritance. *Eur J Hum Genet* **19**, 807-12 (2011).
72. Genomes Project, C. *et al.* An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56-65 (2012).
73. Maurano, M.T. *et al.* Systematic localization of common disease-associated variation in regulatory DNA. *Science* **337**, 1190-5 (2012).
74. Hnisz, D. *et al.* Super-enhancers in the control of cell identity and disease. *Cell* **155**, 934-47 (2013).
75. Forrest, A.R.R. *et al.* A promoter-level mammalian expression atlas. *Nature* **507**, 462-70 (2014).
76. Farh, K.K.H. *et al.* Genetic and epigenetic fine mapping of causal autoimmune disease variants. *Nature* **518**, 337-343 (2015).
77. Dunham, I. *et al.* An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57-74 (2012).
78. Kichaev, G. *et al.* Integrating Functional Data to Prioritize Causal Variants in Statistical Fine-Mapping Studies. *Plos Genetics* **10** (2014).
79. Pickrell, J.K. Joint analysis of functional genomic data and genome-wide association studies of 18 human traits. *Am J Hum Genet* **94**, 559-73 (2014).
80. Gamazon, E.R. *et al.* A gene-based association method for mapping traits using reference transcriptome data. *Nat Genet* **47**, 1091-8 (2015).

3.2

Heritability and genome-wide association study of diffusing capacity of the lung

ABSTRACT

Background: Although several genome wide association studies (GWAS) have investigated the genetics of pulmonary ventilatory function, little is known about the genetic factors that influence gas exchange.

Aim: To investigate the heritability of, and genetic variants associated with the diffusing capacity of the lung.

Methods: GWAS was performed on diffusing capacity, measured by carbon monoxide uptake (DLCO) and per alveolar volume (DLCO/VA) using the single-breath technique, in 8,372 individuals from two population-based cohort studies, the Rotterdam Study and the Framingham Heart Study. Heritability was estimated in related (n=6,246) and unrelated (n=3,286) individuals.

Results: Heritability of DLCO and DLCO/VA ranged between 23% and 28% in unrelated individuals and between 45% and 49% in related individuals. Meta-analysis identified a genetic variant in *ADGRG6* that is significantly associated with DLCO/VA. Gene expression analysis of *ADGRG6* in human lung tissue revealed a decreased expression in patients with COPD and subjects with decreased DLCO/VA.

Conclusion: DLCO and DLCO/VA are heritable traits, with a considerable proportion of variance explained by genetics. A functional variant in *ADGRG6* gene region was significantly associated with DLCO/VA. Pulmonary *ADGRG6* expression was decreased in patients with COPD.

Key words: DLCO, DLCO/VA, Population based, eQTL, gene expression, human lung tissue.

INTRODUCTION

The respiratory system can be separated functionally into two zones. The first one is the conducting zone, which includes the trachea, bronchi, bronchioles, and terminal bronchioles and which is functional in ventilation, i.e. conducting the air in and out of the lungs. The second zone is the respiratory zone, which consists of the respiratory bronchioles, alveolar ducts and alveoli, the site where oxygen and carbon-dioxide are exchanged between the lungs and the blood.

Different pulmonary function tests are available that measure these distinct functions of ventilation and gas exchange. These tests help to evaluate and manage patients with respiratory symptoms and diseases, and include spirometry, measurements of lung volumes, and the diffusing capacity of the lung for carbon monoxide (DLCO). The latter, also known as transfer factor of the lung for CO, provides a quantitative measure of gas transfer in the lung (1, 2) and reflects processes in the alveolar compartment and pulmonary microcirculation.

The DLCO provides clinical insights complimentary to those obtained by spirometry and lung volume measurements, for example, in discriminating asthma from chronic obstructive pulmonary disease (COPD), to identify causes of hypoxemia or dyspnoea, and to monitor patients with interstitial lung disease (3). DLCO is decreased in patients with emphysema due to a decrease in the total surface area of the lung and the loss of capillary beds (1, 4). In contrast to the abundance of genome-wide association studies (GWAS) investigating genetic variation of spirometry measures (5-8), the heritability of, and genetic influences on DLCO, are largely unknown.

Therefore, we first investigated the heritability of DLCO to understand which proportion of the variance in DLCO can be explained by genetics. Next, we performed a GWAS, to identify genetic variants affecting the variability in DLCO, using data from two prospective population-based cohorts; the Rotterdam Study and the Framingham Heart Study. Finally, we investigated the expression of the lead GWAS association in lung tissue of individuals with COPD and (non-smoking and smoking) controls.

METHODS

In this section the methods will be described briefly. Please see **Supplemental Methods** in the Online Data Supplement for more detailed information.

Setting

The present meta-analysis combined results from two population-based studies, i.e. the Rotterdam Study and the Framingham Heart Study. In both cohorts, only individuals of

European ancestry were included in the analyses. The Rotterdam Study (9) is an ongoing prospective population-based cohort study that includes 3 cohorts encompassing 14,926 participants aged ≥ 45 years, living in the Netherlands. DLCO was measured between 2009-2013.

The Framingham heart study is a population-based family study that recruited residents of Framingham, Massachusetts starting from 1948. DLCO was measured at the 8th and 9th examinations of the Offspring Cohort (2005-2008 and 2011-2014) and the 1st and 2nd examinations of the Third Generation Cohort (2002-2005 and 2008-2011). For participants with measurements at both time points, we analyzed the later measurement.

Lung function

DLCO (mmol/min/kPA) and alveolar volume (VA) were measured by the single breath technique in accordance with ERS / ATS guidelines (2). The DLCO per alveolar volume (DLCO/VA; mmol/min/kPA/liter) was calculated by dividing the DLCO by VA. Analyses were restricted to participants with two interpretable and reproducible measurements of DLCO and DLCO/VA.

Heritability analysis

Heritability was defined as the ratio of trait variance due to additive genetic effects to the total phenotypic variance after accounting for covariates. In the Rotterdam Study GCTA software (10) was used to estimated heritability in unrelated individuals. In the Framingham Heart Study SOLAR software (11) was used to estimate heritability based on familial relationships. Analyses were adjusted for age, sex and principal components of genetic relatedness ((PC), in GCTA only). Additional adjustment for current and former smoking were done in a subsequent analysis.

GWAS analyses

A GWAS was performed for both phenotypes DLCO and DLCO/VA using ProbABEL (version 0.4.4). Variants with imputation quality (R^2) < 0.3 and minor allele frequency (MAF) < 0.01 were excluded from the analyses. Linear regression was conducted for each SNP, assuming additive model. All analyses were adjusted for age, sex and PC (Rotterdam Study only) in model 1 and additionally adjusted for smoking, weight and height in model 2. A random effect was added to the model to account for familial relationship in the Framingham Heart Study analyses. Data were meta-analyzed using METAL software and were adjusted for genomic control. Genome-wide significance threshold was set at P-value < 5 $x10^{-8}$ and for suggestive associations at P-value 5 x 10^{-7}. Quantile-quantile plots, Manhattan plots and regional plots were generated using the R software. Analyses were repeated after 1) correction for haemoglobin in the

Rotterdam Study and 2) additional adjustment for FEV_1/FVC, 3) additional adjustment for quantitative emphysema (lower than -950 LAA on CT scan of the lungs) as measured in the Framingham Heart Study within 8 years from the lung function measurement.

Follow-up analyses

Several steps were taken in order to explore the functionality of the variants and genes of interest, and to associate those newly identified loci to clinically relevant disease outcomes. 1) Genetic correlations were investigated, 2) Genetic overlap was investigated with SNPs that are significantly related to COPD (12) and emphysema (13). 3) Posterior probability of causality of the lead SNP was calculated using FINEMAP software (14) . 4) The regulatory function of the lead SNP was explored on the Haploreg server. 5) The effect of the lead SNP on mRNA expression was checked (Expression Quantitative Trait Loci (eQTL) analysis), using lung tissue dataset from Genotype Tissue Expression (GTEx) portal (see **URLs**). 6) Tissue-specific gene expression was checked in GTEx portal and 7) Finally, mRNA expression of the *ADGRG6* gene was analysed in lung tissue (using real-time PCR) of 92 patients with or without COPD.

RESULTS

The study cohorts and participant characteristics

The general characteristics of the study populations (the Rotterdam Study and the Framingham Heart Study) are shown in **Table 1**. The mean age was 67.3 (SD 8.0 yr) years in the Rotterdam Study and 52.8 (SD 14.8 yr) years in the Framingham Heart Study. **Figure 1** shows the study flow of participants that were included in this study.

Heritability

Heritability was estimated in two ways, the first one was by using the Rotterdam Study data with unrelated individuals, with a total number of 3,286 participants with genetic data and interpretable measurements of DLCO. The second was by using data from the Framingham Heart Study to estimate heritability based on familial relationships in 6,246 participants with interpretable measurements of DLCO (**Figure 1**). In **Table 2** heritability estimates for DLCO and DLCO/VA are presented. In unrelated individuals, we found an age- and sex- and PC-adjusted heritability for DLCO of 23%, and a heritability of 28% after additional adjustment for current and past smoking. Similar heritability estimates were found for DLCO/VA with 24% after adjustment for age, sex and PC, and 25% after additional adjustment for smoking. In the Framingham Heart Study, investigating individuals with known familial relationships, we found an age- and sex-adjusted heritability for DLCO of 49%, and a heritability of 47% after

additional adjustment for current and past smoking. Heritability estimates for DLCO/VA were 45% after adjustment for age, sex, and 46% after additional adjustment for current and past smoking.

Table 1: General characteristics of the study populations

	RS	**FHS**
	N= 2,574	**N= 5,798**
Age, years	67.3 (8.0)	52.7 (14.8)
Female, %	51.9	53.9
Weight, Kg	80.5 (14.9)	79.7 (18.5)
Height, cm	170.6 (9.2)	168.9 (9.5)
Former smokers, %	55.4	39.9
Current smokers, %	11.5	10.9
Never smokers, %	33.1	49.2
DLCO (mmol/min/kPA)	8.0 (1.8)	8.3 (2.3)
DLCO corrected for Hb (mmol/min/kPA)	7.9 (1.7)	NA
DLCO/VA (mmol/min/kPA/VA)	1.5 (0.2)	1.5 (0.2)
DLCO/VA corrected for Hb (mmol/min/kPA/VA)	1.5 (0.2)	NA
FEV_1 (L)	2.8 (0.7)	3.1 (0.9)
FEV_1, % predicted	105.3 (19.8)	99.2 (14.9)
FVC (L)	3.7 (1.0)	4.2 (1.1)
FVC, % predicted	110.1 (17.6)	102.7 (13.5)
FEV_1/FVC, %	76.4 (7.1)	75.4 (6.9)

DLCO: Diffusing capacity of the lung for carbon monoxide; DLCO/VA: Diffusing capacity of the lung for carbon monoxide by alveolar volume; FEV1: Forced expiratory volume during the first second; FHS: Framingham Heart Study; FVC: Forced vital capacity; Hb: Haemoglobin; RS: Rotterdam Study

Values are means (standard deviation) for continuous variables or percentages for dichotomous variables.

Table 2: Heritability of diffusing capacity of the lung

Model*	**Rotterdam Study (RS)** *N=3,286 unrelated individuals*				**Framingham Heart Study (FHS)** *N=6,246 with known familial relationships*			
	DLCO		**DLCO/VA**		**DLCO**		**DLCO/VA**	
	h^2 (SE)	**P-value**	**h^2 (SE)**	**P-value**	**h^2 (SE)**	**P-value**	**h^2 (SE)**	**P-value**
1	0.23 (0.10)	0.01	0.24 (0.10)	0.009	0.49 (0.03)	2.3×10^{-106}	0.45 (0.03)	5.0×10^{-82}
2	0.28 (0.10)	0.002	0.25 (0.10)	0.0075	0.47 (0.03)	8.5×10^{-100}	0.46 (0.03)	7.6×10^{-84}

DLCO: Diffusing capacity of the lung for carbon monoxide; DLCO/VA: Diffusing capacity of the lung for carbon monoxide by alveolar volume; FHS: The Framingham Heart Study; h^2: heritability estimate; RS: The Rotterdam Study; SE: Standard error.

* Model 1: adjusted for age, sex and principal components of genetic relatedness (RS only); Model 2: adjusted for age, sex, smoking and principal components of genetic relatedness (RS only).

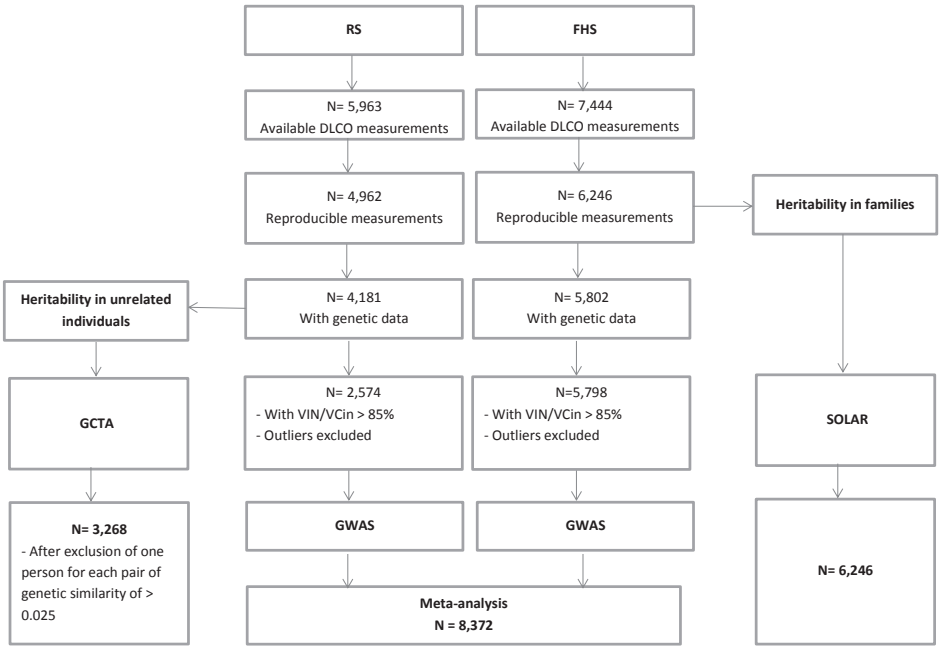

Figure 1 Flowchart of study participants.
DLCO: Diffusing capacity of carbon monoxide; FHS; the Framingham Heart Study; GWAS; genome wide association study; GCTA: Genome-wide Complex Trait Analysis software; N= number of participants; RS: the Rotterdam Study; SOLAR: Sequential Oligogenic Linkage Analysis Routines package; VA: Alveolar volume; VCin: vital capacity measured during maximal inspiration; VIN: inspiratory volume.

Genetic variants associated with diffusing capacity

We performed GWAS on DLCO and DLCO/VA in the Rotterdam Study (n=2,574) and the Framingham Heart Study (n= 5,798), and subsequently meta-analysed both cohorts (n= 8,372). All variants with a p-value below 5×10^{-6} at the meta-analysis stage are presented in **Table 3**. The corresponding quantile-quantile plots are presented in **Figure E1** in the Online Data Supplement**.** GWAS results of the separate cohorts with (P-value $< 5 \times 10^{-6}$), are presented in **Tables E1 and E2** in the Online Data Supplement.

Analyses were adjusted for age, sex and PC in model 1. In model 2 analyses were adjusted for variables in model 1, in addition to weight, height, current and past smoking.

Figure 2 represents the Manhattan-plots of DLCO GWAS at the meta-analysis level. For both DLCO analyses (models 1 and 2), no variant reached genome wide significance threshold. In model 2, two variants at 10q22.1 (rs1665630, gene: *CDH23*, MAF: 0.44, P-value= 2.8×10^{-7}) and at 20p12.3 (rs2423124, close to gene: *GPCPD1*, MAF:0.19, P-value=4.2×10^{-7}) showed a suggestive association with DLCO.

Table 3: Independent genetic variants that are significantly or suggestively associated with DLCO or DLCO/VA at meta-analysis level.

Trait	SNP	Chr:Pos	Gene ‡	A1/A2	RS (n=2,574)		FHS (n=5,798)		RS and FHS (n=8,372)	
					B	P-value	B	P-value	B	P-value
DLCO*	-	-	-	-	-	-	-	-	-	-
	rs1665630	10:73426862	CDH23	T/C	.11	$6.4x10^{-4}$	.10	$9.1x10^{-5}$	0.11	$2.8x10^{-7}$
DLCO†	rs2423124	20:5636945	GPCPD1	T/C	-.20	$1.4x10^{-6}$	-.10	$2.5x10^{-2}$	-.16	$4.2x10^{-7}$
	rs17280293	6:142688969	ADGRG6	A/G	-.06	$3.0x10^{-3}$	-.08	**$6.7x10^{-9}$**	-.07	**$1.4x10^{-10}$**
	rs918606	5:61926379	IPO11	A/G	-.02	$2.0x10^{-3}$	-.02	$5.8x10^{-6}$	-.02	$6.0.x10^{-8}$
DLCO/VA*	rs75834976	4:5231710	STK32B	A/C	-.04	$2.4x10^{-3}$	-.04	$5.5x10^{-5}$	-.04	$6.0x10^{-7}$
	rs56315120	1:165168869	LMX1A	A/G	-.02	0.24	-.06	$1.5x10^{-7}$	-.05	$7.8x10^{-7}$
	rs17280293	6:142688969	ADGRG6	A/G	-.06	$4.3x10^{-3}$	-.07	**$2.3x10^{-9}$**	-.07	**$7.9x10^{-11}$**
DLCO/VA†	rs918606	5:61926379	IPO11	A/G	-.02	$1.2x10^{-3}$	-.02	$1.3x10^{-5}$	-.02	$7.5x10^{-8}$

A1: The first allele; A2: the second allele; B: the effect estimate which are additive effects for each copy of A1; Chr:Pos: Chromosome and position; DLCO: Diffusing capacity of carbon monoxide; DLCO/VA: Diffusing capacity of carbon monoxide by alveolar volume; FHS: the Framingham Heart Study; RS: the Rotterdam Study (Meta-analysis RSI, RSII and RSIII); SNP: single nucleotide polymorphism

*Model 1: Adjusted for age, sex and principal components

†Model 2: Adjusted for age, sex, weight, height, smoking and principal components

‡The Gene name is a label of the region using the closest gene but does not necessarily pinpoint the responsible gene

Bold indicates statistically significant.

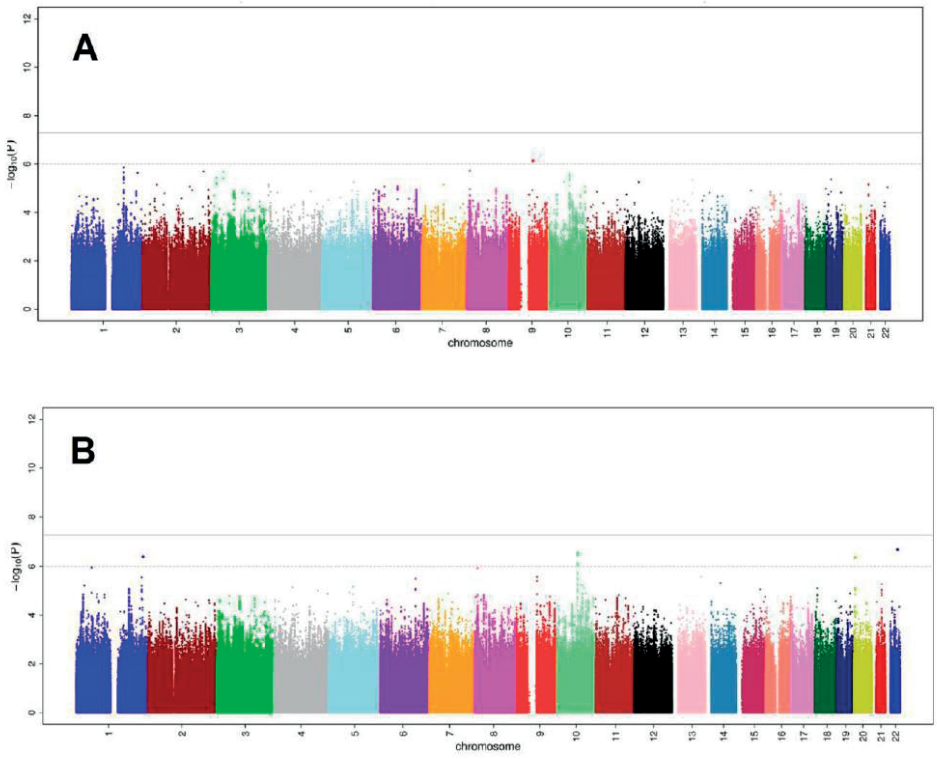

Figure 2 Common genetic variants associated with diffusing capacity of the lung for carbon monoxide (DLCO).
A: Manhattan-plot of the association between common genetic variants and DLCO, adjusted for age, sex and principal components of genetic relatedness.
B: Manhattan-plot of the association between common genetics variants and DLCO, adjusted for age, sex, weight, height, smoking and principal components of genetic relatedness.

Figure 3 represents the Manhattan-plots of DLCO/VA GWAS at the meta-analysis level. Nineteen variants at the same locus at 6q24.1 (top: rs17280293, gene: *ADGRG6*; MAF: 0.03, P-value= 1.4 x 10^{-10}), were significantly associated with DLCO/VA in model 1(see regional plot in **Figure 4).** Of these, six variants at the same locus at 6q24.1, reached the genome-wide significance threshold in model 2. Sensitivity analysis by adjusting for FEV_1/FVC did not explain the effect of the association between rs17280293 and DLCO/VA (beta=-0.07 (SE: 0.01), P-value= 1.51 x 10-10 after adjustment for FEV_1/FVC) versus (beta=-0.07 (SE 0.01), P-value= 7.9 x 10-11 before adjustment for FEV_1/FVC in model 2) (see Figure E2 in Online Data Supplement). Similarly, adjusting for quantitative emphysema (<-950 LAA on CT scan of the lungs) in a subset of the Framingham Heart Study (n=2,176) did not alter the association between rs17280293 and DLCO/VA (beta= -0.06 (SE 0.02), P-value= 0.003 after adjustment for emphysema) versus (beta= -0.06 (SE

0.02), P-value= 0.002 before adjustment for emphysema). Moreover, in both models, a variant at 5q12.1 (rs918606, gene: *IPO11*; MAF: 0.44; P-value model 1= 5.96 x 10^{-8}, P-value-model 2= 7.49 x 10^{-8}) was found to be suggestively associated with DLCO/VA. Additional sensitivity analysis by adjusting for haemoglobin blood concentrations did not materially change the results of the DLCO/VA GWAS (see **Supplemental Results** in the Online Data Supplement).

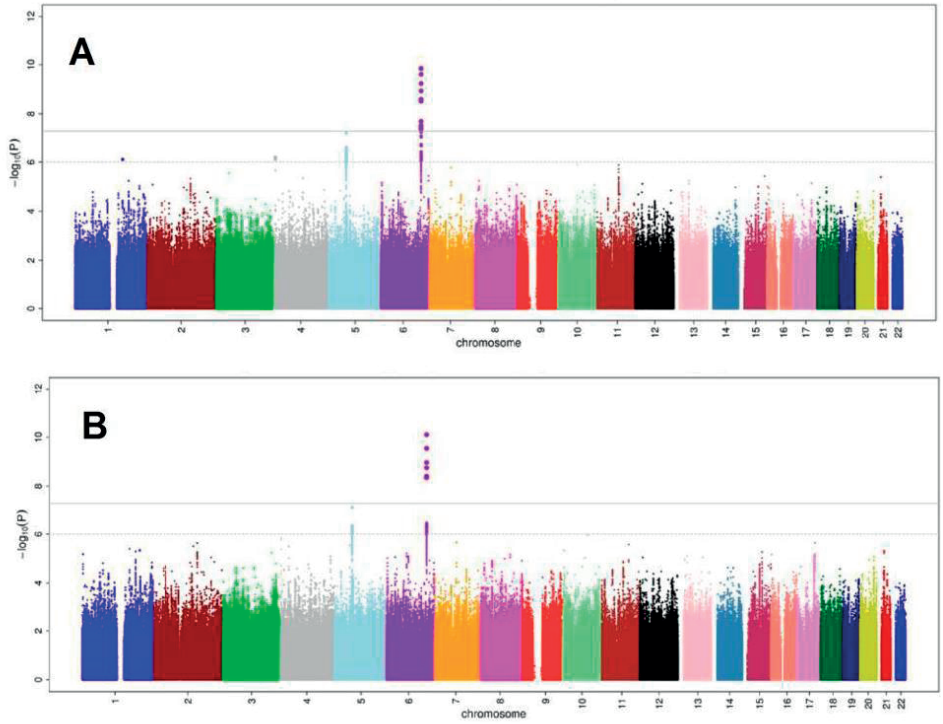

Figure 3 Common genetic variants associated with diffusing capacity of the lung per alveolar volume (DLCO/VA).
A: Manhattan-plot of the association between common genetic variants and DLCO/VA, adjusted for age, sex and principal components of genetic relatedness.
B: Manhattan-plot of the association between common genetics variants and DLCO/VA, adjusted for age, sex, weight, height, smoking and principal components of genetic relatedness.

Interestingly, a more in depth investigation of the *ADGRG6* region (**Figure 4**) revealed the presence of two missense variants; the lead SNP rs17280293 and rs11155242 (MAF 0.19, P-value=2.1 x 10^{-06}). Those two SNPs showed to be in LD with each other, with r^2=0.14 and D'=1.

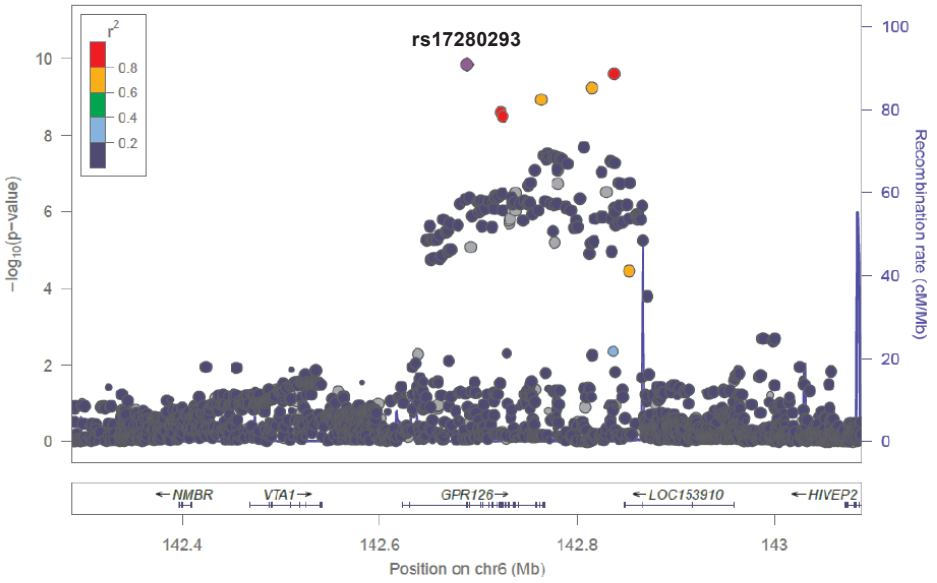

Figure 4 Regional association plot of the genome-wide significant locus in DLCO/VA GWAS.

Follow-up analyses

In this section, the most important findings of the follow-up analyses will be summarized including genetic correlations and gene expression in lung tissue. Additional results on the genetic correlations, overlap with reported COPD and emphysema GWAS associations, posterior probability of causality, functional annotation and gene expression will be presented in the **Supplemental Results and Figures** in the Online Data Supplement.

Genetic correlations

We examined the genetic correlation between DLCO/VA and DLCO using the age, sex, smoking status, weight, height and PC adjusted model. The genetic correlation was 59% (ρgenetic=0.59, P-value=0.04). This was in line with the phenotypic correlation between DLCO and DLCO/VA (r^2=0.46 in the Rotterdam Study and r^2=0.57 in the Framingham Heart Study, P-value < 0.01). We also examined the genetic correlation with FEV_1/FVC and height (see **Supplemental Methods** and **Results** in Online Data Supplement).

ADGRG6 expression

We extracted mRNA from lung resection specimens of 92 patients who underwent surgery for solitary pulmonary tumours or lung transplantation, including 44 patients without COPD and 48 patients with COPD (**Table 4**). The mRNA expression of

ADGRG6 was significantly lower in lung tissue of patients with decreased DLCO/VA compared with patients with normal DLCO/VA (**Figure 5A**) and in subjects with COPD (encompassing different categories of COPD severity according to the GOLD spirometric classification) compared to never smoking controls (**Figure 5B)**. The *ADGRG6* mRNA levels were significantly associated with DLCO/VA after adjustment for age and sex in model 1 (n=67 β=0.85 (95% CI 0.06-1.64)) and after additional adjustment for weight, height and smoking in model 2 n=66 (β=0.75 (95% CI 0.03-1.47)).

Table 4: Characteristics of study individuals for lung mRNA analysis (by RT-PCR) (n=92)

	Never smokers	Smokers without COPD	COPD GOLD II	COPD GOLD III-IV
Number	18	26	34	14
Gender ratio (m/f)	6/12 §	19/7 §	31/3 §	8/6 §
Age (years)	65 (56-70)	63 (55-70)	66 (58-69) ‡	56 (54-60)*† ‡
Current- / ex-smoker	NA	16/10	22/12	0/14
Smoking history (PY)	NA	28 (15-45)*	45 (40-60)* ‡	30 (25-30)*† ‡
FEV_1 post BD (L)	2.7 (2.3-3.2)	2.7 (2.3-3.3)	2.0 (1.8-2.4)* ‡	0.7 (0.7-0.9)*† ‡
FEV_1 post BD (% predicted)	102 (92-116)	95 (93-112)	68 (61-75)* ‡	26 (20-32)*† ‡
FEV_1 / FVC post BD (%)	78 (75-83)	75 (71-79)*	56 (53-60)* ‡	32 (27-35)*† ‡
DLCO (% predicted)	90 (80-105)	80 (61-102)	67 (51-87)*	35 (33-41)*† ‡
DLCO/VA (% predicted)	103 (88-123)	91 (68-107)*	87 (62-108)*	59 (50-65)*† ‡
DLCO (mmol/kPA/min)	21.6 (18.1-26.8)	23.3 (17.0-27.4)	17.2 (14.2-25.0)	2.9 (2.8-3.7) ^°#
DLCO/VA (mmol/kPA/min/VA)	4.6 (3.8-5.3)	3.9 (2.9-4.6)*	3.5 (2.7-4.2)*	0.9 (0.7-0.9) ^°#

DLCO: diffusing capacity of the lung for carbon monoxide; DLCO/VA: diffusing capacity of the lung for carbon monoxide per alveolar volume; f: female; FEV1: forced expiratory volume in 1 second; FVC: forced vital capacity; m: male; NA: not applicable; post-BD: post-bronchodilator; PY: pack years

Data are presented as median (IQR)

Mann-Whitney U test: * $P < 0.05$ versus never smokers; ^ $P < 0.001$ versus never smokers; † $P < 0.05$ versus COPD GOLD II; # $P < 0.001$ versus COPD GOLD II; ‡ $P < 0.05$ versus smokers without COPD; ° $P < 0.001$ versus smokers without COPD; Fisher's exact test: § $P < 0.001$.

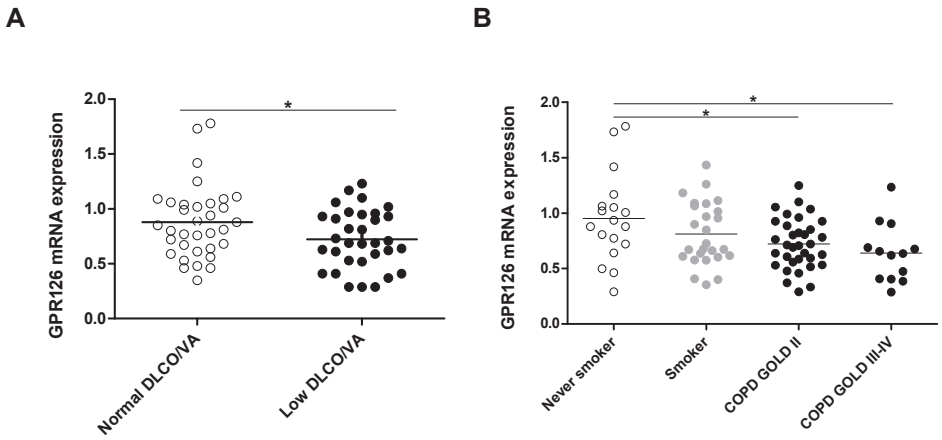

Figure 5 Pulmonary mRNA expression of *ADGRG6* in human subjects
A) mRNA levels of *ADGRG6* in lung tissue of individuals with normal DLCO/VA (n=38) and low DLCO/VA (n=39). mRNA levels were corrected using a calculated normalization factor based on mRNA expression of three reference genes (GAPDH, SDHA, HPRT-1).
B) mRNA levels of *ADGRG6* in lung tissue of never smokers (N=18), smokers without airflow limitation (N=26), patients with COPD GOLD II (N=34) and patients with COPD GOLD III-IV (N=14), as measured by quantitative RT-PCR.
For statistical analysis, Kruskal-Wallis followed by Mann-Whitney U test was used for COPD (*P<0.05 and **P<0.01) and an independent sample t-test was used for DLCO/VA after rank transformation (* p<0.05).

