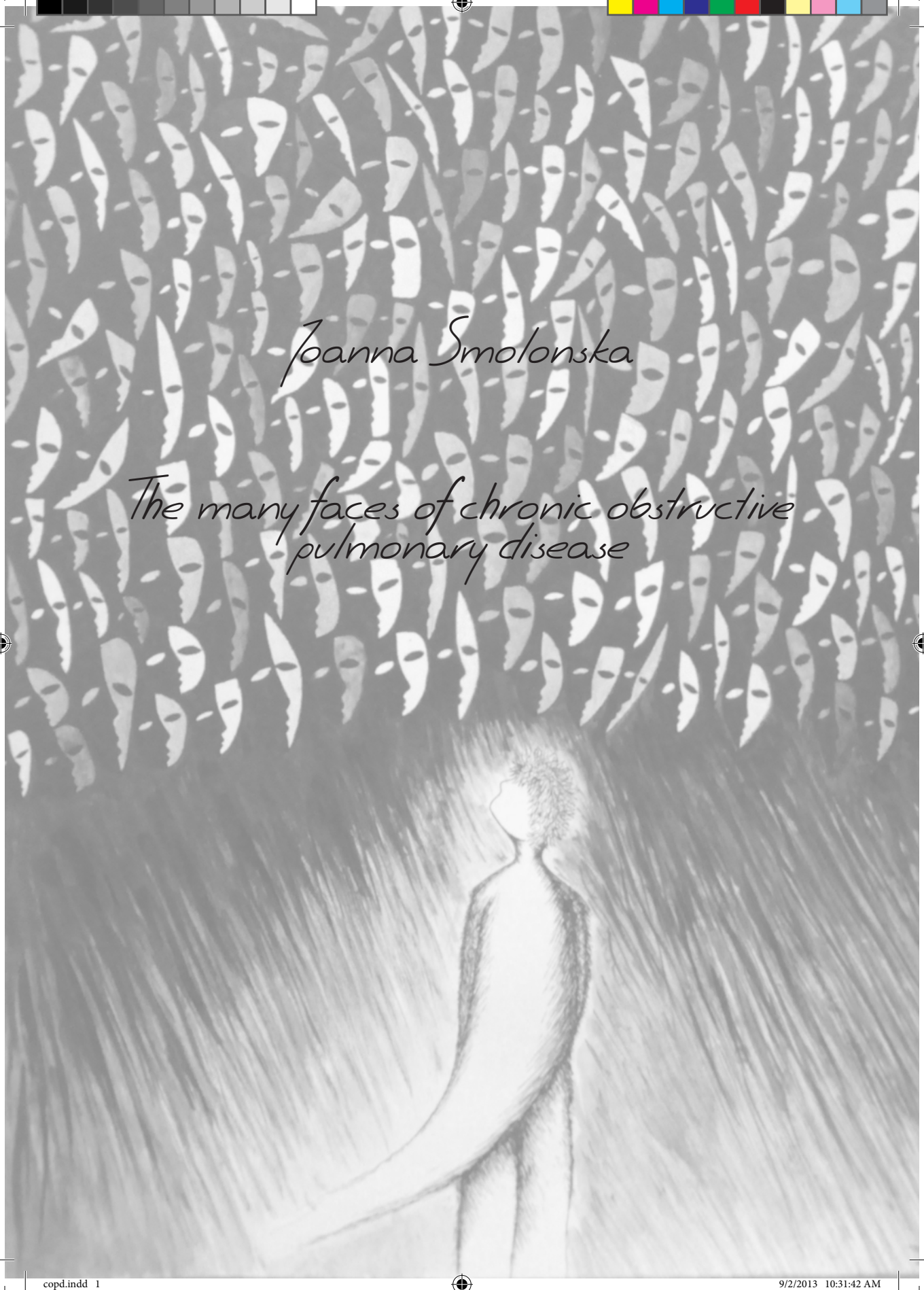
The many faces of chronic obstructive pulmonary disease

Smolonska, Joanna



The many faces of chronic obstructive pulmonary disease
Joanna Smolonska

Phd thesis – Department of Genetics
University Medical Center Groningen, University of Groningen
Groningen, The Netherlands

This study was financially supported by COPACETIC consortium and GUIDE Research Institute, UMCG, Groningen, the Netherlands..

Printing of the thesis was financially supported by University of Groningen, GUIDE and Department of Genetics, UMCG.

Cover design and book layout: Lovebird design and printing solutions
www.lovebird-design.com

Printed by: EIKON Plus

ISBN: 978-90-367-6486-5
ISBN (electronic version): 978-90-367-6485-8

RIJKSUNIVERSITEIT GRONINGEN

The many faces of chronic obstructive pulmonary disease

Proefschrift

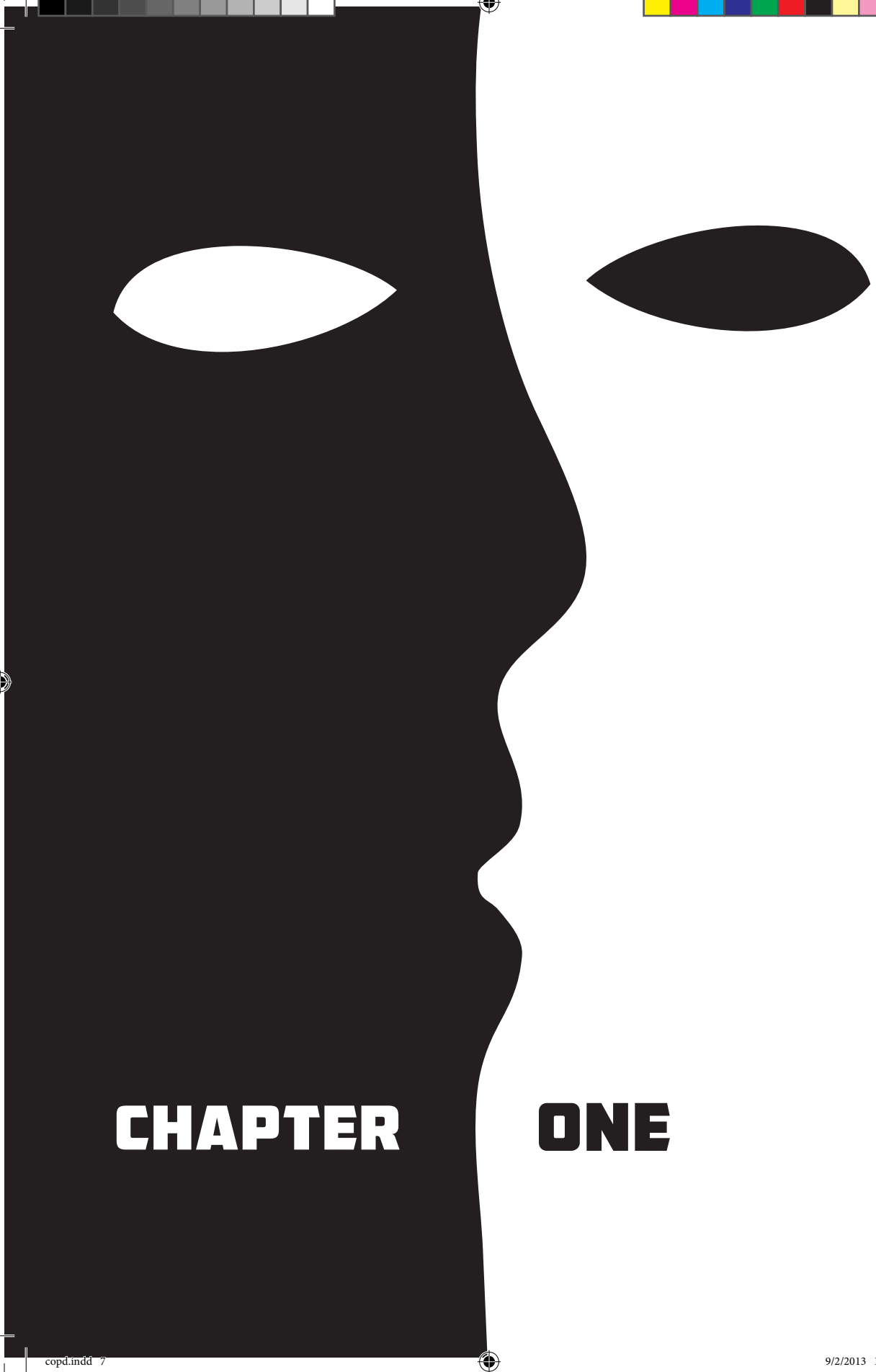
ter verkrijging van het doctoraat in de
Medische Wetenschappen
aan de Rijksuniversiteit Groningen
op gezag van de
Rector Magnificus, dr. E. Sterken,
in het openbaar te verdedigen op
maandag 9 oktober 2013
om 12.45 uur

door

Joanna Smolonska
geboren op 26 juni 1984
Wrocław, Polen

Promotores:	Prof. dr. H.M. Boezen
	Prof. dr. C. Wijmenga
Beoordelingscommissie:	Prof. dr. D. Jarvis
	Prof. dr. P.I.W. de Bakker
	Prof. dr. R.K. Weersma

Contents



CHAPTER ONE

Introduction

What is COPD?

Chronic obstructive pulmonary disease (COPD) is a complex disorder classified as 3rd cause of the death worldwide[1]. The disease is progressive and so far there is no cure. It can be characterized as accelerated lung function decline, ultimately leading to death. But COPD is more than that, it involves inflammation in the lungs, muscle weakness and weight loss decreasing constantly the quality of life of patients[2]. Increased chances to develop COPD come with cigarette smoking mainly, but also air pollution plays an important role. However, only 10–15% of smokers develop COPD[3]. Main phenotypes of COPD are airway obstruction, emphysema and chronic bronchitis. In this thesis the focus is limited to airway obstruction and emphysema. Airway obstruction is a result of inflammation and remodeling of the airways, which leads to airway wall thickening and in consequence to airflow limitation[4]. The diagnosis is made based on spirometry, which measures forced expiratory volume in one second (FEV_1) and forced vital capacity (FVC). The ratio between them is used for COPD diagnosis. According to GOLD criteria[5] $FEV_1/FVC < 0.7$ is a stage I COPD. Emphysema is a result of alveoli enlargement and air trapping resulting in serious tissue damage. It can be diagnosed using X-rays and nowadays mostly with computer tomography (CT) scan of the chest. Diagnosis based on spirometry parameters does not capture individuals with mild emphysema, who do not display decreased lung function yet. Using spirometry alone also does not allow distinguishing between COPD phenotypes, which are airway obstruction alone, emphysema alone and a mixed phenotype with both airway obstruction and emphysema present.

Genetics of COPD – pre-GWAS era

There is also genetic component involved in COPD, as shown in family studies[6]. Linkage analysis identified a locus on chromosome 2q[7], that was later mapped to *SERPINE2* gene[8]. However, this finding was not confirmed by a subsequent larger study[9]. For many years the only genetic factor known to predispose to early-onset COPD was *SERPINA1* gene[10]. Mutation in that gene leads to α-1 antitrypsin deficiency, which results in emphysema. However, the percentage of *SERPINA1* caused COPD was very low (1–2% of COPD cases)[10], therefore it was clear that other genes must be involved as well. For over two decades researchers were looking for these genes using a candidate gene approach. The principle of this method is to select a candidate for further investigation, based on what is already known, so the gene with a plausible function. In case of COPD three main pathways were targeted and most intensively studied: inflammatory, protease/anti protease and oxidative stress pathways. Despite many efforts candidate gene studies performed in COPD resulted in spurious results, making it

difficult to say which gene is associated with COPD and which one is not. Also, they mainly focused on airway obstruction; studies on genetics of emphysema were limited at the time. Studies on genetics of specific phenotypes of COPD, like for example mixed phenotype with airway obstruction and emphysema are lacking.

Genome-wide association studies

With the development of new technology enabling genome-wide association studies (GWAS) a real breakthrough in genetics of complex disorders came. The rationale behind GWAS is straight forward. Genome-wide genotyping is performed and given the presence of linkage disequilibrium (LD), not all the SNPs need to be typed. LD means that many SNPs are dependent on each other, so the genotype of SNP1 is sufficient to predict genotypes of SNP2 and SNP3 (see Figure 1).

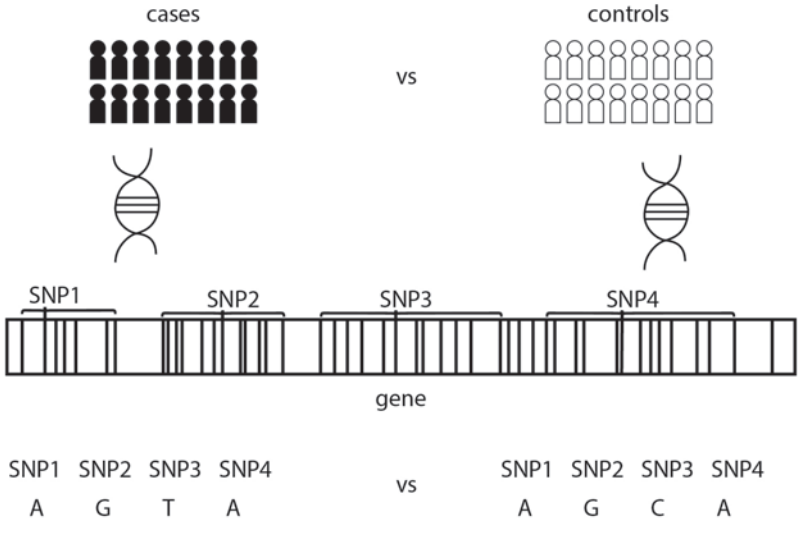


Figure 1. Genome-wide association study (GWAS) workflowLarge number of cases and controls are genotyped. Since there is linkage disequilibrium not all the SNPs need to be genotyped. Here four SNPs need to be genotyped to cover nearly entire gene. Next, genotype distributions are compared between cases and controls in the association test.

If genetic component is involved in a disease, one can expect that patients will have different minor allele frequency (MAF) from controls for SNPs that tag loci related to the disease. Therefore genotyping of a large number of cases and controls is required in order to identify those differences (Figure 2). This number increases as the effect size, odds ratio (OR) decreases. Apart from large numbers needed to detect (small) genetic effects, appropriate and adequate phenotyping of the individuals included in the study is necessary. Mistakes in

determining case-control status can result in spurious and false findings. Once genotyping is done, another crucial and sensitive step is the quality control. It is very important to remove badly genotyped SNPs and samples of lower quality and confidence from the dataset prior to association testing. In studies combining different population factors like population stratification, presence of ethnic outliers, duplicates, relatives, and samples with mismatching gender should be carefully considered during the process of quality control2. This is in fact a crucial step and the more thoroughly it is performed, the better and more confident the results will be. Due to large number of tests performed in such a study the chances of false positive findings are high. There are two ways to minimize this risk: multiple testing correction and replication study. Multiple testing correction is usually done using Bonferroni correction, which determines the significance cut-off by simply dividing nominal level of significance (0.05) by number of tests performed. However, it has been proposed to use a fixed cut-off of 5*10e-08 for GWA studies[11]. Replication study involves independent samples, in which genotyping of a subset of SNPs is performed. This step is very important, as it allows confirming the findings in other cohorts and/or populations.

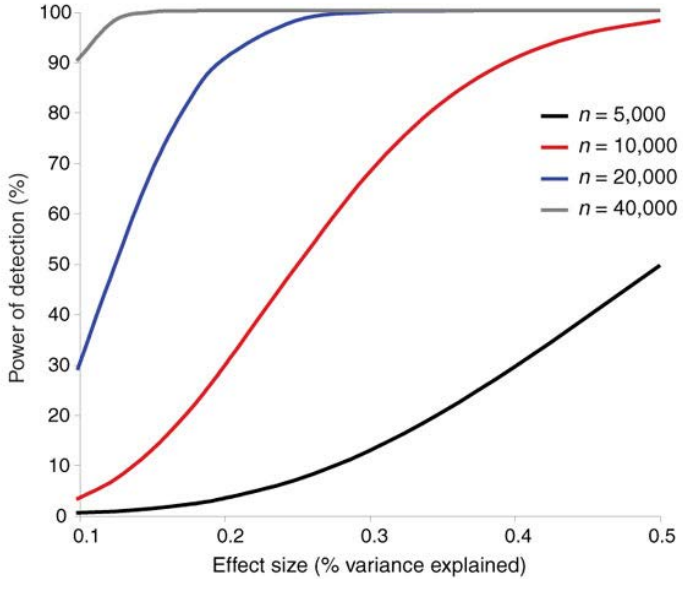


Figure 2. Relation between power of the GWA study and sample size.

COPD genetics – GWAS era

The first GWAS on COPD defined as airway obstruction was published in 2009[12] and reported two loci as associated with the disease: *CHRNA3/5* and *HHIP*. The *CHRNA3/5* locus became controversial, as it was found to be associated with lung cancer[13, 14] and smoking addiction[15] and debate on whether it is independently associated with COPD or acts

via cigarette smoking habits is still ongoing[16]. *HHIP* was previously identified in GWAS on pulmonary measures. It is an interesting finding, as hedgehog interaction protein is known to be involved in lung development. Second GWAS on COPD identified *FAM13A* gene[17], which levels were shown to increase in hypoxia[18]. Same group performed one more GWAS study, in which locus on chromosome 19q13 was identified[19]. This locus contains four genes: *RAB4B, EGLN2, MIA* and *CYP2A6*. Same locus was also associated with smoking behavior[20] and lung cancer[21]. Siedlinski studied smoking behavior of COPD patients and identified two candidate loci associated with age at smoking initiation: 2q21 and 6p21[22].

So far, only one GWAS on emphysema was published[23] and reported *BICD1* gene. This gene is involved in telomere length regulation[24] which is an interesting finding supporting the hypothesis of COPD being a disease of prematurely and accelerated aging lung[25]. There are no GWA studies on COPD patients with airway obstruction and emphysema while we cannot exclude the possibility that different genetic factors underlie this phenotype when compared to pure airway obstruction.

Lung function level, decline and genetics

Lung function is simply a volume of air that can be inhaled and exhaled by the lungs. It is a trait that is heritable in 30–47%[26], however, can be also predicted based on gender, height, weight, age and ethnicity. The lung development starts early in life and continues until the age of 20, when usually the maximal level is reached and plateau phase can be observed. Then around the age of 30–40 lung function level starts to drop, which is an aging related process known as lung function decline. However, in a disease state, like asthma or COPD, the observed decline is more rapid and progressive leading to premature lung function loss.

Multiple studies were conducted to investigate genetics of lung function level. The first GWAS on lung function measures was published in 2007 and used one of the first microarrays available, Affymetrix 80k[27]. Main findings included *GSTO2* associated with mean FEV_1 and mean FVC, *ADARB1* associated with FEV_1/FVC ratio and *SNRPN* associated with FEV_1 decline. Two years later, the same group published results on pulmonary measures, where Affymetrix 550k platform was used, providing a better coverage of human genome[28]. In this study hedgehog-interacting protein (*HHIP*) gene was associated with FEV_1/FVC ratio. In 2009 another study was published by Repapi et al, confirming *HHIP* and identifying three new loci associated with FEV_1 (*GSTCD, TNS1* and *HTR4*) and three with FEV_1/FVC (*AGER, THSD4* and *DAAM2*)[29]. Around the same time a meta-analysis was performed, including nearly 21,000 individuals in a discovery sample and over 16,000 in a replication set[30]. Three new loci were identified, *GPR126* and *ADAM19* were associated with FEV_1/FVC ratio and *INTS12* was associated with FEV_1 level. In 2011 a large GWAS on pulmonary measures was published, including over 48,000 individuals in

a discovery set and over 46,000 in a follow up stage. This study reported 16 new loci (Figure 3)[31].

Only two studies investigated genetics of lung function decline so far. Imboden et al[32] performed a GWAS study on lung function decline in adults with and without asthma. No loci were identified on a genome-wide significant level. Moreover, SNPs previously associated with pulmonary function levels were not associated in this study with the decline. Another study by Hansel et al focused of determining genetic factors responsible for decline in smokers with mild COPD[33]. Two loci were identified with genome-wide significance level however, none of them was replicated. Locus on chromosome 10 contained *TMEM26* and *ANK3* genes and locus on chromosome 14 contained *FOXA1* gene. *TMEM26* and *FOXA1* were shown to be expressed in airway epithelium and parenchyma and *ANK3* in alveolar macrophages. Expression levels were dependent on both, COPD status and lung function.

Aim of the thesis

The general aim of this thesis is to identify genetics underlying chronic obstructive pulmonary disease and related phenotypes. Since COPD is a disease of accelerated lung function decline, we also studied genetics of it, as it could provide some answers. We also focused on specific subphenotypes of COPD, namely airway obstruction and emphysema, as frequently they co-occur. Last, but not least asthma was also a scope of our studies because of some resemblance to COPD.

In **Chapter 2** we tried to summarize the two decades of candidate gene studies and clarify the ambiguous results by performing meta-analyses on 20 polymorphisms in 12 genes that were studied in at least three independent publications.

In **Chapter 3** we describe a GWA study on COPD. We investigated the genetics of airway obstruction and emphysema, as well as an overlap between them, as in the study population we used – the NELSON cohort there are many individuals, who display features of both phenotypes. For replication we used European cohorts, gathered through COPACETIC consortium.

Chapter 4 describes GWA study on lung function decline in a population of heavy smokers. For replication we used four population based cohorts: Doetinchem, Vlagtwedde/Vlaardingen, British 1958 Birth cohort (B58C) and SAPALDIA. Additionally, we investigated gene-by environment interactions in Doetinchem and Vlagtwedde/Vlaardingen cohorts. We identified SNPs interacting with COPD status, smoking and gender.

In **Chapter 5** we investigated shared genetics of asthma and COPD to check validity of the 'Dutch hypothesis' on a genome-wide level. Using the genome-wide genotyping data that was available for cohorts of asthma and COPD patients and controls we performed a meta-analysis. Subsequently, we performed two-staged replication and eQTL study on the most associated SNPs.

In Chapter 6 we describe the first study aiming to identify gene-gene interactions in a genome-wide setup for COPD. We used the NELSON cohort for the discovery and LifeLines cohort for replication.

Chapter 7 discusses the approaches taken in this thesis and future perspectives of genetics of COPD and complex disorders in general.

Study populations

The discovery cohort in all studies described in this thesis was the NELSON cohort. The NELSON cohort recruited Caucasian males of Dutch descent with smoking history of at least 20 pack-years to include them in lung cancer screening trial. University Medical Center Groningen and University Medical Center Utrecht were the study centers where patients were recruited and examined. Each participant underwent spirometry and computer tomography (CT) scan of the chest to enable diagnosis of COPD. Follow-up measurements were taken 3-4 years later. Blood sample was obtained from each individual to enable DNA isolation.

Replication cohorts used included:

1. the Vlagtwedde/Vlaardingen cohort, a population-based cohort that recruited individuals from Vlagtwedde, a rural area, and Vlaardingen, an urban area in the Netherlands. Participants were followed for over 25 years with lung function measurements every 3 years.
2. The Doetinchem cohort, a Dutch population-based cohort followed up for 10-15 years with lung function measurements taken every 5 years.
3. The Copenhagen City Heart Study (CCHS), is a Danish general population cohort. We used the third examination of the CCHs from 1991 through 1994, where 1,212 COPD cases were available and 3,656 controls have been matched randomly
4. Danish Lung Cancer Screening Trial (DLCST), is a cohort of heavy-smokers similar to the NELSON cohort, recruiting heavy-smokers. For this cohort no follow-up measurements were available.
5. Burden of Obstructive Lung Disease (BOLD) is a Polish cohort of 147 COPD patients and 355 matched healthy controls
6. Rucphen is a cohort comprising 158 individuals, recruited from the genetic isolate in the Netherlands, a small village called Rucphen.
7. GLUCOLD (Groningen Leiden Universities Corticosteroids in Obstructive Lung Disease) study recruited 114 COPD cases with GOLD stage II or above.
8. LifeLines study consists of a population based sample of Northern part of the Netherlands including Caucasians of Dutch decent with detailed phenotyping information, including lung function measurements.

References

1. WHO. The global tobacco crisis. 2008
2. Agusti AG. Systemic effects of chronic obstructive pulmonary disease. Proc Am Thorac Soc 2005;**2**:367-370.
3. Fletcher C, Peto R. The natural history of chronic airflow obstruction. Br Med J 1977;**1**:1645-1648.
4. O'Donnell RA, Peebles C, Ward JA et al. Relationship between peripheral airway dysfunction, airway obstruction, and neutrophilic inflammation in COPD. Thorax 2004;**59**:837-842.
5. Pauwels RA, Buist AS, Calverley PM, Jenkins CR, Hurd SS. Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease:NHLBI/WHO Global Initiative for Chronic Obstructive Lung Disease (GOLD) Workshop summary. Am J Respir Crit Care Med 2001;**163**:1256-1276.
6. Palmer LJ, Celedon JC, Chapman HA, Speizer FE, Weiss ST, Silverman EK. Genome-wide linkage analysis of bronchodilator responsiveness and post-bronchodilator spirometric phenotypes in chronic obstructive pulmonary disease. Hum Mol Genet 2003;**12**:1199-1210.
7. Silverman EK, Mosley JD, Palmer LJ et al. Genome-wide linkage analysis of severe, early-onset chronic obstructive pulmonary disease: airflow obstruction and chronic bronchitis phenotypes. Hum Mol Genet 2002;**11**:623-632.
8. DeMeo DL, Mariani TJ, Lange C et al. The SERPINE2 gene is associated with chronic obstructive pulmonary disease. Am J Hum Genet 2006;**78**:253-264.
9. Chappell S, Daly L, Morgan K et al. The SERPINE2 gene and chronic obstructive pulmonary disease. Am J Hum Genet 2006;**79**:184-186.
10. Lieberman J, Winter B, Sastre A. Alpha 1-antitrypsin Pi-types in 965 COPD patients. Chest 1986;**89**:370-373.
11. Pe'er I, Yelensky R, Altshuler D, Daly MJ. Estimation of the multiple testing burden for genomewide association studies of nearly all common variants. Genet Epidemiol 2008;**32**:381-385.
12. Pillai SG, Ge D, Zhu G et al. A genome-wide association study in chronic obstructive pulmonary disease (COPD): identification of two major susceptibility loci. PLoS Genet 2009;**5**:e1000421.
13. Hung RJ, McKay JD, Gaborieau V et al. A susceptibility locus for lung cancer maps to nicotinic acetylcholine receptor subunit genes on 15q25. Nature 2008;**452**:633-637.
14. Amos CI, Wu X, Broderick P et al. Genome-wide association scan of tag SNPs identifies a susceptibility locus for lung cancer at 15q25.1. Nat Genet 2008;**40**:616-622.
15. Thorgeirsson TE, Geller F, Sulem P et al. A variant associated with nicotine dependence, lung cancer and peripheral arterial disease. Nature 2008;**452**:638-642.
16. Budulac SE, Vonk JM, Postma DS, Siedlinski M, Timens W, Boezen MH. Nicotinic acetylcholine receptor variants are related to smoking habits, but not directly to COPD. PLoS One 2012;**7**:e33386.
17. Cho MH, Boutaoui N, Klanderman BJ et al. Variants in FAM13A are associated with chronic obstructive pulmonary disease. Nat Genet 2010;**42**:200-202.
18. Chi JT, Wang Z, Nuyten DS et al. Gene expression programs in response to hypoxia: cell type specificity and prognostic significance in human cancers. PLoS Med 2006;**3**:e47.
19. Cho MH, Castaldi PJ, Wan ES et al. A genome-wide association study of COPD identifies a susceptibility locus on chromosome 19q13. Hum Mol Genet 2012;**21**:947-957.
20. Thorgeirsson TE, Gudbjartsson DF, Surakka I et al. Sequence variants at CHRNB3-CHRNA6 and CYP2A6 affect smoking behavior. Nat Genet 2010;**42**:448-453.
21. London SJ, Idle JR, Daly AK, Coetzee GA. Genetic variation of CYP2A6, smoking, and risk of cancer. Lancet 1999;**353**:898-899.

22. Siedlinski M, Cho MH, Bakke P et al. Genome-wide association study of smoking behaviours in patients with COPD. Thorax 2011;**66**:894-902.

23. Kong X, Cho MH, Anderson W et al. Genome-wide Association Study Identifies BICD1 as a Susceptibility Gene for Emphysema. Am J Respir Crit Care Med 2011;**183**:43-49.

24. Mangino M, Brouilette S, Braund P et al. A regulatory SNP of the BICD1 gene contributes to telomere length variation in humans. Hum Mol Genet 2008;**17**:2518-2523.

25. Lee J, Sandford A, Man P, Sin DD. Is the aging process accelerated in chronic obstructive pulmonary disease? Curr Opin Pulm Med 2011;**17**:90-97.

26. Hukkinen M, Kaprio J, Broms U et al. Heritability of lung function: a twin study among never-smoking elderly women. Twin Res Hum Genet 2011;**14**:401-407.

27. Wilk JB, Walter RE, Laramie JM, Gottlieb DJ, O'Connor GT. Framingham Heart Study genome-wide association: results for pulmonary function measures. BMC Med Genet 2007;**8 Suppl 1**:S8.

28. Wilk JB, Chen TH, Gottlieb DJ et al. A genome-wide association study of pulmonary function measures in the Framingham Heart Study. PLoS Genet 2009;**5**:e1000429.

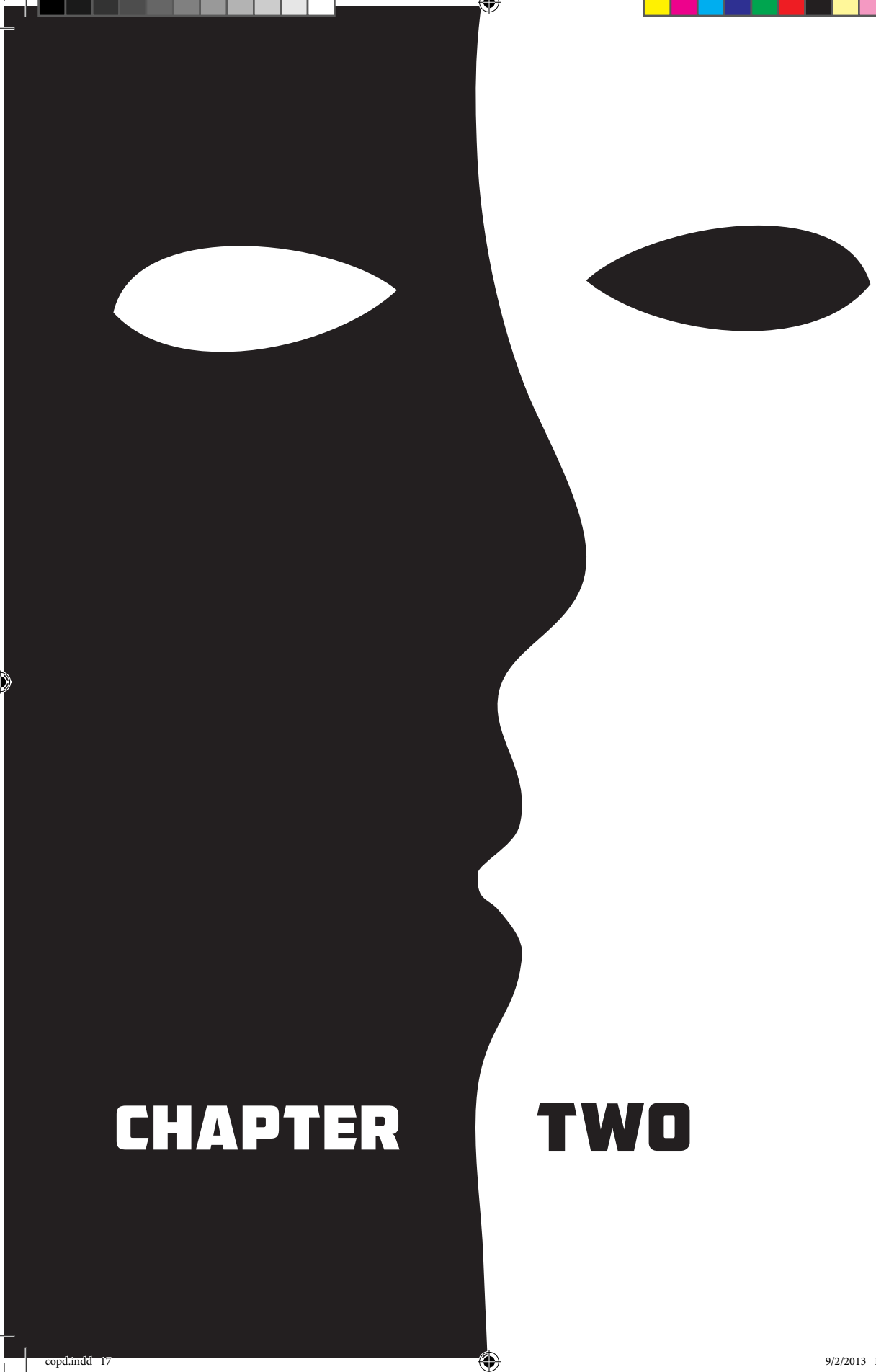
29. Repapi E, Sayers I, Wain LV et al. Genome-wide association study identifies five loci associated with lung function. Nat Genet 2010;**42**:36-44.

30. Hancock DB, Eijgelsheim M, Wilk JB et al. Meta-analyses of genome-wide association studies identify multiple loci associated with pulmonary function. Nat Genet 2010;**42**:45-52.

31. Soler AM, Loth DW, Wain LV et al. Genome-wide association and large-scale follow up identifies 16 new loci influencing lung function. Nat Genet 2011;**43**:1082-1090.

32. Imboden M, Bouzigon E, Curjuric I et al. Genome-wide association study of lung function decline in adults with and without asthma. J Allergy Clin Immunol 2012;**129**:1218-1228.

33. Hansel NN, Ruczinski I, Rafaels N et al. Genome-wide study identifies two loci associated with lung function decline in mild to moderate COPD. Hum Genet 2012.



CHAPTER TWO

Meta-analyses on suspected COPD genes – a summary of 20 years' research

Joanna Smolonska[1], Cisca Wijmenga[1], Dirkje S Postma[2], H Marike Boezen[3]

[1]Department of Genetics, [2] Department of Pulmonology, [3]Department of Epidemiology University Medical Center Groningen, University of Groningen, Groningen, the Netherlands.

Am J Respir Crit Care Med. 2009 Oct 1;180(7):618–31

AT A GLANCE COMMENTARY

Scientific Knowledge on the Subject

For many years only one COPD gene was known. Searching for additional genes was based on candidate gene studies which often were impossible to replicate. Meta-analyses can combine these studies to give them more power.

What This Study Adds to the Field

The choice of populations for a meta-analysis study or replication can strongly affect results, as at least some of the associations have opposite directions.

Abstract

Rationale: Chronic obstructive pulmonary disease (COPD) is a complex disorder with a high mortality worldwide. Studies on the role of candidate genes and their polymorphisms in COPD development have so far produced ambiguous results.

Objectives: The aim of this study was to reveal the role of COPD candidate genes using data collected in previous research.

Methods: We performed meta-analyses on 20 polymorphisms in 12 genes, after searching the PubMed and Embase databases for publications on COPD. These genes involve three main pathways associated with COPD development: the inflammatory, protease-anti-protease balance, and antioxidant pathways.

Measurements and Main Results: We obtained significant results for three *TGFB1* polymorphisms although these were based only on a few studies. The *IL1RN* VNTR polymorphism increases the risk for COPD (OR 1.7, 95%CI 1.09–2.65), whereas the *TNFA* –308 G/A polymorphism does so only in Asian populations (OR 2.01, 95%CI 1.21–3.31). The *GSTP1* I105V polymorphism was protective for COPD in Asian populations only (OR 0.69, 95%CI 0.56–0.85)

Conclusions: These results demonstrate the importance of ethnicity in identifying specific COPD genes.

Keywords: COPD, genes, polymorphism, SNP, meta-analysis, ethnicity

Introduction

Chronic obstructive pulmonary disease (COPD) is a complex disorder characterized by expiratory airway obstruction and lung tissue destruction. In 2002 more than three million people died of COPD and, according to the World Health Organization (WHO), by 2030 COPD is expected to become the third leading cause of death worldwide.1 Smoking is the major risk factor for its development, but genetic susceptibility is thought to play a determining role in the development of the disease since only a relatively small proportion of smokers (10–15%)[2] develops symptomatic COPD and differences in cigarette smoke exposure account for only a small amount of the variation seen in lung function. Family-based and twin studies have shown that lung function is heritable and COPD clusters within families.

A well-known example of a strong genetic factor predictive of the development of COPD is alpha-1-antitrypsin (AAT) deficiency carrier status. AAT deficiency is a very strong genetic trait that is related to severe, early onset COPD. The *SERPINA1* gene (coding AAT – serine protease inhibitor, of which the main target is neutrophil elastase) was for many years the only genetic factor known to be associated with COPD development. However, only 1–2% of the COPD patients inherit this AAT deficiency,[3] hence other genes are thought to play a role in most COPD cases.

Since only 10–15% of smokers develop COPD,[2] the concept of the 'genetically susceptible' smoker has been proposed, i.e. a person who will develop COPD from smoking, whereas others will not. These susceptible smokers show a premature onset of decline in lung function and a more rapid rate of decline later in life compared to non-susceptible smokers. Smoking induces an inflammatory response and oxidative stress in the airways and lung tissue. This ultimately leads to increased numbers of goblet cells, airway remodeling and tissue destruction, phenotypes that accompany COPD. It is thus likely that genetic studies on candidate genes have initially focused on these typical COPD pathways induced by smoking, in addition to studying the enzymes that detoxify cigarette smoke products. Repair mechanisms are important as a response to lung tissue injury induced by cigarette smoke, so that genes involved in the protease/anti-protease balance are also considered to be candidate genes for COPD.

The majority of genetic studies on COPD so far have been performed in relatively small samples of different ethnic populations, and resulted in unclear and sometimes contradictory results. Our aim was to summarize the current evidence for strong candidate genes and perform meta-analyses on published data in order to confirm or refute the genes' importance in COPD. The genes we have studied can be grouped into three categories according to their function: (1) inflammatory genes, (2) genes involved in the protease/anti-protease balance, and (3) antioxidant genes.

Methods

Search strategy and selection criteria

We performed a literature search in PubMed using "COPD [MeSH terms] and gene or polymorphism or pathway" for data to use in meta-analyses. We also searched PubMed and Embase querying for "COPD [MeSH] or pulmonary emphysema [MeSH] or COPD or chronic obstructive pulmonary disease [All fields] and genotype or allele or polymorphism or SNP or variant", using MeSH/EMTREE terms and free text. There were no limitations of any type in our literature searches. Based on the literature found we selected genes and their polymorphisms for which we performed a second literature search, specifically looking for publications investigating their association with COPD. We also manually searched the references of relevant publications. For the meta-analyses we included only publications that described association of a particular polymorphism in a case-control or population-based study with clearly defined COPD with no other accompanying disease. The definition of COPD cases that were included had to follow the ATS/ERS guidelines[4] or the GOLD criteria.[5] Reviews and other publications like editorials and letters were excluded (unless containing original data). In the case of publications that reported only allele frequencies, but in which the authors had checked for Hardy-Weinberg equilibrium (HWE), we calculated genotype frequencies and used them for our meta-analyses. We further checked for HWE in controls among all publications reporting genotype frequencies ~~found~~ using the 'genetics' package for R which uses χ^2 test to compare observed and expected genotype distributions. Publications deviating from HWE ($p<0.05$) were excluded. Information about the 69 publications included in our meta-analyses can be found in Table S1 in the online supplementary information.

Meta-analyses

We included only genes and polymorphisms that were described in at least three different studies in our meta-analyses. This led to 12 genes and 20 polymorphisms in the genes; these are described separately below. We investigated the following genes in relation to COPD: *TNFA, TGFB1, IL-6, IL-1B,* and *IL1RN* in the inflammatory pathways, *MMP9* in the protease/anti-protease pathway, and *GSTP1, GSTM1, GSTT1, EPHX1, SOD2* and *SOD3* in the oxidative stress pathway.

Meta-analyses were performed using R (www.r-project.org) with the rmeta package. All odds ratios (ORs) for individual studies were calculated using this package from the contingency tables. For each meta-analysis a Woolf's test for heterogeneity among publications was performed, and depending on the result, we used a fixed-effect model for non-heterogeneous studies (Mantel-Haenszel method) or a random-effect model for

heterogeneous studies (DerSimonian–Laird method). The model types are indicated in the relevant figures. For all meta-analyses a genetic model was also applied. A dominant model was used for all SNPs and a recessive model for null alleles.

Results

We performed a two-stage literature search, as depicted in Figure 1. We started with a general search for publications on COPD genetics. 694 publications were found and their abstracts were read to determine their relevance. 188 publications passed our criteria and were used to select genes and their polymorphisms for a specific literature search. Of the selected genes, 12 had sufficient numbers of studies and we included them into meta-analyses. From a total of 314 publications we selected only those fulfilling our criteria described in the methods section. We determined genotype data from those publications and checked for HWE in controls. The whole process was performed independently by two investigators (JS and HMB). All discordances were discussed, brought to consensuses and studies were included or excluded unanimously. In the end 69 publications were included in our study.

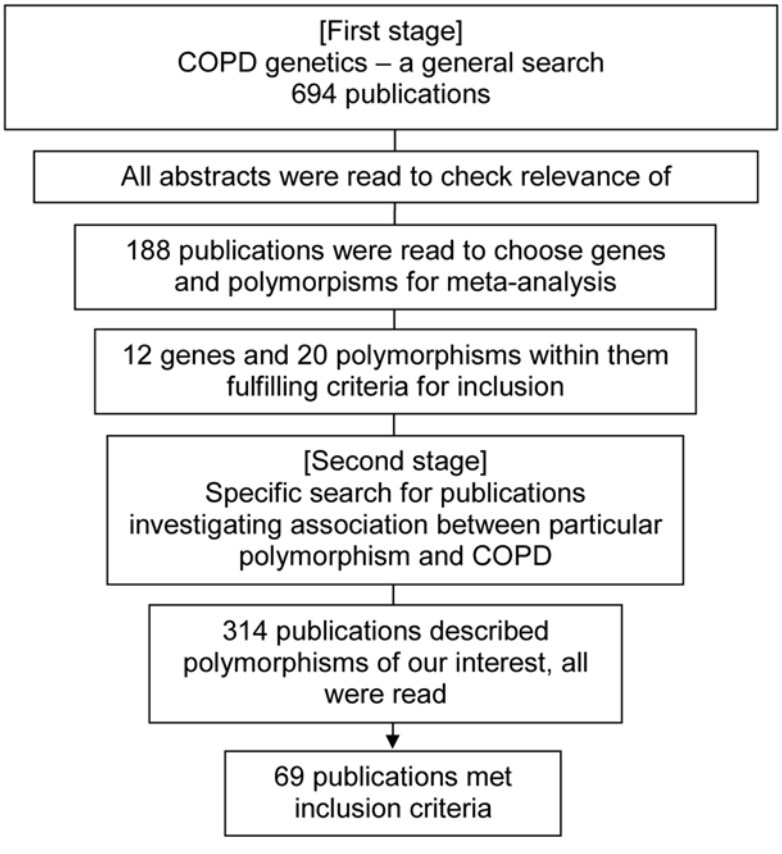


Figure 1: Flow diagram showing literature search strategy

1. Inflammatory genes

Inflammation is a continuous and ongoing process in the lungs of patients with COPD, and it is even more enhanced in severe and very severe cases.[6] Epithelial cells are activated by cigarette smoking and their mediators further activate CD8+ cells and neutrophils, which play a crucial role in inflammation in COPD.[7] Genes encoding these mediators have been studied for association with COPD.

Tumor necrosis factor alpha (TNFA)

Many studies on this group of genes have concerned tumor necrosis factor **α** (TNF**α**), the key cytokine responsible for systemic inflammation, as it drives the production and secretion of many inflammatory mediators. TNF**α** plasma levels are raised in COPD and positively correlate with COPD severity. Thus polymorphisms in *TNFA* might contribute to COPD susceptibility, and indeed, some studies have shown associations between different *TNFA* polymorphisms and COPD,[8–10] or emphysematous changes in the lungs in COPD cases.[11] However, the majority of the studies found no associations in Caucasian[12–17] or Asian[18,19] populations. Our literature search identified 21 studies from 20 publications,[8,10,14–16,18–32] including those recently published in a meta-analysis by Gingo[32] giving a total number of 3,552 cases and 3,659 controls with 1,028 cases and 1,011 controls as carriers of the A allele. We found a non-significant increase of COPD risk with the -308 polymorphism known to increase promoter activity (OR 1.06, 95% CI 0.95–1.17 (figure 2). After performing stratified analyses for Caucasians and Asians separately, the association remained non-significant, and became even weaker in the Caucasian populations (OR 1.02, 95% CI 0.92–1.14). Remarkably, the association was significant in the Asian populations with an OR of 2.01 (95% CI 1.21–3.31) (figure 2). The study by Hung resulted in an outstanding OR when compared to other studies. A sensitivity analysis showed that excluding the Hung study changed the result of this meta-analysis into non-significant (OR =1.42; 95% CI= 0.81–2.5). However, all estimates of the included studies were in the same direction.

Two other *TNFA* polymorphisms have been less well studied, one being a promoter polymorphism -238G/A, possibly increasing promoter activity. We included five publications reporting its genotype or allele frequencies[10,15,29,30,32] in our meta-analysis, which showed an overall increased risk, yet non-significant risk of COPD (OR 1.10, 95% CI 0.87–1.38) (figure 3). The other polymorphism in the TNF**α** gene is located in the first intron at position +489, and its influence on TNF**α** expression is not known.[33] We performed a meta-analysis on six studies from five publications[10,18,19,30,32] and found a non-significant increase of COPD risk (OR 1.10, 95% CI 0.92–1.33) (figure 3). It thus seems unlikely that TNF**α** is associated with COPD susceptibility in Caucasians, although larger studies are needed to elucidate its role in Asians.

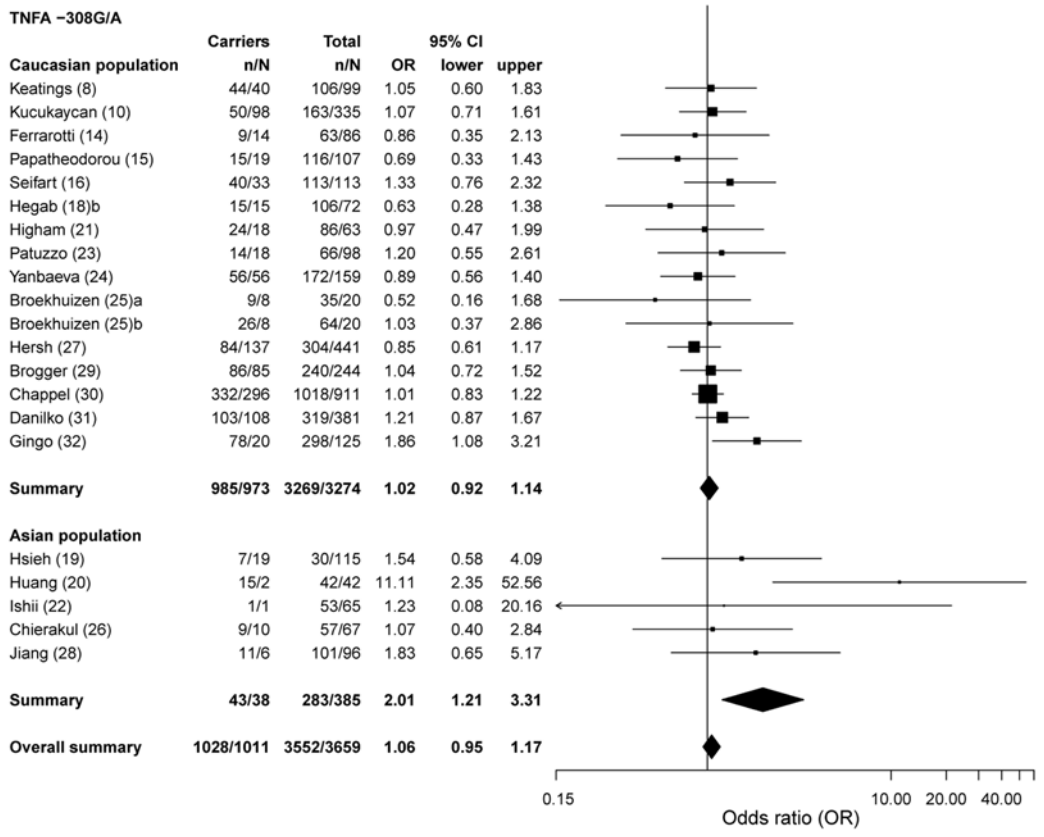


Figure 2: Association of *TNFA* −308 G/A under dominant model. A fixed-effect model was used (n- number of cases, N- number of controls). Heterogeneity among publications: for Caucasians $\chi^2(15)=12.48$ (p=0.64), for Asians $\chi^2(44)=6.68$ (p=0.15) and overall $\chi^2(21)=23.31$ (p=0.27). Broekhuizen a- study with cachectic patients, b- study with non-cachectic patients.

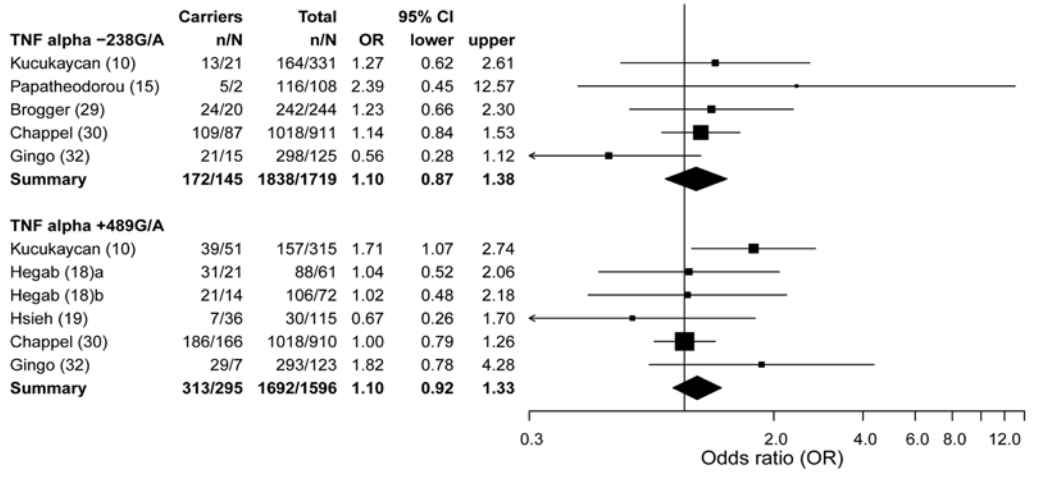


Figure 3: Association of *TNFA* −238 and *TNFA* +489 with COPD under dominant model. For both analyses, a fixed-effect model was used. Test for heterogeneity resulted in $\chi 2(4)=4.83$ (p=0.31) and $\chi^2(5)=6.51$ (p=0.26), respectively.

Transforming growth factor β1 (TGF β1)

Transforming growth factor β1 (TGF β1) is secreted by epithelial cells and macrophages and can cause fibrotic changes in lung tissue. A previous linkage analysis showed an association with chromosome 19 and identified that the association was located in the *TGFB1* region after including additional short tandem repeat markers.[34] Later studies showed inconsistent results and the majority failed to show significant associations. Our search resulted in seven publications describing associations of different *TGFB1* polymorphisms with COPD, [34-40] four investigating associations in Caucasians, three in Asians. We performed meta-analyses on four *TGFB1* polymorphisms (figure 4). The minor alleles of rs2241712 and rs1982073 were protective for COPD with OR 0.84; 95% CI 0.66-1.1 (non-significant) and OR 0.74; 95% CI 0.61-0.89, respectively. The third *TGFB1* SNP, rs1800469, was also protective for COPD (OR 0.75; 95% CI 0.62-0.9). Meta-analysis of the fourth SNP, rs6957, provided a significant increased risk for COPD (OR 1.48, 95% CI 1.13-1.93). A summary of the significant results for *TGFB1* SNPs regarding their putative function and direction of association is shown in Table 1. All these results should be considered with caution, due to the low number of appropriate studies available. In summary, there were too few studies reporting frequencies of *TGFB1* SNPs in COPD patients and healthy controls. Hence it is not possible to draw any definitive conclusions as to their causative role in COPD development.

