THE ROLE OF GENETICS, NUTRITION, AND CIGARETTE SMOKING IN THE LONGITUDINAL CHANGE IN LUNG FUNCTION

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Wenbo Tang

August 2013

THE ROLE OF GENETICS, NUTRITION, AND CIGARETTE SMOKING IN THE LONGITUDINAL CHANGE IN LUNG FUNCTION

Wenbo Tang, Ph.D.

Cornell University 2013

Lung function is an important predictor of population morbidity and mortality. Decline in lung function is a natural part of aging, but accelerated loss in lung function over time is a harbinger of chronic obstructive pulmonary disease (COPD), a leading cause of death globally. Smoking is widely recognized as the key risk factor for reduced lung function and COPD, although additional risk factors, such as genetics and nutrition, have been suggested to also play important roles in contributing to changes in lung function. The overall aim of this research was to investigate the role of, and interaction between, genetics, nutrition, and cigarette smoking in relation to the longitudinal change in lung function, as an indicator of COPD susceptibility.

First, we explored the association between genetic variation within a network of antioxidant enzyme genes and the rate of change in lung function in a prospective cohort study of African and European American elderly adults; this study also investigated gene-by-smoking interaction. Evidence of association was identified for genetic variants in several candidate genes, among which were two novel genes (*mGST3* and *IDH3B*) that interacted with smoking in both races/ethnicities.

Second, to expand the scope of investigation to all common genetic variants

throughout the entire human genome, we conducted a large-scale meta-analysis of genome-wide association studies of longitudinal change in lung function in a consortium of 14 individual cohort studies of adults of European ancestry. We found evidence of association at two novel genetic loci (*IL16/STARD5/TMC3* and *ME3*) in the meta-analysis and performed additional gene expression analyses to demonstrate that both loci harbor candidate genes with biologically plausible functional links to lung function.

Finally, we explored the role of nutrition directly by investigating the relation between overall dietary patterns and longitudinal change in lung function in a prospective cohort of male adults, considering diet-by-smoking interaction. We identified two distinct dietary patterns by applying principal component analysis to food frequency questionnaire data, and found that a prudent diet rich in fruits, vegetables, fish, and poultry attenuated the accelerated decline in lung function in cigarette smokers, but had no association in non-smokers.

BIOGRAPHICAL SKETCH

Wenbo Tang was born in Beijing, China, and enjoyed a care-free and possibly over-privileged childhood as the single child in the entire Tang family. Although not exactly physically gifted, he pursued a broad range of sport activities, particularly competitive swimming throughout his youthful years and was once crowned the unofficial king of frog style by his mostly well-meaning teammates. With the hope to compete in the 2008 Beijing Olympic Games becoming unlikely the moment he stopped growing at 5'7", he made the first life-changing decision to pursue a career in life sciences by attending Peking University Health Science Center in Beijing, China. There he majored in Public Health and received comprehensive training in not only his major concentration but also clinical medicine and basic biomedical sciences. Having allegedly fainted a few times in the surgery room during his clinical internship, Wenbo vowed to continue his study of human health in a fashion that is as far away as possible, both physically and conceptually, from the clinical setting. He then made the second life-changing decision to come to the United States for graduate school and began his studies at the Division of Nutritional Sciences at Cornell University in August, 2008. Since then, he has enjoyed a productive and rewarding learning and research experience as a member in Dr. Pat Cassano's research group, where his work focuses on the impact of genetic, nutritional, environmental factors on lung health and diseases, particularly the longitudinal change in lung function over time among elderly adults.

While at Cornell, Wenbo married his middle and high school sweetheart, Xuan Dong, a Cornell alumna in Statistical Sciences. They have immensely enjoyed their adventures in Ithaca and are embarking on a new journey in Boston, Massachusetts, where Xuan is working

as a statistician in the healthcare industry and Wenbo will be pursuing a career in epidemiology to utilize his interdisciplinary skills in nutrition, epidemiology, and genomics to contribute to public health.

This work is dedicated to:

- My parents, Wei Tang and Yinghong Li, for their consistent and unconditional love, support, and encouragement without which I wound not be who I am today.
- My grandparents and many relatives for their love and support.
- Xuan Dong, the love of my life, my best friend and soul mate. You give colors to my life, bring sunshine to my heart, and inspire me every day to be a better man.

ACKNOWLEDGMENTS

First and foremost, I would like to sincerely thank my advisor, Pat Cassano, for your wonderful mentorship over the past five years. Your enthusiasm for science, research, and knowledge as well as your academic rigor inspire me every day. I attribute my professional development as an epidemiologist to your masterful guidance and training. I respect, value, and admire you not only as an advisor for research, but also as a mentor for life. I will be forever grateful for your immeasurable impact on my life, both professionally and personally.

I would like to sincerely thank my committee members, Marty Wells, Andy Clark, and Marie Caudill, for your guidance and support throughout the course of my dissertation research. I looked forward to every one of our meetings, and have been inspired to be a better scientist by every conversation we have had.

I would like to thank the many members of the Cassano research group, past and present, who are among the first friends I have made after coming to the US and have made my graduate experience at Cornell such an enjoyable journey.

I would like to thank the Cornell Statistical Consulting Unit, particularly Francoise Vermeylen, who graciously answered my many statistical questions, however little sense they made.

I would like to thank the many collaborators that I have gotten the chance to work with in the CHARGE and SpiroMeta consortia, as well as the study participants in the Health, Aging, and Body Composition study and the Respiratory Ancillary Study, who made this dissertation research possible.

Lastly, I would like to acknowledge Cornell University, the Division of Nutritional

Sciences, and the National Institutes of Health, for financially supporting my research and life in the past five years.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

Lung function is an important predictor of morbidity and mortality in the general population (1). Spirometric measures of lung function, such as forced expiratory volume in the first second (FEV_1) and the ratio of FEV_1/forced vital capacity (FEV_1/FVC) are easily measured and reliable indicators of the physiological state of the lungs and airways and provide the basis for diagnosing and staging chronic obstructive pulmonary disease (COPD) (2). COPD, which includes emphysema and chronic bronchitis, is characterized by the development of airflow limitation that is not fully reversible, and includes symptoms such as shortness of breath, decreased exercise capability, wheezing, recurrent respiratory infections, severe cough, and poor overall oxygenation. COPD is a major cause of morbidity and mortality, currently ranked the third leading cause of death in the United States and projected to create an even greater public health burden in the coming years (3-5). Currently, available therapies for COPD are extremely limited in number and effectiveness, and most do not slow the rate of decline in FEV_1 (6, 7). Thus, etiologic research that identifies novel preventive and curative strategies is a high public health priority (8, 9).

Lung function reaches its peak levels during early adulthood, followed by a plateau, and declines subsequently as part of the natural aging process. However, accelerated decline in lung function can occur due to exposures to detrimental factors, such as cigarette smoke, leading to reduced lung function levels that characterize COPD (10, 11). Therefore, the study of longitudinal change in lung function may provide important insights for better

understanding COPD pathogenesis, and contribute to primary and secondary prevention strategies (10-13).

Cigarette smoke, a major source of exogenous oxidants, exposes the lung to elevated levels of oxidative stress, and is widely recognized as the most important risk factor for accelerated loss in lung function and COPD (14, 15). However, only a subset of smokers eventually develop COPD (16), and about 10-15% of COPD cases cannot be explained by smoking (17). In addition, both cross-sectional and longitudinal studies have shown that as much as 90% of the overall variation in lung function remains unexplained after accounting for the effects of age, height, and smoking (18-21). These observations together suggest the importance of other types of risk factors, such as genetics and nutrition, in contributing to variations in lung function and COPD susceptibility.

One prominent theory regarding the etiology of COPD postulates that the imbalance between chronic oxidative stress and antioxidant protection plays a key role in accelerated lung function loss (22, 23). In this context, two major forms of antioxidant defense are endogenous antioxidant enzymes and exogenous antioxidants from the diet, and both of them have been hypothesized to affect lung function and COPD risk, either independently or through interaction with cigarette smoking. On one hand, genetic variation in antioxidant enzymes has been studied through the candidate gene association approach using population data, but published studies have limitations, such as insufficient coverage of plausible genes, limited consideration of gene-by-smoking interaction, and inadequate consideration of longitudinal lung function outcomes to support stronger causal inferences. On the other hand, nutritional epidemiologic studies have also yielded substantial evidence of associations between the intake of numerous individual nutrients and foods and lung function outcomes

(24-28). However, most of these studies employed the "single nutrient" approach that does not sufficiently account for the complexity of dietary intake in terms of the strong correlations among individual nutrients and the presence of interactive or synergistic effects among foods and/or nutrients that are consumed in combinations (29).

Beyond the scope of the oxidant/antioxidant hypothesis, substantial advances regarding the genetic contributions to lung function and COPD have been achieved in recent years as the field of genetic epidemiology entered the genome-wide association study (GWAS) era. To date, more than a dozen GWAS of lung function- and COPD-related outcomes have been published. Among them, three recent large-scale GWAS meta-analyses together identified 26 novel genetic loci in association with cross-sectional lung function levels (30-32). Many implicated genes in these loci possess biological functions that were previously considered irrelevant or unrelated to lung health. Thus these discoveries provide the important first step in the development of new hypothesis regarding the genetic regulation of lung function and genetic susceptibility to COPD. However, the majority of these GWAS focus on cross-sectional lung function outcomes, and to date only one population-based study of longitudinal change in lung function has been published. Due to the small sample size (1,441 asthmatic and 2,667 non-asthmatic participants), the study only reported suggestive evidence of association at one novel locus. Clearly, additional GWAS of longitudinal change in lung function with much greater sample size is needed.

In the above research context, this dissertation aims to examine the hypothesis that genetic and nutritional factors, either independently or through interaction with cigarette smoking, contribute to variability in the longitudinal change in lung function, using both hypothesis-driven and hypothesis-generating strategies.

The first project extended previous work on the association between genetic variation in antioxidant enzymes and lung function outcomes by including a comprehensive set of candidate genes selected following a network-driven approach and studying their association with longitudinal change in lung function. Genotyping was performed for 384 single-nucleotide polymorphisms (SNPs) in 56 candidate genes encoding antioxidant enzymes, in a prospective cohort of 1,281 African American and 1,794 European American elderly adults from the Health, Aging, and Body Composition (Health ABC) study. With the use of linear mixed effects models, SNPs were explored for single-marker association and gene-by-smoking interaction in relation to the rate of change in FEV_1 and FEV_1/FVC, separately in African and European Americans. The results of this hypothesis-driven study are described in Chapter 2.

The second project was designed to further expand the scope of investigation by considering common SNPs throughout the human genome using the GWAS approach and by synthesizing GWAS results from 14 individual cohort studies using meta-analysis. With the use of linear mixed effects models, ~2.5M common SNPs were examined for association with the rate of change in FEV_1 in a combined sample of over 27,000 adults of European ancestry from 14 prospective cohort studies with longitudinal lung function data, and cohort-specific association results were combined using fixed effect meta-analysis. Additional gene expression analyses were performed for candidate genes at the identified loci to demonstrate their functional links to lung function. The results of this hypothesis-generating study are described in Chapter 3.

The third project extended the current literature on the relation between diet and lung health by characterizing dietary intake using overall dietary patterns and investigating their

association with FEV_1 longitudinally. Dietary patterns were derived using principal component analysis based on food frequency questionnaire data collected in 2,560 male adults from the Respiratory Ancillary Study. With the use of linear mixed effects models, identified dietary patterns were explored for association with the rate of change in FEV_1 over ~3 years, adjusting for potential confounders. Given the importance of cigarette smoking, interaction between dietary patterns and smoking were also examined. The results of this hypothesis-driven study are described in Chapter 4.

Overall, the three projects comprising this dissertation research contribute to an improved understanding of the role of, and interaction between, genetics, nutrition, and cigarette smoking in determining the longitudinal change in lung function as an informative indicator of COPD risk. The first two projects present important data regarding the contribution of genetic factors both in and beyond the oxidant/antioxidant system in affecting lung function and disease risk, while the third project demonstrates the value of dietary pattern studies as a new avenue for future research of nutrition and lung health, which could ultimately contribute to studies aimed at investigating the interaction of nutrition with the genome in relation to lung outcomes.

REFERENCES

1. Schunemann HJ, Dorn J, Grant BJ, et al. Pulmonary function is a long-term predictor of mortality in the general population: 29-year follow-up of the Buffalo Health Study. *Chest*. 2000;118(3):656-64.

2. Rabe KF, Hurd S, Anzueto A, et al. Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: GOLD executive summary. *Am J Respir Crit Care Med*. 2007;176(6):532-55.

3. Lopez AD, Murray CC. The global burden of disease, 1990-2020. *Nature medicine*. 1998;4(11):1241-3.

4. Michaud CM, Murray CJ, Bloom BR. Burden of disease--implications for future research. *JAMA*. 2001;285(5):535-9.

5. Minino AM, Murphy SL. Death in the United States, 2010. *NCHS data brief*. 2012(99):1-8.

6. Anthonisen NR, Connett JE, Murray RP. Smoking and lung function of Lung Health Study participants after 11 years. *Am J Respir Crit Care Med*. 2002;166(5):675-9.

7. Croxton TL, Weinmann GG, Senior RM, et al. Clinical research in chronic obstructive pulmonary disease: needs and opportunities. *Am J Respir Crit Care Med*. 2003;167(8):1142-9.

8. Faulkner MA, Hilleman DE. The economic impact of chronic obstructive pulmonary disease. *Expert opinion on pharmacotherapy*. 2002;3(3):219-28.

9. Sullivan SD, Ramsey SD, Lee TA. The economic burden of COPD. *Chest*. 2000;117(2 Suppl):5S-9S.

10. Fletcher C. The natural history of chronic bronchitis and emphysema: an eight-year study of early chronic obstructive lung disease in working men in London. Oxford University Press. 1976.

11. Rennard SI, Vestbo J. Natural Histories of Chronic Obstructive Pulmonary Disease. *Proceedings of the American Thoracic Society*. 2008;5(9):878-83.

12. Wise R. The value of forced expiratory volume in 1 second decline in the assessment of chronic obstructive pulmonary disease progression. *Am J Med*. 2006;119(10):4-11.

13. Han MK, Agusti A, Calverley PM, et al. Chronic obstructive pulmonary disease phenotypes: The future of COPD. *Am J Respir Crit Care Med*. 2010;182(5):598-604.

14. Lindberg A, Bjerg A, Ronmark E, et al. Prevalence and underdiagnosis of COPD by disease severity and the attributable fraction of smoking Report from the Obstructive Lung Disease in Northern Sweden Studies. *Respir Med*. 2006;100(2):264-72.

15. Mannino DM, Buist AS, Petty TL, et al. Lung function and mortality in the United States: data from the First National Health and Nutrition Examination Survey follow up study. *Thorax*. 2003;58(5):388-93.

16. Halbert RJ, Natoli JL, Gano A, et al. Global burden of COPD: systematic review and meta-analysis. *Eur Respir J*. 2006;28(3):523-32.

17. Eisner MD, Anthonisen N, Coultas D, et al. An official American Thoracic Society public policy statement: novel risk factors and the global burden of chronic obstructive pulmonary disease. *Am J Respir Crit Care Med*. 2010;182(5):693-718.

18. Beaty TH, Menkes HA, Cohen BH, et al. Risk factors associated with longitudinal change in pulmonary function. *Am Rev Respir Dis*. 1984;129(5):660-7.

19. Beck GJ, Doyle CA, Schachter EN. Smoking and lung function. *Am Rev Respir Dis*. 1981;123(2):149-55.

20. Kauffmann F, Drouet D, Lellouch J, et al. Twelve years spirometric changes among Paris area workers. *Int J Epidemiol*. 1979;8(3):201-12.

21. Peat JK, Woolcock AJ, Cullen K. Decline of lung function and development of chronic airflow limitation: a longitudinal study of non-smokers and smokers in Busselton, Western Australia. *Thorax*. 1990;45(1):32-7.

22. Rahman I, Adcock IM. Oxidative stress and redox regulation of lung inflammation in COPD. *Eur Respir J*. 2006;28(1):219-42.

23. Rahman I, Biswas S, Kode A. Oxidant and antioxidant balance in the airways and airway diseases. *Eur J Pharmacol*. 2006;533(1-3):222-39.

24. Hu G, Zhang X, Chen J, et al. Dietary vitamin C intake and lung function in rural China. *Am J Epidemiol*. 1998;148(6):594-9.

25. Butland BK, Fehily AM, Elwood PC. Diet, lung function, and lung function decline in a cohort of 2512 middle aged men. *Thorax*. 2000;55(2):102-8.

26. McKeever TM, Scrivener S, Broadfield E, et al. Prospective study of diet and decline in lung function in a general population. *Am J Respir Crit Care Med*. 2002;165(9):1299-303.

27. Ochs-Balcom HM, Grant BJ, Muti P, et al. Antioxidants, oxidative stress, and pulmonary function in individuals diagnosed with asthma or COPD. *Eur J Clin Nutr*. 2006;60(8):991-9.

28. Bentley AR, Kritchevsky SB, Harris TB, et al. Dietary antioxidants and forced expiratory volume in 1 s decline: the Health, Aging and Body Composition study. *Eur Respir J*. 2012;39(4):979-84.

29. National Research Council (U.S.). Committee on Diet and Health. Diet and health : implications for reducing chronic disease risk. Washington, D.C.: National Academy Press; 1989. xiv, 749 p. p.

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*. 2009;42(1):45-52.

31. Repapi E, Sayers I, Wain LV, et al. Genome-wide association study identifies five loci associated with lung function. *Nat Genet*. 2009;42(1):36-44.

32. Soler Artigas M, Loth DW, Wain LV, et al. Genome-wide association and large-scale follow up identifies 16 new loci influencing lung function. *Nat Genet*. 2011;43(11):1082-90.

CHAPTER 2

GENETIC VARIATION IN ANTIOXIDANT ENZYMES, CIGARETTE SMOKING AND LONGITUDINAL CHANGE IN LUNG FUNCTION

Wenbo Tang[1], Amy R. Bentley[1,2], Stephen B. Kritchevsky[3], Tamara B. Harris[4], Anne B. Newman[5], Douglas C. Bauer[6], Bernd Meibohm[7], and Patricia A. Cassano[1,8] for the Health ABC Study

[1]Division of Nutritional Sciences, Cornell University, Ithaca, NY, USA

[2]Center for Research in Genomics and Global Health, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD, USA

[3]Sticht Center on Aging, Wake Forest University School of Medicine, Winston-Salem, NC, USA

[4]Intramural Research Program, National Institute on Aging, Laboratory of Epidemiology, Demography, and Biometry, Bethesda, MD, USA

[5]Center for Aging and Population Health, University of Pittsburgh, PA, USA

[6]Departments of Medicine and Epidemiology & Biostatistics, University of California San Francisco, San Francisco, CA, USA

[7]University of Tennessee, Memphis, TN, USA

[8]Division of Biostatistics and Epidemiology in the Department of Public Health, Weill Cornell Medical College, NY, NY, USA

Corresponding Author: Patricia A. Cassano, Division of Nutritional Sciences, Cornell University, 209 Savage Hall, Ithaca, NY 14853, pac6@cornell.edu, (607) 255-7551 (phone), (607) 255-2691 (fax).

This work is published at the journal of Free Radical Biology and Medicine:

Tang W, Bentley AR, Kritchevsky SB, Harris TB, Newman AB, Bauer DC, Meibohm B, Cassano PA; for the Health ABC study. Genetic variation in antioxidant enzymes, cigarette smoking, and longitudinal change in lung function. Free Radic Biol Med. 2013 May 17;63C:304-312.

ABSTRACT

Rationale: Antioxidant enzymes play an important role in the defense against oxidative stress in the lung and in the pathogenesis of chronic obstructive pulmonary disease (COPD). Sequence variation in genes encoding antioxidant enzymes may alter susceptibility to COPD by affecting longitudinal change in lung function in adults.

Methods: We genotyped 384 sequence variants in 56 candidate genes in 1,281 African American and 1,794 European American elderly adults in the Health, Aging, and Body Composition study. Single-marker associations and gene-by-smoking interactions with rate of change in FEV_1 and FEV_1/ FVC were evaluated using linear mixed effects models, stratified by race/ethnicity.

Results: In European Americans, rs17883901 in *GCLC* was statistically significantly associated with rate of change in FEV_1/FVC; the recessive genotype (*TT*) was associated with a 0.9% per year steeper decline ($P = 4.50 \times 10^{-5}$). Statistically significant gene-by-smoking interactions were observed for variants in two genes in European Americans: the minor allele of rs2297765 in *mGST3* attenuated the accelerated decline in FEV_1/FVC in smokers by 0.45% per year ($P = 1.13 \times 10^{-4}$); for participants with greater baseline smoking pack-years, the minor allele of rs2073192 in *IDH3B* was associated with an accelerated decline in FEV_1/FVC ($P = 2.10 \times 10^{-4}$). For both genes, nominally significant interactions ($P < 0.01$) were observed at the gene-level in African Americans ($P = 0.007$ and 4.60×10^{-4}, respectively). Nominally significant evidence of association was observed for variants in *SOD3* and *GLRX2* in multiple analyses.

Conclusions: This study identifies two novel genes associated with longitudinal lung function phenotypes in both African and European Americans, and confirms a prior finding for *GCLC*.

These findings suggest novel mechanisms and molecular targets for future research and advance the understanding of genetic determinants of lung function and COPD risk.

Keywords: Antioxidant enzymes, cigarette smoking, gene by environment interaction, genetic association, longitudinal change, lung function, oxidative stress

INTRODUCTION

Lung function is an important predictor of morbidity and mortality in the general population (1). Spirometric measures of lung function, such as forced expiratory volume in the first second (FEV_1) and the ratio of FEV_1/forced vital capacity (FEV_1/FVC) are easily measured and reliable indicators of the physiological state of the lungs and airways and provide the basis for diagnosing and staging chronic obstructive pulmonary disease (COPD) (2). Decline in lung function occurs naturally with aging, but accelerated decline can be caused by exposures such as cigarette smoking and can lead to low lung function that characterizes COPD (3, 4). Therefore, longitudinal changes in lung function are informative predictors of COPD risk, and studies of these outcomes provide important insights for understanding disease pathogenesis (3-6).

The imbalance between chronic oxidative stress and antioxidant protection is postulated to play a key role in accelerated lung function loss (7, 8). Cigarette smoke, a major source of exogenous oxidants, exposes the lung to elevated levels of oxidative stress, whereas dietary antioxidants and endogenous antioxidant enzymes are the two major forms of antioxidant defense that counteract these processes. The observation that only a subset of smokers develop COPD (9) and that a substantial proportion of COPD cases cannot be explained by smoking (10) led to the hypothesis that dietary intake of antioxidants and genetic variation in genes encoding antioxidant enzymes both play an important role in modifying antioxidant defense against cigarette smoke in the lung with ultimate effects on COPD risk.