DISCUSSION

This is the first study that has investigated the heritability of, and genome-wide association with, diffusing capacity of the lung using population-based cohort studies. We found a considerable proportion of variance in diffusing capacity of the lung explained by genetics. We also identified one locus on chromosome 6, encompassing the *ADGRG6* gene, that is associated with DLCO/VA and its lead variant showed to have a high posterior probability of causality compared to other SNPs in the same region. Finally, we were able to link the pulmonary expression of *ADGRG6* directly to COPD and to low DLCO/VA (compatible with emphysema in this general population). Here, we demonstrated a differential mRNA expression of *ADGRG6* in lung tissue of COPD patients and patients with decreased DLCO/VA.

Heritability and genetic overlap

Studies on heritability of DLCO in the general population and unrelated individuals are lacking, and so far, DLCO heritability has been studied only in twins (15, 16), with a highest reported estimate of 44%. In our study, we estimated the REML based

heritability of DLCO using the GCTA tool in unrelated individuals of the Rotterdam Study (17), and observed an age- and sex-adjusted heritability of DLCO and DLCO/VA of 23% and 24%, respectively. We also investigated heritability based on known familial relationships in the Framingham Heart Study. Here we found an age and sex-adjusted heritability of DLCO and DLCO/VA of 49% and 45%, respectively. The latter heritability estimates among familial related individuals are in line with the heritability estimates in twin studies and highlight the robustness of our data. Importantly, our study is the first to investigate the lower bound of heritability of DLCO that would be estimated by family and twin studies (18). The advantage of estimating heritability in unrelated individuals using GCTA in addition to the approach based upon family and twin studies is, that GCTA calculates the proportion of heritability that covers the additive effects of commons SNPs only, and does not suffer from bias due to epistatic interactions or shared environment. The latter effects might indeed be present in family and twin studies, leading to an overestimation of the heritability (18-20).

Despite their similar estimates of heritability, DLCO and DLCO/VA showed to have different genetic determinants due to a genetic overlap between the two traits of 59%, explaining why we could not observe the same lead association in the two analyses.

Variation in ADGRG6

The meta-analysis of genetics variants of DLCO/VA yielded one genome-wide significant association, along with a number of suggestive associations that did not reach genome-wide significance. The lead variant (rs17280293) in this study is a missense SNP in *ADGRG6*, with a MAF of 0.03 which is comparable to that in public datasets (0.03 ExAC, 0.02 TOPMED and 0.03 in 1000 genomes). Mutation in this SNP causes an amino acid change (S123G), which is predicted to have a deleterious effect as indicated by both SIFT (21) and Polyphen2 (22). It is therefore likely that this SNP is functional in *ADGRG6*. In this study, we showed that this variant has a high posterior probability of causality compared to other SNPs in the same region and that this SNP is associated with different regulatory chromatin marks, promotor histone marks, and enhancer histone marks in different tissue cell lines including foetal lung fibroblast cell lines and lung carcinoma cell lines. In addition, rs17280293 always co-occurs with another functional SNP in the region (rs11155242, D'=1), which is an eQTL for *ADGRG6* in human lung tissue.

Previous studies have also shown that variation at *ADGRG6* is associated with spirometric measures of lung function (5) (7). Soler Artigas and colleagues observed a strong association between spirometry, particularly FEV_1/FVC, and another SNP rs148274477, which is in strong LD with rs17280293. However, since airflow limitation (i.e. a low FEV1/FVC ratio) might be correlated with low diffusing capacity due to loss of elastic recoil in subjects with emphysema, we assessed the possibility

that the observed association between rs17280293 and DLCO/VA might be driven by FEV_1/FVC. However, this sensitivity analysis indicated an independent association between rs17280293 and DLCO/VA because additional adjustment for FEV_1/FVC did not affect the estimate and no genetic overlap could be proven between DLCO/VA and FEV_1/FVC. Other studies have associated genetic variation in *ADGRG6* with height. In our study, adjustment for height did not affect the association between DLCO/VA and rs17280293, suggesting that the lead association in our GWAS is independent of height. In addition, genetic overlap disappeared after additional adjustment for height in the model, indicating no residual confounding by height in our analyses.

Furthermore, Eichstaedt and colleagues recently used whole genome sequence data from 19 Argentinean highlanders compared to 16 native American lowlanders and showed that rs17280293 might contribute to the physiological adaptations to hypobaric hypoxia (23).

Gene function and expression

The *ADGRG6* gene (adhesion G protein-coupled receptor G6) belongs to the G-protein coupled receptor (GPCR) super family, the largest known receptor family in the human genome. It has been previously shown to be essential in angiogenesis (24). *ADGRG6*, a relatively new adhesion GPCR, has been shown to promote vascular endothelial growth factor (VEGF) signaling, by modulating the expression of endothelial growth factor receptor 2 (VEGFR2). Since *ADGRG6* is involved in angiogenesis, which is critical for the development of pulmonary capillary beds during fetal life, deletion of *ADGRG6* leads to mid-gestation embryonic lethality due to failure in cardiovascular development. GWA studies of spirometric measures of airflow limitation (FEV_1/FVC ratio) have indicated several genes and pathways involved in branching morphogenesis and lung development, implicating an early life origin of complex adult respiratory diseases such as COPD. Intriguingly, this GWA study of diffusing capacity of the lung (DLCO and DLCO/VA) also indicates a gene (*ADGRG6*) which is implicated in cardiopulmonary development during fetal life.

The modulating effect of *ADGRG6* on *VEGFR2* expression was shown to be mediated through the transcriptional activation of *STAT5* and *GATA2* (24). Interestingly *GATA2* was recently linked to pulmonary alveolar proteinosis (25), a rare lung disease, characterized by an abnormal accumulation of pulmonary surfactant in the alveoli, leading to an altered gas exchange.

Moreover, knock down of *ADGRG6* in the mouse retina was shown to result in the suppression of hypoxia-induced angiogenesis (24). This information is interesting in two ways: first it links *ADGRG6* to hypoxia, which is very much related to gas-exchange. Second, processes in the retina might provide a unique insight into lung

microvasculature, since vascular changes in both the retina and the alveoli reflect very much the same process, i.e. micro-angiopathy.

Although there is a good body of evidence that *ADGRG6* is important in lung development and micro-angiopathy, mRNA expression of *ADGRG6* has not been studied in lung diseases such as COPD and decreased diffusing capacity. Therefore we performed an expression analysis of *ADGRG6* in human lung tissue and demonstrated that mRNA expression of *ADGRG6* is decreased significantly in patients with COPD and individuals with a decreased DLCO/VA.

Strengths and limitations

We conducted our analyses using data from two population based studies; the Rotterdam Study and the Framingham Heart Study. The strength of these studies is the population-based setting including data from smokers and non-smokers, and the standardized prospective data collection. We are not aware of other population-based cohort studies that have DLCO data in genotyped individuals available. Therefore, replication in other population-based cohorts was not possible. Yet, the results of the independent analyses in the Rotterdam study and the Framingham Heart Study show, that rs17280293 already reaches genome-wide significance in the Framingham Heart Study and replicates in the Rotterdam Study. Finally, a gene expression analysis on lung tissue was performed in our lab in very well-defined patient groups.

This study has also some limitations. First, for the measurements of diffusing capacity, the single-breath technique was used. This technique is known to underestimate measurements of alveolar volume in individuals with obstructive disease or air trapping, since diffusing capacity cannot be measured in poorly ventilated areas of the lung. It is also known that the underestimation of VA will be greater in more severe COPD and less in milder COPD. However, in our population-based cohorts, there are few individuals with severe COPD; therefore, reducing the impact of the underestimation of VA in our study. Secondly, haemoglobin corrected DLCO measures were only available in the Rotterdam Study. However, the performed sensitivity analysis with or without correction for haemoglobin did not materially change the results within the Rotterdam Study. Third, the high D' between rs17280293 and rs11155242 might suggest linked variant occurrence. However, the high D' between those variants –estimated using data from the 1000 genomes reference panel- could result from the inflated estimation of D' due to the low frequency of the SNPs. For this, it would be helpful to estimate the D' in a bigger reference panel such as the haplotype reference consortium whenever this information would be available in the future. Finally, in this study, we controlled for FEV_1/FVC in our models. We also investigated the genetic correlation between gas exchange and FEV_1/FVC. While controlling for FEV_1/FVC in our analysis is compelling that rs17280293 is independently associated with DLCO/VA, lack of genome-wide

genetic correlation between diffusing capacity and FEV_1/FVC does not exclude the possibility of pleiotropy at this specific locus, given that genetic correlation analyses are influenced by power, and our GWAS has a relatively small sample size. Therefore, caution is warranted by interpreting these results.

In conclusion, DLCO and DLCO/VA are heritable traits with a considerable proportion of variance in diffusing capacity of the lung explained by genetics. We identified a functional variant in *ADGRG6*, a gene which is involved in gas exchange and hypoxia and differentially expressed in lung tissue of patients with COPD and subjects with decreased diffusing capacity. Therefore, experimental studies are needed to investigate the pathophysiological mechanisms and their therapeutic implications.

URLs

METAL, http://www.sph.umich.edu/csg/abecasis/metal/.
GCTA, http://cnsgenomics.com/software/gcta/
Locuszoom plots, http://locuszoom.org/
Genetic correlation-LDscore, https://github.com/bulik/ldsc
Haploreg, http://archive.broadinstitute.org/mammals/haploreg/haploreg.php
GTEx portal, http://www.gtexportal.org/home/
GTEx portal eQTL data, lung tissue set obtained from this location:
javascript:portalClient.browseDatasets.downloadFile('Lung.allpairs.txt.gz','gtex_analysis_v7/single_tissue_eqtl_data/all_snp_gene_associations/Lung.allpairs.txt.gz')

REFERENCES

1. Hegewald MJ. Diffusing capacity. Clin Rev Allergy Immunol. 2009;37(3):159-66.
2. Macintyre N, Crapo RO, Viegi G, Johnson DC, van der Grinten CP, Brusasco V, et al. Standardisation of the single-breath determination of carbon monoxide uptake in the lung. Eur Respir J. 2005;26(4):720-35.
3. Crapo RO, Jensen RL, Wanger JS. Single-breath carbon monoxide diffusing capacity. Clin Chest Med. 2001;22(4):637-49.
4. Bailey KL. The importance of the assessment of pulmonary function in COPD. Med Clin North Am. 2012;96(4):745-52.
5. Hancock DB, Eijgelsheim M, Wilk JB, Gharib SA, Loehr LR, Marciante KD, et al. Meta-analyses of genome-wide association studies identify multiple loci associated with pulmonary function. Nat Genet. 2010;42(1):45-52.
6. Loth DW, Soler Artigas M, Gharib SA, Wain LV, Franceschini N, Koch B, et al. Genome-wide association analysis identifies six new loci associated with forced vital capacity. Nat Genet. 2014;46(7):669-77.
7. 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.
8. Wain LV, Shrine N, Artigas MS, Erzurumluoglu AM, Noyvert B, Bossini-Castillo L, et al. Genome-wide association analyses for lung function and chronic obstructive pulmonary disease identify new loci and potential druggable targets. Nat Genet. 2017;49(3):416-25.
9. Hofman A, Brusselle GG, Darwish Murad S, van Duijn CM, Franco OH, Goedegebure A, et al. The Rotterdam Study: 2016 objectives and design update. Eur J Epidemiol. 2015;30(8):661-708.
10. Yang J, Lee SH, Wray NR, Goddard ME, Visscher PM. GCTA-GREML accounts for linkage disequilibrium when estimating genetic variance from genome-wide SNPs. Proc Natl Acad Sci U S A. 2016;113(32):E4579-80.
11. Almasy L, Blangero J. Multipoint quantitative-trait linkage analysis in general pedigrees. American Journal of Human Genetics. 1998;62(5):1198-211.
12. Hobbs BD, de Jong K, Lamontagne M, Bosse Y, Shrine N, Artigas MS, et al. Genetic loci associated with chronic obstructive pulmonary disease overlap with loci for lung function and pulmonary fibrosis. Nat Genet. 2017;49(3):426-32.
13. Cho MH, Castaldi PJ, Hersh CP, Hobbs BD, Barr RG, Tal-Singer R, et al. A Genome-Wide Association Study of Emphysema and Airway Quantitative Imaging Phenotypes. Am J Respir Crit Care Med. 2015;192(5):559-69.
14. Benner C, Spencer CC, Havulinna AS, Salomaa V, Ripatti S, Pirinen M. FINEMAP: efficient variable selection using summary data from genome-wide association studies. Bioinformatics. 2016;32(10):1493-501.
15. Ghio AJ, Crapo RO, Elliott CG, Adams TD, Hunt SC, Jensen RL, et al. Heritability estimates of pulmonary function. Chest. 1989;96(4):743-6.
16. Hallberg J, Iliadou A, Anderson M, de Verdier MG, Nihlen U, Dahlback M, et al. Genetic and environmental influence on lung function impairment in Swedish twins. Respir Res. 2010;11:92.
17. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. Am J Hum Genet. 2011;88(1):76-82.
18. Plomin R, Deary IJ. Genetics and intelligence differences: five special findings. Mol Psychiatry. 2015;20(1):98-108.

19. Zaitlen N, Kraft P, Patterson N, Pasaniuc B, Bhatia G, Pollack S, et al. Using Extended Genealogy to Estimate Components of Heritability for 23 Quantitative and Dichotomous Traits. Plos Genetics. 2013;9(5).
20. Yang J, Bakshi A, Zhu Z, Hemani G, Vinkhuyzen AA, Lee SH, et al. Genetic variance estimation with imputed variants finds negligible missing heritability for human height and body mass index. Nat Genet. 2015;47(10):1114-20.
21. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat Protoc. 2009;4(7):1073-81.
22. Adzhubei I, Jordan DM, Sunyaev SR. Predicting functional effect of human missense mutations using PolyPhen-2. Curr Protoc Hum Genet. 2013;Chapter 7:Unit7 20.
23. Eichstaedt CA, Pagani L, Antao T, Inchley CE, Cardona A, Morseburg A, et al. Evidence of Early-Stage Selection on EPAS1 and GPR126 Genes in Andean High Altitude Populations. Sci Rep. 2017;7(1):13042.
24. Cui H, Wang Y, Huang H, Yu W, Bai M, Zhang L, et al. GPR126 protein regulates developmental and pathological angiogenesis through modulation of VEGFR2 receptor signaling. J Biol Chem. 2014;289(50):34871-85.
25. Papiris SA, Tsirigotis P, Kolilekas L, Papadaki G, Papaioannou AI, Triantafillidou C, et al. Pulmonary alveolar proteinosis: time to shift? Expert Rev Respir Med. 2015;9(3):337-49.

3.3 Occupational exposure to gases/fumes and mineral dust affect DNA methylation levels of genes regulating expression

ABSTRACT

Background: Many workers are daily exposed to occupational agents like gases/fumes, mineral dust or biological dust, but how these exposures induce adverse health effects is still largely unknown. Epigenetic mechanisms, such as DNA-methylation, have been suggested to play a role. We therefore aimed to identify differentially methylated regions (DMRs) upon occupational exposures in never-smokers and investigated if these DMRs associated with gene expression levels.

Methods: To determine the effects of occupational exposures independent of smoking, 903 never-smokers of the Dutch LifeLines cohort study were included. Illumina 450K arrays were used to obtain genome-wide blood DNA methylation data. We performed three genome-wide methylation analyses, one per occupational exposure being gases/fumes, mineral dust and biological dust, using robust linear regression adjusted for appropriate confounders. DMRs were identified using comb-p in phython. Results were validated in the Rotterdam Study (233 never-smokers) and methylation-expression associations were assessed using Biobank-based Integrative Omics Studies Consortium (BIOS) data (n=2,802).

Results: Of in total 21 significant DMRs, 14 DMRs were associated with gases/fumes and 7 with mineral dust. Three DMRs were associated with both gases/fumes and mineral dust (*RPLP1* and *LINC02169* (2x)). The majority of DMRs were located within transcript start sites of gene expression regulating genes (11 out of 21 DMRs) and some DMRs were located in genes that were previously linked to lung diseases. We replicated three DMRs, two with gases/fumes (*VTRNA2-1* and *GNAS*), and one with mineral dust (*CCDC144NL*). Fifteen DMRs, 9 with gases/fumes and 6 with mineral dust, significantly associated with gene expression levels.

Conclusion: Our data suggest that occupational exposures may induce differential methylation of gene expression regulating genes and thereby may induce adverse health effects. Given the millions of workers that are exposed daily to occupational exposures, further studies on this epigenetic mechanism and health outcomes are warranted.

Keywords: Occupational exposure; DNA methylation; Epigenetic epidemiology; Genome-wide

INTRODUCTION

Daily, millions of workers worldwide are exposed to chemical agents, fumes, and (in) organic dusts.(WHO 2010) The leading occupational causes of death in 2000 were unintentional injuries (41%), chronic obstructive pulmonary disease (COPD, 40%) and lung cancer (13%).(WHO 2010) This is not remarkable, since the skin and the lungs are most directly exposed to occupational pollutants, which could be prevented by implementing protective measures. Studies focusing on specific occupations, like pig farmers, miners, construction and textile workers, found associations between job specific exposures and a faster annual decline in lung function (FEV_1).(Bakke et al. 2004; Iversen and Dahl 2000; Wang et al. 2008) In addition, we have previously shown that exposure to gases/fumes, mineral and biological dust is associated with small and large airways obstruction.(de Jong et al. 2014a, 2014b)

Even though occupational exposures are common, it is still largely unknown how these exposures are involved in (lung) disease development. Epigenetic mechanisms such as DNA methylation have been suggested to play a role and researchers have therefore advocated the importance of epigenetic studies into environmental exposures and lung health.(Melén et al. 2018) Environmental exposures, like occupational exposures, induce changes in DNA methylation levels, which can affect gene expression, possibly aiding in disease development.(Ruiz-Hernandez et al. 2015) DNA methylation is the addition of a methyl group to the DNA without altering its sequence. This usually occurs at sites where a cytosine base is adjacent to a guanine base (CpG) and can have a regulatory function on gene expression.(Jones 2012) Several small studies showed suggestive evidence that specific compounds found in occupational exposures, like cadmium, lead, and mercury affect DNA methylation.(Goodrich et al. 2013; Hossain et al. 2012; Li et al. 2013; Ruiz-Hernandez et al. 2015)

To date, no large hypothesis-free genome-wide DNA-methylation studies assessing the association between occupational exposures and DNA methylation levels have been performed. We therefore aimed to identify differentially methylated CpG-sites (CpGs) and differentially methylated regions (DMRs) associated with occupational exposure to gases/fumes, mineral dust and biological dust, and to assess the effects of these regions on gene expression levels. To determine the effects of occupational exposures independent of smoking exposure, the analyses were restricted to never-smokers.

METHODS

Population and measurements

From the Lifelines Cohort Study, 1,656 unrelated subjects were selected for DNA methylation assessment.(van der Plaat et al. 2018) Subject selection was based on creating relatively equal-sized groups based on age, smoking, occupational exposures, and spirometry. In the current study only never-smokers were included in order to determine the effects of occupational exposures independent of smoking exposure. No, low and high occupational exposure to gases/fumes, mineral dust and biological dust was estimated using the ALOHA+ job exposure matrix (JEM), based on current or last held job.(de Jong et al. 2014b; Matheson et al. 2005) See Figure 1 and supplementary methods for an overview and more detailed information on the methods.

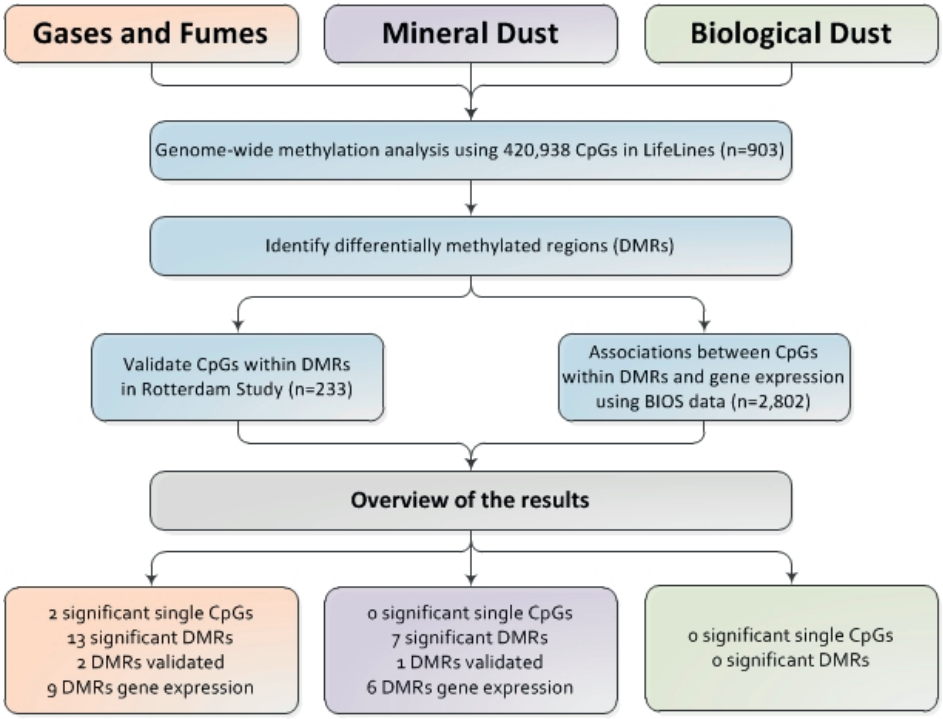

Figure 1: Overview of the performed analyses and results per occupational exposure. All analyses were performed for the three exposures in never-smokers.

Genome-wide methylation data and analysis

Illumina Infinium Human Methylation 450K arrays were used to obtain genome-wide blood DNA methylation data. Samples were processed using the Illumina protocol. Quality control (QC) using *Minfi* and normalization using *DASEN* (*watermelon*) was

performed in R.(Aryee et al. 2014; Pidsley et al. 2013) Quality-control (QC) steps included the removal of samples with >1% of all probes having a detection p-value >0.01, and samples with incorrect sex or SNP prediction. We removed single probes with a detection p-value >0.01, sex chromosome probes, cross-reactive probes(Chen et al. 2013), probes measuring SNPs, and probes where the CpG itself or the single base extension (SBE) site is a SNP. After QC, we had complete data for 420,938 CpG probes in 903 never-smoking subjects.

We performed three genome-wide methylation analyses, one per occupational exposure being gases/fumes, mineral dust and biological dust. We assessed associations between DNA methylation levels (beta-values ranging from 0 to 1) and the three occupational exposures separately using robust linear regression in R [*MASS* package]. Models included low and high exposure dummy-variables (no exposure as reference), and were adjusted for sex, age, technical variances, and differential blood counts (eosinophilic, neutrophilic, and basophilic granulocytes, lymphocytes and monocytes, all obtained using standard laboratory techniques). Single CpGs with a false discovery rate (FDR) adjusted p-value <0.05 for the high exposure dummy-variable were considered genome-wide significant.

Identification of differentially methylated regions (DMRs)

To identify differentially methylated regions (DMRs), *comb-p* in python was used. (Pedersen et al. 2012) Comb-p identifies regions of enrichment by combining adjacent p-values into FDR adjusted regional p-values using auto-correlation and sliding windows. As input we used p-values of the high exposure dummy-variable with the following settings: seed=0.01 and distance=300. Each CpG within a DMR with a Šidák corrected p-value <0.05 was further investigated.(Sidak 1967)

Validation of DMRs

DMRs identified in Lifelines were validated in the baseline assessment of the third Rotterdam Study cohort (RS-III-1, n=722).(Hofman et al. 2015) Blood DNA methylation levels were measured using Illumina 450K arrays and processed similar to LifeLines as described elsewhere.(Ligthart et al. 2016) All CpGs within the DMRs were validated in RS-III-1 and the statistical models (robust linear regression) were similar to LifeLines. Single CpGs and CpGs within the DMRs with a nominal validation p-value <0.05 and same direction of effect in both cohorts were considered significantly replicated.

Association between CpGs within DMRs and gene expression

To assess whether the CpGs within the DMRs were associated with gene expression levels, we used data from four population-based cohorts within Biobank-based Integrative Omics Studies (BIOS), from the Biobanking and Biomolecular Resources

Research Infrastructure for The Netherlands (BBMRI-NL).(BBMRI-NL. 2017) In total, 2,802 subjects were included in the analyses (independent samples of LifeLines (n=727), Rotterdam Study III-2 (n=589), Netherlands Twin Registry (n=900), and Leiden Longevity Study (n=586)).(Hofman et al. 2011; Tigchelaar et al. 2015; Westendorp et al. 2009; Willemsen et al. 2013) In each cohort, probesets (read counts from RNA sequencing) within 1Mb around the CpG were assessed and the linear regression was adjusted for sex, smoking, age and technical variances. Effect estimates of the cohorts were meta-analysed. CpGs with a meta-analysis p-value below the Bonferroni corrected threshold (p=0.05/number of probesets in 1MB window) were considered significant.

RESULTS

Population characteristics

Complete data on all covariates were available for 903 never-smokers in the identification cohort LifeLines and 233 never-smokers in the validation cohort Rotterdam Study (Table 1 and Table S1).

We present the results of our analyses per exposure (i.e. gases/fumes, mineral dust and biological dust). For an overview of all results see Figure 1. The results of all analyses can be found in the supplementary Excel-file and the Manhattan plots are shown in Figure S1.

Table 1. Characteristics of the never-smokers included in the LifeLines cohort study (discovery cohort) and the Rotterdam Study (validation cohort).

	LifeLines	**Rotterdam Study**
N with no missing data	903	233
Males, N (%)	508 (57)	100 (43)
Age (years), median (min-max)	46 (18-80)	57 (47-89)
Occupational exposure, N	*No/Low/High*	*No/Low/High*
Gases/Fumes	637 / 150 / 116	177 / 51 / 5
Mineral dust	673 / 105 / 125	210 / 20 / 3
Biological dust	720 / 69 / 114	N/A

SD = Standard deviation N/A = Not applicable
See table S1 for the characteristics of the LifeLines cohort separately per exposure level.
NB The LifeLines sample is a selected population, not a sample from the general population

Gases/Fumes

Genome-wide methylation analysis

In the genome-wide methylation analysis in never-smokers, two single CpGs were epigenome-wide significantly associated with gases/fumes exposure (FDR<0.05) (Table S2). These CpGs are annotated to Ribosomal Protein L37a (RPL37A) and Grid2 Interacting Protein (GRID2IP).

Identification of DMRs

Thirteen DMRs were significantly associated with exposure to gases/fumes (Table S3). The three most significant DMRs are annotated to long intergenic non-protein coding RNA 2169 (LINC02169), Ribosomal Protein Lateral Stalk Subunit P1 (RPLP1) and leptin (LEP). The genome-wide significant CpG annotated to RPL37A was not located within an identified DMR.

Validation of the DMRs

In the validation analysis, two DMRs contained a significantly replicated CpG and exposure to gases/fumes was associated with lower methylation levels at these CpGs in both cohorts (Table 2 and Table S4). These two DMRs are annotated to Vault RNA 2-1 (*VTRNA2-1*, a.k.a. *MIR886*) and Guanine Nucleotide Binding Protein Alpha Stimulating Activity (*GNAS*) (Figure 2 A/B).

Gene expression analysis

We found that CpGs within nine out of 14 DMRs were significantly associated with differential gene expression, the direction of effect was predominantly negative. Table 3 presents the significant methylation-expression associations of CpGs within replicated DMRs. For the gene expression results of all DMRs identified with gases/fumes see Table S5. The replicated DMRs annotated to *GNAS* was associated with lower expression of *NPEPL1*.

Mineral dust

Genome-wide methylation analysis and identification of DMRs

No CpGs were genome-wide significantly associated with mineral dust exposure (FDR<0.05), but seven DMRs were (Tables S3). The three most significant hits are annotated to *RPLP1*, *LINC02169* and Major Histocompatibility Complex Class I E (*HLA-E*), and the first two mentioned DMRs were also associated with exposure to gases/fumes.

Table 2. Results of the CpGs within the replicated differentially methylated regions (DMRs) associated with occupational exposures in never-smokers.

				N	Annotation			P		LifeLines			Rotterdam Study		
DMR	**Chr**	**Start**	**End**	**probes**	**Gene**	**Feature**	**Island**	**region**	**CpG**	**Beta**	**SE**	**P**	**Beta**	**SE**	**P**
Gases/Fumes															
GN4	chr5	135416331	135416579	7	*VTRNA2-1*	TSS200	Island	$2.68*10^{-5}$	cg18678645	-2.84	0.83	$5.96*10^{-4}$	-5.68	3.49	0.103
									cg06536614	-1.31	0.56	$1.90*10^{-2}$	-3.25	2.47	0.188
									cg26328633	-1.88	0.59	$1.50*10^{-3}$	-5.81	2.67	**0.029**
									cg25340688	-1.96	0.63	$1.80*10^{-3}$	-2.04	2.89	0.479
									cg26896946	-1.10	0.44	$1.18*10^{-2}$	-2.33	2.33	0.318
									cg00124993	-1.59	0.55	$3.75*10^{-3}$	-6.23	3.32	0.061
									cg08745965	-2.21	0.84	$8.84*10^{-3}$	-6.87	4.50	0.127
GN13	chr20	57427713	57427880	6	*GNAS; GNAS-AS1*	Intron; TSS1500	Island	$2.44*10^{-2}$	cg04257105	-1.31	0.52	$1.25*10^{-2}$	-2.45	1.69	0.146
									cg20528838	-1.20	0.42	$4.24*10^{-3}$	-2.52	1.61	0.119
									cg27661264	-1.39	0.46	$2.53*10^{-3}$	-3.15	2.24	0.160
									cg19589727	-1.23	0.45	$6.27*10^{-3}$	-.52	1.88	0.783
									cg10302550	-1.06	0.54	$4.69*10^{-2}$	-4.54	2.31	**0.049**
									cg17414107	-1.32	0.59	$2.54*10^{-2}$	-1.81	2.75	0.510
Mineral dust															
MN7	chr17	20799408	20799694	6	*CCDC144NL; RP11-344E13.3*	TSS200; 5'UTR	Island	$1.13*10^{-3}$	cg08458692	-1.06	0.54	$5.03*10^{-2}$	-6.07	2.30	**0.008**
									cg14560110	-3.36	1.36	$1.37*10^{-2}$	-9.67	6.06	0.111
									cg08288433	-3.05	0.93	$1.05*10^{-3}$	-8.33	7.66	0.277
									cg06809326	-2.81	0.92	$2.14*10^{-3}$	-10.25	5.65	0.070
									cg22570042	-3.20	1.01	$1.48*10^{-3}$	-14.71	7.64	0.054
									cg21980100	-2.05	0.85	$1.62*10^{-2}$	-2.90	3.94	0.461

DMR = Differentially methylated region Chr = Chromosome CpG = DNA-methylation site P = p-value SE = Standard Error

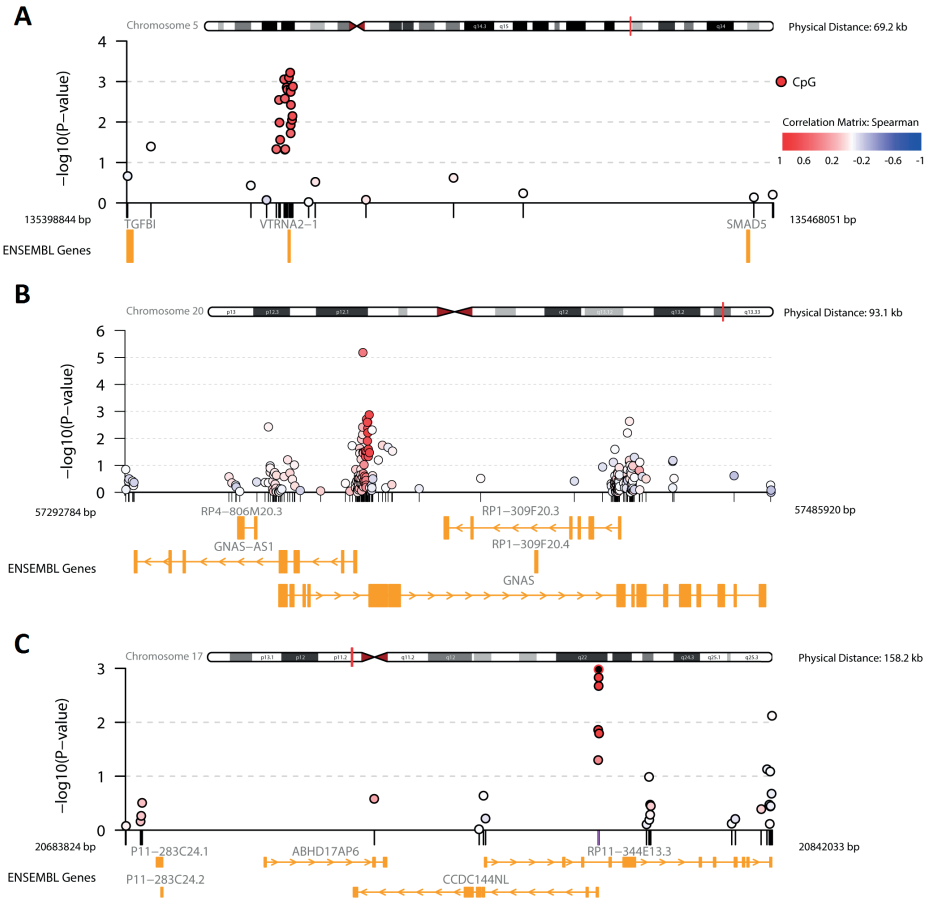

Figure 2: Regional association plots (R package comet) for the three replicated DMRs in never-smokers. A) DMR annotated to *VTRNA2-1*, B) DMR annotated to *GNAS*, and C) DMR annotated to *CCDC144NL*. x-axis = megabase (Mb) position on the chromosome, y-axis = negative log10 of the p-values, dots = CpG-sites, and see inset legend for the correlation explanation between CpGs.