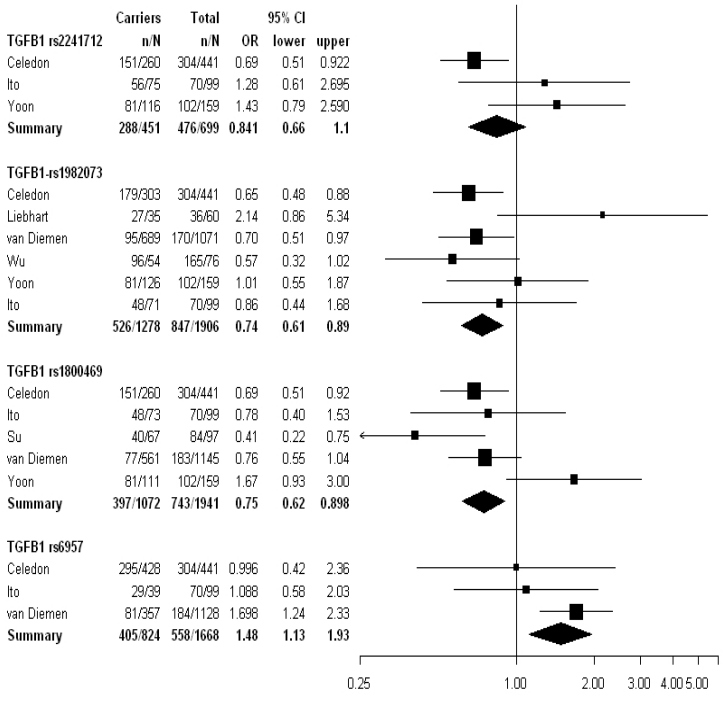


Figure 4: Association of *TGFB* polymorphisms with COPD: rs2241712, rs1982073, rs1800469, and rs6957 under dominant model. A random-effect model was used for rs1800469 (test for heterogeneity $\chi^2(4)=11.65$ ($p=0.02$), $\tau^2=0.1$), and a fixed-effect model was used for rs2241712 ($\chi^2(2)=0.38$, $p=0.83$), rs1982073 ($\chi^2(4)=2.19$, $p=0.7$) and rs6957 ($\chi^2(2)=2.47$, $p=0.29$).

Table 1. The putative function of SNPs associated with COPD and the direction of the association.

TGFB1 SNP	Location/type of polymorrphism	Putative function	Direction of the association
rs2241712	Promoter	Altered gene expresion	Decreased risk
rs1982073 (now rs1800470)	Missense (coding region)	Protein function/ structure changed	Decreased risk
rs6957	3'UTR	mRNA stability	Increased risk

Interleukins (IL)

The most frequently studied IL genes are *IL10*, *IL6* and *IL1*. *IL10* has been identified as a good candidate gene because of its anti-inflammatory potential and decreased levels in COPD. *IL10* polymorphisms were studied in different contexts, but the number of case-control studies for each SNP was too low to perform an informative meta-analysis.

IL6 is a pro-inflammatory cytokine, hence SNPs in the gene may cause prolonged airway inflammation, which may contribute to COPD susceptibility. Promoter polymorphisms of *IL6* have been studied most frequently, as they can lead to altered protein levels. Because of its high levels in stable COPD patients, IL6 has been proposed as a marker of systemic inflammation in COPD. Its function as a predictor of COPD progression is not clear, since there are many sources of this cytokine in the human body besides the lungs[41]. We performed a meta-analysis on the -174G/C polymorphism including four publications studying Caucasian populations[16,25,31,42] (figure 5) and found no significant association with an OR 1.15 (95% CI 0.92-1.43).

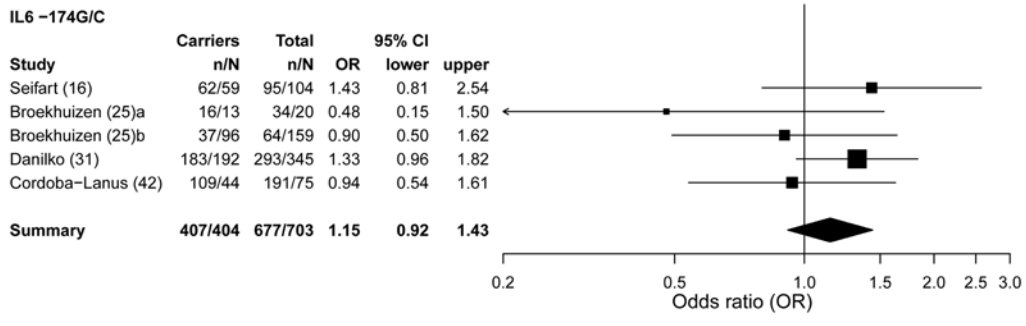


Figure 5: Association of *IL6* -174 G/C polymorphism under dominant model. A fixed-effect model was used ($\chi^2(4)=4.83$, p=0.31).

Finally, another group of candidate genes, members of the proinflammatory interleukin 1 (*IL1*) gene family (encoding two agonists IL1α and IL1β and antagonist receptor IL1RN), have been studied. We found eight relevant studies (four in Caucasians and four in Asians) from

seven publications.[18,19,22,25,31,43,44] We included three *IL1B* polymorphisms into our meta-analyses, i.e. -511 C/T and -31 T/C, both suspected to alter the transcriptional activity and +3954 T/C, a silent substitution known to increase IL1β levels). The *IL1RN* VNTR (variable number of tandem repeats) polymorphism was also studied, always together with *IL1B* polymorphisms. This is not surprising as the IL1β /IL1RN ratio is known to be crucial in determining the severity of inflammatory responses[45] and the IL1RN*2 allele (two repeats) has been shown to increase production of IL1β.[46] We found a strong association between this polymorphism and risk for COPD in Asian populations (OR 1.70, 95% CI 1.09-2.65), as all the publications we could include had studied Asians (figure 6).

We found no significant associations of SNPs in *IL1B* with COPD, i.e. position -511 with OR 1.12 (95% CI 0.74-1.7), position -31 with OR 1.3 (95% CI 0.99-1.71) and position +3954 with OR 0.82 (95% CI 0.63-1.07).

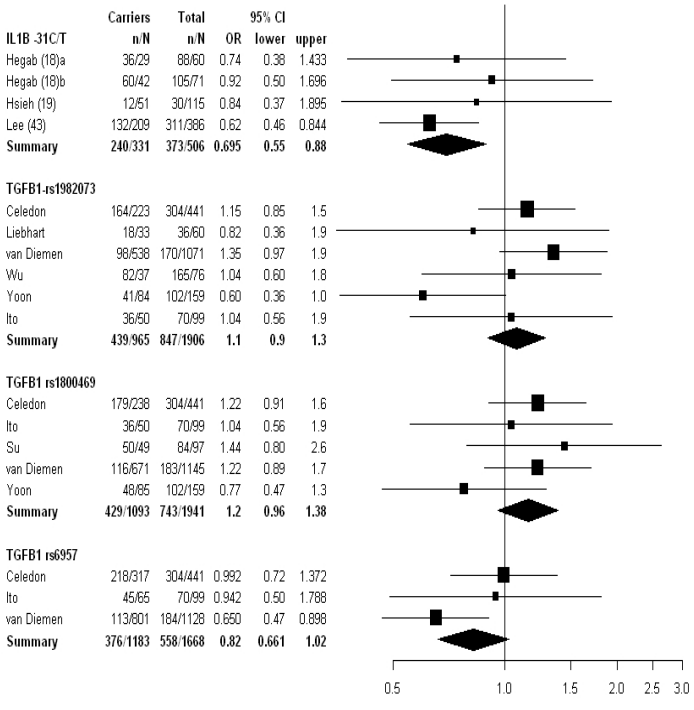


Figure 6: Association of polymorphisms within interleukin-1 gene family: *IL1RN* VNTR polymorphism, *IL1B* -511C/T, -31T/C and +3954T/C under dominant model. A random-effect model was used for *IL1B* -511C/T ($\chi^2(6)=24.18$, p=0.0001), a fixed-effect model was used for two other polymorphisms: -31T/C ($\chi^2(3)=6.22$, p=0.1) and +3954T/C ($\chi^2(3)=2.19$, p=0.53).

2. Genes involved in protease–anti-protease balance

Many proteases secreted by key inflammatory cells are involved in remodeling and progressive degradation of lung tissue. One of these is neutrophil elastase, which is released by neutrophils upon activation. Macrophages produce matrix metalloproteinases (MMPs) and cathepsins, which lead to elastolysis and emphysema. CD8+ cells secrete perforins and granzyme B which, together with TNFα, cause cytolysis and apoptosis of alveolar epithelial cells also contributing to emphysema.[7] Although there is a broad spectrum of proteases involved in tissue damage in COPD, MMPs have been studied the most extensively, especially MMP9. We included five studies (three in Asians and two in Caucasians) on *MMP9* in our meta-analysis.[47–51] The SNP at position –1562 causing a C to T transition, which is known to alter promoter activity,[52] increased the risk for COPD, with an OR of 1.45, but the result was not significant (95% CI 0.84–2.50) (figure 7). More studies are needed to reveal its role.

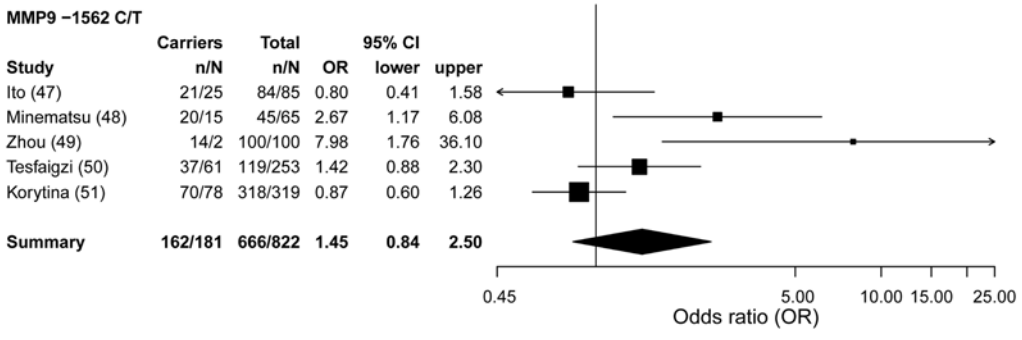


Figure 7: Association of *MMP9* –1562C/T polymorphism under dominant model. A random-effect model was used ($\chi^2(4)=14.32$, $p=0.0063$, $\tau^2=0.25$).

MMPs are inhibited by tissue inhibitors of metalloproteases (TIMP1–TIMP4). *TIMP2* polymorphisms have been studied and shown to be associated with COPD.[53] TIMP1 also plays a role in COPD as its serum concentrations are elevated in COPD patients and correlate inversely with the level of FEV_1 (forced expiratory volume in first second), which is a marker of COPD severity. It did not correlate with emphysema scores, suggesting it only plays a role in the development of airway obstruction.[54]

As mentioned, the *SERPINA1* gene coding AAT was for many years the only known genetic factor associated with COPD. The *SERPINA1* PI*ZZ genotype is known to cause the most severe AAT deficiency. Increased risk was also found for PI*MZ heterozygotes when compared with PI*MM individuals in a meta-analysis of 16 independent studies, although the level of FEV_1 did not differ between these two groups.[55] The S allele is also known to lower AAT levels. With a meta-analysis on SZ and MS carriers, the SZ carriers had an increased risk for COPD compared to normal allele carriers, PI*MM (OR 3.26, 1.24–8.57). The MS carriers had a

non-significant increased risk for COPD (OR 1.19, 1.02–1.38), while the number of individuals with a PI*SS genotype was not sufficient to perform a meta-analysis.[56]

3. Antioxidant genes

Because tobacco smoking is the main cause of COPD, genes involved in utilizing reactive oxygen/nitrogen species (ROS/RNS) delivered with cigarette smoke have also been studied. Two of the most studied genes in the antioxidant pathway are microsomal epoxide hydrolase (*EPHX1*) and glutathione S-transferase (GST) subunits P1, T1 and M1.

Microsomal epoxide hydrolase (EPHX1)

EPHX1 is the first line of defense against toxic epoxides provided with cigarette smoke and it has two polymorphisms that alter its function, Y113H and H139R, corresponding to slow and fast enzyme activity, respectively.[57] A recent meta-analysis on *EPHX*[58] including 1,847 patients and 2,455 controls showed a significant increase in COPD risk (OR 1.59, 95% CI 1.14–2.21) for the Y113H polymorphism and a non-significant reduction in COPD risk (OR 0.90, 95% CI 0.65–1.24) for the H139R polymorphism. We included 18 studies [27,29,59–74] describing both polymorphisms in our meta-analyses, among them 11 in Caucasians and eight in Asians. The Y113H polymorphism was not associated with COPD (OR 1.02, 95% CI 0.91–1.14), also after stratification, in Caucasians (OR 1.02, 95% CI 0.9–1.15) and in Asians (OR 1.02, 95% CI 0.76–1.36) (figure 8). H139R provided a weakly protective role in COPD with a borderline significant OR 0.91 (95% CI 0.83–1.01) (figure 9). Stratification for ethnicity confirmed there was no significant association in Caucasians (OR 0.96, 95% CI 0.85–1.07), but revealed a significant association in Asians (OR 0.76, 95% CI 0.61–0.96).

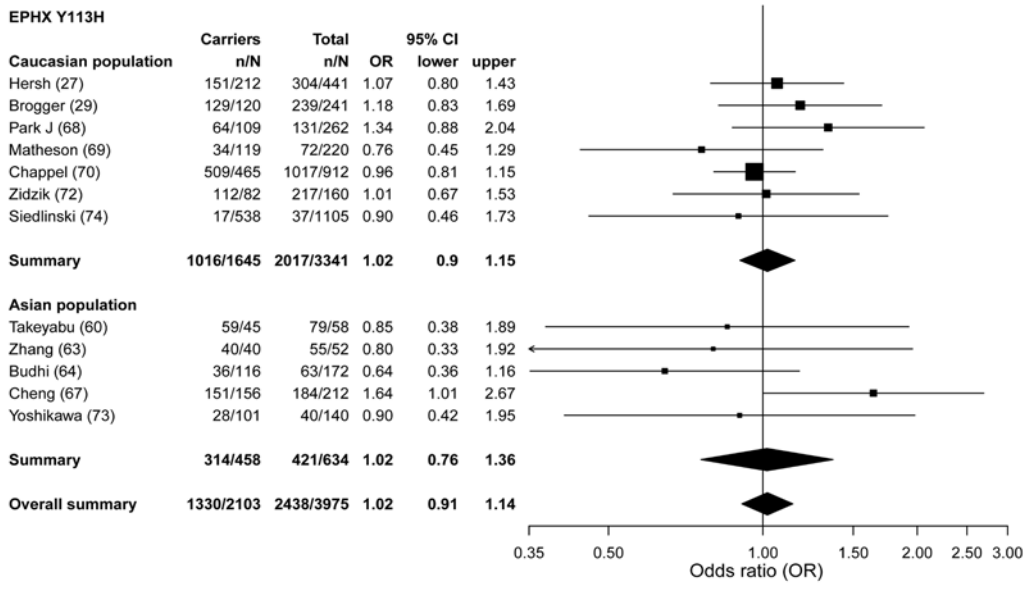


Figure 8. Association for *EPHX* Y113H polymorphism, overall and stratified analyses under dominant model. For all analyses a fixed-effect model was used. Heterogeneity among publications: for Caucasians ($\chi^2(6)=4.08$, p=0.67), for Asians ($\chi^2(4)=6.62$, p=0.16) and overall ($\chi^2(11)=10.7$, p=0.47).

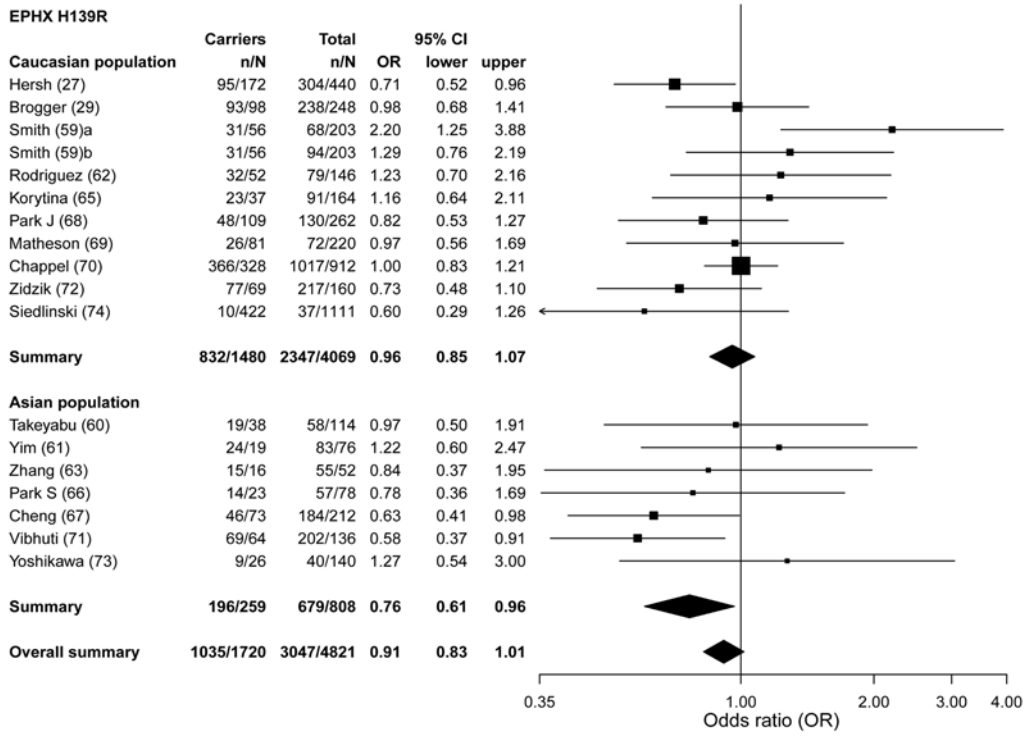


Figure 9. Association for *EPHX* H139R polymorphism, overall and stratified analyses under dominant model. For all analyses a fixed-effect model was used. Heterogeneity among publications: for Caucasians ($\chi^2(10)=18.2$, p=0.052), for Asians ($\chi^2(8)=5.73$, p=0.45) and overall ($\chi^2(17)=26.92$, p=0.059).

Glutathione S-transferase (GST)

We performed meta-analyses on *GSTM1* and *GSTT1* null alleles (figures 10 and 11) under a recessive model, including 14 studies from 11 publications[27,61,64,66,67,72,75-79] and one unpublished work (van Diemen et al) for *GSTM1,* with seven of the studies also covering *GSTT1.* Five of those studies investigated Asian populations and nine Caucasian populations. We found homozygotes for the *GSTM1* null allele to have an increased risk of COPD (OR 1.3, 95% CI 1.07–1.57). Stratified analysis confirmed this effect in Caucasian populations (OR 1.32, 95% CI 1.04–1.68) but not in Asians (OR 1.24, 95% CI 0.85–1.8). For the *GSTT1* null allele we found no significant associations in our overall analysis (OR 1, 95% CI 0.82–1.22) or stratified analyses (OR 1.2, 95%CI 0.63–2.29 and OR 0.93, 95% CI 0.73–1.19 for Caucasian and Asian populations, respectively).

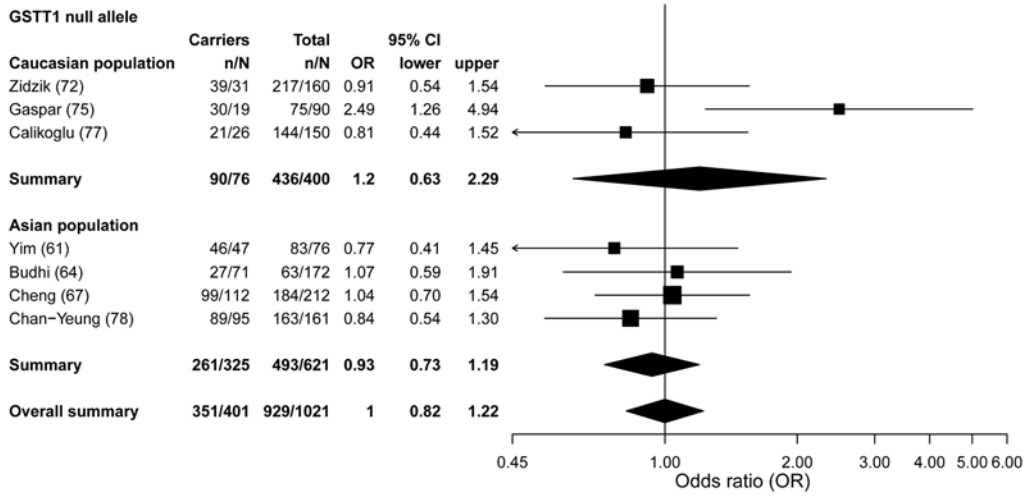


Figure 10: Association of *GSTM1* null allele overall and stratified analyses under recessive model. A random-effect model was used for all analyses. Heterogeneity among publications: for Caucasians ($\chi^2(8)=19.8$, p=0.01, $\tau^2=0.08$), for Asians ($\chi^2(4)=9.94$, p=0.0414, $\tau^2=0.11$) and overall ($\chi^2(14)=29.81$, p=0.005, $\tau^2=0.07$).* unpublished data

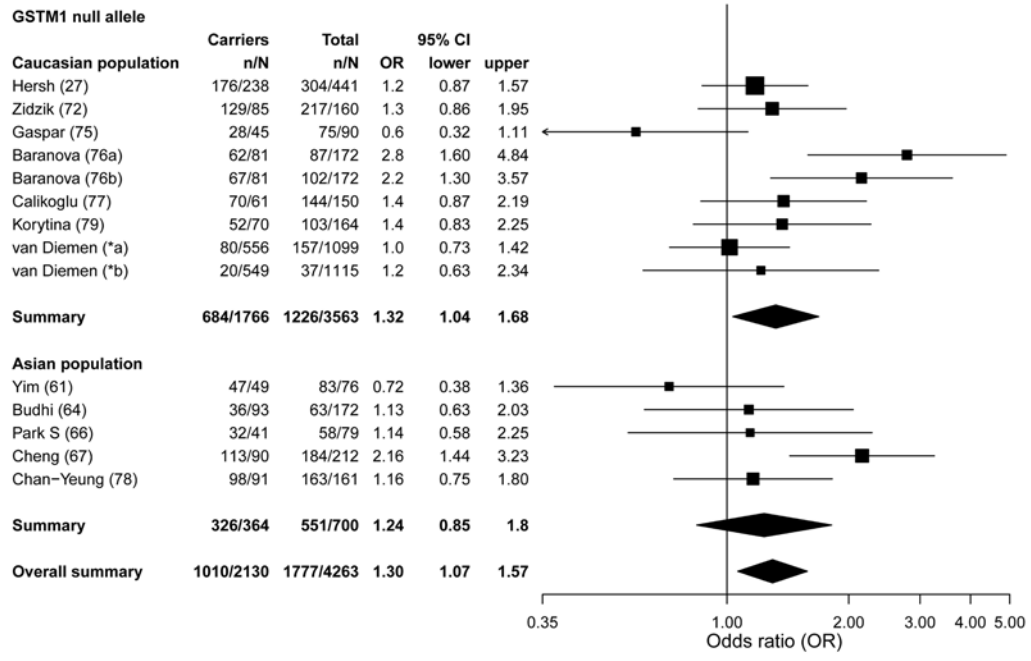


Figure 11. Association of *GSTT1* null allele overall and stratified analyses under recessive model. A random-effect model was used in analysis for Caucasians ($\chi^2(2)=6.81$, p=0.033, $\tau^2=0.23$), and a fixed-effect model was used in analysis for Asians ($\chi^2(3)=1.1$, p=0.78) and overall ($\chi^2(6)=8.74$, p=0.19).

The isoleucine to valine substitution polymorphism at position 105 of *GSTP1* (I105V) increases both its catalytic activity for substrates and the metabolism of diol epoxides of polycyclic aromatic hydrocarbons.[80] A second polymorphism, which has been less well studied, is located in exon 6 at position 114 and causes an alanine to valine substitution, which alters the protein function in the same way as the previous polymorphism. We combined 12 studies (including one unpublished, van Diemen et al) in our meta-analysis, six being conducted in Asians and six in Caucasians,[27,64,67,71,75,78,79,81-84] which showed a non-significant protective effect for the I105V polymorphism (OR 0.89, 95% CI 0.73-1.1) (figure 12). However, stratified analysis showed a significant association in Asians (OR 0.69 95% CI 0.56-0.85), but not in Caucasians (OR 1.14 95% CI 0.96-1.35).

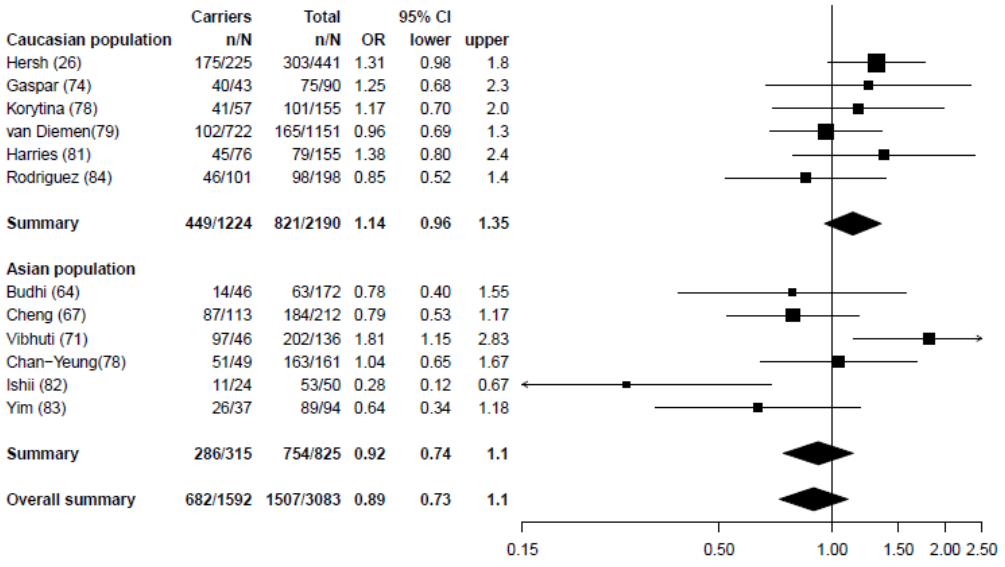


Figure 12: Association of *GSTP1* I105V polymorphism overall and stratified analyses under dominant model. A random-effect model was used in overall analysis ($\chi^2(11)=23.85$, p=0.013, $\tau^2=0.07$). In stratified analyses fixed-effect model was used for Caucasians ($\chi^2(5)=3.82$, p=0.57) and Asians ($\chi^2(5)=7.48$, p=0.19). * unpublished data.

Superoxide dismutases (SOD)

Superoxide anion is delivered with cigarette smoke and is also generated in many resident and attracted inflammatory cells, for example in the defense processes against other ROS. The family of superoxide dismutases (SOD) changes this toxic anion into hydrogen peroxide. There are three members in this family: intracellular copper-zinc SOD (CuZnSOD), mitochondrial manganese SOD (MnSOD), and extracellular SOD (ECSOD), also known as SOD1, SOD2 and SOD3, respectively.[85] Despite their important roles, genes of these enzymes have not yet been widely studied for associations with COPD.

We found four studies investigating the Ala16Val *SOD2* polymorphism in COPD, three in Caucasians and one in Asians.[86-89] Meta-analysis showed an increased risk for COPD (OR 1.07, 95% CI 0.82-1.4) (figure 13). However, the result was non-significant, because the number of publications was low and the largest study,[89] which influenced our meta-analysis the most, showed a non-significant result. More data for this polymorphism is required to reveal its role in COPD.

We also found three publications reporting the association of the *SOD3* R213G polymorphism with COPD in Caucasian populations.[86,89,90] Our meta-analysis had an OR 0.63 (95% CI 0.25-1.60) (figure 14), indicating a possible protective role by the 213G allele, but our result was not significant.

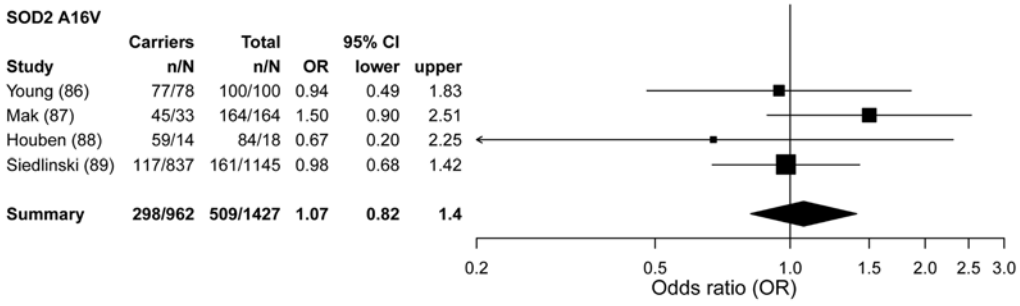


Figure 13: Association of the *SOD2* A16V polymorphism under dominant model. A fixed-effect model was used ($\chi^2(3)=2.6$, p=0.46).

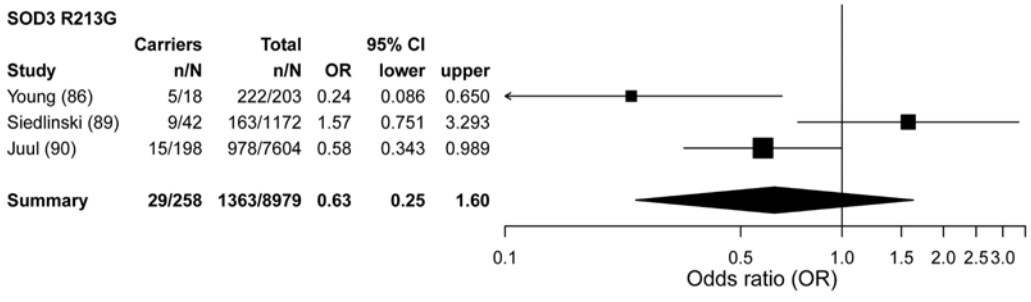


Figure 14: Association of *SOD3* R213G polymorphism under dominant model. A random-effect model was used ($\chi^2(2)=9.45$, p=0.0089, $\tau^2=0.52$).

Recently Dahl et al[91] performed a study on two newly identified SNPs in the *SOD3* (rs8192287 and rs8192288), which were found to be significantly associated with an increased risk for COPD hospitalization and mortality in the large, population-based, Copenhagen City Heart Study. They found a smaller effect in the second large cohort studied, the Copenhagen General Population Study. In both cohorts, associations of these SNPs were found with reduced FEV_1% predicted and FVC% predicted (forced vital capacity). More recently, these results were partially confirmed by Siedlinski et al,[89] who found an association of rs8192288 with reduced FVC, which was of borderline significance. They did not replicate the results for FEV_1, possibly due to the low number of homozygous individuals.

Overlap with other diseases

When considering the association between the polymorphisms in the genes and COPD, it appears that COPD shares some of its genetic background with other complex diseases. According to the HuGE database, up to 28 January 2009, 94 genes had been studied for association with COPD and these are involved in 89 pathways. For asthma 427 genes have been studied, which fall into 125 pathways, and 35 of them overlap with pathways involved in COPD. Table 2 shows the number of publications investigating associations for the genes

studied in the context of COPD with selected number of other complex diseases.

Table 2: Number of publications investigating specific genes for the various diseases. Number of publications for each disease includes all studies ever published according to HuGe database. These numbers contain all types of studies, therefore not only in humans.

Gene	COPD	Asthma	Osteoporosis	Cancer	Alzheimer	Type 2 diabetes	Crohn's disease
TGFB1	9	24	22	94	8	9	4
TNFA	21	49	22	248	30	47	57
MMP9	7	3	-	39	3	4	2
GSTP1	13	34	-	379	4	2	1
EPHX1	16	2	-	98	-	-	1
IL10	8	16	3	168	13	4	16
IL6	8	5	28	120	20	26	7
ADAM33	1	26	-	1	-	-	-
GSTT1	8	21	1	543	3	5	2
GSTM1	9	30	1	687	3	5	2
HMOX1	7	1	-	8	3	2	1
SOD2	2	4	-	83	3	7	-
SOD3	2	-	-	3	1	2	-

What can we infer from this? We can speculate that many diseases have a shared genetic background, but different (combinations of) polymorphisms in the shared genes determine which disease evolves. This has recently become evident for immune-related disorders[92] but might in fact be true for other diseases as well. In addition, the type of disease that evolves from a shared genetic background can be determined by certain environmental factors, at specific windows of exposure. For example, a cigarette smoker with a -308A allele in *TNFA* gene can develop either COPD or Crohn's disease depending on other environmental factors and other polymorphisms. Especially the latter are likely to be important, since complex disorders are determined by polymorphisms in many loci, underlying a specific genetic profile. So far little is known about these profiles, but new techniques like next generation sequencing might well be able to reveal them.

Summary of results

For this review we performed meta-analyses on 20 polymorphisms in 12 functional candidate genes that had previously been studied in at least 3 publications for association with COPD. Only five showed significant associations with COPD. Three of these were located in

TGFB1, two providing a protective effect of the minor allele (rs2241712 and rs1982073) and one (rs6957) gave an increased risk for COPD (OR 1.48). These results should be interpreted with caution as the number of studies was low. Two other polymorphisms showing a significant increase in COPD risk were the *IL1RN* VNTR polymorphism and the *GSTM1* null allele (OR 1.7 and OR 1.3, respectively). For all other polymorphisms we studied, the overall meta-analyses resulted in non-significant associations or had borderline significance, i.e. *IL1B* -31 polymorphism (OR 1.3, 95%CI 0.99–1.71) and H139R *EPHX1* polymorphism (OR 0.91, 95% CI 0.83–1.01). Since ethnicity might influence the outcome of the results, we performed stratified analyses, for the six polymorphisms for which we had a sufficiently large number of studies, namely -308G/A in *TNFA*, Y113H and H139R in *EPHX1*, *GSTM1* and *GSTT1* null alleles, and I105V in *GSTP1*.

Discussion

The aim of this review was to reveal the roles of particular polymorphisms studied for association with COPD in a case-control design or in population-based studies and to summarize their sometimes conflicting results. The published candidate gene studies are all hypothesis-based. We took into account only those candidate genes that had been reported in a sufficient number of studies in relation to a well-defined, specific COPD phenotype. As shown, when critically evaluated in meta-analyses, some promising candidate genes do not seem to hold up at all. For example, the result of our meta-analysis on the *TNFA* -308G/A polymorphism: we included 21 studies and found a non-significant association, whereas after stratifying for ethnicity we did find it to be significant but only in Asians. These results are in contrast with those published by Gingo et al.[32] This can be explained by the fact that all six studies included in our meta-analysis reported non-significant associations. Since Gingo et al did not perform analysis stratified by ethnicity, it cannot be ruled out that the association they observed was mostly due to a true association present in Asian populations, but not in Caucasian populations.

The result for the *TNFA* gene in Asian populations is based on five publications studying a total of 283 cases and 385 controls, with 43 and 38 carrying the A allele, respectively. This group is very small compared to the Caucasian population, where 3269 cases and 3274 controls were present, 985 and 973 carrying the minor A allele, respectively. Is the effect really that strong in the Asian population or is there a publication bias? Results may well be biased, because all Asian studies reported an OR greater than 1. The Y113H polymorphism in *EPHX* was not associated with COPD in the overall analysis, and remained non-significant after analyses on Caucasians and Asians separately. For H139R in *EPHX*, as in our overall analysis, the association was non-significant in Caucasians but showed a significant protective effect in Asian populations, which is of interest because the total number of Caucasians included

was over four times that of Asians (2347 versus 679 cases and 4069 versus 808 controls). Stratified analysis for *GSTM1* confirmed an increased risk for COPD, but only in Caucasians and, again, this group was much larger than the Asian group. The *GSTT1* null allele was not associated with COPD in the overall analysis and this was also true in both Caucasian and Asian populations. By this we confirmed the results published by Hu et al[93] on a *GSTM1* and *GSTT1* meta-analysis. They found significant increase in COPD risk for the *GSTM1* null allele in overall (OR 1.46, 95% CI 1.16–1.83) and stratified analysis in non-Asians only (OR 1.57, 95% CI 1.19–2.07). The *GSTP1* I105V polymorphism was found to have a protective role in Asians, even for the relatively small numbers of subjects studied, suggesting its possible strong effect in this population. Interestingly, the same polymorphism seems to have an opposite effect in Caucasians (figure 12). These results show the importance of ethnicity in genetic studies, as some associations were present in one group but absent in another. Clearly, most of the polymorphisms we described were investigated in Caucasian populations. Thus combined analyses could yield diminished results or, on the contrary, indicate general associations which are in fact present in only one of the groups. We also showed that more studies are needed for most of the polymorphisms investigated here, especially in Asian populations, before more conclusive answers can be given.

There are several limitations to our study. First, despite carefully selecting the studies, we cannot ensure homogeneity in either the case or control groups. Normally, COPD patients are or have been smokers, although this was not always an explicit criterion in the studies included in our meta-analyses (see table E1 in the online supplementary information). Controls should be matched for age, sex and, most of all, smoking history. Specifically the smoking habits, a crucial feature in studies on the etiology of COPD, are often not known, for example if the control groups are blood donors. Even more importantly, lung function measurements were not performed in this group, so individuals with low lung function may have been included. Some of the studies included blood donors as controls and this may have affected the outcome of the study strongly, for example it might have led to the discovery of strong associations, which would be observed due to bias in the participants' smoking status.

Secondly, the number of studies we used for the meta-analyses for each polymorphism is sometimes insufficient to draw reliable conclusions, e.g. on the *TGFB1* polymorphisms. In some cases, when only allele frequencies were given, we calculated genotype frequencies based on HWE. This should not, however, influence the outcome of our analyses, as the number of such studies was relatively small, and the calculated genotype frequencies should not deviate from the actual genotype distributions.

We should mention also the importance of HWE. We excluded 19 studies from some of our meta-analyses due to the lack of HWE in control groups. The most affected was the

meta-analysis on the Y113H polymorphism (eight studies were excluded). This was because of a well-known problem with the first primer set used for genotyping: neighboring polymorphism caused an overestimation of the minor allele in this polymorphism. Including all these studies resulted in significant outcomes in the overall analysis and in the Caucasian populations (data not shown), similar to another recent meta-analysis.[58] Half of the studies we excluded reported that HWE had been checked and that no deviations had been found. In the case of *IL1RN,* the only two studies investigating the role of its polymorphism in Caucasians were also excluded for the same reason, leaving only Asian populations in the meta-analysis.

There are many single studies describing association of a particular gene with specific (related) phenotypic outcomes, like *ADAM33* (a disintegrin and metalloprotease 33) that was first shown to be associated with asthma, then with COPD development in the general population,[94] and later was linked to airway hyperresponsiveness in COPD patients.[95] Such 'loose replications' are sometimes interpreted as further proof for a role of that specific gene in the broader disease definitions, but functional studies are clearly necessary to pinpoint the exact role of such SNPs in genes.

It is worth mentioning recent findings from hypothesis-free genome-wide association (GWA) studies on lung cancer patients, which found an association with the nicotinic acetylcholine receptor subunit (*CHRNA*) gene polymorphisms.[96] There is a debate on whether SNPs in that region have a direct effect on lung cancer or merely mediate smoking habits. Hung et al did not find any association of those SNPs with nicotine dependence,[97] whereas Thorgeisson et al[98] and Young et al[99] investigated genotype frequencies of the most strongly associated SNP found in GWA studies, rs16969968 (which is in strong linkage disequilibrium with rs1051730), in lung cancer patients with and without COPD, and in smoking controls, to check whether this association could be confounded by COPD. They found an increased frequency of the AA genotype in both groups compared to controls, which means that the *CHRNA* gene polymorphism is associated with COPD and with lung cancer. Despite the interesting results, this study shows how important it is to select cases and controls carefully and to avoid possible confounders. The first GWA study on COPD was published recently[100] and identified associations of the same SNPs as in lung cancer, rs8034191 and rs1051730. Similarly, it still remains to be elucidated whether these SNPs are associated with COPD itself or whether this association is mainly seen because of nicotine addiction among COPD patients, although Pillai et al found no association with pack-years smoked in two cohorts. This might be because pack-years may not be the most appropriate indicator of addiction. To answer these questions, more studies are needed, especially studies with quantified addiction data. Clearly, none of the polymorphisms investigated here appeared in the GWA studies. This can be explained in part by the stringent criteria applied to select SNPs for replication and the

relatively small initial cohort. But we cannot rule out that some of these polymorphisms may not be associated with COPD at all.

In summary, we showed that many candidate genes show no association with COPD after combining independent studies. For the majority of the genes more studies will be needed to reveal their role, especially in Asian populations. We show that ethnicity is very important in genetic studies, and therefore should always be taken into account, given that some associations show different directions for different populations.

Acknowledgements

We would like to thank JP Schouten for assisting with the analyses performed for this review, Dr J Fu for translations of articles published in Chinese, T van Ittersum for assisting with the literature search, and J Senior for editing the text.

Conflict of interest statement

We declare we have no conflicts of interest.

Contributions

JS designed the study, performed literature search, extracted, analyzed and interpreted the results, prepared figures and wrote the manuscript. CW designed the study, interpreted the data and critically revised the manuscript. DSP critically revised the manuscript. HMB designed the study, performed literature search, extracted the data, interpreted the results, wrote and critically revised the manuscript.

References

1. WHO. The Global Tobacco Crisis. 2008.
2. Fletcher C, Peto R. The Natural History of Chronic Airflow Obstruction. *Br Med J* 1977;1:1645–1648.
3. Lieberman J, Winter B, Sastre A. Alpha 1-Antitrypsin Pi-Types in 965 COPD Patients. *Chest* 1986;89:370–373.
4. Nici L, Donner C, Wouters E, Zuwallack R, Ambrosino N, Bourbeau J, Carone M, Celli B, Engelen M, Fahy B, et al. American Thoracic Society/European Respiratory Society Statement on Pulmonary Rehabilitation. *Am J Respir Crit Care Med* 2006;173:1390–1413.
5. Pauwels RA, Buist AS, Calverley PM, Jenkins CR, Hurd SS. Global Strategy for the Diagnosis, Management, and Prevention of Chronic Obstructive Pulmonary Disease:NHLBI/WHO Global Initiative for Chronic Obstructive Lung Disease (GOLD) Workshop Summary. *Am J Respir Crit Care Med* 2001;163:1256–1276.
6. Hogg JC, Chu F, Utokaparch S, Woods R, Elliott WM, Buzatu L, Cherniack RM, Rogers RM, Sciurba FC, Coxson HO, et al. The Nature of Small-Airway Obstruction in Chronic Obstructive Pulmonary Disease. *N Engl J Med* 2004;350:2645–2653.
7. Barnes PJ. Mediators of Chronic Obstructive Pulmonary Disease. *Pharmacol Rev* 2004;56:515–548.
8. Keatings VM, Cave SJ, Henry MJ, Morgan K, O'Connor CM, FitzGerald MX, Kalsheker N. A Polymorphism in the Tumor Necrosis Factor-Alpha Gene Promoter Region May Predispose to a Poor Prognosis in COPD. *Chest* 2000;118:971–975.
9. Wood AM, Simmonds MJ, Bayley DL, Newby PR, Gough SC, Stockley RA. The TNFalpha Gene Relates to Clinical Phenotype in Alpha-1-Antitrypsin Deficiency. *Respir Res* 2008; vol. 9. Available from: http://respiratory-research.com/content/9/1/52.
10. Kucukaycan M, Van KM, Pennings HJ, Huizinga TW, Buurman WA, Dentener MA, Wouters EF. Tumor Necrosis Factor-Alpha +489G/A Gene Polymorphism Is Associated With Chronic Obstructive Pulmonary Disease. *Respir Res* 2002; vol. 3. Available from: http://respiratory-research.com/content/pdf/rr194.pdf.
11. Sakao S, Tatsumi K, Igari H, Watanabe R, Shino Y, Shirasawa H, Kuriyama T. Association of Tumor Necrosis Factor-Alpha Gene Promoter Polymorphism With Low Attenuation Areas on High-Resolution CT in Patients With COPD. *Chest* 2002;122:416–420.
12. Tanaka G, Sandford AJ, Burkett K, Connett JE, Anthonisen NR, Pare PD, He JQ. Tumour Necrosis Factor and Lymphotoxin A Polymorphisms and Lung Function in Smokers. *Eur Respir J* 2007;29:34–41.
13. Ruse CE, Hill MC, Tobin M, Neale N, Connolly MJ, Parker SG, Wardlaw AJ. Tumour Necrosis Factor Gene Complex Polymorphisms in Chronic Obstructive Pulmonary Disease. *Respir Med* 2007;101:340–344.
14. Ferrarotti I, Zorzetto M, Beccaria M, Gile LS, Porta R, Ambrosino N, Pignatti PF, Cerveri I, Pozzi E, Luisetti M. Tumour Necrosis Factor Family Genes in a Phenotype of COPD Associated With Emphysema. *Eur Respir J* 2003;21:444–449.
15. Papatheodorou A, Latsi P, Vrettou C, Dimakou A, Chroneou A, Makrythanasis P, Kaliakatsos M, Orfanidou D, Roussos C, Kanavakis E, et al. Development of a Novel Microarray Methodology for the Study of SNPs in the Promoter Region of the TNF-Alpha Gene: Their Association With Obstructive Pulmonary Disease in Greek Patients. *Clin Biochem* 2007;40:843–850.
16. Seifart C, Dempfle A, Plagens A, Seifart U, Clostermann U, Muller B, Vogelmeier C, von WP. TNF-Alpha-, TNF-Beta-, IL-6-, and IL-10-Promoter Polymorphisms in Patients With Chronic Obstructive Pulmonary Disease. *Tissue Antigens* 2005;65:93–100.
17. Demeo DL, Campbell EJ, Barker AF, Brantly ML, Eden E, McElvaney NG, Rennard SI, Sandhaus RA, Stocks JM, Stoller JK, et al. IL10 Polymorphisms Are Associated With Airflow Obstruction in Severe Alpha1-Antitrypsin Deficiency. *Am J Respir Cell Mol Biol* 2008;38:114–120.
18. Hegab AE, Sakamoto T, Saitoh W, Nomura A, Ishii Y, Morishima Y, Iizuka T, Kiwamoto T, Matsuno Y, Massoud HH, et al. Polymorphisms of TNFalpha, IL1beta, and IL1RN Genes in Chronic Obstructive Pulmonary Disease. *Biochem Biophys Res Commun* 2005;329:1246–1252.

19. Hsieh MH, Chong IW, Hwang JJ, Lee CH, Ho CK, Yu ML, Huang CT, Lee CY, Wu MT, Christiani DC. Lack of Associations Between Several Polymorphisms in Cytokine Genes and the Risk of Chronic Obstructive Pulmonary Diseases in Taiwan. *Kaohsiung J Med Sci* 2008;24:126–137.

20. Huang SL, Su CH, Chang SC. Tumor Necrosis Factor-(Alpha) Gene Polymorphism in Chronic Bronchitis. *Am J Respir Crit Care Med* 1997;156:1436–1439.

21. Higham MA, Pride NB, Alikhan A, Morrell NW. Tumour Necrosis Factor-(Alpha) Gene Promoter Polymorphism in Chronic Obstructive Pulmonary Disease. *Eur Respir J* 2000;15:281–284.

22. Ishii T, Matsuse T, Teramoto S, Matsui H, Miyao M, Hosoi T, Takahashi H, Fukuchi Y, Ouchi Y. Neither IL-1(Beta), IL-1 Receptor Antagonist, nor TNF-(Alpha) Polymorphisms Are Associated With Susceptibility to COPD. *Respir Med* 2000;94:847–851.

23. Patuzzo C, Gile LS, Zorzetto M, Trabetti E, Malerba G, Pignatti PF, Luisetti M. Tumor Necrosis Factor Gene Complex in COPD and Disseminated Bronchiectasis. *Chest* 2000;117:1353–1358.

24. Yanbaeva DG, Korytina GF, Viktorova TV. Allelic Variants of TNF Superfamily Genes Serve As Markers of Disease Severity in Patients With Chronic Obstructive Pulmonary Disease and Bronchiectatic Disease. *Genetika* 2004;40:545–551.

25. Broekhuizen R, Grimble RF, Howell WM, Shale DJ, Creutzberg EC, Wouters EF, Schols AM. Pulmonary Cachexia, Systemic Inflammatory Profile, and the Interleukin 1beta -511 Single Nucleotide Polymorphism. *Am J Clin Nutr* 2005;82:1059–1064.

26. Chierakul N, Wongwisutikul P, Vejbaesya S, Chotvilaiwan K. Tumor Necrosis Factor-Alpha Gene Promoter Polymorphism Is Not Associated With Smoking-Related COPD in Thailand. *Respirology* 2005;10:36–39.

27. Hersh CP, Demeo DL, Lange C, Litonjua AA, Reilly JJ, Kwiatkowski D, Laird N, Sylvia JS, Sparrow D, Speizer FE, et al. Attempted Replication of Reported Chronic Obstructive Pulmonary Disease Candidate Gene Associations. *Am J Respir Cell Mol Biol* 2005;33:71–78.

28. Jiang L, He B, Zhao MW, Ning LD, Li XY, Yao WZ. Association of Gene Polymorphisms of Tumour Necrosis Factor-Alpha and Interleukin-13 With Chronic Obstructive Pulmonary Disease in Han Nationality in Beijing. *Chin Med J (Engl)* 2005;118:541–547.

29. Brogger J, Steen VM, Eiken HG, Gulsvik A, Bakke P. Genetic Association Between COPD and Polymorphisms in TNF, ADRB2 and EPHX1. *Eur Respir J* 2006;27:682–688.

30. Chappell S, Daly L, Morgan K, Baranes TG, Roca J, Rabinovich R, Millar A, Donnelly SC, Keatings V, MacNee W, et al. Variation in the Tumour Necrosis Factor Gene Is Not Associated With Susceptibility to COPD. *Eur Respir J* 2007;30:810–812.

31. Danilko KV, Korytina GF, Akhmidishina LZ, Ianbaeva DG, Zagidullin S, Victorova TV. Association of Cytokines Genes (ILL, IL1RN, TNF, LTA, IL6, IL8, IL0) Polymorphic Markers With Chronic Obstructive Pulmonary Disease. *Molecular Biology* 2007;41:26–36.

32. Gingo MR, Silveira LJ, Miller YE, Friedlander AL, Cosgrove GP, Chan ED, Maier LA, Bowler RP. Tumour Necrosis Factor Gene Polymorphisms Are Associated With COPD. *Eur Respir J* 2008;31:1005–1012.

33. Aubier M, Guenegou A, Benessiano J, Leynaert B, Boczkowski J, Neukirch F. !Association of Lung Function Decline With the Microsatellite Polymorphism in the Heme Oxygenase-1 Gene Promoter, in a General Population Sample. Results From the Longitudinal European Community Respiratory Health Survey (ECRHS-France)!. *Bull Acad Natl Med* 2006;190:877–890.

34. Celedon JC, Lange C, Raby BA, Litonjua AA, Palmer LJ, Demeo DL, Reilly JJ, Kwiatkowski DJ, Chapman HA, Laird N, et al. The Transforming Growth Factor-Beta1 (TGFB1) Gene Is Associated With Chronic Obstructive Pulmonary Disease (COPD). *Hum Mol Genet* 2004;13:1649–1656.

35. Wu L, Chau J, Young RP, Pokorny V, Mills GD, Hopkins R, McLean L, Black PN. Transforming Growth Factor-Beta1 Genotype and Susceptibility to Chronic Obstructive Pulmonary Disease. *Thorax* 2004;59:126–129.

36. Su ZG, Wen FQ, Feng YL, Xiao M, Wu XL. Transforming Growth Factor-Beta1 Gene Polymorphisms

Associated With Chronic Obstructive Pulmonary Disease in Chinese Population. *Acta Pharmacol Sin* 2005;26:714–720.

37. van Diemen CC, Postma DS, Vonk JM, Bruinenberg M, Nolte IM, Boezen HM. Decorin and TGF–Beta1 Polymorphisms and Development of COPD in a General Population. *Respir Res* 2006; vol. 7. Available from: http://respiratory-research.com/content/pdf/1465-9921-7-89.pdf.

38. Yoon HI, Silverman EK, Lee HW, Yoo CG, Lee CT, Chung HS, Kim YW, Han SK, Shim YS, Yim JJ. Lack of Association Between COPD and Transforming Growth Factor–Beta1 (TGFB1) Genetic Polymorphisms in Koreans. *Int J Tuberc Lung Dis* 2006;10:504–509.