In support of this hypothesis, observational epidemiologic studies have provided evidence of a positive association between dietary antioxidant intake and lung function, with stronger effects in cigarette smokers (11-15). Genetic variation in antioxidant enzymes has

also been studied in candidate gene association studies using population data, but published studies have limitations (16). First, most studies considered only limited numbers of candidate genes, leaving many biologically relevant genes unstudied. Second, very few studies considered longitudinal lung function phenotypes (17-19). Third, despite compelling evidence for their importance (20, 21), gene-by-smoking interactions are rarely investigated. Finally, very few studies include individuals of non-European ancestry, limiting inference to individuals of European descent. While recent, large-scale genome-wide association studies (GWAS) of lung function phenotypes together identified numerous novel genetic loci, these studies are limited in that they only consider European ancestry and cross-sectional phenotypes (22-24).

We hypothesized that common single nucleotide polymorphisms (SNPs) in genes encoding antioxidant enzymes affect longitudinal decline in lung function. We further hypothesized that gene-by-smoking interactions are present such that some genetic variants affect lung function decline contingent on exposure to cigarette smoke. To investigate these hypotheses, we selected 56 candidate genes that either had putative functional relevance to antioxidant defense in the lung or were previously investigated in relation to COPD-related phenotypes. Functional and tagging SNPs in these genes were genotyped and tested for single-marker associations and gene-by-smoking interactions with rate of change in FEV_1 and FEV_1/FVC in a population of African American and European American elderly adults from the Health, Aging, and Body Composition (Health ABC) study.

MATERIAL AND METHODS

Subjects

The Health ABC study is a longitudinal, prospective cohort study comprising 1,281 African American and 1,794 European American community-dwelling men and women, aged 70-79 years at baseline (1996-1997) and residing in the metropolitan areas of Pittsburgh, PA and Memphis, TN (25). Participants reported self-proclaimed race initially as “Black” or “White”, but the terms “African American” and “European American” are used herein. To be eligible, participants were required to be ambulatory at baseline as confirmed by self-report of no difficulty walking one-quarter of a mile or climbing 10 steps without resting, no difficulty performing basic activities of daily living, and no use of a cane, walker, crutches or other special equipment to ambulate. In addition, participants were required to have no history of active treatment for cancer in the prior 3 years, and no plan to move out of the area in the subsequent 3 years. The Health ABC study was approved by the Institutional Review Boards of the University of Pittsburgh and the University of Tennessee, and the work reported herein was approved by the Institutional Review Board for Human Participants at Cornell University.

Pulmonary Function Testing

Spirometry was completed at four time points (baseline, years 4, 7 and 9) in accordance with standardized guidelines of the American Thoracic Society (ATS), as previously reported (25). The study used a horizontal, dry rolling seal HF6 Spirometer (Sensor Medics Corporation, Yorba Linda, CA, USA) during clinical visits, and the EasyOne Model 2001 diagnostic spirometer (ndd Medizintechnik AG, Zurich, Switzerland) during home visits starting in year 8. The two devices were evaluated for comparability and provided virtually identical values. Consistent with the quality control standard used in recent lung

function GWAS (22-24), all FEV_1 (mL) and FEV_1/FVC (%) measures meeting the ATS criteria for acceptability were included in the current study.

Cigarette Smoking

Participants were classified based on their long-term smoking status during the study follow-up as: (1) never smokers (never smoker at all spirometry time points), who were considered as the reference group in analyses, (2) persistent smokers (current smoker at all time points), (3) former smokers (former smoker at all time points), and (4) intermittent smokers (changing smoking status at different time points). Lifetime smoking dose was quantified as pack-years and calculated at study baseline for current and former smokers.

Candidate Gene Selection and Genotyping

Based on a previous systematic review of genetic association studies and gene expression studies investigating antioxidant enzymes and COPD-related phenotypes (16), we identified 56 candidate genes encoding antioxidant enzymes known to be expressed in lung tissue and postulated to affect the balance of antioxidants/oxidants. 384 functional and tagging SNPs were selected to capture variation across each gene and its regulatory regions (2 kilobases upstream and downstream). Details of the SNP selection strategy are provided elsewhere (26). Separate consideration was given to African Americans and European Americans in SNP selection to maximize coverage in both populations, given differences in linkage disequilibrium (LD) structure and allele frequencies. Details of DNA extraction and genotyping quality, which were excellent, are provided elsewhere (26).

Four genes (*GGT2*, *GSTK1*, *GSTM1*, and *GSTT1*) were excluded from subsequent

analyses due to low genotyping quality or atypical clustering of assayed SNPs. For the remaining SNPs with successful genotyping, Hardy-Weinberg equilibrium (HWE) was tested using the chi-squared goodness-of-fit test, stratified by race. After removing SNPs with genotyping call rate < 95%, minor allele frequency (MAF) < 1%, or p-value < 0.005 for the HWE test, the study included 314 SNPs in 52 genes in the African American analyses and 284 SNPs in the same 52 genes in the European American analyses (Supplementary Table 2.6).

Statistical Analysis

Linear mixed effects models were used to investigate single-marker associations and gene-by-smoking interactions with rate of change in FEV_1 and rate of change in FEV_1/FVC; all analyses were stratified by race/ethnicity. A continuous time variable quantified the time elapsed between each spirometry test and the study baseline. Random intercept and time effects were included at the individual level to differentiate between- and within-individual variation. All models were adjusted for gender, study site, height at each time point, age and smoking pack-years (both at study baseline), smoking status and smoking status × time. To address potential confounding by population substructure, the first two principal component variables for genetic ancestry (27) (computed separately by race/ethnicity; based on data from GWAS completed in Health ABC) were included in all models.

Single-marker associations with change in pulmonary function were tested by evaluating the product term of SNP × time. Gene-by-smoking interactions were tested by evaluating the three-way product term of SNP × smoking × time. Two smoking variables, smoking status during follow-up and baseline smoking pack-years, were tested separately for

interactions, with smoking status during follow-up collapsed into two categories, as follows: smokers (persistent + intermittent) and non-smokers (former + never, which comprised the reference group).

Each SNP was coded by the minor allele and analyzed using an additive genetic model. SNPs with a nominal $P < 0.05$ were further tested using the dominant and recessive genetic models to refine estimates of the underlying genetic effect. The effect estimates for the genetic model with the most significant association were reported. To maintain statistical validity, we presented findings only for SNPs with a participant count ≥ 10 for the least frequent genotype category in the single-marker analyses and for the least frequent genotype-smoking status category in the interaction analyses.

In genetic association studies, the risk of false positives must be minimized without ruling out true associations. GWAS-scale multiple testing adjustments are not appropriate for the hypothesis-based investigation of candidate genes reported herein. Given the presence of LD among analyzed SNPs, we controlled for multiple testing using a Bonferroni adjustment based on the effective number of independent tests (M_{eff}) (28, 29). M_{eff} was computed based on the correlation matrix of genotypes of all analyzed SNPs, and then used in a Bonferroni adjustment at the experiment-wise α level of 0.05. Given the difference in LD patterns, the adjustment was performed separately for each race/ethnicity. For African Americans (M_{eff} = 223), the Bonferroni-corrected significance threshold was $P < 2.3 \times 10^{-4}$; for European Americans (M_{eff} = 171), the analogous threshold of $P < 3.0 \times 10^{-4}$ was used. In addition, nominally significant associations were defined using $P < 0.005$ for single-marker analyses and $P < 0.01$ for gene-by-smoking interaction analyses.

All statistical analyses were conducted using SAS software version 9.1 (SAS Institute, Cary, NC, USA). LD in the Health ABC population was evaluated using Haploview 4.2 (30).

RESULTS

Population Characteristics

After exclusion for missing covariate data, 1,022 African Americans with 2,432 FEV_1 measurements and 1,487 European Americans with 4,157 FEV_1 measurements were included in the FEV_1 analysis (Table 2.1). Similarly, 979 African Americans with 2,244 FEV_1/FVC measurements and 1,469 European Americans with 4,018 FEV_1/FVC measurements were included in the FEV_1/FVC analysis.

We observed statistically significant annual decline in FEV_1 and statistically significant annual decline in FEV_1/FVC in both African Americans and European Americans (Table 2.2). For never smokers, the estimated rate of decline in FEV_1/FVC was about 0.5% per year in both African Americans and European Americans, while the estimated annual decline in FEV_1 was greater in European Americans (about 40 versus 32 mL per year). In general, the effects of smoking on lung function were stronger in European Americans compared to African Americans, consistent with greater smoking doses observed in the former group. Thus, for FEV_1/FVC, while persistent and intermittent smokers had significantly faster declines in both groups compared to never smokers, the effect size of persistent smoking in European Americans was about twice that in African Americans. While the difference in FEV_1/FVC decline between former smokers and never smokers was not significant in African Americans, it was borderline significant in European Americans. Similar patterns were observed for annual decline in FEV_1, although not all associations were

statistically significant; the *P* value for the persistent smoking association was 0.06 in European Americans, and all associations followed expectations for magnitude and size of effect.

Single-Marker Associations

Three genes showed nominal evidence of associations with rate of change in FEV_1 in African Americans and one gene was associated with FEV_1 decline in European Americans, although no associations survived the adjustment for multiple testing (Table 2.3). However, a SNP in *superoxide dismutase 3* (*SOD3*), rs8192287, was marginally statistically significantly associated with a 20 mL per year faster decline in FEV_1 per copy of the minor allele (*T*) in African Americans ($P = 2.45 \times 10^{-4}$).

Four genes were nominally associated with rate of change in FEV_1/FVC in African Americans (Table 2.4). The most statistically significant association, which did not pass the Bonferroni-adjusted threshold, was for a *glutaredoxin 2* (*GLRX2*) SNP, rs35358794; each copy of the minor allele (*A*) was associated with a 0.3% per year slower decline. In European Americans, two genes showed evidence of associations with rate of change in FEV_1/FVC. The most statistically significant association, which survived the Bonferroni adjustment for multiple testing, was for the *glutamate-cysteine ligase catalytic subunit* (*GCLC*) SNP rs17883901 ($P = 4.50 \times 10^{-5}$). The recessive genotype (*TT*) was associated with a 0.9% per year steeper decline compared with the reference genotypes (*CC/CT*).

Gene-by-Smoking Interactions

Potential interactions between SNPs and cigarette smoking were investigated in

relation to rate of change in FEV_1 and FEV_1/FVC separately in African Americans and European Americans. Two smoking variables, smoking status during follow-up and baseline smoking pack-years, were investigated separately for gene-by-smoking interactions (Table 2.5; also Supplementary Tables 2.7 to 2.14).

For rate of change in FEV_1, in African Americans, a nominally significant interaction was identified between rs34552619 in *GLRX2* and smoking status. African American smokers with at least one copy of the minor allele (*C*) had a 32.4 mL per year steeper decline ($P = 2.74 \times 10^{-4}$) than smokers without the minor allele. In European Americans, rs1007991 in *SOD3* had a nominally significant interaction with smoking status; each copy of the minor allele (*C*) attenuated the accelerated decline in FEV_1 in smokers by 17.9 mL per year ($P = 0.002$). In contrast, neither SNP was associated with rate of change in FEV_1 in non-smokers during follow-up.

For rate of change in FEV_1/FVC, two genes had statistically significant interactions with smoking that passed the Bonferroni adjustment for multiple testing in European Americans and gene-level replications were observed for both genes in African Americans. In European Americans, the association between rs2297765 in *microsomal glutathione S-transferase 3* (*mGST3*) and rate of change in FEV_1/FVC differed by smoking status such that each copy of the minor allele (*T*) attenuated the decline in smokers by 0.45% per year, but the SNP had no effect on decline in non-smokers (Table 2.5; $P = 1.13 \times 10^{-4}$; Figure 2.1). In African Americans, a different *mGST3* variant, rs7554034, had a nominally significant interaction with smoking status; compared to the reference genotype, the recessive genotype (*AA*) attenuated the decline in smokers by 0.56% per year, but genotype had no effect on rate of decline in non-smokers (Table 2.5; $P = 0.007$; Figure 2.2). These *mGST3* SNPs were not in

LD in either group (r^2 = 0.002 and 0.01 for African Americans and European Americans, respectively). In European Americans, the association between rs2073192 in *isocitrate dehydrogenase 3 beta* (*IDH3B*) and rate of change in FEV_1/FVC differed by smoking pack-years; in participants with higher smoking dose, the minor allele (*A*) was associated with a faster decline in FEV_1/FVC (Table 2.5, $P = 2.10 \times 10^{-4}$). Two other *IDH3B* SNPs, which were in strong LD with rs2073192 in European Americans ($r^2 \geq 0.92$), showed similar, but less statistically significant, evidence of interaction with smoking pack-years. In African Americans, a nominally significant interaction was observed for one of the *IDH3B* SNPs (rs6115381) and smoking status (Table 2.5; $P = 4.60 \times 10^{-4}$) such that the recessive genotype (*GG*) was associated with a 0.82% per year greater decline (compared to the reference genotype) in smokers only; no such difference was observed across genotypes in non-smokers (Figure 2.3). In African Americans, the rs6115381 and rs6107100 SNPs in *IDH3B* were in moderate LD (r^2 = 0.63), whereas rs2073192 was not in LD with rs6115381 and rs6107100 (r^2 = 0.09 and 0.10, respectively). A nominally significant interaction was also observed between rs2284659 in *SOD3* and smoking status in European Americans for rate of change in FEV_1/FVC (P = 0.004).

DISCUSSION

This study was designed to investigate the hypothesis that genetic variation in candidate genes encoding antioxidant enzymes, which is expected to affect antioxidant defense in the lung, is associated with rate of change in lung function phenotypes, FEV_1 and FEV_1/FVC, and thus contributes to COPD susceptibility, especially in individuals with elevated oxidative stress due to cigarette smoking. Consistent with several recent GWAS of

lung phenotypes there were more findings overall for rate of change in FEV_1/FVC compared to rate of change in FEV_1, although the reasons for this are not yet clear (22-24).

A novel gene, *mGST3*, was associated with rate of change in FEV_1/FVC, with evidence of gene-by-smoking interactions in both European and African Americans. In European Americans, the effect of rs2297765 differed significantly by smoking status, and rs7554034 had a similar interaction in African Americans. The two SNPs are common in both groups, thus the effects on lung function in smokers are of public health interest. MGST3 is a membrane-bound antioxidant enzyme in the microsomal GST family with close links to antioxidant defense. In microarray studies of gene expression, the mouse *Mgst3* gene was up-regulated in the small intestine and liver in response to oxidative stress (31). MGST3 also catalyzes the conjugation reaction that produces leukotriene C4, an important inflammation mediator with a role in allergy and asthma (32, 33). The association of *mGST3* with lung function phenotypes is novel, but microsomal enzymes, as a class, have been linked to lung health in prior studies. Epoxide hydrolase 1 (EPHX1), another microsomal enzyme, detoxifies xenobiotics including products in cigarette-smoke, and genetic variation in *EPHX1* was associated with pulmonary phenotypes including childhood asthma, lung cancer and COPD (34-36).

IDH3B was implicated in a prior study of cross-sectional lung function phenotypes in the Health ABC cohort (gene-by-smoking interactions in African Americans) (26). In the current study of longitudinal lung function phenotypes in the same cohort, SNPs in *IDH3B* had a statistically significant interaction with smoking pack-years in European Americans and a nominally significant interaction with smoking status in African Americans in relation to rate of change in FEV_1/FVC. The IDH enzymes, the majority of which localize to the

mitochondrial matrix, supply the reducing equivalents for the antioxidant activity of the many members of the glutathione and thioredoxin systems. In fibroblasts, decreased expression of *IDH* genes led to higher lipid peroxidation, oxidative DNA damage, intracellular peroxide generation, and increased senescence, indicating an important regulatory role for these genes in the defense against oxidative stress (37).

In the present study, the rs17883901 SNP in *GCLC* was associated with rate of change in FEV_1/FVC in European Americans, but the gene-by-smoking interactions for rs17883901 could not be investigated given the limited number of smokers carrying the minor allele. Although no association was detected in African Americans, there was a considerably lower MAF in African Americans (1%). These findings are consistent with prior reports. The rs17883901 SNP was associated with an increased risk of COPD in a Chinese population (38), and two variants (rs17883901 and a GAG repeat variant (TNR)) were investigated jointly in relation to several pulmonary phenotypes, including change in FEV_1, in two Dutch cohorts (18). Using a nominal significance threshold of $P < 0.05$, the Dutch study reported associations for both variants, including an interaction between TNR and smoking pack-years in relation to rate of change in FEV_1. *GCLC* encodes the catalytic subunit of the heterodimeric enzyme glutamate-cysteine ligase, which catalyzes the *de novo* synthesis of glutathione. *GCLC* is predominantly expressed in lung epithelium (39), and rs17883901 was associated with lower expression of *GCLC* in endothelial cells in vitro (40), suggesting a potential mechanism for the population-level association. Overall, these findings support a role for *GCLC* in longitudinal change in lung function.

SNPs in *SOD3* and *GLRX2* showed nominally significant evidence of associations in multiple analyses, and the *GLRX2* findings are novel. SOD3 is a major extracellular

antioxidant enzyme highly expressed in the lung; SOD3 binds lung matrix components (collagen I, hyaluronan and heparin sulfate) to protect them against oxidative fragmentation and plays a central role in antioxidant defense in lung tissue (41). Genetic variation in *SOD3* has been extensively studied in relation to pulmonary phenotypes at the population level, although primarily in individuals of European descent. Rs1799895, a rare functional SNP in *SOD3*, was associated with lower COPD risk and slower FEV_1 decline in never smokers (42-44), and rs8192287 and rs8192288, which are in strong LD in individuals of European descent, were associated with reduced lung function and increased emphysema risk (45, 46). Rs1799895 was analyzed in the present study for single-marker associations in European Americans, but no statistically significant associations emerged. Rs8192287 was associated with a faster decline in FEV_1 in African Americans, providing novel evidence for an association of rs8192287 with lung function phenotypes in this under-studied group. The observed gene-by-smoking interactions involving other *SOD3* SNPs support effect modification by smoking in this genotype—phenotype association. Novel findings emerged for *GLRX2*, which encodes a mitochondrial antioxidant enzyme in the glutaredoxin family and has been recognized as an important redox regulator (47). *GLRX2* is ubiquitously expressed in various tissues including lung (47), and its over-expression was shown to prevent H_2O_2-induced apoptosis in human lens epithelial cells and to reduce myocardial cell death by preventing apoptosis and necrosis in mice (48, 49).

The study has several strengths. First, this large, epidemiologic cohort study with longitudinal follow-up data on pulmonary function assessed by high-quality spirometry is a unique resource. The long duration allows the estimation of meaningful decline in lung function, making the investigation of the difference in rate of change among individuals

possible. The use of up to 4 repeated measurements per individual provides the data to accurately capture the true trajectory of lung function change over time. Second, the study had high-quality data on important risk and confounding factors, including cigarette smoking and principle component variables for genetic ancestry. The adjustment for genetic ancestry avoids potential confounding due to population substructure in each racial/ethnic group. Multiple forms of smoking data were available, allowing the consideration of both long-term smoking status and lifetime smoking dose in the single-marker analyses and to investigate potential interactions of genotype with these two different aspects of smoking exposure. Third, the Health ABC study includes a sufficiently large sample of African Americans, allowing race-specific analyses to be performed. This is important because African Americans have lower lung function compared with their European American counterparts and they are understudied in pulmonary and genetic epidemiology. Finally, despite the heterogeneity in the frequency and pattern of genetic variation and the challenges in the replication of genetic associations across racial/ethnic groups, this study provides compelling evidence of gene-by-smoking interactions consistent on the gene level between African Americans and European Americans for two novel candidate genes.

A few limitations should be considered when evaluating the study findings. First, despite the goal to comprehensively include genes encoding enzymes in relevant antioxidant pathways in the lung, a few genes did not pass genotyping quality control, and other enzymes with antioxidant activities may have been omitted inadvertently. Second, although the Health ABC study recorded extensive data on smoking behaviors, the statistical modeling of smoking in the study may not fully capture the effect of smoke exposure, possibly due to inaccuracy in participants' self-reports and uncertainty in defining the most relevant aspects of smoking in

affecting pulmonary function. Despite these limitations, and limited power due to sample size, we were able to identify meaningful gene-by-smoking interactions. Third, the analyzed SNPs admittedly provide imperfect coverage of genetic variation in the candidate genes. The SNPs showing significant results are therefore likely "proxies" of the true causal variants. Considering the incomplete linkage between these variants, the true associations of causal variants with the corresponding phenotypes are expected to be greater than what was observed. While the current study focused on an elderly population, given their disproportionately high risk of accelerated lung function loss and consequent morbidity and mortality, the findings may or may not generalize to younger populations, and additional studies are needed to test the reported associations in populations with different characteristics. Finally, due to the risk of false discovery inherent in genetic association studies, we adopted a conservative significance threshold that may be overly conservative.

In conclusion, this study explored genetic variation in candidate genes encoding antioxidant enzymes, cigarette smoking, and longitudinal change in two lung function phenotypes in African American and European American elderly adults. Evidence of association was observed for several novel genes. Of particular importance are the novel findings of gene-by-smoking interactions for *mGST3* and *IDH3B*, which were observed consistently at the gene level in both African Americans and European Americans. The findings for *GCLC* and *SOD3* strengthen existing knowledge and extend the evidence base by the novel consideration of longitudinal phenotypes and African Americans. Future research, especially in the understudied African American population, is warranted to further validate these findings and to elucidate the underlying molecular mechanisms.

Acknowledgements: none

Competing Interest: No competing interests

Funding: This research was supported by NIH R01HL071022 (PAC), and by NIA contracts N01AG62101, N01AG62103, and N01AG62106. The genome-wide association study was funded by NIA grant 1R01AG032098-01A1 to Wake Forest University Health Sciences and genotyping services were provided by the Center for Inherited Disease Research (CIDR). CIDR is fully funded through a federal contract from the National Institutes of Health to the Johns Hopkins University, contract number HHSN268200782096C. Genotyping services specific to this study were provided by the Johns Hopkins University under the U.S. Federal Government contract number N01HV48195. This research was supported in part by the Intramural Research Program of the National Human Genome Research Institute, National Institutes of Health (ARB).

Study Approval and Participant Consent: The Health ABC study was approved by Institutional Review Boards at the University of Pittsburgh (Pittsburgh, PA) and the University of Tennessee (Memphis, TN) and participants provided written informed consent for the study. The Institutional Review Board at Cornell University (Ithaca, NY) approved the current study.

Contributor Statement: WT, ARB and PAC designed this study and TH, SK, ABN, DCB, and BM designed the Health ABC study. WT and PAC analyzed data and wrote the

manuscript. All authors reviewed and approved the final manuscript.