Validation of the DMRs

The DMR annotated to Coiled-Coil Domain Containing 144 Family, N-Terminal Like (*CCDC144NL*) contained a significantly replicated CpG and the association between mineral dust exposure and methylation levels was negative in both cohorts (Tables 2 and S4, and Figure 2C).

Gene expression analysis

In total, CpGs within six out of seven DMRs were significantly associated with differential gene expression and the direction of effect was predominantly negative

(Table S6). The replicated DMR annotated to *CCDC144NL* was associated with lower expression of Abhydrolase Domain Containing 17A Pseudogene 6 (*ABHD17AP6*).

Biological dust

No single CpGs or DMRs were genome-wide significantly associated with biological dust exposure in never-smokers (FDR<0.05). Therefore, no validation of results or methylation-expression analyses were performed.

Table 3. Results of replicated DMRs in never-smokers which were associated with gene expression levels for genes located within 1MB of the CpG (n=2,802)

DMR	CpG	Annotated Gene	Ensembl_ID	Gene	B	SE	P Adjusted
Gases/Fumes							
GN13	cg04257105	*GNAS*	ENSG00000254419	*NPEPL1*	-0.428	0.134	$2.19*10^{-2}$
	cg17414107	*GNAS*	ENSG00000254419	*NPEPL1*	-0.388	0.125	$2.81*10^{-2}$
	cg19589727	*GNAS*	ENSG00000254419	*NPEPL1*	-0.580	0.188	$3.03*10^{-2}$
	cg20528838	*GNAS*	ENSG00000254419	*NPEPL1*	-0.472	0.156	$3.67*10^{-2}$
Mineral dust							
MN7	cg06809326	*CCDC144NL*	ENSG00000226981	*ABHD17AP6*	-1.132	0.251	$9.44*10^{-5}$
	cg08288433	*CCDC144NL*	ENSG00000226981	*ABHD17AP6*	-0.992	0.231	$2.68*10^{-4}$
	cg14560110	*CCDC144NL*	ENSG00000226981	*ABHD17AP6*	-0.977	0.187	$2.57*10^{-6}$
			ENSG00000170298	*LGALS9B*	0.554	0.168	$1.42*10^{-2}$
	cg21980100	*CCDC144NL*	ENSG00000226981	*ABHD17AP6*	-1.327	0.304	$1.92*10^{-4}$
	cg22570042	*CCDC144NL*	ENSG00000226981	*ABHD17AP6*	-1.032	0.216	$2.56*10^{-5}$

DMR = Differentially methylated region CpG = DNA-methylation site B = beta SE = Standard Error P adjusted = Bonferroni correct meta-analysis p-value based on genes with available data located within the 1MB window of the CpG.

DISCUSSION

This is the first genome-wide DNA-methylation study assessing the association between occupational exposures and DNA methylation. Since it is well known that smoking is associated with extensive changes in DNA methylation levels, we restricted our analyses to never-smokers.(Zeilinger et al. 2013) In these never-smokers, occupational exposure to gases/fumes and to mineral dust was associated with 14 and 7 differentially methylated regions (DMRs), respectively. Three of these DMRs were associated with both gases/fumes and mineral dust (1 DMR in *RPLP1* and 2 DMRs in *LINC02169*). We were able to replicate the result of two DMRs associated with gases/fumes and one DMR was associated with mineral dust in the Rotterdam Study. These three DMRs were annotated to *VTRNA2-1*, *GNAS* and *CCDC144NL*. CpGs within the DMRs annotated

to *GNAS* and *CCDC144NL* were significantly associated with lower expression levels of *NPEPL1* and *ABHD17AP6,* respectively. Moreover, 14 out of 21 DMRs were associated with gene expression levels and 11 DMRs were located within the transcript start sites (TSS) of a gene. Together, our data suggest that occupational exposures may induce differential DNA-methylation at specific genomic locations and this may be a mechanism through which occupational exposures affect health.

Interestingly, the majority of identified DMRs were located within the TSS of a gene, including the three replicated DMRs, of which two were also associated with gene expression levels (*GNAS* and *CCDC144NL*). The general idea of the function of DNA methylation at these TSSs is that it blocks the initiation of transcription and thereby lowers gene expression.(Jones 2012) In the current study, we observed that occupational exposure is associated with lower DNA methylation levels which in turn is associated with higher gene expression levels for most DMRs associated with gene expression levels. This observation thus corroborates our knowledge of the function of DNA methylation at TSSs. Moreover, several of the DMRs associated with gene expression were not associated with the annotated gene. This is partly due to the fact that for 11 of our identified DMRs no gene expression data was available for the annotated gene, including the replicated DMR annotated to *VTRNA2-1*. For others, CpGs within a DMR were nominally associated with expression levels of the annotated gene, but did not survive the multiple testing correction (e.g. the replicated DMR annotated *GNAS)*.

Another intriguing observation is that several DMRs that we identified are annotated to or associated with the expression of genes with unknown function, RNA genes or pseudogenes, like *CCDC144NL, ABHD17AP6, NPEPL1, RP11-373N24.2,* and *LINC02169*. It is therefore challenging to understand the relation between these genes and occupational exposures. Long non-coding RNAs (lncRNAs) are known to play a role in gene expression regulation during development, cell differentiation, genomic imprinting, and sex chromosomal dosage compensation.(Fatica and Bozzoni 2014) The gene *ZSCAN26* is a zinc finger (transcription factor) and may therefore also be involved in gene expression regulation.(Klug 2010) In addition, multiple microRNAs and lncRNAs were shown to be key regulators of gene expression in lung diseases such as asthma and COPD.(Booton and Lindsay 2014) These might even be biomarkers or therapeutic targets, but more research into the function of these genes is warranted. Overall, our data seem to suggest that occupational exposures affect regulation of gene expression by changing DNA methylation levels of particular genes that regulate the expression of other genes.

Interestingly, the three DMRs annotated to *RPLP1* and *LINC02169* (2x) were identified in both the gases/fumes and mineral dust analyses. In addition, CpGs annotated to *VTRNA2-1* were also associated with occupational exposure to pesticides.(van der Plaat et al. 2018) *RPLP1* is a ribosomal protein regulating translation and *VTRNA2-1*

is indirectly also related to the innate immune response, since it was shown to inhibit Protein Kinase R (*EIF2AK2*).(Kunkeaw et al. 2013; Martinez-Azorin et al. 2008) This could indicate that different types of occupational exposures affect similar pathways, alternatively it could result from multiple occupational exposures in specific jobs. For example, construction workers can be exposed to mineral dust and gases/fumes at the same time and crop farmers distribute pesticides over their fields using fuelled machines (gases/fumes exposure). Notably, 8 subjects of our cohort were highly exposed to all three occupational exposures and the exposures are moderately to strongly correlated (Correlation between gases/fumes and mineral dust = 0.85, between gases/fumes and biological dust = 0.66, and between mineral dust and biological dust = 0.56, Table S7). Since we used broad categories of occupational exposures, it was not possible to investigate specific exposure molecules. Occupational exposure levels were also estimated based on current or last held job and duration of exposure was not taken into account. It is likely that some subjects classified as non-exposed have changed from an "exposed" to a "non-exposed" job, because they experienced adverse effects from the exposures. Therefore, we may have underestimated the effect of occupational exposures on DNA methylation. However, in our cohort on average 72% of the subjects currently exposed to either gases/fumes, mineral or biological dust had this job for more than 5 years and thus had been exposed for a substantial time period in the same job.

Another restriction of our study is the use of blood DNA-methylation levels. DNA methylation is cell and tissue specific, and the main route of occupational exposure is via inhalation or skin absorption. However, we have validated a number of CpGs associated with cigarette smoke exposure in lung tissue that were originally identified in whole blood.[submitted] Thus using whole blood could be an efficient way to identify differential DNA methylation upon exposures as an accessible proxy for changes in lung tissue. Furthermore, using a JEM does not allow to assess specific chemical compounds present in occupational exposures, nor the effect of lifetime exposure. Different types of jobs classified into the same exposure category might contain different chemical compounds as well. Therefore, our results reflect the effect of current or recent occupational exposure on DNA methylation.

In conclusion, our data suggests that occupational exposures may induce differential methylation of genes that regulate gene expression and therefore occupational exposures may induce adverse health effects via this methylation. Several of our identified differentially methylated regions upon occupational exposure to gases/fumes and mineral dust were associated with gene expression levels. Some regions were even associated with two types of occupational exposure. Given the millions of workers that are exposed daily to occupational exposures, further studies on this epigenetic mechanism and health outcomes are warranted. For example, since 40% of the occupational cause of death is due to COPD, especially in developing countries

without proper precautions, further studies on this epigenetic mechanism could aid in reducing the global burden of COPD.(WHO 2010)

REFERENCES

Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD, et al. 2014. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. Bioinformatics 30:1363–1369; doi:10.1093/bioinformatics/btu049 [doi].

Bakke B, Ulvestad B, Stewart P, Eduard W. 2004. Cumulative exposure to dust and gases as determinants of lung function decline in tunnel construction workers. Occup Environ Med 61: 262–269.

Biobanking and Biomolecular Resources Research Infrastructure for The Netherlands (BBMRI-NL). 2017. Biobank-based Integrative Omics Studies (BIOS). 2017. Available: http://www.bbmri.nl/on_offer/bios/.

Booton R, Lindsay MAA. 2014. Emerging Role of MicroRNAs and Long Noncoding RNAs in Respiratory Disease. Chest 146:193–204; doi:10.1378/chest.13-2736 [doi].

Chen YA, Lemire M, Choufani S, Butcher DT, Grafodatskaya D, Zanke BW, et al. 2013. Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray. Epigenetics 8:203–209; doi:10.4161/epi.23470 [doi].

de Jong K, Boezen HM, Kromhout H, Vermeulen R, Postma DS, Vonk JM, et al. 2014a. Pesticides and other occupational exposures are associated with airway obstruction: the LifeLines cohort study. Occup Environ Med 71:88–96; doi:10.1136/oemed-2013-101639 [doi].

de Jong K, Boezen HM, Kromhout H, Vermeulen R, Vonk JM, Postma DS, et al. 2014b. Occupational exposure to vapors, gases, dusts, and fumes is associated with small airways obstruction. Am J Respir Crit Care Med 189:487–490; doi:10.1164/rccm.201310-1919LE [doi].

Fatica A, Bozzoni I. 2014. Long non-coding RNAs: new players in cell differentiation and development. Nat Rev 15:7–21; doi:10.1038/nrg3606 [doi].

Goodrich JM, Basu N, Franzblau A, Dolinoy DC. 2013. Mercury biomarkers and DNA methylation among Michigan dental professionals. Environ Mol Mutagen 54:195–203; doi:10.1002/em.21763 [doi].

Hofman A, Brusselle GG, Darwish Murad S, van Duijn CM, Franco OH, Goedegebure A, et al. 2015. The Rotterdam Study: 2016 objectives and design update. Eur J Epidemiol 30:661–708; doi:10.1007/s10654-015-0082-x [doi].

Hofman A, van Duijn CM, Franco OH, Ikram MA, Janssen HL, Klaver CC, et al. 2011. The Rotterdam Study: 2012 objectives and design update. Eur J Epidemiol 26:657–686; doi:10.1007/s10654-011-9610-5 [doi].

Hossain MB, Vahter M, Concha G, Broberg K. 2012. Low-level environmental cadmium exposure is associated with DNA hypomethylation in Argentinean women. Environ Health Perspect 120:879–884; doi:10.1289/ehp.1104600 [doi].

Iversen M, Dahl R. 2000. Working in swine-confinement buildings causes an accelerated decline in FEV1: a 7-yr follow-up of Danish farmers. Eur Respir J 16: 404–408.

Jones PA. 2012. Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nat Rev 13:484–492; doi:10.1038/nrg3230 [doi].

Klug A. 2010. The discovery of zinc fingers and their applications in gene regulation and genome manipulation. Annu Rev Biochem 79:213–231; doi:10.1146/annurev-biochem-010909-095056 [doi].

Kunkeaw N, Jeon SH, Lee K, Johnson BH, Tanasanvimon S, Javle M, et al. 2013. Cell death/proliferation roles for nc886, a non-coding RNA, in the protein kinase R pathway in cholangiocarcinoma. Oncogene 32:3722–3731; doi:10.1038/onc.2012.382 [doi].

Li C, Yang X, Xu M, Zhang J, Sun N. 2013. Epigenetic marker (LINE-1 promoter) methylation level was associated with occupational lead exposure. Clin Toxicol (Phila) 51:225–229; doi:10.3109/15563650.2013.782410 [doi].

Ligthart S, Steenaard R V, Peters MJ, van Meurs JB, Sijbrands EJ, Uitterlinden AG, et al. 2016. Tobacco smoking is associated with DNA methylation of diabetes susceptibility genes. Diabetologia 59:998–1006; doi:10.1007/s00125-016-3872-0 [doi].

Martinez-Azorin F, Remacha M, Ballesta JP. 2008. Functional characterization of ribosomal P1/P2 proteins in human cells. Biochem J 413:527–534; doi:10.1042/BJ20080049 [doi].

Matheson MC, Benke G, Raven J, Sim MR, Kromhout H, Vermeulen R, et al. 2005. Biological dust exposure in the workplace is a risk factor for chronic obstructive pulmonary disease. Thorax 60:645–651; doi:60/8/645 [pii].

Melén E, Barouki R, Barry M, Boezen HM, Hoffmann B, Krauss-Etschmann S, et al. 2018. Promoting respiratory public health through epigenetics research: an ERS Environment Health Committee workshop report. Eur Respir J 51:1702410; doi:10.1183/13993003.02410-2017.

Pedersen BS, Schwartz DA, Yang I V, Kechris KJ. 2012. Comb-p: software for combining, analyzing, grouping and correcting spatially correlated P-values. Bioinformatics 28:2986–2988; doi:10.1093/bioinformatics/bts545 [doi].

Pidsley R, Y Wong CC, Volta M, Lunnon K, Mill J, Schalkwyk LC. 2013. A data-driven approach to preprocessing Illumina 450K methylation array data. BMC Genomics 14:293; doi:10.1186/1471-2164-14-293 [doi].

Ruiz-Hernandez A, Kuo CC, Rentero-Garrido P, Tang WY, Redon J, Ordovas JM, et al. 2015. Environmental chemicals and DNA methylation in adults: a systematic review of the epidemiologic evidence. Clin Epigenetics 7:55–015–0055–7. eCollection 2015; doi:10.1186/s13148-015-0055-7 [doi].

Sidak Z. 1967. Rectangular Confidence Regions for the Means of Multivariate Normal Distributions. J Am Stat Assoc 62:626–633; doi:10.2307/2283989.

Tigchelaar EF, Zhernakova A, Dekens JA, Hermes G, Baranska A, Mujagic Z, 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; doi:10.1136/bmjopen-2014-006772 [doi].

van der Plaat DA, de Jong K, de Vries M, van Diemen CC, Nedeljković I, Amin N, et al. 2018. Occupational exposure to pesticides is associated with differential DNA methylation. Occup Environ Med 75:427–435; doi:10.1136/oemed-2017-104787.

Wang X, Zhang HX, Sun BX, Dai HL, Hang JQ, Eisen E, et al. 2008. Cross-shift airway responses and long-term decline in FEV1 in cotton textile workers. Am J Respir Crit Care Med 177:316–320; doi:200702-318OC [pii].

Westendorp RG, van Heemst D, Rozing MP, Frolich M, Mooijaart SP, Blauw GJ, et al. 2009. Nonagenarian siblings and their offspring display lower risk of mortality and morbidity than sporadic nonagenarians: The Leiden Longevity Study. J Am Geriatr Soc 57:1634–1637; doi:10.1111/j.1532-5415.2009.02381.x [doi].

WHO | Chapter 21: Selected occupational risk factors. 2010. WHO.

Willemsen G, Vink JM, Abdellaoui A, den Braber A, van Beek JH, Draisma HH, et al. 2013. The Adult Netherlands Twin Register: twenty-five years of survey and biological data collection. Twin Res Hum Genet 16:271–281; doi:10.1017/thg.2012.140 [doi].

Zeilinger S, Kuhnel B, Klopp N, Baurecht H, Kleinschmidt A, Gieger C, et al. 2013. Tobacco smoking leads to extensive genome-wide changes in DNA methylation. PLoS One 8:e63812; doi:10.1371/journal.pone.0063812 [doi].

3.4 Epigenome-wide association study on diffusing capacity of the lung: replication and meta-analysis

ABSTRACT

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

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

Methods: Epigenome association study (EWAS) was performed on diffusing capacity, measured by carbon monoxide uptake (DLCO) and per alveolar volume (DLCO/VA) using the single-breath technique in 2,674 individuals from two population-based cohort studies, the Rotterdam Study (the discovery panel) and the Framingham Heart Study (the replication panel). We assessed the clinical relevance of our findings by investigating the identified sites in whole blood and lung tissue specific gene expression. Finally, we performed an exploratory epigenome-wide analysis, combining both cohorts.

Results: We identified and replicated two CpG sites (cg05575921 and cg05951221) that were significantly associated with DLCO/VA and one (cg05575921) suggestively associated with DLCO. Furthermore, we found a positive association between *AHRR* (cg05575921) hypomethylation and gene expression of *EXOC3* in whole blood. We confirmed that the expression of *EXOC3* is positively associated with DLCO/VA and DLCO in lung tissue. Finally, several genome-wide associations were identified after meta-analysing the data of both cohorts.

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

INTRODUCTION

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

DLCO and DLCO/VA are also heritable traits, and genetic variation contributes to the variation of these lung function tests (Terzikhan et al). Although genetics play an important role, epigenetic mechanisms such as DNA methylation are important for regulation of gene expression but also essential to understand the interplay between genes, environment and disease (3). Like DLCO and DLCO/VA, DNA methylation is also influenced by smoking (4, 5); however, the question remains whether DNA methylation is associated with gas exchange independently of smoking.

Little is known about the effects of DNA methylation on lung function (6-8) and epigenome wide association studies (EWAS) of pulmonary gas exchange are lacking. Therefore, we aimed to perform an EWAS to investigate which epigenetic variants are related to the phenotypic variation in DLCO and DLCO/VA. In addition, we examined the impact of the discovered epigenetic variants on gene expression. Finally, we combined the data of both cohorts in an exploratory epigenome-wide analysis.

METHODS

Study populations

This EWAS study encompassed a discovery study embedded within the Rotterdam Study and a replication study embedded within the Framingham Heart Study. The Rotterdam Study is an ongoing prospective population-based cohort study in Rotterdam, the Netherlands. The design has been previously described. Briefly, the Rotterdam Study includes four sub-cohorts. For this study, data from the third visit of the second sub-cohort (RSII-3) and the second visit of the third sub-cohort (RSIII-2) was used. The discovery panel consisted of 650 participants from a random subset of 747 individuals of European descent with methylation and diffusing capacity data available.

Replication of the identified CpG sites was performed in 2,114 individuals form the Framingham Heart Study Offspring cohort from which methylation and diffusing capacity data was available. The design of the Framingham Heart Study has been

described extensively before (9). The replication analyses were focused on Offspring cohort participants of European descent who attended the eighth exam (2005-2008). We only included participants over 45 years of age, with methylation and with blood cell counts data available.

DNA methylation

DNA was extracted from whole peripheral blood (stored in EDTA tubes) by standardised salting out methods. Genome-wide DNA methylation levels were measured using the Illumina Human Methylation 450K array (Illumina, San Diego, CA, USA). In summary, samples (500 ng of DNA per sample) were first treated with bisulfite using the Zymo EZ-96 DNA-methylation kit (Zymo Research, Irvine, CA, USA). Subsequently, samples were hybridised to the arrays according to the standardized protocols. The methylation percentage of a CpG site was presented as a β value ranging between 0 (no methylation) to 1 (full methylation). Data processing was performed in the Genetic Laboratory of Internal Medicine, Erasmus University Medical Centre, Rotterdam. Quality control was performed using Genome Studio (v2011.1, methylation module version 1.9.0; Illumina, USA).

For each probe, individuals with methylation levels higher than three times the inter-quartile range (IQR) were excluded. Finally, we excluded cross reactive and polymorphic probes, or probes that have an underlying SNP at the CpG site, in addition to probes within 10 bp of the single base extension (minor allele frequency >1% in European ancestry (EUR) 1000 genomes project data) (10). In total, 363,387 CpGs were included.

Diffusing capacity of the lung: DLCO and DLCO/VA

DLCO (mmol/min/kPA) and alveolar volume (VA) in liter were measured by the single breath technique in accordance with ERS / ATS guidelines (11). The DLCO per alveolar volume (DLCO/VA; mmol/min/kPA/liter) was calculated by dividing the DLCO by VA. Analyses were restricted to participants with two interpretable and reproducible measurements of DLCO and DLCO/VA. Two measurements were considered reproducible if the difference between the first and second DLCO measurement ((highest-lowest value)/highest value) was ≤10% and the difference between the first and second DLCO/VA measurement was ≤15%.

Covariates

Covariates were selected based on known association with DNA methylation and diffusing capacity and included age, sex, smoking status, weight, height and batch effects; array number and position.

Functional analysis

We used data from the BBMRI atlas (12) (**see URLs**) to identify methylation-gene expression associations; the so called expression quantitative trait methylation (eQTM).

Gene Expression in lung tissue

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

RNA was extracted with the miRNeasy Mini kit (Qiagen) from total lung tissue blocks submersed in RNA-later. cDNA was obtained by the miScript II RT kit (Qiagen), following manufacturer's instructions. Expression of target genes ADGRG6 (GPR126) and reference genes Glyceraldehyde-3-phosphate dehydrogenase (GAPDH), Hypoxanthine phosphoribosyltransferase-1 (HPRT-1) and Succinate Dehydrogenase Complex Flavoprotein Subunit A (SDHA) were analyzed using Taqman Gene Expression Assays (Applied Biosystems, Forster City, CA, USA). Real-time PCR reactions were set up in duplicate using diluted cDNA using identical amplification conditions for each of the target and reference genes. A standard curve derived from serial dilutions of a mixture of all samples were included in each run. The amplification conditions consisted of: 10 minutes at 95°C and 60 cycles of 95°C for 10 seconds and 60°C for 15 seconds. Amplifications were performed using a LightCycler 96 detection system (Roche). Data were processed using the standard curve method. Expression of target genes was corrected by a normalization factor that was calculated based on the expression of three reference genes, using the geNorm applet according to the guidelines and theoretical framework previously described (13).

Statistical analysis

Methylation probes were tested for association with DLCO or DLCO/VA using a linear regression model in the Rotterdam Study and linear mixed model in the Framingham Heart Study. Surrogate variable analysis was used to adjust for inflation in the effect estimates by batch effect in the Framingham Heart Study. The first model (Model 1) included age, sex, smoking, white blood cells, array number and position. In the second model (Model 2), we additionally adjusted for weight and height. False discovery rate

of 5% was used to correct for multiple testing (FDR < 0.05 was considered statistically significant). Inverse variance weighted meta-analysis was performed using METAL (14) (see URL). We test the association between mRNA expression of *AHRR* and *EXOC3* genes and gas exchange in a linear regression adjusting for all covariates used in Model 2. For these analyses, outliers (mean ±3 SD) were excluded and mRNA levels were log transformed to make sure the normality assumption was fulfilled. A p-value lower than 0.05 was considered statistically significant.

RESULTS

General characteristics

The discovery set consisted of 659 participants from the Rotterdam Study. Replication was performed in a set of 2,114 participants from the Framingham Heart Study. The general characteristics of the study populations are shown in **Table 1**. The mean age (SD) was 67.4 (5.9) in the Rotterdam Study and 65.6 (8.4) in the Framingham Heart Study. Mean levels of DLCO and DLOC/VA were similar in both study populations.

Table 1 General characteristics of the discovery and validation study populations

	RS		FHS	
N total	659		2,114	
Age (yrs), mean (SD)	67.4	(5.90)	65.6	(8.38)
Female, N (%)	369	(56%)	1159	(55%)
Current smokers, N (%)	70	(11%)	151	(7%)
Past smokers, N (%)	379	(58%)	1,186	(56%)
Weight (kg), mean (SD)	79.9	(13.94)	79.5	(17.83)
Height (cm), mean (SD)	169.8	(9.16)	167.3	(9.54)
DLCOc (mmol/min/kPA), mean (SD)	7.81	(1.64)	7.47	(2.13)
DLCOc/VA (mmol/min/kPA/VA), mean (SD)	1.52	(0.23)	1.42	(0.23)

DLCO: Diffusing capacity of the lung for carbon monoxide; DLCO/VA: Diffusing capacity of the lung for carbon monoxide by alveolar volume; FHS: The Framingham Heart Study; RS: The Rotterdam Study.
Values are means (standard deviation (SD)) for continuous variables or counts (percentages %) for dichotomous variables.

Discovery and replication

Epigenomewide association study (EWAS) of DLCO

No statistically significant associations were found between DNA methylation and DLCO in model 1 adjusted for age, sex, smoking status, white blood cell counts and

batch effects, nor in model 2 additionally adjusted for weight and height. However, cg05575921 (Gene: *AHRR*, chromosome: 5) was suggestively associated with DLCO in both models (FDR_{model1}=0.07, FDR_{model2}= 0.09).

EWAS of DLCO/VA

In model 1, we identified one CpG site associated with DLCO/VA; cg05575921 (Gene: *AHRR*, chromosome: 5, FDR: 0.017). In model 2, we identified two CpG sites associated with DLCO/VA; cg05575921 (Gene: *AHRR*, chromosome: 5, FDR: 0.047) and cg05951221 (Gene: chr2:233283397-233285959, chromosome: 2, FDR: 0.047).

Finally, we replicated these findings using data from the Framingham Heart Study. The statistically significant associations between DNA methylation and DLCO/VA are summarized in **Table 2**.

Table 2 Epigenome-wide associations between genome-wide DNA-methylation and DLCO/VA

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

B: effect estimate; Chr: Chromosome; DLCO/VA: Diffusing capacity of the lung for carbon monoxide by alveolar volume; FDR: false discovery rate; FHS: The Framingham Heart Study; M1: Adjusted for age, sex, white blood cell count and batch effect; M2: adjusted for age, sex, current smoking, former smoking, weight, height, white blood cell count and batch effect; Pos: position; RS: The Rotterdam Study; SE: Standard error.

Functional analysis

We used BBMRI-atlas to identify eQTMs using the identified DLCO/VA associated CpG sites (cg05575921 and cg05951221). Cg05575921 (*AHRR* gene region) was significantly associated with the expression of the *EXOC3* gene (β= 0.15 (SE=0.039), FDR= 3.90E-4). No eQTM was found for Cg05951221.

Gene Expression in lung tissue

Based on the results of the previous section, we were interested in the association between *EXOC3* and DLCO or DLCO/VA independent of smoking. Because AHRR is strongly associated with smoking, we also investigated the association between AHRR and DLCO or DLCO/VA after adjustment for smoking. Therefore, mRNA was extracted from lung resection specimens of 92 patients who underwent surgery for lung transplantation or solitary pulmonary tumours, including 44 patients without COPD

and 48 patients with COPD; and mRNA expression of *AHRR* and *EXOC3* in lung tissue was examined using quantitative RT-PCR. See **Table 3** for the general characteristics of this study population.

Table 3 Characteristics of study individuals for lung mRNA analysis (by RT-PCR) (n=92)

	Never smokers	Smokers without COPD	COPD GOLD II	COPD GOLD III-IV
Number	18	26	34	14
Gender ratio (m/f)	6/12 §	19/7 §	31/3 §	8/6 §
Age (years)	65 (56-70)	63 (55-70)	66 (58-69) ‡	56 (54-60)*† ‡
Current- / ex-smoker	NA	16/10	22/12	0/14
Pack years of smoking	NA	28 (15-45)*	45 (40-60)* ‡	30 (25-30)*† ‡

Data are presented as median (IQR)
Mann-Whitney U test: * $P < 0.05$ versus never smokers; † $P < 0.05$ versus COPD GOLD II; ‡ $P < 0.05$ versus smokers without COPD; Fisher's exact test: § $P < 0.001$

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

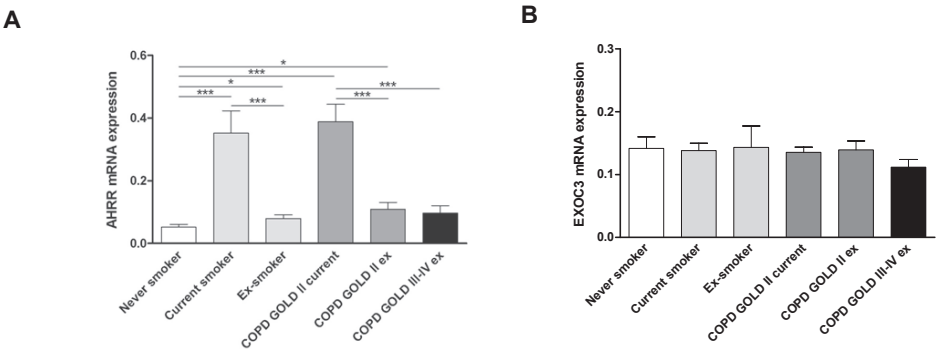

Figure 1 human lung tissue mRNA expression of the **A)** *AHRR* gene and **B)** *EXOC3* gene stratified by smoking and COPD status.
AHRR: Aryl Hydrocarbon Receptor Repressor gene; COPD: Chronic obstructive pulmonary disease; ex: ex-smoker; EXOC3: Exocyst Complex Component 3 gene. GOLD: Global initiative for chronic obstructive lung disease;

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

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

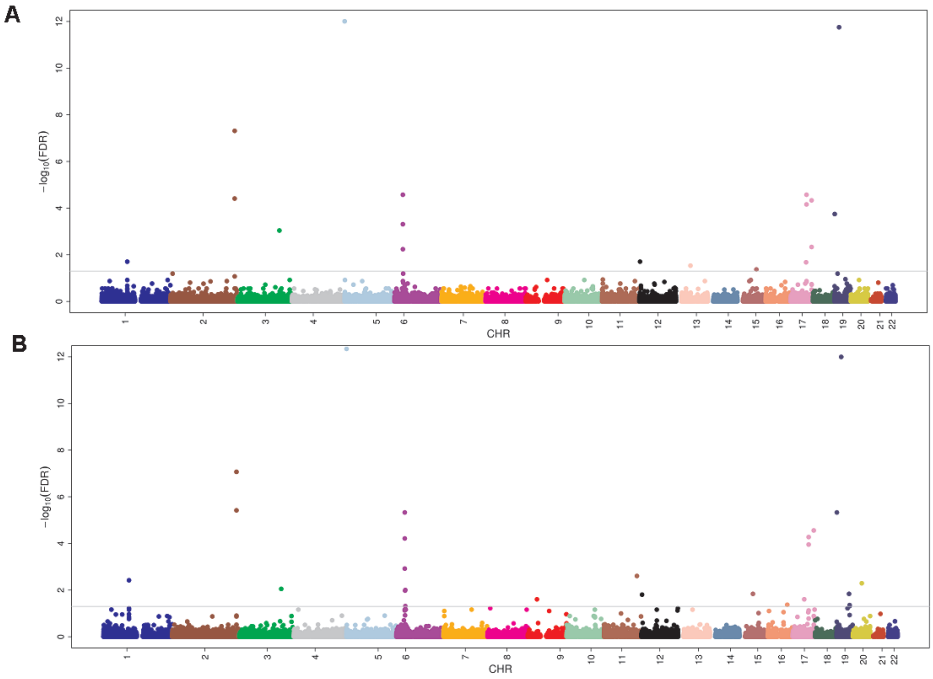

Figure 2 Manhattan plots showing the results of EWAS meta-analysis (RS and FHS) of the association between genome-wide DNA methylation and DLCO . **A)** Model1: adjusted for age, sex, smoking status, white blood cell counts and batch effects. **B)** Model2: adjusted for weight and height in addition to Model1 .