39. Ito M, Hanaoka M, Droma Y, Hatayama O, Sato E, Katsuyama Y, Fujimoto K, Ota M. The Association of Transforming Growth Factor Beta 1 Gene Polymorphisms With the Emphysema Phenotype of COPD in Japanese. *Intern Med* 2008;47:1387–1394.

40. Liebhart J, Polak M, Dobek R, Liebhart E, Dor A, Kulczak A, Medrala W, Barg W, Gladysz U, Lange A. {TGF–Beta1 Gene Polymorphism in Chronic Obstructive Pulmonary Disease}. *Pneumonol Alergol Pol* 2005;73:216–220.

41. Walter RE, Wilk JB, Larson MG, Vasan RS, Keaney JF, Jr., Lipinska I, O'Connor GT, Benjamin EJ. Systemic Inflammation and COPD: the Framingham Heart Study. *Chest* 2008;133:19–25.

42. Cordoba-Lanus E, de-Torres JP, Lopez-Aguilar C, Rodriguez-Perez MC, Maca-Meyer N, Montejo-de-Garcini A, guirre-Jaime A, Perez-Mendez L, Casanova C. Association of IL–6 Gene Polymorphisms and COPD in a Spanish Population. *Respir Med* 2008;102:1805–1811.

43. Lee JM, Kang YR, Park SH, Cha SI, Kim JS, Kang HK, Lee WK, Kim MJ, Kim CH, Kim NS, et al. Polymorphisms in Interleukin–1B and Its Receptor Antagonist Genes and the Risk of Chronic Obstructive Pulmonary Disease in a Korean Population: a Case–Control Study. *Respir Med* 2008;102:1311–1320.

44. Shi YZ, Liu B, Huo JM, Zhang W. {Relationship Between Interleukin–1 Cytokine Gene Polymorphisms and Genetic Susceptibility Ot Chronic Obstructive Pulmonary Diseases}*. *Chin J Clin Rehab* 2006;10:67–69.

45. Dinarello CA. Biologic Basis for Interleukin–1 in Disease. *Blood* 1996;87:2095–2147.

46. Hurme M, Santtila S. IL–1 Receptor Antagonist (IL–1Ra) Plasma Levels Are Co–Ordinately Regulated by Both IL–1Ra and IL–1beta Genes. *Eur J Immunol* 1998;28:2598–2602.

47. Ito I, Nagai S, Handa T, Muro S, Hirai T, Tsukino M, Mishima M. Matrix Metalloproteinase–9 Promoter Polymorphism Associated With Upper Lung Dominant Emphysema. *Am J Respir Crit Care Med* 2005;172:1378–1382.

48. Minematsu N, Nakamura H, Tateno H, Nakajima T, Yamaguchi K. Genetic Polymorphism in Matrix Metalloproteinase–9 and Pulmonary Emphysema. *Biochem Biophys Res Commun* 2001;289:116–119.

49. Zhou M, Huang SG, Wan HY, Li B, Deng WW, Li M. Genetic Polymorphism in Matrix Metalloproteinase–9 and the Susceptibility to Chronic Obstructive Pulmonary Disease in Han Population of South China. *Chin Med J (Engl)* 2004;117:1481–1484.

50. Tesfaigzi Y, Myers OB, Stidley CA, Schwalm K, Picchi M, Crowell RE, Gilliland FD, Belinsky SA. Genotypes in Matrix Metalloproteinase 9 Are a Risk Factor for COPD. *Int J Chron Obstruct Pulmon Dis* 2006;1:267–278.

51. Korytina GF, Akhmadishina LZ, Ianbaeva DG, Viktorova TV. Polymorphism in Promoter Regions of Matrix Metalloproteinases (MMP1, MMP9, and MMP12) in Chronic Obstructive Pulmonary Disease Patients. *Russian Journal of Genetics* 2008;44:242–249.

52. Zhang B, Ye S, Herrmann SM, Eriksson P, de MM, Evans A, Arveiler D, Luc G, Cambien F, Hamsten A, et al. Functional Polymorphism in the Regulatory Region of Gelatinase B Gene in Relation to Severity of Coronary Atherosclerosis. *Circulation* 1999;99:1788–1794.

53. Hirano K, Sakamoto T, Uchida Y, Morishima Y, Masuyama K, Ishii Y, Nomura A, Ohtsuka M, Sekizawa K. Tissue Inhibitor of Metalloproteinases–2 Gene Polymorphisms in Chronic Obstructive Pulmonary Disease. *Eur Respir J* 2001;18:748–752.

54. Higashimoto Y, Yamagata Y, Iwata T, Okada M, Ishiguchi T, Sato H, Masuda M, Itoh H. Increased Serum

Concentrations of Tissue Inhibitor of Metalloproteinase-1 in COPD Patients. *Eur Respir J* 2005;25:885–890.

55. Hersh CP, Dahl M, Ly NP, Berkey CS, Nordestgaard BG, Silverman EK. Chronic Obstructive Pulmonary Disease in Alpha1-Antitrypsin PI MZ Heterozygotes: a Meta-Analysis. *Thorax* 2004;59:843–849.

56. Dahl M, Hersh CP, Ly NP, Berkey CS, Silverman EK, Nordestgaard BG. The Protease Inhibitor PI*S Allele and COPD: a Meta-Analysis. *Eur Respir J* 2005;26:67–76.

57. Hassett C, Aicher L, Sidhu JS, Omiecinski CJ. Human Microsomal Epoxide Hydrolase: Genetic Polymorphism and Functional Expression in Vitro of Amino Acid Variants. *Hum Mol Genet* 1994;3:421–428.

58. Hu G, Shi Z, Hu J, Zou G, Peng G, Ran P. Association Between Polymorphisms of Microsomal Epoxide Hydrolase and COPD: Results From Meta-Analyses. *Respirology* 2008;13:837–850.

59. Smith CAD, Harrison DJ. Association Between Polymorphism in Gene for Microsomal Epoxide Hydrolase and Susceptibility to Emphysema. *Lancet* 1997;350:630–633.

60. Takeyabu K, Yamaguchi E, Suzuki I, Nishimura M, Hizawa N, Kamakami Y. Gene Polymorphism for Microsomal Epoxide Hydrolase and Susceptibility to Emphysema in a Japanese Population. *Eur Respir J* 2000;15:891–894.

61. Yim JJ, Park GY, Lee CT, Kim YW, Han SK, Shim YS, Yoo CG. Genetic Susceptibility to Chronic Obstructive Pulmonary Disease in Koreans: Combined Analysis of Polymorphic Genotypes for Microsomal Epoxide Hydrolase and Glutathione S-Transferase M1 and T1. *Thorax* 2000;55:121–125.

62. Rodriguez F, Jardi R, Costa X, Juan D, Galimany R, Vidal R, Miravitlles M. Detection of Polymorphisms at Exons 3 (Tyr113-->His) and 4 (His139-->Arg) of the Microsomal Epoxide Hydrolase Gene Using Fluorescence PCR Method Combined With Melting Curves Analysis. *Anal Biochem* 2002;308:120–126.

63. Zhang R, Zhang A, He Q, Lu B. [Microsomal Epoxide Hydrolase Gene Polymorphism and Susceptibility to Chronic Obstructive Pulmonary Disease in Han Nationality of North China]*. *Zhonghua Nei Ke Za Zhi* 2002;41:11–14.

64. Budhi A, Hiyama K, Isobe T, Oshima Y, Hara H, Maeda H, Kohno N. Genetic Susceptibility for Emphysematous Changes of the Lung in Japanese. *Int J Mol Med* 2003;11:321–329.

65. Korytina GF, Yanbaeva DG, Viktorova TV. Polymorphisms of the Cytochrome P450 (CYP1A1, CYP2E1) and Microsomal Epoxide Hydrolase (MEPHX) Genes in Cystic Fibrosis and Chronic Respiratory Disease. *Mol Biol (Mosk)* 2003;37:784–792.

66. Park SS, Kim EJ, Son CY, Wi JO, Park KH, Cho GJ, Ju JY, Kim KS, Kim YI, Lim SC, et al. Genetic Polymorphism of Epoxide Hydrolase and GSTM1 in Chronic Obstructive Pulmonary Disease. *Tuberc Respir Dis* 2003;55:88–97.

67. Cheng SL, Yu CJ, Chen CJ, Yang PC. Genetic Polymorphism of Epoxide Hydrolase and Glutathione S-Transferase in COPD. *Eur Respir J* 2004;23:818–824.

68. Park JY, Chen L, Wadhwa N, Tockman MS. Polymorphisms for Microsomal Epoxide Hydrolase and Genetic Susceptibility to COPD. *Int J Mol Med* 2005;15:443–448.

69. Matheson MC, Raven J, Walters EH, Abramson MJ, Ellis JA. Microsomal Epoxide Hydrolase Is Not Associated With COPD in a Community-Based Sample. *Hum Biol* 2006;78:705–717.

70. Chappell S, Daly L, Morgan K, Guetta-Baranes T, Roca J, Rabinovich R, Lotya J, Millar AB, Donnelly SC, Keatings V, et al. Genetic Variants of Microsomal Epoxide Hydrolase and Glutamate-Cysteine Ligase in COPD. *Eur Respir J* 2008;32:931–937.

71. Vibhuti A, Arif E, Deepak D, Singh B, Qadar Pasha MA. Genetic Polymorphisms of GSTP1 and MEPHX Correlate With Oxidative Stress Markers and Lung Function in COPD. *Biochem Biophys Res Commun* 2007;359:136–142.

72. Zidzik J, Slaba E, Joppa P, Kluchova Z, Dorkova Z, Skyba P, Habalova V, Salagovic J, Tkacova R. Glutathione S-Transferase and Microsomal Epoxide Hydrolase Gene Polymorphisms and Risk of Chronic Obstructive Pulmonary Disease in Slovak Population. *Croat Med J* 2008;49:182–191.

73. Yoshikawa M, Hiyama K, Ishioka S, Maeda H, Maeda A, Yamakido M. Microsomal Epoxide Hydrolase Genotypes and Chronic Obstructive Pulmonary Disease in Japanese. *Int J Mol Med* 2000;5:49–53.

74. Siedlinski M, Postma DS, Smit HA, Boezen HM. No Effects of EPHX1 Polymorphisms on the Level or Change of FEV1 in the General Population. *Eur Respir J* 2009;33:446–449.

75. Gaspar P, Moreira J, Kvitko K, Torres M, Moreira A, de Azevedo Weimer T. CYP1A1, CYP2E1, GSTM1, GSTT1, GSTP1, and TP53 Polymorphisms: Do They Indicate Susceptibility to Chronic Obstructive Pulmonary Disease and Non-Small-Cell Lung Cancer? *Genet Mol Biol* 2004;27:133–138.

76. Baranova H, Perriot J, Albuisson E, Ivaschenko T, Baranov VS, Hemery B, Mouraire P, Riol N, Malet P. Peculiarities of the GSTM1 0/0 Genotype in French Heavy Smokers With Various Types of Chronic Bronchitis. *Hum Genet* 1997;99:822–826.

77. Calikoglu M, Tamer L, Ates AN, Karakas S, Ercan B. The Association Between Polymorphic Genotypes of Glutathione S-Transferases and COPD in the Turkish Population. *Biochem Genet* 2006;44:307–319.

78. Chan-Yeung M, Ho SP, Cheung AH, So LK, Wong PC, Chan KK, Chan JW, Ip MS, Mak JC. Polymorphisms of Glutathione S-Transferase Genes and Functional Activity in Smokers With or Without COPD. *Int J Tuberc Lung Dis* 2007;11:508–514.

79. Korytina GF, Yanbaeva DG, Victorova TV. Polymorphism of Glutathione S-Transferase M1 and P1 in Patients With Cystic Fibrosis and Chronic Respiratory Diseases. *Genetika* 2004;40:401–408.

80. Sundberg K, Johansson AS, Stenberg G, Widersten M, Seidel A, Mannervik B, Jernstrom B. Differences in the Catalytic Efficiencies of Allelic Variants of Glutathione Transferase P1–1 Towards Carcinogenic Diol Epoxides of Polycyclic Aromatic Hydrocarbons. *Carcinogenesis* 1998;19:433–436.

81. Harries LW, Stubbins MJ, Forman D, Howard GC, Wolf CR. Identification of Genetic Polymorphisms at the Glutathione S-Transferase Pi Locus and Association With Susceptibility to Bladder, Testicular and Prostate Cancer. *Carcinogenesis* 1997;18:641–644.

82. Ishii T, Matsuse T, Teramoto S, Matsui H, Miyao M, Hosoi T, Takahashi H, Fukuchi Y, Ouchi Y. Glutathione S-Transferase P1 (GSTP1) Polymorphism in Patients With Chronic Obstructive Pulmonary Disease. *Thorax* 1999;54:693–696.

83. Yim JJ, Yoo CG, Lee CT, Kim YW, Han SK, Shim YS. Lack of Association Between Glutathione S-Transferase P1 Polymorphism and COPD in Koreans. *Lung* 2002;180:119–125.

84. Rodriguez F, de la RC, Jardi R, Schaper M, Vidal R, Miravitlles M. Glutathione S-Transferase P1 and Lung Function in Patients With Alpha1-Antitrypsin Deficiency and COPD. *Chest* 2005;127:1537–1543.

85. Kinnula VL, Crapo JD. Superoxide Dismutases in the Lung and Human Lung Diseases. *Am J Respir Crit Care Med* 2003;167:1600–1619.

86. Young R, Hopkins R, Black PN, Eddy C, Wu L, Gamble GD, Mills GD, Garrett JE, Eaton TE, Rees MI. Functional Variants of Antioxidant Genes in Smokers With COPD and in Those With Normal Lung Function. *Thorax* 2006;61:394–399.

87. Mak JC, Ho SP, Yu WC, Choo KL, Chu CM, Yew WW, Lam WK, Chan-Yeung M. Polymorphisms and Functional Activity in Superoxide Dismutase and Catalase Genes in Smokers With COPD. *Eur Respir J* 2007;30:684–690.

88. Houben JMJ, Mercken EM, Ketelslegers HB, Bast A, Wouters EF, Hageman GJ, Schols AM. Telomere Shortening in Chronic Obstructive Pulmonary Disease. *Respir Med* 2009;103:230–236.

89. Siedlinski M, van Diemen CC, Postma DS, Vonk JM, Boezen HM. Superoxide Dismutases, Lung Function and Bronchial Responsiveness in a General Population. *Eur Respir J* 2009;33:986–992.

90. Juul K, Tybjaerg-Hansen A, Marklund S, Lange P, Nordestgaard BG. Genetically Increased Antioxidative Protection and Decreased Chronic Obstructive Pulmonary Disease. *Am J Respir Crit Care Med* 2006;173:858–864.

91. Dahl M, Bowler RP, Juul K, Crapo JD, Levy S, Nordestgaard BG. Superoxide Dismutase 3 Polymorphism

Associated With Reduced Lung Function in Two Large Populations. *Am J Respir Crit Care Med* 2008;178:906–912.

92. Zhernakova A, van Diemen CC, C.Wijmenga. Detecting Shared Pathogenesis From the Shared Genetics of Immune-Related Diseases. *Nat Rev Genet* 2009;10:43–55.

93. Hu G, Yao W, Zhou Y, Hu J, Shi Z, Li B, Ran P. Meta- and Pooled Analyses of the Effect of Glutathione S-Transferase M1 and T1 Deficiency on Chronic Obstructive Pulmonary Disease. *Int J Tuberc Lung Dis* 2008;12:1474–1481.

94. van Diemen CC, Postma DS, Vonk JM, Bruinenberg M, Schouten JP, Boezen HM. A Disintegrin and Metalloprotease 33 Polymorphisms and Lung Function Decline in the General Population. *Am J Respir Crit Care Med* 2005;172:329–333.

95. Gosman MM, Boezen HM, van Diemen CC, Snoeck-Stroband JB, Lapperre TS, Hiemstra PS, Ten Hacken NH, Stolk J, Postma DS. A Disintegrin and Metalloprotease 33 and Chronic Obstructive Pulmonary Disease Pathophysiology. *Thorax* 2007;62:242–247.

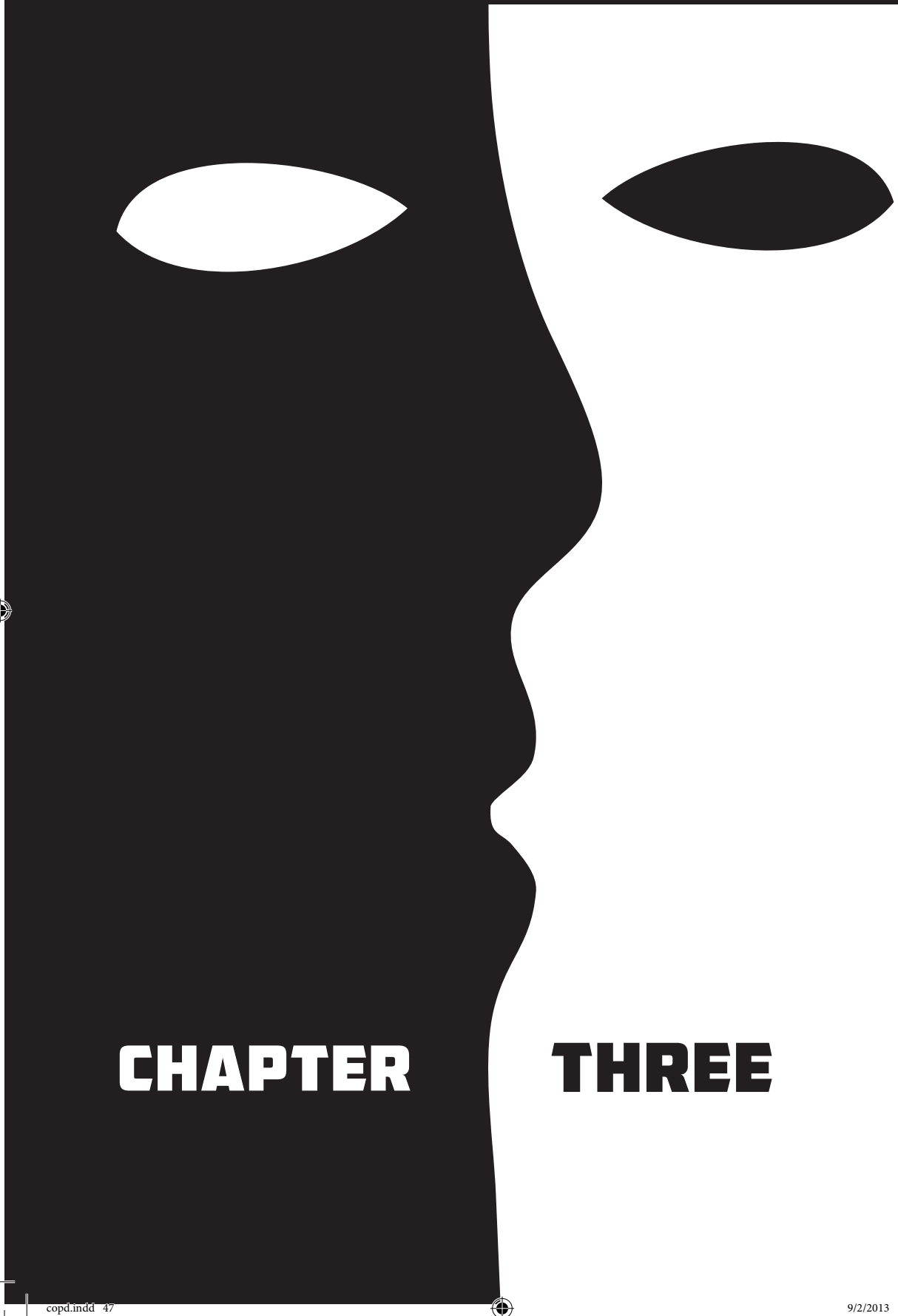
96. Hung RJ, McKay JD, Gaborieau V, Boffetta P, Hashibe M, Zaridze D, Mukeria A, Szeszenia-Dabrowska N, Lissowska J, Rudnai P, et al. A Susceptibility Locus for Lung Cancer Maps to Nicotinic Acetylcholine Receptor Subunit Genes on 15q25. *Nature* 2008;452:633–637.

97. Amos CI, Wu X, Broderick P, Gorlov IP, Gu J, Eisen T, Dong Q, Zhang Q, Gu X, Vijayakrishnan J, et al. Genome-Wide Association Scan of Tag SNPs Identifies a Susceptibility Locus for Lung Cancer at 15q25.1. *Nat Genet* 2008;40:616–622.

98. Thorgeirsson TE, Geller F, Sulem P, Rafnar T, Wiste A, Magnusson KP, Manolescu A, Thorleifsson G, Stefansson H, Ingason A, et al. A Variant Associated With Nicotine Dependence, Lung Cancer and Peripheral Arterial Disease. *Nature* 2008;452:638–642.

99. Young RP, Hopkins RJ, Hay BA, Epton MJ, Black PN, Gamble GD. Lung Cancer Gene Associated With COPD: Triple Whammy or Possible Confounding Effect? *Eur Respir J* 2008;32:1158–1164.

100. Pillai SG, Ge D, Zhu G, Kong X, Shianna KV, Need AC, Feng S, Hersh CP, Bakke P, Gulsvik A, et al. A Genome-Wide Association Study in Chronic Obstructive Pulmonary Disease (COPD): Identification of Two Major Susceptibility Loci. *PLoS Genet* 2009 Mar; vol. 5. Available from: http://www.plosgenetics.org/article/info%3Adoi%2F10.1371%2Fjournal.pgen.1000421.



CHAPTER THREE

Identification of shared genes underlying airway obstruction and emphysema.

Joanna Smolonska[1,2,3], Pieter Zanen[4], Mathieu Platteel[1], Bram van Ginneken[5], Sune F Nielsen[6], Morten Dahl[7], Marianne Benn[8], Borge G Noordestgaard[6,7,8], Filip Mejza[9], Pawel Nastalek[9], Ewa Nizankowska-Mogilnicka[9], Michael Owsijewitsch[10], Hans-Ulrich Kauczor[11], Ulrike von Seydlitz-Kurzbach[11], Kristina Forsman-Semb[12], Kerrie Tosh[13], Magnus Dahlbäck[14], Jorgen Vestbo[15,16], Jesper Holst Pedersen[17], Asger Dirksen[18], Cleo C van Diemen[1], Judith M Vonk[2] W Monique Verschuren[19], H Susan Picavet[19], Henriette A Smit[20], Cornelia M van Duijn[21], Willem Mali[22], Matthijs Oudkerk[23], Harry de Koning[24], Jan-Willem J Lammers[4], Ani Manichaikul[25,26], R Graham Barr[27,28], Harry JM Groen[29], Dirkje S Postma[3,29], Cisca Wijmenga[1], H Marike Boezen[2,3].

[1] University of Groningen, University Medical Center Groningen, Department of Genetics, Groningen, Netherlands
[2] University of Groningen, University Medical Center Groningen, Department of Epidemiology, Groningen, Netherlands
[3] University of Groningen, University Medical Center Groningen, GRIAC research institute, Groningen, The Netherlands
[4] Department of Pulmonology, Division of Heart and Lungs, UMC Utrecht, Utrecht, Netherlands
[5] Department of Radiology, Radboud University, Nijmegen Medical Centre, the Netherlands
[6] Department of Clinical Biochemistry, Herlev Hospital, Copenhagen University Hospital, University of Copenhagen, Denmark
[7] Department of Clinical Biochemistry, Rigshospitalet, Copenhagen University Hospital, University of Copenhagen, Denmark
[8] The Copenhagen City Heart Study, Frederiksberg Hospital, Copenhagen University Hospital, University of Copenhagen, Denmark
[9] Department of Respiratory Diseases, Jagiellonian University Medical College, Krakow, Poland
[10] Department of Diagnostic and Interventional Radiology, University Hospital of Heidelberg, Im Neuenheimer Feld 110, 69120 Heidelberg, Germany
[11] Translational Lung Research Center (TLRC-H), Member of the German Center for Lung Research (DZL), Im Neuenheimer Feld 350, 69120 Heidelberg, Germany
[12] AstraZeneca R&D, S-221 87 Lund, Sweden
[13] AstraZeneca, Alderley Park, Macclesfield, Cheshire, SK10 4TG, United Kingdom
[14] AstraZeneca R&D Mölndal, Mölndal, Sweden
[15] Respiratory and Allergy Research Group, Manchester Academic Sciences Centre, University of Manchester, Manchester, UK
[16] Department of Respiratory Medicine J, Odense University Hospital and University of Southern Denmark, Odense, Denmark
[17] Department of Thoracic Surgery RT, Rigshospitalet, University of Copenhagen, Denmark.
[18] Department of Respiratory Medicine, Gentofte Hospital, University of Copenhagen, Denmark
[19] Center for Prevention and Health Services Research, National Institute for Public Health and the Environment, Bilthoven, the Netherlands
[20] Julius Center for Health Sciences and Primary Care, UMC Utrecht, the Netherlands
[21] Genetic Epidemiology Unit, Department of Epidemiology, Erasmus MC University Medical Center, Rotterdam, The Netherlands
[22] Department of Radiology, UMC Utrecht, Utrecht, Netherlands
[23] University of Groningen, University Medical Center Groningen, Department of Radiology, Groningen, The Netherlands
[24] Department of Public Health, Erasmus MC, Rotterdam, Netherlands
[25] Center for Public Health Genomics, University of Virginia, Charlottesville, Virginia, United States of America
[26] Department of Public Health Sciences, Division of Biostatistics and Epidemiology, University of Virginia, Charlottesville, Virginia, United States of America
[27] Department of Medicine, College of Physicians and Surgeons, Columbia University, New York, New York, United States of America
[28] Department of Epidemiology, Mailman School of Public Health, Columbia University, New York, New York, United States of America
[29], University of Groningen, University Medical Center Groningen, Department of Pulmonology Groningen, Netherlands

Abstract

Introduction

Chronic obstructive pulmonary disease (COPD) is a complex disorder comprising airway obstruction and emphysema. Very often they co-exist in COPD patients. Both phenotypes are known to have genetic components.

Methods

We performed genome-wide association studies (GWAS) and replication studies on airway obstruction and emphysema in the COPACETIC study. Since we wanted to identify the shared genetic component of the two phenotypes, we followed a two-staged approach to identify plausible genes. We used an independent cohort to replicate these findings.

Results

Although the GWA studies on airway obstruction and emphysema did not yield genome-wide significant findings, the overlap analysis using two different approaches pointed to the focal adhesion pathway, which is crucial in epithelial repair mechanisms.

Conclusion

Our results suggest the focal adhesion pathway to be involved in coexistence of airway obstruction and emphysema. Genes composing this pathway individually point towards so called systemic effects of COPD, justifying further attention in future studies.

Introduction

The burden of COPD is increasing and it is expected to become the 3rd cause of death worldwide by 2030.[1] Besides cigarette smoking, an established risk factor for COPD, air pollution and other environmental exposures can contribute to the onset of COPD. Genetic factors are also known to contribute to disease risk,[2] and a large number of candidate gene studies have indicated several genes to be associated with COPD. However, these studies were not able to unequivocally point to causal genes or variants for COPD (reviewed in Smolonska et. al[3]). This might be due to the fact that COPD is a heterogeneous disorder comprising the presence of airway obstruction and emphysema as a single phenotype or both at the same time.[4] Airway obstruction involves inflammation and remodeling of the airways resulting in airway wall thickening in bronchi and bronchioles, leading to airflow limitation. Emphysema is defined as destruction of lung tissue resulting in alveolar enlargement leading to air trapping.[5] It is plausible that specific genetic features may underlie airway obstruction without emphysema and emphysema without airway obstruction, and potentially shared genes underlying the mixed phenotype of these two. Patients are usually diagnosed with COPD merely based on their lung function, according to the GOLD criteria (Global Initiative for Chronic Obstructive Lung Disease).[6] However, this does not elucidate whether the airway obstruction is due to inflammation and remodeling of the airway wall or results from loss of elastic lung tissue as present in emphysema.[7]

In the past few years, much effort has been put into unraveling the genetics of COPD defined based on lung function measurements. Genome-wide association studies (GWAS) have revealed *HHIP, CHRNA3/5* and *FAM13A* loci to be associated with COPD defined as airway obstruction ($FEV_1/VC < 70\%$)[8, 9] and *BICD1*[10] loci to be associated with the emphysema phenotype that is not captured by lung function alone. A few recent studies have confirmed that these genes in fact play a role in both COPD phenotypes. [11, 12]

We have taken a hypothesis-free approach to investigatewhether there is shared genetics underlying airway obstruction and emphysema. If present, this might explain why some COPD patients display features of both sub-phenotypes.

Materials and methods

Study design

Within the COPACETIC consortium framework we performed a genome-wide association (GWA) study on airway obstruction (case-control) and a GWA study on emphysema (quantitative trait) using the NELSON study (see Supplementary Material). Since

these datasets contained a significant proportion of overlapping individuals, a meta-analysis on shared genetics of both phenotypes could not be performed. We followed the analysis strategy shown in Figure 1. For details see Statistical analysis section.

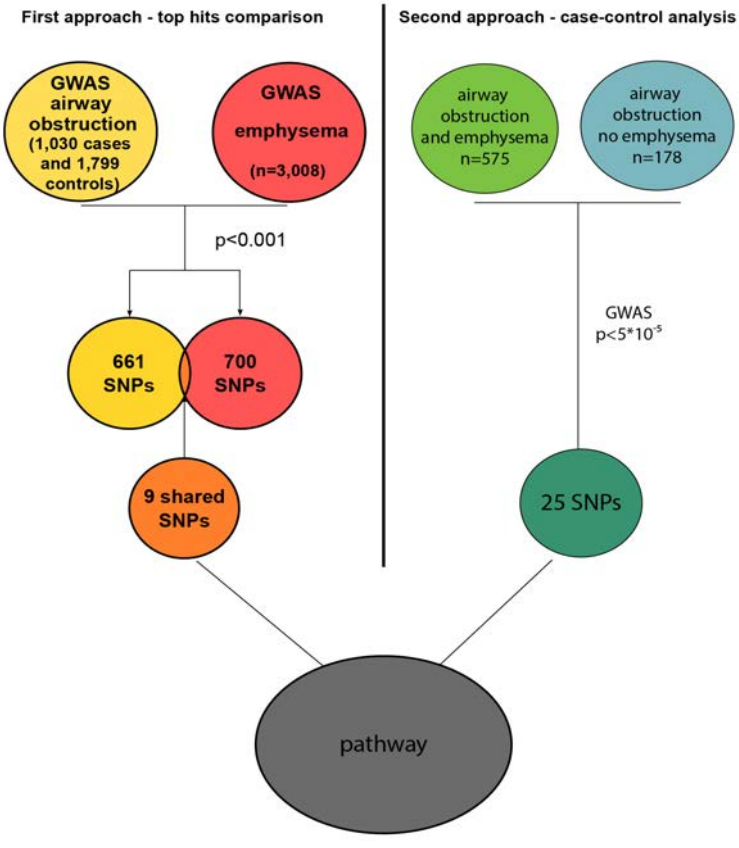


Figure 1. Analytical strategy.

Discovery cohort

We studied participants of the NELSON study which is a population-based cohort established as a lung cancer screening trial that included middle-aged, male, heavy smokers (>20 packyears) from two university medical centers in The Netherlands, Groningen and Utrecht. All participants performed spirometry and underwent computer tomography (CT) scanning to determine the presence of airway obstruction and emphysema respectively. A detailed description of this cohort can be found in van Iersel *et al.*[13]

Replication cohort

The Multi-Ethnic Study of Atherosclerosis (MESA) was used as a separate replication cohort. MESA is a longitudinal study of subclinical cardiovascular disease and risk factors that

predict progression to clinically overt cardiovascular disease or progression of the subclinical disease.[14] Between 2000 and 2002, MESA recruited 6,814 men and women 45 to 84 years of age from Forsyth County, North Carolina; New York City; Baltimore; St. Paul, Minnesota; Chicago; and Los Angeles. Exclusion criteria were clinical cardiovascular disease, weight exceeding 136 kg (300 lb.), pregnancy, and impediment to long-term participation. A small number of additional participants were included from the MESA Air Pollution Study, which recruited in Los Angeles and Rockland County, New York, under the MESA criteria except that participants were ages 50 to 89 years.[15] Given the race/ethnic distribution of NELSON, we limited the MESA sample for replication to Caucasians.

Phenotype definitions

Airway obstruction cases were defined as having a (pre-bronchodilator) $FEV_1/FVC < 0.7$. Controls were defined as having $FEV_1/FVC > 0.7$ and $FEV_1 > 90\%$ predicted.

Firstly, emphysema was investigated as a quantitative trait based on CT scan measurements. We used the 15th percentile of density distribution (p15) in the lungs. To correct for scanner variations, density values in each scan were shifted to ensure that pure air located in the trachea had a density corresponding to -1000 HU.

Secondly, presence and absence of emphysema was defined according to the lowest and highest quartile of the emphysema distribution, being $P_{25<}$<-925HU and P_{75}>-898HU respectively (see Statistical analysis for details on a second approach).

Emphysema measures were obtained on full-lung screening CT scans in NELSON and partial-lung cardiac CT scans in MESA.[16] The latter was adjusted for attenuation of air outside the body and has been previously validated compared to full-lung scans.[17]

Genotyping and quality control

DNA samples were hybridized on Illumina Quad610 BeadChip containing over 620,000 markers providing whole genome coverage. Genotypes were called with the standard algorithm provided by Illumina and implemented in GenomeStudio software. Quality control, as well as analysis, was performed with PLINK[18]. SNPs were excluded if a deviation from Hardy-Weinberg equilibrium was observed ($p<0.0001$) in controls only in the airway obstruction analysis (n=16,003) and in all subjects in the emphysema analysis (n=3,036). SNPs were excluded also in case of missing genotypes in more than 5% of samples (i.e. airway obstruction n=38,842, emphysema n=10,107) or if minor allele frequency was below 5% (i.e. airway obstruction n=53,681, emphysema n=64,251). Samples were excluded if more than 5% of genotype data was missing (i.e. airway obstruction n=24, emphysema n= 19) or in case of

duplicate samples (n= 29); detected as an ethnic outlier (based on genetic distance derived from principal components c1 and c2; n= 49); derived from a relative of another participant (based on IBS estimation, Phat > 0.5, n= 13); or had a diagnosis of lung cancer (n=16).

Participants in the original MESA cohort who consented to genetic analyses were genotyped in 2009 using the Affymetrix Human SNP array 6.0. Data were filtered on SNP level call rate <95%, individual level call rate <95%, heterozygosity >53%, and all monomorphic SNPs were removed.

Statistical analysis

To identify genetic overlap between airway obstruction and emphysema, we selected SNPs with p<0.001 from each GWA analysis and checked if the two resulting lists of top hits contained overlapping SNPs (supplementary data). Next, to check whether findings from this approach are not due to chance, we selected a subset of cases from NELSON participants that had both obstruction and emphysema (FEV_1<0.7 and p15<-925HU) and a subset with airway obstruction only (FEV_1>0.7, FEV_1>90% predicted and p15>-898HU) as controls to perform case-control analysis (second approach). We compared these groups using the Chi-squared test.

For replication of the NELSON findings among Caucasians in MESA, we excluded individuals with top principal components (PCs) of ancestry > 3.5 SD. Based on our examination of principal components within each race/ethnic group, we used 3 PCs to account for population structure within the MESA. Genetic analyses were performed as closely as possible to those performed in the NELSON cohort, with some modifications to account for study-specific confounders in MESA.

To replicate findings of the NELSON study on airway obstruction and emphysema overlap the following analyses were performed in the MESA cohort:

1) Linear regression on p15 adjusted for packyears, age, ex-/current smoking status and gender and logistic regression within cases with (AO+) and without (AO-) airway obstruction to replicate nine SNPs resulting from the top hits lists comparison
2) Logistic regression adjusting for gender within airway obstruction cases with emphysema (FEV_1<0.7 and p15<-925HU; AO+E+) and without emphysema (FEV_1<0.7 and p15>-898 HU; AO+E-) to replicate SNPs found in case-control analysis ($p<5*10^{-05}$).

Networks and pathways

We used publicly available tools to identify networks of genes or proteins and checked if they represent an established pathway. For the first purpose, we used GeneMANIA[19], a web-based tool that finds genes that interact in any way (physically, genetically, etc) with

query genes. We used application version 3.0.4 that was accessed on 26 May. For pathway identification, we used GATHER[20], a web-based annotation tool, that among other functions looks for KEGG pathways and Gene Ontology (GO) terms.

Results

Genetic overlap between airway obstruction and emphysema

We performed GWA studies for airway obstruction and emphysema separately (Supplementary Material). Comparing the top 661 SNPs from the airway obstruction association analysis and top 700 SNPs from the emphysema association analyses respectively (based on p cut-off < 0.001, Figure 1), we found nine SNPs showing moderate associations with both airway obstruction and emphysema (Table 1). These SNPs point to *ABCB1, SLC30A10, ACOT11, KIAA1239,* and *ALPK2* genes that were subsequently investigated in GeneMANIA to identify potential protein networks. The resulting network (Figure 2) is enriched in proteins involved in response to drug (GO:0042493) and focal adhesion (hsa04510) pathways. Interestingly, the network includes *IGLL1* and *PARVA* genes that, among others, were also identified amongst the genes close to the 25 most significant SNPs from our case-control study comparing 575 airway obstruction cases with emphysema, to 178 airway obstruction cases without emphysema (Figure 1, Table 2) from the NELSON cohort. Although this study did not reveal genome-wide significant results, a pathway analysis including the ten most significant SNPs ($p< 2*10^{-5}$) again identified the focal adhesion pathway.

Table 1. 9 SNPs found to be moderately associated with airway obstruction and emphysema with the first approach (top hit lists comparisons).

CHR	SNP	A1	$BETA_{p15}$	P_{p15}	Rank P15	OR_{ao}	P_{ao}	Rank ao	gene	left_gene	right_gene
1	rs300269	G	1.68	5.06E-04	365	0.83	6.45E-04	440	ACOT11	LOC645436	FAM151A
1	rs6700061	A	-3.28	4.32E-04	322	1.43	7.19E-04	489	SLC30A10	LOC728518	LOC728528
1	rs17641611	A	2.74	1.60E-05	14	0.76	3.24E-04	221	NA	LOC643779	LOC100129664
4	rs1469807	A	-2.15	7.43E-05	60	1.23	7.91E-04	536	KIAA1239	LOC100130532	C4orf19
7	rs17064	T	3.19	5.75E-04	411	0.69	7.05E-04	476	ABCB1	ABCB4	RUNDC3B
13	rs1409255	C	1.76	3.28E-04	240	0.81	2.03E-04	143	NA	LOC100129308	DIAPH3
18	rs9789183	G	-2.52	2.50E-07	1	1.21	7.54E-04	512	NA	LOC645355	TTMA
18	rs4271673	A	1.81	9.52E-04	654	0.81	8.28E-04	562	ALPK2	NEDD4L	LOC100128163
20	rs6082843	A	1.62	7.97E-04	561	0.83	9.12E-04	609	NA	KRT18P3	CYB5P4

BETA p15 is derived from linear regression performed on 15th percentile of density distribution, expressed in Hounsfield units, which has negative values. These estimates should be interpreted in the following way: BETA< 0 corresponds to more severe emphysema and BETA > 0 corresponds to a protective effect of the SNP.

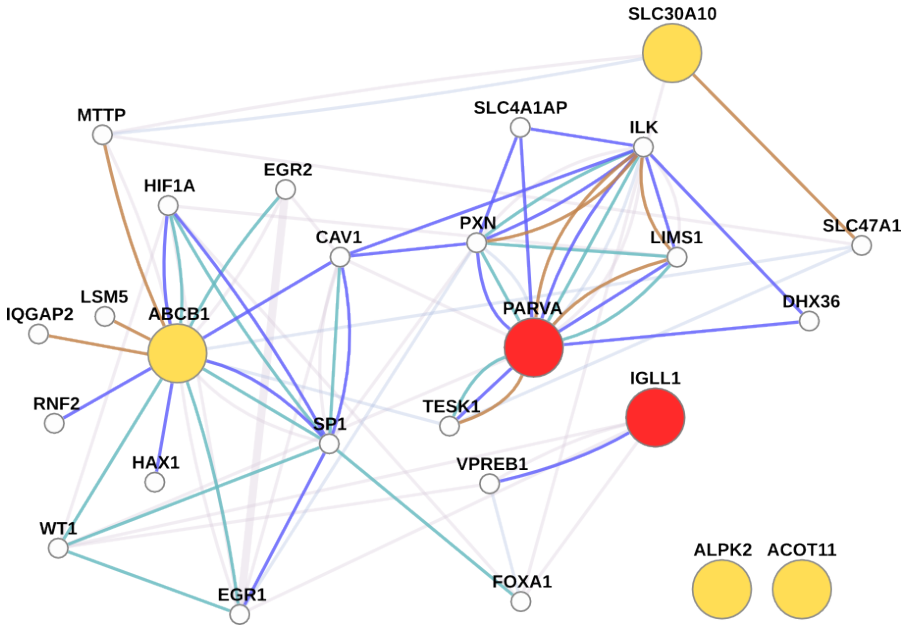


Figure 2. Network analysis for 4 genes found with first approach (comparison). In red genes predicted to be involved in the network and identified with the second approach (case-control analysis).

Table 2. 25 SNPs found to be associated in a second approach – case-control analysis, where cases suffered from airway obstruction and emphysema and controls had airway obstruction only.

CHR	SNP	A1	P	OR	gene	left_gene	right_gene
2	rs16863440	G	1.59E-06	0.34	NA	LOC727982	SOX11
11	rs11022378	C	2.84E-06	1.79	PARVA	MICALCL	TEAD1
22	rs738781	G	9.03E-06	0.50	NA	LOC388882	IGLL1
22	rs915583	A	9.03E-06	0.50	NA	LOC388882	IGLL1
22	rs6003752	G	1.01E-05	0.50	NA	LOC388882	IGLL1
22	rs5996566	C	1.21E-05	0.50	NA	LOC388882	IGLL1
19	rs1613070	G	1.29E-05	0.59	KHSRP	LOC390877	KHSRP
4	rs286451	A	1.39E-05	0.58	NA	PAQR3	ARD1B
11	rs7110901	A	1.84E-05	2.76	LDLRAD3	LOC100128841	C11orf55
2	rs10865212	C	1.95E-05	0.57	PRKCE	SRBD1	EPAS1
1	rs12403920	G	2.01E-05	2.41	NA	RIMKLA	ZMYND12
4	rs4975160	G	2.12E-05	0.58	NA	PAQR3	ARD1B
23	rs2071583	C	2.13E-05	0.27	PHEX	SMS	ZNF645
16	rs7197966	A	2.68E-05	1.71	A2BP1	LOC100129334	LOC440337
4	rs6448032	G	3.26E-05	0.59	KCNIP4	PACRGL	NCRNA00099
19	rs1290617	C	3.35E-05	1.72	CYP4F3	CYP4F8	LOC646575
4	rs6448033	A	3.54E-05	0.59	KCNIP4	PACRGL	NCRNA00099
19	rs1290616	A	3.78E-05	1.71	CYP4F3	CYP4F8	CYP4F3
3	rs12494491	A	4.02E-05	0.59	NA	SPSB4	ACPL2
5	rs17165388	A	4.15E-05	0.53	NA	PPP2CA	CDKL3
11	rs4755440	G	4.44E-05	2.07	LDLRAD3	LOC100128841	C11orf55
16	rs12445310	A	4.63E-05	0.55	NA	CDH8	hCG_1642987
11	rs7107848	G	4.81E-05	0.51	PARVA	MICALCL	TEAD1
12	rs1465934	A	4.82E-05	0.51	PTPRB	KCNMB4	PTPRR
12	rs2165627	C	4.82E-05	0.51	PTPRB	KCNMB4	PTPRR

For the replication we followed two approaches:

1) We set out to replicate the nine SNPs showing moderate associations with both airway obstruction and emphysema among Caucasians in the MESA study. Two out of the nine SNPs, i.e. rs 17641611 and rs1469807 were replicated with nominal p values in the emphysema association analysis (Table 3). rs17641611 (p=0.017) is located on chromosome 1 in a gene desert with pseudogenes and RNA coding genes. The second SNP, rs1469807 (p=0.023) on chromosome 4 annotates to *KIAA1239*. However, none of the overlapping nine SNPs showed an association with airway obstruction (Table 3).

Table 3. 9 overlapping SNPs – replication in MESA cohort.

CHR	SNP	A1	N.p15	Eff.p15	SE.p15	P.p15	N.AO	OR.AO	CI.AO	P.AO
1	rs300269	G	2421	0.43	0.58	0.46	884	0.91	0.73 – 1.13	0.37
1	rs6700061	A	2423	-0.05	1.10	0.96	886	0.80	0.54 – 1.19	0.26
1	rs17641611	A	2419	-1.84	0.77	0.02	884	1.19	0.90 – 1.58	0.21
4	rs1469807	A	2423	-1.49	0.66	0.02	886	0.85	0.67 – 1.08	0.18
7	rs17064	T	2423	0.23	1.17	0.85	886	0.90	0.58 – 1.40	0.64
13	rs1409255	C	2423	0.85	0.59	0.15	886	0.91	0.74 – 1.12	0.36
18	rs9789183	G	2423	-0.41	0.67	0.54	886	1.16	0.91 – 1.48	0.23
18	rs4271673	A	2423	0.59	0.64	0.35	886	0.93	0.73 – 1.17	0.52
20	rs6082843	A	2423	0.21	0.58	0.71	886	1.02	0.82 – 1.26	0.85

OR.AO has been converted from BETA estimate derived from logistic regression to aid comparisons of the results from Table 1.

2) We set out to replicate the 25 SNPs identified by comparing airway obstruction cases with and without emphysema, also among Caucasiansin the MESA study. Two SNPs showed nominal replication: rs10865212 on chromosome 2 (p=0.035) annotating to *PRKCE* and rs12403920 on chromosome 1 (p=0.039) located in the region between *RIMKLA* and *ZMYND12*. The third SNP rs6448032 on chromosome 4 annotating to *KCNIP4* showed suggestive association (p=0.066)(Table 4).

Table 4. Replication results from MESA cohort of the 25 SNPs identified in case-control analysis

CHR	SNP	A1	N.AOemph	OR.AOemph	95% CI	P.AOemph
2	rs10865212	C	243	0.41	0.18 – 0.94	0.04
1	rs12403920	G	243	4.10	1.08 – 15.61	0.04
4	rs6448032	G	243	0.45	0.19 – 1.06	0.07
19	rs1290616	A	243	1.88	0.86 – 4.08	0.11
19	rs1290617	C	243	1.85	0.84 – 4.08	0.12
16	rs7197966	A	243	0.56	0.26 – 1.18	0.13
3	rs12494491	A	243	2.03	0.76 – 5.41	0.15
4	rs286451	A	243	1.79	0.70 – 4.58	0.23
11	rs11022378	C	243	0.60	0.26 – 1.38	0.23
11	rs7110901	A	243	2.32	0.57 – 9.46	0.24
4	rs6448033	A	243	0.64	0.29 – 1.38	0.25
11	rs4755440	G	243	1.60	0.61 – 4.19	0.34
19	rs1613070	G	243	1.44	0.56 – 3.66	0.45
5	rs17165388	A	242	1.37	0.46 – 4.11	0.58
11	rs7107848	G	243	1.48	0.32 –6.92	0.62
2	rs16863440	G	243	0.81	0.22 – 3.00	0.75
22	rs5996566	C	243	0.89	0.35 – 2.23	0.80
22	rs738781	G	243	0.89	0.36 –2.22	0.80
22	rs915583	A	243	0.89	0.36 – 2.22	0.80
16	rs12445310	A	243	0.91	0.38 –2.19	0.84
4	rs4975160	G	243	1.09	0.47 – 2.51	0.84
12	rs1465934	A	243	0.94	0.35 – 2.57	0.91
12	rs2165627	C	243	0.94	0.35 – 2.57	0.91

OR.AOemph has been converted from BETA estimate derived from logistic regression to aid comparisons of the results from Table 2.

Discussion

In the current study we performed an analysis on genetic overlap between the objective presence of airway obstruction and emphysema. We focused on the genes that most significantly pointed towards a network of genes involved in the focal adhesion pathway that is an important component of the epithelial repair. Specifically *ABCB1* as a gene involved in both airway obstruction and emphysema is an interesting candidate. *ABCB1* (MDR-1) belongs to the family of ABC-transporters, just as *ABCC1* (MRP-1) that we previously identified as a candidate gene for COPD[21] with functional relevance. It has been shown that *Mrp1/Mdr1a/1b* triple knock-out mice have a reduced inflammatory response upon cigarette smoke exposure, but the role of MDR-1 in COPD development has to be further explored.[22]

Other genes resulting from the airway obstruction-emphysema overlap analysis seem to be good candidates as well. *ACOT11,* coding for acyl-CoA thioesterase 11 has been shown to increase metabolic activity of brown fat tissue in mice exposed to cold.[23] A specific SNP in this gene, rs300269, could be linked to weight loss which is an important co-morbidity in about half of the COPD patients.[24] Since little is known about the functionality of the other genes, i.e. *ALPK2, SLC30A10* and *KIAA1239,* interpretation of those results in relation to COPD

is difficult.

We suggest some plausible candidates that should be investigated further in future studies. Some interesting pathways may emerge from our study, pathways which were not investigated yet in detail in relation to COPD. It is known already for quite some years, that COPD is not only a lung disease, comprising airway obstruction, chronic bronchitis and emphysema, but there are also so-called "systemic effects" accompanying the disease.[25] These effects include among others muscle weakening and weight loss which usually correspond to poor prognosis in COPD patients, and often in significant drops in quality of life. Most literature presents these effects as a consequence or naturally occurring events during the disease progression, although origins and mechanisms underlying these events are poorly understood. We think our study broadens the spectrum of candidate genes and pathways that could be related to COPD, and stresses their potentially important role in COPD development. However, we cannot rule out the possibility that muscle weakening and higher energy expenditure resulting in weight loss can significantly contribute to the systemic inflammation either, since we do not know the sequence of events.

However, our results should be interpreted with caution, as none of the SNPs reached genome-wide significance. This might be due to a number of reasons. First, we used a single discovery cohort to investigate two phenotypes, hence we could not perform a meta-analysis to combine the results, but used a two-staged approach to strengthen our findings. However, the GWA studies on each of the phenotypes did not result in genome-wide significant findings. The power of each study (1,030 cases and 1,799 controls in airway obstruction and 3,047 individuals in emphysema analyses) might be limited, and be an underlying factor explaining the lack of genome-wide significance. Although the results of the airway obstruction and emphysema overlap analysis were validated in an additional replication step in an independent cohort, further replication was not possible due to lack of cohorts with both spirometry and CT data available. For individual GWAS on airway obstruction (supplementary material) the replication was unsuccessful, most likely due to large heterogeneity between cohorts. Our attempt to account for it decreased the replication sample size and hence the power. Replication cohorts for airway obstruction were mostly general-population cohorts, which are certainly a lot different than the NELSON cohort being a highly selected population of heavy smokers. For replication of individual GWAS on emphysema two cohorts were available with less than 2,500 individuals in total raising again the power problem.

We should also highlight the fact that many of the NELSON participants actually suffer from both airway obstruction and emphysema. Therefore, it cannot be ruled out that associations with airway obstruction already represent SNPs associated with emphysema as well. On the other hand, there is no direct overlap between results from airway obstruction

GWAS and emphysema GWAS (Supplementary data), which implies that the phenotypes are different. Further, it is known that presence of emphysema leads to airway obstruction, therefore "pure emphysema" cases can soon become patients with both phenotypes. It can not be excluded that these patients carry additional genetic load predisposing them to both phenotypes or the genes associated with emphysema lead also to the development of airway obstruction. This question could not be answered by the current study due to very small numbers of individuals with "pure phenotypes". Taking this into account and results of the current study, we think that further research should include all (or as many as possible) aspects of COPD and include patients with both spirometry and CT measurements.

In summary, our current results suggest a possible overlap in genetic components of airway obstruction and emphysema. The SNPs we describe are located within genes pointing to focal adhesion pathway and so-called "systemic effects" of COPD.

Acknowledgments

The COPACETIC study was funded by EU FP7 grant (201379). Authors would like to thank participants of NELSON cohort, Vlagtwedde/Vlaardingen cohort, Doetinchem cohort, GLUCOLD study, Rucphen study, BOLD study, Copenhagen City Heart Study, DLCST study and Lifelines study.