Abbreviations: COPD, chronic obstructive pulmonary disease; FEV_1, forced expiratory volume in the first second; FVC, forced vital capacity; CAT, catalase; G6PD, glucose-6-phosphate dehydrogenase; GCLC, glutamate–cysteine ligase(catalytic subunit); GCLM, glutamate-cysteine ligase (modulatory subunit); GGT1, γ-glutamyl transferase 1; GLRX, glutaredoxin; GPX, glutathione peroxidase; GSR, glutathione reductase; GSS, glutathione synthetase; GST, glutathione S-transferase; HMOX, heme-oxygenase; IDH, isocitrate dehydrogenase; mGST, microsomal glutathione S-transferase; PRDX, peroxiredoxin; SEP, selenoprotein; SOD, superoxide dismutase; TXN, thioredoxin; TXNRD, thioredoxin reductase.

REFERENCES

1. Schunemann HJ, Dorn J, Grant BJ, et al. Pulmonary function is a long-term predictor of mortality in the general population: 29-year follow-up of the Buffalo Health Study. *Chest*. 2000;118(3):656-64.

2. Rabe KF, Hurd S, Anzueto A, et al. Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: GOLD executive summary. *Am J Respir Crit Care Med*. 2007;176(6):532-55.

3. Fletcher C. The natural history of chronic bronchitis and emphysema: an eight-year study of early chronic obstructive lung disease in working men in London. Oxford University Press. 1976.

4. Rennard SI, Vestbo J. Natural Histories of Chronic Obstructive Pulmonary Disease. *Proceedings of the American Thoracic Society*. 2008;5(9):878-83.

5. Wise R. The value of forced expiratory volume in 1 second decline in the assessment of chronic obstructive pulmonary disease progression. *Am J Med*. 2006;119(10):4-11.

6. Han MK, Agusti A, Calverley PM, et al. Chronic obstructive pulmonary disease phenotypes: The future of COPD. *Am J Respir Crit Care Med*. 2010;182(5):598-604.

7. Rahman I, Adcock IM. Oxidative stress and redox regulation of lung inflammation in COPD. *Eur Respir J*. 2006;28(1):219-42.

8. Rahman I, Biswas S, Kode A. Oxidant and antioxidant balance in the airways and airway diseases. *Eur J Pharmacol*. 2006;533(1-3):222-39.

9. Halbert RJ, Natoli JL, Gano A, et al. Global burden of COPD: systematic review and meta-analysis. *Eur Respir J*. 2006;28(3):523-32.

10. Eisner MD, Anthonisen N, Coultas D, et al. An official American Thoracic Society public policy statement: novel risk factors and the global burden of chronic obstructive pulmonary disease. *Am J Respir Crit Care Med*. 2010;182(5):693-718.

11. Hu G, Zhang X, Chen J, et al. Dietary vitamin C intake and lung function in rural China. *Am J Epidemiol*. 1998;148(6):594-9.

12. Butland BK, Fehily AM, Elwood PC. Diet, lung function, and lung function decline in a cohort of 2512 middle aged men. *Thorax*. 2000;55(2):102-8.

13. McKeever TM, Scrivener S, Broadfield E, et al. Prospective study of diet and decline in lung function in a general population. *Am J Respir Crit Care Med*. 2002;165(9):1299-303.

14. Ochs-Balcom HM, Grant BJ, Muti P, et al. Antioxidants, oxidative stress, and pulmonary function in individuals diagnosed with asthma or COPD. *Eur J Clin Nutr*. 2006;60(8):991-9.

15. Bentley AR, Kritchevsky SB, Harris TB, et al. Dietary antioxidants and forced expiratory volume in 1 s decline: the Health, Aging and Body Composition study. *Eur Respir J*. 2012;39(4):979-84.

16. Bentley AR, Emrani P, Cassano PA. Genetic variation and gene expression in antioxidant related enzymes and risk of COPD: a systematic review. *Thorax*. 2008;63(11):956-61.

17. He JQ, Ruan J, Connett JE, et al. Antioxidant gene polymorphisms and susceptibility to a rapid decline in lung function in smokers. *Am J Respir Crit Care Med*. 2002;166(3):323-8.

18. Siedlinski M, Postma DS, van Diemen CC, et al. Lung function loss, smoking, vitamin C intake, and polymorphisms of the glutamate-cysteine ligase genes. *Am J Respir Crit Care Med*. 2008;178(1):13-9.

19. Curjuric I, Imboden M, Schindler C, et al. HMOX1 and GST variants modify attenuation of FEF25-75% decline due to PM10 reduction. *Eur Respir J*. 2010;35(3):505-14.

20. Zhai G, Valdes AM, Cherkas L, et al. The interaction of genes and smoking on forced expiratory volume: a classic twin study. *Chest*. 2007;132(6):1772-7.

21. Gottlieb DJ, Wilk JB, Harmon M, et al. Heritability of longitudinal change in lung function. The Framingham study. *Am J Respir Crit Care Med*. 2001;164(9):1655-9.

22. Hancock DB, Eijgelsheim M, Wilk JB, et al. Meta-analyses of genome-wide association studies identify multiple loci associated with pulmonary function. *Nat Genet*. 2009;42(1):45-52.

23. Repapi E, Sayers I, Wain LV, et al. Genome-wide association study identifies five loci associated with lung function. *Nat Genet*. 2009;42(1):36-44.

24. Soler Artigas M, Loth DW, Wain LV, et al. Genome-wide association and large-scale follow up identifies 16 new loci influencing lung function. *Nat Genet*. 2011;43(11):1082-90.

25. Waterer GW, Wan JY, Kritchevsky SB, et al. Airflow limitation is underrecognized in well-functioning older people. *J Am Geriatr Soc*. 2001;49(8):1032-8.

26. Bentley AR, Kritchevsky SB, Harris TB, et al. Genetic variation in antioxidant enzymes and lung function. *Free Radic Biol Med*. 2012.

27. Price AL, Patterson NJ, Plenge RM, et al. Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics*. 2006;38(8):904-9.

28. Nyholt DR. A simple correction for multiple testing for single-nucleotide polymorphisms in linkage disequilibrium with each other. *Am J Hum Genet*. 2004;74(4):765-9.

29. Li J, Ji L. Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix. *Heredity*. 2005;95(3):221-7.

30. Barrett JC, Fry B, Maller J, et al. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics*. 2005;21(2):263-5.

31. Higgins LG, Hayes JD. Mechanisms of induction of cytosolic and microsomal glutathione transferase (GST) genes by xenobiotics and pro-inflammatory agents. *Drug Metab Rev*. 2011;43(2):92-137.

32. Jakobsson PJ, Mancini JA, Riendeau D, et al. Identification and characterization of a novel microsomal enzyme with glutathione-dependent transferase and peroxidase activities. *J Biol Chem*. 1997;272(36):22934-9.

33. Duroudier NP, Tulah AS, Sayers I. Leukotriene pathway genetics and pharmacogenetics in allergy. *Allergy*. 2009;64(6):823-39.

34. Hu G, Shi Z, Hu J, et al. Association between polymorphisms of microsomal epoxide hydrolase and COPD: results from meta-analyses. *Respirology*. 2008;13(6):837-50.

35. Li X, Hu Z, Qu X, et al. Putative EPHX1 enzyme activity is related with risk of lung and upper aerodigestive tract cancers: a comprehensive meta-analysis. *PLoS ONE*. 2011;6(3):e14749.

36. Tung KY, Tsai CH, Lee YL. Microsomal epoxide hydroxylase genotypes/diplotypes, traffic air pollution, and childhood asthma. *Chest*. 2011;139(4):839-48.

37. Kil IS, Huh TL, Lee YS, et al. Regulation of replicative senescence by NADP+ -dependent isocitrate dehydrogenase. *Free Radic Biol Med*. 2006;40(1):110-9.

38. Liu S, Li B, Zhou Y, et al. Genetic analysis of CC16, OGG1 and GCLC polymorphisms and susceptibility to COPD. *Respirology*. 2007;12(1):29-33.

39. Pierrou S, Broberg P, O'Donnell RA, et al. Expression of genes involved in oxidative stress responses in airway epithelial cells of smokers with chronic obstructive pulmonary disease. *Am J Respir Crit Care Med*. 2007;175(6):577-86.

40. Koide S, Kugiyama K, Sugiyama S, et al. Association of polymorphism in glutamate-cysteine ligase catalytic subunit gene with coronary vasomotor dysfunction and myocardial infarction. *J Am Coll Cardiol*. 2003;41(4):539-45.

41. Petersen SV, Oury TD, Ostergaard L, et al. Extracellular superoxide dismutase (EC-SOD) binds to type i collagen and protects against oxidative fragmentation. *J Biol Chem*. 2004;279(14):13705-10.

42. Juul K, Tybjaerg-Hansen A, Marklund S, et al. Genetically increased antioxidative protection and decreased chronic obstructive pulmonary disease. *Am J Respir Crit Care Med*. 2006;173(8):858-64.

43. Young RP, Hopkins R, Black PN, et al. Functional variants of antioxidant genes in smokers with COPD and in those with normal lung function. *Thorax*. 2006;61(5):394-9.

44. Siedlinski M, van Diemen CC, Postma DS, et al. Superoxide dismutases, lung function and bronchial responsiveness in a general population. *Eur Respir J*. 2009;33(5):986-92.

45. Dahl M, Bowler RP, Juul K, et al. Superoxide dismutase 3 polymorphism associated with reduced lung function in two large populations. *Am J Respir Crit Care Med*. 2008;178(9):906-12.

46. Sorheim IC, DeMeo DL, Washko G, et al. Polymorphisms in the superoxide dismutase-3 gene are associated with emphysema in COPD. *COPD*. 2010;7(4):262-8.

47. Lundberg M, Johansson C, Chandra J, et al. Cloning and expression of a novel human glutaredoxin (Grx2) with mitochondrial and nuclear isoforms. *J Biol Chem*. 2001;276(28):26269-75.

48. Nagy N, Malik G, Tosaki A, et al. Overexpression of glutaredoxin-2 reduces myocardial cell death by preventing both apoptosis and necrosis. *J Mol Cell Cardiol*. 2008;44(2):252-60.

49. Wu H, Xing K, Lou MF. Glutaredoxin 2 prevents H(2)O(2)-induced cell apoptosis by protecting complex I activity in the mitochondria. *Biochim Biophys Acta*. 2010;1797(10):1705-15.

Table 2.1 Characteristics of Study Participants, by Phenotype and Race/Ethnicity, for the Health ABC Study[a]

Characteristic	FEV_1 Phenotype		FEV_1/FVC Phenotype	
	African Americans	European Americans	African Americans	European Americans
No. of participants	1,022	1,487	979	1,469
No. of spirometry measurements	2,432	4,157	2,244	4,018
Males	441 (43.2)	773 (52.0)	431 (44.02)	766 (52.1)
Age at baseline (yr)	73.4 (2.9)	73.8 (2.9)	73.4 (2.9)	73.8 (2.9)
Height at baseline (cm)	165.3 (9.4)	166.7 (9.2)	165.5 (9.5)	166.7 (9.2)
Study site				
Memphis, TN	469 (45.9)	727 (48.9)	439 (44.8)	714 (48.6)
Pittsburgh, PA	553 (54.1)	760 (51.1)	540 (55.2)	755 (51.4)
Smoking status during follow-up				
Never smokers	448 (43.8)	649 (43.6)	423 (43.2)	638 (43.4)
Persistent smokers	112 (11.0)	48 (3.2)	107 (10.9)	48 (3.3)
Intermittent smokers	62 (6.1)	57 (3.8)	60 (6.1)	57 (3.9)
Former smokers	400 (39.1)	733 (49.3)	389 (39.7)	726 (49.4)
Pack-years at baseline[b]	22.0 (1 - 126)	28.5 (1 - 192)	22.0 (1 - 126)	29.0 (1 - 192)
FEV_1 at baseline (mL)	1924.4 (565.4)	2288.6 (645.0)	-	-
FEV_1/FVC at baseline (%)	-	-	75.1 (8.3)	74.2 (7.5)

Abbreviations: FEV_1, forced expiratory volume in 1 s; FVC, forced vital capacity.

[a] Data are presented as n, n (%), mean (standard deviation), or median (range).

[b] Data for ever-smokers only.

Table 2.2 The Association of Smoking with Rate of Change in Spirometry Phenotypes by Race/Ethnicity, for Participants in the Health ABC Study[a]

Population and Variable	Rate of Change in FEV_1			Rate of Change in FEV_1/FVC		
	Effect Estimate (mL/yr)	***P* Value**	***P* Value for set**	**Effect Estimate (%/yr)**	***P* Value**	***P* Value for set**
African Americans						
Time[b]	-32.21 ± 1.71	< 0.0001		-0.50 ± 0.04	< 0.0001	
Smoking status[c]						
Never smokers	Reference			Reference		
Persistent smokers	-5.96 ± 4.84	0.219	0.273	-0.30 ± 0.13	0.019	0.005
Intermittent smokers	-0.45 ± 4.78	0.925		-0.42 ± 0.14	0.004	
Former smokers	2.97 ± 2.47	0.229		-0.05 ± 0.06	0.440	
European Americans						
Time[b]	-39.77 ± 1.35	< 0.0001		-0.52 ± 0.02	< 0.0001	
Smoking status[c]						
Never smokers	Reference			Reference		
Persistent smokers	-12.05 ± 6.47	0.063	0.120	-0.62 ± 0.15	< 0.0001	<0.0001
Intermittent smokers	-6.51 ± 4.66	0.163		-0.34 ± 0.11	0.001	
Former smokers	-2.27 ± 1.75	0.194		-0.07 ± 0.04	0.051	

Abbreviations: FEV_1, forced expiratory volume in 1 s; FVC, forced vital capacity.

[a] Data presented are from the linear mixed effects models for the indicated phenotype and race/ethnicity analyses; all models included the following predictors: gender, study site, height at each time point, age and smoking pack-years (both at study baseline), time, smoking status, and smoking status × time.

[b] Effect estimate for time corresponds to the estimated annual rate of change in the phenotype for never smokers; negative values represent declines in the phenotype.

[c] Effect estimate for each smoking category is for the smoking status × time product term; negative values represent accelerations and positive values represent attenuations of decline, relative to never smokers.

Table 2.3 The Most Statistically Significant Associations for Single Genetic Variants with Rate of Change in FEV_1 by Race/Ethnicity, for Participants in the Health ABC Study[a]

Population and Gene	SNP	Chr	Base Pair Position	Minor Allele	MAF	β (mL/yr)[b]	*P* Value	Genetic Model[c]
African Americans								
mGST3	rs10800120	1	163871934	A	0.20	-5.8 ± 2.0	0.004	Additive
SOD3	rs8192287	4	24405666	T	0.02	-19.8 ± 5.4	2.45×10^{-4}	Additive
GSR	rs8190996	8	30673548	T	0.32	12.3 ± 3.8	0.001	Recessive
European Americans								
GSTA4	rs6904771	6	52964138	G	0.02	-13.2 ± 4.2	0.002	Additive

Abbreviations: FEV_1, forced expiratory volume in 1 s; SNP, single nucleotide polymorphism; Chr, chromosome; MAF, minor allele frequency; β, regression coefficient; *mGST3*, microsomal glutathione S-transferase 3; *SOD3*, superoxide dismutase 3; *GSR*, glutathione reductase; *GSTA4*, glutathione S-transferase A4.

[a] Data shown for associations with $P < 0.005$ for the SNP single-marker effect on rate of change in FEV_1, sorted by race/ethnicity, chromosome and base pair position. Statistically significant associations satisfying the Bonferroni-adjusted threshold (African Americans: $P < 2.3 \times 10^{-4}$; European Americans: $P < 3.0 \times 10^{-4}$) are bolded.

[b] Regression coefficient and standard error for the SNP × time product term in the corresponding mixed effects model.

[c] Genetic model is defined in reference to the minor allele for each SNP.

Table 2.4 The Most Statistically Significant Associations for Single Genetic Variants with Rate of Change in FEV_1/FVC by Race/Ethnicity, for Participants in the Health ABC Study[a]

Population and Gene	SNP	Chr	Base Pair Position	Minor Allele	MAF	β (%/yr)[b]	*P* Value	Genetic Model[c]
African Americans								
GLRX2	rs35358794	1	191336492	A	0.06	0.29 ± 0.09	0.001	Additive
SOD2	rs4342445	6	160018212	A	0.15	-0.53 ± 0.17	0.002	Recessive
TXN2	rs2267337	22	35200417	T	0.22	-0.16 ± 0.05	0.002	Additive
TXN2	rs2281082	22	35202696	T	0.22	-0.15 ± 0.05	0.004	Additive
PRDX4	rs528960	23	23601182	C	0.24	0.19 ± 0.06	0.003	Dominant
European Americans								
GCLC	**rs17883901**	**6**	**53517996**	**T**	**0.09**	**-0.86 ± 0.21**	**4.50×10^{-5}**	**Recessive**
GSTO2	rs157077	10	106027884	C	0.46	-0.09 ± 0.03	3.68×10^{-4}	Additive

Abbreviations: FEV_1, forced expiratory volume in 1 s; FVC, forced vital capacity; SNP, single nucleotide polymorphism; Chr, chromosome; MAF, minor allele frequency; β, regression coefficient; *GLRX2*, glutaredoxin 2; *SOD2*, superoxide dismutase 2; *TXN2*, thioredoxin 2; *PRDX4*, peroxiredoxin 4; *GCLC*, glutamate-cysteine ligase (catalytic subunit); *GSTO2*, glutathione S-transferase O2.

[a] Data shown for associations with $P < 0.005$ for the SNP single-marker effect on rate of change in FEV_1/FVC, sorted by race/ethnicity, chromosome and base pair position. Statistically significant associations satisfying the Bonferroni-adjusted threshold (African Americans: $P < 2.3 \times 10^{-4}$; European Americans: $P < 3.0 \times 10^{-4}$) are bolded.

[b] Regression coefficient and standard error for the SNP × time product term in the corresponding mixed effects model.

[c] Genetic model is defined in reference to the minor allele for each SNP.

Table 2.5 The Most Statistically Significant Gene-by-Smoking Interactions with Rate of Change in FEV_1 and FEV_1/FVC by Race/Ethnicity, for Participants in the Health ABC Study[a]

Phenotype, Population, and Gene	SNP-Smoking Interaction[c]	Chr	Base Pair Position	Minor Allele	MAF	Interaction Effect[d]	*P* Value	Genetic Model[e]
FEV_1								
African Americans								
GLRX2	rs34552619 × smoking status	1	191331738	C	0.08	-32.4 ± 8.8	2.74×10^{-4}	Dominant
IDH1	rs1437410 × smoking pack-years	2	208825562	C	0.22	-1.2 ± 0.4	6.27×10^{-4}	Recessive
European Americans								
SOD3	rs1007991 × smoking status	4	24409783	C	0.34	17.9 ± 5.7	0.002	Additive
GSTZ1	rs2111699 × smoking pack-years	14	76858350	G	0.32	-0.4 ± 0.1	9.55×10^{-4}	Recessive
FEV_1/FVC								
African Americans								
mGST3[b]	rs7554034 × smoking status	1	163877088	A	0.46	0.56 ± 0.21	0.007	Recessive
IDH3B[b]	rs6115381 × smoking status	20	2590376	G	0.37	-0.82 ± 0.23	4.60×10^{-4}	Recessive
SOD1	rs2070424 × smoking status	21	31961191	G	0.19	0.68 ± 0.19	3.62×10^{-4}	Dominant
G6PD	rs2472394 × smoking pack-years	23	153424545	A	0.13	-0.012 ± 0.003	7.75×10^{-4}	Dominant
European Americans								
mGST3	**rs2297765 × smoking**	**1**	**163888831**	**T**	**0.44**	**0.45 ± 0.12**	**1.13×10^{-4}**	**Additive**

	status							
SOD3[b]	rs2284659 × smoking status	4	24403895	T	0.37	0.64 ± 0.22	0.004	Recessive
IDH3B	rs6115381 × smoking pack-years	20	2590376	G	0.07	-0.008 ± 0.002	3.82×10^{-4}	Additive
IDH3B	rs6107100 × smoking pack-years	20	2592685	A	0.07	-0.008 ± 0.002	3.68×10^{-4}	Additive
IDH3B	**rs2073192 × smoking pack-years**	**20**	**2592996**	**A**	**0.07**	**-0.008 ± 0.002**	**2.10×10^{-4}**	**Additive**

Abbreviations: FEV_1, forced expiratory volume in 1 s; FVC, forced vital capacity; SNP, single nucleotide polymorphism; Chr, chromosome; MAF, minor allele frequency; *GLRX2*, glutaredoxin 2; *IDH1*, isocitrate dehydrogenase 1; *SOD3*, superoxide dismutase 3; *GSTZ1*, glutathione S-transferase Z1; *mGST3*, microsomal glutathione S-transferase 3; *IDH3B*, isocitrate dehydrogenase 3B; *SOD1*, superoxide dismutase 1; *G6PD*, glucose-6-phosphate dehydrogenase.

[a] Data shown are for the most statistically significant gene-by-smoking interactions for each phenotype and race/ethnicity analysis. Statistically significant interactions satisfying the Bonferroni-adjusted threshold (African Americans: $P < 2.3 \times 10^{-4}$; European Americans: $P < 3.0 \times 10^{-4}$) are bolded.

[b] These nominally significant ($P < 0.01$) interactions were selectively presented since they represent gene-level replications of most statistically significant interactions in another phenotype and race/ethnicity analysis.

[c] Smoking status was defined as a two-level categorical variable: smokers vs. non-smokers (reference group) during follow-up; smoking pack-years was modeled as a continuous variable.

[d] Beta coefficient and standard error for the SNP × smoking × time product term in the corresponding mixed effects model; mL per year for the effect on rate of change in FEV_1 and % per year for the effect on rate of change in FEV_1/FVC.

[e] Genetic model is defined in reference to the minor allele for each SNP.