Meta-analysis

We meta-analysed the results of the analyses from the Rotterdam Study and the Framingham Heart Study for both phenotypes (DLCO and DLCO/VA) and both models. The results are summarized in the corresponding EWAS Manhattan plots in **Figures 3** and **4**, and in the supplementary tables (**Tables S1-4**). We identified 18 CpG sites which were associated with DLCO in Model 1 and 26 CpG sites in Model 2. For DLCO/VA, we identified 9 CpG sites in Model 1 and 11 CpG sites in Model 2.

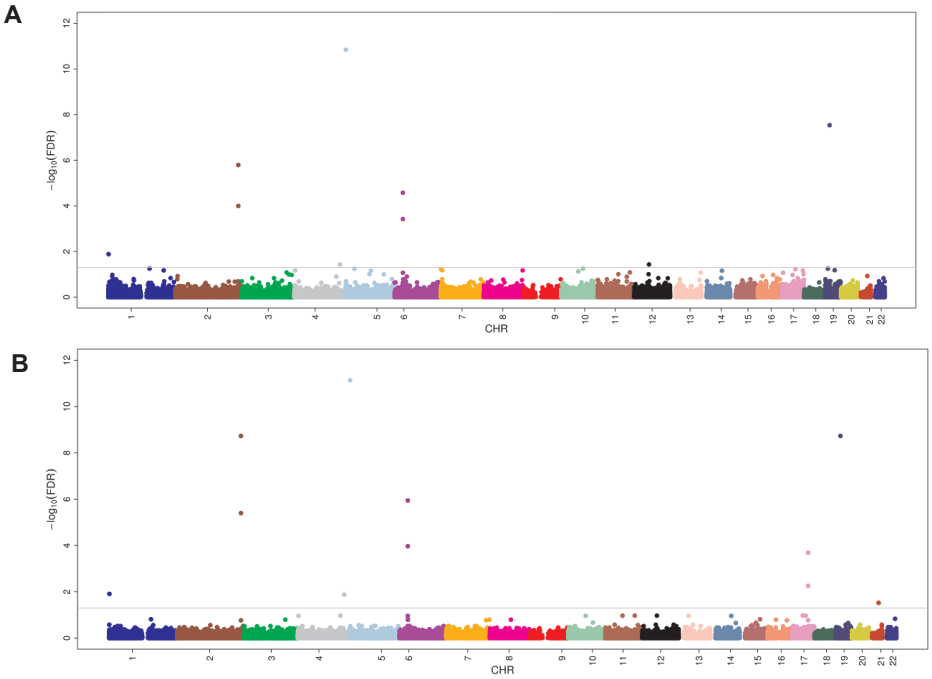

Figure 3 Manhattan plots showing the results of EWAS meta-analysis (RS and FHS) of the association between genome-wide DNA methylation and DLCO/VA. A) Model1: adjusted for age, sex, smoking status, white blood cell counts and batch effects. B) Model2: adjusted for weight and height in addition to Model1.

DISCUSSION

In this EWAS study, we investigated the epigenetic signature of DLCO and DLCO/VA as a measure of gas exchange in the lung. We observed two CpG sites with epigenome-wide associations in the Rotterdam Study, and those were replicated in the Framingham Heart Study. Additional analysis revealed a potential pathway trough which methylation of AHRR might influence gas exchange. Finally, meta-analysis of the two studies revealed many more methylation signals that need to be confirmed in future studies.

This is the first EWAS on diffusing capacity of carbon monoxide uptake in the lungs. Few studies investigated the effect of DNA methylation on lung function, and were restricted to spirometric measures (FEV_1 and FVC). Bell and colleagues (6) investigated age-related phenotypes including FEV_1 and FVC in 172 twin females. Authors reported one association at chromosome 11 in the *WT1* gene region. Another study by Marioni *et al.* (7) performed an EWAS on FEV_1 in 920 individuals and did not find epigenome-wide associations. Qui el al. performed EWAS on FEV1, FVC and COPD and found

many associations, but the analyses were not adjusted for differences in white blood cell count.

In this EWAS, we discovered and confirmed two epigenome-wide associations, one CpG site in the aryl hydrocarbon receptor repressor gene (*AHRR)* which encodes a repressor protein of the aryl hydrocarbon receptor (AHR), and the second CpG site on chromosome 2 (2q37.1). These CpG sites are well described as being strongly associated with smoking behaviour (4, 5, 15). Cg05575921 is located in the AHRR gene region and is involved in the metabolism of smoking-released chemicals, where the AHRR gene supresses the function of the AHR gene –which is responsible for the regulation of smoking related substances- through a negative feedback loop (16). Although the interaction between the genetics and DNA methylation is complex, it is believed that *AHRR* hypomethylation inhibits the translation of the gene by preventing transcription factors from binding to the promotor regions (17, 18). We hypothesise that *AHRR* hypomethylation by smoking might intervene with the elimination of the smoking related substances from the body.

In addition to the association with smoking, *AHRR* has been also recently associated with impaired lung function. Bojesen et al. (15) showed that hypomethylation of cg05575921 was associated with smoking related phenotypes such as COPD, COPD exacerbations and lung cancer. Similarly, Kodal et. al. (19) found that hypomethylation of cg05575921 was associated with low lung function, steeper lung function decline and respiratory symptoms. Whether these associations are independent of smoking effects or confounded by smoking and consequently are the results of the effect of smoking on DNA methylation, remains to be investigated.

A recent study by Li and colleagues, proved the causal effect of smoking on DNA methylation. This information eliminates the possibility for smoking to be an intermediate in the association between DNA methylation of smoking related probes and lung function. Alternatively, smoking can be a confounding factor for which we can adjust in our models. However, residual confounding is still possible as assessment of smoking status is self-reported (20). Therefore the question remains, does residual confounding by smoking explain the whole effect in the association between smoking related probes and lung function? Or is there any significant effect of DNA methylation of those smoking related probes independent of smoking? Both scenarios lead to very interesting hypotheses and might have important clinical implications. In case smoking explains the entire association, the hypothesis might be, as proposed by Li et al. (16) and Kodal et al. (19), that the association between hypomethylation of smoking related probes and a decreased lung function might reflect the smoking induced damage to the lungs through methylation of the smoking related probes. In other words, DNA methylation might act as an intermediate in the effect of smoking on smoking related diseases. The second scenario would be, that smoking affects DNA methylation, but

the effect of the hypomethylation of the smoking related probes on lung function is (partly) independent from smoking. In that case, hypomethylation of the smoking related probes might provide clinically relevant information beyond the effect of smoking on lung function. Unfortunately, it is difficult to investigate these hypotheses in cross-sectional observational studies. An alternative approach might be to study the associations between smoking related probes and lung function in never smokers. However, finding any association in never smokers is still no prove that smoking related probes are independently associated with lung function, as passive smoking or the unmeasured effects of air pollution still might bias these associations.

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

Exocyst complex component 3 (EXOC3), previously known as SEC6, is located on chromosome 5 upstream of AHRR gene and downstream of SLC9A3 gene. EXOC3 is part of the exocyst protein complex and the protein encoded by it, is involved in post Golgi trafficking and essential for biogenesis of epithelial cell surface polarity (21-23). Variation in EXOC3 is linked to variability of cystic fibrosis(24, 25). This makes EXOC3 an interesting candidate to elucidate its role in membrane pathology and gas exchange.

Strength and limitations

The strength of this study is the population-based setting with standardized data collection. Also, this study provides unique data, since we are not aware of other population-based studies with data on gas exchange. Finally, we provided additional results of gene expression analysis on lung tissue, which was performed in well-defined patient groups. However, this study also has some limitations. First, our discovery panel is relatively small, bigger sample sizes may help in the identification of more CpG sites that might give us more insight in pathology of gas exchange. Second, although we were able to replicate findings of the lead association, we cannot exclude the possibility of unmeasured confounding by smoking and air pollution. Third, this study was performed with DNA methylation data in whole blood, future studies should ideally consider performing tissue specific DNA methylation analyses. Finally, the current study analyses the data in a cross-sectional manner. As the epigenome changes over time, longitudinal data-analyses might be more informative about the epigenetic modifications over time, as it may play a crucial role in the aetiology of a decreased gas exchange.

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

REFERENCES

1. Boutou AK, Shrikrishna D, Tanner RJ, Smith C, Kelly JL, Ward SP, et al. Lung function indices for predicting mortality in COPD. Eur Respir J. 2013;42(3):616-25.
2. Hegewald MJ. Diffusing capacity. Clin Rev Allergy Immunol. 2009;37(3):159-66.
3. Shah S, McRae AF, Marioni RE, Harris SE, Gibson J, Henders AK, et al. Genetic and environmental exposures constrain epigenetic drift over the human life course. Genome Res. 2014;24(11):1725-33.
4. Shenker NS, Polidoro S, van Veldhoven K, Sacerdote C, Ricceri F, Birrell MA, et al. Epigenome-wide association study in the European Prospective Investigation into Cancer and Nutrition (EPIC-Turin) identifies novel genetic loci associated with smoking. Hum Mol Genet. 2013;22(5):843-51.
5. Zeilinger S, Kuhnel B, Klopp N, Baurecht H, Kleinschmidt A, Gieger C, et al. Tobacco smoking leads to extensive genome-wide changes in DNA methylation. PLoS One. 2013;8(5):e63812.
6. Bell JT, Tsai PC, Yang TP, Pidsley R, Nisbet J, Glass D, et al. Epigenome-wide scans identify differentially methylated regions for age and age-related phenotypes in a healthy ageing population. PLoS Genet. 2012;8(4):e1002629.
7. Marioni RE, Shah S, McRae AF, Ritchie SJ, Muniz-Terrera G, Harris SE, et al. The epigenetic clock is correlated with physical and cognitive fitness in the Lothian Birth Cohort 1936. Int J Epidemiol. 2015;44(4):1388-96.
8. 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.
9. Kannel WB, Feinleib M, McNamara PM, Garrison RJ, Castelli WP. An investigation of coronary heart disease in families. The Framingham offspring study. Am J Epidemiol. 1979;110(3):281-90.
10. Chen YA, Lemire M, Choufani S, Butcher DT, Grafodatskaya D, Zanke BW, et al. Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray. Epigenetics. 2013;8(2):203-9.
11. Macintyre N, Crapo RO, Viegi G, Johnson DC, van der Grinten CP, Brusasco V, et al. Standardisation of the single-breath determination of carbon monoxide uptake in the lung. Eur Respir J. 2005;26(4):720-35.
12. Bonder MJ, Luijk R, Zhernakova DV, Moed M, Deelen P, Vermaat M, et al. Disease variants alter transcription factor levels and methylation of their binding sites. Nature Genetics. 2017;49(1):131-8.
13. Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, De Paepe A, et al. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biol. 2002;3(7):RESEARCH0034.
14. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics. 2010;26(17):2190-1.
15. Bojesen SE, Timpson N, Relton C, Davey Smith G, Nordestgaard BG. AHRR (cg05575921) hypomethylation marks smoking behaviour, morbidity and mortality. Thorax. 2017;72(7):646-53.
16. Li S, Wong EM, Bui M, Nguyen TL, Joo JE, Stone J, et al. Causal effect of smoking on DNA methylation in peripheral blood: a twin and family study. Clin Epigenetics. 2018;10:18.
17. Bird AP. CpG-rich islands and the function of DNA methylation. Nature. 1986;321(6067):209-13.

18. Machin M, Amaral AF, Wielscher M, Rezwan FI, Imboden M, Jarvelin MR, et al. Systematic review of lung function and COPD with peripheral blood DNA methylation in population based studies. BMC Pulm Med. 2017;17(1):54.
19. Kodal JB, Kobylecki CJ, Vedel-Krogh S, Nordestgaard BG, Bojesen SE. AHRR hypomethylation, lung function, lung function decline and respiratory symptoms. Eur Respir J. 2018;51(3).
20. Shipton D, Tappin DM, Vadiveloo T, Crossley JA, Aitken DA, Chalmers J. Reliability of self reported smoking status by pregnant women for estimating smoking prevalence: a retrospective, cross sectional study. BMJ. 2009;339:b4347.
21. Rodriguez-Boulan E, Macara IG. Organization and execution of the epithelial polarity programme. Nat Rev Mol Cell Biol. 2014;15(4):225-42.
22. Oztan A, Silvis M, Weisz OA, Bradbury NA, Hsu SC, Goldenring JR, et al. Exocyst requirement for endocytic traffic directed toward the apical and basolateral poles of polarized MDCK cells. Mol Biol Cell. 2007;18(10):3978-92.
23. Grindstaff KK, Yeaman C, Anandasabapathy N, Hsu SC, Rodriguez-Boulan E, Scheller RH, et al. Sec6/8 complex is recruited to cell-cell contacts and specifies transport vesicle delivery to the basal-lateral membrane in epithelial cells. Cell. 1998;93(5):731-40.
24. Corvol H, Blackman SM, Boelle PY, Gallins PJ, Pace RG, Stonebraker JR, et al. Genome-wide association meta-analysis identifies five modifier loci of lung disease severity in cystic fibrosis. Nat Commun. 2015;6.
25. Dorfman R. Modifier Gene Studies to Identify New Therapeutic Targets in Cystic Fibrosis. Curr Pharm Design. 2012;18(5):674-82.

3.5

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

ABSTRACT

Background: 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.

Methods: 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.

Results: 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 hyper-responsiveness, 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.

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.

1. INTRODUCTION

Chronic obstructive pulmonary disease (COPD) is a common and complex disease, and one of the leading causes of death worldwide (Lozano R, Naghavi M, 2012). Previous studies provided heritability estimates for COPD of 20% to even 60% (Ingebrigtsen et al., 2010; Zhou et al., 2013). 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 (Bashir, A., Shah, N.N., Hazari, Y.M., Habib, M., Bashir, S., Hilal, N., Banday, M., Asrafuzzaman, S., and Fazili, 2016; Laurell CB and Laurell CB, 1963). 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 (Foreman et al., 2017). In an exome study of severe, early-onset families, Qiao et al identified several genes with rare variants segregating in at least two pedigrees (Qiao et al., 2016). In extended families, genetic linkage studies have found evidence of linkage to chromosomes 2q, 6q, 8p, 12p and 19q, among others (Palmer et al., 2003; Silverman et al., 2002). However, many initially promising findings from linkage or exome sequencing in candidate-gene studies could not be replicated in subsequent analyses (Hersh et al., 2005).

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*) (Hobbs et al., 2017; Wain et al., 2017). 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 (Auer and Lettre, 2015). 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 (Pardo et al., 2005). Therefore, attempting to identify risk genes for COPD in populations that are relatively genetically and environmentally homogeneous could be beneficial (Van Diemen et al., 2010).

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.

2. METHODS

2.1. Study populations

2.1.1. 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 (Pardo et al., 2005). 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 for145 participants, of which 3 did not have genotyping data, resulting in the final sample size for the linkage study of 142 COPD cases.

2.1.2. 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. (Hobbs et al., 2016) (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 (Ikram et al., 2017), 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 (Scholtens et al., 2015). COPD was defined as having pre-bronchodilator FEV1/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 (Van Diemen et al., 2005). 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. (Hobbs et al., 2016) COPD cases were defined as having FEV1/FVC≤0.7 and FEV1≤ 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 (Hobbs et al., 2016).

2.2. Genotyping

2.2.1. DNA isolation

For all participants, DNA was extracted from venous blood using the salting out method (S.A.Miller, 1988).

2.2.2. 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.

2.2.3. Exome-sequencing and genotyping

The sequencing and genotyping in the ERF study have been described elsewhere (Amin et al., 2017). 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 (Iglesias et al., 2017). Haplotype Reference Consortium imputation panel (HRC) (McCarthy et al., 2015) was used for imputation. File preparation and imputation was done as described elsewhere (Iglesias et al., 2017). 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 (De Jong et al., 2015). The Genome of the Netherlands (GoNL) panel was used for imputation of LLS and VlaVla and was done as described elsewhere (Scholtens et al., 2015). 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 (Hobbs et al., 2016).

2.3. Statistical analyses

2.3.1. 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 (Liu et al., 2008). We used Haplopainter (Thiele and Nürnberg, 2005) to illustrate all 27 pedigrees (Supplementary figure 1). We then performed affected-only parametric linkage analysis in MERLIN software (Abecasis et al., 2002) using incomplete penetrance and no phenocopies for both dominant (0, 0.5, 0.5) and recessive models (0, 0, 0.5) (Durner et al., 1999). 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.

2.3.2. 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 (Pardo et al., 2005), 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 (Voorman et al., 2013) 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 (Norušis, 1992) and in LLS, using PLINK (Purcell et al., 2007), 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. (Hobbs et al., 2016). A fixed-effects meta-analysis was performed with the summary statistics from all studies using the "*rmeta*" package in R (Lumley, 2011).

2.3.3. 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) (Ipa, 2015). Furthermore, we consulted the Gene network tool (Fehrmann et al., 2015), 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.

3. RESULTS

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

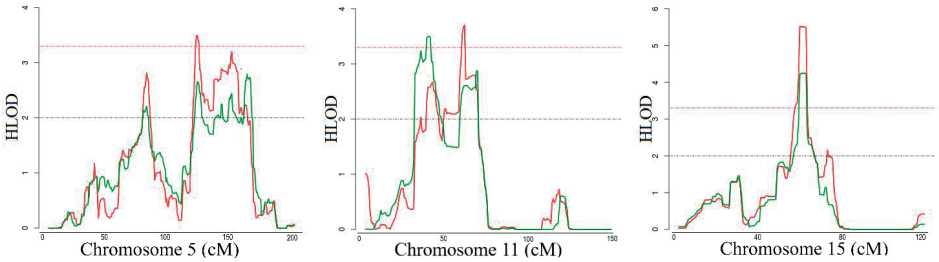

Figure 1. LOD score plot for the regions at (A) chromosomes 5, (B) 11 and (C) 15. X-axis shows the chromosomal position in cM and the Y-axis shows the HLOD score. Red line represents HLOD scores for recessive and green line for dominant model. Dashed red line represents the level of significance (HLOD=3.3),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

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.

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.

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**). 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).

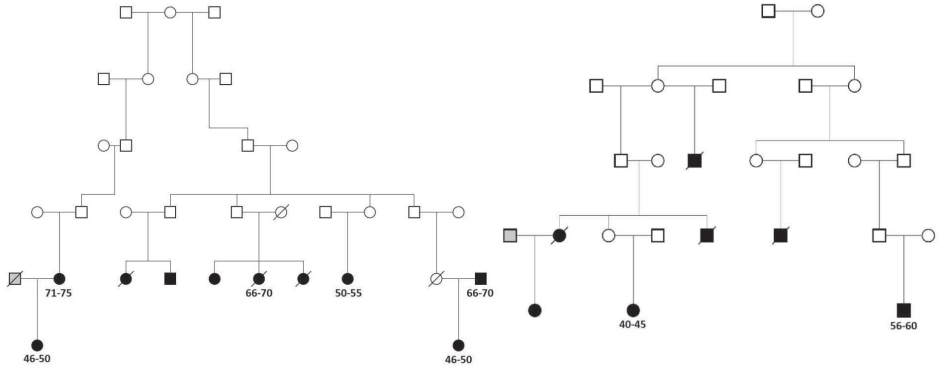

Figure 2. The two sub-families contributing most to the linkage peak on chromosome 11. Squares represent males and circles females. Cases are denoted in black, known controls are denoted in grey and the family members for which we do not have COPD information are denoted in white. Family members with 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.

4. 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 (Thorgeirsson et al., 2008). 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 (Hardin and Silverman, 2014). 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 (van der Plaat et al., 2017), which is in line with the only longitudinal study on the relation between the nicotine receptor variant and annual lung function decline (Budulac et al., 2012). 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 function and airflow obstruction (Hancock et al., 2010; Wilk et al., 2012) was recently associated with COPD by the largest GWAS to date (Hobbs et al., 2017). The *HTR4* gene in 5q32 encodes a serotonin receptor involved in depression and is strongly expressed in respiratory complex neurons (Manzke et al., 2003).

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 (Bosquillon, 2010). *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 (Fehrmann et al., 2015). 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 (Endter et al., 2009). *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 (Bosquillon, 2010).

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 (Olesen et al., 2004). *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 (Fehrmann et al., 2015). 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 (Werynska et al., 2015).

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.

10. REFERENCES

Abecasis, G. R., Cherny, S. S., Cookson, W. O., and Cardon, L. R. (2002). Merlin—rapid analysis of dense genetic maps using sparse gene flow trees. *Nat. Genet.* 30, 97–101. doi:10.1038/ng786.

Amin, N., de Vrij, F. M. S., Baghdadi, M., Brouwer, R. W. W., van Rooij, J. G. J., Jovanova, O., et al. (2017). A rare missense variant in RCL1 segregates with depression in extended families. *Mol. Psychiatry*. doi:10.1038/mp.2017.49.

Auer, P. L., and Lettre, G. (2015). Rare variant association studies: Considerations, challenges and opportunities. *Genome Med.* 7, 16. doi:10.1186/s13073-015-0138-2.

Bashir, A., Shah, N.N., Hazari, Y.M., Habib, M., Bashir, S., Hilal, N., Banday, M., Asrafuzzaman, S., and Fazili, K. M. (2016). Novel variants of SERPIN1A gene: Interplay between alpha1- antitrypsin deficiency and chronic obstructive pulmonary disease. *Respir. Med.* 117, 139–149. doi:https://doi.org/10.1016/j.rmed.2016.06.005.

Bosquillon, C. (2010). Drug Transporters in the Lung—Do They Play a Role in the Biopharmaceutics of Inhaled Drugs? *J. Pharm. Sci.* 99, 4215–4227. doi:10.1002/jps.

Budulac, S. E., Vonk, J. M., Postma, D. S., Siedlinski, M., Timens, W., and Boezen, M. H. (2012). Nicotinic acetylcholine receptor variants are related to smoking habits, but not directly to COPD. *PLoS One* 7, e33386. doi:10.1371/journal.pone.0033386.

De Jong, K., Vonk, J. M., Timens, W., Bossé, Y., Sin, D. D., Hao, K., et al. (2015). Genome-wide interaction study of gene-by-occupational exposure and effects on FEV1levels. *J. Allergy Clin. Immunol.* 136, 1664–1672e14. doi:10.1016/j.jaci.2015.03.042.

Durner, M., Vieland, V. J., and Greenberg, D. A. (1999). Further evidence for the increased power of LOD scores compared with nonparametric methods. *Am J Hum Genet* 64, 281–289. doi:10.1086/302181.

Endter, S., Francombe, D., Gumbleton, M., and Ehrhardt, C. (2009). RT-PCR analysis of ABC, SLC and SLCO drug transporters in human lung epithelial cell models. *J. Pharm. Pharmacol.* 61, 583–591. doi:10.1211/jpp/61.05.0006.

Fehrmann, R. S. N., Karjalainen, J. M., Westra, H., Maloney, D., Simeonov, A., Pers, T. H., et al. (2015). Gene expression analysis identifies global gene dosage sensitivity in cancer. *Nat. Genet.* 47, 115–126. doi:10.1038/ng.3173.

Foreman, M. G., Wilson, C., DeMeo, D. L., Hersh, C. P., Beaty, T. H., Cho, M. H., et al. (2017). Alpha-1 antitrypsin PiMZ genotype is associated with chronic obstructive pulmonary disease in two racial groups. *Ann. Am. Thorac. Soc.* 14, 1280–1287. doi:10.1513/AnnalsATS.201611-838OC.

Hancock, D. B., Eijgelsheim, M., Wilk, J. B., Gharib, S. A., Loehr, L. R., Marciante, K. D., et al. (2010). Meta-analyses of genome-wide association studies identify multiple loci associated with pulmonary function. *Nat. Genet.* 42, 45–52. doi:10.1038/ng.500.

Hardin, M., and Silverman, E. K. (2014). Chronic Obstructive Pulmonary Disease Genetics : A Review of the Past and a Look Into the Future. *J. COPD Found.* 1, 33–46. doi:10.15326/jcopdf.1.1.2014.0120.

Hersh, C. P., DeMeo, D. L., Lange, C., Litonjua, A. A., Reilly, J. J., Kwiatkowski, D., et al. (2005). Attempted replication of reported chronic obstructive pulmonary disease candidate gene associations. *Am. J. Respir. Cell Mol. Biol.* 33, 71–78. doi:10.1165/rcmb.2005-0073OC.

Hobbs, B. D., De Jong, K., Lamontagne, M., Bossé, Y., Shrine, N., Artigas, M. S., 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. doi:10.1038/ng.3752.

Hobbs, B. D., Parker, M. M., Chen, H., Lao, T., Hardin, M., Qiao, D., et al. (2016). Exome array analysis identifies a common Variant in IL27 associated with chronic obstructive pulmonary disease. *Am. J. Respir. Crit. Care Med.* 194, 48–57. doi:10.1164/rccm.201510-2053OC.

Iglesias, A. I., van der Lee, S. J., Bonnemaijer, P. W. M., Höhn, R., Nag, A., Gharahkhani, P., et al. (2017). Haplotype reference consortium panel: Practical implications of imputations with large reference panels. *Hum. Mutat.* 38, 1025–1032. doi:10.1002/humu.23247.

Ikram, M. A., Brusselle, G. G. O., Murad, S. D., van Duijn, C. M., Franco, O. H., Goedegebure, A., et al. (2017). The Rotterdam Study: 2018 update on objectives, design and main results. *Eur. J. Epidemiol.* 32, 807–850. doi:10.1007/s10654-017-0321-4.

Ingebrigtsen, T., Thomsen, S. F., Vestbo, J., van der Sluis, S., Kyvik, K. O., Silverman, E. K., et al. (2010). Genetic influences on Chronic Obstructive Pulmonary Disease. *a Twin Study* 104, 1890–1895. doi:10.1016/j.rmed.2010.05.004.

Ipa, Q. (2015). Redwood City, CA. *CA Available online http//www.ingenuity.com/products/ipa*.

Laurell CB, E. S., and Laurell CB, E. S. (1963). The electrophoretic alpha 1-globulin pattern of serum in alpha 1-antitrypsin deficiency. *Scand J Clin Lab Invest* 15, 132–40. doi:https://doi.org/10.1080/00365516309051324.

Liu, F., Kirichenko, A., Axenovich, T. I., van Duijn, C. M., and Aulchenko, Y. S. (2008). An approach for cutting large and complex pedigrees for linkage analysis. *Eur. J. Hum. Genet.* 16, 854–860. doi:10.1038/ejhg.2008.24.

Lozano R, Naghavi M, F. K. (2012). 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* 380, 2095–2128. doi:https://doi.org/10.1016/S0140-6736(12)61728-0.

Lumley, T. (2011). rmeta: Meta-analysis. R package version 2.16. 2009. *Cited on*.

Manzke, T., Guenther, U., Ponimaskin, E. G., Haller, M., Dutschmann, M., Schwarzacher, S., et al. (2003). 5-HT4(a) receptors avert opioid-induced breathing depression without loss of analgesia. *Science (80-.).* 301, 226–229. doi:10.1126/science.1084674.

McCarthy, S., Das, S., Kretzschmar, W., Durbin, R., Abecasis, G., and Marchini, J. (2015). A reference panel of 64,976 haplotypes for genotype imputation. *bioRxiv* 48, 35170. doi:10.1101/035170.

Norušis, M. J. (1992). *SPSS/PC+ professional statistics version 5.0*. SPSS, Inc.

Olesen, C., Moller, M., and Byskov, A. G. (2004). Tesmin Transcription is Regulated Differently during Male and Female Meiosis. *Mol. Reprod. Dev.* 67, 116–126. doi:10.1002/mrd.20007.

Palmer, L. J., Celedón, J. C., Chapman, H. A., Speizer E., F. E., Weiss, S. T., and Silverman, E. K. (2003). Genome-wide linkage analysis of bronchodilator responsiveness and post-bronchodilator spirometric phenotypes in chronic obstructive pulmonary disease. *Hum. Mol. Genet.* 12, 1199–1210. doi:10.1093/hmg/ddg125.

Pardo, L. M., MacKay, I., Oostra, B., van Duijn, C. M., and Aulchenko, Y. S. (2005). The effect of genetic drift in a young genetically isolated population. *Ann. Hum. Genet.* 69, 288–295. doi:10.1046/j.1529-8817.2005.00162.x.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., et al. (2007). PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *Am. J. Hum. Genet.* 81, 559–575. doi:10.1086/519795.

Qiao, D., Lange, C., Beaty, T. H., Crapo, J. D., Barnes, K. C., and Bamshad, M. (2016). Exome Sequencing Analysis in Severe , Early-Onset Chronic Obstructive Pulmonary Disease. *Am. J. Respir. Crit. Care Med.* 193, 1–61. doi:https://doi.org/10.1164/rccm.201506-1223OC.

S.A.Miller, D. D. D. and H. F. P. (1988). A simple salting out procedure for extractin DNA from humam nnucleated cells. *Nucleic Acids Res.* 15, 1215.

Scholtens, S., Smidt, N., Swertz, M. A., Bakker, S. J. L., Dotinga, A., Vonk, J. M., et al. (2015). Cohort Profile: LifeLines, a three-generation cohort study and biobank. *Int. J. Epidemiol.* 44, 1172–1180. doi:10.1093/ije/dyu229.

Silverman, E. K., Mosley, J. D., Palmer, L. J., Barth, M., Senter, J. M., Brown, A., et al. (2002). Genome-wide linkage analysis of severe , early-onset chronic obstructive pulmonary disease : airflow obstruction and chronic bronchitis phenotypes. *Hum. Mol. Genet.* 11, 623–632. Available at: http://dx.doi.org/10.1093/hmg/11.6.623.

Thiele, H., and Nürnberg, P. (2005). HaploPainter: A tool for drawing pedigrees with complex haplotypes. *Bioinformatics* 21, 1730–1732. doi:10.1093/bioinformatics/bth488.

Thorgeirsson, T. E., Geller, F., Sulem1, P., Rafnar1, T., Wiste, A., Magnusson, K. P., et al. (2008). A variant associated with nicotine dependence, lung cancer and peripheral arterial disease. *Nat.* 452, 638–642. doi:10.1038/nature06846.

van der Plaat, D. A., de Jong, K., Lahousse, L., Faiz, A., Vonk, J. M., van Diemen, C. C., et al. (2017). Genome-wide association study on the FEV1/FVC ratio in never-smokers identifies HHIP and FAM13A. *J. Allergy Clin. Immunol.* 139, 533–540. doi:10.1016/j.jaci.2016.06.062.

Van Diemen, C. C., Postma, D. S., Aulchenko, Y. S., Snijders, P. J. L. M., Oostra, B. A., Van Duijn, C. M., et al. (2010). Novel strategy to identify genetic risk factors for COPD severity: A genetic isolate. *Eur. Respir. J.* 35, 768–775. doi:10.1183/09031936.00054408.