The MESA Lung/SHARe Study was funded by NIH grant RC1HL100543. The MESA Lung and MESA COPD Studies are funded by NIH grants R01HL077612, and R01HL093081. MESA and the MESA SHARe project are conducted and supported by contracts N01-HC-95159 through N01-HC-95169 and RR-024156 from the National Heart, Lung, and Blood Institute (NHLBI). MESA Air is conducted and supported by the United States Environmental Protection Agency (EPA) in collaboration with MESA Air investigators, with support provided by grant RD83169701. Funding for MESA SHARe genotyping was provided by NHLBI Contract N02 HL 6 4278. MESA Family is conducted and supported in collaboration with MESA investigators; support is provided by grants and contracts R01HL071051, R01HL071205, R01HL071250, R01HL071251, R01HL071252, R01HL071258, R01HL071259, M01-RR00425, UL1RR033176, and UL1TR000124. The authors thank the participants of the MESA study, the Coordinating Center, MESA investigators, and study staff for their valuable contributions. A full list of participating MESA investigators and institutions can be found at http://www.mesa-nhlbi.org.

References

1. World Health organisation. http://www who int/whr/2004/annex/topic/en/annex_2_en pdf, 2004.
2. Steiling K, Lenburg ME, Spira A. Airway gene expression in chronic obstructive pulmonary disease. Proc Am Thorac Soc 2009;**6**:697-700.
3. Smolonska J, Wijmenga C, Postma DS, Boezen HM. Meta-analyses on suspected chronic obstructive pulmonary disease genes: a summary of 20 years' research. Am J Respir Crit Care Med 2009;**180**:618-631.
4. Bafadhel M, Umar I, Gupta S et al. The role of CT scanning in multidimensional phenotyping of COPD. Chest 2011;**140**:634-642.
5. Hogg JC, Timens W. The pathology of chronic obstructive pulmonary disease. Annu Rev Pathol 2009;**4**:435-459.
6. Vestbo J, Hurd SS, Agusti AG et al. Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: GOLD executive summary. Am J Respir Crit Care Med 2013;**187**:347-365.
7. O'Donnell RA, Peebles C, Ward JA et al. Relationship between peripheral airway dysfunction, airway obstruction, and neutrophilic inflammation in COPD. Thorax 2004;**59**:837-842.
8. Cho MH, Boutaoui N, Klanderman BJ et al. Variants in FAM13A are associated with chronic obstructive pulmonary disease. Nat Genet 2010;**42**:200-202.
9. Pillai SG, Ge D, Zhu G et al. A genome-wide association study in chronic obstructive pulmonary disease (COPD): identification of two major susceptibility loci. PLoS Genet 2009;**5**:e1000421.
10. Kong X, Cho MH, Anderson W et al. Genome-wide Association Study Identifies BICD1 as a Susceptibility Gene for Emphysema. Am J Respir Crit Care Med 2011;**183**:43-49.
11. Pillai SG, Kong X, Edwards LD et al. Loci identified by genome-wide association studies influence different disease-related phenotypes in chronic obstructive pulmonary disease. Am J Respir Crit Care Med 2010;**182**:1498-1505.
12. Lambrechts D, Buysschaert I, Zanen P et al. The 15q24/25 susceptibility variant for lung cancer and chronic obstructive pulmonary disease is associated with emphysema. Am J Respir Crit Care Med 2010;**181**:486-493.
13. van Iersel CA, de Koning HJ, Draisma G et al. Risk-based selection from the general population in a screening trial: selection criteria, recruitment and power for the Dutch-Belgian randomised lung cancer multi-slice CT screening trial (NELSON). Int J Cancer 2007;**120**:868-874.
14. Bild DE, Bluemke DA, Burke GL et al. Multi-ethnic study of atherosclerosis: objectives and design. Am J Epidemiol 2002;**156**:871-881.
15. Kaufman JD, Adar SD, Allen RW et al. Prospective study of particulate air pollution exposures, subclinical atherosclerosis, and clinical cardiovascular disease: The Multi-Ethnic Study of Atherosclerosis and Air Pollution (MESA Air). Am J Epidemiol 2012;**176**:825-837.
16. Carr JJ, Nelson JC, Wong ND et al. Calcified coronary artery plaque measurement with cardiac CT in population-based studies: standardized protocol of Multi-Ethnic Study of Atherosclerosis (MESA) and Coronary Artery Risk Development in Young Adults (CARDIA) study. Radiology 2005;**234**:35-43.
17. Hoffman EA, Jiang R, Baumhauer H et al. Reproducibility and validity of lung density measures from cardiac CT Scans--The Multi-Ethnic Study of Atherosclerosis (MESA) Lung Study. Acad Radiol 2009;**16**:689-699.
18. Purcell S, Neale B, Todd-Brown K et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 2007;**81**:559-575.
19. Mostafavi S, Ray D, Warde-Farley D, Grouios C, Morris Q. GeneMANIA: a real-time multiple association network integration algorithm for predicting gene function. Genome Biol 2008;**9 Suppl 1**:S4.
20. Chang JT, Nevins JR. GATHER: a systems approach to interpreting genomic signatures. Bioinformatics 2006;**22**:2926-2933.

21. Siedlinski M, Boezen HM, Boer JM, Smit HA, Postma DS. ABCC1 polymorphisms contribute to level and decline of lung function in two population-based cohorts. Pharmacogenet Genomics 2009;**19**:675-684.

22. van der Deen M, Timens W, Timmer-Bosscha H et al. Reduced inflammatory response in cigarette smoke exposed Mrp1/Mdr1a/1b deficient mice. Respir Res 2007;**8**:49.

23. Adams SH, Chui C, Schilbach SL et al. BFIT, a unique acyl-CoA thioesterase induced in thermogenic brown adipose tissue: cloning, organization of the human gene and assessment of a potential link to obesity. Biochem J 2001;**360**:135-142.

24. Agust AG, Gari PG, Sauleda J, Busquets X. Weight loss in chronic obstructive pulmonary disease. Mechanisms and implications. Pulm Pharmacol Ther 2002;**15**:425-432.

25. Agusti AG. Systemic effects of chronic obstructive pulmonary disease. Proc Am Thorac Soc 2005;**2**:367-370.

Supplementary material.

Supplementary methods.

COPACETIC Discovery cohort and phenotype definition

The work described in the current manuscript was performed within the COPACETIC consortium framework (See Figure S1 for set up of the study).

The GWA on airway obstruction was performed 1,030 cases and 1,799 controls (including 846 controls from the blood bank). Airway obstruction cases were defined as having a (pre-bronchodilator) $FEV_1/FVC < 0.7$. Controls were defined as having $FEV_1/FVC > 0.7$ and $FEV_1 > 90\%$ predicted.

GWA study on quantitative emphysema was performed on 3,008 individuals from the NELSON cohort. Emphysema was investigated as a quantitative trait based on CT scan measurements. We used the 15th percentile of density distribution (p15) adjusted for peak in the trachea to account for possible center differences. We checked the normality of the distribution.

Statistical analysis

The GWA on airway obstruction was performed in the identification cohort, NELSON, with the Chi-squared test with no additional adjustments. Likewise, each of the nine replication cohorts was analyzed separately, with additional adjustment for gender. Results were combined in a directional fixed effects (Mantel-Haenszel) meta-analysis.

The GWA on emphysema was performed using a linear regression additive model with adjustment for age and pack-years smoked. The two cohorts available for replication (LUSI and DLCST) were analyzed likewise. Final results were obtained in a directional meta-analysis.

COPACETIC Replication cohorts

<u>COPACETIC Replication of airway obstruction GWAS results</u>

Nine independent cohorts were used as a replication sample for airway obstruction. The Vlagtwedde/Vlaardingen study (n=1,293) and the Doetinchem study (n= 898) are longitudinal, population-based cohorts of Caucasians of Dutch decent with a follow-up of 25 and 15 years, respectively.[1, 2] The Rucphen study (n= 158) and GLUCOLD study (n= 315) are patient populations with mild to moderate COPD (GOLD stage I and II, respectively) and their controls, also of Dutch descent.[3, 4] Burden of Obstructive Lung Disease (BOLD) is a Polish

cohort of 147 COPD patients and 355 matched healthy controls[5]. The Copenhagen City Heart Study (CCHS,n= 8,627) is a population-based cohort of Danish descent with over 1,000 COPD cases.[6] The Danish Lung Cancer Screening Trial (DLCST, n= 1,202) is a population-based cohort, recruited in Denmark for a lung cancer screening trial.[7] NELSON Extra is an additional sample of the NELSON cohort (n=488). The LifeLines study consists of a population based sample of Northern part of the Netherlands including Caucasians of Dutch decent (n=3,724).[8]

COPACETIC Replication of emphysema GWAS results

Two cohorts had CT measurements and were available for emphysema associations replication: DCLST (n= 1,465) and German Lung Cancer Screening Intervention Trial (LUSI; n=1,013) which comprises individuals screened for lung cancer and COPD

MESA Replication

Among Caucasians enrolled into MESA with genotypic data, 886 had valid measures of pulmonary function and emphysema measures were available for 2,323. Airway obstruction cases were defined as in COPACETIC, and there were 243 such cases among Caucasians in MESA.

The statistical analysis was performed as in COPACETIC, except that the following additional variables were included in the regression models to account for scanner effects: scanner, site, and milliamperes.

Replication Genotyping

Replication genotyping was done with a custom made Illumina GoldenGate array (airway obstruction) and Veracode assay (emphysema) containing 312 and 71 SNPs with $p<5*10^{-5}$, respectively. SNPs were clustered in GenomeStudio software with a standard algorithm and adjusted manually separately for each replication cohort. Quality control was performed as described before (main manuscript).

Accounting for heterogeneity in airway obstruction association replication

To reduce heterogeneity to a minimum we followed the approach shown in Supplementary Figure 2. We performed a sensitivity analysis for every SNP separately, by excluding one cohort at a time. If homogeneity was achieved cohorts were included in a final meta-analysis. If no homogeneity was observed we performed sensitivity analysis excluding two cohorts at a time, using all possible pair -wise comparisons. These steps were repeated with three, four, five and six cohorts excluded at a time. Prior to the sensitivity analysis the minimum number of replication cohorts was set to be three, to ensure robustness of our results.

Results

Airway obstruction

In a final association analysis, 492,339 SNPs were tested in 1,030 cases and 1,799 controls. The genomic inflation factor (**λ**) equaled 1.02, indicating an appropriate quality control and no population stratification. Genome-wide significance was not reached by any of the SNPs (Supplementary Figure 3). We selected 312 SNPs with the lowest p values ($p<0.0005$) for replication in nine independent cohorts giving a total of 3,236 cases and 13,956 controls. None of the SNPs exceeded the Bonferroni-corrected p value in a replication phase ($1.60 \cdot 10^{-4}$, 0.05/312; Supplementary Table 1) or reached genome-wide significance level after combining the results of the identification cohort and the replication cohorts ($1.02 \cdot 10^{-7}$, 0.05/492,339).

Emphysema

We included 3,047 individuals and 522,104 SNPs in a linear regression on emphysema. We did not observe an abnormal deviation of association signals, as **λ** equaled 1.03. There were no genome-wide significant signals at this stage (Supplementary Figure 4). 71 SNPs with p $<5 \cdot 10^{-5}$ were selected for replication in two population-based cohorts. 2,476 individuals and 66 SNPs passed quality checks in a replication phase. None of the SNPs reached the Bonferroni-corrected p value ($7.58 \cdot 10^{-4}$, 0.05/66; Supplementary Table 2) or genome-wide significance level after combining all the results ($9.58 \cdot 10^{-8}$, 0.05/522,104).

Concluding remark

We found no genome-wise significant signals in the discovery population when performing the genome-wide association study, trying to identify genes involved in airway obstruction and emphysema, two major phenotypes of COPD. We found some SNPs with a trend of association in the replication cohorts, but no significance after multiple testing corrections, neither for airway obstruction nor for emphysema. In both cases, this can be explained by the fact, that these phenotypes of COPD are complex and vary much between all the cohorts used in this study, introducing a lot of heterogeneity. Especially in the emphysema phenotype, where computer tomography is used to assess its presence, differences in type of CT-scanner and software used can introduce noise and make replication more difficult. Moreover, the size of our replication sample for the emphysema phenotype was small and thus had limited power. In addition, the differences between populations studied as well as confounding factors that were not accounted for could be the reason that major signals found in the discovery set were not replicated. In the GWA on airway obstruction we included blood

bank controls to increase the power. However, except for gender, there is no information available for these individuals, hence we could not account for smoking habits in this analysis. Also, taking into account the relatively small size of the discovery sets (1,030 cases and 1,799 controls in the airway obstruction and 3,047 individuals in the emphysema analyses) the initial signals from the identification cohort could simply have been false positives.

However in light of findings made in Soler Artigas et al[9] in lung function parameters we think our results are valid, even despite lack of genome-wide significance. Soler Artigas in their meta-analysis of 23 cohorts summing up to over 48,000 individuals and replication sample of over 46,000 individuals identified 16 novel loci associated with lung function, among them transforming growth factor β (*TGFB2*). In the current study, GWAS in airway obstruction identified rs1317681 (combined $p=2.17*10^{-06}$) located also in *TGFB2* gene.

In conclusion, the current study showed no genome-wide significant findings however, our results are suggestive and strengthened by findings from previous studies by other groups. Future studies should include higher numbers of cases and controls, which can be only done within consortia with detailed phenotyping data enabling sub-phenotype studies.Supplementary

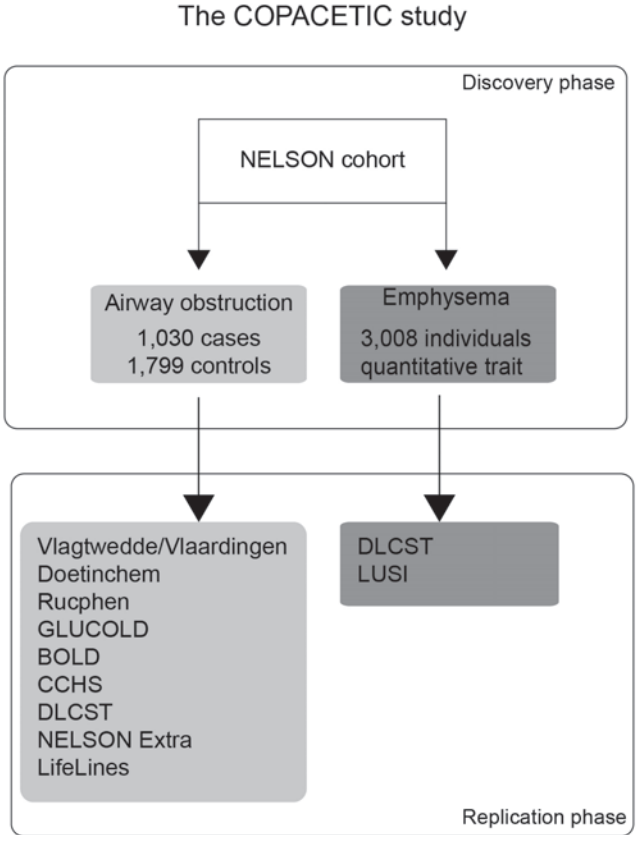
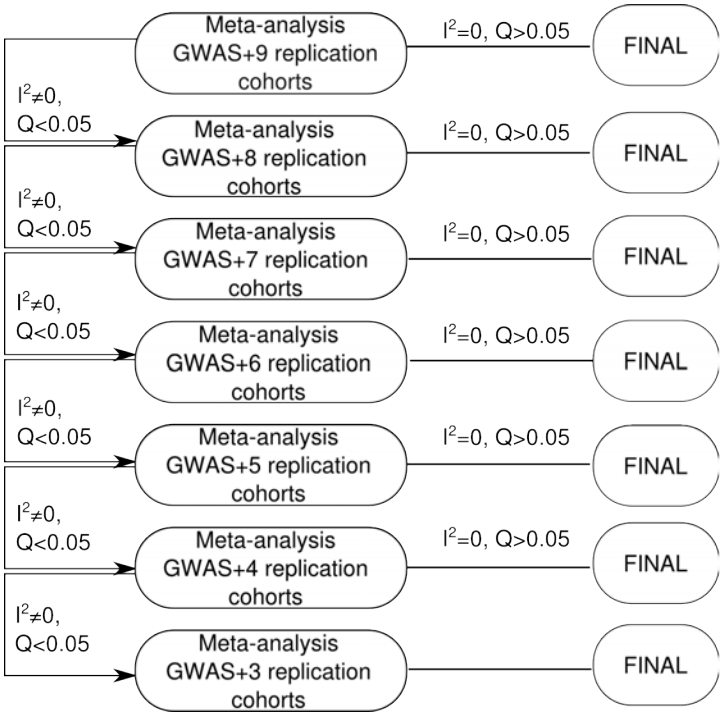
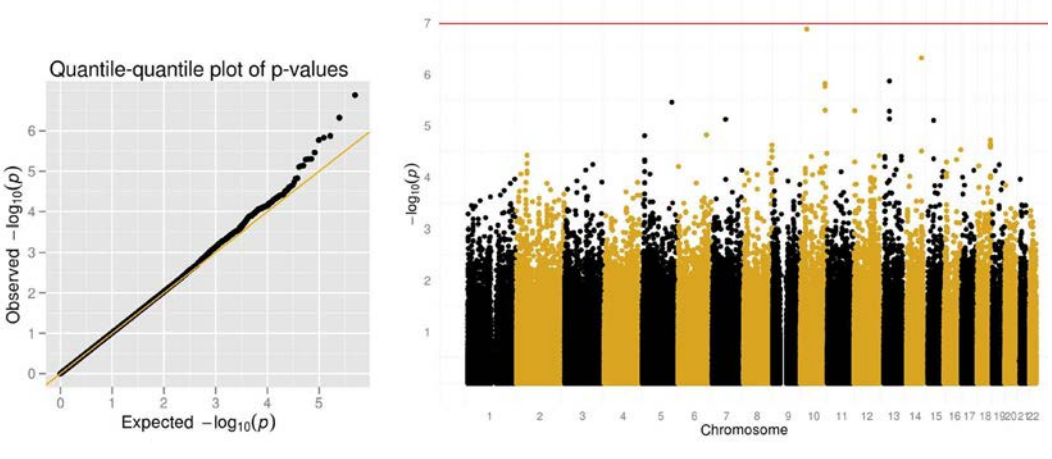


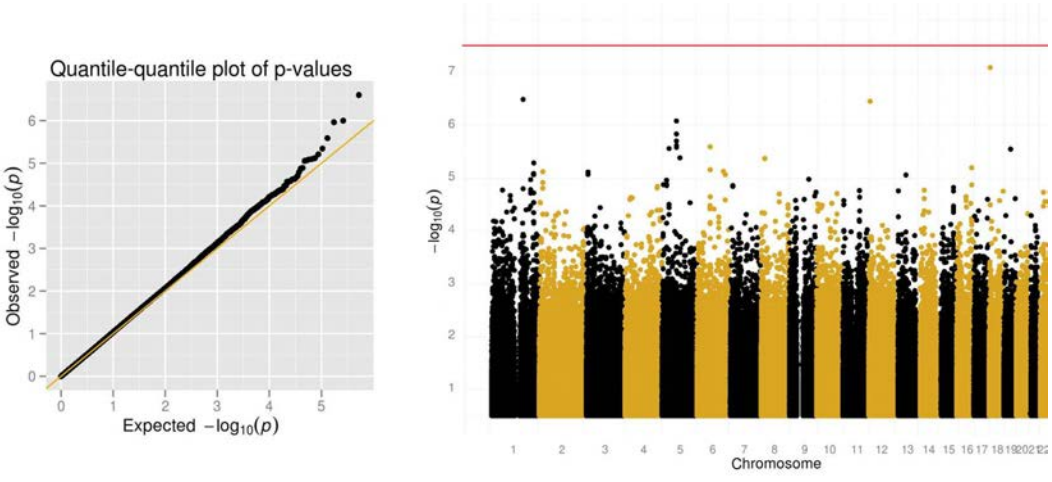
Figure 1. The COPACETIC study set up.



Supplementary Figure 2. Accounting for heterogeneity – a workflow.



Supplementary Figure 3. QQ and Manhattan plots – airway obstruction association results



Supplementary Figure 4. QQ and Manhattan plots – emphysema association results

Supplementary Table 1. Six SNPs with the lowest p value in airway obstruction analysis.

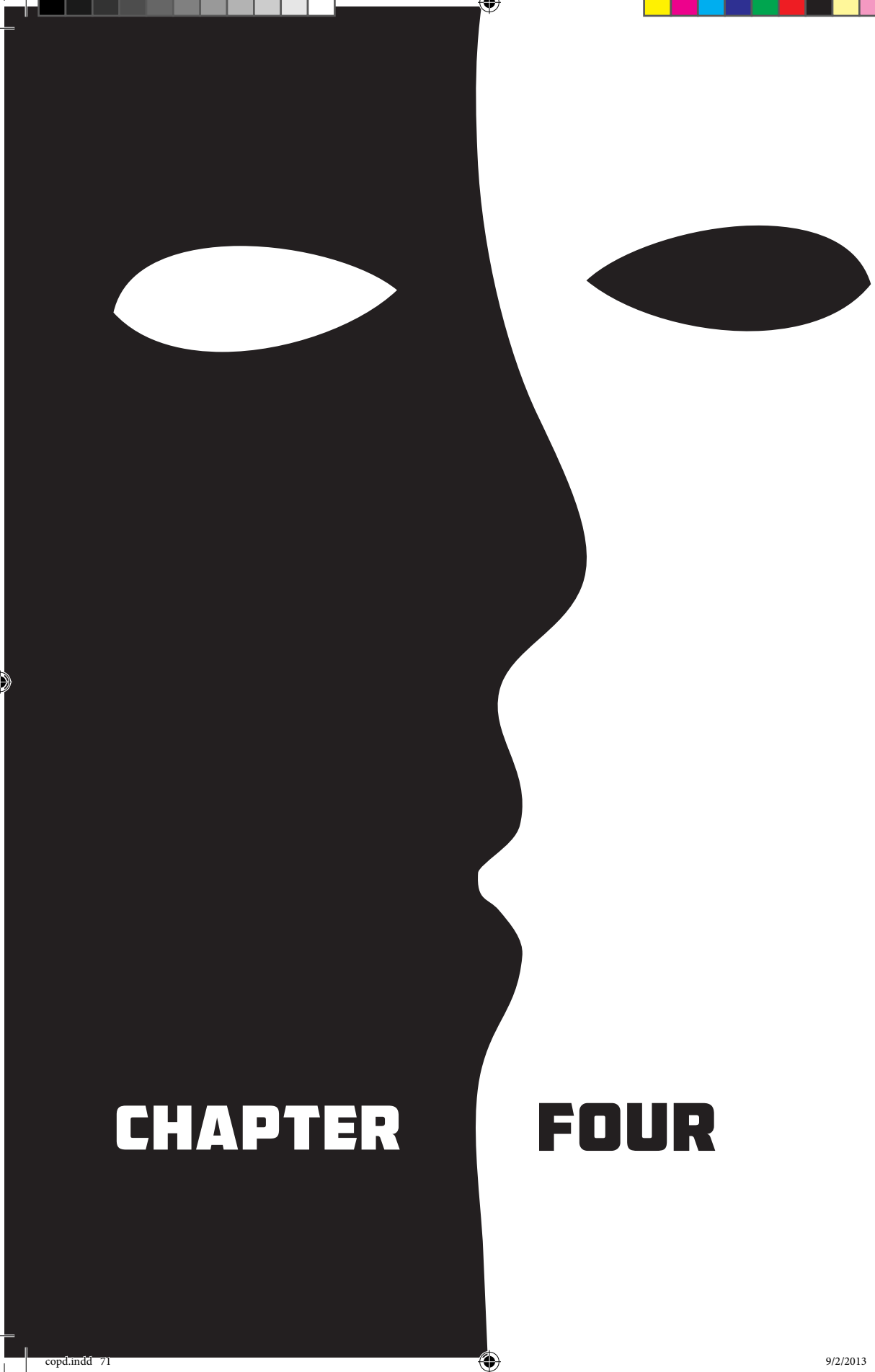
				GWAS		Replication		Combined	
CHR	SNP	Gene	A1	P	OR	P	OR	P	OR
13	rs7337088	NA	G	7.04E-05	1.45	7.01E-04	1.47	1.87E-07	1.46
8	rs189509	PNOC	G	1.13E-04	1.55	7.39E-04	1.52	3.25E-07	1.54
1	rs1317681	TGFB2	A	4.47E-04	0.75	1.13E-03	0.68	2.17E-06	0.73
12	rs7956804	CD4	G	2.25E-04	0.77	1.27E-03	0.86	2.38E-06	0.83
14	rs17091195	SERPINA13	A	3.07E-05	1.37	1.39E-02	1.18	4.04E-06	1.26
18	rs930027	NA	A	3.09E-04	0.73	6.40E-03	0.81	9.40E-06	0.78

Supplementary Table 2. Five SNPs with lowest p values in p15 (emphysema) analysis.

				GWAS		Replication		Combined	
CHR	SNP	Gene	A1	P	BETA	P	BETA	P	BETA
5	rs11948188	GFM2	A	1.26E-05	2.65	4.51E-03	4.19	2.91E-07	2.87
7	rs10486076	NA	A	3.53E-05	-2.39	9.60E-03	-3.53	1.38E-06	-2.57
4	rs4697618	SEL1L3	A	1.04E-05	-2.48	5.50E-02	-1.41	2.96E-06	-2.09
6	rs1224526	NA	A	4.72E-05	-3.12	3.04E-02	-2.09	5.71E-06	-2.72
12	rs3764875	SCNN1A	G	9.66E-06	2.25	0.2824	1.35	6.41E-06	2.12

References

1. van Diemen CC, Postma DS, Vonk JM, Bruinenberg M, Schouten JP, Boezen HM. A disintegrin and metalloprotease 33 polymorphisms and lung function decline in the general population. Am J Respir Crit Care Med 2005;**172**:329-333.
2. Verschuren WM, Blokstra A, Picavet HS, Smit HA. Cohort profile: the Doetinchem Cohort Study. Int J Epidemiol 2008;**37**:1236-1241.
3. Sleegers K, de K, I, Aulchenko YS et al. Cerebrovascular risk factors do not contribute to genetic variance of cognitive function: the ERF study. Neurobiol Aging 2007;**28**:735-741.
4. Lapperre TS, Snoeck-Stroband JB, Gosman MM et al. Dissociation of lung function and airway inflammation in chronic obstructive pulmonary disease. Am J Respir Crit Care Med 2004;**170**:499-504.
5. Nizankowska-Mogilnicka E, Mejza F, Buist AS et al. Prevalence of COPD and tobacco smoking in Malopolska region--results from the BOLD study in Poland. Pol Arch Med Wewn 2007;**117**:402-410.
6. Thomsen M, Nordestgaard BG, Sethi AA, Tybjaerg-Hansen A, Dahl M. beta2-adrenergic receptor polymorphisms, asthma and COPD: two large population-based studies. Eur Respir J 2012;**39**:558-566.
7. Ashraf H, Tonnesen P, Holst PJ, Dirksen A, Thorsen H, Dossing M. Effect of CT screening on smoking habits at 1-year follow-up in the Danish Lung Cancer Screening Trial (DLCST). Thorax 2009;**64**:388-392.
8. Stolk RP, Rosmalen JG, Postma DS et al. Universal risk factors for multifactorial diseases: LifeLines: a three-generation population-based study. Eur J Epidemiol 2008;**23**:67-74.
9. Soler AM, Loth DW, Wain LV et al. Genome-wide association and large-scale follow up identifies 16 new loci influencing lung function. Nat Genet 2011;**43**:1082-1090.



CHAPTER FOUR

Common genes underlying asthma and chronic obstructive pulmonary disease. A genome-wide association study on the Dutch hypothesis

J. Smolonska[1,2], G.H. Koppelman[3], C. Wijmenga[1], J.M. Vonk[2], P. Zanen[4], M. Bruinenberg[1], I. Curjuric[5,6], M. Imboden[5,6], G.A. Thun[5,6], L. Franke[1], N.M. Probst-Hensch[5,6], P. Nürnberg[7], R. Riemersma[8], O. van Schayck[9], H.M. Groen[10], J.W Lammers[4], W. Mali[11] H.M. Boezen[2], D.S. Postma[12] *

[1] University of Groningen, University Medical Center Groningen, Department of Genetics, Groningen, the Netherlands
[2] University of Groningen, University Medical Center Groningen, GRIAC Research Institute, Department of Epidemiology, Groningen, the Netherlands
[3] University of Groningen, University Medical Center Groningen, GRIAC Research Institute, Department of Pediatric Pulmonology and Pediatric Allergology, Beatrix Children's Hospital, Groningen, the Netherlands
[4] University Medical Center Utrecht, Department of Pulmonology, Utrecht, the Netherlands
[5] Department of Epidemiology and Public Health, Swiss Tropical and Public Health Institute SwissTPH, Basel, Switzerland
[6] University of Basel, Basel, Switzerland
[7] Cologne Center for Genomics, University of Cologne, Cologne, Germany
[8] University of Groningen, University Medical Center Groningen, Department of Pulmonology, Groningen, the Netherlands
[9] Department of General Practice, CAPHRI School for Public Health and Primary Care, Maastricht University Medical Centre
[10] University Medical Center Utrecht, Department of Radiology, Utrecht, the Netherlands
[11] University of Groningen, University Medical Center Groningen, GRIAC Research Institute, Department of General Practice, Groningen, the Netherlands
[12] University of Groningen, University Medical Center Groningen, GRIAC Research Institute, Department of Pulmonology, Groningen, the Netherlands
* and the DAG study and the "LifeLines Cohort Study", see acknowledgements

Abstract

BACKGROUND

Asthma and chronic obstructive pulmonary disease (COPD) are thought to share a genetic background (the "Dutch hypothesis").

METHODS

We investigated whether asthma and COPD have common underlying genetic factors, performing genome-wide association studies for both asthma and COPD and by combining results in meta-analyses.

RESULTS

Three loci showed potential involvement in both diseases: chr2p24.3, chr5q23.1and chr13q14.2, containing *DDX1, COMMD10* (both participating in the NF**κβ** pathway) and *GNG5P5,* respectively, as most likely candidate genes to harbor the causal variants. SNP rs9534578 in *GNG5P5* reached genome-wide significance after first stage replication ($p=9.96 \cdot 10^{-9}$). The second stage replication in two independent cohorts provided similar direction of associations, without significant replication. eQTL analysis performed on the top seventeen associated SNPs identified 3 SNPs in *COMMD10* influencing gene expression levels.

CONCLUSIONS

Although inflammatory processes differ in asthma and COPD, they are unequivocally mediated by NF**κβ**, and as suggested by our results, they could be driven by the same underlying genes, *COMMD10* and *DDX1,* implying that the Dutch hypothesis may have some validity. Since none of the SNPs reached genome-wide significance further investigation of the loci should be performed to assess their role in both asthma and COPD.

Introduction

Asthma and Chronic Obstructive Pulmonary Disease (COPD) are two common respiratory diseases, their estimated prevalences ranging from approximately 1% to 18% in different countries[1-3]. Both diseases may lead to airways obstruction, which is reversible in asthma in contrast to COPD. However, the diagnosis cannot rely on reversibility, as it can disappear with asthma progression, making asthma and COPD harder to distinguish. The immune mechanisms underlying the two diseases are thought to be very different, but similarities in inflammatory processes have been reported in both disease entities recently[4]. Classically inflammation in asthma is represented by elevated numbers of CD4+ lymphocytes and eosinophils, and by elevated CD8+ lymphocytes, macrophages and neutrophils in COPD[5]. However, severe asthma can be accompanied by neutrophilia[6] and COPD exacerbation by eosinophilia[7].

Over fifty years ago, the so called 'Dutch hypothesis' was formulated by Orie and colleagues[8] stating that asthma and COPD are two features of one disease entity, referred to as chronic non-specific lung disease (CNSLD). CNSLD was defined to result from the interplay of endogenous factors like genetic predisposition, and exogenous factors like viral infections, air pollution, tobacco smoking and allergen exposures. The timing of this interplay would then determine which clinical syndrome one developed during a lifetime, i.e. asthma, or COPD, or features of both asthma and COPD.

So far this hypothesis has neither been confirmed nor refuted completely[9], but several common environmental exposures have been unequivocally identified as shared risk factors for both asthma and COPD, e.g. maternal smoking during pregnancy, air pollution and active smoking[10]. Genetic factors have been associated with either asthma or COPD using linkage[11-15], candidate gene[16-19] and genome-wide association studies (GWAS)[20, 21]. These studies elucidated genetic factors unique either to asthma or COPD, but additionally potentially shared genetic risk factors including *TGFB1, TNFA, GSTP1, IL13*[22] and *SERPINE2*[23]. *ADAM33* has been linked to the presence of asthma[24], COPD and accelerated lung function decline in the general population and in asthma[25, 25], suggesting common underlying genetic factors of both onset and course of asthma and COPD[26]. So far, hypothesis free GWAS studies aiming to identify novel genes underlying both asthma and COPD in the same source population are lacking. The aim of our study was to identify shared genetic risk factors for asthma and COPD using an unbiased GWAS approach. We first performed a GWAS on asthma and COPD separately using individuals of Dutch descent and subsequently combined these in a meta-analysis, followed by 3 replication studies.

Methods

Study populations

For the identification phase (meta-analysis1), subjects were recruited as participants of the following cohorts:

1. The Dutch Asthma GWAS (DAG) Study: a cohort screened for genetic studies, characterized by the presence of a doctor diagnosis of asthma and hyperresponsiveness.
2. The NELSON cohort study[27]: a population-based cohort screening for lung cancer, including current or ex-smokers with at least 20 pack-years. To increase power of the COPD set, blood bank controls from Amsterdam and Utrecht without clinical data except for age (range 18-65), were added.

For the 1st replication phase (meta-analysis2) subjects were participants of the LifeLines cohort study (LifeLines1).

The 2nd replication phase (meta-analyses3+4) evaluated the top 17 SNPs; subjects were participants of LifeLines (LifeLines2), an independent sample of LifeLines, and the SAPALDIA cohort.

There were no overlapping subjects in all cohorts used. All participants signed informed consent; studies were approved by institutional ethics committees. Detailed information and characteristics of the study populations are shown in Supplementary Appendix (Supplementary table 3 and methods).

Asthma and COPD phenotype definition

In all cohorts, asthma was defined as having a doctor diagnosis of asthma ever, or use of asthma medication (beta-agonists, steroids, anticholinergics, cromoglycate, montelukast, theophyllines) while having 2 or more of the following symptoms: wheeze without a cold, an attack of breathlessness while resting, waking up with an attack of breathlessness, ever. Controls were defined as not having asthma.

In all cohorts, COPD was defined as a pre-bronchodilator $FEV_1/FVC<0.7$ (asthma cases were excluded), and controls (except for blood bank controls) were defined as having an $FEV_1/FVC>0.7$ and $FEV_1>90\%$ predicted.

Genotyping, quality control and imputation

All cohorts were genotyped with Illumina arrays with different SNP content. Genotypes were called and quality control was performed for each dataset to ensure the best quality (Supplementary methods).

Study design and statistical analyses

The analytic work flow is shown in Figure 1. Genome-wide associations on asthma (2,004,043 SNPs) and COPD (1,872,289 SNPs) were performed using χ^2- test on a genetic additive model (0, 1, and 2). The results presented are crude odds ratios. The reasons for not adjusting are several:

1. NELSON study consisted of males only that were heavy-smokers with 20 pack-years at least. This forces already a specific age group, which is in this case age range of 51-79 years of age. In addition, a use of blood bank controls does not allow making any adjustments, except for gender, which was not included as females were just a subset of controls.
2. DAG study and LifeLines asthma cohorts were not adjusted for any covariates to ensure the similarity of the analysis, following the NELSON cohort. LifeLines COPD cohorts included smoking patients, in line with the NELSON cohort characteristics.
3. SAPALDIA cohort is a multi-center study, which forces use of specific covariates, like area. Smoking adjustment was necessary, as analysis in COPD included also non-smokers (in contrast with other cohorts, as shown in Table S3), as they constituted the 50% of an already small group of COPD patients.

Next, results were combined in a directional meta-analysis on 1,811,026 SNPs shared between the asthma and COPD datasets (meta-analysis1). 2,048 SNPs showing $p<0.001$ were selected for *in silico* replication in a second set of asthma and COPD case-control groups derived from the LifeLines cohort (LifeLines1). These markers were analyzed with χ^2- tests and then combined in a second directional meta-analysis (meta-analysis2). The top 17 SNPs with $p\leq0.001$ from meta-analysis2 were investigated in the second stage replication consisting of meta-analysis3 and meta-analysis4 in LifeLines2 and SAPALDIA, respectively.

In SAPALDIA genetic associations with asthma and COPD were tested using logistic regression. Models were controlled for pack-years smoking, study area and principal components capturing inter-European population structure. Results were combined in meta-analysis4.

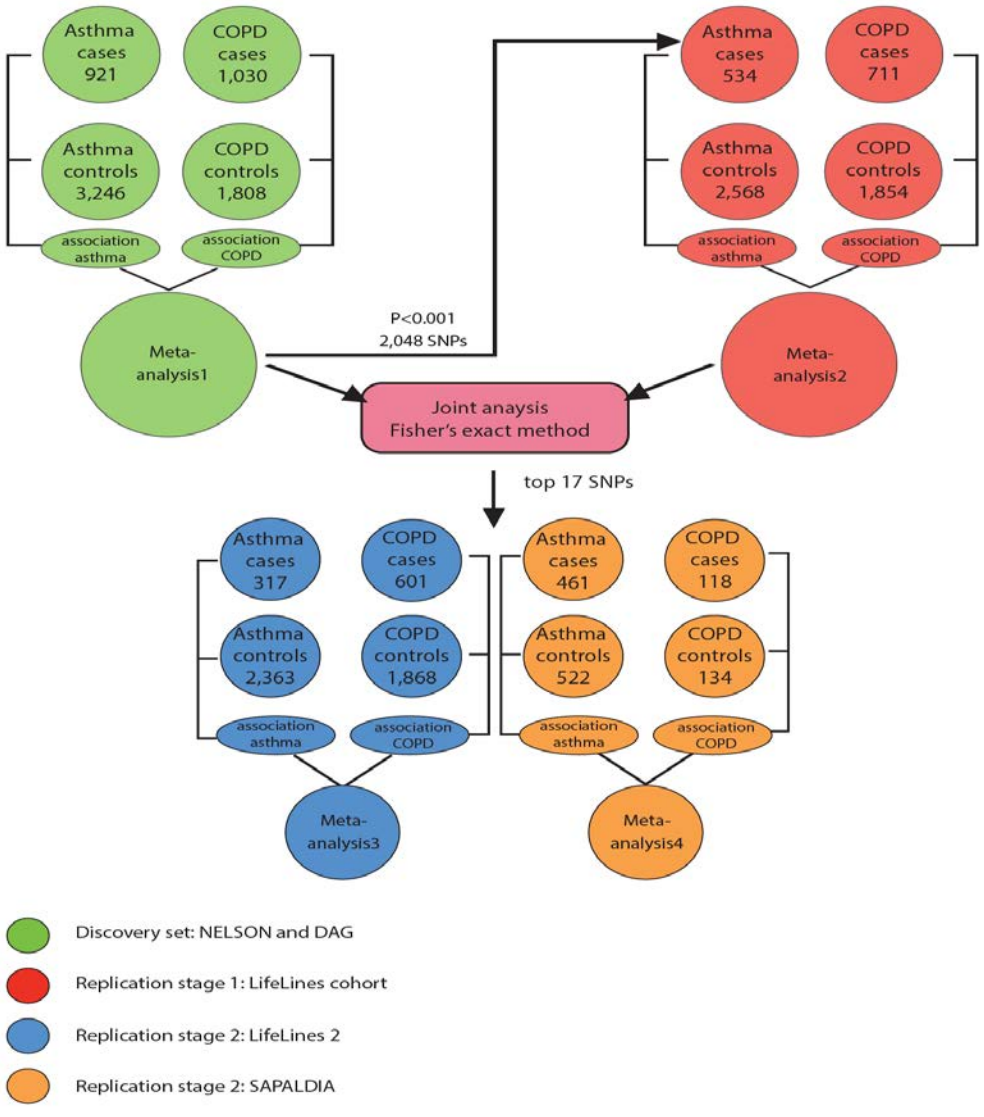


Figure 1. Analytic work flow

Cis-eQTL mapping

Cis-eQTL (expression quantitative trait locus) mapping was performed as described previously by Fehrmann et al.[28] Briefly, each probe on the expression chip was mapped and correlated with SNPs in the vicinity of 250kb. Principal component analysis was applied to the data prior to the analysis to ensure that signals detected as eQTLs are not due to e.g. batch effects. Analysis involved non-parametric Spearman's rank correlation test. Because two different expression chips were used, when probes were present on both, the final result came from meta-analysis. False discovery rate was applied to account for multiple testing.

Network analysis

Gene network was constructed using GeneMANIA. The gene set resulting from this approach was investigated with GATHER to identify enriched pathways.
More details are presented in Supplementary methods.

Results

Genome-wide association and meta-analyses

GWAS were performed on both asthma (921 cases, 3,246 controls) and COPD (1,030 cases, 1,808 controls). Genomic inflation factors (**λ**) equaled 1.01 for both asthma and COPD, indicating no population stratification and presence of true association signals (Figure S1). Individual p-values and odds ratios (ORs) were combined in a directional meta-analysis using a fixed-effects model (meta-analysis1, Figure 1; this data is publicly available at The European Genome-phenome Archive (EGA), accession number EGAS00000000130). All 2,048 single nucleotide polymorphisms (SNPs) with $p \leq 0.001$ were selected for a first replication analysis in asthma and COPD cohorts separately, respectively 534 and 711 cases and 2,568 and 1,854 controls. Subsequently results were combined in a meta-analysis (meta-analysis2, Figure 1).

Seventeen SNPs replicated at $p<0.001$ (Table 1) in the combined meta-analysis1 and meta-analysis2, one SNP reaching genome wide significance. Sixteen of these 17 SNPs map to three genomic locations: 2p24.3, 5q23.1, and 13q14.2 (Table S1).

The chromosome 2p24.3 locus spans ~380 kb and contains genes encoding functional units, like processed transcripts, pseudogenes and RNA genes (Figure 2). The nearest gene with a known function, *DEAD-box polypeptide 1* (*DDX1*), is ~139kb away from the top associated 2p24.3 SNP rs1477253. The locus on chromosome 5 is ~328 kb and contains a single gene: *COMM domain containing 10* (*COMMD10)* (Figure 2). The locus on chromosome 13 spans ~320 kb and only contains a pseudogene: *guanine nucleotide binding protein (G protein), gamma 5 pseudogene 5* (*GNG5P5*) (Figure 2). SNP rs9534578 in *GNG5P5* reached genome-wide significance ($p = 9.96*10^{-9}$).

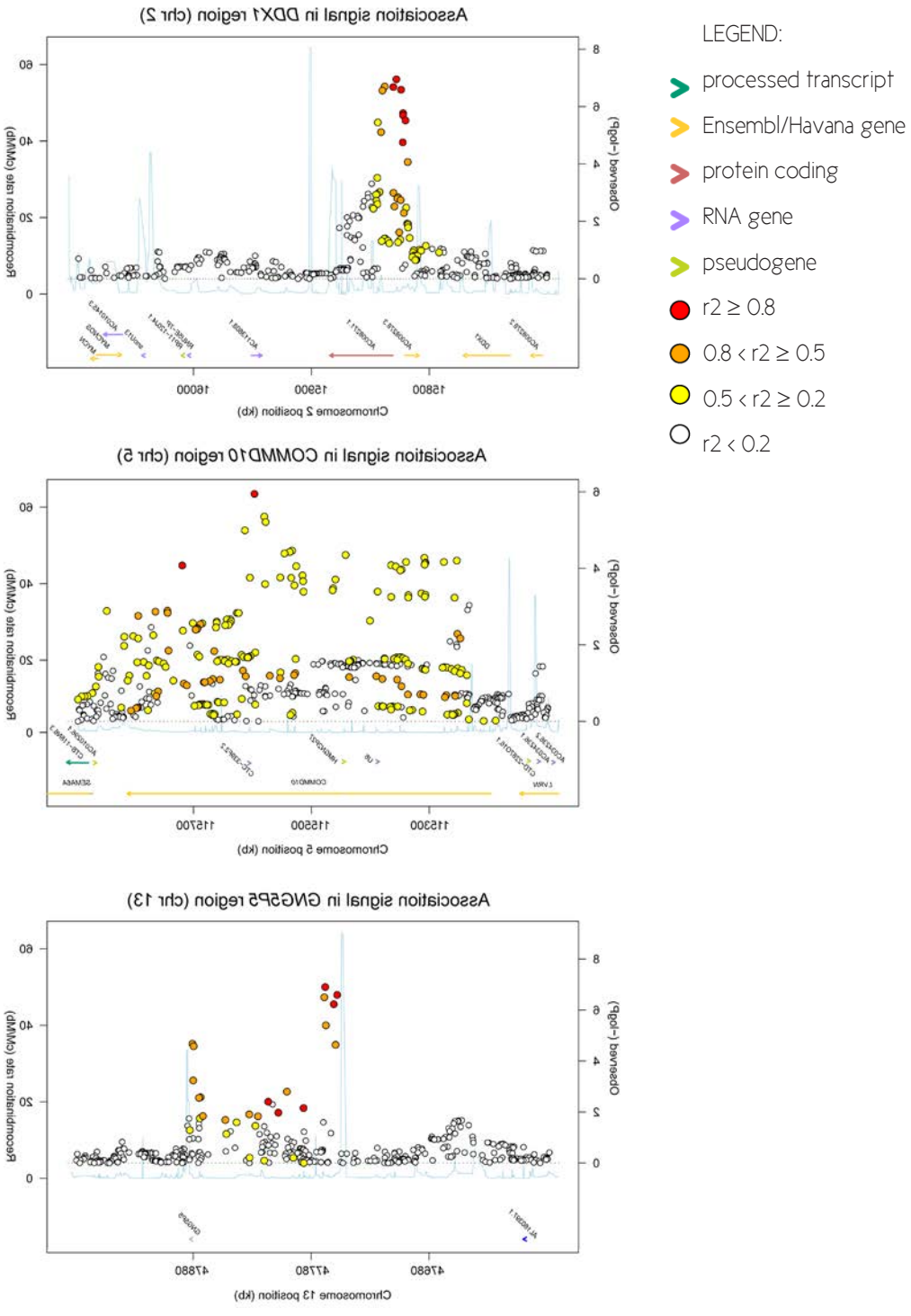


Figure 2. Regional association plots for *DDX1, COMMD10* and *GNG5P5* loci.
The plots were generated using R and regional association plot script from BROAD institute

Table 1. Top 17 SNPs resulting from the identification meta-analysis1 and 1st phase replication meta-analysis2.

CHR	BP	SNP	A1*	A2†	MAF‡	$P_{Nelson+DAG}$	$OR_{Nelson+DAG}$	$P_{LifeLines1}$	$OR_{LifeLines1}$	$P_{Sapaldia}$	$OR_{Sapaldia}$	$P_{LifeLines2}$	$OR_{LifeLines2}$	$P_{Nelson+DAG+Lifelines1}$	$P_{overall}$	Locus
2	15820130	rs2112101	T	C	0.35	1.32E-004	1.16	5.42E-03	1.14	0.54	0.95	0.44	1.04	3.02E-06	3.13E-05	DDX1
2	15822156	rs6728667	G	A	0.36	1.64E-004	1.16	2.67E-03	1.15	0.37	0.92	0.31	1.06	1.68E-06	9.69E-06	DDX1
2	15822185	rs6728750	G	A	0.36	1.01E-004	1.16	4.57E-03	1.14	0.49	0.94	0.36	1.05	1.99E-06	1.88E-05	DDX1
2	15823917	rs2544534	T	C	0.37	3.14E-005	1.18	1.78E-03	1.16	0.47	0.94	0.42	1.04	2.55E-07	3.69E-06	DDX1
2	15827908	rs1477253	T	C	0.36	7.28E-006	1.19	2.52E-03	1.15	0.61	0.96	0.43	1.04	1.11E-07	3.63E-06	DDX1
2	15830470	rs2693008	G	C	0.35	1.78E-005	1.19	2.26E-03	1.15	0.85	0.98	0.40	1.05	2.06E-07	1.36E-05	DDX1
2	15837774	rs2544523	T	C	0.34	2.39E-005	1.18	1.74E-03	1.16	0.67	1.04	0.22	1.07	1.98E-07	7.12E-06	DDX1
2	15839739	rs2693019	T	C	0.34	2.85E-005	1.18	2.04E-03	1.16	0.64	1.04	0.22	1.07	2.75E-07	7.88E-06	DDX1
2	215298827	rs280621	C	T	0.10	6.17E-004	1.25	6.42E-03	1.24	0.71	0.95			1.37E-05	1.64E-04	BARD1
5	115623770	rs10036292	G	A	0.08	4.04E-004	0.78	4.27E-03	0.75	0.41	0.88	0.26	1.12	6.12E-06	3.10E-05	COMMD10
5	115624947	rs10043228	T	C	0.08	3.43E-004	0.78	3.58E-03	0.75	0.40	0.87	0.29	1.11	4.40E-06	2.29E-05	COMMD10
5	115633819	rs254149	G	C	0.17	9.76E-005	0.82	2.81E-03	0.83	0.24	0.88	0.25	1.09	1.13E-06	3.36E-06	COMMD10
5	115642406	rs7727882	G	C	0.08	3.74E-004	0.78	7.20E-03	0.78	0.36	0.86			1.01E-05	3.52E-05	COMMD10
13	46738025	rs17069787	A	G	0.08	1.58E-006	1.58	7.51E-03	1.27	0.23	1.18	0.98	1.00	1.26E-07	5.49E-07	GNG5P5
13	46739001	rs7994542	T	C	0.13	3.29E-005	1.31	2.06E-03	1.22	0.96	0.99	0.96	1.00	3.14E-07	2.61E-05	GNG5P5
13	46741378	rs9534578	A	C	0.08	6.17E-007	1.62	1.81E-03	1.29	0.38	1.13	0.64	0.96	9.96E-09	2.14E-07	GNG5P5
13	46877470	rs7989394	T	A	0.07	6.85E-004	1.34	8.74E-03	1.29	0.15	0.82			2.09E-05	1.65E-05	GNG5P5

* A1 is a minor allele and the risk allele

† A2 is a major allele

‡ MAF is minor allele frequency: calculated in the discovery sample

Replication phase 2 of top 17 SNPs

The top 17 markers from the combined analysis were further evaluated in an independent sample of the LifeLines cohort (LifeLines2; 317 asthma and 2,363 controls; 601 COPD and 1,868 controls, 14 SNPs present) and the SAPALDIA cohort (461 asthma, 118 COPD, 656 controls). None of the SNPs replicated at a nominal p-value <0.05 in LifeLines2 or SAPALDIA although nine and eight SNPs, respectively, showed similar effect estimates. Meta-analysis of all cohorts together did not result in genome-wide significant associations (Table 1).

eQTL analysis of top 17 SNPs

Three of the top 17 SNPs from the combined analysis showed a cis-eQTL effect, when correlating the genotypes with gene expression levels in 1,469 peripheral blood mononuclear cell samples with both GWAS and genome-wide gene expression data available[28]. The three SNPs were located in *COMMD10*. The risk allele (G) of rs254149 decreased *COMMD10* expression levels; risk alleles of rs10043228 (T) and rs10036292 (G), which are in perfect LD (r^2=1), increased *COMMD10* expression levels (Figure 3). Out of these SNPs, rs254149 showed the strongest association and was in rather weak LD with two remaining SNPs (r^2=0.292 for both, Table S1).

Network analysis

The genes found were investigated with GeneMANIA which does not support pseudogenes. Hence we queried only *COMMD10* and *DDX1*. This gene enrichment approach resulted in a set of genes, two genes (*RAD50* and *MRE11A)* being involved in regulation of mitotic recombination (Bayes factor 11, p<0.0001) and telomere maintenance (Bayes factor 6, p<0.0001), possibly implicating COPD as a disease of rapidly aging lungs[29]. Another gene (*BICD1*) involved in telomere maintenance was previously reported in emphysema[30].

Moreover, products of *DDX1* and *COMMD10* interact with NF**κβ**2. COMMD10 has a direct interaction, while DDX1 interacts with RELA and RELB, known to interact directly with NF**κβ**2 and to function in the same pathway (Figure 4).