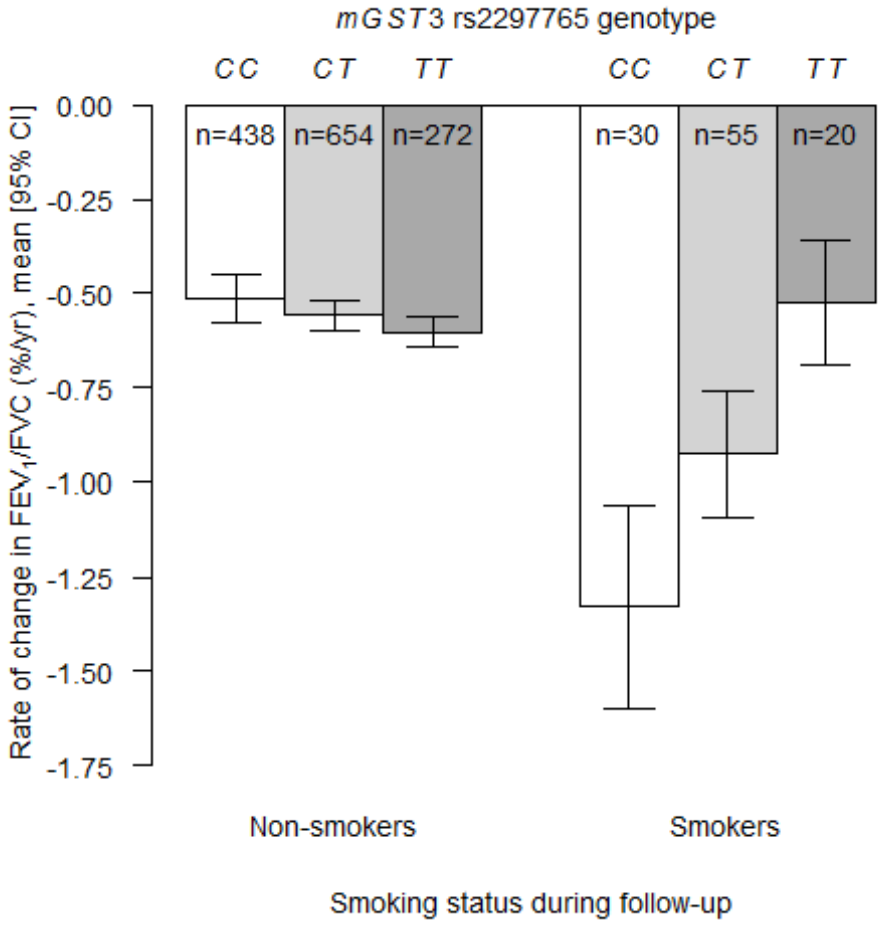


Figure 2.1 Estimated rates of change in FEV_1/FVC (% per year) according to smoking status during follow-up and *mGST3* rs2297765 genotype for European American participants in the Health ABC study. Open bars represent the *CC* genotype, light shaded bars represent the *CT* genotype and dark shaded bars represent the *TT* genotype. The estimates were computed from the linear mixed effects model that was adjusted for all covariates and included the SNP × smoking status × time product term ($P = 1.13 \times 10^{-4}$ following an additive genetic effect model). CI = confidence interval.

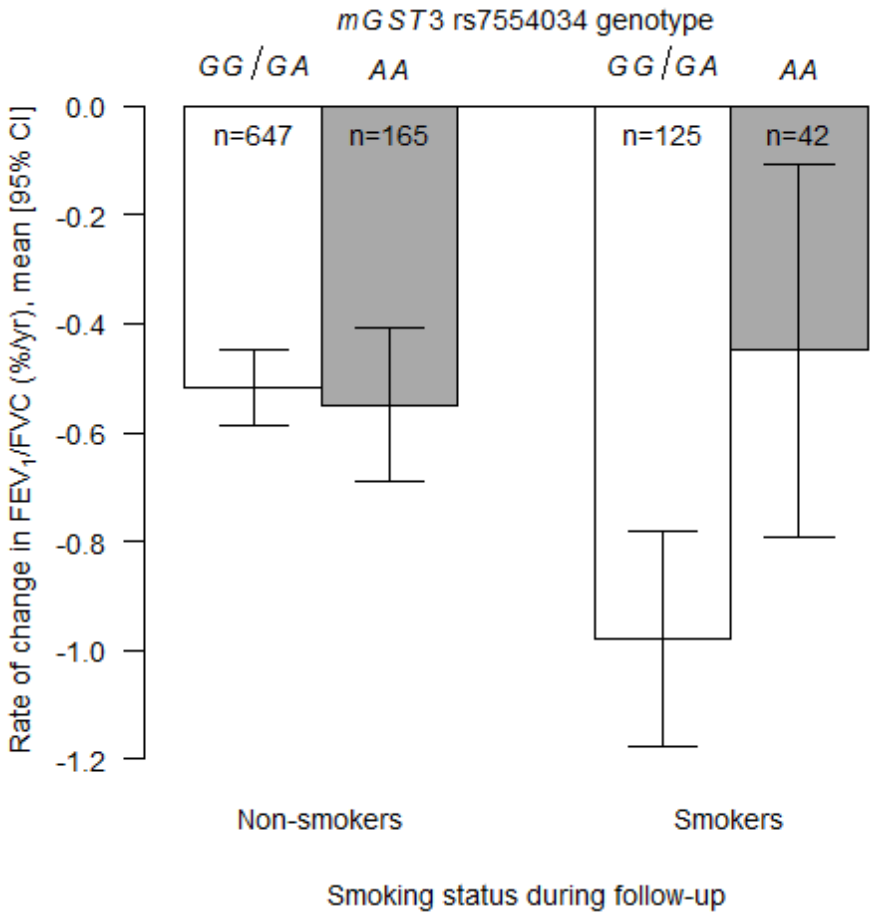


Figure 2.2 Estimated rates of change in FEV_1/FVC (% per year) according to smoking status during follow-up and *mGST3* rs7554034 genotype for African American participants in the Health ABC study. Open bars represent the *GG/GA* genotypes and shaded bars represent the *AA* genotype. The estimates were computed from the linear mixed effects model that was adjusted for all covariates and included the SNP × smoking status × time product term (*P* = 0.007 following a recessive genetic effect model). CI = confidence interval.

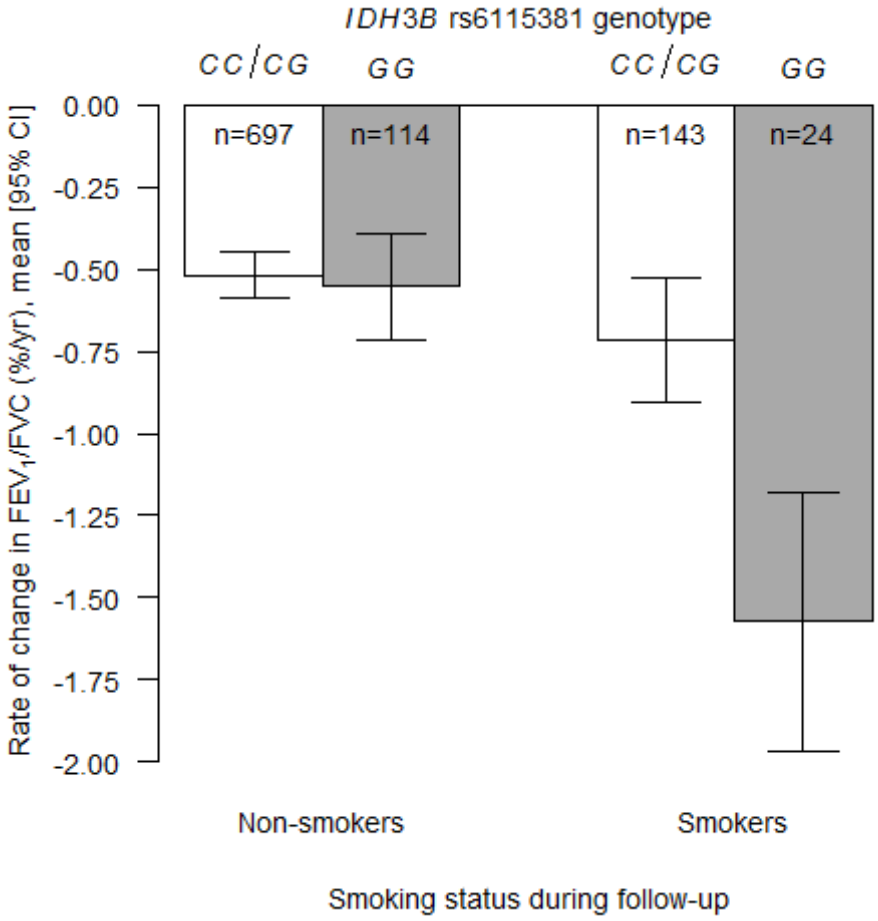


Figure 2.3 Estimated rates of change in FEV_1/FVC (% per year) according to smoking status during follow-up and *IDH3B* rs6115381 genotype for African American participants in the Health ABC study. Open bars represent the *CC/CG* genotypes and shaded bars represent the *GG* genotype. The estimates were computed from the linear mixed effects model that was adjusted for all covariates and included the SNP x smoking status x time product term ($P = 4.60 \times 10^{-4}$ following a recessive genetic effect model). CI = confidence interval.

Supplementary Table 2.6: Description of the 52 Candidate Genes Analyzed in the Study[a]

Gene Code	Gene Name	Functional Group	Chr	Gene Length (kb)	SNPs Genotyped	
					All	Non-Synonymous
CAT	Catalase	Catalase	11	33	7	0
G6PD	Glucose-6-Phosphate Dehydrogenase	Reducing Equivalents	23	16	1	0
GCLC	Glutamate-cysteine ligase (catalytic subunit)	GSH Synthesis	6	48	15	0
GCLM	Glutamate-cysteine ligase (modulatory subunit)	GSH Synthesis	1	22	6	0
GGT1	Gamma-glutamyl Transferase 1	GSH Synthesis	22	45	3	0
GLRX	Glutaredoxin	Disulfide Reductase	5	9	5	0
GLRX2	Glutaredoxin 2	Disulfide Reductase	1	10	6	0
GPX1	Glutathione Peroxidase 1	Peroxidase Activity	3	1.2	3	0
GPX2	Glutathione Peroxidase 2	Peroxidase Activity	14	3.7	5	0
GPX3	Glutathione Peroxidase 3	Peroxidase Activity	5	8.5	4	0
GPX4	Glutathione Peroxidase 4	Peroxidase Activity	19	2.9	5	0
GPX7	Glutathione Peroxidase 7	Peroxidase Activity	1	6.7	5	0
GSR	Glutathione Reductase	Disulfide Reductase	8	49	9	1
GSS	Glutathione Synthetase	GSH Synthesis	20	11	7	0
GSTA1	Glutathione S-Transferase A1	Glutathione S-Transferase	6	12	3	0
GSTA2	Glutathione S-Transferase A2	Glutathione S-Transferase	6	13	3	1
GSTA3	Glutathione S-Transferase A3	Glutathione S-Transferase	6	13	6	1
GSTA4	Glutathione S-Transferase A4	Glutathione S-Transferase	6	17	6	0
GSTA5	Glutathione S-Transferase A5	Glutathione S-Transferase	6	14	7	1
GSTM2	Glutathione S-Transferase M2	Glutathione S-Transferase	1	7.2	3	0
GSTM3	Glutathione S-Transferase M3	Glutathione S-Transferase	1	6.5	6	1
GSTM4	Glutathione S-Transferase M4	Glutathione S-Transferase	1	9.4	4	0

Gene Code	Gene Name	Functional Group	Chr	Gene Length (kb)	SNPs Genotyped	
					All	Non-Synonymous
GSTO1	Glutathione S-Transferase O1	Glutathione S-Transferase	10	13	4	0
GSTO2	Glutathione S-Transferase O2	Glutathione S-Transferase	10	31	9	1
GSTP1	Glutathione S-Transferase P1	Glutathione S-Transferase	11	2.8	8	2
GSTZ1	Glutathione S-Transferase Z1	Glutathione S-Transferase	14	8.1	6	2
HMOX1	Heme-Oxygenase 1	Heme-Oxygenase	22	13	5	0
HMOX2	Heme-Oxygenase 2	Heme-Oxygenase	16	34	6	0
IDH1	Isocitrate Dehydrogenase 1	Reducing Equivalents	2	19	7	1
IDH2	Isocitrate Dehydrogenase 2	Reducing Equivalents	15	19	7	0
IDH3A	Isocitrate Dehydrogenase 3A	Reducing Equivalents	15	21	5	0
IDH3B	Isocitrate Dehydrogenase 3B	Reducing Equivalents	20	5.8	4	0
IDH3G	Isocitrate Dehydrogenase 3G	Reducing Equivalents	23	8.7	5	0
mGST1	Microsomal Glutathione S-Transferase 1	Glutathione S-Transferase	12	17	8	0
mGST2	Microsomal Glutathione S-Transferase 2	Glutathione S-Transferase	4	39	11	0
mGST3	Microsomal Glutathione S-Transferase 3	Glutathione S-Transferase	1	24	11	0
PRDX1	Peroxiredoxin 1	Peroxidase Activity	1	11	3	0
PRDX2	Peroxiredoxin 2	Peroxidase Activity	19	5	6	0
PRDX3	Peroxiredoxin 3	Peroxidase Activity	10	11	6	0
PRDX4	Peroxiredoxin 4	Peroxidase Activity	23	19	6	0
PRDX5	Peroxiredoxin 5	Peroxidase Activity	11	3.7	6	1
PRDX6	Peroxiredoxin 6	Peroxidase Activity	1	11	6	0
SEPP1	Selenoprotein P 1	Selenoprotein	5	12	6	1
SEPW1	Selenoprotein W 1	Selenoprotein	19	6.1	7	0
SOD1	Superoxide Dismutase 1	Superoxide Dismutase	21	9.3	6	0
SOD2	Superoxide Dismutase 2	Superoxide Dismutase	6	14	6	1

Gene Code	Gene Name	Functional Group	Chr	Gene Length (kb)	SNPs Genotyped	
					All	Non-Synonymous
SOD3	Superoxide Dismutase 3	Superoxide Dismutase	4	6.4	10	1
TXN	Thioredoxin 1	Thioredoxin	9	12	5	0
TXN2	Thioredoxin 2	Thioredoxin	22	15	5	0
TXNRD	Thioredoxin Reductase 1	Disulfide Reductase	12	63	8	0
TXNRD	Thioredoxin Reductase 2	Disulfide Reductase	22	66	12	1
TXNRD	Thioredoxin Reductase 3	Disulfide Reductase	3	52	4	0

[a] Four additional genes (*GGT2*, *GSTK1*, *GSTM1*, and *GSTT1*) were genotyped but had to be excluded from further analyses due to low genotyping quality or atypical clustering of assayed SNPs.

Supplementary Table 2.7: Summary of Significant Gene by Smoking Status Interactions ($P < 0.01$) with Rate of Change in Forced Expiratory Volume in 1 s for African American Participants in the Health ABC Study

Gene	SNP	Chr	Base Pair Position	Minor Allele	MAF	β (mL/yr)[a]	*P* Value	Genetic Model[b]
PRDX6	rs34977864	1	171712466	G	0.07	-31.8 ± 10.9	3.59×10^{-4}	Additive
GLRX2	rs34552619	1	191331738	C	0.08	-32.4 ± 8.8	2.74×10^{-4}	Dominant
IDH1	rs6435435	2	208820796	G	0.13	-20.8 ± 6.8	0.002	Additive
GPX1	rs8179172	3	49368271	T	0.09	-25.7 ± 8.8	0.003	Additive
GSR	rs2978296	8	30694482	G	0.29	-21.9 ± 6.9	0.002	Dominant
PRDX2	rs10404253	19	12767789	C	0.23	-20.9 ± 6.9	0.003	Dominant
GGT1	rs16978740	22	23313264	T	0.10	27.8 ± 8.7	0.001	Dominant
PRDX4	rs557914	23	23596108	G	0.37	-21.0 ± 7.1	0.003	Dominant

Abbreviations: SNP, single nucleotide polymorphism; Chr, chromosome; MAF, minor allele frequency

[a] Beta coefficient and standard error for the SNP × smoking status × time product term in the mixed effects model; smoking status defined as a two-level categorical variable: smokers vs. non-smokers (reference group).

[b] Genetic model is defined in reference to the minor allele for each SNP.

Supplementary Table 2.8: Summary of Significant Gene by Smoking Pack-years Interactions ($P < 0.01$) with Rate of Change in Forced Expiratory Volume in 1 s for African American Participants in the Health ABC Study

Gene	SNP	Chr	Base Pair Position	Minor Allele	MAF	β (mL/yr)[a]	*P* Value	Genetic Model[b]
IDH1	rs1437410	2	208825562	C	0.22	-1.2 ± 0.4	6.27×10^{-4}	Recessive

Abbreviations: SNP, single nucleotide polymorphism; Chr, chromosome; MAF, minor allele frequency.

[a] Beta coefficient and standard error for the SNP × smoking pack-years × time product term in the mixed effects model.

[b] Genetic model is defined in reference to the minor allele for each SNP.

Supplementary Table 2.9: Summary of Significant Gene by Smoking Status Interactions ($P < 0.01$) with Rate of Change in Forced Expiratory Volume in 1 s for European American Participants in the Health ABC Study

Gene	SNP	Chr	Base Pair Position	Minor Allele	MAF	β (mL/yr)[a]	*P* Value	Genetic Model[b]
SOD3	rs2284659	4	24403895	T	0.37	14.5 ± 5.3	0.006	Additive
SOD3	rs1007991	4	24409783	C	0.34	17.9 ± 5.7	0.002	Additive
SOD3	rs2855262	4	24411074	T	0.37	15.2 ± 5.2	0.004	Additive
mGST2	rs8191997	4	140804481	A	0.16	20.5 ± 7.7	0.008	Dominant
GSTZ1	rs2111699	14	76858350	G	0.32	-14.6 ± 5.6	0.009	Additive
IDH3G	rs2071122	23	152705101	A	0.28	-25.2 ± 9.7	0.009	Recessive
IDH3G	rs2071123	23	152705843	A	0.28	-25.2 ± 9.7	0.009	Recessive

Abbreviations: SNP, single nucleotide polymorphism; Chr, chromosome; MAF, minor allele frequency

[a] Beta coefficient and standard error for the SNP × smoking status × time product term in the mixed effects model; smoking status defined as a two-level categorical variable: smokers vs. non-smokers (reference group) during follow-up.

[b] Genetic model is defined in reference to the minor allele for each SNP.

Supplementary Table 2.10: Summary of Significant Gene by Smoking Pack-years Interactions ($P < 0.01$) with Rate of Change in Forced Expiratory Volume in 1 s for European American Participants in the Health ABC Study

Gene	SNP	Chr	Base Pair Position	Minor Allele	MAF	β (mL/yr)[a]	*P* Value	Genetic Model[b]
mGST2	rs795589	4	140819575	C	0.42	-0.2 ± 0.1	0.001	Dominant
GSTZ1	rs2111699	14	76858350	G	0.32	-0.4 ± 0.1	9.55×10^{-4}	Recessive
GSTZ1	rs2363643	14	76858661	A	0.32	-0.4 ± 0.1	0.001	Recessive

Abbreviations: SNP, single nucleotide polymorphism; Chr, chromosome; MAF, minor allele frequency

[a] Beta coefficient and standard error for the SNP × smoking pack-years × time product term in the mixed effects model.

[b] Genetic model is defined in reference to the minor allele for each SNP.

Supplementary Table 2.11: Summary of Significant Gene by Smoking Status Interactions ($P < 0.01$) with Rate of Change in the Ratio of Forced Expiratory Volume in 1 s/Forced Vital Capacity for African American Participants in the Health ABC Study

Gene	SNP	Chr	Base Pair Position	Minor Allele	MAF	β (%/yr)[a]	*P* Value	Genetic Model[b]
GPX7	rs6588431	1	52840174	T	0.40	0.53 ± 0.18	0.003	Dominant
mGST3	rs7554034	1	163877088	A	0.46	0.56 ± 0.21	0.007	Recessive
HMOX2	rs8055559	16	4480228	G	0.26	0.48 ± 0.18	0.008	Dominant
IDH3B	rs6115381	20	2590376	G	0.37	-0.82 ± 0.23	4.60×10^{-4}	Recessive
SOD1	rs4998557	21	31956763	A	0.39	0.36 ± 0.12	0.004	Additive
SOD1	rs2070424	21	31961191	G	0.19	0.68 ± 0.19	3.62×10^{-4}	Dominant
GGT1	rs2154611	22	23319920	C	0.32	-0.86 ± 0.29	0.003	Recessive

Abbreviations: SNP, single nucleotide polymorphism; Chr, chromosome; MAF, minor allele frequency

[a] Beta coefficient and standard error for the SNP × smoking status × time product term in the mixed effects model; smoking status defined as a two-level categorical variable: smokers vs. non-smokers (reference group) during follow-up.

[b] Genetic model is defined in reference to the minor allele for each SNP.

Supplementary Table 2.12: Summary of Significant Gene × Smoking Pack-years Interactions ($P < 0.01$) with Rate of Change in the Ratio of Forced Expiratory Volume in 1 s/Forced Vital Capacity for African American Participants in the Health ABC Study

Gene	SNP	Chr	Base Pair Position	Minor Allele	MAF	β (%/yr)[a]	*P* Value	Genetic Model[b]
SEPP1	rs230813	5	42834790	G	0.45	0.010 ± 0.003	0.001	Recessive
GSR	rs1002149	8	30705280	T	0.24	-0.008 ± 0.003	0.004	Dominant
SEPW1	rs10427074	19	52976136	G	0.12	0.010 ± 0.003	0.001	Additive
HMOX1	rs6518952	22	34112513	T	0.32	0.008 ± 0.003	0.003	Dominant
G6PD	rs2472394	23	153424545	A	0.13	-0.012 ± 0.003	7.75×10^{-4}	Dominant

Abbreviations: SNP, single nucleotide polymorphism; Chr, chromosome; MAF, minor allele frequency

[a] Beta coefficient and standard error for the SNP × smoking pack-years × time product term in the mixed effects model.

[b] Genetic model is defined in reference to the minor allele for each SNP.

Supplementary Table 2.13: Summary of Significant Gene × Smoking Status Interactions ($P < 0.01$) with Rate of Change in the Ratio of Forced Expiratory Volume in 1 s/Forced Vital Capacity for European American Participants in the Health ABC Study

Gene	SNP	Chr	Base Pair Position	Minor Allele	MAF	β (%/yr)[a]	*P* Value	Genetic Model[b]
GCLM	rs7549683	1	94126037	T	0.35	0.33 ± 0.12	0.004	Additive
GCLM	rs769211	1	94132695	A	0.26	0.36 ± 0.12	0.003	Additive
GCLM	rs7517826	1	94137922	A	0.35	0.33 ± 0.12	0.004	Additive
GCLM	rs3827715	1	94142211	C	0.26	0.36 ± 0.12	0.003	Additive
mGST3	rs2297765	1	163888831	T	0.44	0.45 ± 0.12	1.13×10^{-4}	Additive
SOD3	rs2284659	4	24403895	T	0.37	0.64 ± 0.22	0.004	Recessive
GPX4	rs757228	19	1052992	G	0.45	-0.54 ± 0.18	0.003	Dominant
GPX4	rs3746165	19	1053211	G	0.45	-0.54 ± 0.18	0.003	Dominant

Abbreviations: SNP, single nucleotide polymorphism; Chr, chromosome; MAF, minor allele frequency

[a] Beta coefficient and standard error for the SNP × smoking status × time product term in the mixed effects model; smoking status defined as a two-level categorical variable: smokers vs. non-smokers (reference group) during follow-up.

[b] Genetic model is defined in reference to the minor allele for each SNP.