Van Diemen, C. C., Postma, D. S., Vonk, J. M., Bruinenberg, M., Scheuten, J. P., and Boezen, H. M. (2005). A disintegrin and metalloprotease 33 polymorphisms and lung function decline in the general population. *Am. J. Respir. Crit. Care Med.* 172, 329–333. doi:10.1164/rccm.200411-1486OC.

Voorman, A., Brody, J., and Lumley, T. (2013). skatMeta: an R Package for meta analyzing region-based tests of rare DNA variants June. *R Packag. version* 1, 1–17.

Wain, L. V., Shrine, N., Artigas, M. S., Erzurumluoglu, A. M., Noyvert, B., Bossini-Castillo, L., et al. (2017). Supplementary: Genome-wide association analyses for lung function and chronic obstructive pulmonary disease identify new loci and potential druggable targets. *Nat. Genet.* 49, 416–425. doi:10.1038/ng.3787.

Werynska, B., Pula, B., Kobierzycki, C., Dziegiel, P., and Podhorska-Okolow, M. (2015). Metallothioneins in the lung cancer. *Folia Histochem. Cytobiol.* 53, 1–10. doi:10.5603/FHC.a2015.0009.

Wilk, J. B., Shrine, N. R. G., Loehr, L. R., Zhao, J. H., Manichaikul, A., Lopez, L. M., et al. (2012). Genome-wide association studies identify CHRNA5/3 and HTR4 in the development of airflow obstruction. *Am. J. Respir. Crit. Care Med.* 186, 622–632. doi:10.1164/rccm.201202-0366OC.

Zhou, J. J., Cho, M. H., Castaldi, P. J., Hersh, C. P., Silverman, E. K., and Laird, N. M. (2013). Heritability of Chronic Obstructive Pulmonary Disease and Related Phenotypes in Smokers. *Am. J. Respir. Crit. Care Med.* 188, 941–947. doi:10.1164/rccm.201302-0263OC [doi].

Clinical Epidemiology

4

4.1

Pulmonary artery to aorta ratio and risk of all-cause mortality in the general population: the Rotterdam Study

ABSTRACT

Background: The pulmonary artery to aorta ratio (PA:A)>1 is a proxy of pulmonary hypertension. It is unknown whether this measure carries prognostic information in the general population and in individuals with COPD.

Methods: Between 2003-2006, 2,197 participants from the population-based Rotterdam Study (mean±SD age 69.7±6.7 years, 51.3% women), underwent cardiac CT scan with PA:A quantification, defined as the ratio between the pulmonary artery and aorta diameters. COPD was diagnosed based on spirometry or clinical presentation and obstructive lung function measured by a treating physician. Cox regression was used to investigate the risk of mortality.

Results: We observed no association between 1-SD increase of PA:A and mortality in the general population. Larger PA:A was associated with an increased risk of mortality in individuals with COPD, particularly in moderate to severe COPD (HR=1.36, 95%CI=1.03-1.79). We demonstrated that the risk of mortality in COPD was driven by severe COPD and that this risk increased with decreasing diffusing capacity.

Conclusion: Larger PA:A is not associated with mortality in an older general population, but is an independent determinant of mortality in moderate to severe COPD. Measuring PA:A in CT scans obtained for other indications may yield important prognostic information in individuals with COPD.

Key words: mortality, COPD, pulmonary hypertension, epidemiology, computed tomography, population-based

INTRODUCTION

Chronic obstructive pulmonary disease (COPD) is a common cause of death worldwide. In Europe, approximately 18 per 100,000 individuals die due to COPD each year [1]. The high mortality-rate of COPD stresses the need for early identification of COPD patients at high risk of adverse outcomes, in order to guide and improve patient management.

One of the complications in the advanced stages of COPD, which is known to contribute to mortality, is pulmonary hypertension (PH). Currently, the gold standard for diagnosis of PH is an invasive pressure measurement in the pulmonary artery [2]. Alternatively, PH may be assessed using echocardiography. However, this method has proven to be often inconclusive in the advanced stages of COPD, particularly in those persons with emphysema or obesity, due to air and adipose tissue obscuring echocardiographic examination [3, 4].

An alternative, and more novel suggested approach to assess PH, is to measure the ratio between the diameter of the pulmonary artery and that of the aorta (PA:A) using computed tomography. Given its correlation with the mean pulmonary artery pressure, a PA:A larger than 1 is suggested to be a reliable indicator of PH, and may thus directly indicate a worse clinical outcome [5]. Indeed, in patients with COPD, PA:A ratios larger than 1 are related to an increased rate of severe exacerbations requiring hospitalization [6]. However, other than its relationship with exacerbations, the implications of larger PA:A remain unclear. In addition, large-scale population-based data on the utility of PA:A with regard to clinical end points in the general population and specifically in individuals with COPD are lacking. Such data may contribute to the development of targeted additional therapeutic or preventive strategies for exacerbations and mortality in COPD patients.

Therefore, in a population-based setting, we investigated the association between larger PA:A and mortality, with a specific focus on individuals with COPD.

METHODS

Setting

The present study was embedded within the Rotterdam Study, an ongoing prospective population-based cohort study aimed at investigating the occurrence and risk factors of chronic diseases in the general population. The objective and methods of the Rotterdam Study have been published in great detail previously [7]. Briefly, the Rotterdam Study includes 3 cohorts encompassing 14,926 participants aged ≥ 45 years, living in Ommoord, a well-defined suburb of the city of Rotterdam, the Netherlands. Baseline data were collected between 1990 and 1993 (n = 7,983), between 2000 and 2003 (n = 3,011), and between 2006 and 2008 (n = 3,932); thereafter, examinations

have been conducted every 4 to 5 years in all cohorts. Between 2003 and 2006, all participants that visited the research center were invited to undergo multi-detector computed tomography (MDCT)-study as part of a large project on vascular calcification (n = 2,524). The cardiac scan that was performed in this protocol was used for the assessment of PA:A. Therefore, this visit represents the baseline for the current analyses. Figure 1 shows the study flow of participants that were included in this study.

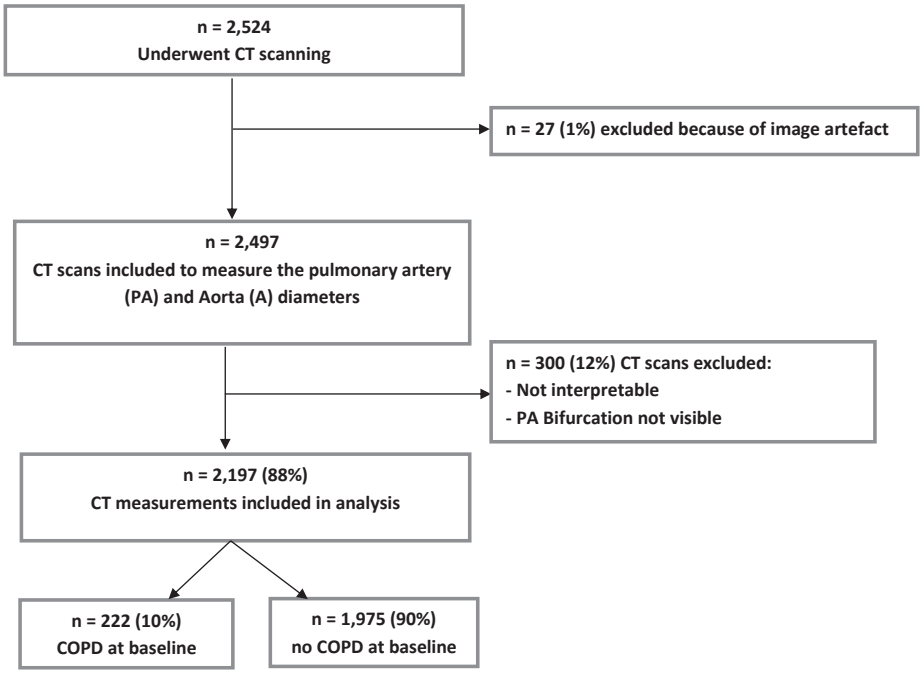

Figure 1 Flow chart of participants in the study.
COPD: chronic obstructive pulmonary disease, CT: computed tomography, n= number of individuals.

Follow-up for mortality

Information on mortality of study participants was obtained from the local municipality in Rotterdam, and additionally validated with data from medical records kept by the general practitioner's, as described in detail previously [8]. Mortality data was complete until March 2015.

Assessment of PA:A

Non-contrast CT-scanning was performed with a 16-slice (n = 785) or 64-slice (n = 1,739) MDCT scanner (Somatom Sensation 16/64, Siemens, Forchheim, Germany). An ECG-gated cardiac imaging protocol was used to visualize the heart and the proximal

part of the great vessels, including the pulmonary artery and the aorta. Detailed information on imaging parameters of this scan has been published previously [9]. Two reviewers measured the PA:A under the supervision of two radiologists. The diameters of the main pulmonary artery and of the ascending aorta were measured at the level of the bifurcation of the pulmonary artery on the same CT-image (Figure 2), according to the procedure as described by Wells et al. [6]. The reviewers were blinded to the clinical status of the participants. Additionally, no information on the status of the lungs could be obtained from the cardiac scans, given that the field-of-view was optimized for visualisation of the heart and the great vessels. The kappa values for the inter-observer and intra-observer agreement for the diameters of the pulmonary artery and the aorta (n=100) were 0.91 and 0.98, and 0.94 and 0.99, respectively.

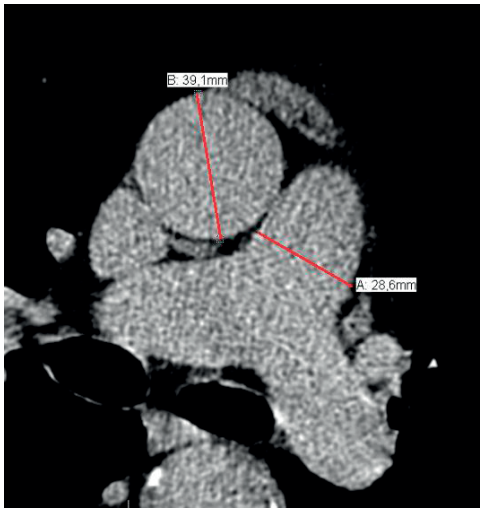

Figure 2 Measurement of the diameter of the pulmonary artery and aorta.
This figure shows an example of the measurement of the diameter of the pulmonary artery (A: 28.6mm) and of the ascending aorta (B: 39.1mm) at the level of bifurcation of the pulmonary artery on the same slice of a non-contrast CT examination.

COPD diagnosis

The diagnosis of COPD was based on an obstructive pre-bronchodilator spirometry (FEV1/FVC < 0.70) according to the Global initiative for Obstructive Lung Disease (GOLD) guidelines [10] Spirometry was performed according to the ATS/ERS guidelines by qualified medical personnel using a portable spirometer (SpiroPro; Erich Jaeger GmbH; Hoechberg, Germany). Spirometry results which did not meet ATS/ERS criteria for acceptability were classified as not interpretable.

In absence of an interpretable study-acquired spirometry, the medical records kept by the general practitioners, including outpatient clinic reports and discharge

letters from medical specialists, were reviewed for all patients who used medication for obstructive lung disease for at least six months (Anatomical Therapeutic Chemical Classification codes: R03) [11]. COPD cases were then defined as having an obstructive lung function measured by their treating physician and clinical events [11].

COPD severity groups were based on GOLD classifications, where $FEV_1/FVC<70$ and $FEV_1 > 80\%$ predicted was defined as mild COPD, while moderate to severe COPD was defined as $FEV_1/FVC<70$ and $FEV_1 < 80\%$ predicted [10].

Covariables

Information on relevant covariables was obtained using interview, physical examinations, and blood sampling [7]. Smoking status was assessed by interview and persons were categorized as current smoker, former smoker or never smoker. Body mass index (BMI) was calculated by dividing body weight in kilograms by height in meters squared. Obesity was defined as a BMI ≥ 30 kg/m^2. Diabetes mellitus was defined as a fasting glucose level of ≥ 7.0 mmol/L or ≥ 11.1 mmol/L if fasting samples were unavailable, or use of blood glucose-lowering medication [12]. Hypertension was defined as a systolic blood pressure ≥ 140 mmHg, a diastolic blood pressure ≥ 90 mmHg, or the use of blood pressure-lowering drugs. As a measure of left ventricular systolic function we used left ventricular fractional shortening at the endocardium in the parasternal long axis window (defined as left ventricular end-diastolic dimension minus left ventricular end-systolic dimension divided by the left ventricular end-diastolic dimension). For this, resting transthoracic echocardiograms were acquired by trained echo-cardiographers according to a standardized protocol. [13]. Clinical diagnosis of heart failure was based on active follow-up using the medical records of the participants [8]. Of note, the routinely collected echocardiography data in the Rotterdam Study was considered in the heart failure adjudication process. Pulmonary artery systolic pressure (PASP) was estimated from echocardiographic measurements using the recommendations by the ASE/EAE/CSE as the sum of the estimated right atrial pressure (based on inferior vena cava diameter and forced respiratory collapse) and the pressure gradient over the tricuspid valve. The pressure gradient was computed from the highest Doppler tricuspid regurgitation velocity gathered from several windows using the simplified Bernoulli equation ($4v^2$, where v is tricuspid regurgitation peak velocity in m/sec) [14]. PH was defined as having PASP > 40 mmHg. Finally, diffusing capacity of carbon monoxide per alveolar volume (DLCO/VA (mmol/min/kPa/L)) was measured by single breath technique and was corrected for haemoglobin values. PASP and DLCO were measured between 2009 and 2012.

Statistical analysis

We determined the association between PA:A (per 1-SD increase) and all-cause mortality in the general population, using Cox proportional-hazard models. For the Cox models we adjusted for covariates that were considered biologically relevant and also changed the point estimates of the univariate association with mortality by at least 10%. In the first model we adjusted for age and sex. In the second model we additionally adjusted for BMI, smoking, diabetes mellitus, left ventricular systolic function and COPD. In addition, we examined the relation between PA:A and mortality in the individuals with COPD using the same Cox proportional hazard models. Furthermore, analyses were also performed in mild and moderate to severe COPD. As a sensitivity analysis, we also analysed the data in moderate COPD and severe COPD separately. We explored non-linearity by using fractional polynomials and constructing quartiles of PA:A (Q1 ≤ 0.64, Q2 = 0.65-0.70, Q3 = 0.71-0.77 and Q4 > 0.77). The cut-offs for the PA:A quartiles were based on the study data. Finally, for the association between PA:A and all-cause mortality in individuals with COPD, we conducted a sensitivity analysis additionally adjusting for COPD severity as measured by FEV_1. We also tested the association between 1 SD increase in PA:A and PASP or PH in COPD and non-COPD, using univariate linear regression or logistic regression models. Finally, we analysed DLCO/VA data by testing whether the mean DLCO/VA was statistically significantly different between COPD and non-COPD using an independent sample t-test. Subsequently, we tested whether the risk of mortality per 1-SD increase of PA:A differs in individuals with and without COPD with different values of DLCO/VA. This was done by adding an interaction term to the second model. Missing data on covariables were imputed using the Expectation Maximization (EM) method. We used SPSS version 21 (IBM Corp, Armonk, NY, USA) for all analyses.

RESULTS

Figure 1 shows the study flow of the participants that were included in this study. The scanned population (n=2,524) was not different from the total RS population [15]. The baseline characteristics of the population with interpretable PA:A measurements (n=2,197) are presented in Table 1. The mean age was 69.7 (SD 6.7) years and 51.3% were women. The prevalence of COPD at the time of CT scanning was 10% (n=222). The 90^{th} percentile of the PA:A in our population was 0.84. The maximum PA:A was 1.27, and only 17 out of 2,197 persons had a ratio of 1 or higher. Additional information about the diameter of the pulmonary artery, the aorta and the PA:A by disease status in the general population is presented in Table 2. During 17,751 person-years of follow-up (median: 8.8 years), 423 (19.3%) persons died [mortality rate 23.8 per 1,000 person-

Table 1 Baseline characteristics of the study population

	Total population n = 2,197		No COPD n = 1,975		COPD n = 222	
Age, years	69.7	(6.7)	69.5	(6.7)	71.2	(7.1)
Female	51.3%		52.3%		42.3%	
Ever smoker	68.5%		67.1%		81.1%	
Pack-years of smoking*	22.6	(21.5)	21.4	(21.1)	32.1	(21.6)
Body mass index, kg/m^2	27.9	(4.0)	27.9	(4.0)	27.2	(3.7)
Hypertension	73.9%		73.4%		78.4%	
Diabetes Mellitus	12.6%		11.8%		19.8%	
Heart failure	3.1%		2.5%		6.5%	
Left ventricular systolic function, %	40.0	(6.2)	40.2	(6.1)	38.1	(7.5)
DLCOc/VA, mmol/min/kPa/L†	1.5	(0.2)	1.5	(0.2)	1.3	(0.3)
PASP, mmHg†	26.0	(7.0)	25.7	(6.7)	29.2	(8.7)
FEV_1, % predicted	103.3	(19.8)	106.1	(17.7)	79.8	(21.0)
Pulmonary artery, mm	26.0	(3.7)	25.9	(3.6)	26.7	(4.1)
Aorta, mm	36.9	(3.9)	36.9	(3.9)	37.5	(3.7)
PA:A	0.71	(0.10)	0.71	(0.10)	0.72	(0.11)

DLCOc/VA: diffusing capacity of carbon monoxide per alveolar volume, corrected for haemoglobin, FEV_1: Forced expiratory volume in 1 second, PA:A: pulmonary artery to aorta ratio, PASP: Pulmonary artery systolic pressure.

Values are mean (standard deviation) for continuous variables or percentages for dichotomous variables.

Data represent original data without imputed values. In de total population missing values were present for smoking (2.2%), pack years of smoking (4.6%), body mass index (0.9%), hypertension (0.3%), diabetes mellitus (6.0%), heart failure (1.9%), left ventricular systolic function (3.4%), DLCOc/VA (40%), PASP (42%), FEV1, % pred (11.3%).

*Pack-years are presented for former and current smokers only.

† DLCOc/VA and PASP measures were performed between 2009 and 2012.

years (95% confidence interval (CI) 21.6-26.2)]. The mortality rate in individuals with prevalent COPD was 51 per 1,000 person-years (95% CI 40.6-63.1). The main causes of death in the COPD group were cardiovascular events (41.8%), bronchial carcinoma (16.4%), other malignancies (13.4%) and pulmonary complications from COPD (6.0%).

PA:A and the risk of all-cause mortality

We explored the association between PA:A and the risk of mortality by a linear model. The log relative hazard between PA:A ratio and mortality in individuals with or without COPD are plotted in Figure 3. P-value for non-linearity in those groups was 0.76 and 0.10, respectively.

Table 2 Mean values of A (aorta), PA (pulmonay artery) and PA:A, stratified by the presence of absence of risk factors (n = 2,197)

		N	Mean (SD)		
			A(mm)	PA (mm)	PA:A
Gender					
	Men	1,070	38.1 (3.9)	26.4 (3.8)	0.70 (0.1)
	Women	1,127	35.8 (3.5)	25.6 (3.6)	0.72 (0.1)
Obesity					
	Yes	562	37.6 (4.0)	27.3 (3.8)	0.73 (0.1)
	No	1,635	36.7 (3.8)	25.6 (3.6)	0.70 (0.1)
Smoking					
	Current	381	37.1 (4.0)	26.0 (3.9)	0.71 (0.1)
	Former	1,124	37.2 (3.8)	26.1 (3.6)	0.71 (0.1)
	Never	644	36.4 (3.9)	25.8 (3.7)	0.71 (0.1)
COPD					
	Prevalent*	222	37.5 (3.7)	26.7 (4.1)	0.72 (0.1)
	Mild	98	37.7 (3.8)	26.1 (4.2)	0.70 (0.1)
	Moderate to severe	107	37.5 (3.6)	27.3 (4.1)	0.73 (0.1)
	Absent	1,975	36.9 (3.9)	26.0 (3.6)	0.71 (0.1)
Hypertension					
	Present	1,624	37.2 (3.9)	26.2 (3.7)	0.71 (0.1)
	Absent	566	36.1 (3.7)	25.5 (3.6)	0.71 (0.1)
DM					
	Present	277	36.7 (3.9)	26.7 (3.7)	0.73 (0.1)
	Absent	1,786	37.0 (3.9)	25.9 (3.7)	0.71 (0.1)
Heart failure					
	Present	69	38.1 (4.7)	28.9 (4.9)	0.77 (0.1)
	Absent	2,128	36.9 (3.8)	25.9 (3.6)	0.71 (0.1)

COPD: chronic obstructive pulmonary disease; DM: diabetes mellitus; PA:A: pulmonary artery to aorta ratio; SD: standard deviation.
Data are presented in mean (SD).
* The total number of COPD cases in the sub-groups do not add up to the total number of all COPD n=222, since COPD diagnosis in 15 out of 222 was confirmed after reviewing medical charts and specialist letters, but severity could not be adjudicated.

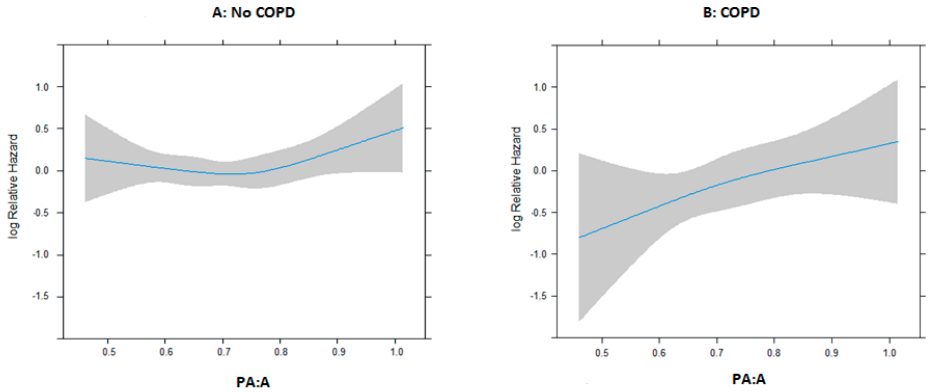

Figure 3 Log relative hazard plotted against PA:A values
Estimates for the log relative hazard in A: individuals without COPD, and B: individuals with COPD were derived using 3 restricted cubic splines, with a P for non-linearity for A = 0.10 and for B = 0.76.

A modest, but statistically non-significant association was found between PA:A with the risk of mortality in the general population (adjusted hazard ratio (HR) per 1-SD increase in PA:A: 1.08 (95%CI:0.98-1.18)) and in individuals without COPD (HR 1.04 (95% CI 0.93-1.15)) (Table 3). In persons with prevalent COPD, a larger PA:A was statistically significantly associated with a higher risk of mortality (HR per 1-SD increase in PA:A: 1.21 (95% CI 1.01-1.46)). The risk of mortality was higher in individuals with moderate to severe COPD compared to individuals with mild COPD (HR per 1-SD

Table 3 The association between PA:A and all-cause mortality.

			Hazard Ratio per 1-SD increase in PA:A (95% CI)			
			Model1*		Model2†	
Total population‡	N=2,197		1.09	(1.00-1.20)	1.08	(0.98-1.18)
No COPD	n=1,975		1.05	(0.95-1.17)	1.04	(0.93-1.15)
All COPD¶	n=222		1.17	(0.98-1.40)	**1.21**	**(1.01-1.46)**
Mild		n=98	1.08	(0.77-1.52)	1.09	(0.76-1.56)
Moderate-severe		n=109	1.22	(0.94-1.58)	**1.36**	**(1.03-1.79)**

CI: Confidence interval; COPD: chronic obstructive pulmonary disease; PA:A: pulmonary artery to aorta ratio; SD: standard deviation.
*Model1: adjusted for age and sex.
†Model2: adjusted for age, sex, body mass index, smoking, diabetes mellitus and left ventricular systolic function.
‡ Additionally adjusted for COPD in Model2.
¶ The total number of COPD cases in the sub-groups do not add up to the total number of all COPD n=222, since COPD diagnosis in 15 out of 222 was confirmed after reviewing medical charts and specialist letters, but severity could not be adjudicated.

Table 4 Risk of mortality per quartile increase of PA:A*

		Hazard Ratio (95% CI)								
PA:A quartiles		Total population†		No COPD		All COPD‡		Mild COPD		Moderate to severe COPD	
		N=2,197		N=1,975		N=222		N=98		N=109	
Q1	(≤ 0.64)	Reference		Reference		Reference		Reference		Reference	
Q2	(0.65-0.70)	0.97	(0.74-1.28)	0.93	(0.68-1.26)	1.30	(0.65-2.60)	0.46	(0.11-1.98)	1.73	(0.64-4.70)
Q3	(0.71-0.77)	1.03	(0.78-1.35)	0.93	(0.68-1.26)	1.64	(0.81-3.33)	0.91	(0.28-2.94)	1.95	(0.69-5.46)
Q4	(> 0.77)	1.19	(0.91-1.55)	1.06	(0.79-1.43)	**2.03**	**(1.06-3.88)**	1.33	(0.41-4.31)	**2.78**	**(1.07-7.23)**
P-for-trend:		0.20		0.74		**0.02**		0.62		**0.03**	

CI: confidence interval; COPD: chronic obstructive pulmonary disease; PA:A: pulmonary artery to aorta ratio.
*Adjusted for age, sex, BMI, smoking, diabetes mellitus, and left ventricular systolic function.
† Additionally adjusted for COPD.
‡ The total number of COPD cases in the sub-groups do not add up to the total number of all COPD n=222, since COPD diagnosis in 15 out of 222 was confirmed after reviewing medical charts and specialist letters, but severity could not be adjudicated.

increase in PA:A: 1.36 (95% CI 1.03-1.79) versus (HR per 1-SD increase in PA:A: 1.09 (0.76-1.56), respectively) (Table 3). Sensitivity analysis showed that the association in individuals with moderate-to-severe COPD is mainly driven by individuals with severe COPD with a HR per 1-SD increase in PA:A: 3.01 (95% CI 1.26-7.17). The association in prevalent COPD did not materially change after adjustment for COPD severity by FEV1 (L) (HR per 1-SD increase in PA:A: 1.22 (95% CI 1.02-1.48)).

Table 4 represents the results of the association of PA:A quartiles with the risk of mortality in all groups. We observed a statistically significant trend over the quartiles in persons with COPD (P-for-trend = 0.02), particularly in those with moderate to severe COPD (P-for-trend =0.03). The corresponding Kaplan-Meier curve in individuals with COPD is presented in Figure 4.

Individuals with moderate to severe COPD in the highest PA:A quartile, had an almost 3-fold increased risk of mortality (HR 2.78 (95% CI 1.07-7.23)) compared to the persons in the lowest quartile of the PA:A.

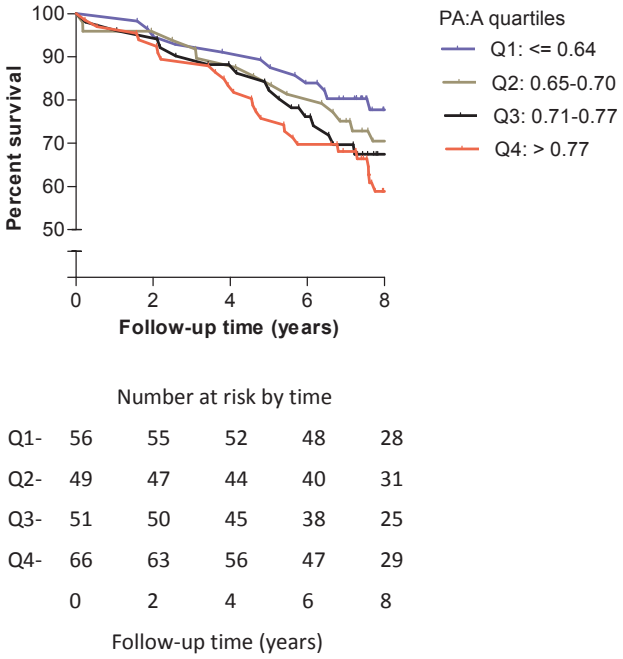

Figure 4 Kaplan-Meier survival curve of individuals with COPD (n=222) by quartiles of PA:A.

PA:A and Pulmonary Artery Systolic Pressure (PASP)

A statistically significant association was found between PA:A ratio and PASP, as estimated by echocardiography, both in non-COPD and COPD subjects. We observed that 1-SD increase of PA:A was associated with 0.61 (SE 0.23) mmHg increase in PASP in individuals without COPD (95% CI 0.16-1.06), and with 2.33 (SE 0.77) mmHg increase in PASP in individuals with COPD (95% CI 0.80-3.85).

PA:A and Pulmonary Hypertension

The association between PA:A and pulmonary hypertension (defined as PASP >40 mmHg) were in line with that between PA:A and PASP. PA:A was significantly associated with pulmonary hypertension in both groups. The risk of PH in COPD per 1-SD increase in PA:A was OR: 2.59 (95% CI 1.07-6.32), whereas the risk of PH in non-COPD per 1-SD increase of PA:A was OR: 1.86 (1.17-2.94).

PA:A and diffusing capacity of the lung

The mean diffusing capacity of the lung per alveolar volume (DLCO/VA) in COPD subjects was statistically significantly different from the mean DLCO/VA in non-COPD subjects (1.32 versus 1.49 mmol/min/kPa/L, P-value < 0.001).

We tested for interaction between PA:A and DLCO/VA in the Cox model to investigate whether the relationship between PA:A and mortality is different with different values of DLCO/VA, separately for individuals with COPD and without. In individuals with COPD, we observed a statistically significant interaction between PA:A and DLCO/VA (P-for-interaction= 0.01), while no interaction was observed in individuals without COPD (P-for-interaction= 0.14).

DISCUSSION

In this large population-based study, we observed no association between PA:A and mortality in an older general population. However, persons with COPD with higher PA:A ratios were at increased risk of mortality, particularly those with moderate to severe COPD.

To our knowledge, this is the first population-based study to investigate the association between the PA:A and mortality in the general population and in individuals with COPD, specifically. Although several studies have been performed with PA:A as proxy of PH, the variation in the use of PA:A is considerable, and the results have been inconsistent. There have been several studies that used the PA:A as a binary variable with 0.9 or 1 as cut-off for normal (< 0.9 or 1) versus abnormal (> 0.9 or 1). For example, Ersoy and colleagues [2] found no association between PA:A >1 and the risk of mortality risk in a group of 106 patients admitted to the intensive care because of an acute COPD exacerbation. In contrast, another retrospective study conducted in 1,326 patients with suspected coronary artery disease undergoing CT angiography [16], demonstrated that PA:A >0.9 was associated with an increased risk of mortality. In addition Shin et al [17], demonstrated that PA:A >1 was significantly associated with an increased risk of mortality in very severe COPD patients undergoing evaluation for lung transplantation.

An important consideration with regard to these studies cut-offs, is that these results are based on findings in selected groups such as COPD patients with severe exacerbations requiring hospitalization [2], patients with suspected ischemic heath disease [16] or patients with very severe COPD undergoing evaluation for lung transplantation [17]. In contrast to these clinic-based populations, our cohort represents the general population, including smokers and non-smokers, with and without various comorbidities. Importantly, only 17 out of 2,197 persons in our cohort had a PA:A >1. Therefore, instead of dichotomizing the PA:A on a value of 1, we analysed PA:A as a continuous measure and explored potential linear associations with risk of mortality.

We observed that PA:A was not associated with mortality in the general population and in individuals without COPD. When analysing quartiles of PA:A, we found that

only persons with moderate to severe COPD who were in the highest quartile of PA:A were at statistically significantly increased risk of mortality, and that this association was driven by individuals with severe COPD. Pulmonary artery pressure and thereby PH is known to increase in advanced COPD due to combined effects of hypoxaemia and loss of capillaries in severe emphysema [18, 19]. In this study, we demonstrated that the magnitude of increase in pulmonary arterial systolic pressures (PASP) per 1 SD increase in PA:A is much larger in persons with COPD compared to those without COPD. We also demonstrated that the risk of mortality per 1 SD increase of PA:A was most pronounced in subjects with reduced diffusing capacity of the lung, suggesting that the identified association in individuals with COPD may be driven by emphysema. Further longitudinal studies are necessary to address potential causality.