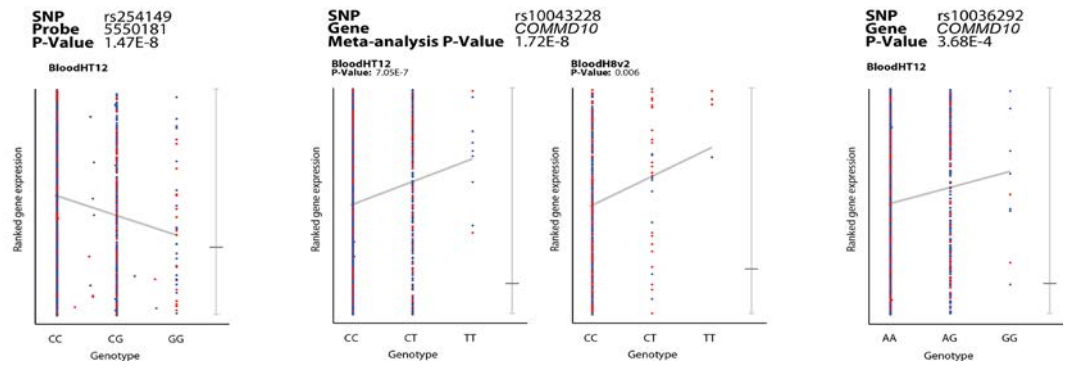


Figure 3. eQTLs identified for *COMMD10* SNPs. Order on x-axis is from non-risk homozygote, heterozygote and risk homozygote for all three eQTLs

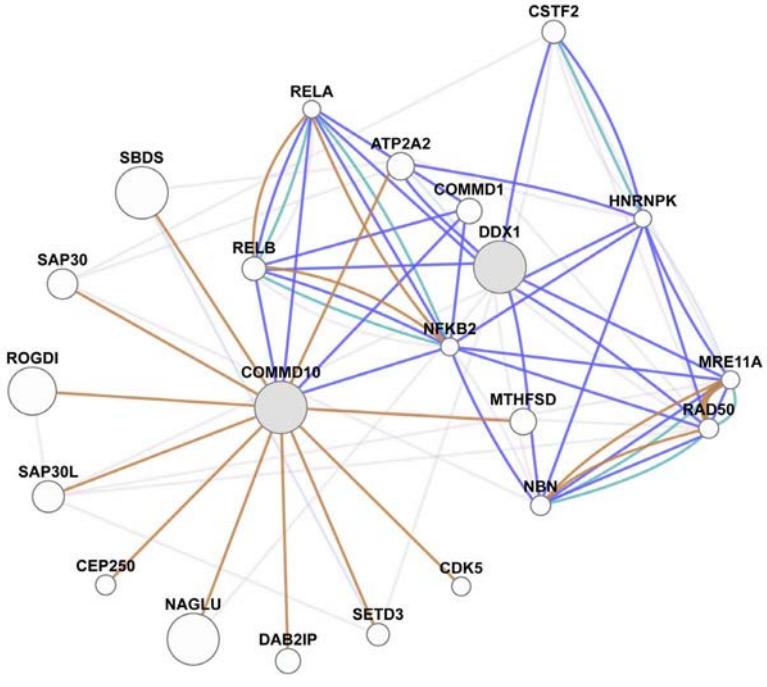


Figure 4. Gene enrichment plot using *DDX1* and *COMMD10* genes as a query

Legend:

- physical interactions
- predicted interactions
- pathway
- co-expression

Discussion

This is the first investigation of shared genetics of asthma and COPD in a hypothesis-free manner using a genome-wide screening in well-defined asthma and COPD cohorts and a large population-based sample. We report three novel loci as potentially shared genetic factors between asthma and COPD, none reaching genome-wide significance in the discovery set or two replication cohorts. None of these three loci were previously reported to be associated with either asthma or COPD. However, *DDX1* locus was reported in a recently published meta-analysis of lung function[31], with a p-value of $9*10^{-6}$. The T allele of rs2544527 in *DDX* was associated with lower lung function and in our study with a higher risk of both asthma and COPD.

The shared 5q23.1 risk locus contains the *COMMD10* gene. COMMD10 is a member of COMM domain containing proteins[32] with a largely unknown function. COMMD10 has shown to form a complex with COMMD1, another member of this family of proteins, which regulates copper metabolism and sodium uptake and inhibits NF**κβ** activation[33]. Copper and sodium levels are inversely regulated, i.e. when copper levels increase, sodium import in cells is inhibited and vice versa. Both ion levels can be regulated by COMMD1, with sodium control mediated through epithelial sodium channels (ENaCs) that are abundantly present in lung epithelial cells[34]. Sodium is crucial for maintaining a fluidic layer in the alveolar part of the lungs and ENaCs play a crucial role in this process[35]. It is tempting to speculate that COMMD10 is involved in this maintenance either through interaction with COMMD1, or independently by displaying similar functions as COMMD1. Also, its function in inhibition of NF**κβ** activation could play a role in regulating inflammatory processes in airways diseases.

The 13q14.2 locus contains the guanine nucleotide binding protein (G protein), gamma 5 pseudogene 5 (*GNG5P5*). Poliseno et al recently showed that pseudogenes can have a pronounced role in regulation of their putative transcripts by competing in non-coding RNA binding[36]. It needs to be tested whether *GNG5P5* can affect GNG5 levels, but it is interesting to note that the pseudogene is processed and has a transcript (ENST00000420444). The biological consequence of a change in GNG5 levels in relation to asthma and COPD pathology is unclear but it is well established that G proteins play a crucial role in signal transduction from cell surface to its interior. It is also known that G-protein coupled receptors (GPCRs) are involved in asthma and more generally are a target of many of the currently used asthma drugs[37].

A third locus on 2p24.3 is bordered by the *DDX1* gene, encoding DEAD-box protein 1, RNA helicase I, and the *MYCN* genes whereas the locus itself contains non-protein coding genes including lincRNAs, ncRNAs, pseudogenes, processed transcripts and one newly discovered, protein-coding gene. Theoretically, any of these could be involved in asthma and

COPD, hindering interpretation of our findings. However, the regional association plot (Figure 2) shows that the signal is mostly confined to *AC008278.3* and *AC008271.1*. Further refinement of the region and functional assessment of the associated variants could help to potentially pin-point the actual causal gene. *DDX1* is a plausible candidate for both asthma and COPD since it interacts with RELA, one of NF**κβ** subunits, upon which it acts as a co-activator of NF**κβ**-mediated transcription[38]. Since this is a central and common pathway of inflammation present in the airways of both asthma and COPD, this may signify a unifying underlying mechanism of both disease entities. Further studies are needed to confirm this hypothesis.

The strengths of our study are clearly the data quality of the cohorts involved, the design of the study and the analysis strategy of the discovery and replication phases. There are some limitations to our study as well. First, we investigated Caucasians of Dutch descent only for discovery and used Caucasians of Swiss and Dutch descent for replication, thus our results are only representative of part of the Caucasian population. However, the markers identified were reasonably common (MAF >5%), with a likely generality to a population of broader ancestry. Secondly, we found no replication in an independent sample of LifeLines (LifeLines2), which might raise the question whether our findings are robust. One explanation for the lack of replication might be the somewhat lower prevalence of asthma in LifeLines2 (7.5% versus 8.5% in LifeLines1) due to the average older age of the subjects included in LifeLines2. This could reflect a cohort effect or some asthma remission at elderly ages[39]. Since we did observe nominal associations with COPD in LifeLines2, but not with asthma (Table S2) the joint analysis did not yield nominal associations. This could be due to use of patients with self-reported doctor diagnosis of asthma ever. It cannot be excluded that this group consists in part of individuals diagnosed with asthma in childhood, who now are in complete remission. Replication of the top 17 SNPs in SAPALDIA was also unsuccessful, possibly due to limited statistical power. Important limitations common to most GWAS remain the heterogeneity of the phenotypes assessed, and heterogeneity between discovery and replication samples. Table S5 shows heterogeneity per meta-analysis performed, i.e. for each asthma-COPD meta-analysis. It differs substantially and due to specificity of the study we could not account for the heterogeneity between meta-analyses. We did not want to perform meta-analysis of meta-analyses, hence we chose Fisher's exact method to combine p values. As highlighted by the Dutch hypothesis the importance of both type and temporal sequences of environmental exposures contribute to the occurrence of either phenotype. This may have affected the phenotypic outcome considerably and hence a crude covariate adjustment may represent an underestimated challenge to identify common genetic determinants of asthma and COPD.

Recent efforts to characterize substantial number of patients diagnosed with both asthma and COPD[40] show the increasing scientific interest in the phenotypic overlap between

asthma and COPD. Future studies on the underlying genetics in this group of overlap patients would be of interest, specifically comparing outcomes with our results.

Overall, our results are suggestive of a role of the NFκβ pathway, a key transcription factor in the inflammatory response, in both asthma and COPD, suggesting that the Dutch hypothesis may have some validity. Since none of the SNPs reached genome-wide significance further investigation of the loci should be performed to assess their role in both asthma and COPD. Although inflammatory processes differ in asthma and COPD, they are unequivocally mediated by NFκβ, and as suggested by our current results, they could be driven by the same underlying genes, *COMMD10* and *DDX1*. Our eQTL study supported the role of *COMMD10*, since we established that three SNPs in *COMMD10* region influence expression of this gene. There were opposite effects of these SNPs, therefore the natural next step is to perform genome-wide epistatic analysis in large cohorts of asthma and COPD patients to reveal the complex nature of interactions between SNPs and loci and their impact on the ultimate phenotype.

Acknowledgments

Authors would like to thank participants of NELSON, DAG, LifeLines and SAPALDIA cohorts.

Nelson We thank H de Koning, M Oudkerk, and W Mali for their efforts in patient and data collection.

DAG Patients participated from the Groningen cohorts (DS Postma, GH Koppelman), SiTA study (R Riemersma, T van der Molen), SGO asthma (EF Knol, C Bruynzeel-Koomen C, R Gerth van Wijk, JGR de Monchy), Prevasc (O van Schayck), ELON (M Kerkhof), Van Lookeren cohort (J Vonk).

LifeLines Cohort Study

Expanded Banner or Group Author:

LifeLines Cohort Study: Behrooz Z Alizadeh (1), Rudolf A de Boer (2), H Marike Boezen (1), Marcel Bruinenberg (3), Lude Franke (4), Pim van der Harst (2), Hans L Hillege (1,2), Melanie M van der Klauw (5), Gerjan Navis (6), Johan Ormel (7), Dirkje S Postma (8), Judith GM Rosmalen (7), Joris P Slaets (9), Harold Snieder (1), Ronald P Stolk (1), Bruce HR Wolffenbuttel (5), Cisca Wijmenga (4)

1. Department of Epidemiology, University of Groningen, University Medical Center Groningen, The Netherlands
2. Department of Cardiology, University of Groningen, University Medical Center Groningen, The Netherlands
3. LifeLines Cohort Study, University of Groningen, University Medical Center Groningen, The Netherlands
4. Department of Genetics, University of Groningen, University Medical Center Groningen, The Netherlands
5. Department of Endocrinology, University of Groningen, University Medical Center Groningen, The Netherlands
6. Department of Internal Medicine, Division of Nephrology, University of Groningen, University Medical Center Groningen, The Netherlands
7. Interdisciplinary Center of Psychopathology of Emotion Regulation (ICPE), Department of Psychiatry, University of Groningen, University Medical Center Groningen, The Netherlands
8. Department of Pulmonology, University of Groningen, University Medical Center Groningen, GRIAC research institute, The Netherlands
9. University Center for Geriatric Medicine, University of Groningen, University Medical Center Groningen, The Netherlands

The LifeLines Cohort Study, and generation and management of GWAS genotype data for the LifeLines Cohort Study is supported by the Netherlands Organization of Scientific Research NWO (grant 175.010.2007.006), the Economic Structure Enhancing Fund (FES) of the Dutch government, the Ministry of Economic Affairs, the Ministry of Education, Culture and Science, the Ministry for Health, Welfare and Sports, the Northern Netherlands Collaboration of Provinces (SNN), the Province of Groningen, University Medical Center Groningen, the University of Groningen, Dutch Kidney Foundation and Dutch Diabetes Research Foundation.

We thank Behrooz Alizadeh, Annemieke Boesjes, Marcel Bruinenberg, Noortje Festen, Pim van der Harst, Ilja Nolte, Lude Franke, Mitra Valimohammadi for their help in creating the GWAS database, and Rob Bieringa, Joost Keers, René Oostergo, Rosalie Visser, Judith Vonk for their work related to data-collection and validation. The authors are grateful to the study participants, the staff from the LifeLines Cohort Study and Medical Biobank Northern Netherlands, and the participating general practitioners and pharmacists.

SAPALDIA: The study could not have been done without the help of the study participants, technical and administrative support and the medical teams and field workers at the local study sites. Local fieldworkers : Aarau: M Broglie, M Bünter, D Gashi, Basel: R Armbruster, T Damm, U Egermann, M Gut, L Maier, A Vögelin, L Walter, Davos: D Jud, N Lutz, Geneva: M Ares, M Bennour, B Galobardes, E Namer, Lugano: B Baumberger, S Boccia Soldati, E Gehrig-Van Essen, S Ronchetto, Montana: C Bonvin, C Burrus, Payerne: S Blanc, AV Ebinger, ML Fragnière, J Jordan, Wald: R Gimmi, N Kourkoulos, U Schafroth. Administrative staff: N Bauer, D Baehler, C Gabriel, R Gutknecht. SAPALDIA Team: *Study directorate:* T Rochat, NM Probst Hensch, N Künzli, C Schindler, JM Gaspoz. *Scientific team:* JC Barthélémy, W Berger, R Bettschart, A Bircher, G Bolognini, O Brändli, C Brombach, M Brutsche, L Burdet, M Frey, U Frey, MW Gerbase, D Gold, E de Groot, W Karrer, R Keller, B Knöpfli, B Martin, D Miedinger, U Neu, L Nicod, M Pons, F Roche, T Rothe, E Russi, P Schmid-Grendelmeyer, A Schmidt-Trucksäss, A Turk, J Schwartz, D. Stolz, P Straehl, JM Tschopp, A von Eckardstein, E Zemp Stutz. *Scientific team at coordinating centers:* M Adam, E Boes, PO Bridevaux, D Carballo, E Corradi, I Curjuric, J Dratva, A Di Pasquale, L Grize, D Keidel, S Kriemler, A Kumar, M Imboden, N Maire, A Mehta, F Meier, H Phuleria, E Schaffner, GA Thun, A Ineichen, M Ragettli, M Ritter, T Schikowski, G Stern, M Tarantino, M Tsai, M Wanner.

The work described in this manuscript was supported by EU project, COPACETIC: COPD, Addressing Critical Gaps, Early Treatment and Diagnosis and Innovative Concepts, Grant Agreement No. 201379

References

1. Masoli M, Fabian D, Holt S, Beasley R. The global burden of asthma: executive summary of the GINA Dissemination Committee report. Allergy 2004;59:469–478.
2. Beasley R. The Global Burden of Asthma Report, Global Initiative for Asthma (GINA). 2004
3. Halbert RJ, Isonaka S, George D, Iqbal A. Interpreting COPD prevalence estimates: what is the true burden of disease? Chest 2003;123:1684–1692.
4. Kraft M. Asthma and chronic obstructive pulmonary disease exhibit common origins in any country! Am J Respir Crit Care Med 2006;174:238–240.
5. Plusa T. {Overlap syndrome--asthma and chronic obstructive pulmonary disease}. Pneumonol Alergol Pol 2011;79:351–356.
6. Monteseirin J. Neutrophils and asthma. J Investig Allergol Clin Immunol 2009;19:340–354.
7. Papi A, Bellettato CM, Braccioni F et al. Infections and airway inflammation in chronic obstructive pulmonary disease severe exacerbations. Am J Respir Crit Care Med 2006;173:1114–1121.
8. The host factor in bronchitis. Groninen, the Netherlands: 1961.
9. Barnes PJ. Against the Dutch hypothesis: asthma and chronic obstructive pulmonary disease are distinct diseases. Am J Respir Crit Care Med 2006;174:240–243.
10. Postma DS, Kerkhof M, Boezen HM, Koppelman GH. Asthma and chronic obstructive pulmonary disease: common genes, common environments? Am J Respir Crit Care Med 2011;183:1588–1594.
11. Postma DS, Bleecker ER, Amelung PJ et al. Genetic susceptibility to asthma--bronchial hyperresponsiveness coinherited with a major gene for atopy. N Engl J Med 1995;333:894–900.
12. A genome-wide search for asthma susceptibility loci in ethnically diverse populations. The Collaborative Study on the Genetics of Asthma (CSGA). Nat Genet 1997;15:389–392.
13. Xu J, Meyers DA, Ober C et al. Genomewide screen and identification of gene-gene interactions for asthma-susceptibility loci in three U.S. populations: collaborative study on the genetics of asthma. Am J Hum Genet 2001;68:1437–1446.
14. Silverman EK, Palmer LJ, Mosley JD et al. Genomewide linkage analysis of quantitative spirometric phenotypes in severe early-onset chronic obstructive pulmonary disease. Am J Hum Genet 2002;70:1229–1239.
15. Palmer LJ, Celedon JC, Chapman HA, Speizer FE, Weiss ST, Silverman EK. Genome-wide linkage analysis of bronchodilator responsiveness and post-bronchodilator spirometric phenotypes in chronic obstructive pulmonary disease. Hum Mol Genet 2003;12:1199–1210.
16. Nicolaides NC, Holroyd KJ, Ewart SL et al. Interleukin 9: a candidate gene for asthma. Proc Natl Acad Sci U S A 1997;94:13175–13180.
17. Duetsch G, Illig T, Loesgen S et al. STAT6 as an asthma candidate gene: polymorphism-screening, association and haplotype analysis in a Caucasian sib-pair study. Hum Mol Genet 2002;11:613–621.
18. Nicolae D, Cox NJ, Lester LA et al. Fine mapping and positional candidate studies identify HLA-G as an asthma susceptibility gene on chromosome 6p21. Am J Hum Genet 2005;76:349–357.
19. Castaldi PJ, Cho MH, Cohn M et al. The COPD genetic association compendium: a comprehensive online database of COPD genetic associations. Hum Mol Genet 2010;19:526–534.
20. Zhang Y, Moffatt MF, Cookson WO. Genetic and genomic approaches to asthma: new insights for the origins. Curr Opin Pulm Med 2012;18:6–13.
21. Wain LV, Artigas MS, Tobin MD. What can genetics tell us about the cause of fixed airflow obstruction? Clin Exp Allergy 2012;42:1176–1182.
22. Smolonska J, Wijmenga C, Postma DS, Boezen HM. Meta-analyses on suspected chronic obstructive

pulmonary disease genes: a summary of 20 years' research. Am J Respir Crit Care Med 2009;180:618–631.

23. Himes BE, Klanderman B, Ziniti J et al. Association of SERPINE2 with asthma. Chest 2011;140:667–674.

24. Van EP, Little RD, Dupuis J et al. Association of the ADAM33 gene with asthma and bronchial hyperresponsiveness. Nature 2002;418:426–430.

25. van Diemen CC, Postma DS, Vonk JM, Bruinenberg M, Schouten JP, Boezen HM. A disintegrin and metalloprotease 33 polymorphisms and lung function decline in the general population. Am J Respir Crit Care Med 2005;172:329–333.

26. Jongepier H, Boezen HM, Dijkstra A et al. Polymorphisms of the ADAM33 gene are associated with accelerated lung function decline in asthma. Clin Exp Allergy 2004;34:757–760.

27. van Iersel CA, de Koning HJ, Draisma G et al. Risk-based selection from the general population in a screening trial: selection criteria, recruitment and power for the Dutch-Belgian randomised lung cancer multi-slice CT screening trial (NELSON). Int J Cancer 2007;120:868–874.

28. Fehrmann RS, Jansen RC, Veldink JH et al. Trans-eQTLs reveal that independent genetic variants associated with a complex phenotype converge on intermediate genes, with a major role for the HLA. PLoS Genet 2011;7:e1002197.

29. Lee J, Sandford A, Man P, Sin DD. Is the aging process accelerated in chronic obstructive pulmonary disease? Curr Opin Pulm Med 2011;17:90–97.

30. Kong X, Cho MH, Anderson W et al. Genome-wide association study identifies BICD1 as a susceptibility gene for emphysema. Am J Respir Crit Care Med 2011;183:43–49.

31. Soler AM, Loth DW, Wain LV et al. Genome-wide association and large-scale follow up identifies 16 new loci influencing lung function. Nat Genet 2011;43:1082–1090.

32. Burstein E, Hoberg JE, Wilkinson AS et al. COMMD proteins, a novel family of structural and functional homologs of MURR1. J Biol Chem 2005;280:22222–22232.

33. de BP, van de Sluis B, Klomp L, Wijmenga C. The many faces of the copper metabolism protein MURR1/COMMD1. J Hered 2005;96:803–811.

34. Handy RD, Eddy FB, Baines H. Sodium-dependent copper uptake across epithelia: a review of rationale with experimental evidence from gill and intestine. Biochim Biophys Acta 2002;1566:104–115.

35. Eaton DC, Helms MN, Koval M, Bao HF, Jain L. The contribution of epithelial sodium channels to alveolar function in health and disease. Annu Rev Physiol 2009;71:403–423.

36. Poliseno L, Salmena L, Zhang J, Carver B, Haveman WJ, Pandolfi PP. A coding-independent function of gene and pseudogene mRNAs regulates tumour biology. Nature 2010;465:1033–1038.

37. Postma DS, Koppelman GH. Confirmation of GPRA: a putative drug target for asthma. Am J Respir Crit Care Med 2005;171:1323–1324.

38. Ishaq M, Ma L, Wu X et al. The DEAD-box RNA helicase DDX1 interacts with RelA and enhances nuclear factor kappaB-mediated transcription. J Cell Biochem 2009;106:296–305.

39. Vonk JM, Postma DS, Boezen HM et al. Childhood factors associated with asthma remission after 30 year follow up. Thorax 2004;59:925–929.

40. Hardin M, Silverman EK, Barr RG et al. The clinical features of the overlap between COPD and asthma. Respir Res 2011;12:127.

Supplementary Appendix

Supplementary methods

Study populations

For the identification phase (meta-analysis1), subjects were recruited as participants of the following cohorts:

1. The Dutch Asthma GWAS (DAG) Study: a cohort screened for genetic studies, characterized by the presence of a doctor diagnosis of asthma and hyperresponsiveness and extensive phenotyping.
2. The NELSON cohort study[1]: a population-based cohort screening for lung cancer, including current or ex-smokers with at least 20 pack-years. Blood bank controls, collected in Amsterdam and Utrecht, were added to increase power of the COPD set, and except for the age (range 18-65) no clinical data were available.

For the 1st replication phase (meta-analysis2, 2,048 SNPs) subjects were recruited as participants of the LifeLines cohort study.

LifeLines[2] is a biobanking initiative collecting random individuals from the general population and their relatives in order to perform intense phenotyping which will be used to better define disease type and/or subtype and relate it to genetic and environmental factors. Amongst others spirometry is performed (following ATS guidelines) and data on respiratory symptoms are collected (using standardized questionnaires) enabling a diagnosis of asthma and/or COPD. For purpose of current study all relatives were excluded.

The 2nd replication phase (meta-analyses3+4) evaluated the top 17 SNPs; subjects were recruited as participants of the LifeLines cohort study (LifeLines2), an independent sample of Lifelines and the SAPALDIA cohort.

The SAPALDIA cohort: a multi-center study in eight geographic areas representing the range of environmental, meteorological and socio-demographic conditions in Switzerland. It was initiated in 1991 (SAPALDIA1) with a follow-up assessment in 2002 (SAPALDIA2) and 2011 (SAPALDIA3). This study has specifically been designed to investigate longitudinally lung function, respiratory and cardiovascular health, and to study and identify the associations of these health indicators with individual long term exposure to air pollution, other toxic inhalants, life style and molecular factors. SAPALDIA2 was used in current study.

Genotyping, quality control and imputation

The Dutch Asthma GWAS cohort was genotyped on Illumina Hap300 and Hap370 platforms. The NELSON COPD cohort was genotyped using Illumina 610 Quad BeadChip containing 620,901 probes. The blood bank controls for the NELSON COPD cohort were

genotyped with Illumina 670 Quad BeadChip containing custom and tagging SNPs selected by WTCCC2.

The LifeLines cohort was genotyped on CytoChip containing a selection of 299,140 SNPs to tag the whole genome and capture regions most frequently used in cytogenetic screening. Genotypes were called with an algorithm provided by Illumina and implemented in Genome Studio. Quality control was performed for each dataset to ensure the best quality. SNPs were removed if call rates were <95%, the minor allele frequency was below 5% and/or Hardy-Weinberg equilibrium was not met (HWE; $p<0.0001$). Samples were removed when more than 5% of genotypes were missing, samples were duplicated (PI_HAT value - 1) or from individuals related to another individual in the dataset (PI_HAT>0.5) or ethnic outliers (based on components from multidimensional scaling C1 and C2). To enable uniform analyses across all datasets imputation was performed using BEAGLE 3.0[3] and HapMap CEU as reference panel (HapMap 2, release 24, genome build 36). Genotype dosages were converted to regular genotypes with a best-guess method. SNPs imputed with less confidence were removed (correlation $r2<0.5$).

In the framework of the European GABRIEL study on asthma genetics, genotyping in the SAPALDIA cohort was done on the Illumina Human610 Quad platform at the Centre National de Génotypage in Evry, France on 663 asthmatics and a random sample of 997 non-asthmatic participants. 567589 successfully genotyped autosomal SNPs were imputed to 2.5 Mio by MACH v 1.0 software[4] using the HapMap v22 CEPH panel of Utah residents with ancestry from northern and Western Europe as reference. Samples with <97% genotyping success rate, non-European origin, cryptic relatedness or sex inconsistencies were excluded from analysis. After quality control procedures, genetic and covariate data from 461 cases of doctor diagnosed asthma, 118 cases of COPD and 656 controls was available.

Statistical analysis

Case-control analysis was performed using $\mathbf{X}^2$ test for both asthma and COPD. Association results were combined in a directional meta-analysis. Same approach was used for later replication stages. All above mentioned analyses were performed using PLINK[5] in case of NELSON, DAG and LifeLines cohorts. SAPALDIA performed the analysis using ProbABEL software version 0.1.3[6].

Directions of the signals were compared between the meta-analyses and if they were the same results were combined with Fishers' method using R[7]. We used Fishers' method to combine p values resulting from meta-analyses, as meta-analysis of meta-analysis is statically inappropriate.

Network analysis

To perform gene enrichment base on networks we used publically available tool: GeneMANIA[8] (http://www.genemania.org/). Because GeneMANIA does not support pseudogenes, we queried two genes resulting from our analysis: *DDX1* and *COMMD10*. The settings were the following: we chose to show up to 20 genes in the network, we used automatic weighting for the network. For the network creation we used only physical interactions, predicted interaction, pathways and coexpression. GeneMANIA was accessed on 28th September 2011.

In order to identify the overrepresented pathways in the above mentioned network we used GATHER[9]. GATHER is web-based tool (http://gather.genome.duke.edu/), which performs annotations, among others, to pathways, based on the query genes. The resulting annotations are given Bayes factor and p value to indicate the strength of the annotation with the gene list provided. We queried 22 genes in total: *DDX1, COMMD10* and 20 genes resulting from gene enrichment in GeneMANIA: *SBDS, NAGLU, ROGDI, SAP30L, SAP30, ATP2A2, MTHFSD, COMMD1, DAB2IP, RELB, CSTF2, SETD3, CEP250, NBN, RAD50, CDK5, MRE11A, NFKB2, RELA* and *HNRNPK*. GATHER was last accessed on August 3rd 2012.

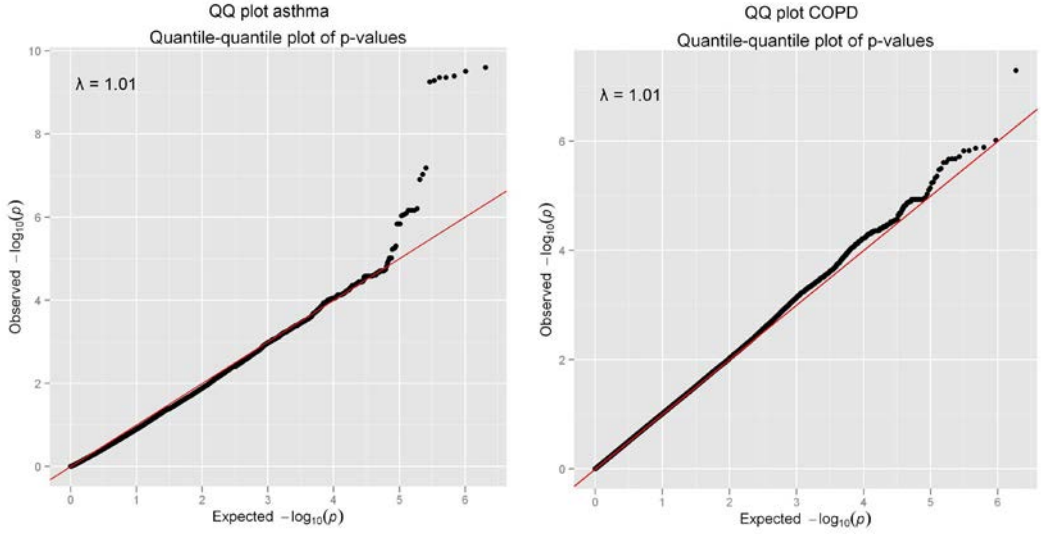


Figure S1. Quantile-quantile plots for asthma and COPD GWAS

Table S1. Linkage disequilibrium between top 17 SNPs in asthma–COPD meta-analysis with highest significance level

SNP	Proxy	Distance	RSquared	DPrime	Chromosome	Coordinate_HG18
rs2112101	rs6728750	2055	0.867	1	chr2	15822185
rs2112101	rs2544534	3787	0.863	0.963	chr2	15823917
rs2112101	rs1477253	7778	0.863	0.963	chr2	15827908
rs2112101	rs6728667	2026	0.838	1	chr2	15822156
rs2112101	rs2693008	10340	0.828	0.927	chr2	15830470
rs2112101	rs2544523	17644	0.609	0.84	chr2	15837774
rs2112101	rs2693019	19609	0.609	0.84	chr2	15839739
rs6728667	rs6728750	29	0.966	1	chr2	15822185
rs6728667	rs2544534	1761	0.9	1	chr2	15823917
rs6728667	rs1477253	5752	0.9	1	chr2	15827908
rs6728667	rs2112101	2026	0.838	1	chr2	15820130
rs6728667	rs2693008	8314	0.804	0.962	chr2	15830470
rs6728667	rs2544523	15618	0.604	0.915	chr2	15837774
rs6728667	rs2693019	17583	0.604	0.915	chr2	15839739
rs6728750	rs6728667	29	0.966	1	chr2	15822156
rs6728750	rs2544534	1732	0.932	1	chr2	15823917
rs6728750	rs1477253	5723	0.932	1	chr2	15827908
rs6728750	rs2112101	2055	0.867	1	chr2	15820130
rs6728750	rs2693008	8285	0.833	0.963	chr2	15830470
rs6728750	rs2544523	15589	0.627	0.916	chr2	15837774
rs6728750	rs2693019	17554	0.627	0.916	chr2	15839739
rs2544534	rs1477253	3991	1	1	chr2	15827908
rs2544534	rs6728750	1732	0.932	1	chr2	15822185
rs2544534	rs6728667	1761	0.9	1	chr2	15822156
rs2544534	rs2112101	3787	0.863	0.963	chr2	15820130
rs2544534	rs2693008	6553	0.83	0.927	chr2	15830470
rs2544534	rs2544523	13857	0.676	0.918	chr2	15837774
rs2544534	rs2693019	15822	0.676	0.918	chr2	15839739
rs1477253	rs2544534	3991	1	1	chr2	15823917
rs1477253	rs6728750	5723	0.932	1	chr2	15822185
rs1477253	rs6728667	5752	0.9	1	chr2	15822156
rs1477253	rs2112101	7778	0.863	0.963	chr2	15820130
rs1477253	rs2693008	2562	0.83	0.927	chr2	15830470
rs1477253	rs2544523	9866	0.676	0.918	chr2	15837774
rs1477253	rs2693019	11831	0.676	0.918	chr2	15839739
rs2693008	rs6728750	8285	0.833	0.963	chr2	15822185
rs2693008	rs1477253	2562	0.83	0.927	chr2	15827908
rs2693008	rs2544534	6553	0.83	0.927	chr2	15823917
rs2693008	rs2112101	10340	0.828	0.927	chr2	15820130
rs2693008	rs6728667	8314	0.804	0.962	chr2	15822156
rs2693008	rs2544523	7304	0.766	0.96	chr2	15837774
rs2693008	rs2693019	9269	0.766	0.96	chr2	15839739
rs2544523	rs2693019	1965	1	1	chr2	15839739
rs2544523	rs2693008	7304	0.766	0.96	chr2	15830470
rs2544523	rs1477253	9866	0.676	0.918	chr2	15827908
rs2544523	rs2544534	13857	0.676	0.918	chr2	15823917
rs2544523	rs6728750	15589	0.627	0.916	chr2	15822185
rs2544523	rs2112101	17644	0.609	0.84	chr2	15820130
rs2544523	rs6728667	15618	0.604	0.915	chr2	15822156

rs2693019	rs2544523	1965	1	1	chr2	15837774
rs2693019	rs2693008	9269	0.766	0.96	chr2	15830470
rs2693019	rs1477253	11831	0.676	0.918	chr2	15827908
rs2693019	rs2544534	15822	0.676	0.918	chr2	15823917
rs2693019	rs6728750	17554	0.627	0.916	chr2	15822185
rs2693019	rs2112101	19609	0.609	0.84	chr2	15820130
rs2693019	rs6728667	17583	0.604	0.915	chr2	15822156
rs10036292	rs10043228	1177	1	1	chr5	115624947
rs10036292	rs7727882	18636	1	1	chr5	115642406
rs10036292	rs254149	10049	0.292	1	chr5	115633819
rs10043228	rs10036292	1177	1	1	chr5	115623770
rs10043228	rs7727882	17459	1	1	chr5	115642406
rs10043228	rs254149	8872	0.292	1	chr5	115633819
rs254149	rs7727882	8587	0.292	1	chr5	115642406
rs254149	rs10043228	8872	0.292	1	chr5	115624947
rs254149	rs10036292	10049	0.292	1	chr5	115623770
rs7727882	rs10043228	17459	1	1	chr5	115624947
rs7727882	rs10036292	18636	1	1	chr5	115623770
rs7727882	rs254149	8587	0.292	1	chr5	115633819
rs17069787	rs9534578	3353	0.901	1	chr13	46741378
rs17069787	rs7989394	139445	0.49	0.7	chr13	46877470
rs17069787	rs7994542	976	0.489	0.894	chr13	46739001
rs7994542	rs7989394	138469	0.611	1	chr13	46877470
rs7994542	rs9534578	2377	0.551	1	chr13	46741378
rs7994542	rs17069787	976	0.489	0.894	chr13	46738025
rs9534578	rs17069787	3353	0.901	1	chr13	46738025
rs9534578	rs7994542	2377	0.551	1	chr13	46739001
rs9534578	rs7989394	136092	0.548	0.78	chr13	46877470
rs7989394	rs7994542	138469	0.611	1	chr13	46739001
rs7989394	rs9534578	136092	0.548	0.78	chr13	46741378
rs7989394	rs17069787	139445	0.49	0.7	chr13	46738025

Table S2. Top 17 SNPs replication in LifeLines2 sample*

LifeLines 2 association signals					asthma		COPD	
CHR	SNP	BP	A1†	A2‡	P	OR	P	OR
2	rs2112101	15820130	T	C	3.46E-01	0.9197	9.07E-02	1.123
2	rs6728667	15822156	G	A	3.85E-01	0.9261	5.01E-02	1.143
2	rs6728750	15822185	G	A	3.24E-01	0.9163	5.64E-02	1.139
2	rs2544534	15823917	T	C	2.31E-01	0.9013	5.28E-02	1.139
2	rs1477253	15827908	T	C	2.86E-01	0.9113	6.78E-02	1.131
2	rs2693008	15830470	G	C	2.08E-01	0.8948	4.21E-02	1.148
2	rs2544523	15837774	T	C	4.14E-01	0.9301	2.79E-02	1.161
2	rs2693019	15839739	T	C	3.82E-01	0.9253	2.56E-02	1.164
5	rs10036292	115623770	G	A	4.23E-01	0.8735	5.07E-02	1.266
5	rs10043228	115624947	T	C	4.13E-01	0.8709	5.91E-02	1.256
5	rs254149	115633819	G	C	6.48E-01	0.9482	7.11E-02	1.172
13	rs17069787	46738025	A	G	9.37E-01	1.013	9.24E-01	0.9869
13	rs7994542	46739001	T	C	5.92E-01	0.9349	6.26E-01	1.049
13	rs9534578	46741378	A	C	6.60E-01	0.9336	7.96E-01	0.9688

* only 14 SNPs were genotyped/imputed in LifeLines2
† A1 is a minor allele and the risk allele
‡ A2 is a major allele

Table S3. Characteristics of participants

	DAG	NELSON	LifeLines1		LifeLines2		SAPALDIA2	
Phenotype	asthma	COPD	asthma	COPD	asthma	COPD	asthma	COPD
N	920	1030	534	711	317	601	461	118
Age, yrs, mean (SD)	34 (16)	63.3 (5.6)	44.8 (9.7)	54 (10.6)	46.7 (11.2)	56.7 (10.8)	49.0 (11.8)	58.3 (10.0)
Height, meters, mean (SD)	1.69 (0.17)	1.78 (0.06)	1.74 (0.1)	1.76 (0.09)	1.73 (0.09)	1.74 (0.09)	1.68 (0.09)	1.69 (0.09)
Gender male n (%)	430 (47)	1030 (100)	214 (40)	369 (52)	120 (37.9)	282 (46.9)	212 (46.0)	67 (56.8)
Smoking habits								
Current smoker, n (%)	147 (16.0)	410 (39.8)	106 (19.9)	363 (51.1)	41(12.9)	231(38.4)	95 (20.6)	44 (37.3)
Never smoker, n (%)	544 (59.1)	0 (0%)	293 (54.9)	0 (0)	171(53.9)	0 (0)	215 (46.6)	49 (41.5)
Ex smoker, n (%)	226 (24.6)	620 (60.2)	135 (25.3)	348 (48.9)	105 (33.1)	370 (61.6)	151 (32.8)	25 (21.2)
Pack-years, median (p25 – p75)*	7.9(2.1 -17.3)	38.7(29.7-49.5)	10.8(4.9 – 20.5)	16.8(8.5 – 26.7)	7.4(3 – 15.5)	15.2(7 – 25.2)	16.3 (4.9-32.9)	37.0 (15.4-52.7)
Steroid use in last 12 months (%)								
Oral steroids, n (%)	51 (5.5)	n.a.	3 (0.6)	0(0)	0 (0)	0 (0)	n.a.	n.a.
Inhaled steroids, n (%)	506 (55.0)	n.a.	254 (47.6)	30 (4.2)	136 (42.9)	34 (5.7)	n.a.	n.a.
Oral and inhaled steroids, n (%)	23 (2.5)	n.a.	22 (4.1)	2 (0.3)	21 (6.6)	0 (0)	n.a.	n.a.
No steroids, n (%)	325 (35.3)	n.a.	255 (47.8)	679 (95.5)	160 (50.5)	567(94.3)	n.a.	n.a.
Lung function variables								
FEV1 pre med, L, mean (SD)†	2.81 (0.95)	2.76 (0.70)	3.27 (0.78)	2.89 (0.77)	3.29 (0.86)	2.84 (0.72)	3.02 (0.95)	2.83 (0.79)
FVC pre med, L, mean (SD)‡	3.35 (1.44)	3.31 (1.61)	4.33 (1.01)	4.5 (1.07)	4.34 (1.06)	4.4 (1.04)	4.17 (1.11)	4.37 (1.17)
FEV1 percent pred, mean (SD)§	86 (21)	82.3 (18.3)	89.28 (12.9)	81.6 (13.9)	91.8 (13.8)	84.3 (13.0)	90.0	89.3
FEV1/FVC, mean (SD)¶	79.72 (10.6)	60.28 (8.5)	75.7 (7.4)	63.9 (5.7)	75.9 (7.2)	64.4 (5.4)	72.1 (9.7)	64.6 (4.9)

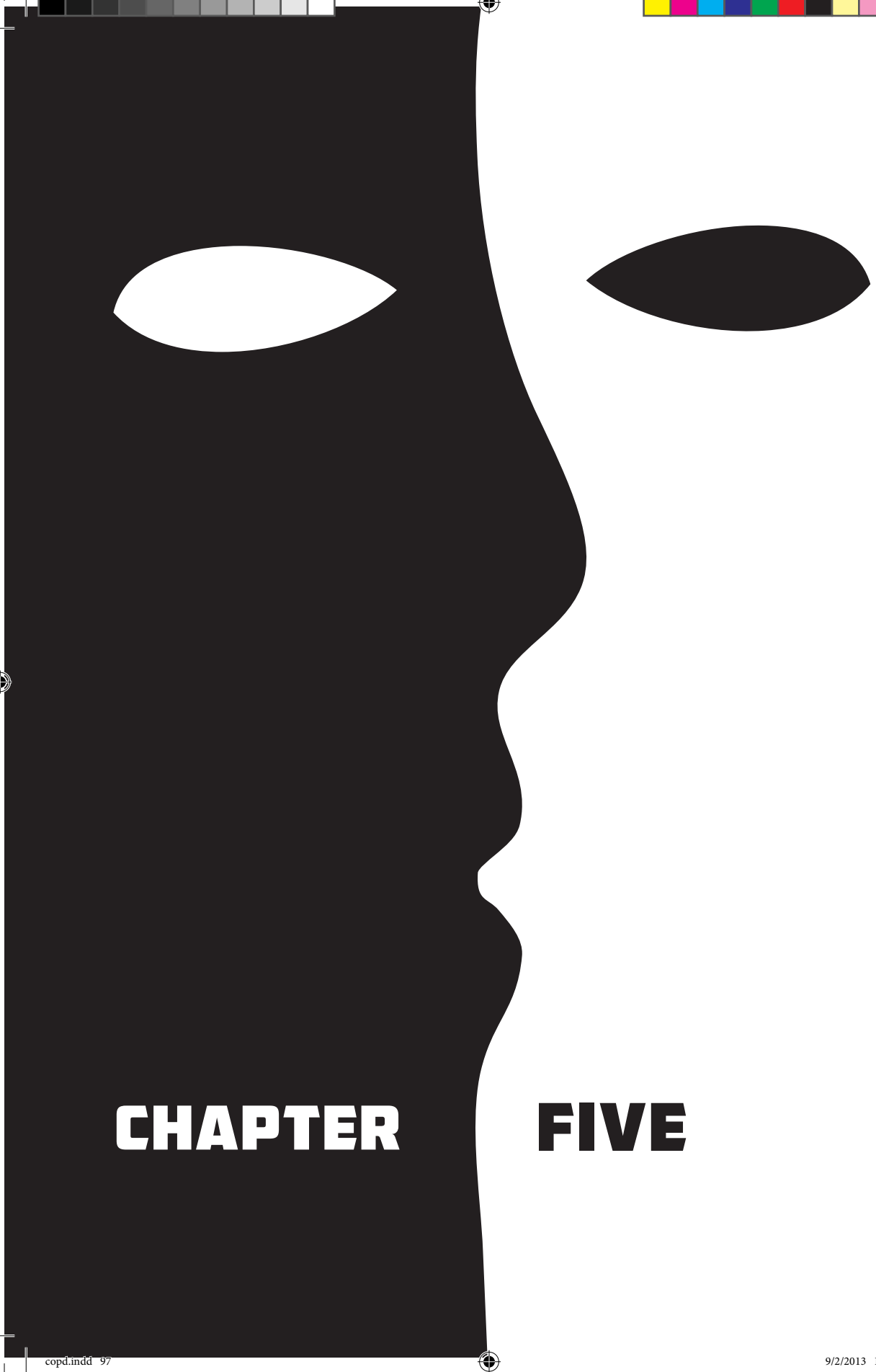
* calculated in ever smokers
† forced expiratory volume in one second
‡ forced vital capacity
§FEV1 percent predicted
¶ FEV1/FVC ratio, used for COPD diagnosis

Table S4. Heterogeneity between asthma and COPD cohorts.

			Meta1 (DAG and NELSON)		Meta1 (LifeLines 1)		Meta1 (LifeLines 2)		Meta3 (SAPALDIA)	
CHR	BP	SNP	Q	I	Q	I	Q	I	Q	I
2	15822156	rs6728667	0.60	0.00	0.56	0.00	0.06	71.84	0.86	0.00
2	15822185	rs6728750	0.61	0.00	0.57	0.00	0.05	73.63	0.95	0.00
2	15823917	rs2544534	0.58	0.00	0.24	28.27	0.03	78.03	0.97	0.00
2	15827908	rs1477253	0.58	0.00	0.28	15.12	0.05	74.04	0.97	0.00
2	15830470	rs2693008	0.57	0.00	0.72	0.00	0.03	79.96	1.00	0.00
2	15837774	rs2544523	0.50	0.00	0.90	0.00	0.05	74.58	0.81	0.00
2	15839739	rs2693019	0.44	0.00	0.94	0.00	0.04	76.24	0.78	0.00
2	215298827	rs280621	0.63	0.00	0.14	55.22	NA	NA	0.08	66.46
5	115623770	rs10036292	0.52	0.00	0.72	0.00	0.07	68.61	0.06	72.49
5	115624947	rs10043228	0.49	0.00	0.79	0.00	0.08	67.80	0.06	72.04
5	115633819	rs254149	0.73	0.00	0.40	0.00	0.15	52.30	0.03	77.59
5	115642406	rs7727882	0.57	0.00	0.81	0.00	NA	NA	0.07	69.11
13	46738025	rs17069787	0.92	0.00	0.59	0.00	0.90	0.00	0.52	0.00
13	46739001	rs7994542	0.45	0.00	0.14	54.75	0.47	0.00	0.65	0.00
13	46741378	rs9534578	0.61	0.00	0.57	0.00	0.85	0.00	0.60	0.00
13	46877470	rs7989394	0.04	76.36	0.04	76.51	NA	NA	0.25	23.28

References

1. van Iersel CA, de Koning HJ, Draisma G et al. Risk-based selection from the general population in a screening trial: selection criteria, recruitment and power for the Dutch-Belgian randomised lung cancer multi-slice CT screening trial (NELSON). Int J Cancer 2007;120:868-874.
2. Stolk RP, Rosmalen JG, Postma DS et al. Universal risk factors for multifactorial diseases: LifeLines: a three-generation population-based study. Eur J Epidemiol 2008;23:67-74.
3. Browning BL, Browning SR. A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. Am J Hum Genet 2009;84:210-223.
4. Li Y, Willer CJ, Ding J, Scheet P, Abecasis GR. MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. Genet Epidemiol 2010;34:816-834.
5. Purcell S, Neale B, Todd-Brown K et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 2007;81:559-575.
6. Aulchenko YS, Struchalin MV, van Duijn CM. ProbABEL package for genome-wide association analysis of imputed data. BMC Bioinformatics 2010;11:134.
7. R: A Language and Environment for Statistical Computing. 2011.
8. Mostafavi S, Ray D, Warde-Farley D, Grouios C, Morris Q. GeneMANIA: a real-time multiple association network integration algorithm for predicting gene function. Genome Biol 2008;9 Suppl 1:S4.
9. Chang JT, Nevins JR. GATHER: a systems approach to interpreting genomic signatures. Bioinformatics 2006;22:2926-2933.



CHAPTER FIVE

Genome-wide association study and gene-environment interactions in lung function decline.

Joanna Smolonska[1,2], Firdaus A.A. Mohamed Hoesein[3], Pieter Zanen[3], Harry J.M. Groen[4,5], Salome Scholtens[2], Judith M Vonk[2] , Cleo C van Diemen[1], Ivan Curjuric[6,7], H Susan Picavet[8], Gian-Andri Thun[6,7], Willem Mali[9], Matthijs Oudkerk[10], Harry de Koning[11], WM Verschuren[8], Medea Imboden[6,7], Dirkje S. Postma[12,] H.A. Smit[13], David P Strachan[14], Nicole Probst-Hensch[6,7], Cisca Wijmenga[1], H. Marike Boezen[2,5]

[1]Department of Genetics, University of Groningen, UMC Groningen, the Netherlands
[2]Department of Epidemiology, University of Groningen, UMC Groningen, the Netherlands
[3]Department of Pulmonology, Division of Heart and Lungs, UMC Utrecht, the Netherlands
[4]Department of Pulmonary Diseases, University of Groningen, UMC Groningen, the Netherlands
[5]GRIAC Research Institute, University of Groningen, UMC Groningen, the Netherlands
[6]Department of Epidemiology and Public Health, Swiss Tropical and Public Health Institute SwissTPH, Basel, Switzerland
[7]University of Basel, Basel, Switzerland
[8] Centre for Prevention and Health Services Research, National Institute for Public Health and the Environment, Bilthoven, the Netherlands
[9] Department of Radiology, UMC Utrecht, the Netherlands
[10]Department of Radiology, UMC Groningen, the Netherlands
[11] Department of Public Health, Erasmus MC, Rotterdam, the Netherlands
[12]Department of Pulmonology, University of Groningen, UMC Groningen, the Netherlands
[13] Julius Center for Health Sciences and Primary Care, UMC Utrecht, the Netherlands
[14] Division of Population Health Sciences and Education, St George's, University of London, United Kingdom

AT A GLANCE COMMENTARY

Scientific Knowledge on the Subject

Genetics of lung function decline have scarcely been studied and outcomes were inconsistent probably due to differences in study populations and environmental factors.

What This Study Adds to the Field

The current study identified multiple genes for lung function decline, interacting with smoking habits and COPD status, which indicates differential effects of the same loci in several subgroups. Future studies on lung function decline should include analyses stratified by risk factors and health status in order to detect signals that might otherwise be missed.

This article has an online data supplement, which is accessible from this issue's table of content online at www.atsjournals.org

Abstract

RATIONALE

Lung function decreases gradually with time due to ageing however, some subjects experience an accelerated FEV_1 decline leading to COPD.

OBJECTIVES

The aim of this study was to identify genetic factors and gene-environment interactions associated with this accelerated FEV_1 decline.

METHODS

We performed a genome-wide association study on accelerated FEV_1 decline in the NELSON cohort, a population of heavy smoking males, with replication of the 50 most significantly associated SNPs in four independent replication cohorts. These SNPs were also investigated for interactions with COPD status and smoking history.

RESULTS

Two SNPs were replicated at nominal level (rs7182547 and rs2458416 annotating to *TM7SF4* and *TRPM1*), but genome-wide significance for lung function decline was not reached. We identified SNPs interacting with COPD status (rs10928087) and smoking (rs2492321 annotating to *JUN*, rs12671330 and rs12705973 annotating to *FOXP2* and rs4723295 annotating to *BBS9*).

CONCLUSIONS

The current study identified multiple genes interacting with smoking habits and COPD status, which suggests a differential effect of the same loci in several subgroups. Future studies on lung function decline should include stratified analyses in order to detect signals that might otherwise be missed.

Introduction

Human lung growth and development starts prenatally and continues until reaching a plateau. After the plateau phase which ends between 18–25 years of age[1], lung function decreases gradually with time due to ageing. Rapid decline is observed in some smokers and COPD patients, which can result in early lung function impairment[1,2] (Figure 1). Notably, the decline in FEV_1 does not always start from an optimal FEV_1, since not all subjects reach their maximal attainable level at young adulthood. Lung function impairment at elderly age can therefore be due to a) sub-maximally attained level of lung function, b) an abnormally early onset of decline, c) an abnormally high rate of decline, or d) a combination of any of these (see Figure 1). Therefore, the genetics of these different patterns are best studied using a longitudinal population-based study design, covering the time span during which these different patterns and their underlying causes can evolve[3].