Supplementary Table 2.14: Summary of Significant Gene × Smoking Pack-years Interactions ($P < 0.01$) with Rate of Change in the Ratio of Forced Expiratory Volume in 1 s/Forced Vital Capacity for European American Participants in the Health ABC Study

Gene	SNP	Chr	Base Pair Position	Minor Allele	MAF	β (%/yr)[a]	*P* Value	Genetic Model[b]
GSTA4	rs6904771	6	52964138	G	0.02	0.012 ± 0.004	9.89×10^{-4}	Dominant
IDH3B	rs6115381	20	2590376	G	0.07	-0.008 ± 0.002	3.82×10^{-4}	Additive
IDH3B	rs6107100	20	2592685	A	0.07	-0.008 ± 0.002	3.68×10^{-4}	Additive
IDH3B	rs2073192	20	2592996	A	0.07	-0.008 ± 0.002	2.10×10^{-4}	Additive

Abbreviations: SNP, single nucleotide polymorphism; Chr, chromosome; MAF, minor allele frequency

[a] Beta coefficient and standard error for the SNP × smoking pack-years × time product term in the mixed effects model.

[b] Genetic model is defined in reference to the minor allele for each SNP.

CHAPTER 3

GENOME-WIDE ASSOCIATION STUDIES OF LONGITUDINAL CHANGE IN ADULT LUNG FUNCTION

Detailed information on authors, affiliations, correspondence, contributions, and sources of funding is presented at the end of the chapter.

ABSTRACT

Background: Genome-wide association studies (GWAS) have identified numerous loci influencing cross-sectional lung function, but less is known about genes influencing longitudinal change in lung function.

Methods: We performed GWAS of the rate of change in forced expiratory volume in the first second (FEV_1) in 14 longitudinal, population-based cohort studies comprising 27249 adults of European ancestry using linear mixed effects model and combined cohort-specific results using fixed effect meta-analysis to identify novel genetic loci associated with longitudinal change in lung function. As a secondary aim, we estimated the rate of decline in FEV_1 by smoking pattern across these 14 studies using meta-analysis.

Results: The overall meta-analysis produced suggestive evidence for association at the novel *IL16/STARD5/TMC3* locus on chromosome 15 ($P = 5.71 \times 10^{-7}$). In addition, meta-analysis using the five cohorts with ≥3 FEV_1 measurements per participant identified the novel *ME3* locus on chromosome 11 ($P = 2.18 \times 10^{-8}$) at genome-wide significance. Neither locus was associated with FEV_1 decline in two additional cohort studies. We confirmed gene expression of *IL16*, *STARD5*, and *ME3* in multiple lung tissues. Publicly available microarray data confirmed differential expression of all three genes in lung samples from COPD patients compared with controls. The combined estimate for FEV_1 decline was 26.9, 29.2 and 35.7 mL/year in never, former, and persistent smokers, respectively.

Conclusions: In this large-scale GWAS, we identified two novel genetic loci in association with the rate of change in FEV_1 that harbor candidate genes with biologically plausible functional links to lung function.

Key words: Genome-wide association study; GWAS; lung function; respiratory function test; forced expiratory volume in the first second; FEV_1; longitudinal change; cigarette smoking

INTRODUCTION

Forced expiratory volume in the first second (FEV_1) is a reliable spirometric parameter that reflects the physiological state of the lungs and airways. Reduced FEV_1 relative to forced vital capacity (FVC), is a defining feature of chronic obstructive pulmonary disease (COPD), a leading cause of death globally (1). FEV_1 is also a predictor of morbidity and mortality in the general population (2, 3). Lung function reaches its peak in early adulthood, followed by a plateau, and then subsequently declines. As first reported by Fletcher and Peto (4), decline in lung function is accelerated in smokers, leading to increased risks of COPD and premature death. While cigarette smoking is a key risk factor for accelerated loss of lung function, genetic variation is hypothesized to also play an important role (5, 6). Family and twin studies of the longitudinal change in lung function report heritability estimates between 10 and 39% (7, 8).

Recent large-scale genome-wide association studies (GWAS) identified 26 novel loci for cross-sectional lung function (9-11), demonstrating the power of GWAS with large sample size to identify common genetic variants with modest effect sizes. However, cross-sectional measurements in adults reflect the combination of maximal attained lung growth and subsequent decline. GWAS that specifically study the longitudinal change in lung function are needed to distinguish the genetic contributions to age-related decline. To date, only one population-based GWAS meta-analysis of longitudinal change in lung function has been reported (12). Separate analyses were conducted in 1441 asthmatic and 2667 non-asthmatic participants; association was found at one novel locus in each analysis, though only the locus in non-asthmatics replicated.

In this study, we conducted primary GWAS of the rate of change in FEV_1 in each of 14 population-based cohort studies from the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) and SpiroMeta consortia, comprising 27249 adult participants of European ancestry and 62130 FEV_1 measurements. We then performed meta-analysis of the cohort-specific results, followed up our most statistically significant associations in the AGES-Reykjavík cohort study and the Lung Health Study (LHS) for corroborative evidence, and explored the biological basis for identified associations using cell-specific gene expression studies, and expression quantitative trait loci (eQTL) analysis.

METHODS

All 14 cohort studies are members of the CHARGE or SpiroMeta Consortium (Table 3.1). The respective local Institutional Review Boards approved all study protocols, and written informed consent for genetic studies was obtained from all participants.

Spirometry tests were performed at baseline and at least one follow-up time point by trained technicians and in accordance with the American Thoracic Society or European Respiratory Society recommendations (details in online supplement) (13). FEV_1 measurements meeting inclusion criteria were included in the current study.

Studies performed genotyping following standard quality control measures; imputation was conducted based on the HapMap CEU reference panel to generate genotype dosages for ~ 2.5 million autosomal single nucleotide polymorphisms (SNPs) (Supplementary Table 3.5).

Each cohort study performed the GWAS using a linear mixed effects model. The model included a random intercept and a random slope, and fixed effects for time (a

continuous variable quantifying the time distance between each FEV_1 measurement and baseline), SNP and its interaction with time (SNP-by-time), baseline age, gender, standing height, smoking pattern during follow-up and its interaction with time (smoking-by-time), baseline smoking pack-years, study site, and principal components for genetic ancestry (as needed). Cohort-specific results for the SNP-by-time interaction term, which estimates the effect of genotype on the rate of change in FEV_1, were shared, and two meta-analyses, one using all 14 studies and the other using the five studies with ≥3 FEV_1 measurements per participant, were performed using METAL software (14) with inverse variance weighting to combine effect estimates after applying genomic control correction.

We sought corroborative evidence for SNPs with $P < 1 \times 10^{-5}$ in the AGES-Reykjavík cohort study (n = 1494), and in LHS (n = 4048), a clinical cohort study of smokers with mild COPD, in which a longitudinal GWAS was recently reported (15).

Expression profiles of genes at the novel loci were evaluated in human lung tissues and primary cell samples using RT-PCR (Supplementary Table 3.11). Using publicly available data from the Lung Genomics Research Consortium (LGRC), expression profiles of these genes were compared in lung specimens of 219 COPD patients and 137 controls, and sentinel (most associated) SNPs at the novel loci were also searched against an eQTL database of lymphoblastoid cell lines (16).

RESULTS

The majority of the 14 cohort studies had FEV_1 at two times, but five studies (BHS, CARDIA, CHS, FHS, Health ABC) had ≥3 FEV_1 measurements per participant. The maximum length of follow-up ranged from 4 to 29 years. Studies with older participants

generally had fewer current smokers and more former smokers, and had lower mean baseline FEV_1.

All 14 studies implemented a preliminary mixed model adjusted for all specified variables except the SNP genotypes and reported the estimated rate of change in FEV_1 by smoking pattern (Table 3.2). The rate of decline in FEV_1 in never smokers ranged from 10.0 to 39.7 mL/year, and was generally steeper in studies with older participants, as expected (4). Across all 14 studies, the meta-analyzed rate of change in FEV_1 was a decline of 26.9±0.3 mL/year in never smokers, and was 8.8±0.7, 2.6±0.6, and 2.3±0.5 mL/year steeper in persistent, intermittent, and former smokers, respectively (Table 3.2). We repeated the meta-analyses in the five cohort studies with ≥3 FEV_1 measurements per participant, and found similar, although less statistically significant results.

Study-specific genomic inflation factors (λ_{gc}) were calculated for the SNP-by-time interaction term and used for study-level genomic control prior to the meta-analyses. Study-specific λ_{gc} values ranged from 0.96 to 1.11 (Supplementary Table 3.5) and the meta-analysis λ_{gc} was 1.01 for both the 14-study and five-study meta-analyses. Supplementary Figures 3.3 and 3.4 present the Manhattan and quantile-quantile (QQ) plots.

In the meta-analysis including all 14 cohort studies, 15 SNPs at nine independent loci were associated with the rate of change in FEV_1 at $P < 1 \times 10^{-5}$, and none reached the genome-wide significance threshold of $P < 5 \times 10^{-8}$. The association results for the sentinel SNPs at these nine loci are presented in Table 3.3, and more detailed results for all 15 SNPs are included in Supplementary Table 3.6. The most statistically significant association, and the only one that reached $P < 1 \times 10^{-6}$, was for rs4077833, an intronic SNP located in the novel *IL16/STARD5/TMC3* gene region on chromosome 15 ($P = 5.71 \times 10^{-7}$; Figure 3.1). The

C allele of rs4077833, with a frequency of 10%, was associated with an attenuation of the rate of decline in FEV_1 by 2.3 mL/year.

For estimation of longitudinal trajectory in lung function, having more than two measurements over time provides greater precision (4). We performed a further meta-analysis with the five cohort studies (BHS, CHS, CARDIA, FHS, Health ABC) having ≥3 FEV_1 measurements per participant, with a combined sample size of 10476 participants and 32054 FEV_1 measurements (online supplement for details). A novel region on chromosome 11 had a genome-wide significant association ($P < 5 \times 10^{-8}$) with the rate of change in FEV_1 (Table 3.4). The most statistically significant finding at this locus was for rs507211, an intronic SNP located in *ME3* (Figure 3.2). Six other SNPs, which are in linkage disequilibrium (LD) with rs507211 and are located in *ME3*, were identified at $P < 1 \times 10^{-6}$ (Supplementary Table 3.7). The rs507211 A allele, with a frequency of 25%, was associated with an attenuation of the rate of decline in FEV_1 by 2.09 mL/year ($P = 2.18 \times 10^{-8}$). Besides the *ME3* locus, 17 SNPs from four other chromosomal regions had *P* values between 5×10^{-8} and 1×10^{-5} for associations with the rate of change in FEV_1 (Tables 3.4 and Supplementary Table 3.7).

Corroborative evidence was sought for the sentinel SNP at each of the 14 loci associated at $P < 1 \times 10^{-5}$ (from both the 14-study and five-study meta-analyses) in 1494 adults from the AGES-Reykjavík population-based cohort study (Supplementary Table 3.8). A *P* value of 0.004, representing the Bonferroni correction for 14 tests at the $\alpha = 0.05$ level, was selected *a priori* as the threshold for statistical significance. No SNPs achieved this threshold. The lowest *P* value was for rs740577 in *CACNG4* ($P = 0.08$), which showed consistent effect direction and magnitude with the original meta-analysis.

These same 14 SNPs were further examined in LHS, a clinical cohort study of 4,048 smokers with mild COPD for evidence of consistent association between healthy and diseased individuals (17). None of the 14 SNPs were associated with the rate of change in FEV_1 in LHS at $P < 0.004$ (Supplementary Table 3.8).

Previous meta-analyses in the CHARGE and SpiroMeta consortia identified 26 novel loci associated with cross-sectional FEV_1 and/or FEV_1/FVC at genome-wide significance (9-11). We examined the sentinel SNPs from these loci in the meta-analysis of the 14 cohort studies for association with the rate of change in FEV_1 (Supplementary Table 3.9). Given the *a priori* association with cross-sectional lung function, a *P* value threshold of 0.05 was used. Sentinel SNPs in *PID1*, *HHIP*, *GPR126*, and *CFDP1* showed association with the rate of change in FEV_1 ($0.005 \leq P \leq 0.048$).

Three genes (*IL16*, *STARD5*, and *TMC3*) at the novel chromosome 15 locus and *ME3* at the novel chromosome 11 locus were selected for follow-up mRNA expression profiling in human lung tissue, and primary cultures of human bronchial epithelial and airway smooth muscle cells, together with control tissues (peripheral blood mononuclear cells and brain). Transcripts of *STARD5* and *ME3* were found in all lung-derived tissues, transcripts of *IL16* were found in lung tissue and smooth muscle cells, but not in epithelial cells, and *TMC3* was not expressed in any of the lung-derived tissues (Supplementary Table 3.10).

Using the public LGRC data repository, we found that the expression profiles of *IL16*, *STARD5*, and *ME3* in human lung samples showed statistically significant differences ($P < 0.05$) between COPD patients and controls (Supplementary Figure 3.5). Lower levels of *IL16* ($P = 0.004$) were observed in COPD patients compared with controls, whereas higher levels

of *STARD5* ($P = 3.22 \times 10^{-9}$) and *ME3* ($P = 0.044$) were observed in COPD patients compared with controls. Data on *TMC3* expression were not available.

We performed additional follow-up analysis of the sentinel SNPs at the two novel loci using an eQTL database of lymphoblastoid cell lines (Supplementary Table 3.12). Trans-eQTL associations were observed between rs4077833 at the *IL16/STARD5/TMC3* locus and a nuclear receptor, *NR1I2* (chromosome 3; $P = 6.84 \times 10^{-4}$) and between rs507211 at the *ME3* locus and *KIAA1109* (chromosome 4; $P = 5.20 \times 10^{-4}$), which is part of a gene cluster (*KIAA1109-TENR-IL2-IL21*) that encodes two interleukins (IL2 and IL21) (18).

DISCUSSION

Although the genetic contribution to cross-sectional lung function phenotypes has been addressed by large-scale GWAS, much less information is available for longitudinal lung function phenotypes. To identify novel loci that specifically affect lung function change over time, we performed a large-scale GWAS of the rate of change in FEV_1 in 27249 participants from 14 population-based cohort studies. We identified a novel locus (*IL16/STARD5/TMC3*) on chromosome 15 with suggestive evidence for association with the rate of change in FEV_1. Given the greater precision to estimate longitudinal trends with more measurements, a meta-analysis of the five cohort studies with $\geq$3 FEV_1 measurements per participant was performed, and it identified a second novel locus (*ME3*) on chromosome 11 at genome-wide statistical significance. For both loci, the minor allele was protective, and the magnitude of the association with the rate of change in FEV_1 was similar to that of being an intermittent or former smoker versus a never-smoker.

The sentinel SNP at the novel chromosome 15 locus is located in *TMC3*, although two neighboring genes, *IL16* and *STARD5* both harbor SNPs that are in modest LD with the sentinel SNP (Figure 3.1). TMC3, a member of the transmembrane channel-like gene family, likely functions as an ion channel, transporter, or modifier (19), and has been associated with deafness (20) and skin cancer (21). IL16 is a pleiotropic immunomodulatory cytokine that acts as a chemoattractant for $CD4^+$ cells and contributes to their recruitment and activation in response to inflammation (22). Notably, asthma was the first disease where increased *IL16* expression was observed (23). Subsequent studies confirmed that in the non-diseased state *IL16* is almost exclusively expressed by T lymphocytes in lymphatic tissue, whereas in asthmatic patients IL16 is also synthesized by airway epithelial cells to inhibit airway inflammation (24-26). A promoter polymorphism (T-295C) in *IL16* was associated with asthma in a Caucasian population in England (27), although this finding was not confirmed in an Australian study (28). STARD5 belongs to the steroidogenic acute regulatory lipid transfer domain protein superfamily, and is involved in the trafficking of cholesterol and other lipids between intracellular membranes (29). Recent *in vitro* studies showed increased *STARD5* expression and protein redistribution as a protective mechanism in response to induced endoplasmic reticulum (ER) stress and consequent over-accumulation of intracellular free cholesterol (30). We confirmed the expression of *STARD5* in all human lung tissues examined and of *IL16* in human lung smooth muscle cells, but not epithelial cells, in line with previous observations. In contrast, no expression of *TMC3* was detected in any of the tested human lung tissues. We also found significantly lower levels of *IL16* in whole lung samples from COPD patients compared with controls, in contrast to its increased expression in asthma, and significantly higher levels of *STARD5* in COPD patients compared with controls. Taken

together, these results suggest *IL16* as the most likely candidate accounting for the observed association, but further investigation is needed to elucidate underlying mechanisms.

The sentinel SNP at the novel chromosome 11 locus is located in *ME3*, whose protein product is a mitochondrial NADP(+)-dependent malic enzyme that catalyzes the oxidative decarboxylation of malate to pyruvate using NADP+ as a cofactor (31). Mitochondrial malic enzymes play a role in the energy metabolism in tumors, and are considered potential therapeutic targets in cancer (32, 33). We performed independent expression profiling of *ME3* and confirmed its expression in all human lung tissues examined, and found significantly higher levels of *ME3* in lung samples from COPD patients compared with controls. In addition, the sentinel SNP in *ME3* was associated with airway obstruction at $P = 0.049$ in a recent GWAS of airway obstruction (34). Taken together, these results support *ME3* as a biologically plausible candidate in the regulation of lung function and pathogenesis of COPD.

The identification of trans-eQTL associations for the sentinel SNPs at both the *IL16/STARD5/TMC3* and *ME3* loci is interesting, and while the interpretation of trans-eQTL associations is ambiguous (35), the regions these SNPs regulate merit further study.

Besides the GWAS meta-analyses, the assembly of 14 longitudinal cohort studies allowed us to meta-analyze the association of cumulative smoking patterns with the rate of change in FEV_1 in the general population. The meta-analyzed estimate for the rate of decline in FEV_1 in never smokers was 26.9 mL/year, and the annual decline was steeper in persistent, intermittent, and former smokers by 8.8, 2.6, and 2.3 mL/year, respectively. These findings provide a reference point for the effect of cigarette smoking on longitudinal lung function change in the general population.

There is phenotypic variation among the 14 cohort studies in aspects such as baseline age and cigarette smoking, and in factors that are of special importance to this longitudinal GWAS, such as the number of FEV_1 measurements per participant and follow-up duration. Phenotypic heterogeneity represents a general challenge in genetic epidemiology, particularly in the investigation of longitudinal phenotypes. Thus, we performed a meta-analysis using the subset of cohort studies with $\geq$3 FEV_1 measurements per participant, given that longitudinal trajectories are best estimated over longer time periods and with more measurements. There was little overlap between the top loci identified in the two meta-analyses at $P < 1 \times 10^{-5}$, suggesting that phenotypic heterogeneity affected the association results. Future meta-studies of lung function decline should aim to increase sample size while maintaining high phenotypic comparability among participating studies.

We sought corroborative evidence in a single cohort study of 1,494 participants. This sample size is much smaller and arguably insufficient compared with replications applied to previous studies of cross-sectional lung function phenotypes. Thus, despite the lack of corroboration for the two novel loci identified in the meta-analyses, results from the complementary gene expression analyses provide compelling evidence for biologically plausible roles of the implicated genes in the longitudinal change in lung function.

None of the 14 sentinel SNPs were associated with the rate of change in FEV_1 in the COPD patient-based LHS cohort. Similarly, a previous population-based GWAS of lung function decline noted a high degree of heterogeneity in findings when analyses were stratified by presence/absence of asthma (12). The observed discrepancy of association results suggests that the genetic determination of lung function decline may be different in healthy individuals compared with COPD patients, may contribute differentially in a pre-diseased vs.

post-diseased state in which medications may influence the rates of decline, or that LHS was underpowered for confirming our findings.

In summary, we performed GWAS of the longitudinal change in lung function and subsequent meta-analyses, using harmonized data from more than 27000 participants of European ancestry to identify genetic loci influencing the rate of change in FEV_1. We identified the novel *ME3* locus on chromosome 11 at genome-wide significance and found suggestive evidence for association at the novel *IL16/STARD5/TMC3* locus on chromosome 15. Additional expression analyses confirmed the expression of *ME3*, *IL16*, and *STARD5* in multiple lung tissues, and found differential expression profiles of these three genes in the lungs of COPD patients compared to non-COPD controls. These results support the involvement of these implicated genes in the longitudinal change in lung function in the general population.

REFERENCES

1. Rabe KF, Hurd S, Anzueto A, et al. Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: GOLD executive summary. *Am J Respir Crit Care Med*. 2007;176(6):532-55.

2. Schunemann HJ, Dorn J, Grant BJ, et al. Pulmonary function is a long-term predictor of mortality in the general population: 29-year follow-up of the Buffalo Health Study. *Chest*. 2000;118(3):656-64.

3. Young RP, Hopkins R, Eaton TE. Forced expiratory volume in one second: not just a lung function test but a marker of premature death from all causes. *Eur Respir J*. 2007;30(4):616-22.

4. Fletcher C, Peto R. The natural history of chronic airflow obstruction. *Br Med J*. 1977;1(6077):1645-8.

5. Halbert RJ, Natoli JL, Gano A, et al. Global burden of COPD: systematic review and meta-analysis. *Eur Respir J*. 2006;28(3):523-32.

6. Eisner MD, Anthonisen N, Coultas D, et al. An official American Thoracic Society public policy statement: novel risk factors and the global burden of chronic obstructive pulmonary disease. *Am J Respir Crit Care Med*. 2010;182(5):693-718.

7. Gottlieb DJ, Wilk JB, Harmon M, et al. Heritability of longitudinal change in lung function. The Framingham study. *Am J Respir Crit Care Med*. 2001;164(9):1655-9.

8. Finkel D, Pedersen NL, Reynolds CA, et al. Genetic and environmental influences on decline in biobehavioral markers of aging. *Behav Genet*. 2003;33(2):107-23.

9. Hancock DB, Eijgelsheim M, Wilk JB, et al. Meta-analyses of genome-wide association studies identify multiple loci associated with pulmonary function. *Nat Genet*. 2009;42(1):45-52.

10. Repapi E, Sayers I, Wain LV, et al. Genome-wide association study identifies five loci associated with lung function. *Nat Genet*. 2009;42(1):36-44.

11. Soler Artigas M, Loth DW, Wain LV, et al. Genome-wide association and large-scale follow up identifies 16 new loci influencing lung function. *Nat Genet*. 2011;43(11):1082-90.

12. Imboden M, Bouzigon E, Curjuric I, et al. Genome-wide association study of lung function decline in adults with and without asthma. *The Journal of allergy and clinical immunology*. 2012;129(5):1218-28.

13. Miller MR, Hankinson J, Brusasco V, et al. Standardisation of spirometry. *Eur Respir J*. 2005;26(2):319-38.

14. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics*. 2010;26(17):2190-1.

15. Hansel NN, Ruczinski I, Rafaels N, et al. Genome-wide study identifies two loci associated with lung function decline in mild to moderate COPD. *Hum Genet*. 2013;132(1):79-90.