Our finding suggests that the PA:A is an independent determinant of mortality in individuals with moderate to severe COPD. Especially, given that non-contrast CT scans of the chest are often performed in current clinical practice for various indications (e.g. screening for lung cancer in subjects at risk), our finding might help in identifying persons with COPD who are at increased risk of death and thereby provide guidance for more targeted therapeutic decision making.

The strength of our study is the population-based setting and the long follow-up period, the prospective, standardized data collection for COPD and mortality, and the standardized, CT-based assessment of PA:A, with excellent inter- and intra-observer correlation coefficients.

Our study also has some limitations. First, in individuals with COPD, analyses of cause-specific mortality as endpoint were not performed due the limited number of cases per cause. Second, the observed risks of mortality may be underestimated due to healthy volunteer effect [20]. Third, although we have carefully considered the potential confounders of the association between PA:A and mortality, it is possible that part of the observed associations might be explained by residual confounding. Finally, the Rotterdam Study comprises a homogenous sample of white participants, which may limit the generalizability of our results to other ethnic groups.

CONCLUSION

Larger PA:A is an independent indicator for mortality in individuals with COPD, particularly in moderate to severe COPD. Measuring PA:A in scans, including those obtained for other indications, may yield important prognostic information usable to tailor patient management in clinical practice.

REFERENCES

1. ERS, *Chapter 13: Chronic obstructive pulmonary disease*, in *ERS whitebook*. 2013.
2. Ortac Ersoy, E., et al., *Measurement of pulmonary artery to aorta ratio in computed tomography is correlated with pulmonary artery pressure in critically ill chronic obstructive pulmonary disease patients.* J Crit Care, 2016. **33**: p. 42-6.
3. Arcasoy, S.M., et al., *Echocardiographic assessment of pulmonary hypertension in patients with advanced lung disease.* Am J Respir Crit Care Med, 2003. **167**(5): p. 735-40.
4. Fisher, M.R., et al., *Estimating pulmonary artery pressures by echocardiography in patients with emphysema.* Eur Respir J, 2007. **30**(5): p. 914-21.
5. Mahammedi, A., et al., *Pulmonary artery measurements in pulmonary hypertension: the role of computed tomography.* J Thorac Imaging, 2013. **28**(2): p. 96-103.
6. Wells, J.M., et al., *Pulmonary arterial enlargement and acute exacerbations of COPD.* N Engl J Med, 2012. **367**(10): p. 913-21.
7. Hofman, A., et al., *The Rotterdam Study: 2016 objectives and design update.* Eur J Epidemiol, 2015. **30**(8): p. 661-708.
8. Leening, M.J., et al., *Methods of data collection and definitions of cardiac outcomes in the Rotterdam Study.* Eur J Epidemiol, 2012. **27**(3): p. 173-85.
9. Odink, A.E., et al., *Association between calcification in the coronary arteries, aortic arch and carotid arteries: the Rotterdam study.* Atherosclerosis, 2007. **193**(2): p. 408-13.
10. Vestbo, J., 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): p. 347-65.
11. Terzikhan, N., et al., *Prevalence and incidence of COPD in smokers and non-smokers: the Rotterdam Study.* Eur J Epidemiol, 2016.
12. Ligthart, S., et al., *Lifetime risk of developing impaired glucose metabolism and eventual progression from prediabetes to type 2 diabetes: a prospective cohort study.* Lancet Diabetes Endocrinol, 2016. **4**(1): p. 44-51.
13. Kardys, I., et al., *Echocardiographic parameters and all-cause mortality: the Rotterdam Study.* Int J Cardiol, 2009. **133**(2): p. 198-204.
14. Rudski, L.G., et al., *Guidelines for the echocardiographic assessment of the right heart in adults: a report from the American Society of Echocardiography endorsed by the European Association of Echocardiography, a registered branch of the European Society of Cardiology, and the Canadian Society of Echocardiography.* J Am Soc Echocardiogr, 2010. **23**(7): p. 685-713; quiz 786-8.
15. Bos, D., et al., *Intracranial carotid artery atherosclerosis: prevalence and risk factors in the general population.* Stroke, 2012. **43**(7): p. 1878-84.
16. Nakanishi, R., et al., *Mortality risk as a function of the ratio of pulmonary trunk to ascending aorta diameter in patients with suspected coronary artery disease.* Am J Cardiol, 2013. **111**(9): p. 1259-63.
17. Shin, S., et al., *Pulmonary artery size as a predictor of pulmonary hypertension and outcomes in patients with chronic obstructive pulmonary disease.* Respir Med, 2014. **108**(11): p. 1626-32.
18. Chaouat, A., R. Naeije, and E. Weitzenblum, *Pulmonary hypertension in COPD.* Eur Respir J, 2008. **32**(5): p. 1371-85.
19. Wells, J.M. and M.T. Dransfield, *Pathophysiology and clinical implications of pulmonary arterial enlargement in COPD.* Int J Chron Obstruct Pulmon Dis, 2013. **8**: p. 509-21.

20. Leening, M.J., et al., *Healthy volunteer effect and cardiovascular risk*. Epidemiology, 2014. **25**(3): p. 470-1.

4.2

Chronic Obstructive Pulmonary Disease is associated with an increased risk of peripheral artery disease and mortality

ABSTRACT

Introduction: Patients with COPD commonly present with multimorbidity. We aimed to investigate the association between COPD and the development of peripheral arterial disease (PAD) in the general population, and how this might affect mortality among individuals with COPD.

Methods: We included 3,123 participants of the population-based Rotterdam Study without PAD at baseline (mean age 65yr; 57.4% females). The association between COPD at baseline and PAD (Ankle Brachial Index ≤ 0.9) during follow-up at the research center, was studied using logistic regression. Cox regression was used for mortality analysis. Interaction terms were used to investigate mortality risk modification by PAD.

Results: Presence of COPD was associated with incident PAD, (Odds $Ratio_{adjusted}$ 1.9 (95% Confidence Interval 1.1-3.2)). Mortality rates per 100,000 PY were: 10.0 in individuals without COPD and PAD, 18.4 in those with only COPD, 16.1 in those with only PAD, and 30.1 in individuals with both COPD and PAD. No statistical interaction was found between PAD and COPD and risk of dying.

Conclusion: Individuals with COPD have an almost doubled risk of developing PAD. Although PAD does not modify the association between COPD and mortality, people suffering from both diseases have substantially higher mortality rates.

Key words: PAD, COPD, epidemiology, population-based, effect modification

INTRODUCTION

Chronic Obstructive Pulmonary Disease (COPD) is the third major cause of death worldwide (1). Beyond the airflow limitation and respiratory impairment, COPD patients often suffer from multi-morbidities, of which most prominently vascular diseases (2) (3).

Atherosclerosis is known as a major cause of symptomatic vascular disease. One of the manifestations of atherosclerosis is peripheral arterial disease (PAD), which refers to the occlusion of the arteries in the lower limbs. PAD is often asymptomatic but can be present in advanced stages in the form of intermittent claudication, leg pain at rest and rarely non-healing wounds, ulcerations and gangrene (4).

Patients with PAD have comparable – but not identical - risk factor profiles compared to patients with coronary heart disease, of which systemic inflammation, diabetes and smoking are most predominant in PAD (5-8).

The prevalence of PAD in individuals with COPD shows a wide range of variation depending on the COPD severity in the study population. Its prevalence was found to be 8% in Asian COPD patients, 8.8% in the German COSYCONET study of COPD patients followed in secondary care and 36.8% in COPD patients hospitalized for a severe exacerbation (9, 10 , 11).

A previous study by Pecci and colleagues (10) demonstrated that PAD was asymptomatic in a large proportion of COPD patients, and was associated with more severe lung disease than in COPD subjects without PAD. However, longitudinal studies investigating the association between COPD and incident PAD are lacking. In addition, although PAD is known to have a significant impact on mortality, it is unknown whether the risk of mortality is higher in patients with both COPD and PAD. Therefore, the objectives of our study were to investigate the association between COPD and incident PAD in a longitudinal cohort, and to elucidate the effect of PAD on mortality in subjects with COPD in a community-dwelling population of middle-aged and older subjects.

METHODS

Setting

The present study was embedded within the Rotterdam Study, an ongoing prospective population-based cohort which was enrolled to investigate the occurrence of, and risk factors for, chronic diseases in the general population. The aims and methods of the Rotterdam Study have been published in detail previously (12). In short, the Rotterdam Study started in 1990, where inhabitants of the Ommoord district in Rotterdam were invited every 3-4 years to the research centre for follow-up examination. Participants were additionally monitored continuously for morbidity and mortality through linkage

of general practitioners and municipality records to the study base. The present study was conducted using data from the first cohort of the Rotterdam Study (RS-I) and comprises of two parts; the first longitudinal analysis was performed in order to study the association between COPD (measured at baseline, between 1990-1993) and newly diagnosed PAD (measured between 1996-2000) (**Figure 1**). In short, there were 6,450 individuals with a measurement of the ankle-brachial index (ABI) at baseline as described below of whom 336 had prevalent PAD, 1,881 died, and 1,110 did not have a second measurement of the ABI. The Rotterdam Study has been approved by the Medical Ethics Committee of the Erasmus MC and by the Dutch Ministry of Health, Welfare and Sport. All participants provided written informed consent to participate in the study and to have their information obtained from treating physicians.

COPD diagnosis

We validated whether participants had COPD by reviewing the medical charts, including outpatient clinic reports and hospital discharge letters of all participants who used medication for obstructive lung disease for at least six months (Anatomical Therapeutic Chemical Classification codes: R03) or who stated to have COPD in the questionnaire. COPD cases were defined as having a medical COPD diagnosis supported by clinical presentation and/or obstructive lung function (13). The index date was defined as the date of diagnosis of COPD as described in the medical charts or the date of a first prescription of COPD medication in patients with COPD, whichever came first. In this study COPD data was used at baseline for the first objective and at follow-up for the second objective.

Assessment of Peripheral Artery Disease (PAD)

The method of measuring the Ankle Brachial Index (ABI) has been described previously (14). In short, blood pressure in the arm was obtained by calculating the mean of two successive measurements with a random-zero sphygmomanometer at the right brachial artery while the participant was in a sitting position (14). In addition, in both the left and the right leg, the systolic blood pressure level of the posterior tibial artery was measured using a random-zero sphygmomanometer and an 8-MHz continuous-wave Doppler probe (Huntleigh 500 D, Huntleigh Technology) while the participant was in the supine position. The ABI was defined as the ratio of the systolic blood pressure at the ankle to the systolic blood pressure at the arm and was calculated for each leg. The presence of PAD was defined as an ABI of ≤ 0.90 in at least one leg (5). Patients with prevalent PAD at baseline were excluded and only newly diagnosed (incident) PAD during follow-up was considered as outcome. Severe PAD was defined as ABI < 0.6 (11). Besides the ABI measurements, we used the criteria of the World Health

Organization/Rose questionnaire which was incorporated in the home interview (15) to investigate intermittent claudication to capture symptomatic PAD.

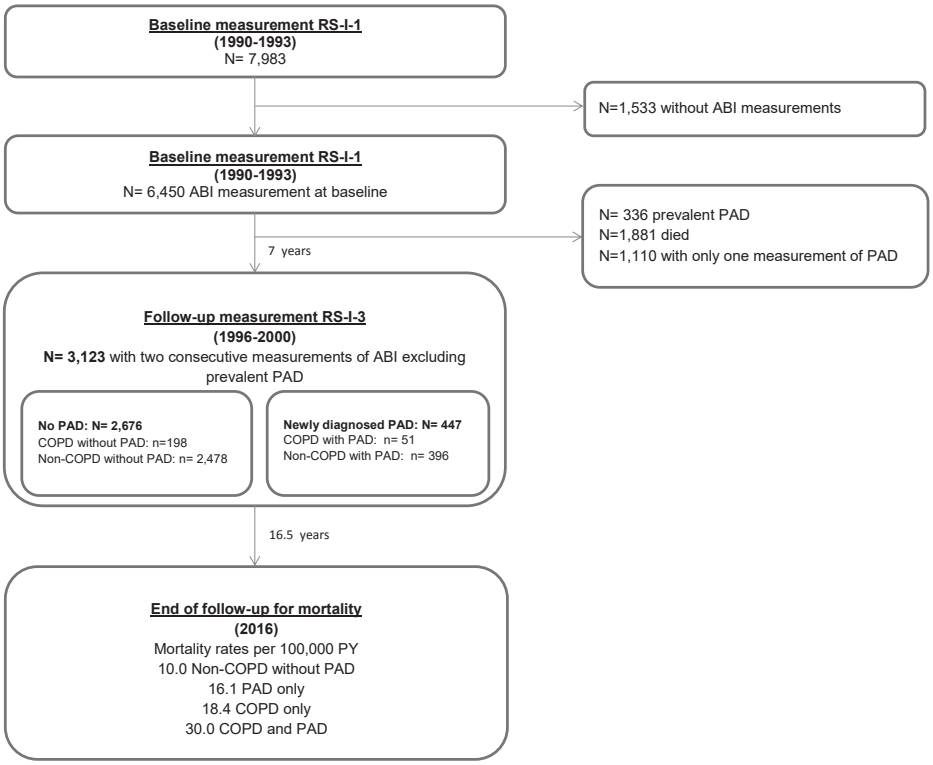

Figure 1: Flow chart to investigate the association between COPD and newly diagnosed PAD. ABI: Ankle Brachial Index; COPD: Chronic obstructive pulmonary disease; N: total sample size; PAD: Peripheral arterial disease; RS-I-1: The first visit of the first wave of the Rotterdam Study. RS-I-3: The third visit of the first wave of the Rotterdam Study.

Follow-up for mortality

To investigate whether PAD increases the risk of all-cause mortality in subjects with or without COPD, 3,123 individuals with 2 ABI measurements, were followed from the second PAD measurement on until mortality or the last visit to the study centre.

Information on mortality of study participants was obtained from the municipality of Rotterdam, and additionally validated with data from medical records kept by the general practitioner's, as described in detail previously (16). Mortality data was complete until August 2016. Cause specific mortality data was complete until January 2014.

Covariables

Relevant covariables at baseline of each analysis were obtained using interview, blood sampling, and physical examination (17). Smoking status was assessed by interview and categorized as never, former or current smoker, or as ever and never smoker. Body mass index (BMI) was calculated by dividing body weight in kilograms by height in meters squared. Hypertension was defined as a systolic blood pressure ≥ 140 mmHg, a diastolic blood pressure ≥ 90 mmHg, or the use of blood pressure-lowering drugs. Subjects were classified as statin users if they had received at least one prescription for statins at baseline. Information on statin use was obtained from interview and from pharmacies. Type 2 diabetes mellitus was defined according to recent WHO guidelines, as a fasting blood glucose ≥ 7.0 mmol/L, a non-fasting blood glucose ≥ 11.1 mmol/L (in absence of fasting samples), or the use of blood glucose lowering medication (16). Concentrations of HDL cholesterol and total cholesterol were measured from the blood samples using an automated enzymatic method (18). HDL to total cholesterol ratio was calculated by dividing HDL blood concentrations by total cholesterol concentrations. Ethnicity was based on genetic ancestry data, or interview data in absence of genetic information. High sensitivity C-reactive protein (hs-CRP) was measured using Rate Near Infrared Particle Immunoassay (Immage Immunochemistry System, Beckman Coulter, Fullerton, CA).

Statistical analysis

To study population characteristics the chi-square test and the independent sample t-test were used to test differences between individuals with or without COPD. We determined the association between baseline COPD and the development of PAD (assessed during follow-up visit), using logistic regression after exclusion of individuals with prevalent PAD at baseline. For the association between COPD and newly diagnosed PAD, we adjusted for covariables that were considered risk factors for atherosclerosis and cardiovascular disease. The following potential confounders where considered: age, sex, smoking, pack years of smoking, BMI, hypertension, hs-CRP, statin use, HDL to cholesterol ratio, ethnicity and diabetes. Missing data on covariables were imputed using the Expectation Maximization (EM) method. Effect modification was tested with interaction terms. Additional sensitivity analyses were performed by adjusting for hs-CRP (as a marker for inflammation) and by excluding incident COPD or individuals with ABI > 1.4 (as a marker for arterial stiffening (14)) from the analysis.

As secondary objective, we aimed to study the impact of PAD on the risk of mortality among individuals with COPD. For this analysis four groups were defined. 1) individuals without COPD nor PAD, 2) individuals with COPD only, 3) individuals with PAD only and 4) individuals with both COPD and PAD all assessed at the second PAD round. Mortality rates were obtained by dividing the number of deaths by the total number of

person years (PY) of subjects at risk and were presented per 100,000 PY. Confidence intervals (CI) around these rates were calculated using a Poisson distribution. Follow-up time for all subjects was defined as the time period between the start of the second analysis (where incident PAD was measured) and death or the last visit to the study centre. Sensitivity analysis was performed by censoring for incident COPD during follow-up. Median follow-up time was estimated using the reverse Kaplan-Meier method. Crude survival curves were obtained using the Kaplan-Meier survival method. The Cox proportional hazard model was used to calculate mortality hazard ratios and was adjusted for age, sex, pack years of smoking, hypertension, HDL to cholesterol ratio and diabetes mellitus. We used SPSS version 21 (IBM Corp, Armonk, NY, USA) and R software for all analyses.

RESULTS

The characteristics of the population for follow-up (n=3,123) are presented in **Table 1**. The mean age was 65.1 (SD 6.6) years and 57.2% were women. The prevalence of ever smokers was 65% in the total population, and 67.1% in individuals with incident PAD (**Tables 1** and **2**). Within individuals with PAD, almost all men (89.1%) and 48.3% of the women were ever smoker. In total, 1.8% (56/3,123) had severe PAD based on an ABI < 0.6. Only 21.4% (12/56) of the subjects with severe PAD indicated to have intermittent claudication on questionnaire.

COPD and the association with newly diagnosed PAD

During a median follow-up of 7.4 years, 447 (12.9%) patients developed incident PAD. A statistically significant association was found between COPD and newly diagnosed PAD during follow-up. The age- and sex adjusted odds ratio (OR) was 1.89 (95% CI 1.13-3.17), and 1.87 (95% CI 1.10-3.18) after additional adjustment for pack years of smoking, hypertension, HDL to cholesterol ratio and diabetes mellitus (**Table 2**). This association remained significant upon additional adjustment for baseline hsCRP levels ($OR_{adjusted}$ 1.86 (95% CI 1.10-3.16), exclusion of individuals (n=186) who developed COPD during follow-up ($OR_{adjusted}$1.93 (95% CI 1.13-3.29)) and upon exclusion of individuals with ABI >1.40 (n=495) ($OR_{adjusted}$ 1.95 (95% CI 1.13-3.36)). The association between COPD and incident PAD was stronger in males ($OR_{adjusted}$ 2.59; 95% CI 1.04-5.16) than females (OR $1.22_{adjusted}$; 95% CI: 0.53-2.87). However, test for interaction between COPD and sex was not significant (p for interaction=0.18). Stratified analysis by smoking revealed that the overall association between COPD and newly developed PAD was driven by the group of individuals who had ever smoked (P for interaction = 0.41) (**Table 2**).

Table 1 Baseline characteristics of all subjects without PAD at baseline, stratified by COPD status.

	Overall	no COPD	COPD	p-value
N (%)	3,123 (100)	3,038 (97)	85 (3)	
Age (mean (sd))	65.14 (6.6)	65.12 (6.6)	65.69 (5.6)	0.43
Female, n (%)	1785 (57.2)	1745 (57.4)	40 (47.1)	0.07
BMI (mean (sd))	26.27 (3.5)	26.28 (3.5)	26.17 (3.1)	0.78
Pack years (mean (sd))	25.5 (21.3)	25.3 (21.2)	33.5 (24.1)	< 0.01
Smoking status,n (%)				< 0.01
Never	1069 (35.4)	1054 (35.8)	15 (18.5)	
Former	1394 (46.1)	1348 (45.8)	46 (56.8)	
Current	560 (18.5)	540 (18.4)	20 (24.7)	
Ever smoking, n (%)	1954 (64.6)	1888 (64.2)	66 (81.5)	< 0.01
Diabetes, n (%)	380 (12.6)	367 (12.5)	13 (15.9)	0.47
HDL/Chol ratio (mean (sd))	0.21 (0.1)	0.21 (0.1)	0.22 (0.1)	0.13
Cholesterol lowering drugs, n (%)	72 (2.3)	70 (2.3)	2 (2.4)	0.98
Hypertension, n (%)	1481 (47.4)	1446 (47.6)	35 (41.2)	0.29
BP lowering drugs, n (%)	770 (24.7)	743 (24.5)	27 (31.8)	0.12
hs-CRP (mean (sd))	2.44 (3.3)	2.42 (3.3)	3.02 (3.0)	0.11
Inhaled therapy (R03), n (%)	1.24 (4.0)	64 (2.1)	60 (70.6)	< 0.01

BP: Blood pressure; BMI: Body Mass Index; COPD: Chronic Obstructive Pulmonary Disease; hs-CRP: High sensitivity C-reactive protein, HDL/Chol ratio: High density lipoprotein to cholesterol ratio; PAD: Peripheral arterial disease; SD: standard deviation

P-values represents the difference between COPD groups. Data represent original data without imputed values. In de total population, missing values were present for smoking (3.2%), pack years of smoking (7.1%), body mass index (0.2%), and diabetes mellitus (3.5%).

*Pack years are presented for former and current smokers only

Disease status and risk of all-cause mortality

We investigated the impact of PAD on mortality starting follow-up for all subjects from the time when the second PAD measurement was performed (n= 3,123). The general characteristics of the population which was followed-up for mortality are presented in **Table 3**. The prevalence of PAD in COPD was 20.5% (51/(198+51)) *versus* PAD in non-COPD 13.8% (396/(2,478+396)). We stratified the study population into the following four groups: 1) No COPD or PAD (N=2,478), 2) COPD only (N= 198) 3) PAD only (N= 396) and 4) comorbid COPD and PAD (N= 51). Median follow-up time for mortality was 16.5 years. At the end of follow-up, 1,805 individuals (58%) had died. The mortality rate in the group without COPD nor PAD was 10.0 per 100,000 PY (95% CI 9.47-10.62); subjects with COPD and subjects with PAD had comparable mortality rates: 18.4 (95% CI 15.49-21.58) and 16.1 (95% CI 14.17-18.16) per 100,000 PY,

Table 2 The association between COPD and the development of PAD additionally stratified by sex and smoking status

		N (%)						Model 1*		Model 2†		P int.‡
		Overall		**No PAD**		**PAD case**		**OR**	**95% CI**	**OR**	**95% CI**	
Overall		3,123	(100)	2,676	(100)	447	(100)					
non-COPD		3,038	(97.0)	2,611	(97.6)	427	(95.5)	**Ref.**	-	**-**	-	
COPD		85	(3.0)	65	(2.4)	20	(5.0)	**1.89**	1.13-3.18	**1.87**	1.10-3.18	
Stratified	***M***											0.18
non-COPD		1,293	(41.4)	1,122	(41.9)	171	(38.3)	**Ref.**	-	**-**	-	
COPD		45	(1.4)	32	(1.2)	13	(2.9)	**2.61**	1.33-5.12	**2.59**	1.04-5.16	
	F											
non-COPD		1,745	(55.9)	1,489	(55.6)	256	(57.3)	**Ref.**	-	-	-	
COPD		40	(1.3)	33	(1.2)	7	(1.6)	1.24	0.54-2.87	1.22	0.53-2.87	
	NS											0.41
non-COPD		1,066	(34.1)	921	(34.4)	145	(32.4)	**Ref.**	-	-	-	
COPD		15	(0.5)	13	(0.5)	2	(0.4)	1.00	0.21-4.47	1.20	0.26-5.57	
	ES											
non-COPD		1,972	(63.1)	1,690	(63.2)	282	(63.1)	**Ref.**	-	**-**	-	
COPD		70	(2.2)	52	(1.9)	18	(4.0)	**2.00**	1.14-3.50	**1.98**	1.12-3.49	

CI: Confidence interval; F: Female; M: male; N: total number; ES: Ever smoker; NS: Never smoker; OR: odds ratio; PAD: peripheral arterial disease
*Model1: adjusted for age and sex
†Model2: adjusted for age, sex, pack years of smoking, hypertension, HDL to cholesterol ratio and diabetes
#Excluding the covariate sex from the analysis
‡ P value for interaction, calculated by adding an interaction term to Model 2.
Data for smoking represent imputed values.

respectively. In subjects with both COPD and PAD, the mortality rate was the highest: 30.01 per 100,000 PY (95% CI 22.01-40.11). **Figure 2** represents the Kaplan-Meier survival curves in the different groups. Age and sex adjusted mortality hazard ratios compared to individuals without COPD nor PAD were, 1.41 (95% CI 1.24-1.61), 1.50 (95% CI 1.26-1.77) and 2.23 (95% CI 1.66-3.00) for subjects with PAD, COPD and subjects with both COPD with PAD, respectively. After additional adjustment for pack years of smoking, hypertension, HDL to cholesterol ratio and diabetes mellitus the mortality rates were 1.36 (95% CI 1.19-1.54), 1.52 (95% CI 1.28-1.80) and 2.30 (95% CI 1.71-2.09), for subjects with PAD, COPD and subjects with both COPD with PAD, respectively. We did not find effect modification by PAD on the multiplicative scale (p for interaction= 0.82). In addition, sensitivity analysis after censoring for incident COPD during follow-up did not materially change the results.

Table 3 Characteristics of all subjects at the second PAD assessment: individuals without COPD nor PAD, subjects with COPD only, subjects with PAD only and subjects with both COPD and PAD at the second PAD assessment . This table describes baseline characteristics of individuals who were followed for mortality.

	Overall	**No COPD or PAD**	**COPD only#**	**PAD only**	**COPD and PAD#**
N	3123	2478	198	396	51
Age (mean (SD))	71.73 (6.58)	71.13 (6.36)	72.79 (5.95)	74.58 (7.21)	74.84 (6.89)
Women (%)	1785 (57.20)	1435 (57.90)	87 (43.90)	246 (62.10)	17 (33.30)
BMI (mean (SD))	26.87 (3.89)	26.94 (3.86)	26.91 (3.78)	26.55 (4.08)	25.83 (3.72)
Ever smoker, (%)	2,049 (65.60)	1,580 (63.80)	163 (82.30)	261 (65.9)	45 (88.2)
Pack years (mean (SD))	25.47 (22.50)	23.99 (22.01)	34.13 (26.62)	26.49 (19.85)	39.78 (25.63)
Diabetes Mellitus (%)	816 (26.10)	608 (24.50)	63 (31.80)	130 (32.80)	15 (29.40)
HDL/Chol ratio (mean (SD))	0.24 (0.07)	0.25 (0.07)	0.25 (0.08)	0.23 (0.06)	0.25 (0.08)

Data represent imputed values

#The total number of individuals with COPD at follow-up as presented in this table (N=249, N=198 COPD only + N=51 COPD and PAD)) represent COPD at baseline (N=85, see Table 1) plus the individuals that acquired COPD during follow-up

*Pack years are presented for former and current smokers only

BMI: Body Mass Index; COPD: Chronic Obstructive Pulmonary Disease; HDL/Chol ratio: High density lipoprotein to cholesterol ratio; PAD: Peripheral arterial disease; SD: standard deviation.

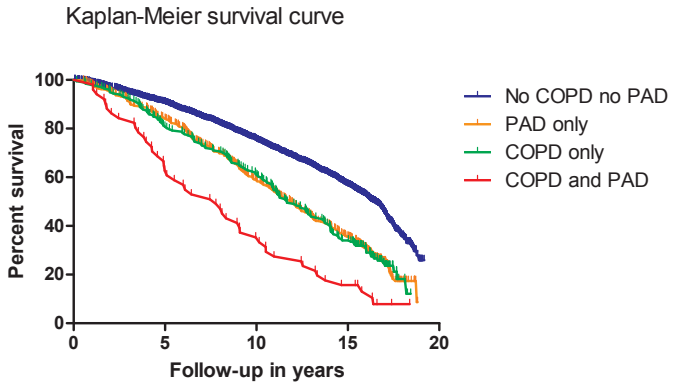

Figure 2: Kaplan-Meier curves of mortality in different groups according to the presence or absence of COPD and Peripheral Arterial Disease (PAD).
COPD: Chronic obstructive pulmonary disease; PAD: Peripheral arterial disease

DISCUSSION

To our knowledge, this is the first population based study that investigated longitudinally the association between COPD and incident PAD, and the influence of PAD on mortality rates in individuals with COPD. We observed that individuals with COPD have a higher risk of developing PAD, and that the risk of mortality was highest in people with both diseases.

Smoking is a well-known risk factor for both COPD and PAD. However in our study, the association between COPD and PAD remained statistically significant upon adjustment for cumulative smoking history (i.e. pack years). Additionally, increased CRP levels are known to be an important risk factor for PAD, and CRP is also elevated in a subgroup of patients with COPD (19). However, adjusting for baseline hsCRP levels did not change the overall association substantially. Despite the potential for residual confounding or other unmeasured mediators of systemic inflammation, such as Tumor Necrosis Factor-ć(TNF-ǵ, Interleukin-1β (IL-1β) or matrix metalloproteinases, we hypothesize that COPD-related mechanisms beyond smoking and systemic inflammation might contribute to the onset of PAD.

COPD and PAD have a well-known impact on mortality, however, no study in the literature has investigated the impact of PAD on the association between COPD and all cause-mortality. We observed a higher risk of mortality in individuals when COPD and PAD co-occur, although formally there was no statistical interaction between both diseases on a multiplicative scale. The observed mortality rates in individuals with PAD or with COPD are in line with literature reports. The ERS white book (https://www.erswhitebook.org) reported an age-standardised (to the European Standard Population) mortality rate for COPD of approximately 18 per 100,000 inhabitants per year. Also,

unstandardized all-cause mortality in the subgroup of individuals with only COPD was 18.4 per 100,000 PY. In addition, in a group of patients with asymptomatic PAD, Diehm *et. al.* (20) reported a mortality rate ratio of 1.4 after adjustment for known cardiovascular risk factors. We observed a similar adjusted hazard ratio of 1.4 in our population, which highlights the robustness of our data.

Our findings have important implications for disease management (21). As COPD is associated with PAD development and PAD is often asymptomatic, patients with COPD might benefit from routine ABI screening for the timely diagnosis of PAD. Apart from its cost-effectiveness, early targeted screening and treatment of asymptomatic PAD is likely to improve health by preventing future cardiovascular disease (22).

The strengths of this study are the prospective longitudinal design, and the standardised data collection. However, our study has also some limitations. First, information on PAD was only gathered at the study centre (during study visits), implying that actual dates of PAD onset were missing which prevented us to conduct time to PAD development analysis. Second, since spirometry measurements were not routinely performed in the early rounds of the Rotterdam Study data, the COPD incident date was based on a clinical diagnosis of COPD only, and not all clinically diagnosed COPD subjects were still alive to be confirmed by spirometry at the research centre later on. Third, although we have evaluated all available known potential confounders of the association between COPD and PAD development, residual confounding might still explain part of this association. Finally, including patients with two ABI measurements at two subsequent visits might lead to a selection towards healthier individuals.

Conclusion

Subjects with COPD have a higher risk of developing PAD. People with both COPD and PAD have a substantially increased risk of death. Consequently, early detection of PAD and preventive actions in people with COPD should receive more attention in clinical respiratory care.