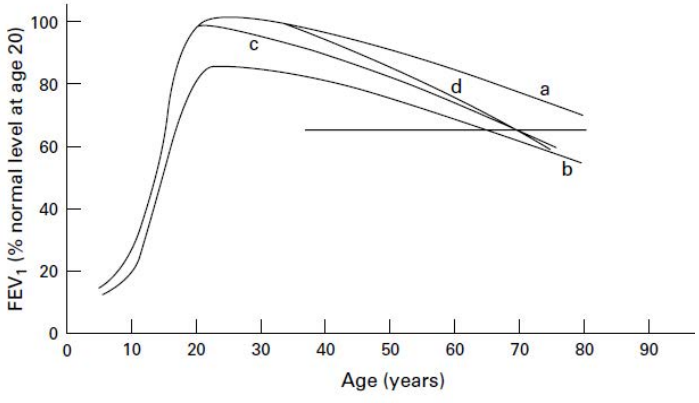


Figure 1. Lung function decline in relation to age. From Kerstjens et al (with permission)[1].
a – healthy normal subjects; b – submaximal growth but normal decline; c – premature or early decline; d – an accelerated decline

Apart from genetics, the rate of the decline is dependent on many other factors, specifically cigarette smoking. Cigarette smoking may have different effects at different periods of life. Moreover, smoking might interact with distinct genes throughout different phases in life. Therefore, taking such potential gene-by-smoking interactions into account may lead to identification of specific genetic variants playing a role in lung function decline, other than those identified in genome wide association studies (GWAS)[3] focusing on main gene effects only. Studying genetics of FEV_1 decline in both smokers and non-smokers thus seems to be a crucial issue.

Most GWAS so far have focused on the level and not on decline of lung function [4–8]. Only two GWAS addressing FEV_1 decline were published so far. The first study focused on decline in asthmatic and non-asthmatic subjects[9]. Although the top hits in this study did

not reach genome-wide significance, the study clearly showed that the main genetic signals identified differed between asthmatic and non-asthmatic subjects. The second GWAS identified two loci associated with decline in smokers with mild COPD[10], which were distinct from those identified in the GWAS in asthmatics and non-asthmatics. Together, these findings are in line with the hypothesis that different genetic factors may affect biologically different pathways leading to lung function decline in different subgroups like asthmatics, subjects with COPD and individuals without COPD.

In the current study we aimed to identify novel genes associated with lung function decline in smokers and non-smokers, and studied interactions with smoking and COPD status.

Methods

Study populations

The discovery cohort was derived from the NELSON study[11] which consisted of heavy-smoking Caucasian males of Dutch descent. Longitudinal data, with a follow-up FEV_1 measurement three to four years after the baseline FEV_1 measurement were available for 1,942 males with age ranging from 47.0 to 76.9 years. We performed replication in four population-based cohorts, being:

1. The Vlagtwedde/Vlaardingen cohort, a Dutch population-based cohort including individuals from Vlagtwedde, a rural area, and Vlaardingen, an urban area in the Netherlands. Participants were followed for over 25 years with lung function measurements every 3 years. The age of participants ranged from 35.0 to 79.0 years. Details have been described previously[12].
2. The Doetinchem cohort, a Dutch population-based cohort followed up for 10-15 years with lung function measurements every 5 years. Participants were aged from 31.2 to 70.9 years. Details have been described previously[13].
3. The SAPALDIA cohort, a multi-center study in eight geographic areas representing the range of environmental, meteorological and socio-demographic conditions in Switzerland. It was initiated in 1991 (SAPALDIA 1) with a follow-up assessment in 2002 (SAPALDIA 2) and 2011 (SAPALDIA3). Only non-asthmatic participants of SAPALDIA cohort were used for current study with age ranging from 18.30 to 61.40 years.
4. The British 1958 birth cohort (B58C) is an ongoing follow-up of all persons born in Britain during one week in 1958. Lung function was measured at ages 35 and 45 years on a subsample of the cohort with a history of chest illness in childhood, and a control sample with no such history, as described in detail elsewhere[14]. The main characteristics of the cohorts used in the current study are shown in Table 1.

SAPALDIA and B58C were part of previous GWA study on lung function decline [9].
More details can be found in the supplementary methods.

Genotyping

Samples from the discovery cohort NELSON were genotyped using the Illumina Quad610 BeadChip. The replication cohorts Vlagtwedde/Vlaardingen and Doetinchem were genotyped using custom-made VeraCode assay from Illumina. The SAPALDIA cohort and the B58C performed *in silico* replication using their genome-wide genotyping data (See supplementary methods).

Quality control

Quality control was performed for both the genome-wide and replication genotyping data. Genome-wide data were passed through standard filtering to remove single nucleotide polymprphisms (SNPs) and individual samples of lower quality. SNPs with call rate below 95%, minor allele frequency below 5% and deviating from Hardy-Weinberg equilibrium (HWE; $p<0.0001$) were removed. Samples with more than 5% of genotypes missing were removed, as well as duplicates, related individuals or ethnic outliers (based on components from multidimensional scaling C1 and C2). In the replication cohorts the same filters were applied, except for HWE ($p<0.001$). All filtering steps analyses were performed using PLINK[15]. Plots were generated using R[16].

Phenotype definition

FEV_1 was measured following ATS/ERS guidelines in all cohorts (see online supplement for detailed information regarding the specificities). Lung function decline was defined as the baseline FEV_1 minus the last FEV_1 measurement during the follow-up, divided by the time between measurements. The decline in FEV_1 is expressed in milliliters per year.

Statistical analysis

The association between genotypes and the annual decline in FEV_1 in the NELSON cohort was assessed using linear regression models. Covariates included in the model were: baseline FEV_1, packyears smoked and current/ex- smoking status. We did not adjust for gender, since NELSON consisted of males only, or age, given the cohort's restricted age range resulting in a high correlation with packyears. Baseline FEV_1 was included in order to assure that genetic effects detected will refer to lung FEV_1 decline solely, irrespective of the initial level.

The analyses in the different replication cohorts were performed as in the discovery cohort (i.e. with adjustment for baseline FEV_1, packyears smoked and current/ex- vs. never-smoking status) and additional adjustment for gender. The analysis in SAPALDIA was additionally adjusted for area.

Within the Doetinchem and the Vlagtwedde/Vlaardingen cohort we performed stratified analyses according to:

1. presence or absence of COPD (FEV_1/FVC<0.7 according to GOLD criteria)
2. smoking habits, dichotomized at 20 packyears

For each stratum linear regression was performed and regression coefficients of the groups were compared (Z-score comparison) to test for interaction between gene and 'environment' (GxE), i.e. gene-by-COPD interaction and gene-by-smoking interaction.

Additionally, GxE analyses were performed according to COPD status separately in the Doetinchem cohort, i.e. gene-by-smoking interaction within COPD controls and COPD cases. Due to an insufficient number of COPD cases these analysis could not be performed in the Vlagtwedde/Vlaardingen cohort study.

Results

All estimates of genetic effects presented in the current manuscript represent decline of lung function expressed in milliliters per year. The values above zero reflect accelerated FEV_1 decline, while values below zero indicate attenuated FEV_1 decline due to a particular allele.

SNPs associated with annual FEV_1 decline in NELSON

Quality control indicated there was no population stratification (see Q-Q plot, Figure 2). After quality control 522,004 SNPs and 1,942 individuals remained. The genomic inflation factor (**λ**) equaled to 1. The genome-wide screen for variants associated with FEV_1 decline did not identify any SNPs at a genome-wide significance threshold (being $5{\cdot}10^{-8}$, Figure 3). Fifty SNPs with $p<5{\cdot}10^{-5}$ were selected for further replication in the replication cohorts.

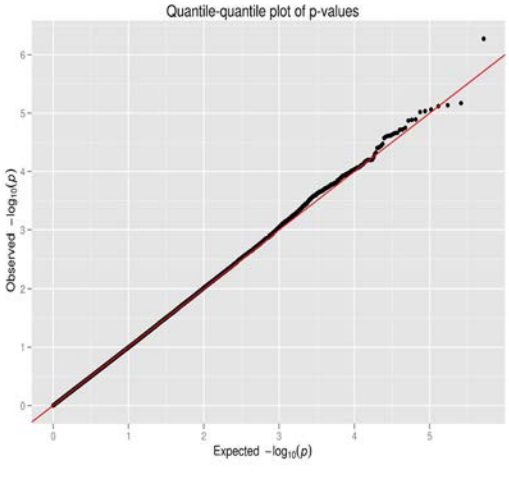


Figure 2. Q-Q plot. Observed p values are derived from linear regression analysis of lung function decline.

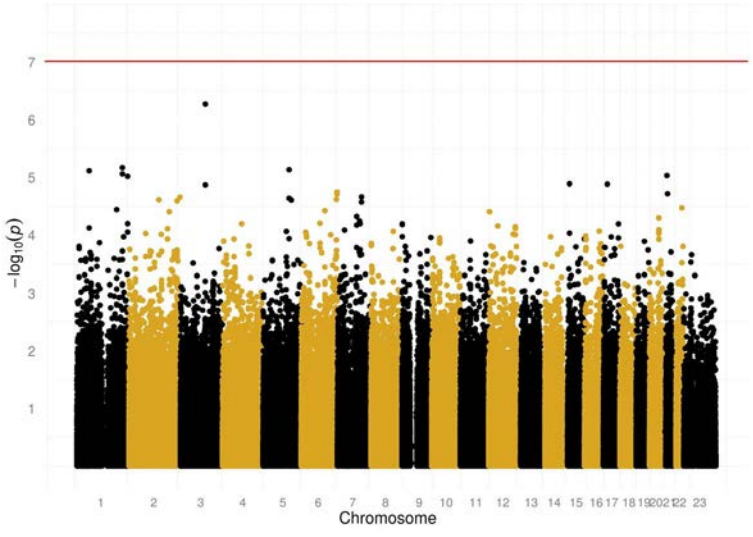


Figure 3. Manhattan plot. Distribution of FEV_1 decline association signals across the genome.

Replication of SNPs associated with annual FEV_1 decline in Vlagtwedde/Vlaardingen, Doetinchem, SAPALDIA and B58C

After quality control in the Vlagtwedde/Vlaardingen cohort 1,056 individuals and 43 SNPs and in the Doetinchem cohort 1,093 individuals and 41 SNPs were included. SAPALDIA and B58C had genotypes available for all top 50 SNPs in 788 individuals and 834 individuals respectively. Characteristics of the cohorts used are shown in Table 1.

Table 1. Cohorts characteristics.

Cohort	NELSON	Vlagtwedde/Vlaardingen	Doetinchem	SAPALDIA	B58C
N	1942	1056	1093	788	834
Age at baseline *(Y, SD)*	59.8 (5.4)	50.7 (10.4)	46.2 (9.9)	41.63 (11.10)	35.0 (0.2)
Height (cm, *SD)*	178.7 (6.3)	170.3 (9.3)	173.2 (9.0)	169.1 (9.32)	169.7 (9.3)
Follow-up time *(Y, SD)*	2.99 (0.44)	19.83 (4.45)	8.36 (2.33)	10.96 (0.23)	10.2 (0.5)
FEV1 (l, *SD)*	3.44 (0.73)	3.05 (0.76)	3.51 (0.81)	3.63 (0.81)	3.62 (0.75)
FVC (l, *SD)*	4.22 (1.29)	3.92 (0.96)	4.65 (1.04)	4.57 (1.01)	4.53 (0.92)
FEV1/FVC *(%, SD)*	72 (9.5)	77.84 (5.22)	75.77 (7.95)	79.68 (7.10)	80.1 (6.5)
FEV1 decline (ml, *SD)*	199.44 (242.32)	12.97 (21.52)	31.13 (38.67)	33.87 (28.22)	33.8 (32.1)
FEV1%pred *(%, SD)*	98.9 (18.5)	98.13 (13.02)	N/A	101.64 (12.43)	N/A
% male	100	48.6	50.4	49.24	48.4
N (%) never smoker	0 (0)	366 (34.6)	337 (30.8)	339 (43.02)	423 (50.7)
N (%) current smoker	867 (44.6)	298(28.2)	339 (38.1)	243 (30.84)	159 (19.1)
N (%) former smoker	1075 (55.4)	392?	416 (311.0)	206 (26.14)	252 (30.2)
Pack-years smoking*	37.95 (21.5-133.2)	5.62 (0 - 128.0)	6 (0 - 118)	19.06 (20.87)	14.5 (0.1 - 58.0)

All values expressed as mean (SD), unless specified otherwise. *median (range)

Table 2. Main results of genome-wide association study on lung function decline and replication in the

				Nelson			Vlagtwedde/Vlaardingen			Doetinchem		
CHR	SNP	GENE	A1	BETA	SE	P	BETA	SE	P	BETA	SE	P
15	rs7182547	TRPM1	A	15.78	3.82	3.69E-05	0.99	0.86	0.25			
8	rs2458416	TM7SF4	G	23.39	5.92	8.17E-05	1.53	1.32	0.25	0.39	2.69	0.89
6	rs4947133	NA	A	20.72	5.20	6.91E-05						
7	rs12667992	NA	A	25.63	6.29	4.73E-05	2.03	1.43	0.16	2.44	2.64	0.36
2	rs4664511	CACNB4	G	20.23	5.44	0.000203	1.22	1.36	0.37	5.42	2.83	0.06
7	rs1916602	NA	G	24.31	5.90	3.89E-05	1.28	1.31	0.33	1.96	2.53	0.44
7	rs12705973	FOXP2	G	-12.96	3.25	7.00E-05	-1.04	0.80	0.19	-0.08	1.62	0.96
5	rs11949657	PHF15	C	22.70	5.46	3.37E-05	2.49	1.18	0.04	-1.80	2.56	0.48
7	rs12671330	NA	G	-13.10	3.26	5.91E-05	-0.83	0.79	0.29	-0.10	1.61	0.95
17	rs17581728	UNC13D	A	15.28	3.87	8.33E-05						
21	rs13049780	NA	G	25.99	5.70	5.32E-06	0.05	1.35	0.97	0.41	2.68	0.88
1	rs2492321	LOC645757	A	-18.15	3.98	5.41E-06	-1.10	0.98	0.26	2.66	1.91	0.16
17	rs2258367	SLC5A10	C	23.68	5.46	1.49E-05	0.60	1.28	0.64	-2.51	2.52	0.32
5	rs1366331	NA	A	22.87	5.79	8.10E-05	0.18	1.46	0.90	3.44	2.94	0.24
2	rs892867	NA	G	-13.09	3.57	0.000251	-0.56	0.84	0.51	-3.13	1.72	0.07
1	rs6604558	SPATA17	A	18.62	4.04	4.37E-06	-0.03	0.95	0.98	2.40	1.90	0.21
4	rs1394339	NA	A	-12.57	3.34	0.000171	-0.42	0.79	0.59	-0.54	1.59	0.73
1	rs6694862	SPATA17	G	18.46	4.05	5.53E-06	0.09	0.95	0.92	2.24	1.90	0.24
6	rs221749	PDE10A	G	-15.62	3.81	4.22E-05	-0.75	0.91	0.41	0.56	1.80	0.75
22	rs139156	PARVG	A	25.34	5.98	2.35E-05						
21	rs2827251	NA	A	26.56	6.27	2.39E-05	0.85	1.54	0.58	1.20	3.04	0.69
2	rs11679243	NA	A	16.72	3.88	1.75E-05						
6	rs1498426	CDKAL1	G	-13.45	3.37	6.67E-05	1.22	0.80	0.13	-1.74	1.59	0.28
7	rs9791835	NA	A	22.70	5.66	6.38E-05	1.60	1.28	0.21	-1.96	2.59	0.45
12	rs11058897	NA	A	29.88	7.37	5.22E-05						
12	rs10847254	NA	G	30.14	7.37	4.47E-05						
12	rs11058900	NA	G	29.58	7.44	7.22E-05						
5	rs1862404	NA	G	-15.66	3.52	9.11E-06	2.63	0.84	0.00	0.96	1.68	0.57
5	rs4836058	NA	G	-14.17	3.35	2.51E-05	2.13	0.80	0.01	0.18	1.62	0.91
6	rs12528430	NA	A	13.80	3.34	3.79E-05	-1.32	0.80	0.10	0.10	1.65	0.95
6	rs17182135	NA	G	13.84	3.36	3.88E-05	-1.23	0.80	0.13	0.16	1.65	0.92
9	rs10962687	NA	G	26.59	6.75	8.42E-05	-1.26	1.59	0.43	0.36	3.52	0.92
9	rs10962689	NA	G	26.55	6.75	8.69E-05	-1.16	1.55	0.45	0.64	3.51	0.85
3	rs1554328	NA	A	25.60	5.28	1.35E-06	-0.05	1.23	0.96	-4.34	2.46	0.08
20	rs1013562	NA	G	18.08	4.41	4.29E-05	-1.28	0.99	0.20	2.72	2.14	0.21
3	rs938111	NA	C	21.08	5.08	3.47E-05	-0.09	1.18	0.94	-1.20	2.36	0.61
2	rs10928087	LRP1B	G	24.24	5.96	4.97E-05	0.20	1.51	0.90	0.65	2.84	0.82
1	rs927743	NA	A	-13.54	3.42	7.96E-05	-0.97	0.82	0.23	3.02	1.62	0.06
20	rs6073768	NA	G	18.90	4.45	2.25E-05	-0.65	1.00	0.52	2.43	2.14	0.26
2	rs2690751	MGC13057	A	17.01	3.98	1.97E-05	-0.56	0.96	0.56	-2.60	1.89	0.17
12	rs4639992	NELL2	G	19.08	4.87	9.35E-05	1.31	1.13	0.24	-3.32	2.30	0.15
1	rs11589174	NA	C	18.78	4.81	9.72E-05	-1.03	1.15	0.37	2.21	2.46	0.37
12	rs16946677	NA	C	20.51	5.27	0.000102	-1.10	1.18	0.35	0.92	2.48	0.71
20	rs6073862	WFDC10B	G	18.10	4.41	4.21E-05	-1.09	0.97	0.26	3.12	2.14	0.15
1	rs6677308	NA	A	20.33	4.76	2.03E-05	-1.54	1.16	0.19	2.32	2.40	0.34
7	rs4723295	BBS9	G	20.84	5.08	4.32E-05	-0.42	1.18	0.72	-3.88	2.40	0.11
1	rs12093170	NA	C	17.27	4.50	0.000128	-0.50	1.08	0.64	0.26	2.29	0.91
2	rs12467492	NA	A	14.14	3.35	2.57E-05	-1.43	0.77	0.06	-0.11	1.60	0.95
1	rs920307	NA	A	-20.74	5.01	3.63E-05	0.20	1.05	0.85			
2	rs3770111	ITGA4	A	21.74	5.45	6.98E-05	-1.15	1.28	0.37	0.07	2.53	0.98

total population.

SAPALDIA			B58C			Replication		Overall	
BETA	SE	P	BETA	SE	P	BETA	P	BETA	P
2.98	1.34	0.03	1.24	1.67	0.46	1.52	0.02	1.93	5.68E-06
4.89	2.10	0.02	1.23	2.68	0.65	2.05	0.03	2.60	1.76E-05
6.28	3.12	0.04	2.28	3.25	0.48	4.36	0.05	6.95	2.86E-05
2.71	2.08	0.19	-2.88	2.90	0.32	1.65	0.10	2.26	5.45E-05
2.92	2.11	0.17	0.05	2.73	0.98	1.95	0.05	2.53	5.73E-05
2.71	1.87	0.15	-2.48	2.63	0.35	1.25	0.17	1.81	1.09E-04
-0.92	1.25	0.46	-0.47	1.51	0.76	-0.81	0.16	-1.18	1.42E-04
2.34	2.54	0.36	-4.12	2.43	0.09	0.98	0.28	1.58	2.24E-04
-0.74	1.25	0.55	-0.55	1.51	0.72	-0.68	0.23	-1.05	2.33E-04
-0.42	1.49	0.78	3.88	1.76	0.03	1.36	0.23	2.47	2.82E-04
1.34	2.54	0.60	0.66	2.72	0.81	0.39	0.70	1.18	4.78E-04
-2.64	1.92	0.17	1.87	1.78	0.29	-0.28	0.70	-0.86	4.82E-04
3.23	2.23	0.15	-2.70	5.99	0.65	0.54	0.59	1.30	5.71E-04
2.61	2.27	0.25	-1.76	3.25	0.59	0.94	0.38	1.66	6.56E-04
-1.14	1.42	0.42	1.53	1.64	0.35	-0.71	0.25	-1.06	6.81E-04
-1.37	1.58	0.38	-0.22	1.86	0.91	0.01	0.99	0.54	1.14E-03
0.70	1.35	0.60	-2.15	1.52	0.16	-0.48	0.41	-0.83	1.17E-03
-1.45	1.58	0.36	-0.47	1.86	0.80	0.00	0.99	0.53	1.29E-03
0.97	1.45	0.50	-1.15	1.71	0.50	-0.29	0.66	-0.73	1.36E-03
2.56	2.65	0.33	-1.80	2.88	0.53	0.56	0.77	2.94	1.41E-03
2.35	3.05	0.44	-4.00	2.67	0.13	0.23	0.84	1.07	1.73E-03
0.27	1.63	0.87	0.04	1.75	0.98	0.16	0.89	1.59	1.74E-03
-1.83	1.28	0.15	-1.31	1.51	0.39	-0.15	0.80	-0.53	2.67E-03
-1.15	2.22	0.61	-1.61	2.72	0.55	0.22	0.82	0.84	2.80E-03
-0.72	2.27	0.75	2.45	3.37	0.47	0.27	0.89	2.08	3.07E-03
-0.90	2.31	0.69	2.42	3.42	0.48	0.13	0.94	2.03	3.33E-03
-0.71	2.27	0.75	2.45	3.37	0.47	0.27	0.88	2.04	3.62E-03
0.41	1.50	0.79	0.81	1.61	0.62	1.76	0.00	1.23	0.04
0.22	1.39	0.87	1.14	1.55	0.46	1.39	0.02	0.92	0.11
-0.99	1.27	0.43	-1.40	1.54	0.36	-1.09	0.06	-0.65	0.25
-0.89	1.27	0.48	-1.40	1.54	0.36	-1.01	0.08	-0.58	0.31
-4.21	3.15	0.18	-5.23	3.23	0.11	-2.07	0.09	-1.17	0.33
-4.21	3.15	0.18	-5.23	3.23	0.11	-1.96	0.10	-1.08	0.36
1.64	2.04	0.42	4.66	2.36	0.05	-0.92	0.41	-0.92	0.41
0.38	1.91	0.84	-3.46	2.05	0.09	-0.82	0.28	-0.28	0.70
-0.20	1.98	0.92	-4.34	2.21	0.05	-0.90	0.30	-0.29	0.73
-2.21	2.45	0.37	-1.30	2.65	0.63	-0.45	0.68	0.33	0.76
0.45	1.26	0.72	1.54	1.53	0.31	0.21	0.72	-0.17	0.76
0.45	1.98	0.82	-2.48	2.11	0.24	-0.33	0.67	0.22	0.77
0.24	1.46	0.87	-0.70	1.65	0.67	-0.67	0.32	-0.18	0.79
-1.15	1.77	0.52	-5.98	2.25	0.01	-0.76	0.35	-0.22	0.79
-0.20	1.87	0.92	-0.73	2.30	0.75	-0.44	0.61	0.14	0.87
-1.19	1.95	0.54	0.20	2.23	0.93	-0.68	0.43	-0.13	0.88
0.54	1.92	0.78	-3.35	2.04	0.10	-0.63	0.40	-0.11	0.88
0.24	1.89	0.90	-0.28	2.24	0.90	-0.52	0.54	0.12	0.89
-0.34	2.02	0.87	0.57	2.28	0.80	-0.71	0.41	-0.11	0.90
-1.58	1.68	0.35	0.04	2.05	0.99	-0.57	0.47	-0.05	0.95
2.40	1.24	0.05	-1.06	1.48	0.48	-0.43	0.45	-0.03	0.96
2.36	1.98	0.23				0.67	0.47	-0.04	0.97
-0.17	2.29	0.94	0.14	2.42	0.95	-0.61	0.51	0.03	0.97

Table 2 shows the results for each replication cohort, for the replication cohorts

combined, and for the replication cohorts combined with the discovery cohort NELSON. Overall, two of the top SNPs were replicated with nominal p values after combining all replication cohorts (rs7182547 p=0.02 and rs2458416 p=0.03). Genome-wide significance was not achieved after combining the discovery and replication cohorts.
Gene annotations for all SNPs selected for replication are also shown in Table 2. The two replicated SNPs annotated to *TM7SF4* (rs2458416) and *TRPM1* (rs7182547) genes.

Stratified analyses of SNPs associated with annual FEV_1 decline and gene-environment interaction (GxE) in Vlagtwedde/Vlaardingen and Doetinchem

We performed stratified analyses based on presence or absence of COPD and smoking in Vlagtwedde/Vlaardingen and Doetinchem cohorts for the 50 top hits for FEV_1 decline. Beta estimates differed in all sub-groups, indicating that both factors play a role in a genetic association with lung function decline and suggesting possible interactions (i.e. COPD cases vs COPD controls and smokers with ≥20 packyears vs smokers with <20 packyears).

Stratification according to COPD status and gene-by-COPD interaction

In Doetinchem, the effect estimates for FEV_1 decline differed in the COPD cases and in subjects without COPD. In subjects with COPD, some of the SNPs effect estimates were in line with those observed in the NELSON (identification) cohort (Supplementary table 1). However, the analysis in within COPD subjects reduced the sample size, and hence the power. In the Vlagtwedde/Vlaardingen cohort the stratified analysis according to COPD status was not feasible since there were only 13 COPD cases.

The interaction in the Doetinchem cohort was significant at the nominal level for one SNP: rs10928087 on chromosome 2 (p=0.031) annotating to *LRP1B* gene (Table 3). Two other SNPs, rs1862404 and rs1366331 on chromosome 5 showed borderline significance for the interaction with p=0.051 and p=0.089, respectively. None of these SNPs showed any association with FEV_1 decline in COPD controls (n=833) and only two showed borderline association in COPD cases (n=250; rs10928087, p=0.086 and rs1862404, p=0.068).

Table 3. Gene-by-COPD status interaction in the Doetinchem cohort (COPD defined as FEV_1/FVC < 70%), GOLD criteria)

		Doetinchem all			Doetinchem COPD			Doetinchem non-COPD			GxE	
CHR	SNP	N	BETA	SE	N	BETA	SE	N	BETA	SE	Z_GXE	P_GXE
2	rs10928087	1092	0.6451	2.836	250	12.58	6.84	833	-3.759	3.235	-2.16	0.03079
5	rs1862404	1093	0.9597	1.678	251	7.84	4.305	833	-1.307	1.868	-1.949	0.05129
5	rs1366331	1093	3.439	2.942	251	10.76	7.053	833	-2.525	3.36	-1.701	0.08891

Stratification according to smoking habits and gene-by-smoking interactions

There was a clear difference in genetic effects between subjects smoking less and subjects smoking 20 packyears or more in both cohorts. Especially Doetinchem heavy-smokers were the most alike the NELSON cohort despite inclusion of females in the Doetinchem cohort. In Doetinchem five SNPs showed significant interaction with smoking ($p<0.05$, Table 4). One of them remains significant after multiple testing correction ($p<0.001$): rs2492321 annotating to *LOC100131060*, an uncharacterized gene located next to *JUN* gene. In Vlagtwedde/Vlaardingen two SNPs showed nominal interaction with smoking, of which one was also identified in the Doetinchem cohort, i.e. rs4723295 on chromosome 7 (Tables 4 and 5, highlighted with bold), but with the opposite effect.

Table 4. Gene-by-pack years interaction in the Doetinchem cohort (with pack years dichotomized at 20)

CHR	SNP	N<20py	BETA<20py	SE<20py	P<20py	N>20py	BETA>20py	SE>20py	P>20py	Z_GXE	P_GXE
1	rs2492321	807	−0.8699	2.164	0.6877	286	15.96	4.536	0.0005037	−3.35	0.0008093
7	rs12671330*	807	−3.743	1.872	0.04597	285	8.385	3.502	0.01731	−3.054	0.00226
7	rs12705973*	806	−3.562	1.886	0.05929	286	7.87	3.552	0.0275	−2.843	0.004473
7	**rs4723295**	807	−8.503	2.783	0.002323	286	6.051	5.273	0.2522	−2.441	0.01466
2	rs10928087	807	−3.97	3.421	0.2462	285	10.05	5.754	0.08177	−2.094	0.03622
20	rs1013562	805	5.247	2.44	0.0318	286	−5.391	5.027	0.2845	1.904	0.05695
20	rs6073862	806	5.71	2.446	0.01983	286	−4.687	4.937	0.3432	1.887	0.05914
2	rs4664511*	807	1.78	3.32	0.592	286	13.58	6.111	0.02705	−1.697	0.08972
2	rs2690751	806	−0.4001	2.178	0.8543	286	−8.389	4.308	0.05247	1.655	0.09793

* SNPs identified also in the similar analysis, but restricted to COPD cases (Table 6)

Table 5. Gene-by-pack years interaction in the Vlagtwedde/Vlaardingen cohort (with pack years dichotomized at 20).

CHR	SNP	N<20py	BETA<20py	SE<20py	P<20py	N>20py	BETA>20py	SE>20py	P>20py	Z_GXE	P_GXE
7	**rs4723295**	806	2.063	1.68	0.2199	250	−5.252	2.385	0.0286	2.507	0.01217
8	rs2458416	806	2.939	1.816	0.1059	250	−4.179	2.986	0.1629	2.037	0.04167
2	rs3770111	806	2.542	1.849	0.1696	250	−2.983	2.515	0.2368	1.77	0.07676

Stratification according to smoking habits within COPD strata and gene-by-smoking interactions in the Doetinchem cohort

We performed a gene-by-smoking interaction analysis in the groups of individuals with and without COPD of the Doetinchem cohort (Tables 6 and 7). Four and five SNPs interacted significantly with smoking in groups with and without COPD, respectively. These SNPs did not overlap. Three of the SNPs identified in the group with COPD were also identified in gene-by-smoking interaction analysis in the total Doetinchem cohort (marked with asterisk in Tables 4 and 7, full results in Supplementary Tables 2 and 3).

Table 6. Gene-by-pack years interaction in the Doetinchem cohort in individuals without COPD (with pack years dichotomized at 20).

CHR	SNP	N<20py	BETA<20py	SE<20py	P<20py	N>20py	BETA>20py	SE>20py	P>20py	Z_GXE	P_GXE
1	rs2492321	659	-1.612	2.23	0.4701	180	14.39	5.282	0.0071	-2.79	0.005266
7	rs4723295	659	-8.674	2.88	0.0027	180	7.651	5.752	0.1852	-2.538	0.01116
20	rs6073768	658	3.2	2.583	0.2159	180	-10.2	5.784	0.07957	2.115	0.03442
20	rs6073862	658	4.378	2.576	0.08971	180	-8.076	5.803	0.1658	1.961	0.04982
20	rs1013562	657	3.959	2.565	0.1232	180	-8.076	5.803	0.1658	1.897	0.05785

Table 7. Gene-by-pack years interaction in the Doetinchem cohort in individuals with COPD only (with pack years dichotomized at 20).

CHR	SNP	N<20py	BETA<20py	SE<20py	P<20py	N>20py	BETA>20py	SE>20py	P>20py	Z_GXE	P_GXE
2	*rs4664511	148	-16.76	9.436	0.0776	106	22.89	8.783	0.01051	-3.076	0.0021
1	rs6677308	148	-4.606	7.012	0.5123	106	26.56	9.621	0.006819	-2.618	0.008848
7	*rs12671330	148	-5.28	5.321	0.3227	105	13.29	6.244	0.03563	-2.264	0.02357
7	*rs12705973	148	-4.751	5.514	0.3903	106	12.86	6.361	0.04574	-2.092	0.03641
1	rs12093170	148	-6.546	7.094	0.3576	106	18.71	9.929	0.06232	-2.07	0.03849
1	rs11589174	148	-2.635	7.3	0.7187	106	20.75	10.78	0.05685	-1.797	0.07235

* SNPs identified also in the similar analysis, but in the entire cohort (Table 3)

Discussion

In the current study we performed a genome-wide search for SNPs associated with FEV_1 decline. Additionally, we studied gene-by-smoking interactions and gene-by-COPD status interactions. Our GWA study did not identify genome-wide significant SNPs with lung function decline. Three SNPs showed nominal level of association in the replication sample:

rs1862404 (p=0.009), rs12467492 (p=0.012) and rs4836058 (p=0.027), however, the direction of the association was opposite to the discovery cohort.

In the Doetinchem cohort four SNPs had a significant interaction with smoking habits with respect to FEV_1 decline: rs2492321 annotating to *LOC100131060*, next to *JUN*, rs12671330 and rs12705973 annotating to *FOXP2* and rs4723295 located in *BBS9* gene. While nothing is known about *LOC100131060*, Jun is a proto oncogene, known to interact with many proteins and involved in many pathways. One of them is the JNK (c-Jun N terminal kinase) pathway which has been shown to be activated in human endothelial cells upon cigarette smoke exposure[17]. *FOXP2* (forkhead box P2) encodes forkhead/winged-helix (FOX) transcription factor that is mainly expressed in fetal brain and interestingly in fetal lungs as well. *FOXP2* was previously shown to repress gene expression in epithelium of the lungs, together with FOXP1[18]. Further, studies in mice identified Foxp2 and Foxp1 as crucial regulators of lung development and specifically peripheral epithelial cells[19], which signifies its putative role in FEV_1 decline.

Lung function decline is a heterogeneous phenotype depending on many internal and external factors. One of the reasons for not replicating our findings might be the fact that the follow-up in the discovery cohort in our current study is shorter than in our replication cohorts, hence we might be studying not entirely the same type of decline. In case of complex phenotypes, like lung function decline, even when we adjust for covariates we know to influence the phenotype like duration of follow-up, this can simply be insufficient to account for heterogeneity within and especially between cohorts. We take differences in smoking habits and history, gender distribution and baseline FEV_1 levels into account in our analyses. However, there may have been other components that might result in heterogeneity we cannot measure, do not have access to or we are not aware of and thus cannot adjust for, yet they may contribute to cohort heterogeneity (for example physical activity, occupational exposure, etc.). In addition, as shown in Figure 1, the course of the decline can be different. The NELSON cohort consists of selected individuals, heavy-smokers that probably will display mostly the accelerated decline (Figure 1d). The replication cohorts are population based hence individuals can display any course of decline shown in Figure 1. This could be a main reason for the lack of replication in our study.

We performed gene-by-smoking and gene-by-COPD status interaction analyses in order to identify differential effects of smoking and COPD-status within strata. Alternatively, construction of a more complicated model like three-way interactions would have been feasible. However, interpretation of such results is more problematic compared to results from two-way interactions within e.g. COPD strata. The more specific the phenotype is, the lower is the available number of subjects. Indeed, if the analyses are restricted to specific phenotypes such as the Doetinchem COPD cases with ≥ 20 packyears, the effect estimates are comparable to

those in the discovery cohort (Table 7). However, even for two-way interactions the power of the current study might not be sufficient, as only one of the interactions remained significant after multiple testing correction.

The limited number of GWA studies on lung function decline performed previously, like ours, showed ambiguous results. A recent meta-analysis published by Imboden et al.[9] did not yield genome-wide significant SNPs despite a large number of individuals included in their study (the discovery set comprising 1,441 asthmatic and 2,677 non-asthmatic participants and the replication set comprising 1,160 asthmatic and 10,858 non-asthmatic participants). Lack of replication of results from a GWA study on lung function decline was also shown recently in the study by Hansel et al[10]. This study reported two genome-wide loci, but none of them was even nominally associated in the replication cohorts. This stresses the importance of optimal use of the limited resources for such genetic studies, i.e. there might be lack of replication cohorts with specific phenotypes, like lack of mild COPD patients in case of the Hansel study and heavy smoking COPD patients in case of our study. All the studies performed so far on genetics of FEV_1 decline, including our, lack overlap in reported associations and also lack overlap with GWA studies on lung function level. This stresses the complexity of the FEV_1 decline phenotype and the difficulties related to cohort heterogeneity mentioned before.

In summary, the current study reports no genome-wide significant SNPs associated with lung function decline. We identified multiple genes interacting with smoking habits and with COPD status, which indicates differential effects of the same loci in several subgroups. Future studies on lung function decline should include stratified analyses in order to detect signals that are otherwise missed in the full data set due to adding up the opposite effects of relevant loci. Efforts should be directed towards deep phenotyping and collecting detailed data on environmental exposures data.

Acknowledgements

NELSON: Authors would like to thank participants of NELSON cohort.

Vlagtwedde/Vlaardingen:
The study could not have been performed without J.P. Schouten (Unit Chronic airway diseases, Dept. of Epidemiology, University Medical Center Groningen, University of Groningen, Groningen, the Netherlands) who performed the continuous management of the Vlagtwedde-Vlaardingen data. R. van der Lende initiated the cohort in the mid-sixties of the previous century, and B. Rijcken contributed to its continuation. A big thanks to the participants of the Vlagtwedde-Vlaardingen studies for their loyal participation for over 25 years.

Doetinchem:
The study could not have been performed without the epidemiologists and fieldworkers of the Municipal Health Services in Doetinchem (the Netherlands) and their important contribution to the data collection of the Doetinchem Study, as well as J. Seidell, M. Verschuren and B. Bueno-de Mesquita (all National Institute of Public Health and the Environment (RIVM), Bilthoven, the Netherlands), and A. Blokstra and P. Vissink (both Centre for Prevention and Health Services Research, RIVM) who performed the logistic and data management.

British 1958 birth cohort: We acknowledge the use of phenotype and genotype data from the British 1958 Birth Cohort DNA collection, funded by the Medical Research Council grant G0000934 and the Wellcome Trust grant 068545/Z/02. (http://www.b58cgene.sgul.ac.uk/). Genotyping for the B58C-WTCCC subset was funded by the Wellcome Trust grant 076113/B/04/Z. The B58C-T1DGC genotyping utilized resources provided by the Type 1 Diabetes Genetics Consortium, a collaborative clinical study sponsored by the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), National Institute of Allergy and Infectious Diseases (NIAID), National Human Genome Research Institute (NHGRI), National Institute of Child Health and Human Development (NICHD), and Juvenile Diabetes Research Foundation International (JDRF) and supported by U01 DK062418. B58C-T1DGC GWAS data were deposited by the Diabetes and Inflammation Laboratory, Cambridge Institute for Medical Research (CIMR), University of Cambridge, which is funded by Juvenile Diabetes Research Foundation International, the Wellcome Trust and the National Institute for Health Research Cambridge Biomedical Research Centre; the CIMR is in receipt of a Wellcome Trust Strategic Award (079895). The B58C-GABRIEL genotyping was supported by a contract from the European Commission Framework Programme 6 (018996) and grants from the French Ministry of Research.

SAPALDIA: The study could not have been done without the help of the study participants, technical and administrative support and the medical teams and field workers at the local study sites. Local fieldworkers : Aarau: M Broglie, M Bünter, D Gashi, Basel: R Armbruster, T Damm, U Egermann, M Gut, L Maier, A Vögelin, L Walter, Davos: D Jud, N Lutz, Geneva: M Ares, M Bennour, B Galobardes, E Namer, Lugano: B Baumberger, S Boccia Soldati, E Gehrig-Van Essen, S Ronchetto, Montana: C Bonvin, C Burrus, Payerne: S Blanc, AV Ebinger, ML Fragnière, J Jordan, Wald: R Gimmi, N Kourkoulos, U Schafroth. Administrative staff: N Bauer, D Baehler, C Gabriel, R Gutknecht. SAPALDIA Team: *Study directorate:* T Rochat, NM Probst Hensch, N Künzli, C Schindler, JM Gaspoz. *Scientific team:* JC Barthélémy, W Berger, R Bettschart, A Bircher, G Bolognini, O Brändli, C Brombach, M Brutsche, L Burdet, M Frey, U Frey, MW Gerbase, D Gold, E de Groot, W Karrer, R Keller, B Knöpfli, B Martin, D Miedinger, U Neu, L Nicod, M Pons, F Roche, T Rothe, E Russi, P Schmid-Grendelmeyer, A Schmidt-Trucksäss, A Turk, J Schwartz, D. Stolz, P Straehl, JM Tschopp, A von Eckardstein, E Zemp Stutz. *Scientific team at coordinating centers:* M Adam, E Boes, PO Bridevaux, D Carballo, E Corradi, I Curjuric, J Dratva, A Di Pasquale, L Grize, D Keidel, S Kriemler, A Kumar, M Imboden, N Maire, A Mehta, F Meier, H Phuleria, E Schaffner, GA Thun, A Ineichen, M Ragettli, M Ritter, T Schikowski, G Stern, M Tarantino, M Tsai, M Wanner.

ReferenceS

1. Kerstjens, H. A., B. Rijcken, J. P. Schouten, and D. S. Postma. 1997. Decline of FEV1 by age and smoking status: facts, figures, and fallacies. *Thorax* 52:820–827.
2. Fletcher, C. and R. Peto. 1977. The natural history of chronic airflow obstruction. *Br.Med.J.* 1:1645–1648.
3. Boezen, H. M. 2009. Genome-wide association studies: what do they teach us about asthma and chronic obstructive pulmonary disease? *Proc.Am.Thorac.Soc.* 6:701–703.
4. Wilk, J. B., R. E. Walter, J. M. Laramie, D. J. Gottlieb, and G. T. O'Connor. 2007. Framingham Heart Study genome-wide association: results for pulmonary function measures. *BMC.Med.Genet.* 8 Suppl 1:S8.
5. Wilk, J. B., T. H. Chen, D. J. Gottlieb, R. E. Walter, M. W. Nagle, B. J. Brandler, R. H. Myers, I. B. Borecki, E. K. Silverman, S. T. Weiss, et al. 2009. A genome-wide association study of pulmonary function measures in the Framingham Heart Study. *PLoS.Genet.* 5:e1000429.
6. Repapi, E., I. Sayers, L. V. Wain, P. R. Burton, T. Johnson, M. Obeidat, J. H. Zhao, A. Ramasamy, G. Zhai, V. Vitart, et al. 2010. Genome-wide association study identifies five loci associated with lung function. *Nat.Genet.* 42:36–44.
7. Hancock, D. B., M. Eijgelsheim, J. B. Wilk, S. A. Gharib, L. R. Loehr, K. D. Marciante, N. Franceschini, Y. M. van Durme, T. H. Chen, R. G. Barr, et al. 2010. Meta-analyses of genome-wide association studies identify multiple loci associated with pulmonary function. *Nat.Genet.* 42:45–52.
8. Soler, A. M., D. W. Loth, L. V. Wain, S. A. Gharib, M. Obeidat, W. Tang, G. Zhai, J. H. Zhao, A. V. Smith, J. E. Huffman, et al. 2011. Genome-wide association and large-scale follow up identifies 16 new loci influencing lung function. *Nat.Genet.* 43:1082–1090.
9. Imboden, M., E. Bouzigon, I. Curjuric, A. Ramasamy, A. Kumar, D. B. Hancock, J. B. Wilk, J. M. Vonk, G. A. Thun, V. Siroux, et al. 2012. Genome-wide association study of lung function decline in adults with and without asthma. *J.Allergy Clin.Immunol.* 129:1218–1228.
10. Hansel, N. N., I. Ruczinski, N. Rafaels, D. D. Sin, D. Daley, A. Malinina, L. Huang, A. Sandford, T. Murray, Y. Kim, et al. 2012. Genome-wide study identifies two loci associated with lung function decline in mild to moderate COPD. *Hum.Genet.*
11. van Iersel, C. A., H. J. de Koning, G. Draisma, W. P. Mali, E. T. Scholten, K. Nackaerts, M. Prokop, J. D. Habbema, M. Oudkerk, and R. J. van Klaveren. 2007. Risk-based selection from the general population in a screening trial: selection criteria, recruitment and power for the Dutch-Belgian randomised lung cancer multi-slice CT screening trial (NELSON). *Int.J.Cancer* 120:868–874.
12. van Diemen, C. C., D. S. Postma, J. M. Vonk, M. Bruinenberg, J. P. Schouten, and H. M. Boezen. 2005. A disintegrin and metalloprotease 33 polymorphisms and lung function decline in the general population. *Am.J.Respir.Crit Care Med.* 172:329–333.
13. Verschuren, W. M., A. Blokstra, H. S. Picavet, and H. A. Smit. 2008. Cohort profile: the Doetinchem Cohort Study. *Int.J.Epidemiol.* 37:1236–1241.
14. Marossy, A. E., D. P. Strachan, A. R. Rudnicka, and H. R. Anderson. 2007. Childhood chest illness and the rate of decline of adult lung function between ages 35 and 45 years. *Am.J.Respir.Crit Care Med.* 175:355–359.
15. Purcell, S., B. Neale, K. Todd-Brown, L. Thomas, M. A. Ferreira, D. Bender, J. Maller, P. Sklar, P. I. de Bakker, M. J. Daly, et al. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am.J.Hum.Genet.* 81:559–575.
16. R Development Core Team. R: A Language and Environment for Statistical Computing. 2009. R Foundation for Statistical Computing.
17. Hoshino, S., M. Yoshida, K. Inoue, Y. Yano, M. Yanagita, H. Mawatari, H. Yamane, T. Kijima, T. Kumagai, T. Osaki, et al. 2005. Cigarette smoke extract induces endothelial cell injury via JNK pathway. *Biochem.Biophys.Res.Commun.* 329:58–63.
18. Shu, W., H. Yang, L. Zhang, M. M. Lu, and E. E. Morrisey. 2001. Characterization of a new subfamily of winged-helix/forkhead (Fox) genes that are expressed in the lung and act as transcriptional repressors. *J.Biol.Chem.*

276:27488–27497.

19. Shu, W., M. M. Lu, Y. Zhang, P. W. Tucker, D. Zhou, and E. E. Morrisey. 2007. Foxp2 and Foxp1 cooperatively regulate lung and esophagus development. *Development* 134:1991–2000.

Supplementary methods

Study populations and genotyping

NELSON: the NELSON cohort was established as a lung cancer screening trial, in the beginning only males of Dutch descent smoking more than 20 pack-years were recruited. At later stage females were recruited as well, but they are a minority of the cohort. Individuals filled the questionnaires, underwent spirometry and CT scan of the chest in 2004 Blood was drawn to enable DNA (and in some cases RNA) extraction. The second survey took place three to four years later and same measurements were taken. Samples were genotyped using Illumina 610 Quad BeadChip, genotypes were called using a standard algorithm implemented in GenomeStudio software from Illumina.

Vlagtwedde/Vlaardingen cohort: is a Dutch population-based cohort including individuals from Vlagtwedde, a rural area, and Vlaardingen, an urban area in the Netherlands. The cohort was established in 1965 to study differences among others in lung function decline in relation to two types of environment. Participants were followed for over 25 years, with last surveys performed in 1989/1990, when lung function and other measurements taken every three years. Information on area of residence, age, sex, smoking habits and respiratory symptoms was collected by the Dutch version of the UK MRC standard questionnaire. Also blood and serum were collected for DNA extraction. Samples were genotyped using Illumina Veracode assay with custom content. Genotypes were called with standard algorithm implemented in GenomeStudio software from Illumina and genotype clusters were adjusted manually if necessary.

Doetinchem cohort: is a prospective part of the MORGEN (Monitoring of Risk Factors and Health in The Netherlands) study. FEV_1 was according to the European Respiratory Society (ERS) guidelines. Samples were genotyped using Illumina Veracode assay with custom content. Genotypes were called with standard algorithm implemented in GenomeStudio software from Illumina and genotype clusters were adjusted manually if necessary.

SAPALDIA: The SAPALDIA cohort, a multi-center study in eight geographic areas representing the range of environmental, meteorological and socio-demographic conditions in Switzerland. It was initiated in 1991 (SAPALDIA 1) with a follow-up assessment in 2002 (SAPALDIA 2) and 2011 (SAPALDIA3). This study has specifically been designed to investigate longitudinally lung function, respiratory and cardiovascular health; and to study and identify the associations of these health indicators with individual long term exposure to air pollution, other toxic inhalants, life style and molecular factors. Genome-wide genotyping was done on the Illumina Human 610quad BeadChip in the framework of the EU-funded GABRIEL study. More details can be found in Moffatt et al[E1].

B58C: The British 1958 birth cohort (B58C) is an ongoing follow-up of all persons born in Britain during one week in 1958. Lung function was measured at ages 35 and 45 years on a subsample of the cohort with a history of chest illness in childhood, and a control sample with no such history, as described in detail elsewhere [E2]. Genotypes were collated from data deposited by the Wellcome Trust Case-Control Consortium [E3], the Type 1 Diabetes Genetics Consortium [E4] and the GABRIEL asthma genetics consortium [E1]. Genotypes measured by the Illumina 550K v1 array (WTCCC), Illumina 550K v3 array (T1DGC) or Illumina 610K array (GABRIEL) were combined and imputed against the HapMap CEU release 21 reference haplotypes using MACH version 1.0.16.a.

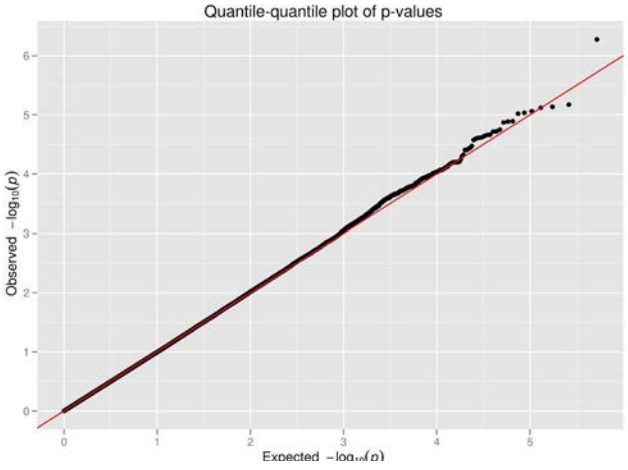


Figure E1. Q-Q plot. Observed p values are derived from linear regression analysis in lung function decline.

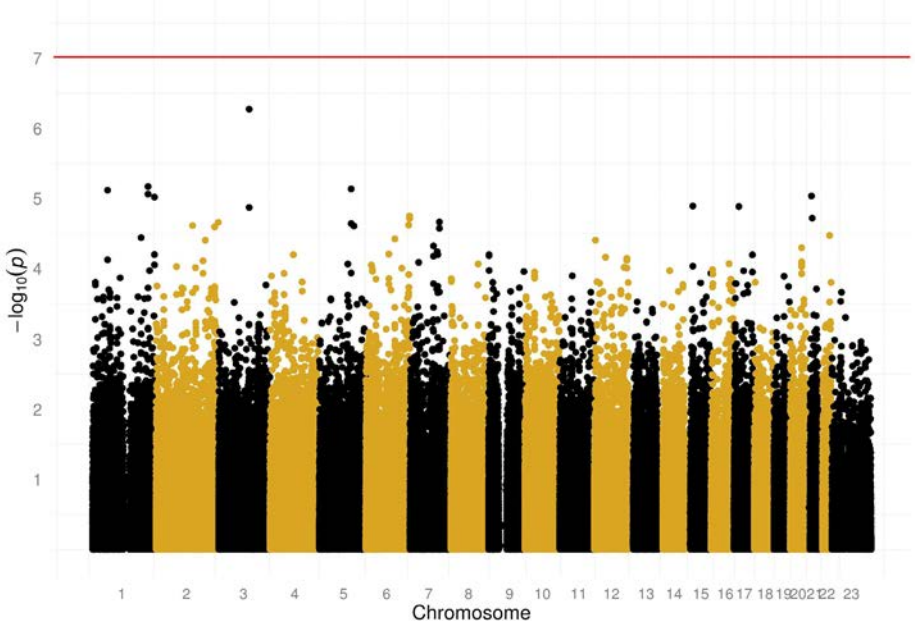


Figure E2. Manhattan plot. Distribution of FEV_1 decline association signals across the genome.

Supplementary Table 1. Beta estimates in Doetinchem: whole cohort, COPD, non-COPD and GxE test.