16. Dixon AL, Liang L, Moffatt MF, et al. A genome-wide association study of global gene expression. *Nat Genet*. 2007;39(10):1202-7.

17. Anthonisen NR, Connett JE, Kiley JP, et al. Effects of smoking intervention and the use of an inhaled anticholinergic bronchodilator on the rate of decline of FEV1. The Lung Health Study. *JAMA*. 1994;272(19):1497-505.

18. van Heel DA, Franke L, Hunt KA, et al. A genome-wide association study for celiac disease identifies risk variants in the region harboring IL2 and IL21. *Nat Genet*. 2007;39(7):827-9.

19. Kurima K, Yang Y, Sorber K, et al. Characterization of the transmembrane channel-like (TMC) gene family: functional clues from hearing loss and epidermodysplasia verruciformis. *Genomics*. 2003;82(3):300-8.

20. Vreugde S, Erven A, Kros CJ, et al. Beethoven, a mouse model for dominant, progressive hearing loss DFNA36. *Nat Genet*. 2002;30(3):257-8.

21. Ramoz N, Rueda LA, Bouadjar B, et al. Mutations in two adjacent novel genes are associated with epidermodysplasia verruciformis. *Nat Genet*. 2002;32(4):579-81.

22. Cruikshank WW, Kornfeld H, Center DM. Signaling and functional properties of interleukin-16. *International reviews of immunology*. 1998;16(5-6):523-40.

23. Bellini A, Yoshimura H, Vittori E, et al. Bronchial epithelial cells of patients with asthma release chemoattractant factors for T lymphocytes. *The Journal of allergy and clinical immunology*. 1993;92(3):412-24.

24. Cruikshank WW, Long A, Tarpy RE, et al. Early identification of interleukin-16 (lymphocyte chemoattractant factor) and macrophage inflammatory protein 1 alpha (MIP1 alpha) in bronchoalveolar lavage fluid of antigen-challenged asthmatics. *Am J Respir Cell Mol Biol*. 1995;13(6):738-47.

25. Krug N, Cruikshank WW, Tschernig T, et al. Interleukin 16 and T-cell chemoattractant activity in bronchoalveolar lavage 24 hours after allergen challenge in asthma. *Am J Respir Crit Care Med*. 2000;162(1):105-11.

26. Laberge S, Ernst P, Ghaffar O, et al. Increased expression of interleukin-16 in bronchial mucosa of subjects with atopic asthma. *Am J Respir Cell Mol Biol*. 1997;17(2):193-202.

27. Burkart KM, Barton SJ, Holloway JW, et al. Association of asthma with a functional promoter polymorphism in the IL16 gene. *The Journal of allergy and clinical immunology*. 2006;117(1):86-91.

28. Akesson LS, Duffy DL, Phelps SC, et al. A polymorphism in the promoter region of the human interleukin-16 gene is not associated with asthma or atopy in an Australian population. *Clinical and experimental allergy : journal of the British Society for Allergy and Clinical Immunology*. 2005;35(3):327-31.

29. Rodriguez-Agudo D, Ren S, Hylemon PB, et al. Human StarD5, a cytosolic StAR-related lipid binding protein. *Journal of lipid research*. 2005;46(8):1615-23.

30. Rodriguez-Agudo D, Calderon-Dominguez M, Medina MA, et al. ER stress increases StarD5 expression by stabilizing its mRNA and leads to relocalization of its protein from the nucleus to the membranes. *Journal of lipid research*. 2012;53(12):2708-15.

31. Chang GG, Tong L. Structure and function of malic enzymes, a new class of oxidative decarboxylases. *Biochemistry*. 2003;42(44):12721-33.

32. Moreadith RW, Lehninger AL. The pathways of glutamate and glutamine oxidation by tumor cell mitochondria. Role of mitochondrial NAD(P)+-dependent malic enzyme. *J Biol Chem*. 1984;259(10):6215-21.

33. Teller JK, Fahien LA, Davis JW. Kinetics and regulation of hepatoma mitochondrial NAD(P) malic enzyme. *J Biol Chem*. 1992;267(15):10423-32.

34. Wilk JB, Shrine NR, Loehr LR, et al. Genome Wide Association Studies Identify CHRNA5/3 and HTR4 in the Development of Airflow Obstruction. *Am J Respir Crit Care Med*. 2012.

35. Montgomery SB, Dermitzakis ET. From expression QTLs to personalized transcriptomics. *Nat Rev Genet*. 2011;12(4):277-82.

Table 3.1 Baseline characteristics of cohort studies included in the meta-analysis*

Cohort:	ARIC	B58C	BHS	CARDIA	CHS	FHS	Health ABC
No. of participants	8,242	827	1,009	1,492	3,159	3,230	1,586
No. of FEV_1 measurements	15,582	1,653	3,073	6,140	7,140	11,275	4,426
No. of FEV_1 per person	2	2	7	5	3	5	4
Follow-up duration, yr	5.6	10	29	20.1	7.9	14.7	9.5
Males, %	46.5	48.6	41.6	46.9	39	47	52.7
Baseline age, yr	54.6 (5.7)	35.0 (0.2)	37.5 (12.8)	27.5 (2.3)	72.3 (5.4)	50.9 (10.3)	73.8 (2.8)
Baseline height, cm	168.7 (9.4)	170.1 (9.5)	168.1 (8.9)	171.2 (9.3)	164.6 (9.4)	168.4 (9.3)	166.8 (9.3)
Current smokers, %	20.2	27.1	20.9	24.8	10.8	24.6	6.4
Former smokers, %	32.6	41.5	16.5	17.3	35.7	39.8	49.9
Baseline pack-years†	25.9 (21.7)	7.5 (11.4)	8.2 (17.8)	6.0 (6.5)	33.2 (27.0)	25.4 (21.3)	36.8 (32.2)
Baseline FEV_1, mL	2972 (758)	3631 (744)	3230 (927)	3818 (781)	2123 (652)	2989 (806)	2308 (649)
Baseline FEV_1/FVC, %	74.1 (7.1)	80.6 (5.8)	78.2 (9.2)	81.6 (6.5)	70.5 (10.5)	75.7 (8.0)	74.7 (7.8)

Cohort:	KORA	LBC1921	LBC1936	PIVUS	RS	SAPALDIA	SHIP
No. of participants	890	512	1,002	818	1,321	1,401	1,760
No. of FEV_1 measurements	1,597	706	1,790	1,469	2,016	2,692	2,571
No. of FEV_1 per person	2	2	2	2	2	2	2
Follow-up duration, yr	3.2	8.9	4.8	5.8	8.3	10.9	7.9
Males, %	47.2	41.4	50.8	49.9	45.1	48	49.4
Baseline age, yr	53.8 (4.5)	79.1 (0.6)	69.6 (0.8)	70.2 (0.2)	74.4 (5.6)	41.1 (11.2)	52.4 (13.6)
Baseline height, cm	169.3 (9.3)	163.2 (9.4)	166.5 (8.9)	169.0 (9.3)	167.3 (9.1)	169.4 (9.1)	169.5 (9.7)
Current smokers, %	20.5	7.0	12.9	10.2	11.1	26.9	32.8
Former smokers, %	40.9	50.4	42.6	39.6	56.7	25.8	23.8
Baseline pack-years†	11.2 (17.1)	15.3 (22.3)	16.9 (25.8)	14.3 (15.8)	25.7 (21.3)	17.4 (18.0)	11.3 (11.9)
Baseline FEV_1, mL	3280 (792)	1887 (625)	2371 (687)	2452 (682)	2215 (652)	3516 (861)	3238 (876)
Baseline FEV_1/FVC, %	77.5 (6.2)	79.0 (11.8)	78.3 (10.2)	76.0 (10.0)	74.8 (7.9)	78.5 (8.2)	83.1 (6.6)

Definition of abbreviations: ARIC = Atherosclerosis Risk in Communities; B58C = British 1958 Birth Cohort; BHS = Busselton Health Study; CARDIA = Coronary Artery Risk Development in Young Adults; CHS = Cardiovascular Health Study = FHS, Framingham Heart Study; Health ABC = Health, Aging, and Body Composition; KORA = Cooperative Health Research in the Region of Augsburg; LBC1921 = Lothian Birth Cohort 1921; LBC1936 = Lothian Birth Cohort 1936; PIVUS = Prospective Investigation of the Vasculature in Uppsala Seniors; RS = Rotterdam Study; SAPALDIA = Swiss Study on Air Pollution and Lung Diseases in Adults; SD = standard deviation; SHIP = Study of Health in Pomerania.

[*] Data are presented as mean (SD) unless otherwise indicated; total no. participants = 27249, total no. FEV_1 measurements = 62130.

[†] Pack-years are calculated among current and former smokers at study baseline**.**

Table 3.2 Model estimates for the rate of change in FEV_1 in never smokers and effects of other smoking patterns (compared with never smokers) on the rate of change in FEV_1 (mL/year)*

Study	Annual FEV_1 change in never smokers (referent group)		Additional Effect† of smoking patterns on annual FEV_1 change					
			Persistent smokers		Intermittent smokers		Former smokers	
	β	SE	β	SE	β	SE	β	SE
ARIC	-14.0	1.3	-12.4	1.7	-5.5	2.1	-5.3	1.4
B58C	-29.6	1.5	-9.4	2.8	-2.2	3.4	-3.0	3.0
BHS	-23.0	1.0	-20.0	3.0	-8.0	2.0	-9.0	2.0
CARDIA	-26.4	0.5	-6.7	1.3	-0.2	1.0	1.0	1.2
CHS	-35.0	1.1	-2.2	3.3	-4.6	2.2	-2.4	1.7
FHS	-26.0	0.6	-8.1	1.3	-2.9	1.0	-1.1	0.8
Health ABC	-39.7	1.3	-12.9	6.1	-6.8	4.4	-2.6	1.7
KORA	-22.1	3.7	2.2	7.2	-10.4	9.3	2.8	5.2
LBC1921	-10.0	3.6	-11.6	15.7	2.8	14.4	-18.8	4.9
LBC1936	-32.3	3.6	-19.0	9.9	40.1	16.8	4.3	5.3
PIVUS	-21.1	2.5	-15.9	8.2	-21.7	13.4	-3.9	3.9
RS	-27.5	3.7	-1.8	9.0	9.3	8.6	-4.6	4.5
SAPALDIA	-29.7	1.2	-7.4	2.3	-2.0	2.6	-2.8	2.1
SHIP	-31.8	2.8	-0.4	10.9	-0.1	3.9	-15.0	7.3
14-cohort meta-analyzed estimate	-26.9	0.3	-8.8	0.7	-2.6	0.6	-2.3	0.5

Definition of abbreviations: ARIC = Atherosclerosis Risk in Communities; B58C = British 1958 Birth Cohort; BHS = Busselton Health Study; CARDIA = Coronary Artery Risk Development in Young Adults; CHS = Cardiovascular Health Study; FHS = Framingham Heart Study; Health ABC = Health, Aging, and Body Composition; KORA = Cooperative Health Research in the

Region of Augsburg; LBC1921 = Lothian Birth Cohort 1921; LBC1936 = Lothian Birth Cohort 1936; PIVUS = Prospective Investigation of the Vasculature in Uppsala Seniors; RS = Rotterdam Study; SAPALDIA = Swiss Study on Air Pollution and Lung Diseases in Adults; SE = standard error; SHIP = Study of Health in Pomerania.

* Data shown are the effect estimates (β and SE) of the time and smoking-by-time interaction terms in the preliminary mixed effects model fully adjusted for all specified variables except the SNP terms. Time represents the rate of change in FEV_1 in never smokers and the smoking-by-time interaction term represents the effects of the other three smoking patterns on the rate of change in FEV_1, compared with never smokers. Smoking categories are defined as persistent (smoke throughout follow-up), intermittent (stop and/or start smoking during follow-up) and former (smoke only prior to start of follow-up).

† Effect estimates in smoking categories are added to estimate in never smokers to compute the actual rate of change in each group (for example, in ARIC, the point estimate of the rate of change in FEV_1 in persistent smokers was -14.0 - 12.4 = -26.4 mL/year).

Table 3.3 Association of the most statistically significant SNPs with the rate of change in FEV_1 (mL/year) in the meta-analysis of 14 cohort studies (n = 27249)*

SNP	Chr	Position	Closest Gene(s)	Coded Allele	Frequency	β	SE	*P* Value
rs12137475	1	44059735	*ST3GAL3*	T	0.11	-3.5	0.8	3.90×10^{-6}
rs766488	1	61583103	*NFIA*	A	0.31	1.4	0.3	6.60×10^{-6}
rs17698444	1	215483178	*ESRRG/GPATCH2*	C	0.89	-2.2	0.5	2.62×10^{-6}
rs12692550	2	159958017	*BAZ2B*	T	0.17	-1.7	0.4	5.16×10^{-6}
rs2260722	13	113236292	*TMCO3*	A	0.72	-1.5	0.3	1.83×10^{-6}
rs4077833	15	79419738	*IL16/STARD5/TMC3*	C	0.10	2.3	0.5	5.71×10^{-7}
rs8027498	15	89595638	*SV2B*	A	0.25	1.4	0.3	9.41×10^{-6}
rs8051319	16	15794449	*MYH11*	T	0.60	1.7	0.3	5.12×10^{-6}
rs740557	17	62451139	*CACNG4*	C	0.85	-2.3	0.5	3.59×10^{-6}

Definition of abbreviations: Chr = chromosome; SE = standard error; SNP = single-nucleotide polymorphism.

* Data reported are the meta-analysis results of the SNP-by-time interaction term from the GWAS mixed effects model. A positive β-coefficient indicates an attenuation of FEV_1 decline and a negative β-coefficient an acceleration of FEV_1 decline.

Table 3.4 Association of the most statistically significant SNPs with the rate of change in FEV_1 (mL/year) in the meta-analysis of the five cohort studies with ≥3 FEV_1 measurements per participant (n = 10476)

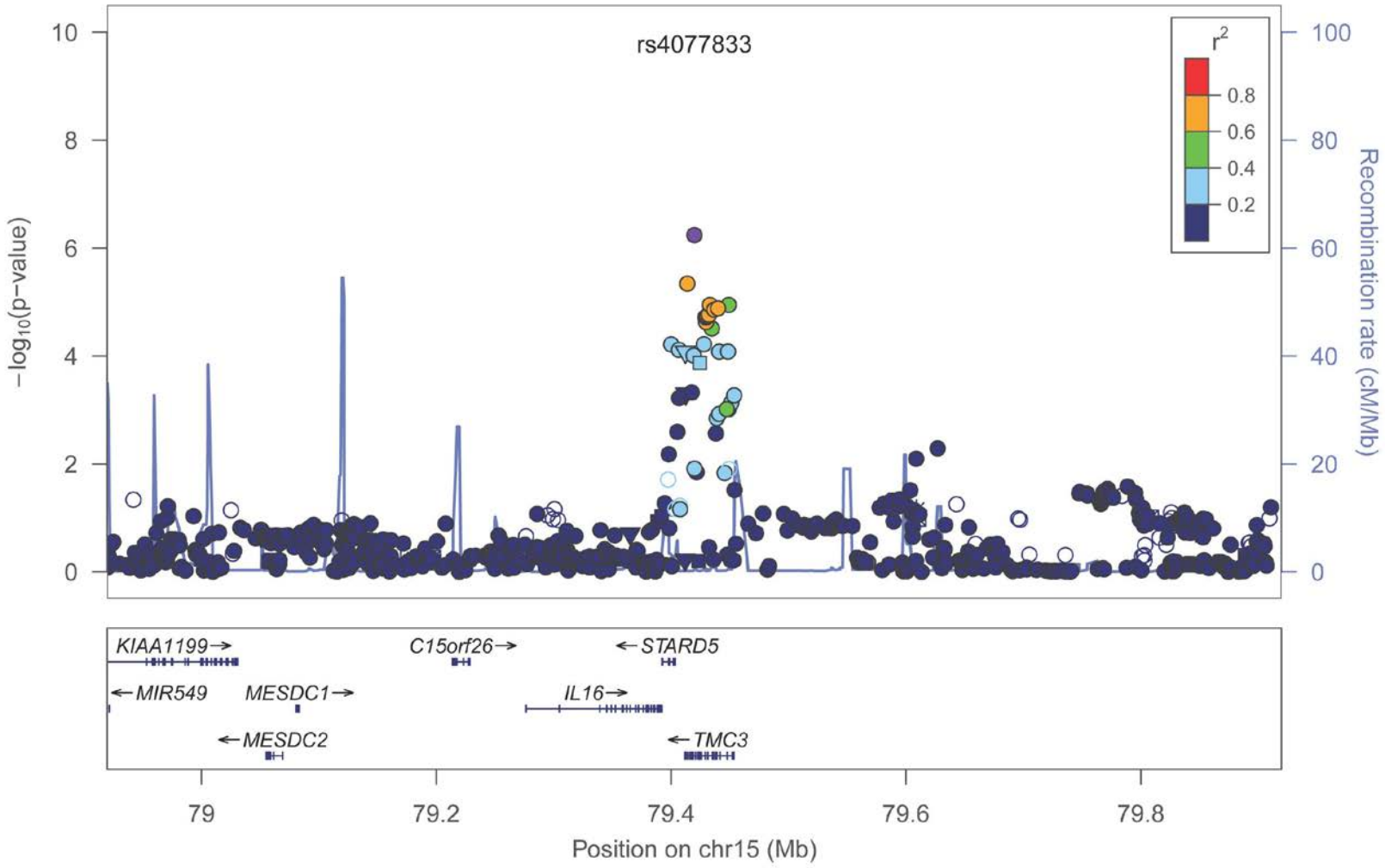
SNP	Chr	Position	Closest Gene(s)	Coded Allele	Frequency	β	SE	*P* Value
rs10209501	2	28536881	*FOSL2/PLB1*	A	0.33	1.6	0.4	7.09×10^{-6}
rs12692550	2	159958017	*BAZ2B*	T	0.18	-2.0	0.4	2.02×10^{-6}
rs1729588	3	110790025	*FLJ25363/MIR4445*	A	0.30	1.6	0.4	8.38×10^{-6}
rs10764053	10	19863644	*C10orf112*	T	0.47	1.5	0.3	4.15×10^{-6}
rs507211	11	86054387	*ME3*	A	0.25	2.1	0.4	2.18×10^{-8}

Definition of abbreviations: Chr = chromosome; SE = standard error; SNP = single-nucleotide polymorphism.

* Data reported are the meta-analysis results of the SNP-by-time interaction term from the GWAS mixed effects model. A positive β-coefficient indicates an attenuation of FEV_1 decline and a negative β-coefficient an acceleration of FEV_1 decline.

Figure 3.1 Association of the chromosome 15 locus with the rate of change in FEV_1 in the meta-analysis of 14 cohort studies. **A)** Regional association plot, where the X-axis is Megabase (Mb) position and Y-axes are the negative log of the *P* value on the left and recombination rate on the right. The sentinel SNP is colored in purple and linkage disequilibrium to the sentinel SNP is depicted by degree of color according to the legend. **B)** Forest plot for rs4077833, where the size of the square for each study represents its contributing weight to the meta-analysis.

A)


rs4077833
r²
0.8
0.6
0.4
0.2
−log10(p−value)
Recombination rate (cM/Mb)
KIAA1199
C15orf26
STARD5
MIR549
MESDC1
IL16
MESDC2
TMC3
79
79.2
79.4
79.6
79.8
Position on chr15 (Mb)

B)