REFERENCES

1. Lozano R, Naghavi M, Foreman K, Lim S, Shibuya K, Aboyans V, 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. 2012;380(9859):2095-128.
2. Barnes PJ. Senescence in COPD and Its Comorbidities. Annu Rev Physiol. 2017;79:517-39.
3. Negewo NA, Gibson PG, McDonald VM. COPD and its comorbidities: Impact, measurement and mechanisms. Respirology. 2015;20(8):1160-71.
4. Patel MR, Conte MS, Cutlip DE, Dib N, Geraghty P, Gray W, et al. Evaluation and treatment of patients with lower extremity peripheral artery disease: consensus definitions from Peripheral Academic Research Consortium (PARC). J Am Coll Cardiol. 2015;65(9):931-41.
5. Authors/Task Force M, Aboyans V, Ricco JB, Bartelink MEL, Bjorck M, Brodmann M, et al. 2017 ESC Guidelines on the Diagnosis and Treatment of Peripheral Arterial Diseases, in collaboration with the European Society for Vascular Surgery (ESVS). Eur J Vasc Endovasc Surg. 2017.
6. Bah F, Bhimji SS. Peripheral Arterial Disease. 2017.
7. Hirsch AT, Criqui MH, Treat-Jacobson D, Regensteiner JG, Creager MA, Olin JW, et al. Peripheral arterial disease detection, awareness, and treatment in primary care. JAMA. 2001;286(11):1317-24.
8. Tunstall-Pedoe H, Peters SAE, Woodward M, Struthers AD, Belch JJF. Twenty-Year Predictors of Peripheral Arterial Disease Compared With Coronary Heart Disease in the Scottish Heart Health Extended Cohort (SHHEC). J Am Heart Assoc. 2017;6(9).
9. Lin MS, Hsu KY, Chen YJ, Chen CR, Chen CM, Chen W. Prevalence and risk factors of asymptomatic peripheral arterial disease in patients with COPD in Taiwan. PLoS One. 2013;8(5):e64714.
10. Pecci R, Aguado JD, Rivo ABS, Conde PS, Abelaira MDC. Peripheral arterial disease in patients with chronic obstructive pulmonary disease. Int Angiol. 2012;31(5):444-53.
11. Houben-Wilke S, Jorres RA, Bals R, Franssen FM, Glaser S, Holle R, et al. Peripheral Artery Disease and Its Clinical Relevance in Patients with Chronic Obstructive Pulmonary Disease in the COPD and Systemic Consequences-Comorbidities Network Study. Am J Respir Crit Care Med. 2017;195(2):189-97.
12. Ikram MA, Brusselle GGO, Murad SD, van Duijn CM, Franco OH, Goedegebure A, et al. The Rotterdam Study: 2018 update on objectives, design and main results. Eur J Epidemiol. 2017;32(9):807-50.
13. 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;31(8):785-92.
14. Meijer WT, Hoes AW, Rutgers D, Bots ML, Hofman A, Grobbee DE. Peripheral arterial disease in the elderly: The Rotterdam Study. Arterioscler Thromb Vasc Biol. 1998;18(2):185-92.
15. Rose G, H B, RF G, RJ P. Cardiovascular Survey Methods. Geneva, Switzerland: World Health Organization. 1982.
16. Leening MJ, Kavousi M, Heeringa J, van Rooij FJ, Verkroost-van Heemst J, Deckers JW, et al. Methods of data collection and definitions of cardiac outcomes in the Rotterdam Study. Eur J Epidemiol. 2012;27(3):173-85.
17. Hofman A, Brusselle GG, Darwish Murad S, van Duijn CM, Franco OH, Goedegebure A, et al. The Rotterdam Study: 2016 objectives and design update. Eur J Epidemiol. 2015;30(8):661-708.

18. Braun KVE, Dhana K, de Vries PS, Voortman T, van Meurs JBJ, Uitterlinden AG, et al. Epigenome-wide association study (EWAS) on lipids: the Rotterdam Study. Clin Epigenetics. 2017;9:15.
19. Broekhuizen R, Wouters EF, Creutzberg EC, Schols AM. Raised CRP levels mark metabolic and functional impairment in advanced COPD. Thorax. 2006;61(1):17-22.
20. Diehm C, Allenberg JR, Pittrow D, Mahn M, Tepohl G, Haberl RL, et al. Mortality and vascular morbidity in older adults with asymptomatic versus symptomatic peripheral artery disease. Circulation. 2009;120(21):2053-61.
21. Alahdab F, Wang AT, Elraiyah TA, Malgor RD, Rizvi AZ, Lane MA, et al. A systematic review for the screening for peripheral arterial disease in asymptomatic patients. J Vasc Surg. 2015;61(3 Suppl):42S-53S.
22. Vaidya A, Joore MA, Ten Cate-Hoek AJ, Ten Cate H, Severens JL. Screen or not to screen for peripheral arterial disease: guidance from a decision model. BMC Public Health. 2014;14:89.

General discussion

5

In this thesis the genetic and non-genetic determinants and clinical consequences of impaired lung function were studied. Especially, genetic epidemiological studies had a great share in this work. Here, comprehensive study designs were used including GWAS, EWAS, and Genome-wide linkage to investigate which genetic and epigenetic changes are significant contributors to disease onset and progression. For this, we integrated population genetics with experimental clinical research. In this chapter, the main findings, research challenges, clinical relevance and future directions are discussed.

MAIN FINDINGS AND CHALLENGES

Epidemiology of COPD

Epidemiological research is crucial to gain insight into the occurrence of diseases, to support informed decision making and develop public health interventions and strategies to address risk factors (1). Alongside with other research disciplines, epidemiological research is essential and aims to answer questions like: how big is the problem? What is the risk of acquiring the disease in a certain amount of time? And which risk factors are important?

In **chapter 2**, we studied the epidemiology of COPD. COPD is the third leading cause of death worldwide and thereby a major public health problem (2). In de last decade, the number of COPD cases worldwide increased by 28% and this number is expected to grow in the future (3). We investigated the epidemiology of COPD in 15,000 participants of the Rotterdam Study who were followed for up to 25 years. In the absence of a clinical diagnosis of asthma in the electronic medical records, COPD was diagnosed based on a pre-bronchodilator obstructive spirometry ($FEV_1/FVC < 0.70$). In absence of an interpretable spirometry within the Rotterdam Study, cases were defined as having COPD diagnosed by a physician on the basis of clinical presentation and obstructive lung function measured by the general practitioner or respiratory physician. We reported on the prevalence and the incidence of the disease, and investigated differences by sex and smoking status.

One of the challenges in epidemiological cohorts studies is, that researchers often deal with historical changes that they have to consider in the sake of a good interpretation of the study results. In the Rotterdam study for example, spirometry was introduced after 2001, meaning that for a subset of COPD cases the diagnosis was based on medical chart records only. Should we have presented the prevalence and incidence data of COPD in the total population only, then we would have missed an interesting piece of information that enriches our knowledge about the incidence of COPD in different settings. Subgroup analyses were required to highlight different study

results (**Table 1**). A higher incidence of COPD was found when diagnosis was based on spirometry (only) as compared to a clinically based diagnosis, which can be attributed to the inclusion of asymptomatic COPD subjects who frequently still have mild airflow limitation. Since mild COPD cases rarely seek medical attention, the incidence of COPD is frequently underestimated in clinical settings.

Table 1: Prevalence and incidence data according to different classification methods in the total cohort and in the sub-groups (spirometry versus medical charts group)

	Spirometry data	Medical records data	Combined data
	N=7,153	N=7,466	N=14,619
COPD GOLD			
Prevalence	5.3%	4.2%	4.7%
Incidence	11.7/1,000PY	5.8/1,000PY	8.9/1,000PY
COPD LLN			
Prevalence	3.4%	4.2%	3.8%
Incidence	5.2/1,000PY	5.8/1,000PY	5.5/1,000PY

COPD: Chronic obstructive pulmonary disease; GOLD: Global initiative for Obstructive Lung Disease; LLN: lower limit of normal; PY: person years

Another historical change is the introduction of "better" (i.e. more accurate) cut-off values of FEV_1/FVC to define airflow limitation. In the same chapter, we presented the incidence figures of COPD according to either the fixed 0.70 ratio of FEV_1/FVC (as recommended by the Global initiative for Obstructive Lung Disease [GOLD]) or the lower limit of normal (LLN; as recommended by the Global Lung function Initiative [GLI]) (**Table 1**) (4-7). The question remains whether we really have to replace the fixed ratio with the LLN. Several studies claim that using the fixed 0.70 FEV_1/FVC ratio overestimates the diagnosis of COPD in older individuals due to the inclusion of asymptomatic subjects with mild COPD (GOLD I) (6, 8, 9), while others state that mild COPD might be a good marker for preclinical disease (6). Importantly, the new criteria does not consider subjects with mild COPD (GOLD I) as "true" COPD cases, even though we have shown in the Rotterdam Study that even subjects with mild airflow limitation (i.e. mild COPD) have an increased risk of cardiovascular mortality (10), all-cause mortality (11) (**Figure 1**), and future severe exacerbations compared with control subjects with normal spirometry (12).

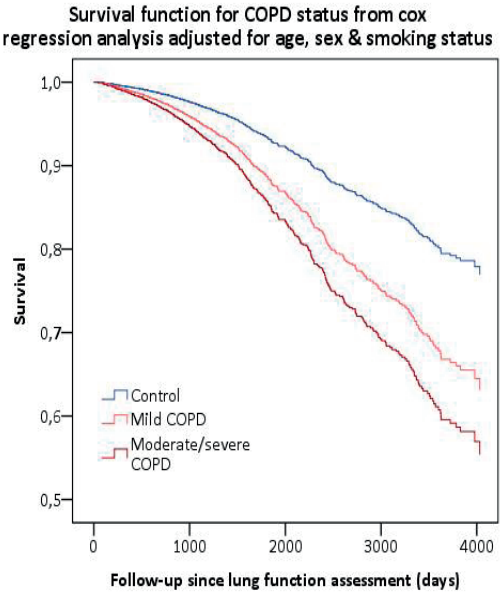

Figure 1 All-cause mortality in COPD in The Rotterdam Study, stratified by COPD severity

The global initiative for chronic obstructive lung diseases (GOLD) defined COPD as a preventable and treatable chronic respiratory disease (13, 14). Although research efforts broadened our understanding of the disease, we are still unable to unravel the exact pathophysiological mechanisms of this disease, nor we are able to define accurate biomarkers of COPD (15-21). Thereby, we are left with a progressive disease that forms a huge burden on the quality of life of millions of people worldwide (14). The only bright spot for patients with COPD are medications for symptoms relief that help them improve their quality of life (22). However, this change in COPD treatment forms a methodological challenge in epidemiological research of the 'natural course' of COPD, because the more treatment options are available, the more difficult it is to capture the differences in severity that arise, due to effect of treatment.

Finally, COPD is an age-related disease. Patients with COPD often present with multi-morbidity including cardiovascular and cerebrovascular disease (22). For this very reason, researchers encounter difficulties studying associations between COPD and other age related diseases, such as dementia, because the more advanced the disease stage is, the more researchers will deal with informative missingness in their data, leading to a possible bias in the generated results (23).

Genome-wide association studies (GWAS)

GWAS associations can typically tag multiple genetic variants in the same region that are highly correlated (or in other words, in linkage disequilibrium (LD)). Correlated

variants have an equal likelihood to be causally associated with the disease of interest, when assuming no independent effects. The idea behind LD relates to the non-random association between one or more loci on the genome (24). The LD between SNPs in the same chromosomal region is calculated based on information of the observed haplotype frequency of the two SNPs of interest and the frequency of each individual SNP (24-26). Although there are many ways proposed to calculate LD (27, 28), it is nowadays typically expressed as a measure of squared correlation (r^2), or as the relative measure of disequilibrium compared to its maximum, as proposed by Lewontin , the D prime (D') (26). However, many web based applications, that provide summaries and visualizations of GWAS results often base their LD calculations on r^2 only. It is important to realize that r^2 can only reach its maximum if the allele frequency of the two SNPs are similar and the two haplotype frequencies are bigger than zero (25, 29). While D' can be large if any allele of the first locus is seen with one allele of the other, and thereby does not depend on allele frequencies as it is the case with r^2. As we are now in the era of conducting GWA studies with better imputation of rare variants, it is desirable to realize that the use of r^2 might lead to misleading results about the true extent of the LD, simply because of its properties. D' might be more useful in case we are interested in LD measures using a less frequent variant. Nevertheless caution is warranted in the interpretation of D' as well, since D' can be inflated when the sample size is small or if one of the two alleles is extremely rare (29, 30).

The statistical power to detect associations in GWAS depends on several factors such as sample size, effect size of the genetic variant(s), the heterogeneity of the phenotype of interest, measurement errors, the distribution of the causal variants in the studied population, their frequency and the correlation (LD) with the observed genotyped variants (31, 32).

In **chapter 3**, we presented two GWA studies of lung function. In the first GWA study we investigated a multi-ethnic cohort with more than 90,000 individuals from 22 studies, of whom 60,000 of European descent. Here, we calculated the heritability and genetic determinants of FEV_1, FVC and FEV_1/FVC, and identified over 50 novel loci, of which sixteen genes encode proteins with predicted or established drug targets (33). In the second GWA study, we calculated the heritability of and studied the genome-wide associations with diffusing capacity of carbon monoxide in approximately 8,000 individuals. Here, we found that the *ADGRG6*, locus was significantly associated with DLCO/VA and that the expression of this gene was decreased in patients with COPD and in subjects with decreased diffusing capacity (DLCO/VA) (34).

Sample size

The two GWA studies show roughly similar heritability estimates but differ considerably in the number of identified loci (50 novel loci versus 1). This example is a very nice

illustration of how sample size (60,000 versus ~8,000 individuals) can increase the power to detect (novel) genome-wide significant associations.

Sample size is one of the most important factors that influence the power to find an association in GWA studies. However, since we are dealing with age related traits of lung function, the question remains whether it is justified to add individuals from all age groups in order to increase sample size and thereby gain power, or to restrict the study population to specific age groups (e.g. older individuals) in order to decrease trait variability and thereby to increase the power.

Heterogeneity

Phenotypic heterogeneity reduces the power to detect statistically significant associations and decreases the magnitude of the risk estimates that are ascribed to genetic variations (35). In the GWAS of diffusing capacity of carbon monoxide, we had a relatively small sample size to begin with, however we tried to gain more power by reducing the heterogeneity of the trait by using not only interpretable but also reproducible measurements in the analyses, so that only high quality lung function measurements were included (32, 35).

Intuitively, the greatest power might be obtained when these two important factors (sample size and heterogeneity) are in balance. Manchia and colleagues suggested that accurate phenotype delineation is maybe even more critical for new genetic discoveries than increasing the sample size (35).

Genetic correlation

So far, GWAS had an important role in genetic discoveries beyond detecting trait-SNP associations (31). GWAS summary statistics can be used in Mendelian randomisation studies to test causal relationships, or to quantify the genomic architecture by assessing the LD structure. These can also be used to detect and quantify pleiotropy by means of polygenic risk scores or genetic correlation estimations. In **chapters 3.1** and **3.2**, we used summary statistics of the GWASs to test genetic correlations between lung function and other traits. An important challenge in genetic correlation studies is, that they are sample size dependent (36, 37). In other words, lack of statistically significant correlation does not mean lack of pleiotropy, but could be due to lack of power to detect pleiotropy. Another challenge in genetic correlations studies is, that genetic correlation estimates might be counter intuitive, for example in the case of a negative genetic correlation between two lung function measures; FEV_1 and DLCO/VA (as measured using data from the Rotterdam Study: -0.35 (SE 0.13), p-value=0.009). Several reasons are possible to explain negative genetic correlations (**Figure 2**), these include: A) direct effect of the causal SNP on the trait, where the SNP has a positive effect on one trait and a negative effect on the other, B) a causal SNP that has a positive effect on the first

trait, which might be in LD with another causal SNP that has a negative effect on the second trait, C) a causal SNP that influences the traits via a mediator that has a positive effect on the first trait and a negative one on the second, and finally as an alternative for explanation C, D) a causal SNP that affects the first trait which in turn affects the second trait in an opposite manner. It is not straightforward to elucidate the exact reason of a negative genetic correlation based on the GWAS results only. Therefore, many methods are now used to detect genome-wide, regional or single-variant pleiotropy (37-39). These methods provide a quantitative assessment of pleiotropy, and have an important role in understanding shared genetic effects.

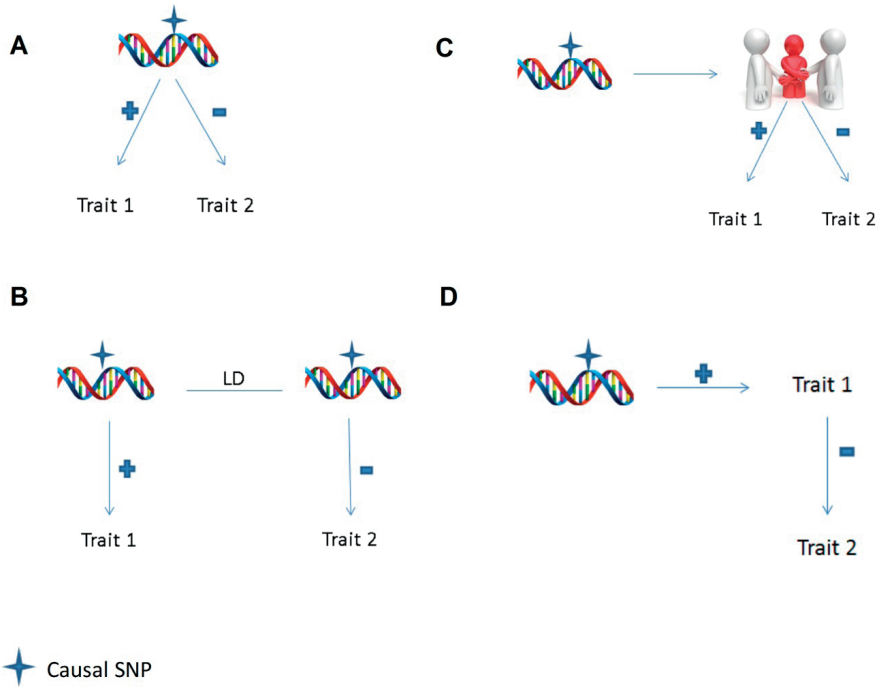

Figure 2 Interpretation of a negative genetic correlation between traits with positive phenotypic correlation

Epigenome-wide association studies (EWAS)

In **chapters 3.3** and **3.4**, we studied epigenetic associations of blood DNA methylation with occupational exposures and diffusing capacity, respectively. Analyses of the first study (**chapter 3.3**) were restricted to 903 never-smokers from **LifeLines**. The association between occupational exposure to gases/fumes, mineral or biological dust and blood DNA methylation was assessed. Several associations were identified and the results were validated in the Rotterdam Study. Finally, methylation-expression associations were assessed. In the second study (**chapter 3.4**), the association between

DLCO and DLCO/VA and blood DNA methylation was assessed, after adjustment for several covariates including smoking. Data from the Rotterdam Study was used for discovery and data from the Framingham Heart Study served as replication panel. Two methylation sites were identified and replicated and the role of smoking related genes on lung function was discussed. In both studies we hypothesised that DNA methylation might act as an intermediate between occupational exposures (**chapter 3.3**) or smoking (**chapter 3.4**) and lung function impairment.

A considerable challenge in EWAS is, that EWAS associations can be causal as well as consequential, implicating major challenges in study design and interpretations of the findings (40, 41). GWAS and EWAS findings are hypothesis generating and need replication in independent population-based or clinic-based cohorts. Replicated findings are interesting candidates for further experimental research.

In the **EWAS** of diffusing capacity of carbon monoxide (DLCO), we hypothesised that the causal effect of smoking would affect methylation levels of the CpG site in the ***AHRR*** gene. The hypomethylation of the *AHRR* gene is associated with the low expression of another gene, ***EXOC3***. We were able to associate the down regulation of mRNA expression of this gene with decreased diffusing capacity in human lung tissue (**Figure 3**). It is worthwhile to mention that although we were able to demonstrate clinical relevance by linking the methylation status to gene expression and consequently lung function decline, the proposed hypothesis is the result of conclusions based on multiple steps where each individual step in the triangle is prone to errors that might lead to biased results and conclusions. For example, residual confounding by smoking might lead to bias in any direction. Also, the biological samples to perform the experimental work on, are often derived from a smaller sample of selected clinical patients (e.g. undergoing thoracic surgery for lung tumours). Therefore caution must be warranted when dealing with such datasets and a well-designed analysis plan is encouraged to avoid bias as much as possible.

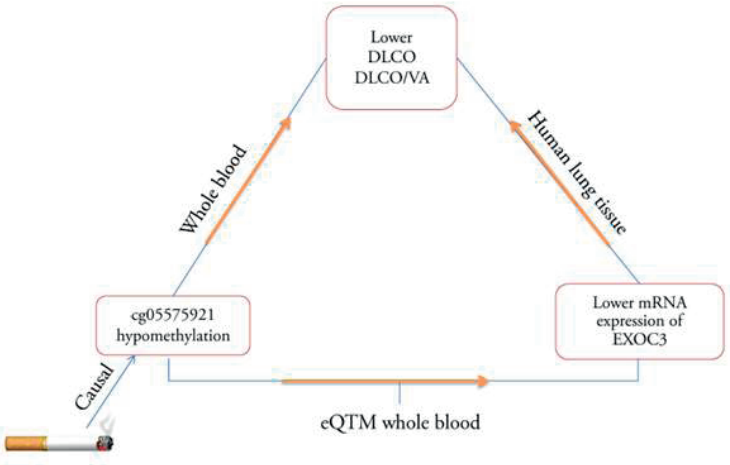

Figure 3 Possible pathway through which hypomethylation of cg05575921 (*AHRR gene)* affects diffusing capacity.

Genome-wide linkage scan

Although larger studies and more accurate data are needed to increase the power for discovery in GWAS, it is unclear whether the findings will help to explain the biggest proportion of the missing heritability. Therefore rare variant analyses, including genome-wide linkage scans might be interesting, as they can be informative for discoveries beyond the ones revealed by GWAS so far (**Table 2**).

Table 2 Possible sources of genetic heritability that might explain the missing heritability problem (42, 43)

Sources of heritability	Causes of missing heritability
Common genetic variants	- Lack of power in GWAS
Rare genetic variants	- Lack of power in WES and WGS - Rare variants might explain only a small part of the heritability - Underrepresentation of rare variants in SNP arrays
Epigenetic variants	- Difficulty to control for epigenetic changes that contribute to variable gene expression
Interaction and epistasis	- GWAS assume additive effects without interaction - Twin studies do take interaction and epistasis into account
Copy number variations, insertions, deletions, inversions	- These variants are largely under investigated - Large sample sizes are needed - Often considered as unapproachable by the current genotyping and sequencing technologies.

GWAS: Genome-wide association study; WES: Whole exome sequencing; WGS: Whole genome sequencing

COPD is a complex disease with genetics contributing to a substantial part of its variability. In **chapter 3.5**, a genome-wide **linkage** scan was performed to discover rare genetic variants underlying COPD in 142 cases clustered in 27 families from an isolated population in The Netherlands; The Erasmus Rucphen Family study (**ERF**) (44). Several linkage peaks were detected (**Figure 4**). Potential causal variants were selected for further association testing with COPD in ERF and the Rotterdam Study. The strengths of this study are that it provides the opportunity to investigators to point-out genetically interesting regions that are associated with COPD. However, a major limitation is often the lack of sequencing information in the affected family members, implicating that no further analysis could be performed (yet) with two of the three discovered linkage peaks.

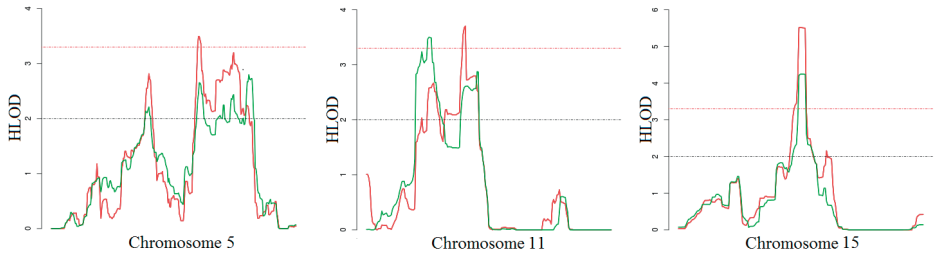

Figure 4 Heterogeneity logarithm of the odds (HLOD) score plot for the chromosomes 5, 11 and 15. X-axis shows chromosome positions and Y-axis shows HLOD score. Red line represents HLOD scores for recessive model and green line for dominant model. Dashed red line represents the level of significance (HLOD=3).

Consequences of impaired lung function

In the fourth chapter of this thesis, we tried to understand the link between impaired lung function or COPD and other comorbidities, including pulmonary hypertension and peripheral arterial disease (PAD).

Pulmonary hypertension

Pulmonary Hypertension is a serious complication in patients with COPD. It evolves due to hypoxic vasoconstriction of the small pulmonary arteries, inflammation, mechanical stress of hyperinflated lungs and loss of pulmonary capillary bed due to emphysema. Pulmonary hypertension may lead to structural changes in the pulmonary arteries and can eventually lead to right ventricular hypertrophy and heart failure (45-47).

Pulmonary artery to aorta (**PA:A**) ratio, defined as the ratio between the diameters of the pulmonary artery and the aorta on transversal images of CT-scans of the thorax, has been shown to be a good surrogate marker of pulmonary hypertension in patients with COPD and a predictor for an increased risk of severe COPD exacerbations requiring hospitalisation (48). In 2,197 participants from the Rotterdam study, we assessed the

association between 1-SD increased of PA:A and mortality in the general population and in individuals with COPD (**chapter 4.1**). We demonstrated that higher PA:A ratio -as a proxy for pulmonary hypertension- has predictive value on **mortality** in individuals with **COPD**, especially in individuals with low **diffusing capacity**. Importantly, even though an increased PA:A ratio is not completely indicative of pulmonary hypertension (49, 50), the PA:A ratio might be a good proxy for the impact of pulmonary hypertension related mechanisms. Other mechanisms for an increased PA:A ratio include left heart disease, pulmonary inflammation, airway remodelling and vasoconstriction, loss of capillary beds in the pulmonary circulation and hyperinflation. In conclusion, this imaging-based marker - the pulmonary artery to aorta (**PA:A**) ratio - has possibly important clinical utility, as it might help in identifying those COPD patients who are at greater risk of severe exacerbations, hospitalisations and mortality, and thereby provide guidance for a more targeted therapeutic decision making (51).

Peripheral artery disease

Patients with COPD present with multi-morbidity of which most prominently vascular disease e.g. atherosclerosis. Peripheral arterial disease (PAD) is one of the manifestations of atherosclerosis and appears to be often asymptomatic in individuals with COPD. In **chapter 4.2**, we aimed to investigate the association between COPD and future PAD (Ankle Brachial Index ≤ 0.9) longitudinally, and to elucidate the effect of PAD on mortality in individuals with COPD. Therefore , we included 3,123 participants from the Rotterdam study. We found that, individuals with COPD have a higher (almost doubled) risk of developing PAD. People suffering from both COPD and PAD had higher mortality rates compared to people without both diseases. Although the increased mortality could not be directly attributed to PAD as cause, the occurrence of PAD frequently heralds vascular disease in other vascular beds resulting in cardiovascular disease (e.g. ischemic heart disease) and cerebrovascular disease (e.g. stroke).

Future directions

In this section, I discuss my ideas about future research in the field of pulmonary and genetic epidemiology.

Where do we stand now?

Today, we are dealing with two opposite movements in science. On the one hand, the increasing interest of scientific journals to support the open data policies. On the other hand, the General Data Protection Regulation (GDPR), that came into force on the 25th of May 2018, that forces attention and action for personal data protection (52-57). At first sight, open data policy and more data protection seem to oppose each other. There are several examples in the scientific field, where data sharing has been practiced

successfully under the umbrella of ethical and legal norms to the individual privacy. For example, the Global Alliance for Genomics and Health (GA4GH), established in 2013, aims to catalyse genomic science by developing the right legal frameworks to assure responsible global sharing of clinical and genetic data (58-62). Such frameworks are needed, since they offer legal dimensions to research and data sharing between the involved institutes and governments, to protect responsible data sharing. Whether GDPR will prohibit data sharing within and between countries is yet to be discovered; several issues in GDPR are subject for own interpretation and need to be clarified first (54). The good news is that these uncertainties will definitely lead to global discussion about future data sharing in research.

COPD research

Despite the enormous effort that is done so far to understand the causes of COPD, comprehensive knowledge of the disease pathophysiology of COPD is still lacking and epidemiological research in older COPD individuals with more advanced disease is challenging. Early identification of individuals who will develop COPD, will not only help with preventing disease onset or progression, it would also solve important challenges which epidemiological COPD research faces, due to the effect of aging and disease severity. Therefore, future research must focus on understanding disease pathophysiology and finding a suitable biomarker for COPD. For this, results of large (epi)genome-wide association studies must also be considered, as these provide simultaneous information on important disease pathways that are involved in the onset and/or progression of this heterogeneous and complex disease.

Genetic discoveries

In the last decade, many consortia of genetic association studies have been set up (63-65), in order to facilitate the possibility of meta-analysing genetic data across many population-based and clinic based cohorts. We witnessed an increase in genetic discoveries by increasing sample sizes of those studies (up to 3 million in the GIANT consortium with Height and BMI phenotypes). It is likely to assume that this trend of cohorts with increasing sample sizes will continue and that it will probably cumulatively explain a substantial amount of the missing heritability in lung function traits and respiratory diseases. Importantly, these large meta-analysis cannot be well performed without a well-designed analysis plan supported by guidelines for observational studies (66, 67), especially harmonization of the data across the cohorts is needed in order to obtain meaningful and unbiased results.

Another approach to increase the chance of discovery, is to use either better imputation reference panels, such as the haplotype reference consortium (HRC) panel, or to use the more expensive deeper density coverage of variation in the genome using

whole exome sequencing (WES) or whole genome sequencing (WGS) techniques. In this thesis, we present two GWAS in which we used 1,000 genomes as imputation reference panel. Even though we are now in the era of using WGS data with better density of coverage of variation in the genome and larger MAF spectrum, the associated costs are still a hurdle. Therefore, it might be a reasonable next step to use the HRC imputation panel in future GWAS of lung function, given its availability, low costs and reasonable accurate imputations of variants with frequency of 0.1% and higher (68).

While there is a long way ahead of us to find out what we do not yet know in lung function genetics (the so called missing heritability; and whether the discovered genetic variants are causal), there is also a huge need for translational research to elucidate how the – presumably causal - genetic variants contribute to the disease pathogenesis. Previous GWAS findings are excellent candidates for experimental studies to provide insight in the pathophysiology of disease, including knockout or knockdown animal models, experimental studies in human (lung) tissue or studies using relatively novel gene editing techniques, such as the CRISPR-cas9 technique (69). These studies are essential to understand the pathophysiology of the genetic findings and shows the path that we need to follow to link genetic discovery with clinical relevance, and most importantly the identification and validation of novel therapeutic targets (**Figure 5**) (70).

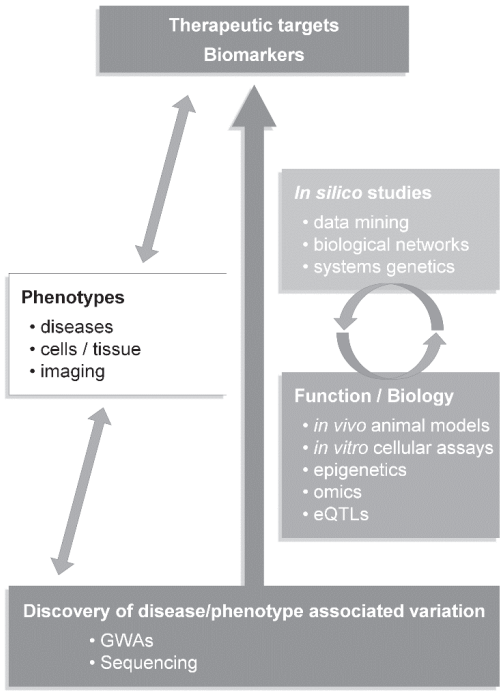

Figure 5 The journey from genetic discoveries based on genetic association studies to finding therapeutic targets and biomarkers for a particular trait or disease.

Clinical observational studies

Regarding the clinical epidemiological studies, our findings could have important implications for disease management in patients with COPD. Indeed, routine measurement of Pulmonary artery to aorta (PA:A) ratio on the chest CT scan (**chapter 3.1**) and/or Ankle Brachial Index screening (**chapter 3.2**) might contribute to timely diagnosis of vascular disease and its consequences.

Implementing PA:A ratio as a prognostic marker for mortality in patients with moderate-to-severe COPD might help with risk stratification and better disease management. There is a significant increase in using chest CT imaging in clinical practice, either in the context of screening for lung cancer in (heavy) smokers or for establishing a diagnosis of pulmonary embolism or ischemic heart disease. On these (screening or clinical) CT scans, also the PA:A ratio should be measured systematically since it provides prognostic information, especially in patients with COPD. Still, several questions remain unanswered. It is interesting to know whether PA:A ratio also applies for patients with mild COPD. Finally, repeated measurement of PA:A ratio in the population could be useful to understand trajectories of PA:A change and their accompanied effect on mortality in COPD.

There is increasing evidence that COPD and peripheral artery disease (PAD) are associated, and our study suggests that COPD might contribute to the development of PAD. However, in a next step it would be interesting to investigate the bi-directionality of the association between COPD and PAD, and specifically to investigate the vascular hypothesis in the development of COPD and emphysema (71). It is indeed tempting to speculate that not only pulmonary and systemic inflammation, but also alterations in the lung circulation including pulmonary vessel remodelling and endothelial cell death are involved in the pathophysiology of COPD (72).