				Doetinchem all			Doetinchem COPD			Doetinchem non-COPD			GxE	
CHR	SNP	BP	A1	N	BETA	SE	N	BETA	SE	N	BETA	SE	Z_GXE	P_GXE
1	rs2492321	59124034	A	1093	2.66	1.91	251	9.73	5.21	833	0.91	2.09	-1.57	0.12
1	rs927743	59141535	A	1090	3.02	1.62	251	2.26	4.41	830	2.86	1.78	0.12	0.90
1	rs6604558	216001829	A	1093	2.40	1.90	251	2.89	4.99	833	2.69	2.12	-0.04	0.97
1	rs6694862	216033717	G	1093	2.24	1.90	251	2.19	4.99	833	2.69	2.12	0.09	0.93
1	rs11589174	240756312	C	1093	2.21	2.46	251	2.67	6.17	833	1.19	2.78	-0.22	0.83
1	rs12093170	240757607	C	1093	0.26	2.29	251	-0.23	5.89	833	0.41	2.57	0.10	0.92
1	rs6677308	240763170	A	1091	2.32	2.40	251	4.69	5.85	831	0.98	2.74	-0.57	0.57
2	rs892867	79218487	G	1093	-3.13	1.72	251	0.58	4.32	833	-4.23	1.92	-1.02	0.31
2	rs10928087	141458573	G	1092	0.65	2.84	250	12.58	6.84	833	-3.76	3.24	-2.16	0.03
2	rs4664511	152522710	G	1093	5.42	2.83	251	7.90	6.64	833	3.20	3.27	-0.63	0.53
2	rs3770111	182090561	A	1093	0.07	2.53	251	-4.71	6.59	833	1.37	2.81	0.85	0.40
2	rs2690751	190763511	A	1092	-2.60	1.89	251	-1.25	4.59	832	-3.41	2.15	-0.43	0.67
2	rs12467492	227127070	A	1092	-0.11	1.60	250	-2.46	4.23	833	0.60	1.76	0.67	0.50
3	rs1554328	118325495	A	1073	-4.34	2.46	248	-10.12	6.26	816	-3.46	2.77	0.97	0.33
3	rs938111	118330571	C	1092	-1.20	2.36	250	-5.04	5.77	833	-1.77	2.70	0.51	0.61
4	rs1394339	90492121	A	1093	-0.54	1.59	251	-3.21	4.10	833	0.02	1.78	0.72	0.47
5	rs1366331	110330169	A	1093	3.44	2.94	251	10.76	7.05	833	-2.53	3.36	-1.70	0.09
5	rs4836058	123375799	G	1093	0.18	1.62	251	5.62	4.36	833	-2.00	1.78	-1.62	0.11
5	rs1862404	123378142	G	1093	0.96	1.68	251	7.84	4.31	833	-1.31	1.87	-1.95	0.05
5	rs11949657	133897834	C	1093	-1.80	2.56	251	3.24	6.78	833	-5.34	2.82	-1.17	0.24
6	rs1498426	21116132	G	1092	-1.74	1.59	250	4.26	4.22	833	-2.58	1.75	-1.50	0.13
6	rs221749	165783026	G	1091	0.56	1.80	250	-4.54	4.84	832	1.85	1.98	1.22	0.22
6	rs17182135	169147239	G	1093	0.16	1.65	251	-4.96	4.34	833	0.19	1.82	1.09	0.27
6	rs12528430	169151607	A	1093	0.10	1.65	251	-4.79	4.38	833	0.19	1.82	1.05	0.29
7	rs4723295	33591734	G	1093	-3.88	2.40	251	-1.43	6.72	833	-4.19	2.61	-0.38	0.70
7	rs12667992	91183013	A	1093	2.44	2.64	251	-1.07	7.02	833	4.41	2.91	0.72	0.47
7	rs1916602	91199648	G	1093	1.96	2.53	251	-3.77	6.92	833	3.95	2.76	1.04	0.30
7	rs9791835	105324494	A	1092	-1.96	2.59	251	1.86	7.33	832	-4.41	2.81	-0.80	0.42
7	rs12705973	114100435	G	1092	-0.08	1.62	251	3.00	4.28	832	-1.99	1.79	-1.08	0.28
7	rs12671330	114121279	G	1092	-0.10	1.61	250	2.59	4.16	833	-2.00	1.78	-1.01	0.31
8	rs2458416	105420643	G	1092	0.39	2.69	250	2.63	7.07	833	0.10	2.98	-0.33	0.74
9	rs10962687	1690272	G	1093	0.36	3.52	251	3.59	8.19	833	-0.43	4.11	-0.44	0.66
9	rs10962689	1690460	G	1093	0.64	3.51	251	3.59	8.19	833	0.20	4.07	-0.37	0.71
12	rs4639992	43209601	G	1093	-3.32	2.30	251	-2.55	5.71	833	-3.76	2.61	-0.19	0.85
12	rs16946677	115386413	C	1093	0.92	2.48	251	4.61	6.54	833	-3.27	2.72	-1.11	0.27
17	rs2258367	18796532	C	1093	-2.51	2.52	251	-9.09	6.20	833	-1.49	2.84	1.12	0.26
20	rs6073768	43563560	G	1091	2.43	2.14	251	7.27	5.45	832	0.53	2.40	-1.13	0.26
20	rs1013562	43744158	G	1091	2.72	2.14	251	4.72	5.54	831	1.54	2.38	-0.53	0.60
20	rs6073862	43765119	G	1092	3.12	2.14	251	5.25	5.43	832	1.86	2.39	-0.57	0.57
21	rs13049780	19638152	G	1093	0.41	2.68	251	-2.29	6.77	833	3.02	3.01	0.72	0.47
21	rs2827251	22429921	A	1093	1.20	3.04	251	-1.92	9.23	833	1.35	3.24	0.33	0.74

BETA > 0 annual FEV_1 decline acceleration
BETA > 0 annual FEV_1 decline attenuation.

Supplementary Table 2. Beta estimates in Doetinchem: whole cohort, >20 pack-years, <20 pack-years and GxE test

				Doetinchem all			Doetinchem >20py			Doetinchem <20py			GxE	
CHR	SNP	BP	A1	N	BETA	SE	N	BETA	SE	N	BETA	SE	Z_GXE	P_GXE
1	rs2492321	59124034	A	1093	2.66	1.91	286	15.96	4.54	807	-0.87	2.16	-3.35	8.09^{-04}
1	rs927743	59141535	A	1090	3.02	1.62	286	5.26	3.69	804	2.01	1.87	-0.78	0.43
1	rs6604558	216001829	A	1093	2.40	1.90	286	-1.11	4.13	807	3.81	2.23	1.05	0.30
1	rs6694862	216033717	G	1093	2.24	1.90	286	-1.18	4.09	807	3.58	2.24	1.02	0.31
1	rs11589174	240756312	C	1093	2.21	2.46	286	6.57	5.56	807	0.14	2.84	-1.03	0.30
1	rs12093170	240757607	C	1093	0.26	2.29	286	6.70	5.10	807	-2.14	2.67	-1.54	0.12
1	rs6677308	240763170	A	1091	2.32	2.40	286	8.28	5.24	805	-0.24	2.81	-1.43	0.15
2	rs892867	79218487	G	1093	-3.13	1.72	286	-2.61	3.99	807	-2.85	1.96	-0.05	0.96
2	rs10928087	141458573	G	1092	0.65	2.84	285	10.05	5.75	807	-3.97	3.42	-2.09	0.04
2	rs4664511	152522710	G	1093	5.42	2.83	286	13.58	6.11	807	1.78	3.32	-1.70	0.09
2	rs3770111	182090561	A	1093	0.07	2.53	286	4.05	5.57	807	-1.96	2.95	-0.95	0.34
2	rs2690751	190763511	A	1092	-2.60	1.89	286	-8.39	4.31	806	-0.40	2.18	1.66	0.10
2	rs12467492	227127070	A	1092	-0.11	1.60	285	-5.00	3.62	807	1.64	1.84	1.64	0.10
3	rs1554328	118325495	A	1073	-4.34	2.46	282	-5.35	5.20	791	-4.40	2.92	0.16	0.87
3	rs938111	118330571	C	1092	-1.20	2.36	285	0.84	4.99	807	-3.14	2.80	-0.70	0.49
4	rs1394339	90492121	A	1093	-0.54	1.59	286	-4.14	3.38	807	0.73	1.88	1.26	0.21
5	rs1366331	110330169	A	1093	3.44	2.94	286	8.43	7.45	807	0.45	3.29	-0.98	0.33
5	rs4836058	123375799	G	1093	0.18	1.62	286	1.04	3.59	807	-0.22	1.89	-0.31	0.76
5	rs1862404	123378142	G	1093	0.96	1.68	286	2.83	3.65	807	0.02	1.97	-0.68	0.50
5	rs11949657	133897834	C	1093	-1.80	2.56	286	-1.10	6.18	807	-3.37	2.89	-0.33	0.74
6	rs1498426	21116132	G	1092	-1.74	1.59	285	2.70	3.53	807	-2.50	1.85	-1.31	0.19
6	rs221749	165783026	G	1091	0.56	1.80	284	3.14	4.17	807	-0.18	2.06	-0.71	0.48
6	rs17182135	169147239	G	1093	0.16	1.65	286	1.64	3.62	807	-1.68	1.92	-0.81	0.42
6	rs12528430	169151607	A	1093	0.10	1.65	286	1.75	3.65	807	-1.68	1.92	-0.83	0.41
7	rs4723295	33591734	G	1093	-3.88	2.40	286	6.05	5.27	807	-8.50	2.78	-2.44	0.01
7	rs12667992	91183013	A	1093	2.44	2.64	286	2.59	5.97	807	2.59	3.05	0.00	1.00
7	rs1916602	91199648	G	1093	1.96	2.53	286	1.95	5.71	807	1.72	2.92	-0.04	0.97
7	rs9791835	105324494	A	1092	-1.96	2.59	285	-2.26	5.69	807	-2.95	3.01	-0.11	0.92
7	rs12705973	114100435	G	1092	-0.08	1.62	286	7.87	3.55	806	-3.56	1.89	-2.84	4.47^{-03}
7	rs12671330	114121279	G	1092	-0.10	1.61	285	8.39	3.50	807	-3.74	1.87	-3.05	2.26^{-03}
8	rs2458416	105420643	G	1092	0.39	2.69	286	-0.53	5.93	806	0.52	3.15	0.16	0.88
9	rs10962687	1690272	G	1093	0.36	3.52	286	2.44	6.89	807	-0.75	4.33	-0.39	0.69
9	rs10962689	1690460	G	1093	0.64	3.51	286	3.14	6.82	807	-0.54	4.33	-0.45	0.65
12	rs4639992	43209601	G	1093	-3.32	2.30	286	-5.98	4.92	807	-1.96	2.72	0.71	0.48
12	rs16946677	115386413	C	1093	0.92	2.48	286	-2.72	6.44	807	0.15	2.74	0.41	0.68
17	rs2258367	18796532	C	1093	-2.51	2.52	286	0.44	6.17	807	-3.36	2.83	-0.56	0.58
20	rs6073768	43563560	G	1091	2.43	2.14	286	-2.65	4.94	805	4.16	2.46	1.24	0.22
20	rs1013562	43744158	G	1091	2.72	2.14	286	-5.39	5.03	805	5.25	2.44	1.90	0.06
20	rs6073862	43765119	G	1092	3.12	2.14	286	-4.69	4.94	806	5.71	2.45	1.89	0.06
21	rs13049780	19638152	G	1093	0.41	2.68	286	0.79	5.20	807	0.59	3.33	-0.03	0.97
21	rs2827251	22429921	A	1093	1.20	3.04	286	7.96	6.84	807	-2.18	3.51	-1.32	0.19

BETA > 0 annual FEV_1 decline acceleration
BETA > 0 annual FEV_1 decline attenuation.

Supplementary table 3. Beta estimates in Vlagtwedde/Vlaardingen: whole cohort, >20 pack-years, <20 pack-years and GxE test

				Vlagtwedde/ Vlaardingen all			Vlagtwedde/ Vlaardingen >20py			Vlagtwedde/ Vlaardingen <20py			GxE	
CHR	SNP	BP	A1	NMISS	BETA	SE	NMISS	BETA	SE	NMISS	BETA	SE	Z_GXE	P_GXE
1	rs2492321	59124034	A	1056	-1.10	0.98	250	0.62	2.03	806	-2.02	1.39	-1.07	0.28
1	rs927743	59141535	A	1053	-0.97	0.82	248	0.49	1.87	805	-1.11	1.12	-0.73	0.46
1	rs920307	188985655	A	1044	0.20	1.05	247	-2.82	2.18	797	1.02	1.49	1.46	0.15
1	rs6604558	216001829	A	1056	-0.03	0.95	250	0.37	2.11	806	-1.85	1.32	-0.89	0.37
1	rs6694862	216033717	G	1056	0.09	0.95	250	0.38	2.10	806	-1.61	1.32	-0.80	0.42
1	rs11589174	240756312	C	1049	-1.03	1.15	249	-1.55	2.31	800	-1.65	1.64	-0.03	0.97
1	rs12093170	240757607	C	1055	-0.50	1.08	250	1.83	2.20	805	-1.46	1.55	-1.22	0.22
1	rs6677308	240763170	A	1055	-1.54	1.16	250	-0.87	2.34	805	-1.73	1.67	-0.30	0.77
2	rs892867	79218487	G	1055	-0.56	0.84	250	-0.76	1.91	805	0.24	1.15	0.45	0.65
2	rs10928087	141458573	G	1056	0.20	1.51	250	-0.50	3.44	806	-0.96	2.08	-0.11	0.91
2	rs4664511	152522710	G	1056	1.22	1.36	250	2.10	3.05	806	0.46	1.87	-0.46	0.65
2	rs3770111	182090561	A	1056	-1.15	1.28	250	-2.98	2.52	806	2.54	1.85	1.77	0.08
2	rs2690751	190763511	A	1055	-0.56	0.96	250	0.85	2.19	805	-0.66	1.32	-0.59	0.55
2	rs12467492	227127070	A	1056	-1.43	0.77	250	0.73	1.65	806	-2.04	1.08	-1.40	0.16
3	rs1554328	118325495	A	1048	-0.05	1.23	249	-1.23	3.22	799	1.87	1.64	0.85	0.39
3	rs938111	118330571	C	1056	-0.09	1.18	250	0.85	2.87	806	0.16	1.60	-0.21	0.83
4	rs1394339	90492121	A	1056	-0.42	0.79	250	-1.02	1.70	806	-0.16	1.10	0.42	0.67
5	rs1366331	110330169	A	1056	0.18	1.46	250	1.35	3.43	806	0.82	1.99	-0.13	0.89
5	rs4836058	123375799	G	1056	2.13	0.80	250	3.16	1.86	806	1.34	1.10	-0.85	0.40
5	rs1862404	123378142	G	1056	2.63	0.84	250	4.13	1.96	806	1.27	1.15	-1.26	0.21
5	rs11949657	133897834	C	1056	2.49	1.18	250	2.10	2.71	806	4.01	1.62	0.60	0.55
6	rs1498426	21116132	G	1055	1.22	0.80	250	-0.26	1.71	805	1.01	1.12	0.62	0.53
6	rs221749	165783026	G	1055	-0.75	0.91	249	0.75	1.99	806	-0.94	1.27	-0.72	0.47
6	rs17182135	169147239	G	1055	-1.23	0.80	250	0.29	1.70	805	-0.95	1.13	-0.61	0.54
6	rs12528430	169151607	A	1055	-1.32	0.80	249	-0.18	1.68	806	-0.76	1.12	-0.29	0.77
7	rs4723295	33591734	G	1056	-0.42	1.18	250	-5.25	2.39	806	2.06	1.68	2.51	0.01
7	rs12667992	91183013	A	1056	2.03	1.43	250	2.85	3.05	806	0.57	2.02	-0.62	0.53
7	rs1916602	91199648	G	1056	1.28	1.31	250	3.25	2.80	806	-1.41	1.83	-1.40	0.16
7	rs9791835	105324494	A	1055	1.60	1.28	250	-0.13	2.82	805	2.53	1.77	0.80	0.42
7	rs12705973	114100435	G	1056	-1.04	0.80	250	-0.79	1.80	806	1.41	1.10	1.05	0.30
7	rs12671330	114121279	G	1054	-0.83	0.79	250	0.73	1.77	804	-1.51	1.09	-1.08	0.28
8	rs2458416	105420643	G	1056	1.53	1.32	250	-4.18	2.99	806	2.94	1.82	2.04	0.04
9	rs10962687	1690272	G	1056	-1.26	1.59	250	-3.25	4.78	806	0.24	2.08	0.67	0.50
9	rs10962689	1690460	G	1056	-1.16	1.55	250	-2.79	4.65	806	0.35	2.02	0.62	0.54
12	rs4639992	43209601	G	1056	1.31	1.13	250	-1.59	2.60	806	2.03	1.54	1.20	0.23
12	rs16946677	115386413	C	1056	-1.10	1.18	250	2.23	2.92	806	-1.05	1.58	-0.99	0.32
15	rs7182547	29084964	A	1051	0.99	0.86	249	1.92	1.79	802	0.00	1.22	-0.89	0.37
17	rs2258367	18796532	C	1045	0.60	1.28	248	-2.88	2.79	797	0.65	1.79	1.06	0.29
20	rs6073768	43563560	G	1055	-0.65	1.00	250	-2.95	2.27	805	-0.86	1.37	0.79	0.43
20	rs1013562	43744158	G	1048	-1.28	0.99	247	-4.12	2.19	801	-1.04	1.37	1.19	0.23
20	rs6073862	43765119	G	1056	-1.09	0.97	250	-3.65	2.13	806	-1.09	1.36	1.02	0.31
21	rs13049780	19638152	G	1055	0.05	1.35	250	0.08	2.91	805	1.58	1.89	0.43	0.66
21	rs2827251	22429921	A	1055	0.85	1.54	250	3.39	3.52	805	1.18	2.12	-0.54	0.59

BETA > 0 annual FEV_1 decline acceleration
BETA > 0 annual FEV_1 decline attenuation.

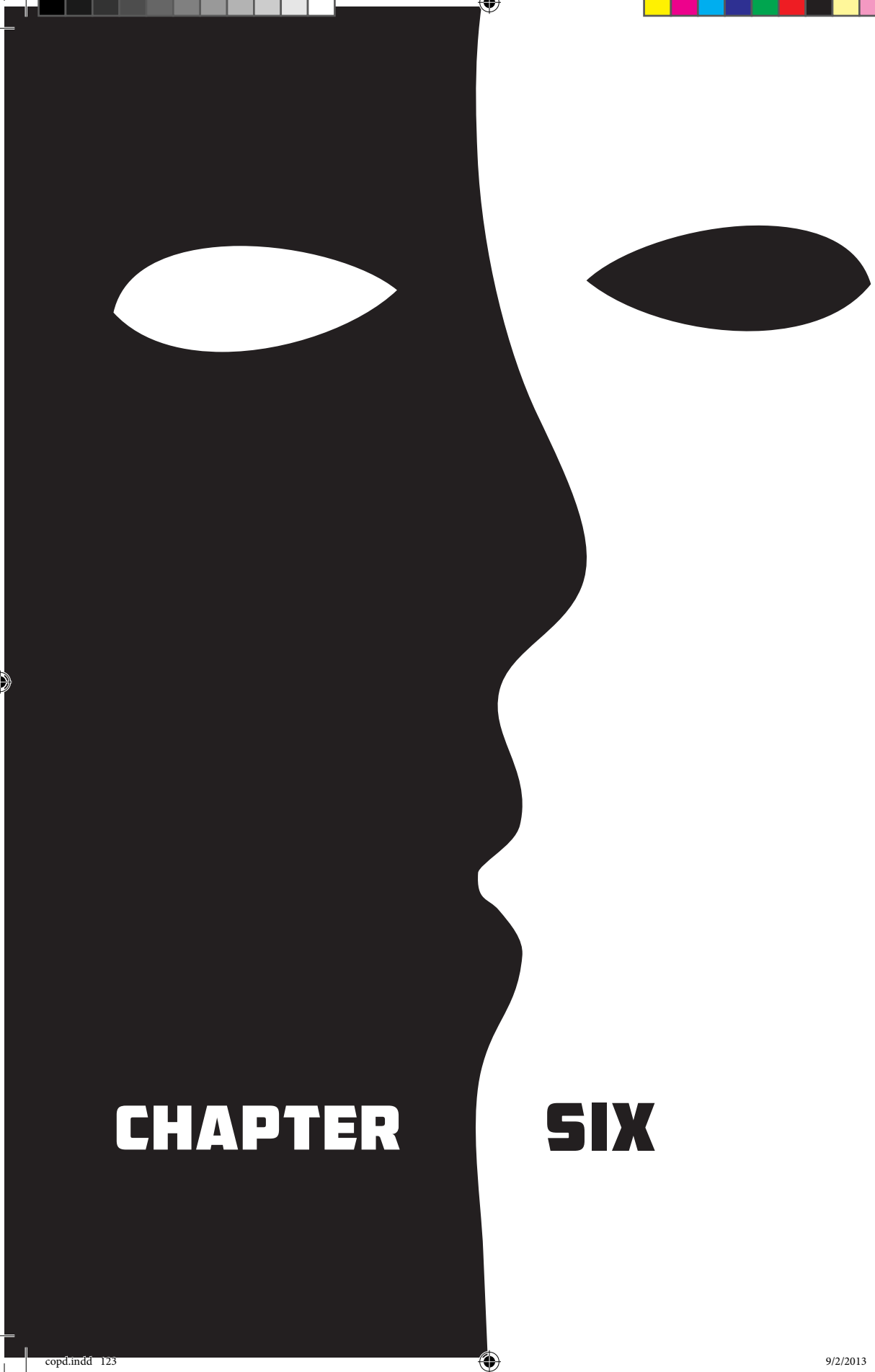
References

E1. Moffatt MF, Gut IG, Demenais F et al. A large-scale, consortium-based genomewide association study of asthma. N Engl J Med 2010;363:1211–1221.

E2. Marossy AE, Strachan DP, Rudnicka AR, Anderson HR. Childhood chest illness and the rate of decline of adult lung function between ages 35 and 45 years. Am J Respir Crit Care Med 2007;175:355–359.

E3. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 2007;447:661–678.

E4. Barrett JC, Clayton DG, Concannon P et al. Genome-wide association study and meta-analysis find that over 40 loci affect risk of type 1 diabetes. Nat Genet 2009;41:703–707.



CHAPTER SIX

Genome-wide epistatic analysis in COPD

Joanna Smolonska[1,2], Dirkje S Postma[3,4], Pieter Zanen[5], Morris Swertz[1], Judith M Vonk[2,4], Eleonora AM Festen[1,6], Pieter Neerincx[1], Willem Mali[7], Matthijs Oudkerk[8], Harry de Koning[9], The LifeLines Cohort Study, Harry JM Groen[3], Cisca Wijmenga[1*], H Marike Boezen[2,4*#]

*authors contributed equally

[1] Department of Genetics, University of Groningen, UMC Groningen, Groningen, the Netherlands
[2] Department of Epidemiology, University of Groningen, UMC Groningen, Groningen, the Netherlands
[3] Department of Pulmonology, University of Groningen, UMC Groningen, Groningen, the Netherlands
[4] GRIAC Research Institute, University of Groningen, UMC Groningen, Groningen, the Netherlands
[5] Department of Pulmonology, UMC Utrecht, Utrecht, the Netherlands
[6] Department of Gastroenterology, University of Groningen, UMC Groningen, Groningen, the Netherlands
[7] Department of Radiology, UMC Utrecht, Utrecht, the Netherlands
[8] Department of Radiology, UMC Groningen, Groningen, the Netherlands
[9] Department of Public Health, Erasmus MC, Rotterdam, the Netherlands

Corresponding author: H Marike Boezen: h.m.boezen@umcg.nl

Keywords: epistasis, COPD, genome-wide, interaction

Abstract

Chronic obstructive pulmonary disease (COPD) is a complex disorder involving multiple genes. However, despite genome-wide association studies (GWAS) identifying many disease genes, we know a significant proportion of a trait's heritability remains unexplained. We therefore performed the first genome-wide SNP-SNP interaction (epistasis) analysis in COPD, as we suspect epistasis holds the key to at least part of the missing heritability. We used the NELSON cohort as a discovery set, comprising 1,030 cases and 1,808 controls, and the LifeLines cohort as a replication set, with 711 cases and 3,013 controls. Our results show SNP-SNP interactions suggesting that cell adhesion and cell communication pathways could be involved in COPD development. SNPs involved in these interactions are annotated to genes that include *SERPINB9, CNTN5, PTPRM* and *SIGLEC1,* and which, except for *CNTN5,* were not previously associated with COPD. Our findings open up new avenues towards understanding the course of events leading to COPD, but they should be considered with caution, as genome-wide significance was not reached. Further replication is needed to confirm our findings.

Chronic obstructive pulmonary disease (COPD, MIM 606963) is a respiratory disease caused by both genetic and environmental factors, including cigarette smoking and air pollution. The role of genetics was first seen in studies on families with early-onset COPD[1]. Linkage studies identified a locus on chromosome 2q that mapped to the *SERPINE2* (MIM 177010) gene[2]. *SERPINA1* (MIM 107400) is known to predispose to early onset emphysema, one of the phenotypes in COPD[3]. Candidate gene studies have focused mostly on three pathways that are involved in COPD: the inflammatory, anti-oxidant and protease-anti-protease pathways. They have identified the following genes: *TNFA* (MIM 191160), *TGFB1* (MIM 190180), *GSTP1* (MIM 134660) and *MMP9* (MIM 120361), and others, as associated with COPD (reviewed in Smolonska *et al*[4]). However, these associations have not yet been confirmed by genome-wide association studies (GWAS). GWAS have identified *CHRNA3/5* (MIM 118503, 118505), *HHIP* (MIM 606178)[5], *FAM13A* (MIM 613299)[6] and *BICD1* (MIM 602204)[7] as loci associated with COPD, the last with emphysema in particular. Since COPD is a complex disorder, these variants are unlikely to explain all the "missing heritability" for COPD. However, gene-gene interactions (epistasis) have been proposed as a potential mechanism that might explain a large part of this missing heritability[8]; their contribution to COPD development has not yet been investigated. We aimed to perform the first genome-wide epistatic analysis, using 1,030 COPD cases and 1,808 controls (Patients characteristics in Table S1).

Samples were genotyped on the Illumina 610Quad BeadChip and the data was subjected to detailed quality control (Supplemental Material) that left 492,339 SNPs for the analysis. We used EpiBLASTER[9], software enabling a full genome epistasis scan in a time-efficient manner. EpiBLASTER follows a two-stage approach. In the first stage, the difference in Pearson's correlation coefficients between cases and controls for all possible SNP pairs are computed and the significance of this difference is indicated. Using an arbitrary $p < 1*10^{-6}$ as a cut-off for the first stage, we ended up with 120,425 unique SNP pairs exceeding this threshold. In the second stage, full rank logistic regression was performed on the top 1,000 SNP pairs (arbitrary number) from the first stage to obtain the effect estimate and p-value for the interaction term. Next, we performed likelihood ratio tests to test if the effect estimates from logistic regression were significantly different from zero. To confirm our findings we performed a replication analysis of the same 1,000 SNP pairs using 711 COPD cases and 3,013 controls from the LifeLines cohort[10]. This cohort was genotyped using the Illumina Human CytoSNP. Imputation was performed with BEAGLE 3.0 software[11] using HapMap 2 CEU population as a reference panel and, after imputation, SNPs with $r^2 < 0.3$ were removed, leaving 764 SNP pairs of the top 1,000 SNPs for analysis. Fifteen SNP pairs were replicated with $p < 0.05$ (Table 1). P-values were combined using Fisher's method. None of the SNP pairs exceeded the genome-wide significance threshold: $4.12*10^{-13}$ ($0.05/492{,}339^2/2$). However, taking into account that: (1) SNPs included in the epistatic analysis

were not pruned for linkage disequilibrium (LD) beforehand, hence the tests were not entirely independent, and (2) EpiBLASTER uses a two-stage approach, in which the first is a filtration step preventing some 120 billion tests from being performed, it did not seem justified to use a strict Bonferroni correction. We therefore applied the false discovery rate (FDR) method to account for multiple testing in the replication phase, but none of the 764 SNP pairs showed significant epistatic effect. This might be due to (1) a power issue, i.e. a much larger sample size is required to replicate interactions, rather than traditional genotypic associations, or (2) because the identified SNP pairs are false-positives. We should also highlight the fact that the replication cohort had imputed genotypes, which can give good approximations for single polymorphism analysis, but in epistatic analysis, where two genotypes are taken together, can give a much higher deviation from the actual distribution. The minor allele frequency (MAF) may well be under- or over-estimated. The differences can thus seem insignificant, yet the sum of frequencies of genotypes involved in epistatic effect differed by nearly 3% (F_{epi}, Supplemental Material). We should therefore interpret both positive and negative findings with caution when using imputed genotypes for epistatic analysis.

Since we were interested whether epistasis could identify plausible pathways, we performed annotations for the 30 SNPs with $p < 0.05$ for interaction in the replication phase using SCANdb[12] (Table 1). Ten SNP pairs could not be translated into gene-gene interactions, because at least one of the SNPs was not annotated to a gene. Genes presented in Table 1 were used for gene enrichment analysis in GeneMANIA[13] (Figure 1) and annotation with GATHER[14]. We observed a clear over-representation of genes involved in cell adhesion (GO:0007155) and cell communication (GO:0007154) pathways ($p < 0.0001$ for both). Cell adhesion plays a crucial role in epithelial barrier repair after tissue injury, abundantly present in COPD.

One of the interesting gene pairs is *SERPINB9* (MIM 601799) and *CNTN5* (MIM 607219). *SERPINB9* belongs to the serine inhibitor protein family, with two members already associated with COPD (*SERPINA1* and *SERPINE2*). Furthermore, it specifically inhibits granzyme B[15], a cytolytic enzyme released by lymphocytes during the inflammatory response initiating target cell apoptosis[16].

Another interesting pair of genes is *PTPRM* (MIM 176888) and *SIGLEC1* (MIM 600751), both with cell adhesion-related functions. PTPRM (protein tyrosine phosphatase receptor type M) levels have been shown to be reduced after injury of epithelial cells. Lower expression of PTPRM results in reduced adhesion of epithelial cells[17]. *SIGLEC1* (sialoadhesin, sialic acid binding Ig-like lectin 1) is one of the few receptors that are expressed solely on the monocyte and macrophage surfaces[18]. Macrophages are key players in the development of COPD and are recruited to the lung upon cigarette smoke inhalation. Interestingly, the ligands of sialoadhesin are mainly expressed on neutrophils[19], which also play an important role in the pathogenesis of COPD.

Table 1. Top 15 SNP pairs identified and replicated in the current COPD epistasis study.

								EpiBLASTER stage1			EpiBLASTER stage 2			Replication		Combined
CHR1	BP1	SNP1	Gene1	CHR2	BP2	SNP2	gene2	PC_diff	Z	P	BETA	SE	P_{int}	BETA	P_{rep}	P_{comb}
1	56686929	rs12736058	---	12	27653244	rs11049068	PPFIBP1	-0.23	-5.79	7.02E-09	-0.6	0.11	4.96E-09	-0.25	0.04186	4.84E-09
1	161459203	rs10917702	---	21	42433102	rs220179	UMODL1	-0.23	-5.9	3.59E-09	-0.67	0.12	2.70E-09	-0.36	0.01916	1.28E-09
1	163453368	rs12068601	LMX1A	20	8381514	rs2221695	PLCB1	0.23	5.87	4.34E-09	0.88	0.16	8.75E-09	0.35	0.04355	8.65E-09
3	70529586	rs9825323	---	10	115306802	rs4918842	HABP2	0.23	5.83	5.53E-09	0.74	0.13	3.46E-09	0.56	0.002452	2.25E-10
3	113114254	rs1282948	PHLDB2	9	116488534	rs7853853	---	-0.23	-5.81	6.38E-09	-0.66	0.11	2.74E-09	-0.28	0.01268	8.71E-10
4	7711046	rs11737898	SORCS2	12	71762662	rs11179538	---	-0.23	-5.83	5.44E-09	-0.68	0.12	5.06E-09	-0.25	0.03555	4.22E-09
4	91446704	rs1037926	FAM190A	7	32362185	rs412876	---	0.22	5.76	8.51E-09	0.47	0.08	2.18E-08	0.2	0.02451	1.19E-08
5	32107024	rs17512228	PDZD2	12	70554471	rs12580478	TBC1D15	0.23	5.88	4.22E-09	-0.67	0.12	8.94E-08	-1.24	0.04562	8.29E-08
6	2838007	rs318489	SERPINB9	11	98930610	rs1494470	CNTN5	-0.23	-5.78	7.65E-09	-0.51	0.09	6.03E-09	-0.22	0.01997	2.87E-09
6	2838007	rs318489	SERPINB9	11	98932301	rs2047166	CNTN5	-0.23	-5.78	7.65E-09	-0.51	0.09	6.03E-09	-0.22	0.02059	2.96E-09
7	25572040	rs2952522	---	20	8861264	rs7270760	---	-0.23	-5.88	4.11E-09	0.53	0.09	5.47E-09	0.18	0.04807	6.06E-09
7	35391794	rs10257155	---	10	25768499	rs11014558	GPR158	-0.23	-5.75	8.73E-09	0.6	0.1	9.26E-09	0.63	0.002265	5.37E-10
8	88868935	rs7813504	---	9	14840256	rs16932354	FREM1	0.23	5.94	2.87E-09	0.99	0.17	3.87E-09	1.21	0.004831	4.81E-10
16	13242472	rs2042416	---	22	37817611	rs139272	RP4-742C19.12	0.23	5.78	7.33E-09	0.63	0.11	2.36E-08	0.35	0.003111	1.79E-09
18	7761132	rs4798593	PTPRM	20	3623333	rs3746638	SIGLEC1	-0.23	-5.92	3.13E-09	-0.49	0.08	2.45E-09	-0.19	0.04288	2.52E-09

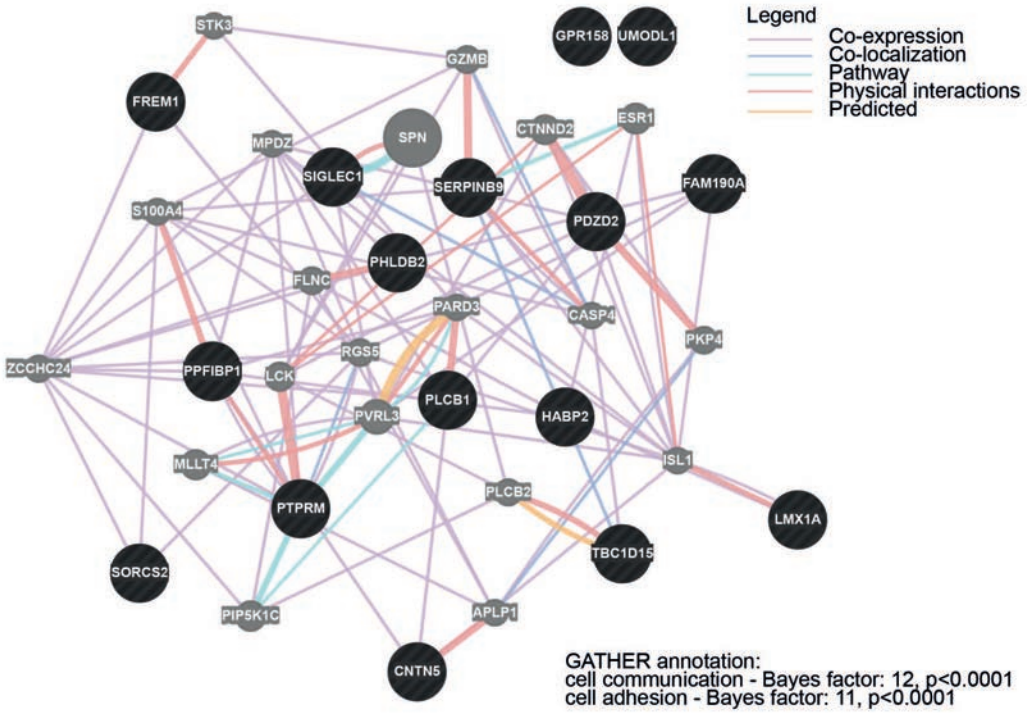


Figure 1. Gene enrichment in GeneMANIA based on top gene–gene interactions.
Black circles: query genes (from Table 2); Grey circles: genes enriched in GeneMANIA

We further investigated the regulatory potential of SNPs not in the direct proximity of a gene, using HaploReg[20] (Table 2). Six out of 11 SNPs showed regulatory potential at least in one ENCODE[21] experiment type: three SNPs cause an alteration in transcription factor binding sites, two SNPs are located in DNAse I footprints (which correlate with open-state chromatin and transcription factor binding sites) and one SNP is situated within an enhancer. One of the interesting findings is that rs2042416 alters the Foxo motif, where Forkhead box O (FOXO) transcription factors bind. FOXO transcription factors are crucial to all aspects of T-cell biology, immune response and oxidative stress[22] and FOXO3 levels have been shown to be reduced in the lungs of smokers and COPD patients[23].

Table 2. Regulatory potential of the SNPs not in the direct proximity of the gene.

Chr	pos (hg19)	variant	Ref	Al	CEU freq	Enhancer histone marks	DNAse	Proteins bound	Motifs changed	GENCODE genes	RefSeq genes
7	35425269	rs10257155	G	T	0.1					110kb 3' of AC007652.1	9.2kb 5' of LOC401324
1	163192579	rs10917702	G	A	0.6				NF-I	RP11-267N12.3	20kb 3' of RGS5
12	73476395	rs11179538	T	G	0.3					292kb 5' of AC131213.1	417kb 5' of TRHDE
1	56914341	rs12736058	T	C	0.5	HepG2				33kb 3' of RP4-710M16.2	46kb 3' of PPAP2B
16	13334971	rs2042416	G	A	0.2				Foxo	5.4kb 3' of SHISA9	697bp 5' of SHISA9
7	25605515	rs2952522	T	C	0.4		A549, AoSMC, CLL, Fibrobl, HeLa-S3, HTR8svn, T-47D, Urothelia	CTCF		4.8kb 5' of AC091705.1	337kb 3' of NPVF
7	32395660	rs412876	G	T	0.6					57kb 5' of PDE1C	57kb 3' of PDE1C
20	8913264	rs7270760	C	T	0.3		FibroP			PLCB1	48kb 5' of PLCB1
8	88799819	rs7813504	T	G	0.1					32kb 5' of AF121898.3	83kb 3' of DCAF4L2
9	117448713	rs7853853	G	A	0.1					15kb 5' of RP11-402G3.3	4.3kb 3' of LOC100505478
3	70446896	rs9825323	A	C	0.6				Hoxa5, CHX10, Dlx2, Bsx	146kb 5' of AC134025.1	429kb 5' of MITF

In summary, we describe the first complete, genome-wide epistatic analysis in COPD, without pre-selection of markers. We tested for SNP-SNP interactions and performed replication in an independent cohort. We replicated 15 SNP pairs associated with COPD ($p < 0.05$) but none of the SNPs involved in these 15 interactions was even marginally associated with COPD on its own(data not shown). This is in line with the definition of epistasis. Annotation of these SNPs indicates cell adhesion and cell communication pathways, both plausible in COPD pathogenesis. The overall picture of COPD pathogenesis emerging from the genes involved in interactions, like *SERPINB9, CNTN5, PTPRM* and *SIGLEC,* opens new avenues for further research and may be crucial for understanding the course of events leading to clinical expression of COPD. Our results must be interpreted with caution because none of the SNP-pairs reached

genome-wide significance threshold, probably due to the limited replication sample size and use of imputed genotypes. Our findings should be replicated in a large COPD cohort with non-imputed, genome-wide genotype data.

Acknowledgments

We thank Dr Tony Kam-Thong for assistance in running EpiBLASTER. This work was funded by the EU project, COPACETIC: COPD, Addressing Critical Gaps, Early Treatment and Diagnosis and Innovative Concepts, Grant Agreement No. 201379. We thank participants and researchers involved in sample and data collection in the NELSON and LifeLines cohorts. EAMF is supported by a Mandema Stipend 78220 UMCG, University of Groningen. We thank Jackie Senior for editing the final manuscript.

Web Resources

The URLs for data presented are as follows:

Online Mendelian Inheritance in Man (OMIM), www.omim.org

GeneNetwork, www.genenetwork.nl

GeneMANIA, http://www.genemania.org/

GATHER, http://gather.genome.duke.edu/

References

1. Silverman EK, Mosley JD, Palmer LJ et al. Genome-wide linkage analysis of severe, early-onset chronic obstructive pulmonary disease: airflow obstruction and chronic bronchitis phenotypes. Hum Mol Genet 2002;**11**:623-632.
2. DeMeo DL, Mariani TJ, Lange C et al. The SERPINE2 gene is associated with chronic obstructive pulmonary disease. Am J Hum Genet 2006;**78**:253-264.
3. Lieberman J, Winter B, Sastre A. Alpha 1-antitrypsin Pi-types in 965 COPD patients. Chest 1986;**89**:370-373.
4. Smolonska J, Wijmenga C, Postma DS, Boezen HM. Meta-analyses on suspected chronic obstructive pulmonary disease genes: a summary of 20 years' research. Am J Respir Crit Care Med 2009;**180**:618-631.
5. Pillai SG, Ge D, Zhu G et al. A genome-wide association study in chronic obstructive pulmonary disease (COPD): identification of two major susceptibility loci. PLoS Genet 2009;**5**:e1000421.
6. Cho MH, Boutaoui N, Klanderman BJ et al. Variants in FAM13A are associated with chronic obstructive pulmonary disease. Nat Genet 2010;**42**:200-202.
7. Kong X, Cho MH, Anderson W et al. Genome-wide association study identifies BICD1 as a susceptibility gene for emphysema. Am J Respir Crit Care Med 2011;**183**:43-49.
8. Maher B. Personal genomes: The case of the missing heritability. Nature 2008;**456**:18-21.
9. Kam-Thong T, Czamara D, Tsuda K et al. EPIBLASTER-fast exhaustive two-locus epistasis detection strategy using graphical processing units. Eur J Hum Genet 2011;**19**:465-471.
10. Stolk RP, Rosmalen JG, Postma DS et al. Universal risk factors for multifactorial diseases: LifeLines: a three-generation population-based study. Eur J Epidemiol 2008;**23**:67-74.
11. Browning BL, Browning SR. A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. Am J Hum Genet 2009;**84**:210-223.
12. Gamazon ER, Zhang W, Konkashbaev A et al. SCAN: SNP and copy number annotation. Bioinformatics 2010;**26**:259-262.
13. Mostafavi S, Ray D, Warde-Farley D, Grouios C, Morris Q. GeneMANIA: a real-time multiple association network integration algorithm for predicting gene function. Genome Biol 2008;**9 Suppl 1**:S4.
14. Chang JT, Nevins JR. GATHER: a systems approach to interpreting genomic signatures. Bioinformatics 2006;**22**:2926-2933.
15. Sun J, Bird CH, Sutton V et al. A cytosolic granzyme B inhibitor related to the viral apoptotic regulator cytokine response modifier A is present in cytotoxic lymphocytes. J Biol Chem 1996;**271**:27802-27809.
16. Hatcher VB, Oberman MS, Lazarus GS, Grayzel AI. A cytotoxic proteinase isolated from human lymphocytes. J Immunol 1978;**120**:665-670.
17. Heijink IH, van OA, Kapus A. Epidermal growth factor receptor signalling contributes to house dust mite-induced epithelial barrier dysfunction. Eur Respir J 2010;**36**:1016-1026.
18. Hartnell A, Steel J, Turley H, Jones M, Jackson DG, Crocker PR. Characterization of human sialoadhesin, a sialic acid binding receptor expressed by resident and inflammatory macrophage populations. Blood 2001;**97**:288-296.
19. Crocker PR, Freeman S, Gordon S, Kelm S. Sialoadhesin binds preferentially to cells of the granulocytic lineage. J Clin Invest 1995;**95**:635-643.
20. Ward LD, Kellis M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. Nucleic Acids Res 2012;**40**:D930-D934.
21. Schaub MA, Boyle AP, Kundaje A, Batzoglou S, Snyder M. Linking disease associations with regulatory information in the human genome. Genome Research 22, 1748-1759. 2012.

22. Hedrick SM, Hess MR, Doedens AL, Goldrath AW, Stone EL. FOXO transcription factors throughout T cell biology. Nat Rev Immunol 2012;**12**:649–661.

23. Hwang JW, Rajendrasozhan S, Yao H et al. FOXO3 deficiency leads to increased susceptibility to cigarette smoke–induced inflammation, airspace enlargement, and chronic obstructive pulmonary disease. J Immunol 2011;**187**:987–998.

Supplemental Material

Quality control of genome-wide genotyping data.

SNPs with a call rate <95%, minor allele frequency <5% and SNPs deviating from Hardy-Weinberg equilibrium (HWE; p<0.0001) were removed. Samples were removed if 5% or more of the genotypes were missing, if they were duplicates, if they were derived from individuals related to another one in the dataset or if they were ethnic outliers (based on components from multidimensional scaling C1 and C2). These quality control steps were performed using PLINK[1].

Effects of MAF underestimation on epistasis as a result of imputation

Imputed genotypes can be a good approximation in case of single polymorphism analysis but in epistatic analysis, where two genotypes are taken together, the deviation from the actual distribution can be much higher. It can happen that minor allele frequency (MAF) is under- or overestimated. If we assume SNP1 has MAF_1=0.3 and SNP2 has MAF_2=0.08 we will observe distributions as shown in Table S2A. However, if SNP1 is genotyped and MAF_1 is not changed and SNP2 is imputed and underestimated MAF_2 equals to 0.05 the distribution is different (Table S2B). Differences can seem insignificant, however the sum of frequencies of genotypes involved in epistatic effect differs by nearly 3%.

Table S1. Patients characteristics.

Cohort	NELSON		Lifelines{COPD}	
	Cases	Controls	Cases	Controls
N	1,030	844*+ **964**[a]	711	3,013
Age {mean(sd)}	61.3 (5.6)	59.1 (5)	63.9 (5.7)	43.2 (8.6)
% male	100	100	51.9	43.6
Height (cm) {mean(sd)}	178.3 (6.4)	178.2 (6.6)	175.6 (8.9)	175.5 (9.2)
% current smoker	39.8	64.5	49.9	42.4
% former smoker	60.2	35.5	50.1	57.6
Pack-years smoking[b]	42.6 (17.9)	39.4 (16.9)	19.2 (13.8)	10.8 (9)
FEV1/FVC(%) {mean(sd)}	60.3 (8.5)	76.5 (4.9)	63.9 (5.7)	79.3 (4.9)

* Blood bank controls – no demographic data available

[a] Characteristics below for number of individuals in bold

[b] Calculated only within ever-smokers {mean(sd)}

COPD cases were defined as airway obstruction according to GOLD[2] criteria (FEV_1/FVC<0.7).

Table S2 Effects of MAF underestimation on epistasis as a result of imputation.

A) Actual distribution of genotype combinations

	SNP1, MAF=0.3		
SNP2, MAF=0.08	AA1	AB1	BB1
AA2	0.414736	0.355488	0.076176
AB2	0.072128	0.061824	0.013248
BB2	0.003136	0.002688	0.000576

Shaded cells – frequencies of genotype combinations involved in epistasis, F_{epi}=0.078336

B) Distribution of genotype combinations when SNP2 is imputed and MAF is underestimated.

	SNP1, MAF=0.3		
SNP2, MAF=0.05	AA1	AB1	BB1
AA2	0.442225	0.37905	0.081225
AB2	0.04655	0.0399	0.00855
BB2	0.001225	0.00105	0.000225

Shaded cells – frequencies of genotype combinations involved in epistasis, F_{epi} = 0.049725

Supplemental References

1. Purcell S, Neale B, Todd-Brown K et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 2007;81:559-575.
2. From the Global Strategy for the Diagnosis, Management and Prevention of COPD, Global Initiative for Chronic Obstructive Lung Disease (GOLD). http://www.goldcopd.org/, 2011.

GENERAL DISCUSSION & FUTURE PERSPECTIVES

Discussion

Genetics of respiratory phenotypes

The aim of this thesis was to gain insight into the genetic factors underlying COPD. Back in 2008 when this project started, little was known about these factors. There were many candidate gene studies reporting associations with COPD, but the results were inconsistent. We therefore started with a systematic review and meta-analysis of previous findings (Chapter 2). When the GWAS era started (in 2005), several studies were published on respiratory phenotypes, including lung function level (FEV_1, FEV_1/VC), lung function decline, and COPD. The overview of findings from these studies and our own research is shown in Figure 1.

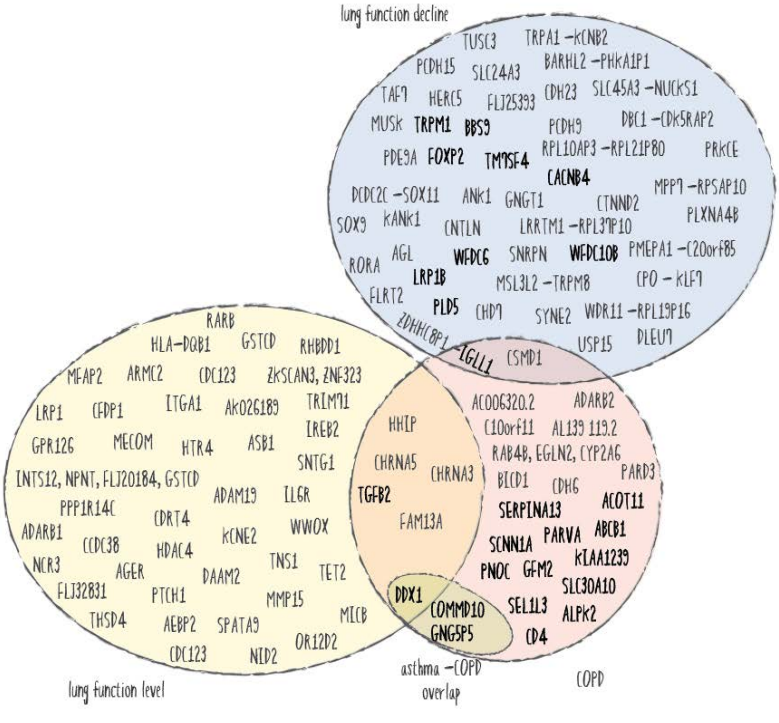


Figure 1. Genetics of lung function level, lung function decline, and COPD – findings from previous studies and studies described in this thesis (bold).

Since COPD is basically a disease of progressive lung function loss, we cannot discuss the genetics of COPD without discussing the genetics of lung function and its decline. Indeed, there are now indications from several studies that genes associated with these respiratory phenotypes show an overlap (Figure 1). This is also one of the reasons why we investigated the shared genetics of airway obstruction and emphysema (Chapter 3) and the shared genetics of asthma and COPD (Chapter 5). We thought that studying multiple aspects of a phenotype would yield more knowledge. However, we kept in mind that airway obstruction and emphysema

(Chapter 3) are separate entities and that lung function decline (Chapter 4) not only occurs in COPD patients, and we therefore studied the specific genetics of these phenotypes as well. We were the first to investigate gene-gene interactions and their contribution to COPD in a genome-wide, hypothesis-free manner (Chapter 6). We identified focal adhesion pathways (Chapter 3) and cell adhesion (Chapter 6) pathways by investigating the genetics of COPD using different approaches. This could be an indication that cell communication is important in COPD. Airway epithelium is known to be damaged in COPD patients and by exposure to smoke.[1] Alterations in focal adhesion and cell adhesion pathways can interfere with wound repair in airway epithelium.[2] However, these pathways can also be altered in other cell types and lead to COPD development. Adhesion is not only crucial for maintaining multicellular structures, like tissue, but it also plays a key role in cell-matrix and cell-cell boundaries and communication. However, it still needs to be established where exactly these pathways are altered and how they may lead to COPD.

Complex genetics of complex disorders

The genetics of complex disorders has indeed become very complex over the last few years. Genome-wide association studies (GWAS) have identified loci associated with many complex phenotypes. These loci differ in size and content: on the one hand, sometimes they are huge and harbor many genes, making it difficult to determine which gene is actually the causal one. On the other hand, sometimes these loci are so-called 'gene deserts', harboring no protein-encoding genes and thus increasing the difficulty of interpreting the results further. Especially in the light of recent advances by the ENCODE project, which so far has mapped 80% of the genome to regulatory elements, like transcription factor binding sites, DNAseI hypersensitive sites, open chromatin, etc. and taking into account the linkage disequilibrium structure of the loci identified by GWAS, we can expect that most of the SNPs will be annotated to a regulatory element. This might also be the case for the SNPs that are annotated to the genes and it may turn out that the gene itself has nothing to do with the disease. So were the GWA studies helpful? Yes, they were. They helped us identify loci that would never have been discovered using a candidate gene approach. If we consider genetics as the map of a disease, GWA studies would be the grid of this map, composed of identified loci and pathways. Now we need to fill in the grid to complete the map, which will not be an easy task given the multiple levels of information required. One of the tools we can use for this are microarrays with a content specific for a particular disease group, like the a metabochip or immunochip for performing fine-mapping studies. The aim is to refine the association signal in the locus, which can help to narrow down the region of interest. Another tool that can be used is next-generation sequencing (NGS), a powerful technique to detect mutations in

the genome. NGS has been successfully applied in a small number of individuals to identify mutations leading to rare Mendelian disorders. But, contrary to its application in Mendelian disorders, in complex disorders, much larger numbers of individuals need to be sequenced, but with the falling price of NGS such studies are becoming feasible. Other omics techniques, like gene expression, proteomics and metabolomics will add an extra layer to this map, making it more comprehensive and easier to read. The exact route, however, will be drawn by functional studies, which will provide information on the exact course of events leading to disease and on the tissue and cell types in which these events actually take place. Another level of information will come from studying both gene-environment and gene-gene interaction.