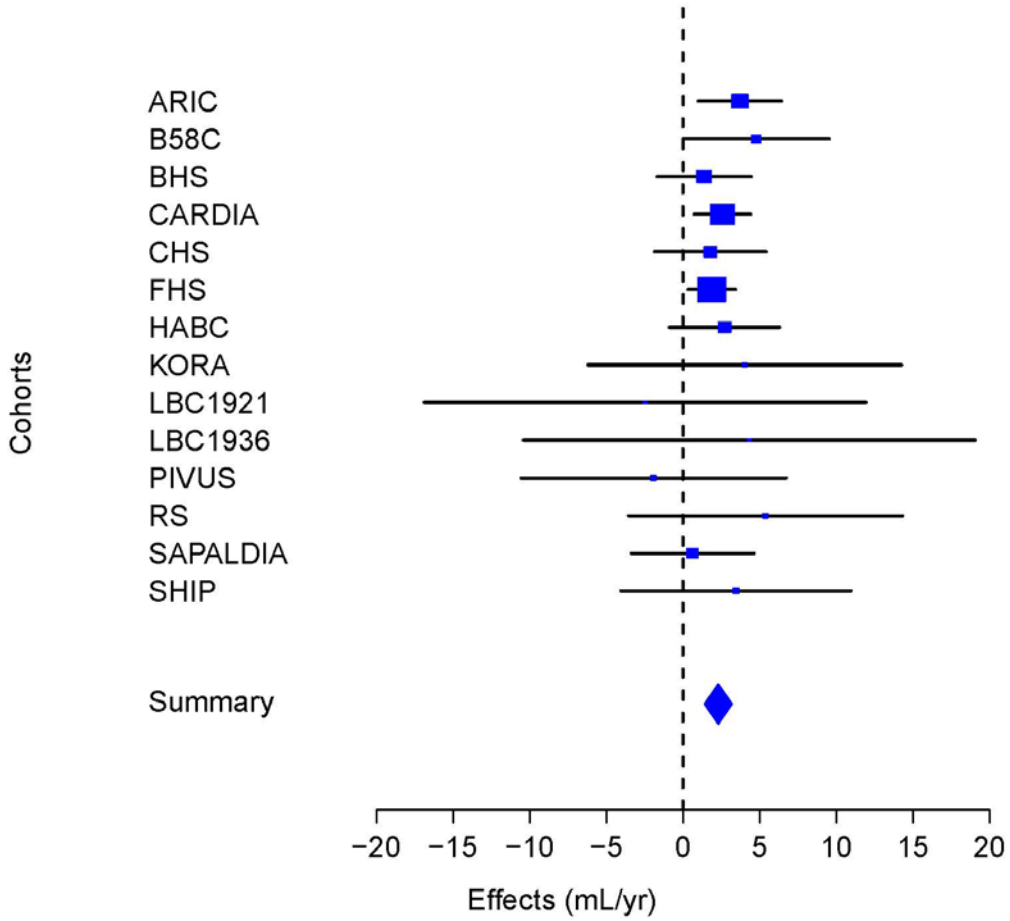
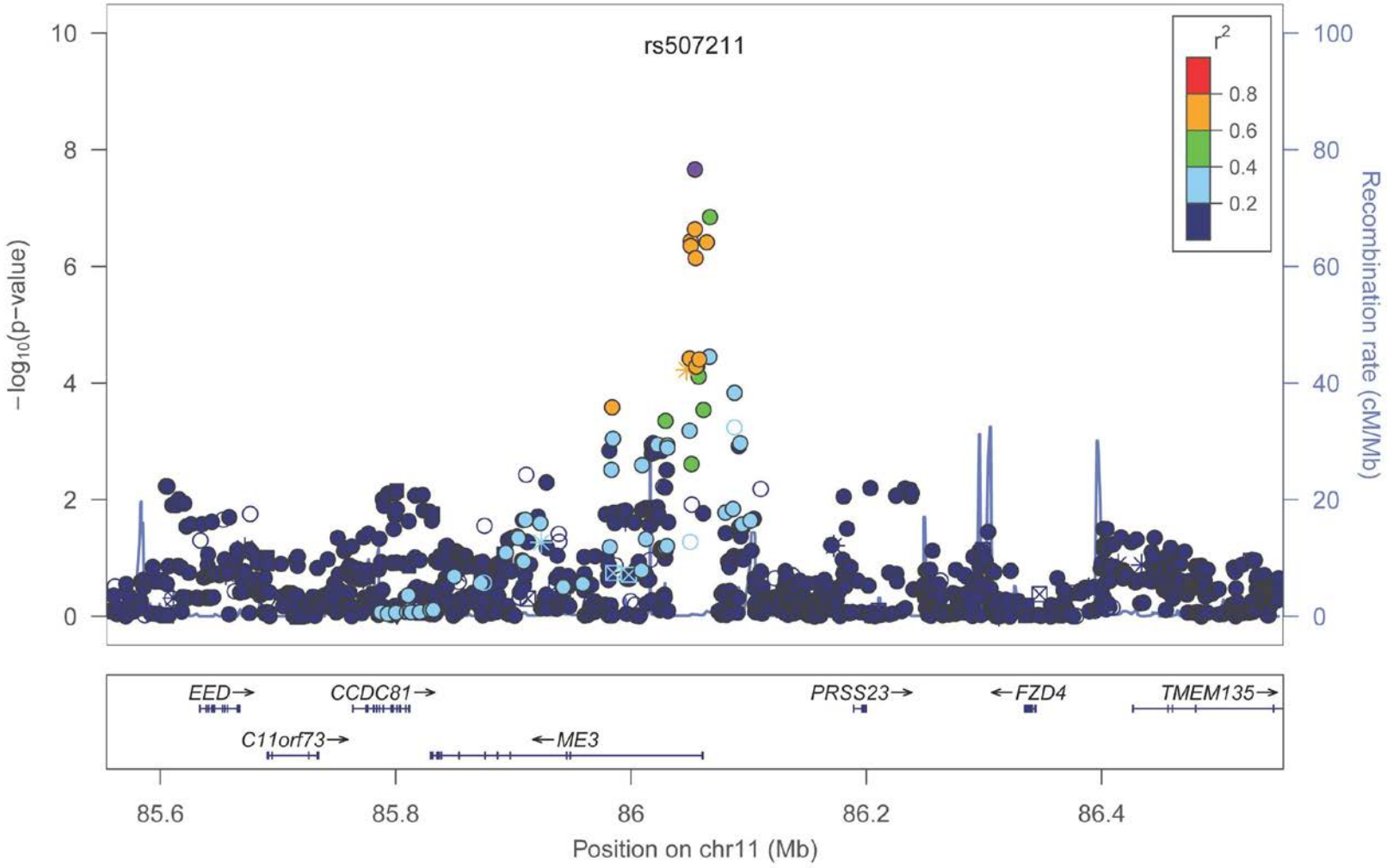
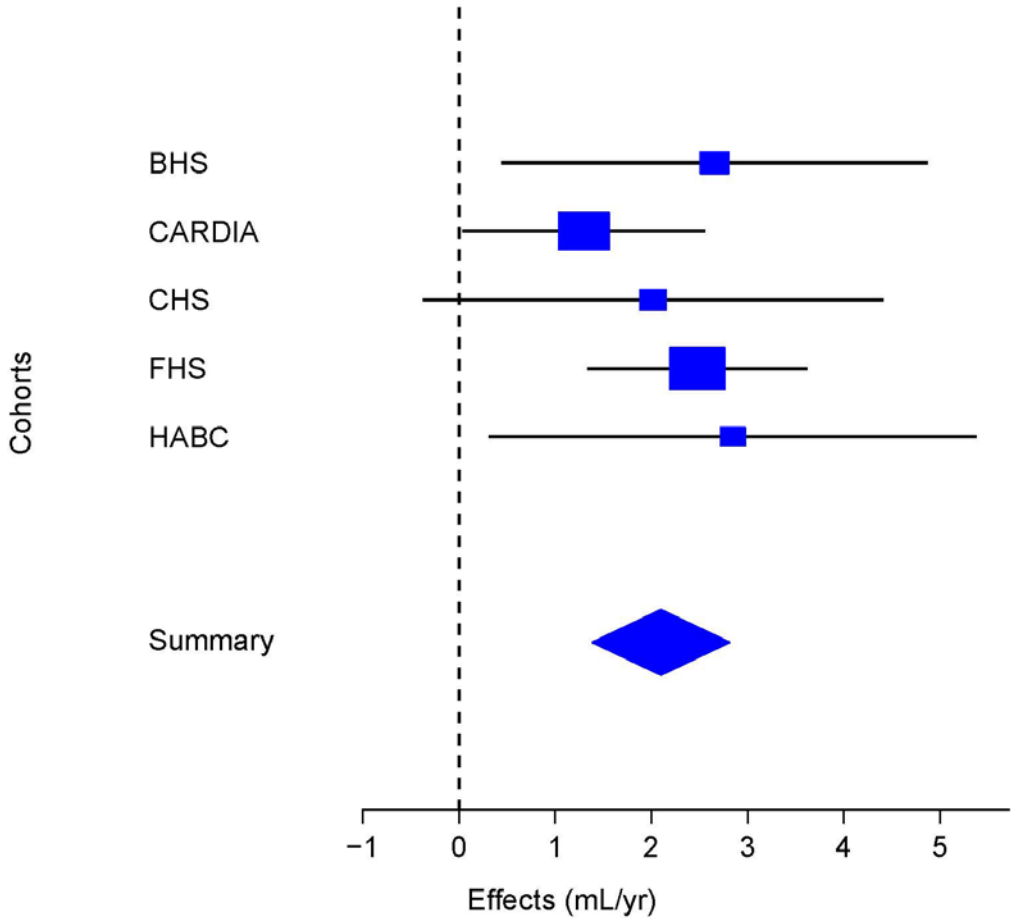

ARIC
B58C
BHS
CARDIA
CHS
FHS
HABC
KORA
LBC1921
LBC1936
PIVUS
RS
SAPALDIA
SHIP
Summary
Cohorts
−20
−15
−10
−5
0
5
10
15
20
Effects (mL/yr)

Figure 3.2 Association of the chromosome 11 locus with the rate of change in FEV_1 in the meta-analysis of the five cohort studies with ≥3 FEV_1 measurements per participant. **A)** Regional association plot, where the X-axis is Megabase (Mb) position, and the Y-axes are the negative log of the *P* value on the left and recombination rate on the right. The sentinel SNP is colored in purple and linkage disequilibrium to the sentinel SNP is depicted by degree of color according to the legend. **B)** Forest plot for rs507211, where the size of the square for each study represents its contributing weight to the meta-analysis.

A)


rs507211
r2
0.8
0.6
0.4
0.2
−log10(p−value)
Recombination rate (cM/Mb)
10
8
6
4
2
0
100
80
60
40
20
0
EED→
CCDC81→
PRSS23→
←FZD4
TMEM135→
C11orf73→
←ME3
85.6
85.8
86
86.2
86.4
Position on chr11 (Mb)

B)


BHS
CARDIA
CHS
FHS
HABC
Summary
Cohorts
−1
0
1
2
3
4
5
Effects (mL/yr)

ADDITIONAL STUDY INFORMATION

Authors:

Wenbo Tang[1]*, Matthew Kowgier[2]*, Daan W. Loth[3,4]*, María Soler Artigas[5,6]*, Bonnie R. Joubert[7]*, Emily Hodge[8]*, Sina A. Gharib[9]*, Albert V. Smith[10,11], Ingo Ruczinski[12], Vilmundur Gudnason[10,11], Rasika A. Mathias[13], Tamara B. Harris[14], Nadia N. Hansel[13], Lenore J. Launer[14], Kathleen C. Barnes[13], Joyanna G. Hansen[1], Eva Albrecht[15], Melinda C. Aldrich[16], Michael Allerhand[17], R. Graham Barr[18,19], Guy G. Brusselle[3,20,21], David J. Couper[22], Ivan Curjuric[23,24], Gail Davies[17,25,26], Ian J. Deary[17,26], Josée Dupuis[27,28], Tove Fall[29], Millennia Foy[30], Nora Franceschini[31], Wei Gao[27], Sven Gläser[32], Xiangjun Gu[30], Dana B. Hancock[7,33], Joachim Heinrich[34], Albert Hofman[3,35], Medea Imboden[23,24], Erik Ingelsson[29,36], Alan James[37], Stefan Karrasch[38,39,40], Beate Koch[32], Stephen B. Kritchevsky[41], Ashish Kumar[23,24,36], Lies Lahousse[3,20], Guo Li[42], Lars Lind[43], Cecilia Lindgren[36,44], Yongmei Liu[45], Kurt Lohman[46], Thomas Lumley[47], Wendy L. McArdle[48], Bernd Meibohm[49], Andrew P. Morris[36], Alanna C. Morrison[50], Bill Musk[37], Kari E. North[31], Lyle Palmer[2,51,52], Nicole M. Probst-Hensch[23,24], Bruce M. Psaty[42,53,54,55], Fernando Rivadeneira[35,56], Jerome I. Rotter[57], Holger Schulz[34], Lewis J. Smith[58], Akshay Sood[59], John M. Starr[17,60], David P. Strachan[61], Alexander Teumer[62], André G. Uitterlinden[35,56], Henry Völzke[63], Arend Voorman[64], Louise V. Wain[6,65], Martin T. Wells[66], Jemma B. Wilk[28,67], O. Dale Williams[68], Susan R. Heckbert[42,53,54], Bruno H. Stricker[3,4], Stephanie J. London[7], Myriam Fornage[30,50]**, Martin D. Tobin[5,6]**, George T. O'Connor[28,69]**, Ian P. Hall[8]**, Patricia A. Cassano[1,70]**

*starred first authors

**starred last authors

Affiliations:

1. Division of Nutritional Sciences, Cornell University, Ithaca, NY, USA
2. Ontario Institute for Cancer Research, Toronto, Ontario, Canada
3. Department of Epidemiology, Erasmus Medical Center, Rotterdam, the Netherlands
4. Netherlands Healthcare Inspectorate, The Hague, the Netherlands

5. University of Leicester, Genetic Epidemiology Group, Department of Health Sciences, Leicester, UK
6. National Institute for Health Research (NIHR) Leicester Respiratory Biomedical Research Unit, Glenfield Hospital, Leicester, UK
7. Epidemiology Branch, National Institute of Environmental Health Sciences, National Institutes of Health, U.S. Department of Health and Human Services, Research Triangle Park, NC, USA
8. Division of Therapeutics and Molecular Medicine, Nottingham Respiratory Biomedical Research Unit, University Hospital of Nottingham, Nottingham, UK
9. Computational Medicine Core, Center for Lung Biology, Division of Pulmonary & Critical Care Medicine, Department of Medicine, University of Washington, Seattle, WA, USA
10. Icelandic Heart Association, Kopavogur, Iceland
11. University of Iceland, Reykjavik, Iceland
12. Department of Biostatistics, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD, USA
13. Department of Medicine, School of Medicine, Johns Hopkins University, Baltimore, MD, USA
14. Laboratory of Epidemiology, Demography, and Biometry, National Institute on Aging, National Institutes of Health, Bethesda, MD, USA
15. Institute of Genetic Epidemiology, Helmholtz Zentrum München - German Research Center for Environmental Health, Neuherberg, Germany
16. Department of Thoracic Surgery and Division of Epidemiology, Vanderbilt University Medical Center, Nashville, TN, USA
17. Centre for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh, Edinburgh, UK
18. Division of General Medicine, Pulmonary, Allergy and Critical Care, Department of Medicine, College of Physicians and Surgeons, Columbia University, New York, NY, USA

19. Department of Epidemiology, Mailman School of Public Health, Columbia University, New York, NY, USA
20. Department of Respiratory Disease, University Hospital Ghent, Ghent, Belgium
21. Department of Respiratory Medicine, Erasmus Medical Center, Rotterdam, the Netherlands
22. Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA
23. Swiss Tropical and Public Health Institute, Basel, Switzerland
24. University of Basel, Basel, Switzerland
25. Medical Genetics Section, University of Edinburgh Molecular Medicine Centre and MRC Institute of Genetics and Molecular Medicine, Western General Hospital, Edinburgh, UK
26. Department of Psychology, University of Edinburgh, Edinburgh, UK
27. Biostatistics Department, Boston University School of Public Health, Boston, MA, USA
28. The National Heart, Lung, and Blood Institute's Framingham Heart Study, Framingham, MA, USA
29. Department of Medical Sciences, Molecular Epidemiology and Science for Life Laboratory, Uppsala University, Uppsala, Sweden
30. Institute of Molecular Medicine, University of Texas Health Science Center at Houston, Houston, TX, USA
31. Gillings School of Global Public Health, Department of Epidemiology, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA
32. Department of Internal Medicine B; Pneumology, Cardiology, Intensive Care Medicine; Field of Research: Pneumology and Pneumological Epidemiology, University Medicine Greifswald, Germany
33. Behavioral Health Epidemiology Program, Research Triangle Institute, Research Triangle Park, NC, USA
34. Institute of Epidemiology I, Helmholtz Zentrum München - German Research Center for Environmental Health, Neuherberg, Germany
35. Netherlands Consortium for Healthy Aging, Rotterdam, the Netherlands

36. Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK
37. School of Medicine and Pharmacology, University of Western Australia, Perth, Western Australia, Australia
38. Institute and Outpatient Clinic for Occupational, Social and Environmental Medicine, Ludwig-Maximilians-Universität, Munich, Germany
39. Institute of General Practice, University Hospital Klinikum rechts der Isar, Technische Universität München, Munich, Germany
40. Institute of Epidemiology I, Helmholtz Zentrum München - German Research Center for Environmental Health, Neuherberg, Germany
41. Sticht Center on Aging, Wake Forest School of Medicine, Winston-Salem, NC, USA
42. Cardiovascular Health Research Unit, University of Washington, Seattle, WA, USA
43. Department of Medical Sciences, Uppsala University, Uppsala, Sweden
44. Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA
45. Department of Epidemiology and Prevention, Division of Public Health Sciences, Wake Forest School of Medicine, Winston-Salem, NC, USA
46. Department of Biostatistical Sciences, Division of Public Health Sciences, Wake Forest School of Medicine, Winston-Salem, NC, USA
47. Department of Statistics, University of Auckland, Auckland, New Zealand
48. School of Social and Community Medicine, University of Bristol, Bristol, UK
49. College of Pharmacy, University of Tennessee Health Science Center, Memphis, TN, USA
50. Human Genetics Center, School of Public Health, University of Texas Health Science Center at Houston, Houston, TX, USA
51. Epidemiology and Obstetrics & Gynaecology, University of Toronto, Toronto, Ontario, Canada
52. Samuel Lunenfeld Research Institute, Toronto, Ontario, Canada
53. Department of Epidemiology, University of Washington, Seattle, WA, USA
54. Group Health Research Institute, Group Health Cooperative, Seattle, WA, USA
55. Department of Medicine, University of Washington, Seattle, WA, USA
56. Department of Internal Medicine, Erasmus Medical Center, Rotterdam, the Netherlands

57. Institute for Translational Genomics and Population Sciences, Los Angeles Biomedical Research Institute at Harbor-UCLA Medical Center, Torrance, CA, USA
58. Northwestern University Feinberg School of Medicine, Chicago, IL, USA
59. University of New Mexico, NM, USA
60. Alzheimer Scotland Dementia Research Centre, University of Edinburgh, Edinburgh, UK
61. Division of Population Health Sciences and Education, St George's, University of London, London, UK
62. Department for Genetics and Functional Genomics, Interfaculty Institute for Genetics and Functional Genomics, University Medicine Greifswald, Greifswald, Germany
63. Institute for Community Medicine, Study of Health In Pomerania (SHIP)/Clinical Epidemiological Research, University Medicine Greifswald, Greifswald, Germany
64. Department of Biostatistics, University of Washington, Seattle, WA, USA
65. Genetic Epidemiology Group, Department of Health Sciences, University of Leicester, Leicester, UK
66. Department of Statistical Science, Cornell University, Ithaca, NY, USA
67. Division of Aging, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA
68. Florida International University, Miami, FL, USA
69. Section of Pulmonary, Allergy, and Critical Care Medicine, Department of Medicine, Boston University School of Medicine, Boston, MA, USA
70. Department of Public Health, Division of Biostatistics and Epidemiology, Weill Cornell Medical College, New York, NY, USA

Correspondence: Address all correspondence to Dr. Patricia A Cassano, 209 Savage Hall, Division of Nutritional Sciences, Ithaca NY or pac6@cornell.edu, phone 607-255-7551, fax 607-255-1033.

Author contributions:

AGES Study concept and design: TBH, VG, LJL; Genotype data acquisition/QC: AVS; Data analysis: AVS; **ARIC** Study concept and design: DJC, NF, DBH, BRJ, SJL, ACM, KEN;

Phenotype data acquisition/QC: DJC; Genotype data acquisition/QC: ACM, KEN; Data analysis: BRJ, SJL; **B58C** Study concept and design: DPS; Phenotype data acquisition/QC: DPS; Genotype data acquisition/QC: WLM; Data analysis: DPS, LVW; **BHS** Study concept and design: AJ, BM; Phenotype data acquisition/QC: AJ, BM; Genotype data acquisition/QC: AJ, BM, LP; Data analysis: MK, LP; **CARDIA** Study concept and design: MF, AS; Phenotype data acquisition/QC: LJS, AS, ODW; Genotype data acquisition/QC: MF, MF, XG; Data analysis: MF, MF, XG; **CHS** Study concept and design: SRH, SAG, BMP; Phenotype data acquisition/QC: BMP; Genotype data acquisition/QC: TL, BMP, JIR; Data analysis: SAG, SRH, GL, TL, AV; **FHS** Study concept and design: JD, GTO, JBW; Phenotype data acquisition/QC: GTO; Genotype data acquisition/QC: GTO; Data analysis: JD, WG, JBW; **Health ABC** Study concept and design: PAC, SBK, WT; Phenotype data acquisition/QC: PAC, SBK, BM, WT; Genotype data acquisition/QC: SBK, YL, KL; Data analysis: PAC, YL, KL, WT, MTW; **KORA** Study concept and design: JH, HS; Phenotype data acquisition/QC: JK, SK, HS; Data analysis: EA; **LBC** Study concept and design: IJD, JMS; Phenotype data acquisition/QC: IJD, JMS; Genotype data acquisition/QC: GD; Data analysis: MA, GD; **LHS** Study concept and design: KCB, NNH, RAM; Genotype data acquisition/QC: KCB, RAM, IR; Data analysis: KCB, NNH, RAM, IR; **RS** Study concept and design: GGB, AH, FR, BHS, AGU; Phenotype data acquisition/QC: GGB, LL, DWL, BHS; Genotype data acquisition/QC:FR, AGU; Data analysis: LL, DWL; **SAPALDIA** Study concept and design: MI, NMP-H; Phenotype data acquisition/QC: IC, MI, NMP-H; Genotype data acquisition/QC: IC, AK, MI, NMP-H; Data analysis: MI; **SHIP** Study concept and design: BK, SG, HV; Phenotype data acquisition/QC: BK, SG, HV; Genotype data acquisition/QC: BK, SG, AT, HV; Data analysis: AT; **SpiroMeta** Study concept and design: IPH, MDT; Genotype data acquisition/QC: IPH, MDT; Data analysis: MSA, IPH, MDT; PIVUS Phenotype data acquisition/QC: EI, LL; Genotype data acquisition/QC: EI, LL, APM; Data analysis: TF;

Critical revision of manuscript: All authors

Sources of funding:

The AGES-Reykjavik Study is funded by NIH contract N01-AG-12100, the NIA Intramural Research Program, Hjartavernd (the Icelandic Heart Association), and the Althingi (the Icelandic Parliament).

The Atherosclerosis Risk in Communities Study is carried out as a collaborative study supported by National Heart, Lung, and Blood Institute contracts (HHSN268201100005C, HHSN268201100006C, HHSN268201100007C, HHSN268201100008C, HHSN268201100009C, HHSN268201100010C, HHSN268201100011C, and HHSN268201100012C), R01HL087641, R01HL59367 and R01HL086694; National Human Genome Research Institute contract U01HG004402; and National Institutes of Health contract HHSN268200625226C. The authors thank the staff and participants of the ARIC study for their important contributions. Infrastructure was partly supported by Grant Number UL1RR025005, a component of the National Institutes of Health and NIH Roadmap for Medical Research. Work was supported in part by the Division of Intramural Research, National Institute of Environmental Health Sciences ZO1 ES43012.

We acknowledge 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.

The 1994 Busselton follow-up Health Study was supported by Healthways, Western Australia. The Busselton Health Study is supported by The Great Wine Estates of the Margaret River region of Western Australia. The study gratefully acknowledges the assistance of the Western Australian DNA Bank (NHMRC Enabling Facility) with DNA samples and the support provided by The Ark at University of Western Australia for this study.

The Coronary Artery Risk Development in Young Adults (CARDIA) study was funded by contracts N01-HC-95095, N01-HC-48047, N01-HC-48048, N01-HC-48049, N01-HC-48050, N01-HC-45134, N01-HC-05187, N01-HC-45205, and N01-HC-45204 from NHLBI to the CARDIA investigators. Genotyping of the CARDIA participants was supported by grants U01-HG-004729, U01-HG-004446, and U01-HG-004424 from the NHGRI. Statistical analyses were supported by grants U01-HG-004729 and R01-HL-084099 to MF.

The Cardiovascular Health Study was supported by NHLBI contracts HHSN268201200036C, N01 HHSN268200800007C, N01HC55222, N01HC85079, N01HC85080, N01HC85081, N01HC85082, N01HC85083, N01HC85086; and NHLBI grants HL080295, HL087652, HL105756 with additional contribution from the National Institute of Neurological Disorders and Stroke (NINDS). Additional support was provided through AG023629 from the National Institute on Aging (NIA). A full list of CHS investigators and institutions can be found at http://www.chs-nhlbi.org/pi.htm. DNA handling and genotyping at Cedars-Sinai Medical Center was supported in part by the National Center for Research Resources grant UL1RR033176, now at the National Center for Advancing Translational Sciences CTSI grant UL1TR000124; the National Institute of Diabetes and Digestive and Kidney Disease grant DK063491 to the Southern California Diabetes Endocrinology Research Center; and the Cedars-Sinai Board of Governors' Chair in Medical Genetics (JIR).