CONCLUSION

In conclusion, in this thesis we presented our work on the epidemiology of impaired lung function and its consequences on morbidity and mortality. An integrated approach to the study of genetic and non-genetic determinants of impaired lung function is necessary in order to elucidate the pathophysiology of COPD and other respiratory diseases.

REFERENCES

1. Tripepi G, Jager KJ, Dekker FW, Zoccali C. Measures of effect in epidemiological research. Nephron Clin Pract. 2010;115(2):c91-3.
2. Lozano R, Naghavi M, Foreman K, Lim S, Shibuya K, Aboyans V, 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. 2012;380(9859):2095-128.
3. Disease GBD, Injury I, Prevalence C. Global, regional, and national incidence, prevalence, and years lived with disability for 328 diseases and injuries for 195 countries, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet. 2017;390(10100):1211-59.
4. 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;31(8):785-92.
5. Housset B. [The lower limit of normal: toward a new COPD detection strategy?] La limite inferieure de la normale : vers une nouvelle strategie de detection de la BPCO ? Rev Mal Respir. 2010;27(9):991-2.
6. Mohamed Hoesein FA, Zanen P, Lammers JW. Lower limit of normal or FEV1/FVC < 0.70 in diagnosing COPD: an evidence-based review. Respir Med. 2011;105(6):907-15.
7. van Dijk WD, Gupta N, Tan WC, Bourbeau J. Clinical relevance of diagnosing COPD by fixed ratio or lower limit of normal: a systematic review. COPD. 2014;11(1):113-20.
8. van Dijk W, Tan W, Li P, Guo B, Li S, Benedetti A, et al. Clinical relevance of fixed ratio vs lower limit of normal of FEV1/FVC in COPD: patient-reported outcomes from the CanCOLD cohort. Ann Fam Med. 2015;13(1):41-8.
9. Izquierdo Alonso JL, De Lucas Ramos P, Rodriguez Glez-Moro JM, grupo de estudio C. The use of the lower limit of normal as a criterion for COPD excludes patients with increased morbidity and high consumption of health-care resources. Arch Bronconeumol. 2012;48(7):223-8.
10. Mattila T, Vasankari T, Kanervisto M, Laitinen T, Impivaara O, Rissanen H, et al. Association between all-cause and cause-specific mortality and the GOLD stages 1-4: A 30-year follow-up among Finnish adults. Resp Med. 2015;109(8):1012-8.
11. Odeyemi YE, Lewis O, Ngwa J, Dodd K, Gillum RF, Mehari A. Does Low FEV1 in Addition to Fixed Ratio and/or Lower Limit of Normal of FEV1/FVC Improve Prediction of Mortality in COPD? The NHANES-III-linked-mortality Cohort. J Natl Med Assoc. 2018.
12. Santibanez M, Garrastazu R, Ruiz-Nunez M, Helguera JM, Arenal S, Bonnardeux C, et al. Predictors of Hospitalized Exacerbations and Mortality in Chronic Obstructive Pulmonary Disease. PLoS One. 2016;11(6):e0158727.
13. Rodriguez-Roisin R, Rabe KF, Vestbo J, Vogelmeier C, Agusti A, all p, et al. Global Initiative for Chronic Obstructive Lung Disease (GOLD) 20th Anniversary: a brief history of time. Eur Respir J. 2017;50(1).
14. Rabe KF, Watz H. Chronic obstructive pulmonary disease. Lancet. 2017;389(10082):1931-40.
15. Agusti A, Sin DD. Biomarkers in COPD. Clin Chest Med. 2014;35(1):131-41.
16. Faner R, Tal-Singer R, Riley JH, Celli B, Vestbo J, MacNee W, et al. Lessons from ECLIPSE: a review of COPD biomarkers. Thorax. 2014;69(7):666-72.
17. Goncalves I, Guimaraes MJ, van Zeller M, Menezes F, Moita J, Simao P, et al. Clinical and molecular markers in COPD. Pulmonology. 2018.
18. Agusti A, Bafadhel M, Beasley R, Bel EH, Faner R, Gibson PG, et al. Precision medicine in airway diseases: moving to clinical practice. Eur Respir J. 2017;50(4).

19. Agusti A, Celli B, Faner R. What does endotyping mean for treatment in chronic obstructive pulmonary disease? Lancet. 2017;390(10098):980-7.
20. Agusti A, Gea J, Faner R. Biomarkers, the control panel and personalized COPD medicine. Respirology. 2016;21(1):24-33.
21. Jones PW, Agusti AG. Outcomes and markers in the assessment of chronic obstructive pulmonary disease. Eur Respir J. 2006;27(4):822-32.
22. Vogelmeier CF, Criner GJ, Martinez FJ, Anzueto A, Barnes PJ, Bourbeau J, et al. Global Strategy for the Diagnosis, Management, and Prevention of Chronic Obstructive Lung Disease 2017 Report. GOLD Executive Summary. Am J Respir Crit Care Med. 2017;195(5):557-82.
23. Hardy SE, Allore H, Studenski SA. Missing data: a special challenge in aging research. J Am Geriatr Soc. 2009;57(4):722-9.
24. Slatkin M. Linkage disequilibrium--understanding the evolutionary past and mapping the medical future. Nat Rev Genet. 2008;9(6):477-85.
25. Wray NR. Allele frequencies and the r2 measure of linkage disequilibrium: impact on design and interpretation of association studies. Twin Res Hum Genet. 2005;8(2):87-94.
26. Lewontin RC. On measures of gametic disequilibrium. Genetics. 1988;120(3):849-52.
27. Chen Y, Lin CH, Sabatti C. Volume measures for linkage disequilibrium. BMC Genet. 2006;7:54.
28. Devlin B, Risch N. A comparison of linkage disequilibrium measures for fine-scale mapping. Genomics. 1995;29(2):311-22.
29. Teare MD, Dunning AM, Durocher F, Rennart G, Easton DF. Sampling distribution of summary linkage disequilibrium measures. Ann Hum Genet. 2002;66(Pt 3):223-33.
30. Tenesa A, Wright AF, Knott SA, Carothers AD, Hayward C, Angius A, et al. Extent of linkage disequilibrium in a Sardinian sub-isolate: sampling and methodological considerations. Hum Mol Genet. 2004;13(1):25-33.
31. Visscher PM, Wray NR, Zhang Q, Sklar P, McCarthy MI, Brown MA, et al. 10 Years of GWAS Discovery: Biology, Function, and Translation. Am J Hum Genet. 2017;101(1):5-22.
32. Sham PC, Purcell SM. Statistical power and significance testing in large-scale genetic studies. Nat Rev Genet. 2014;15(5):335-46.
33. Wyss AB, Sofer T, Lee MK, Terzikhan N, Nguyen JN, Lahousse L, et al. Multiethnic meta-analysis identifies ancestry-specific and cross-ancestry loci for pulmonary function. Nat Commun. 2018;9(1):2976.
34. Terzikhan N, Sun F, Verhamme FM, Adams HHH, Loth D, Bracke KR, et al. Heritability and genome-wide association study of diffusing capacity of the lung. Eur Respir J. 2018.
35. Manchia M, Cullis J, Turecki G, Rouleau GA, Uher R, Alda M. The impact of phenotypic and genetic heterogeneity on results of genome wide association studies of complex diseases. PLoS One. 2013;8(10):e76295.
36. Github. Genetic correaltions ldscore 2017 [updated 22-04-2017. Available from: https://github.com/bulik/ldsc/wiki/FAQ.
37. Hackinger S, Zeggini E. Statistical methods to detect pleiotropy in human complex traits. Open Biol. 2017;7(11).
38. Ligthart S, Vaez A, Hsu YH, Inflammation Working Group of the CC, Pmi Wg XCP, LifeLines Cohort S, et al. Bivariate genome-wide association study identifies novel pleiotropic loci for lipids and inflammation. BMC Genomics. 2016;17:443.
39. Schaid DJ, Tong X, Larrabee B, Kennedy RB, Poland GA, Sinnwell JP. Statistical Methods for Testing Genetic Pleiotropy. Genetics. 2016;204(2):483-97.

40. Rakyan VK, Down TA, Balding DJ, Beck S. Epigenome-wide association studies for common human diseases. Nat Rev Genet. 2011;12(8):529-41.
41. Saffari A, Silver MJ, Zavattari P, Moi L, Columbano A, Meaburn EL, et al. Estimation of a significance threshold for epigenome-wide association studies. Genet Epidemiol. 2018;42(1):20-33.
42. Eichler EE, Flint J, Gibson G, Kong A, Leal SM, Moore JH, et al. Missing heritability and strategies for finding the underlying causes of complex disease. Nat Rev Genet. 2010;11(6):446-50.
43. Blanco-Gomez A, Castillo-Lluva S, Del Mar Saez-Freire M, Hontecillas-Prieto L, Mao JH, Castellanos-Martin A, et al. Missing heritability of complex diseases: Enlightenment by genetic variants from intermediate phenotypes. Bioessays. 2016;38(7):664-73.
44. Nedeljkovic I, Terzikhan N, Vonk JM, van der Plaat DA, Lahousse L, van Diemen CC, et al. A Genome-Wide Linkage Study for Chronic Obstructive Pulmonary Disease in a Dutch Genetic Isolate Identifies Novel Rare Candidate Variants. Front Genet. 2018;9:133.
45. Rosenkranz S, Gibbs JS, Wachter R, De Marco T, Vonk-Noordegraaf A, Vachiery JL. Left ventricular heart failure and pulmonary hypertension. Eur Heart J. 2016;37(12):942-54.
46. GOLD. Gobal strategy for the diagnosis, management and prevention of chronic obstructive pulmonary disease. 2018 report. 2018.
47. Bazan IS, Fares WH. Pulmonary hypertension: diagnostic and therapeutic challenges. Ther Clin Risk Manag. 2015;11:1221-33.
48. Wells JM, Washko GR, Han MK, Abbas N, Nath H, Mamary AJ, et al. Pulmonary arterial enlargement and acute exacerbations of COPD. N Engl J Med. 2012;367(10):913-21.
49. Zouk AN, Wells JM. In Rotterdam, size really does matter: implications of pulmonary artery enlargement on mortality. Eur Respir J. 2017;49(6).
50. Wells JM, Dransfield MT. Pathophysiology and clinical implications of pulmonary arterial enlargement in COPD. Int J Chron Obstruct Pulmon Dis. 2013;8:509-21.
51. Terzikhan N, Bos D, Lahousse L, Wolff L, Verhamme KMC, Leening MJG, et al. Pulmonary artery to aorta ratio and risk of all-cause mortality in the general population: the Rotterdam Study. Eur Respir J. 2017;49(6).
52. Townend D. Conclusion: harmonisation in genomic and health data sharing for research: an impossible dream? Hum Genet. 2018;137(8):657-64.
53. Knoppers BM, Joly Y. Introduction: the why and whither of genomic data sharing. Hum Genet. 2018;137(8):569-74.
54. Nyren O, Stenbeck M, Gronberg H. The European Parliament proposal for the new EU General Data Protection Regulation may severely restrict European epidemiological research. Eur J Epidemiol. 2014;29(4):227-30.
55. Chassang G. The impact of the EU general data protection regulation on scientific research. Ecancermedicalscience. 2017;11:709.
56. Morrison M, Bell J, George C, Harmon S, Munsie M, Kaye J. The European General Data Protection Regulation: challenges and considerations for iPSC researchers and biobanks. Regen Med. 2017;12(6):693-703.
57. Haug CJ. Turning the Tables - The New European General Data Protection Regulation. N Engl J Med. 2018;379(3):207-9.
58. Knoppers BM. Framework for responsible sharing of genomic and health-related data. Hugo J. 2014;8(1):3.

59. Rahimzadeh V, Dyke SO, Knoppers BM. An International Framework for Data Sharing: Moving Forward with the Global Alliance for Genomics and Health. Biopreserv Biobank. 2016;14(3):256-9.
60. Terry SF. The global alliance for genomics & health. Genet Test Mol Biomarkers. 2014;18(6):375-6.
61. Knoppers BM. International ethics harmonization and the global alliance for genomics and health. Genome Med. 2014;6(2):13.
62. Lawler M, Siu LL, Rehm HL, Chanock SJ, Alterovitz G, Burn J, et al. All the World's a Stage: Facilitating Discovery Science and Improved Cancer Care through the Global Alliance for Genomics and Health. Cancer Discov. 2015;5(11):1133-6.
63. consortium E. ENIGMA consortium: Enhancing neuroimaging genetics through meta-analysis 2018 [Available from: http://enigma.ini.usc.edu/.
64. consortium C. CHARGE consortium: cohorts for heart and aging research in genomic epidemiology 2018 [Available from: http://www.chargeconsortium.com/.
65. consortium G. GIANT: Genetic investigation of anthropometric triats 2018 [Available from: https://portals.broadinstitute.org/collaboration/giant/index.php/Main_Page.
66. von Elm E, Altman DG, Egger M, Pocock SJ, Gotzsche PC, Vandenbroucke JP, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: guidelines for reporting observational studies. Int J Surg. 2014;12(12):1495-9.
67. Thelle DS. STROBE and STREGA: instruments for improving transparency and quality of reporting scientific results. Eur J Epidemiol. 2009;24(1):7-8.
68. McCarthy S, Das S, Kretzschmar W, Delaneau O, Wood AR, Teumer A, et al. A reference panel of 64,976 haplotypes for genotype imputation. Nat Genet. 2016;48(10):1279-83.
69. van de Vrugt HJ, Cornel MC, Wolthuis RMF. [CRISPR/Cas: technique to repair DNA errors: is a clinical breakthrough near?] CRISPR/Cas: techniek om DNA-fouten te repareren. Ned Tijdschr Geneeskd. 2018;162.
70. Center DM, Schwartz DA, Solway J, Gail D, Laposky AD, Lin QS, et al. Genomic medicine and lung diseases. Am J Respir Crit Care Med. 2012;186(3):280-5.
71. Okpechi SC, Ghonim MA, Lammi MR. Advances in Chronic Obstructive Pulmonary Disease Therapy: A Vascular-Targeted Approach. Clin Med Insights-Th. 2017;9.
72. Voelkel NF, Gomez-Arroyo J, Mizuno S. COPD/emphysema: The vascular story. Pulm Circ. 2011;1(3):320-6.

6

Summary

Samenvatting

SUMMARY

Lung function tests and genetics are important to understand lung health and to investigate pathways that lead to an impaired lung function. The aim of this thesis was to study the epidemiology of impaired lung function including COPD, to unravel the genetic and non-genetic determinants, and finally to study the impact of impaired lung function on adverse health outcomes.

In **chapter 2**, we studied the epidemiology of COPD and investigated its prevalence and incidence in the general population. Here, we found a higher incidence of COPD in men and smokers.

Chapter 3 includes the genetic epidemiological studies, including heritability and genome-wide association studies in **chapters 3.1** and **3.2**, epigenome-wide association studies in **chapters 3.3** and **3.4** and finally genome-wide linkage scan in **chapter 3.5**. In all these studies, interesting genes and pathways were discovered and where possible replicated, making these findings very interesting candidates for further experimental research.

Chapter 4 includes the clinical epidemiological studies and highlights the consequences of an impaired lung function. In **chapter 4.1** we were able to demonstrate that a higher pulmonary artery to aorta ratio increases the risk of mortality in individuals with COPD, especially in those with a low diffusing capacity. In **chapter 4.2** we found an association between COPD and incident peripheral arterial disease (PAD). We also showed that individuals with both COPD and PAD had higher mortality rates compared to people without both disease, although the increased mortality rate could not be attributed to PAD as cause.

In **chapter 5**, we discuss the most important findings, challenges and future directions.

SAMENVATTING

Longfunctietesten en genetica zijn belangrijk om de physiologie van de long te begrijpen en om "pathways" te onderzoeken die leiden tot een verminderde longfunctie. Het doel van dit proefschrift was om de epidemiologie van een gedaalde longfunctie, zoals bijvoorbeeld bij patiënten met COPD, te bestuderen, om de genetische en niet-genetische determinanten te ontrafelen en om uiteindelijk de impact van een gestoorde longfunctie op slechtere gezondheidsuitkomsten te bestuderen.

In **hoofdstuk 2** bestudeerden we de epidemiologie van COPD en onderzochten we de prevalentie en incidentie van COPD in de algemene populatie. Hier vonden we een hogere incidentie van COPD bij mannen en rokers.

Hoofdstuk 3 omvat de genetisch epidemiologische studies, waaronder de erfbaarheid en genoom-brede associatiestudies (**hoofdstukken 3.1** en **3.2**), epigenoom-brede associatiestudies (**hoofdstukken 3.3** en **3.4**) en tot slot genoomwijde linkage scan in **hoofdstuk 3.5**. In al deze studies werden interessante genen en "pathways" ontdekt en waar mogelijk gerepliceerd. Deze bevindingen worden gezien als zeer interessante kandidaten voor verder experimenteel onderzoek.

In **hoofdstuk 4** onderzochten we, aan de hand van klinisch epidemiologisch onderzoek, de relatie tussen een gedaalde longfunctie en het risico op klinische eindpunten. In **hoofdstuk 4.1** konden we aantonen dat een hogere "longslagader tot hoofdslagader ratio" het risico op mortaliteit bij patiënten met COPD verhoogt vooral bij individuen met een lage diffusiecapaciteit. In **hoofdstuk 4.2** vonden we een verband tussen COPD en incident perifeer arteriëel lijden (PAD). We toonden ook aan dat personen met zowel COPD als PAD hogere sterftecijfers hadden in vergelijking met mensen zonder beide ziekten, hoewel dit verhoogd sterftecijfer niet aan PAD kon toegeschreven worden aan .

In **hoofdstuk 5** bespreken we de belangrijkste bevindingen, uitdagingen en toekomst visies.

A

List of publications

PhD portfolio

Words of gratitude

About the author

JOURNAL PUBLICATIONS

1. Davies G, Lam M, Harris SE, ..., **Terzikhan N.** et al. Study of 300,486 individuals identifies 148 independent genetic loci influencing general cognitive function. Nat Commun 2018; 9:2098.
2. Nedeljkovic I, **Terzikhan N**, Vonk JM, et al. A Genome-Wide Linkage Study for Chronic Obstructive Pulmonary Disease in a Dutch Genetic Isolate Identifies Novel Rare Candidate Variants. Front Genet 2018; 9:133.
3. **Terzikhan N**, Bos D, Lahousse L, et al. Pulmonary artery to aorta ratio and risk of all-cause mortality in the general population: the Rotterdam Study. Eur Respir J 2017; 49.
4. **Terzikhan N**, Sun F, Verhamme FM, et al. Heritability and genome-wide association study of diffusing capacity of the lung. Eur Respir J 2018; 52.
5. **Terzikhan N**, Verhamme KM, Hofman A, et al. Prevalence and incidence of COPD in smokers and non-smokers: the Rotterdam Study. Eur J Epidemiol 2016; 31:785-792.
6. Vincent Chouraki, Sven J. van der Lee, Benjamin Grenier-Boley, ..., **Terzikhan N.** et al. Genome-wide Association Study Links APOEε4 and BACE1 Variants with Plasma Amyloid β Levels. 2017.
7. Wyss AB*, Sofer T*, Lee MK*, **Terzikhan N***, et al. Multiethnic meta-analysis identifies ancestry-specific and cross-ancestry loci for pulmonary function. Nat Commun 2018; 9:2976.
8. Xu J, Bartz TM*, Chittoor G*, Eiriksdottir G*, Manichaikul, AW*, Sun, F*, **Terzikhan, N*** et al. Meta-analysis across Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) consortium provides evidence for an association of serum vitamin D with pulmonary function. Br J Nutr 2018:1-12.
9. Xu J, Gaddis NC*, Bartz TM*, Hou R*, Manichaikul AW*, Pankratz N*, Smith AV*, Sun F*, **Terzikhan N***. et al. Omega-3 Fatty Acids and Genome-wide Interaction Analyses Reveal DPP10-Pulmonary Function Association. Am J Respir Crit Care Med 2018.
10. **Terzikhan N**, Lahousse L, Verhamme KMC, Franco OH, et al. Chronic Obstructive Pulmonary Disease is associated with an increased risk of peripheral artery disease and mortality, accepted in the *Eur Resp J Open Research 2018.*

Manuscripts

11. Karimi L, Lahousse L, **Terzikhan N**, et al. ADRB2 polymorphisms and risk of COPD exacerbations: the Rotterdam Study, 2018.
12. van der Plaat D, Vonk JM*, **Terzikhan N***, et al. Occupational exposures affect lung function via methylation of genes regulating expression, 2018.

13. **Terzikhan N***, Xu H, Bracke KR, et al. Epigenome-wide association study on diffusing capacity of the lung (DLCO): replication and meta-analysis, 2018.
14. **Terzikhan N**, Bos D, et al. Stronger association between pulmonary gas exchange and cerebral circulation in dyspneic women, 2018.
15. Clark DW, .. , **Terzikhan N** et al. Effects of autozygosity on a broad range of human phenotypes, 2018.

Editorials dedicated to our work

16. Hobbs BD, and Cho MH. Dissecting respiratory disease heterogeneity through the genetics of diffusing capacity. Eur Respir J 2018; 52.
17. Zouk AN, and Wells JM. In Rotterdam, size really does matter: implications of pulmonary artery enlargement on mortality. Eur Respir J 2017; 49.

European Food Safety Authority (EFSA) supporting publications

18. **Terzikhan N**, Doets EL, and Vonk Noordegraaf-Schouten M. Extensive literature search and review as preparatory work for the evaluation of the essential composition of total diet replacement products for weight control. EFSA Supporting Publications 2015.
19. Eeuwijk J, Oordt A, **Terzikhan N**, et al. Literature search and review related to specific preparatory work in the establishment of Dietary Reference values for Niacin, Biotin and Vitamin B6. EFSA Supporting Publications 2012.

PHD PORTFOLIO

Name PhD candidate: ir. Natalie Terzikhan
Erasmus MC department: Epidemiology
UGent department: Respiratory Medicine
Research schools: NIHES (NL) and Doctoral School (BE)
Promotors: Prof. dr. Guy Brusselle, Prof. dr. Bruno Stricker
Co-promotors: dr. Katia Verhamme, dr. Lies Lahousse

PhD training	Year	ECTs
General courses		
Master of Science in Health Sciences (Genetic Epidemiology, NIHES)	2015-2017	120
Doctoral School – Ghent University	2014-2018	---
Research Integrity	2016	0.3
Employability outside the academia	2018	1.0
International conferences/consortium meetings		
U4 meeting – Stockholm, Sweden	2015	3.0
European Respiratory society (ERS) congress – Amsterdam, the Netherlands	2015	1.5
ERS congress – London, UK	2016	1.5
HD-READy consortium meeting – Rotterdam, the Netherlands	2017	2.0
CHARGE meeting – New York, USA	2017	0.6
ERS congress – Milan, Italy	2017	1.5
CHARGE meeting – Boston, USA	2017	0.6
CHARGE meeting – Rotterdam, the Netherlands	2018	0.6
ERS congress – Paris, France	2018	1.5
American Society of Human Genetics (ASHG) congress- San Diego, USA	2018	0.6
Seminars, symposia and research visits		
Molecular epidemiology research meetings	2014-2018	2.0
Epidemiology research seminars	2014-2018	1.0
Nutrition and epigenetics: a focus at B vitamins	2018	0.3
Frequent research visits to the University of Gent	2014-2018	1
Research visit to the University of Groningen	2016	0.5
Presentations		
ERS congress – Amsterdam, the Netherlands – poster presentation	2015	
ERS congress – London, UK – poster presentation	2016	
HD-READy consortium meeting – Rotterdam, the Netherlands – oral presentation	2016	
CHARGE meeting – New York, USA- poster presentation	2017	
ERS congress – Milan, Italy – poster discussion	2017	
St. Franciscus Gasthuis, Rotterdam, the Netherlands – oral presentation	2017	

ERS congress – Paris, France – oral presentation	2018	
ASHG congress – San Diego, USA - poster presentation		
Teaching activities		
Supervising students		
Junior Med School: Salma Bouahiri & Aiswarya Govindara Title: *Luchtwanddike bij deelnemers van de ERGO-studie*	2017	0.8
Medical student: Mirjam van Elderen Title: *Adherence to the Dutch dietary guidelines and its association with poor lung function and mortality risk in COPD: the Rotterdam Study*	2018	1.5
Other teaching activities		
Teaching assistant- Principle of Research in Medicine and Epidemiology	2015	0.2
Teaching assistant- Pharmaco-epidemiology and drug safety	2015-2018	0.2
Lecturer: Sustainable development goals	2018	0.2
Scholarships and awards		
Young investigators sponsorship- (ERS congress)	2018	
Erasmus MC grant for 2018	2018	
Other		
Member of the seminar committee	2015-2016	

WORDS OF GRATITUDE

"If I had seen further, it is by standing on the shoulders of giants". These words by Isaac Newton say it all. This thesis would not have been possible without the support of many people, especially those giants:

Dear **professor Brusselle**, thank you so much for giving me the opportunity to be part of your group. You are not only the best professor I met so far, you are a leader and the best role model. You truly believe in the power of collaboration. I have seldom seen a person that combines the knowledge from experimental and epidemiological research so perfectly as you do. Thank you so much for believing in me and for the luxury that we (your PhDs) had. You have been always available, no matter how busy your agenda was. Thank you for being such a great promotor and inspiration.

Dear **professor Stricker**, I was always impressed by your knowledge and didactic skills. During my PhD, you have been of great support to me. Thank you so much for giving me the freedom to work independently. You have always been eager to teach your students the right concepts of epidemiology, not only by organizing weekly meetings, but also by arranging book discussions and paper reviews. Thank you so much for everything, I learned a lot from working with you.

Dear (professor) dr. Verhamme, dear **Katia**, it is always very impressive to see how humble you are. You are a highly valuable additional to my supervision team with your experience and high sense of responsibility towards your students. For me, you are the person with a helicopter view, who sees things that others do not even consider as important. In every single paper you provided me with very unique and valuable comments from very different angles. Thank you so much for your time and effort to make me better.

Dear (professor) dr. Lahousse, dear **Lies,** in the last four years you have been my daily support and supervisor. I have known you as the punctual, very hard working and dedicated person. I am happy to be your first PhD student, we have been a great team together. I have never waited longer than 2 days to receive feedback from you. Thank you so much for being such a dedicated co-promotor. You are a strong advocate of female presence in leading positions, and I think you are already a role model in that.

I would like to extend my deepest gratitude to the reading and plenary committee members **prof. dr. Arfan Ikram**, **prof. dr. Marike Boezen**, **prof. dr. Guy Joos**, **prof. dr. ir. Cock van Duijn**, **prof. dr. Louise Wain.**

Also special thanks to you **Hieab**. I couldn't wish myself a better introduction to the field of genetic epidemiology. Your passion for this work and dedication gives me the power to become a better scientist with correct work ethics. I remember you saying, "you don't have to reinvent the wheel, we should build on each other's knowledge", and that's what I call leadership. You are such a talented man with a vision and I believe

in you. Hieab, your support to me in the last period is of unmeasurable value. Thank you for being the critical eyes, the best friend and the partner in crime for me.

My dear paranimphs: **Brenda**, I have known you since the first day I started this PhD. You were always ready to help everyone, without expecting anything in return. We shared lots of happy and sad moments, and listened to each other's concerns. Your support during the last year for me is extremely valuable. Thank you so much for being there for me, I really feel privileged to have a person like you in my life.

Wasma, we have known each other from Wageningen times, for 10 years now. I think you would agree with me that it is unbelievable how many similarities we share: We speak the same languages (also literatly), we share similar passion for music and have comparable opinion about almost everything. The fact that we also share the same birthday forces me to start reconsidering believing in horoscopes :P. Our conversations can easily take hours. It is amazing to see how devoted you are to have a positive impact on this world. Wasma, you are like an artist. You are always in the spotlights wherever you are, and this will help you to get closer to your dream of making this world a better place. Thank you Wasma, you make me feel strong and proud every time again.

My roommates, Eliza and Emmely. How lucky we are to have shared a room together with real walls :). **Eliza**, you have such a big heart. I admire your work attitude and commitment. It is amazing to see the hunger for knowledge in your eyes. Your perfectionism and positively are your strengths. **Emmely**, you are not only a great and passionate doctor, you are a talented epidemiologist as well. We had very interesting discussions about study designs and ethics. You have done an incredible work in validating asthma data in the Rotterdam Study. Thank you for listening to my complains every day, and for being a great friend.

Dear **Bart Coucquyt,** Bart the great, I really appreciate your help during the past four years. It was a true honor for me to work with you. Thank you!

I would like to express my deepest appreciation to the participants of the Rotterdam Study, to my colleagues behind the scenes **Jolande, Frank, Nano, René, Maurice, Timo, Henny, Gabriëlle** and **Erica**, and to all my colleagues at the study centre in Ommoord, especially the "lung function" ladies **Bernardina**, **Edith** and **Henriëtte**. Without you there is no Rotterdam Study and no research, your work is highly appreciated.

A special thanks to those who always believed in me and criticized me most: **Emile Engel**, **Frank Planting**, **Louis Wagenaar, Eveline Verheij, Naomi Bakels, Ellen Kampman**, **Nicole de Roos, Nicolette de Jong** and ofcourse to my dearest family:

Papa

حبيبي بابا، أشكرك على ثقتك بي و تشجيعك لي. انت قدوتي و مثلي الأعلى. اقدر التضحية التى قدمتها لنا انت و ماما لحياة أفضل. يعجبني فيك عقلك المنفتح و يعجبني فيي انني أشبهك في أغلب الصفات.
كثيرون كانوا يذكرون جدتي، و يَرَوْن لي عن مدّا تشابه صفاتنا عن حب العلم و العمل الدؤوب. اشكر جدتي التى لم اعرفها و أتمنى ان أكون قد حققت جزئ من الذي كانت تتمنى ان تحققه هي.
أشكرك ببساطة لأنك ابي و انتمنى ان أكون قد نجحت برفع اسم عائلتنا عاليا. و أنتمنى أيضا ان أكون قد أوفيت بجزئ بسيط من ما أعطيتنا إياه. ها و انت قد (ختيرت)، أتمنى لك الصحة و العمر الطويل.

Mama

Սիրելի մամի, իմ հրեշտակս: Շնորհակալություն ամեն ինչի համար: դուն մեր տան սյունն ես: Շնորհակալություն մեզ միասին պահելու համար: Դուն իմ ամենալավ օրինակն ես: ես միշտ կը մնամ ձեր փոքրիկ նաթալին. Սիրով:

Lieve **Caroline, Jean, Philip** en **Michael**, wat zijn jullie samen een mooi en een warm gezinnetje. Carawetta ☺, samen hebben we mooie herinneringen gedeeld, muziek gemaakt, gebasketbald, gelachen en gehuild. Bedankt voor alle mooie en betekenisvolle momenten. Ik wens je veel geluk en gezonde jaren toe samen met Jimmy en de kinderen. Jullie kunnen altijd op me rekenen.

ABOUT THE AUTHOR

Natalie Terzikhan was born in Aleppo, Syria, for a Syrian father and an Armenian mother. She grew up bilingual speaking both Arabic and Armenian. She started playing piano and singing at the age of 4. At the age of 10, Natalie won the first price of the Syrian National Music Competition for her age category. Two years later she was selected to sing at the Syrian Music Festival on the theater inside the iconic citadel of Aleppo. In 2000, the family left Syria and moved to The Netherlands, where Natalie started an active career as a musician under the supervision of the pianist and conductor Emile Engel. She won several prices of which the best player of the modern classical music during the princess Christina Competition. Natalie was then enrolled in the young talented class at the conservatory of Arnhem and later in Amsterdam, under the supervision of the pianist Mila Baslawskaja. In 2006, Natalie decided to change career and move to academia to do Nutrition and Health studies, and obtained a bachelor degree in 2009 and a master degree in 2011 from Wageningen University. After graduation, Natalie worked for several years as an Epidemiologist. In 2014, she started a joint PhD program at the University of Gent in Belgium and The Erasmus University in the Netherlands, under the supervision of prof. Guy Brusselle and prof. Bruno Stricker. Coupled with her PhD Natalie obtained her second master degree in Genetic Epidemiology from the National Institute of Health Sciences. After her PhD Natalie will continue working at the department of Epidemiology of the Erasmus MC as a postdoctoral researcher and a manager. Natalie is a proud mother of Melina (7) and Ella (5).