Missing heritability, hidden heritability, phantom heritability, and epistasis

By now (April 2013), the GWAS era has yielded 9,872 SNPs associated with over 1,400 traits and described in 1,563 publications in little more than five years. Despite this abundance, only a small proportion of heritability of many traits can so far be explained. Some time ago the scientific community had already started wondering how the remainder of the heritability could be found and where. In the beginning, copy number variations (CNV) and rare variants were thought to account for a substantial proportion of the *missing heritability*, but it was soon proven this was not the case.[3] The concept of *hidden heritability* was born, as it was thought that some relevant variants are hidden below conservative genome-wide significance thresholds or because the variants are so rare that they cannot be captured with GWAS.[4] The next suspect was epistasis, in other words gene-gene interaction. Epistasis is a phenomenon in which two genes together affect the phenotype, either by the masking of one or acting synergistically. Epistasis is frequent in all organisms, starting with yeast [5] and up to humans.[6] It has been proposed that by not taking epistasis into account while calculating the proportion of explained heritability, so-called *phantom heritability* is created.[7] In other words, the genetic variants identified in GWA studies could explain much more of the heritability, but not under the additive model assumptions. Therefore the statement that rare variants or even more common variants must underlie common diseases might not be valid. In fact, it is likely that, for many disorders, the major susceptibility loci have already been identified.

For many years, studies on epistasis were restricted to a subset of genes, most frequently the candidate genes, or later to genes identified in GWAS. This restriction was due to a large number of possible interactions. If we want to test all the possible pair-wise interactions, assuming we genotyped 500,000 SNPs, we would have to test $1.25*10^9$ SNP pairs for interaction. Such an analysis is exhaustive computationally, not to mention the time required to run it. Not so long ago there were basically no tools that could enable a whole genome epistatic analysis to be completed in less than a year! With the development of

graphical processing units (GPUs) this has become feasible and more studies describing epistasis in complex disorders have appeared recently. Full genome scans for epistasis has already been were performed in prostate cancer [8] and serum uric acid levels,[9] for example. Both studies were underpowered to detect significant interactions, as the required sample size is much larger than for GWAS, or it could also be due to applying a strict Bonferroni correction. Therefore novel approaches need to be implemented to reduce the limitations of multiple testing, like a two-stage approach, as implemented in EpiBLASTER.[10]

Complex genetics are getting more and more complex

Gene, pseudogene, gene desert, junk DNA: these are just a few terms for which the definitions have changed in the past few years. Genes used to be defined as DNA units composed of a sequence containing protein code. We now know that genes not only encode for proteins, but also for many sorts of RNA genes, including miRNAs (micro RNAs), siRNAs (small interfering RNA), lincRNAs (long intergenic non-coding RNAs), to name a few. Pseudogenes were thought to be an inactive copy of a parent gene but recently it was shown that some pseudogenes are not so 'pseudo' and are either expressed [11] or at least transcribed,[12] and the transcript itself may have a profound biological function.[13] Gene deserts have turned out not to be deserts at all, but harbor many newly discovered genes, not only for protein coding, but also regulatory elements, etc. The concept of junk DNA simply does not hold anymore, as proven by the ENCODE consortium.[14]

These definitions have changed and been improved with the time and effort put into many studies and, as a result, our understanding has increased somewhat. But they also mean genetics is much more complex than we thought. If we take a complex disorder and a change in SNP X that is associated with it, what does it mean? Let's start from the SNP, which by definition a single letter in the genome composed of 3.2 billion such letters. There are many ways through which this single letter can lead to a phenotype, irrespective of its location (Figure 2). A minority of the SNPs identified for complex disorders are located within exons of protein-coding genes, which means that, for the remaining majority, the function cannot be inferred easily. And at the end, we have a phenotype, a complex disorder, which means that not only SNP X will be responsible for it, but also that there might be perhaps hundreds of other SNPs contributing to the phenotype. This is the case for height, for example.

Future perspectives

The 21st century has already brought significant advancements in technologies that can be used to construct tools to enable scientists to make significant scientific progress. This holds for all disciplines, but starting from 2005, the field of genetics has exploded. GWA

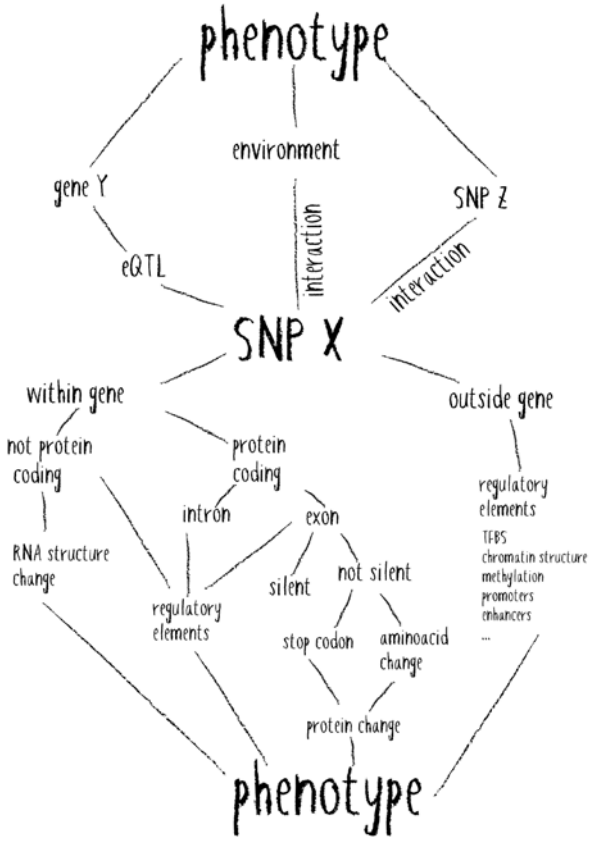


Figure 2. Possible ways through which SNPs can lead to a phenotype.
eQTL – expression quantitative trait locus, TFBS – transcription factor binding site

studies are just a grid for our map of complex disorders. Maybe we still need to create a few of them, for example, for COPD there are still some gaps to be filled. So far there have been no GWA studies on COPD patients who have never smoked and it would be extremely interesting to see if the same or a completely distinct set of genes is responsible for COPD development in this group. However, research resources are limited, and new cohorts take time, effort and money to collect and then to characterize. The near future will probably see fewer and fewer new GWA studies and only a few meta-analyses by consortia. Meta-analyses do have the potential to discover new loci, as shown by many studies, for example, for ulcerative colitis, where pooling six studies yielded 29 new loci.[15] No large meta-analysis has yet been performed for COPD.

COPD remains a chronic and disabling disease affecting millions of people worldwide. The more genetic studies can shed light on its origin and development, the more likely we are to discover leads towards better prognoses and treatment. Many questions about COPD are left to be answered. GWAS era generated many hypotheses that need to be proven or refuted. In the end, finding a gene is just a beginning.

Reference List

1. Barnes PJ. Mediators of chronic obstructive pulmonary disease. Pharmacol Rev 2004;**56**:515–548.

2. Puchelle E, Zahm JM, Tournier JM, Coraux C. Airway epithelial repair, regeneration, and remodeling after injury in chronic obstructive pulmonary disease. Proc Am Thorac Soc 2006;**3**:726–733.

3. Maher B. Personal genomes: The case of the missing heritability. Nature 2008;**456**:18–21.

4. Gibson G. Hints of hidden heritability in GWAS. Nat Genet 2010;**42**:558–560.

5. Gerke J, Lorenz K, Cohen B. Genetic interactions between transcription factors cause natural variation in yeast. Science 2009;**323**:498–501.

6. Mejstrikova E, Janda A, Hrusak O et al. Skin lesions in a boy with X-linked lymphoproliferative disorder: comparison of 5 SH2D1A deletion cases. Pediatrics 2012;**129**:e523–e528.

7. Zuk O, Hechter E, Sunyaev SR, Lander ES. The mystery of missing heritability: Genetic interactions create phantom heritability. Proc Natl Acad Sci U S A 2012;**109**:1193–1198.

8. Tao S, Feng J, Webster T et al. Genome-wide two-locus epistasis scans in prostate cancer using two European populations. Hum Genet 2012;**131**:1225–1234.

9. Wei W, Hemani G, Hicks AA et al. Characterisation of genome-wide association epistasis signals for serum uric acid in human population isolates. PLoS One 2011;**6**:e23836.

10. Kam-Thong T, Czamara D, Tsuda K et al. EPIBLASTER-fast exhaustive two-locus epistasis detection strategy using graphical processing units. Eur J Hum Genet 2011;**19**:465–471.

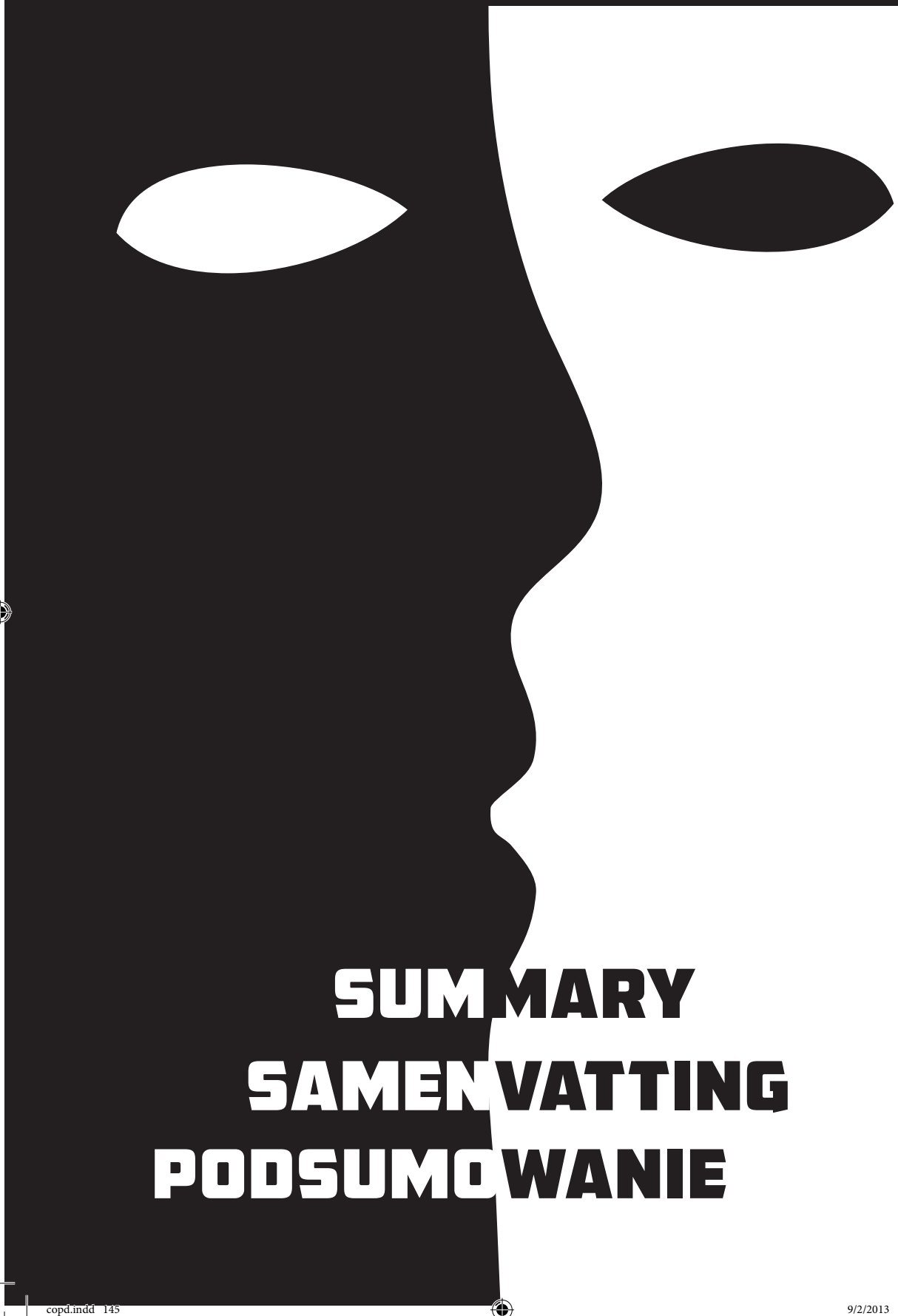
11. Kalyana-Sundaram S, Kumar-Sinha C, Shankar S et al. Expressed pseudogenes in the transcriptional landscape of human cancers. Cell 2012;**149**:1622–1634.

12. Harrison PM, Zheng D, Zhang Z, Carriero N, Gerstein M. Transcribed processed pseudogenes in the human genome: an intermediate form of expressed retrosequence lacking protein-coding ability. Nucleic Acids Res 2005;**33**:2374–2383.

13. Poliseno L, Salmena L, Zhang J, Carver B, Haveman WJ, Pandolfi PP. A coding-independent function of gene and pseudogene mRNAs regulates tumour biology. Nature 2010;**465**:1033–1038.

14. Schaub MA, Boyle AP, Kundaje A, Batzoglou S, Snyder M. Linking disease associations with regulatory information in the human genome. Genome Res 2012;**22**:1748–1759.

15. Anderson CA, Boucher G, Lees CW et al. Meta-analysis identifies 29 additional ulcerative colitis risk loci, increasing the number of confirmed associations to 47. Nat Genet 2011;**43**:246–252.



SUMMARY
SAMENVATTING
PODSUMOWANIE

Summary

The work described in this thesis was done within the framework of the COPACETIC consortium, which was established to determine the genetic background of chronic obstructive pulmonary disease (COPD) and lung function decline. We started with genome-wide screening that yielded dense genotype information.

Chapter 3 describes the main outcome of the consortium, which are genome-wide association (GWA) studies on airway obstruction and emphysema, as well as an attempt to identify shared genetic components. These GWA studies did not yield any genome-wide significant findings, despite large sized replication cohorts (especially in airway obstruction), which might have been due to phenotype heterogeneity and limited power in the discovery cohort. The overlap between airway obstruction and emphysema was difficult to study, due to use of a single discovery cohort. Using two different approaches we were able to point to the same pathway which was focal adhesion pathway, known to be crucial in epithelial repair mechanisms.

In **Chapter 4** we investigated the so-called 'Dutch hypothesis' on a genome-wide level. Briefly, the Dutch hypothesis stated, that asthma and COPD are different manifestations of the same disease unit, hence share genetic background and environmental factors. Only the timing and specific interplay between genetics and environment will determine the final phenotype. To prove or refute this hypothesis we performed meta-analysis using asthma and COPD GWA studies. We found three loci associated with both phenotypes. The genes located in these three loci, *COMMD10* and *DDX1*, point towards NF**κβ** which mediates inflammatory process in both, asthma and COPD. These processes are different, but could be mediated by *COMMD10* and *DDX1* in both diseases, implying that the Dutch hypothesis can be valid. At this stage the Dutch hypothesis can not be proven or refuted with certainty.

Chapter 5 describes GWA study on lung function decline. Our discovery cohort was a population of heavy smokers and our replication cohorts were population-based. What we noticed immediately were differences of the decline extent itself and genetic effects sizes. Heavy-smokers displayed much larger lung function decline per year, compared to the general population and we were not able to replicate the findings from the GWA study in the population-based replication cohorts. However, we performed some additional gene-environment interaction analyses in two out of three replication cohorts, which strongly suggested that the genetic effects are diluted in the replication cohorts due to heterogeneity of the phenotypes that were present. We have shown, that if the replication sample is restricted to heavy smokers with COPD, the genetic effects are much more alike the discovery cohort (heavy-smokers, 50% of them with COPD). However, this restriction reduced the sample size significantly therefore the conclusion is that more phenotype-specific cohorts are needed to

perform proper lung function decline GWA studies.

So far three GWA studies have been published on COPD, however the number of loci identified is much lower than in other phenotypes. There is a debate ongoing on so-called 'missing heritability' or 'hidden heritability' phenomena. Most likely, genetic associations will not explain the entire heritability of a phenotype. Therefore in Chapter 6 we performed the first genome-wide gene-gene interaction analysis in COPD. Due to large number of tests performed our results did not reach genome-wide significance.

We observed a clear over-representation of genes involved in cell adhesion and cell communication pathways. Cell adhesion plays a crucial role in epithelial barrier repair after tissue injury, abundantly present in COPD. These results should be further validated in a large COPD cohort.

In summary, this thesis is an investigation of several phenotypes of COPD: airway obstruction, emphysema, lung function decline. We were interested not only in these phenotypes, but also their shared genetic background. Airway obstruction and emphysema can occur together, asthma and COPD despite being separate diseases share many features, and therefore we searched for common genes underlying these phenotypes. GWA studies were the main tools used, but we also looked beyond GWA studies and used other tools, like analysis of epistasis. The pathways that emerge from our findings are worth investigating further, as not all of them were considered to be involved in COPD before.

Samenvatting

Het in deze thesis gepresenteerde werk is uitgevoerd binnen het COPACETIC consortium, welke opgericht is om de genetische achtergrond van chronische obstructieve longaandoeningen (chronic obstructive pulmonary disease, COPD) en het verval van longfunctie te onderzoeken. We zijn gestart met genoomwijde genotyperingen, welke uitgebreide genotype informatie opleverden.

Hoofdstuk 3 beschrijft de voornaamste uitkomst van het consortium. Dit zijn genoomwijde associatie (GWA) studies op luchtwegobstructie en emfyseem, evenals een poging om gedeelde genetische componenten te identificeren. In de GWA studies werden geen genoomwijd significant geassocieerde bevindingen gedaan, ondanks het gebruik van een groot replicatie cohort (vooral voor luchtwegobstructie). Dit wordt waarschijnlijk veroorzaakt door een grote heterogeniteit in de fenotypes en een gebrek aan statistische power in het initiële cohort. De overlap tussen luchtwegobstructie en emfyseem was moeilijk te bestuderen, aangezien het initiële cohort hetzelfde was voor beide studies. Door middel van twee verschillende methoden konden we gezamelijke pathway aanwijzen, het focale adhesie pathway, bekend om zijn cruciale rol in epitheliale herstel mechanismen.

In hoofdstuk 4 hebben we de validiteit van de zogenaamde "Nederlandse hypothese" op een genoomwijde niveau getest. De Nederlandse hypothese gaat er van uit dat astma en COPD twee verschillende manifestaties zijn van eenzelfde ziekte, en daardoor een genetische achtergrond en omgevingsfactoren delen. De timing en specifieke interacties tussen genetische factoren en de omgeving bepalen het uiteindelijke fenotype. Om deze hypothese te bewijzen, dan wel te verwerpen, hebben we een meta-analyse verricht op astma en COPD studies. We vonden drie loci die met beide fenotypes geassocieerd zijn. Echter, we konden deze bevindingen niet repliceren in populatie gebaseerde cohorten met een minder sterk uitgesproken fenotype. De genen in deze loci, COMMD10 en DDX1 wijzen richting NFkB, welke imflammatoire processen in zowel astma en COPD reguleert. De onderliggende processen zijn verschillend in beide aandoeningen, maar kunnen in beide aandoeningen gereguleerd worden door COMMD10 en DDX1, wat er op wijst dat de Nederlandse hypothese waar zou kunnen zijn. Op dit moment is het echter nog onmogelijk de Nederlandse hypothese definitief te bewijzen of te verwerpen.

Hoofdstuk 5 beschrijft een GWA studie op verval van longfunctie. Ons initiële cohort is een van zware rokers en de replicatie cohorten zijn populatie gebaseerd. Wat ons direct opviel waren de verschillen in de ernst van het verval en de effect grootte van de genetische varianten. Zware rokers laten per jaar een sterker verval in longfunctie zien ten opzichte van de algehele bevolking. Daardoor is het niet verrassend dat we de bevindingen van de GWA studie niet konden repliceren.

Daarnaast hebben we ook gen-omgeving interactie analyses uitgevoerd in twee van de drie replicatie cohorten. Deze suggereren dat de genetische effecten in de replicatie cohorten worden afgezwakt doordat er een grote variatie is in fenotypes. We toonden aan dat, als we alleen zware rokers met COPD voor de replicatie gebruikenn, de genetische effecten veel meer lijken op die gevonden in het initiële cohort (zware rokers, 50% heeft COPD). Echter, dat vermindert de grootte van het cohort sterk, en daardoor concluderen we dat er meer fenotype-specifieke cohorten nodig zijn om goede GWA studies op verval van longfunctie uit te voeren.

Tot nu toe zijn er drie GWA studies op COPD gepubliceerd. Het aantal gevonden loci is echter

veel lager wanneer we dit vergelijken met andere fenotypes. Er is een discussie gaande over het zogenoemde fenomeen "missende erfelijkheid" of "verborgen erfelijkheid". Zeer waarschijnlijk zullen genetische associaties niet de gehele erfelijkheid van een fenotype kunnen verklaren. Daarom hebben we in hoofdstuk 6 de eerste genoomwijde gen-gen interactie analyse in COPD uitgevoerd. Door het grote aantal uitgevoerde testen konden onze resultaten niet de grens van genoomwijde significantie bereiken. Het feit dat we geïmputeerde genotypes hebben gebruikt in het replicatie cohort kan ook het resultaat hebben beïnvloed. Echter, we vonden wel een duidelijke overrepresentatie van genen die betrokken zijn bij cel adhesie en cel communicatie pathways. Cel adhesie speelt een cruciale rol in epitheliale barriere herstel na weefselschade, hetgeen veel aanwezig is in COPD. Deze resultaten moeten nog verder gevalideerd worden in een groot COPD cohort.

Samenvattend vormt deze thesis een onderzoek van vele fenotypes van COPD: luchtwegobstructie, emfyseem en verval van longfunctie. We waren niet alleen in deze fenotypes geïnteresseerd, maar ook in hun gedeelde genetische achtergrond. Luchtwegobstructie en emfyseem kunnen samen voorkomen en ondanks dat astma en COPD aparte aandoeningen zijn, delen ze toch vele karakteristieken. Daarom hebben we gezocht naar gemeenschappelijke genen die ten grondslag liggen aan deze fenotypes. We hebben vooral gebruikt gemaakt van GWA studies, maar we hebben ook verder gekeken middels epistase analyse. Het is van belang de gevonden pathways verder te onderzoeken, aangezien deze eerder nog niet waren aangemerkt als relevant voor COPD.

Podsumowanie

Praca przedstawiona w niniejszym manuskrypcie została wykonana w ramach konsorcjum COPACETIC, które powstało w celu odkrycia podłoża genetycznego przewlekłej obturacyjnej choroby płuc (POChP, z ang. Chronic obstructive pulmonary disease, COPD) oraz spadku funkcji płuc. Rozpoczęliśmy od genotypowania całego genomu, które dostarczyło szczegółowych informacji o genotypach w dużej rozdzielczości.

Rozdział 3 opisuje główne wyniki prac konsorcjum, którymi są genomowe badania asocjacyjne na fenotypach obstrukcji dróg oddechowych i rozedmy oraz próbę identyfikacji wspólnych dla tych fenotypów czynników genetycznych. Oba genomowe badania asocjacyjne nie wykazały żadnych istotnych statystycznie wyników, pomimo replikacji wyników w dużych kohortach (szczególnie w przypadku obstrukcji dróg oddechowych), co może wynikać z różnorodności fenotypów oraz ograniczonej mocy głównej kohorty. Badanie wspólnych czynników genetycznych dla obu fenotypów było trudne, wobec faktu użycia jednej kohorty. Z użyciem dwóch różnych metod udało nam się wskazać na tą samą ścieżkę metaboliczną, którą była adhezja fokalna, kluczowa w mechanizmach naprawy bariery nabłonkowej.

W Rozdziale 4 rozpatrywaliśmy tak zwaną "hipotezę holenderską" na poziomie całego genomu. W skrócie, hipoteza holnderska zakładała, że astma i POChP są różnymi manifestacjami tej samej jednostki chorobowej, a zatem mają wspólne podłoże genetyczne i czynniki środowiskowe. Jedynie czas i specyficzne oddziaływanie czynników genetycznych i środowiskowych determinuje ostateczny fenotyp. Aby udowodnić lub odrzucić tę hipotezę przeprowadziliśmy meta-analizę używając genomowch badań asocjacyjnych przeprowadzonych na astmi i POChP. Zidentyfikowaliśmy trzy loci zasocjowane z obydwoma fenotypami. Geny w tych loci, COMMD10 oraz DDX1, wskazują na NFκβ, który odpowiada za proces zapalny zarówno w astmie jak i w POChP. Procesy te różnią się, ale mogłyby zachodzić za pośrednictwem COMMD10 i DDX1 w obu chorobach, sugerując że hipoteza holenderska może być prawdziwa. Na tym etapie hipoteza ta nie może zostać ani potwierdzona ani odrzucona z pewnością.

Rozdział 5 opisuje genomowe badanie asocjacyjne zaniku funkcji płuc. Główną kohortą była populacja nałogowych palaczy a kohorty replikacyjne składały się z przedstawicieli populacji ogólnej. Co dało zauważyć się natychmiast, to różnice w skali zaniku funkcji płuc oraz efektów czynników genetycznych. Palacze wykazywali znacznie większy zanik funkcji płuc w ciągu roku, w porównaniu do populacji ogólnych i nie byliśmy w stanie zreplikować wyników genomowego badania asocjacyjnego. Jednakże przeprowadziliśmy dodatkowe analizy interakcji czynników genetycznych i środowiskowych w dwóch spośród trzech kohort replikacyjnych, które sugerowały „rozmycie" efektów genetycznych w kohortach replikacyjnych z powodu heterogenności fenotypów. Pokazaliśmy, że ograniczenie próby

replikacyjnej do palaczy chorych na PoChP pozwala uzyskać efekty genetyczne podobne do tych z kohorty głównej (palacze, 50% chorych na PoChP). Jednakże, taki zabieg znacząco redukuje liczebność kohorty, zatem naszym wnioskiem jest, że potrzebnych jest więcej kohort specyficznych dla tego fenotypu aby przeprowadzić prawidłowe genomowe badanie asocjacyjne zaniku funkcji płuc.

Dotychczas opublikowane został trzy genomowe badania asocjacyjne dla POChP, jednakże liczba odkrytych loci jest znacznie niższa w porównaniu z innymi fenotypami. Trwa debata na temat tak zwanej „brakującej dziedziczności" lub „ukrytej dziedziczności". Najprawdopodobniej asocjacje genetyczne nie wyjaśnią całkowitej dziedziczności fenotypu. Wobec tego w Rozdziale 6 opisaliśmy pierwszą genomową analizę interakcji czynników genetycznych. Z powodu dużej liczby przeprowadzonych testów, żaden z wyników nie osiągnął istotności statystycznej.

Zaobserwowaliśmy zdecydowaną przewagę genów związanych z adhezją komórkową i komunikacją komórkową. Adhezja komórkowa odgrywa istotną rolę w naprawie bariery nabłonkowej po uszkodzeniu tkanki, która obficie występuje u chorych na POChP. Te wyniki powinny zostać potwierdzone w dużej kohorcie POChP.

Podsumowując, niniejsza praca rozpatruje kilka fenotypów POChP: obstrukcję dróg oddechowych, rozedmę oraz zanik funkcji płuc. Interesowały nas nie tylko same fenotypy, ale również ich wspólne podłoże genetyczne. Obstrukcja dróg oddechowych i rozedma mogą występować jednocześnie, astma i POChP mimo iż są osobnymi chorobami dzielą wiele cech, wobec czego poszukiwaliśmy wspólnych genów stanowiących podłoże tych fenotypów. Genomowe badania asocjacyjne były głównym narzędziem używanym w niniejszej pracy, używaliśmy również innych narzędzi, jak analiza epistazy. Ścieżki metaboliczne, które wyłaniają się na podstawie naszych wyników powinny być rozpatrywane w przyszłości, jako że nie wszystkie z nich były brane pod uwagę w kontekście POChP.

List of Publications

1 Coenen MJ, Trynka G, Heskamp S, Franke B, van Diemen CC, **Smolonska J**, van Leeuwen M, Brouwer E, Boezen MH, Postma DS, Platteel M, Zanen P, Lammers JW, Groen HJ, Mali WP, Mulder CJ, Tack GJ, Verbeek WH, Wolters VM, Houwen RH, Mearin ML, van Heel DA, Radstake TR, van Riel PL, Wijmenga C, Barrera P, Zhernakova A. Common and different genetic background for rheumatoid arthritis and coeliac disease. Hum Mol Genet 2009; 18(21):4195-203.

2 **Smolonska J**, Wijmenga C, Postma DS, Boezen HM. Meta-analyses on suspected chronic obstructive pulmonary disease genes: a summary of 20 years' research. Am J Respir Crit Care Med 2009; 180(7):618-31. *In this thesis.*

3 Fehrmann RS, Jansen RC, Veldink JH, Westra HJ, Arends D, Bonder MJ, Fu J, Deelen P, Groen HJ, **Smolonska A**, Weersma RK, Hofstra RM, Buurman WA, Rensen S, Wolfs MG, Platteel M, Zhernakova A, Elbers CC, Festen EM, Trynka G, Hofker MH, Saris CG, Ophoff RA, van den Berg LH, van Heel DA, Wijmenga C, Te Meerman GJ, Franke L. Trans-eQTLs reveal that independent genetic variants associated with a complex phenotype converge on intermediate genes, with a major role for the HLA. PLoS Genet 2011; 7(8):e1002197.

4 Janse M, Lamberts LE, Franke L, Raychaudhuri S, Ellinghaus E, Muri BK Muri Boberg K, Melum E, Folseraas T, Schrumpf E, Bergquist A, Björnsson E, Fu J, Jan Westra H, Groen HJ, Fehrmann RS, **Smolonska J**, van den Berg LH, Ophoff RA, Porte RJ, Weismüller TJ, Wedemeyer J, Schramm C, Sterneck M, Günther R, Braun F, Vermeire S, Henckaerts L, Wijmenga C, Ponsioen CY, Schreiber S, Karlsen TH, Franke A, Weersma RK. Three ulcerative colitis susceptibility loci are associated with primary sclerosing cholangitis and indicate a role for IL2, REL, and CARD9. Hepatology 2011; 53(6):1977-85.

5 Hancock DB, Artigas MS, Gharib SA, Henry A, Manichaikul A, Ramasamy A, Loth DW, Imboden M, Koch B, McArdle WL, Smith AV, **Smolonska J**, Sood A, Tang W, Wilk JB, Zhai G, Zhao JH, Aschard H, Burkart KM, Curjuric I, Eijgelsheim M, Elliott P, Gu X, Harris TB, Janson C, Homuth G, Hysi PG, Liu JZ, Loehr LR, Lohman K, Loos RJ, Manning AK, Marciante KD, Obeidat M, Postma DS, Aldrich MC, Brusselle GG, Chen TH, Eiriksdottir G, Franceschini N, Heinrich J, Rotter JI, Wijmenga C,Williams OD, Bentley AR, Hofman A, Laurie CC, Lumley T, Morrison AC, Joubert BR, Rivadeneira F, Couper DJ, Kritchevsky SB, Liu Y, Wjst M, Wain LV, Vonk JM, Uitterlinden AG, Rochat T, Rich SS, Psaty BM, O'Connor GT, North KE, Mirel DB, Meibohm B, Launer LJ, Khaw KT, Hartikainen AL, Hammond CJ, Gläser S, Marchini J, Kraft P, Wareham NJ, Völzke H, Stricker BH, Spector TD, Probst-Hensch NM, Jarvis D, Jarvelin MR, Heckbert SR, Gudnason V, Boezen HM, Barr RG, Cassano PA, Strachan DP, Fornage M, Hall IP, Dupuis J, Tobin MD, London SJ. Genome-wide joint meta-analysis of SNP and SNP-by-smoking interaction identifies novel loci for pulmonary function. PLoS Genet 2012; 8(12):e1003098.

6 Wilk JB, Shrine NR, Loehr LR, Zhao JH, Manichaikul A, Lopez LM, Smith AV, Heckbert SR, **Smolonska J**, Tang W, Loth DW, Curjuric I, Hui J, Cho MH, Latourelle JC, Henry AP, Aldrich M, Bakke P, Beaty TH, Bentley AR, Borecki IB, Brusselle GG, Burkart KM, Chen TH, Couper D, Crapo JD, Davies G, Dupuis J, Franceschini N, Gulsvik A, Hancock DB, Harris TB, Hofman A, Imboden M, James AL, Khaw KT, Lahousse L, Launer LJ, Litonjua A, Liu Y, Lohman KK, Lomas DA, Lumley T, Marciante KD, McArdle WL, Meibohm B, Morrison AC, Musk AW, Myers RH, North KE, Postma DS, Psaty BM, Rich SS, Rivadeneira F, Rochat T, Rotter JI, Artigas MS, Starr JM, Uitterlinden AG, Wareham NJ, Wijmenga C, Zanen P, Province MA, Silverman EK, Deary IJ, Palmer LJ, Cassano PA, Gudnason V, Barr RG, Loos RJ, Strachan DP, London SJ, Boezen HM, Probst-Hensch N, Gharib SA, Hall IP,

O'Connor GT, Tobin MD, Stricker BH. Genome Wide Association Studies Identify CHRNA5/3 and HTR4 in the Development of Airflow Obstruction. Am J Respir Crit Care Med 2012.

7 Mohamed Hoesein FA, Wauters E, Janssens W, Groen HJ, **Smolonska J**, Wijmenga C, Postma DS, Boezen HM, De Jong PA, Decramer M, Lammers JW, Lambrechts D, Zanen P. Variants in the 15q24/25 locus associate with lung function decline in active smokers. PLoS One 2013; 8(1):e53219.

8 Hansel NN, Ruczinski I, Rafaels N, Sin DD, Daley D, Malinina A, Huang L, Sandford A, Murray T, Kim Y, Vergara C, Heckbert SR, Psaty BM, Li G, Elliott WM, Aminuddin F, Dupuis J, O'Connor GT, Doheny K, Scott AF, Boezen HM, Postma DS, **Smolonska J**, Zanen P, Mohamed Hoesein FA, de Koning HJ, Crystal RG, Tanaka T, Ferrucci L, Silverman E, Wan E, Vestbo J, Lomas DA, Connett J, Wise RA, Neptune ER, Mathias RA, Paré PD, Beaty TH, Barnes K.C.et al. Genome-wide study identifies two loci associated with lung function decline in mild to moderate COPD. Hum Genet 2013; 132(1):79-90.

9 van Setten J, Isgum I, **Smolonska J**, Ripke S, De Jong PA, Oudkerk M, de Koning H, Lammers JW, Zanen P, Groen HJ, Marike Boezen H, Postma DS, Wijmenga C, iergever MA, Mali WP, de Bakker PI. Genome-wide association study of coronary and aortic calcification implicates risk loci for coronary artery disease and myocardial infarction. Atherosclerosis 2013 {Epub ahead of print}

Thank you/Dziękuję!

Before I start listing All of You, I want to thank you All, and if by any chance your name is not there, please forgive me.

Dear Supervisors, thank you for your time and effort you invested to make me a with a PhD degree.
Cisca, thank you for learning me to question things and to be critical. Marike, thank you for always being so calm, even when my adrenaline levels were sky high ;-) and for going with me through all my heart attacks, and we know there were quite a few of them.

Dear People from the COPACETIC consortium, without your contribution and effort it would not be possible to do this project. It was always very nice to meet you in Amsterdam!
Dirkje, thank you for all the warm words during all the meetings we had, your input in all the papers we have written together and your patience.
Pieter, without you our COPACETIC meetings, both in Groningen and in Amsterdam would not be the same. I always smiled seeing your A4-sized emails, whatever was the subject :-)
Dear Borge, Marianne, Sune and Morten, thank you for hosting me and Mathieu in your Department in Herlev Hospital, we both enjoyed it and I fell in love with Kobenhavn :-) It was a real pleasure to work with you all the way through my PhD.
Dear Jorgen, thank you for your invaluable input in the COPACETIC project, I always enjoyed your presentations.
Dear Kerrie, Morten and Kristina, the Astra-Zeneca people / it was a real pleasure to work with you, thank you for all the efforts to genotype samples.
Dear Hans, Ulrike, Monika, Michael and Suzanne, thank you for all your hard work you put to have samples and measurements ready.
Dear Ewa, Paweł and Filip, thank you for your contribution to this project, it is nice that I was not the only Polish accent in it. Dear Ewa, special thank you for all the nice talks we had during COPACETIC meetings and for bringing sweets from Poland :-)
Dear Firdaus, you were the first to defend with COPACETIC project, congratulations and good luck in your future carreer. It was nice to work with you.
Dear Akkelies, thank you for always trying to cheer me up, when I was down a bit quite frequently :-), for your patience with me and your sense of humor. Good luck with everything!

Dear People from Genetica, I am not sure where to start, as there are so many of you! Some of you became my friends and thanks to you I will have some nice memories as well.
First I met you Mathieu, my dearest Grandpa / I will never forget our awesome time in the lab, all the evil things we did together]:-) and all the fun we had! You were the one to introduce me to everyone, to show me the lab and to teach me how to do the GWAS from the basics! We were the dream team when it comes to Illumina assays / Further, I always enjoyed teasing you and I will continue doing that :-D It was also a lot of fun to visit your home, you have a great wife and cool kids :-) Good luck Grandpa :*

Dear Soesie J, you are a real superstar in this department! I admire your high-heel walking skills. I always enjoyed our chit-chats, more or less serious conversations and of course dancing together under the supervision of Justyn :D Good luck with everything, especially with your two cute little heart-breakers ;-)

Dear Astrid, thank you for your involvement in everything!

Dearest Secretaries, Helene, Bote, Joke and Hentje, thank you for your invaluable help in everything else than science, so in real life J Helene, thank you for reminding me about all the meetings that I forgot. Bote, thanks for all your help with Agata's books, it was a real drama, but we did it ;)
Dear Hayo, thank you for all nice talks we had in vicinity of the coffee machine and for your everyday smile :-)
Dear Marina, thank you for your patience in explaining mysteries of a Dutch bureaucracy.

Dear Jackie, thank you for all your help, not only the editorial one and for all the talks we had.

Dear Gerard, thank you for explaining statistical mysteries to me so well, that I was leaving your office always convinced that I understand and know what to do. Later it turned out I did not, but anyway: thank you! Also for reminding me frequently that smoking and piercing are bad for me :-)

My Dear, countless room mates, I had four offices in genetics, so if you were not my room mate, please move to the next section, as this one will be rather long.

My first office room mates, Paul and Annemike, it was nice sharing an office with you. I think Paul, you enjoyed it as much as I did, hence you moved to the second office with me ;) Annemike, we were both newbees back then, sharing fears and some problems. I enjoyed our lunch time together later on :-) Good luck with your PhD, I am sure you will be done soon, and your further scientific career.

My Second office room mates, Paul, Mats, Gosia, Jihane, Agata, Lude. Paul, it was a real fun, I enjoyed your jokes and sense of humor. Mats, you sure talk a lot, but it was still a pleasure :-) good luck in Sweden! Gosia, I hated your squirrel or whatever that shit was :-D Thanks for all the help with GWAS/Linux and good luck in Boston. Jiji, thanks for everything! You were there when I cried, when I laughed, always trying to calm me down when necessary. I will always remember you walking with your awesome hair, shining in the sun, with one of your countless bags and sunglasses of course! Gosia and Jiji, thanks for printing your books with Eikon, that was foundation for my company!
Agata, as a room mate I nearly hated you, and you already know that, so find yourself in the next section, as in this one I have nothing nice to say ;-P

My third office room mates, Yun-cute, Rodrigo, Senapati, Mitja, Rajendra and Omid. Guys, I had fun sharing this office with you, I think I helped each of you with something, so in fact you should consider yourselves lucky to have me there ;-D Rodrigo: pills!
Yunia, my Yun-cute, you were always looking for recipes exactly when I was hungry, recipes

with pictures of course, to make me suffer even more :-D But even though I enjoyed sharing the office with you :D Good luck in Rotterdam and further, wherever you go!

My fourth and last office room mates, Barbara, Pumpkin and Cleo, you had this bad luck of sharing an office with me when the time was the most difficult – the end of my endless PhD. I thank you all and one by one for going through all the dramas with me, whether you wanted or not. You bravely dealt with all the swearing, tears, more swearing and more tears.

B, you are, I think, one of the most positive persons I know. No matter what happens you always see the bright side and you don't worry that much, I would love to have this skill. So keep it this way as long as you can. STERKTE and GEOEDE luck with everything!

Pumpkin, my Evil Pumpkin, I have no words to express how severely disturbed you are :-D You are the craziest person I have met in the Netherlands, you are unpredictable in both, your words and actions :-D But even with these minor [honestly, they are not so minor] flaws ;-) you are simply a good friend, who will listen, understand and help to find a solution. I will miss you Pumpkin! Take care of Shieldpad ;-)

Dear Shieldpad, thanks to you a "Genius of the day" trophy was born :-D It was always fun to go for lunch with you and Pumpkin and talk about all this more and less [mostly less] serious stuff :-)

Dear Isis, I am so proud of you! Thank you for participating in the Glamour Day, despite your disability to wear high heels :-) Supervising you in a non-scientific matters was a real pleasure. And you brought pigs for me, for which I will be grateful forever :-D

Anna P. thanks for all the delicious things you brought from Hungary and for your help when Marcin was arranging the stay in Budapest! Remember: Polak, Węgier dwa bratanki!

Dear Duco, you left to a different department, I blame Agata for that :-D it is a miracle that you have been able to stay in the same office with her + two other crazy girls for so long ;-) You are a really nice guy, thing I found uncommon in the Netherlands. So do not change! And good luck with your PhD!

Dear Helga, thank you for your LAUGH, whenever I heard it I was laughing myself, it is one of the most infectious laughs I have heard in my life :-D I like your mean and evil sense of humor, similar to mine :-D

Dear Dineke, even though you think yoga is nothing, I still would like to thank you for discussions over lunch and all the fun associated with it od luck with your career, both, as a scientist and as a mom /

Dear Cleo, you are certainly the coolest post-doc I have ever met! I think I saw you angry only once during these five years, you take things with ease and do not make that much drama... Thanks for the positive energy!

I would also like to thank remaining people from the bright side of the corridor: Ettje, Javier, Olga, Alejandro, Anna D, Eva, Ellen, Helen, Rene, Wytske, Michiel. Thank you for nice time, I enjoyed eating and smoking with you :-D Also those who moved from Groningen already, Ana F, Maria and Kaushal all the best! Robert thank you for your jokes, smile and shirts.

Dear People from the Dark Side of the Corridor :-D

Dear George, always a pleasure talking to you about anything, thanks for all the help with EpiBLASTER, together with Marcel!

Jiha! Thanks for your help with Lovebird's web and for making the coolest contact form ever! I am also grateful you introduced me to climbing, it really helped me at the difficult period of the never-approaching end of my PhD ;-)

Jing-yan! Thank you for always finding time to help me solve the mysteries of R and for translating the Chinese article for my review :-)

Freerk! Thanks for all the talks, those with sense and those with complete lack of it, and the nice trip to Berlin it was fun! Good luck with your career and see you at some festival one day I guess ;-)

Harm-Jan and Lude, thanks for all the help with imputations, eQTLs and other hard-core genetic stuff.

Alex and Despoina, thanks for your crazy sense of humor and huge distance to the surrounding reality!

Dear Epi people,

Thank you for the nice atmosphere at the unit meetings, all your help and time, especially Judith and Marjan, Simona, Kim, Nienke and Despo, we had a lot of fun in Barcelona, I enjoyed it a lot, Niloofar, good luck with your thesis, I keep my fingers crossed. Petra, you were always very helpful and I think one of the most positive and optimistic people I have met in the Netherlands! Thanks for your everyday smile!

Salome, thanks for always listening to me, I could always talk to you about the dark sides of science and you understood it very well. I hope I was not too irritating visiting the "Monster" form time to time ;-)

Dear Jan, thank you for introducing me to the R world, for all your help, for many tea cups we had together and for your patience with me.

Dear Marjolein, I would like to thank you for all the yoga classes, you have no idea how much they helped me to survive the hardest moments in the Netherlands. You introduced me to Ashtanga yoga, for which I will be grateful forever, as it became my real passion. Thank you for always trying to come up with something extra challenging for me! :-)

Dziękuję rodzicom, za ich wsparcie, za zawsze ciepłe powitanie przy krótkich powrotach do domu. Dziękuję Mamuś za tysiące pierogów i krokietów, które ulepiłaś dla mnie i dla innych, bo wiesz, że zawsze się dzielę! Tato, dziękuję za pomoc przy rowerze, przy innych sprawach, nie tylko technicznych i za sukienkę na obronę!

Pawełku, Garbiku, dziękuję, że zawsze odwoziłeś mnie z tobołami na lotnisko i odradzałeś powrót do Polski :-P wróciłam, bo się nie słucham młodszych!

Bobki, dzięki za wszystkie przygody, te w Holandii, w Polsce i w innych miejscach na świecie! Dzięki za pomoc przy przeprowadzkach, wsparcie mentalne i po prostu piękną przyjaźń, której nawet pobyt w tym przebrzydłym kraju nie zniszczył!

Czubu i Lalu, Wy zawsze znalazłyście dla mnie czas, żeby spotkać się choć na chwilę. Dziękuję za wszystkie wieczorylnoce przegadane do rana, za wysłuchiwanie moich narzekań, za zwalczanie mojej depresji!

Dziękuję też innym, którzy zawsze umilali mi czas w Polsce. Aśka W ,teraz już S., dzięki że zawsze wpadałaś na piwo, dopóki mogłaś, że wysłuchałaś. Dziadu! U Ciebie zawsze czas jakoś inaczej płynie, wpada człowiek na chwilę, a tu noc go zastaje :-D Dzięki za powrót do dzieciństwa, czasu dinozaurów, Czarodziejki z Księżyca i innych! Michał, Krzyw, Natalia, Rychu, Grzesiek, Andrzej: dzięki za wszystkie wieczory z grami, hulankami i swawolami, sylwestry, woodstocki i inne :-)

Dziękuję rodzicom Marcina, Elżbiecie i Czesławowi, za to, że pozwolili mi porwać ich syna ;-) oraz całej rodzinie Kępińskich za akceptację, wsparcie i gościnę oraz za miłe spędzony razem czas: Kasi, Mirkowi, Wojtkowi, Ewie i Bartkowi, cioci Ani i wujkowi Jankowi oraz babci Józefie.

Justyn, dzięki za wspólne mieszkanie, narzekanie oraz znoszenie mojego narzekania, którego było zawsze najwięcej :-D, za rysowanie, razem miałyśmy ten jakże owocny "okres holenderski" w twórczości :-D i za Reggeaton i za break dance i za wszystkie kawy i fajki na przerwach. Tonka! Tobie również dzięki za wspólne mieszkanie, za poranne dyskusje, wiem, że czasem było ciężko ;-) ale dałyśmy radę, Pamiętaj, że masz przepiękne włosy! Fajnie było poznać Twoich rodziców i posłuchać opowieści o Chorwacji! Przy okazji pozdrowienia!

Gosia, dzięki za wspólnie spędzony czas, czy to w NL czy to w PL, zawsze fajnie wyjść razem na piwko :-) Dbaj o naszą Babcię Snoepcię, Tooohohohoho.

Albert, Dimer and Remco, thanks for all the biertjes, parties, dinners, concerts and stuff! It was always a lot of fun. Albert, thanks for teaching me the basics of HTML and CSS and for rescuing the computer in Epi :-)

Anno! Zawsze miło było zjeść z Tobą lunch i pogadać o dramatach większych i mniejszych :-) przebrnęłyśmy przez ten doktorancki koszmar! Dzięki też za druk z Lovebird! Zaszalałyśmy z groszkami!

Sąsiedzi! Paulina i Łukasz, Ciocia i Wujek samo zło! Dzięki za wszystkie imprezy spędzone razem, te niezliczone ilości alkoholu wypite i te przetańczone na macie godziny :-D zawsze był z Wami ubaw, mam nadzieję, że będziemy kontynuować w Polsce. Wspólne narzekanie i naśmiewanie się z holendrów było czystą przyjemnością! Złośliwce :-P Paulinko, Tobie szczególnie dziękuję za rycie mi beretu podczas jazdy rowerem! nie prędko Ci to zapomnę :-P

Hahdo i Lamusie, although we met in fact few times, I always enjoyed it and had a lot of fun with you, Good luck my sweetest Lamusy! Kisses :*

Snoepje, moja Ty wielokrotnie invited speaker, moja Ty Noblistko, mój Ty Człowieku-Ekierko! nie wiem co Ci mam tu pisać, wszystko wiesz już dawno, a może lepiej nie pisać tu nic, żebyś mi później nie pier$%#nęła tą książką w twarz :-D a napiszę, co mi tam! Jesteś najbardziej porytą osobą jaką znam :-D I za to Cię uwielbiam! Twój fazowy śmiech zawsze doprowadza mnie do łez :-D

Canard, despite the fact that you are Frunch, I always enjoyed quacking with you :-D I think I have never talked about so much bullshit with anyone else :-D it was like therapy! I hope one day you will make friends with email or other means of contacting people and that we can be in touch! How is Lulu? :P ,

Valentin, thanks for taking care of Agata, I know this is not so easy sometimes ;-) I really like your sense of humor, your passion for bigos and your elephantness :-D Thanks for driving with me this summer and for really nice holidays!

Agata, Habibąku! Tobie to osobny rozdział musiałabym chyba poświęcić, żeby Ci podziękować za wszystko. Przede wszystkim dziękuję, za znoszenie moich szlochów, ile też Ci ja przed Koleżanką wylałam! Za słuchanie o patologiach mego życia i umysłu, za wsparcie, za opieprzanie kiedy trzeba, a trzeba było parę razy :-P Za wszystkie wieczory z "pazgnokciami" :-D za ziółka, za zupę z pieca, za Twoją dysleksję Ci nawet dziękuję, bo niejednokrotnie ubawiła mnie do łez. Dziękuję za wspólne trwonienie pieniędzy jak były :-D za wszystkie koncerty, za wszystkie wypady, za sesje edukacyjne o chorobach przenoszonych drogą płciową, o drugiej wojnie światowej i wielu innych. Dziękuję po prostu za przyjaźń, za to, że zawsze jesteś i zawsze wysłuchasz. Mam nadzieję, że znajdziesz wkrótce wymarzoną pracę i że po prostu będzie dobrze, bo kiedyś musi być.

Dziękuję również Zofii, za zrozumienie bez słów jeśli chodzi o Habibąkowe sprawy :-D I ogólny ubaw jaki zwykle razem mamy, Dziękuję Małgorzacie, mamie Agaty, za zawsze miłe spędzony czas i gościnę i Krzysztofowi, za miłe rozmowy na balkonie oraz cierpliwość podczas szukania Agaty na lotnisku!

Misiaku Mój :* Tobie dziękuję najbardziej ze wszystkich, bo to Ty pomogłeś mi przetrwać te pięć lat. Przede wszystkim dziękuję, że nie posłuchałeś mnie na imprezie w kinie i że znając mnie tak krótko zdecydowałeś się na przyjazd do Holandii ze mną, czego pewnie czasem żałowałeś.. Pamiętam początki, które nie były łatwe, ale daliśmy radę, jak ze wszystkim! Twoja cierpliwość do mnie, mnie samą zadziwia. Dziękuję Ci za ciągłe wsparcie, podnoszenie mnie na duchu i przywracanie wiary we własne siły i umiejętności, za Twój optymizm zwalczający mój skrajny realizm zwany potocznie pesymizmem. Dziękuję, za to, że dbasz o mnie (ą), że pozwalasz zabrać całą kołdrę i że jesteś taki, a nie inny. Uwielbiam Cię najbardziej na świecie! :*

I would like to make special thanks for everyone who helped me with my company, Lovebird design & printing solutions and all the people who decided to print their books with me!

Dziękuję drukarni Eikon Plus za owocną współpracę, zwłaszcza Dominikowi Sieńko, za wprowadzenie mnie w tajniki druku oraz Pawłowi Gietce, za cierpliwość i odpowiadanie na moje niezliczone pytania, zawsze ogromne wsparcie przy realizacji moich nieszablonowych pomysłów (udało się tylko z wizytówką ;-P), a także za ubaw na Skype :-)

Nobody else seems to do that, but I would also like to thank myself, for being stubborn or stupid enough (time will show) to finish this PhD.

Once again, thank you all, do not take it personally if you are not listed here.
Dziękuję, Dank je, Gracias, etc.

Asia