Framingham Heart Study (FHS) research was conducted in part using data and resources of the NHLBI and Boston University School of Medicine. The analyses reflect intellectual input and resource development from the FHS investigators participating in the SNP Health Association Resource (SHARe) project. This work was partially supported by

NHLBI (contract no. N01-HC-25195) and its contract with Affymetrix for genotyping services (contract no. N02-HL-6-4278). A portion of this research utilized the Linux Cluster for Genetic Analysis (LinGA-II) funded by the Robert Dawson Evans Endowment of the Department of Medicine at Boston University School of Medicine and Boston Medical Center. JBW was supported by a Young Clinical Scientist Award from the Flight Attendant Medical Research Institute (FAMRI).

The Health, Aging, and Body Composition Study was supported by NIA contracts N01AG62101, N01AG2103, and N01AG62106, NIA grant R01-AG028050, NINR grant R01-NR012459, and in part by the Intramural Research Program of the NIA, NIH. The genome-wide association study was funded by NIA grant 1R01AG032098–01A1 to Wake Forest Health Sciences, and genotyping services were provided by the Center for Inherited Disease Research, which is fully funded through a federal contract from the National Institutes of Health to The Johns Hopkins University, contract number HHSN268200782096C. This research was further supported by RC1AG035835.

The KORA study was funded by the Helmholtz Zentrum München - German Research Center for Environmental Health, German Federal Ministry of Education and Research, State of Bavaria, Munich Center of Health Sciences (MC Health), Ludwig-Maximilians-Universität, as part of LMUinnovativ, and Competence Network ASCONET, subnetwork COSYCONET (FKZ 01GI0882).

The Lothian Birth Cohorts 1921 and 1936 were funded by the Lifelong Health and Wellbeing Initiative (BBSRC, EPSRC, ESRC and MRC). The authors thank the cohort participants and team members who contributed to these studies. Phenotype collection in the Lothian Birth Cohort 1921 was supported by the BBSRC, The Royal Society and The Chief Scientist Office of the Scottish Government. Phenotype collection in the Lothian Birth Cohort 1936 was supported by Research Into Ageing (continues as part of Age UK The Disconnected Mind project). Genotyping of the cohorts was funded by the UK Biotechnology and Biological Sciences Research Council (BBSRC). The work was undertaken by the University of Edinburgh Centre for Cognitive Ageing and Cognitive Epidemiology, part of the cross council Lifelong Health and Wellbeing Initiative (G0700704/84698). Funding from the

BBSRC, Engineering and Physical Sciences Research Council (EPSRC), Economic and Social Research Council (ESRC), and MRC is gratefully acknowledged.

The Lung Health Study (LHS) was supported by GENEVA (U01HG 004738), and by the Mary Beryl Patch Turnbull Scholar Program (KCB, in part).

The PIVUS study was funded by the Swedish Foundation for Strategic Research (ICA08-0047), the Swedish Research Council (2012-1397), the Swedish Heart-Lung Foundation (20120197), the Swedish Society of Medicine, and Uppsala University. CML is a Wellcome Trust Research Career Development Fellow (086596/Z/08/Z). The computations were performed on resources provided by SNIC through Uppsala Multidisciplinary Center for Advanced Computational Science (UPPMAX) under Project p2013056. APM acknowledges funding from the Wellcome Trust under grants WT098017, WT064890 and WT090532.

The Rotterdam Studies were funded by the Netherlands Organization of Scientific Research NWO Investments, nr. 175.010.2005.011, 911-03-012, Research Institute for Diseases in the Elderly, 014-93-015; RIDE2, the Netherlands Genomics Initiative (NGI)/Netherlands Organization for Scientific Research (NWO) project nr. 050-060-810, the Ministry of Education, Culture and Science, the Ministry for Health, Welfare and Sports, the European Commission (DG XII) Municipality of Rotterdam.

SAPALDIA was supported by the Swiss National Science Foundation (grants no 33CS30-134276/1, 33CSCO-108796, 3247BO-104283, 3247BO-104288, 3247BO-104284, 3247-065896, 3100-059302, 3200-052720, 3200-042532, 4026-028099, 3233-054996, PDFMP3-123171), the Federal Office for Forest, Environment, and Landscape, the Federal Office of Public Health, the Federal Office of Roads and Transport, the canton's government of Aargau, Basel-Stadt, Basel-Land, Geneva, Luzern, Ticino, Valais, Zurich, the Swiss Lung League, the canton's Lung League of Basel Stadt/Basel Landschaft, Geneva, Ticino, Valais and Zurich, Schweizerische Unfallversicherungsanstalt (SUVA), Freiwillige Akademische Gesellschaft, UBS Wealth Foundation, Talecris Biotherapeutics GmbH, Abbott Diagnostics. Genotyping in the GABRIEL framework was supported by grants European Commission 018996 and Wellcome Trust WT 084703MA.

SHIP is part of the Community Medicine Research net of the University of Greifswald, Germany, which is funded by the Federal Ministry of Education and Research, the Ministry of

Cultural Affairs as well as the Social Ministry of the Federal State of Mecklenburg-West Pomerania, and the network ‘Greifswald Approach to Individualized Medicine (GANI_MED)’ funded by the Federal Ministry of Education and Research, and the German Asthma and COPD Network (COSYCONET), under 01ZZ9603, 01ZZ0103, 01ZZ0403, 03IS2061A, and BMBP 01GI0883. Genome-wide data have been supported by the Federal Ministry of Education and Research and a joint grant from Siemens Healthcare, Erlangen, Germany and the Federal State of Mecklenburg- West Pomerania (03ZIK012). The University of Greifswald is a member of the ‘Center of Knowledge Interchange’ program of the Siemens AG and the Caché Campus program of the InterSystems GmbH.

The research undertaken by MDT, LVW and MSA was partly funded by the National Institute for Health Research (NIHR). The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health. MDT holds a Medical Research Council Senior Clinical Fellowship (G0902313).

The expression analysis undertaken by IPH and EH was funded by the Medical Research Council of UK (grant no. G1000861).

ONLINE DATA SUPPLEMENT

Methods

Pulmonary function assessment

Study-specific information on spirometry protocols was reported previously (1). Additional descriptions of spirometry tests specific to this longitudinal study are provided below.

ARIC measurements were completed at two time points (baseline and approximately 3 years later) in accordance with the standardized guidelines of the American Thoracic Society (ATS), as previously described (2). Measurements were made with a Collins Survey II water-seal spirometer (Collins Medical, Inc.) and Pulmo-Screen II software (PDS Healthcare Products, Inc.).

The **British 1958 birth cohort (B58C)** is a longitudinal study of all people born in England, Scotland and Wales during one week in 1958. At age 35 years, spirometry was performed in 1156 cohort members with a history of asthma, wheezy bronchitis or pneumonia during childhood or early adult life, and a subsample of 293 cohort members with no history of any of these conditions (3). At age 45 years, measurements of ventilatory function were repeated as part of a more general biomedical examination of the entire cohort (4). The results contributed to the present meta-analysis relate to 827 individuals with valid spirometry on both occasions (5), plus genome-wide genotyping performed on DNA samples collected at the 45-year follow-up. On both occasions, spirometry was performed in the home by trained research nurses, after daily calibration of the instrument using a 1L syringe. Measurements were taken in the standing position, without noseclips. At age 35 years, at least three (and up to eight) forced expiratory maneuvers were recorded by dry bellows spirometer (Vitalograph

R-model or S-model; Vitalograph, Buckingham, UK) until two technically satisfactory measurements of forced expiratory volume in the first second (FEV_1) within 5% of each other were obtained. At age 45 years, at least three (and up to five) spirograms were recorded by pneumotachograph (Micro; Vitalograph, Buckingham, UK) until three technically satisfactory blows had been obtained. On each occasion, the highest technically satisfactory values of FEV_1 and forced vital capacity (FVC) were used in the analysis.

The **Busselton Health Study (BHS)** is a longitudinal survey of the town of Busselton in the south-western region of Western Australia that began in 1966. Lung function was measured at 9 time points. Spirometric measures of forced expired volume in one second (FEV_1) and forced vital capacity (FVC) were assessed as described previously (6, 7).

Spirometry was performed at **CARDIA** years 0, 2, 5, 10 and 20 examination visits, adhering to the ATS guidelines. At CARDIA years 0, 2, 5 and 10 examination visits, spirometry was performed using the Collins Survey 8-liter water-sealed spirometer and the Eagle II microprocessor (Warren E. Collins, Inc., Braintree, MA) in a sitting position with noseclips, as per the 1979 ATS criteria (8). Specifically, each subject performed a minimum of three trials with expirations recorded to the FVC plateau, which occurs after six seconds of expiration and was maintained for at least one second before terminating the forced expiratory maneuver. If, at the end of the three trials, there were at least three acceptable tracings, and with the maximum FVC and FEV_1 reproduced to within 5% or 100 mL, whichever is greater, no more trials were performed. At CARDIA year 20 examination visit, a dry rolling-seal SensorMedics model 1022 spirometer fitted by OMI (Viasys Corp, Loma Linda, CA) was used for spirometry testing in a standing position with noseclips. The criteria for reproducibility were changed at the year 20 visit - the two largest FVC values were to agree

within 150 ml, and the two largest FEV_1 values were also to agree within 150 ml, consistent with the 1994 update by the ATS (9). A comparability study performed on 25 volunteers at the LDS Hospital (Salt Lake City, UT) demonstrated excellent consistency between the old and new machines; the average difference between the Collins Survey and OMI spirometer was 6 mL for FVC and 21 mL for FEV_1.

In **CHS**, spirometry was completed on three occasions (baseline and after 4 and 7 years of follow-up) for the original cohort recruited in 1989-90. The spirometry procedures for pulmonary function testing have been previously described (10, 11). Briefly, spirometry technicians were centrally trained and certified prior to recruitment of participants. A standard spirometry system, including a Collins Survey I water-seal spirometer (Collins Medical, Inc., Braintree, Massachusetts) and software from S&M Instruments (Doylestown, Pennsylvania), was used by technicians at all four recruitment centers. Stringent quality assurance procedures for spirometry testing exceeded ATS recommendations (10).

In **FHS**, spirometry at the 5^{th}, 6^{th}, and 7^{th} Offspring Cohort examinations was performed using a Collins Survey II spirometer (Collins Medical, Inc., Braintree, MA), interfaced to pulmonary function data acquisition and quality control software (S and M Instruments, Doylestown, PA) and calibrated daily. Spirometry at the 8^{th} Offspring Cohort examination was performed using a Collins CPL system (nSpire Health Inc., Longmont, CO) which was calibrated daily. Spirometric maneuvers were performed according to contemporaneous ATS (9, 12) or European Respiratory Society (ERS)-ATS standards (13).

Spirometry in **Health ABC** was completed at four time points (baseline, years 4, 7 and 9) in accordance with standardized guidelines of the ATS, as previously reported (14). The study used a horizontal, dry rolling seal HF6 Spirometer (Sensor Medics Corporation, Yorba

Linda, CA, USA) during clinical visits, and the EasyOne Model 2001 diagnostic spirometer (ndd Medizintechnik AG, Zurich, Switzerland) during home visits starting in year 8. The two devices were evaluated for comparability and provided virtually identical values. All FEV_1 measures meeting the ATS criteria for acceptability were included in the current study.

The **KORA** studies (Cooperative Health Research in the Region of Augsburg) are a series of independent population based studies from the general population living in the region of Augsburg, southern Germany (15, 16). KORA F4 including 3,080 individuals was conducted from 2006-08. Baseline lung function tests were performed in random subsample of subjects born between 1946 and 1965 (age range 41-63 years, n = 1,321). Spirometry was performed in line with the ATS/ ERS recommendations (13, 17) using a pneumotachograph-type spirometer (Masterscreen PC, CardinalHealth, Würzburg, Germany) before and after inhalation of 200 µg salbutamol. The spirometer was calibrated daily using a calibration pump (CardinalHealth, Würzburg, Germany), and additionally, an internal control (examiner) was used to ensure constant instrumental conditions. Under the guidance of the experienced examiners at least 3 and at most 8 trials were recorded to obtain at least 2 acceptable and reproducible flow-volume curves. After completion of each test, the curves were visually inspected, maneuvers with artifacts excluded and results evaluated according to ATS/ERS recommendations [3]. The present study is based on maximum values of FEV_1 measured before bronchodilation. On average 3.1 years later subjects were reexamined by spirometry (KORA F4L, n = 1,050, response rate 79.5%). Conditions for lung function measurements, including the examiners and the data evaluation, were the same in both cohorts. Both studies were approved by the Ethics Committee of the Bavarian Medical Association and informed

consent was obtained from the study participants. Genotypes were available for 890 of those individuals.

In the **Lothian Birth Cohorts (LBC)** spirometry was completed at two time points; at 70 and 73 years of age in LBC1936 and at 79 and 87 years of age in LBC1921. For both cohorts a Micro Medical Spirometer was used, assessments were conducted sitting down without noseclips. The accuracy of the spirometer is ± 3% (following ATS recommendations in standardization of spirometry 1994 update for flows and volumes).

Spirometry was performed at two time-points (baseline and year 5) in the **Prospective Investigation of the Vasculature in Uppsala Seniors (PIVUS)** (18). The study was performed in accordance with ATS recommendations (9) using the α spirometer; Vitalograph Ltd; Buckingham, UK. The best value from three recordings was used. The Ethics Committee of the University of Uppsala approved the study, and the participants gave their informed consent.

The **Rotterdam Study (RS)** is a prospective population-based cohort study founded in 1990 in a suburb of Rotterdam, the Netherlands. The first cohort (RS I) consists of 7,983 participants, aged 55 years and over. Performing of spirometry was introduced in 2004. Spirometry was performed by trained paramedical personnel using a SpiroPro® portable spirometer (Erich Jaeger, Hoechberg, Germany) and using a Carefusion MasterScreen PFT (as of 2009), according to ATS/ERS guidelines. Measurements were done at the visit of the Rotterdam Study's research facility. FEV_1, FVC and FEV_1/FVC ratio were measured. All spirometry measures were validated by two researchers, of which one is a specialist in respiratory medicine, by assessment of all applicable flow-volume and flow-time curves.

Spirometry in **SAPALDIA** was completed at two time points 11 years apart. Identical spirometry devices and protocols were used at both examinations for SAPALDIA (19, 20) (Sensormedics model 2200, Yorba Linda, USA). At least three forced expiratory lung function maneuvers were performed by each participant and a minimum of two acceptable forced expiratory flows, FVC and FEV_1 complying with ATS criteria were obtained (9). Expiratory flow measures with the highest sum of FVC and FEV_1 were taken from the same flow-volume curves to calculate the ratio of FEV_1/FVC. No bronchodilator was administered. Participants were requested not to use beta-2-agonists or anticholinergic inhalers four hours prior to and long-acting beta agonists, oral beta-2-agonists, theophyline or oral antimuscarinic medication eight hours prior to the time of appointment of the examination.

Spirometry in **SHIP** was conducted using a variable pressure bodyplethysmograph equipped with a pneumotachograph (VIASYS Healthcare, MasterScreen Body/Diff., JAEGER, Hoechberg, Germany) which met the ATS criteria (12). The procedures were conducted in a sitting position with subjects wearing a noseclip. The volume signal was calibrated with a 3.0 litre syringe connected to the pneumotachograph, in accordance with the manufacturer's recommendations, and at least once daily. Barometric pressure, temperature and relative humidity were registered every morning. Volume calibration referred to ATP-conditions (Ambient Temperature Pressure) but resulting lung volumes were expressed as BTPS-corrected (Body Temperature Pressure Saturated) (12, 21). The tests were performed in accordance with ATS and ERS recommendations (21, 22) in the following order: 1) determination of static lung volumes, 2) forced spirometry.

Statistical analysis

In each cohort study, a linear mixed effects model was constructed to model the longitudinal trajectory of FEV_1, which includes two components: the baseline FEV_1 levels and the rate of change in FEV_1 during follow-up. A continuous time (slope) variable was included in the model to quantify the time (in years) elapsed between each FEV_1 measurement point and the study baseline and the coefficient estimate for time conveys the rate of change in FEV_1 over time (mL/year). To allow for variation in baseline FEV_1 and the rate of change in FEV_1 across participants, the intercept and time variable were specified as both fixed and random effects. This model is commonly used for repeated measurements over time on the same participant, and takes the simple form:

$$FEV_{1ij} = \alpha + \beta * t_{ij} + a_i + b_i * t_{ij} + e_{ij},$$

where α and β are fixed population effects for intercept and time (slope), respectively; a_i and b_i are the individual intercept and slope for participant i; t_{ij} is the time value corresponding to the j^{th} FEV_1 measurement for participant i; and FEV_{1ij} is the FEV_1 value for participant i at the j^{th} measurement time; finally, e_{ij} represents the independent random error term and its values are assumed to be normally distributed (mean 0, variance σ^2). It is assumed that the random effects, a_i and b_i, are each independently, normally distributed with mean 0 and variance $\sigma_a{}^2$ and $\sigma_b{}^2$, respectively; a_i and b_i are not independent and their covariance is σ_{ab}. Therefore, the mixed effects model explicitly estimates the covariate effects, random effects, the variance for each random effect, and the covariance between each pair of random effects. In addition, the mixed effects model provides flexibility in handling unbalanced longitudinal data, which is common in longitudinal cohort studies, and allows participants with only one FEV_1 measurement to be included in the analysis as well as those with more than one measurement.

During the model development process, we performed exploratory analyses in the Health ABC study to demonstrate that it was sufficient to model the longitudinal change in FEV_1 using the linear effect of the time variable. This was also the only available option for the other cohort studies with only two repeated measurements per participant.

In each cohort study, a preliminary mixed effects model was constructed based on the above modeling framework and was additionally adjusted for the following covariates: gender, standing height at each time point, baseline age and baseline smoking pack-years, smoking status during study follow-up, the product term of smoking status x time, and study site and principal component variables for genetic ancestry as needed (family structure was accounted for where needed, for example in the Framingham Heart Study). Smoking status during study follow-up was defined as a four-level categorical variable as follows: never smokers at all time points (referent group), persistent smokers defined as current smokers at all time points, former smokers at all time points, and intermittent smokers defined as inconsistent smoker status across time points (switching between current and former status). The above main effects estimated how these covariates affected baseline FEV_1 and the product term of smoking status x time estimated the effects of the three smoking statuses on the rate of change in FEV_1 in comparison to never smokers.

Using the above preliminary model, each cohort study evaluated the presence and magnitude of covariance between the random intercept and time effects. When a meaningful covariance was present, the mixed effects model was specified to account for the covariance explicitly. Otherwise, the two random effects were specified as independent for model parsimony. Subsequently, residual diagnosis was performed in each cohort study based on the same preliminary model to exclude FEV_1 measurements detected as outliers (i.e.,

|standardized residual| > 3 or 4, as determined by each study). Effect estimates of important covariates were compiled from all cohort studies for quality control and summary of overall patterns. The above preliminary modeling analyses were performed using Proc Mixed in SAS or lme4 or pedigreemm in R in each cohort study. Effect estimates for selected covariates across all cohort studies are presented in Table 1 and Supplementary Table E2.

For the final GWAS model, a SNP main effect and SNP x time interaction term were included in the above preliminary model; these terms estimated the effects of SNP genotype on baseline FEV_1 and the rate of change in FEV_1, respectively. All cohort studies performed the final GWAS analysis using either lme4 or pedigreemm (in FHS to account for familial correlation) in R. Study-specific results for both the SNP main and SNP x time interaction effects were shared and meta-analyses were performed using METAL software (23) with the inverse variance weighting method to combine effect estimates after applying genomic control correction (24).

Rationale for the meta-analysis of five cohort studies

There is noticeable phenotypic heterogeneity among the 14 cohort studies in the meta-analysis, as reflected by several important aspects of the longitudinal study design (Table 1). Two important factors related to the quality of the outcome measurement are the number of repeated measurements of lung function and the follow-up duration. Five cohort studies, namely BHS, CARDIA, CHS, FHS, and Health ABC, had ≥3 repeated measurements and relatively long follow-up durations in comparison to other studies, allowing for more precise estimation of the longitudinal trajectory in pulmonary function. This was supported empirically by the preliminary mixed effects model results, and the observation that these five

cohort studies reported the smallest standard errors for the estimated rate of change in FEV_1 in the reference group of never-smokers (Table 2). In light of this, we performed a second meta-analysis based on these five cohort studies with the goal to reduce noise due to phenotypic heterogeneity and thus gain improved statistical association signals.

SNP selection for follow-up analyses

The most statistically significant SNPs were first identified using a significance threshold of $P \leq 1 \times 10^{-5}$ in either meta-analysis; all of the most statistically significant SNPs had a minor allele frequency above 5% (Tables E2 and E3). These SNPs were evaluated for regional linkage disequilibrium (LD), and multiple SNPs from a region were selected only when they met the p-value criterion and had an r^2 estimate less than 0.7 with each other, according to HapMap CEU as implemented in SNAP (25). Among SNPs in a given region with r^2 values of 0.7 or greater with each other, the SNP with the lowest p-value was selected for follow-up analysis.

Expression Profiling

The mRNA expression profiles of the implicated genes at the novel chromosome 11 and 15 loci were determined using reverse transcription polymerase chain reaction (RT-PCR). Human airway smooth muscle (HASM) cells were isolated from healthy bronchial tissue of patients (with no previous history of asthma) undergoing surgery, and cultured as described previously (26). Written consent was given by the patients and approval was granted by the Nottingham Local Ethical Research Committee (ref. EC00/165). Undifferentiated human bronchial epithelial cells (HBEC) (Lonza/Clonetics, Slough, UK) were maintained in culture

as described previously (27). Expression profiling by PCR used cDNA templates synthesised from total RNA, using the Superscript First-Strand Synthesis System for RT-PCR (Invitrogen/ Life Technologies Ltd., Paisley, UK), from cultured HASM and HBEC, as well as commercially available total lung and brain tissue (Ambion/Life Technologies Ltd.) and PBMC (3H Biomedical AB, Uppsala, Sweden). Primer sequences for the genes of interest are given in Table E7. PCR used the following cycling conditions: 35 cycles of 94°C for 1.5 minutes, 55°C for 1.5 minutes, and 72°C for 1.5 minutes, followed by 72°C for 10 minutes. Amplicons were extracted from agarose gels using the StrataPrep DNA Gel Extraction Kit (Agilent Technologies UK Ltd., Cheshire, UK) and validated by sequencing, using the BigDye Terminator v3.1 Cycle Sequencing Kit, in conjunction with an ABI PRISM 310 Genetic Analyser (Applied Biosystems/ Life Technologies Ltd.).

In addition, publicly available gene expression profiles of lung specimens from COPD patients (N=219) and controls (N=137) were obtained from the Lung Genomics Research Consortium (LGRC) site (http://www.lung-genomics.org/). Mean expression levels of the implicated genes at the novel chromosome 11 and 15 loci were compared between COPD patients and controls. The *P* value for the difference in means between the two groups was calculated using the two-sample t-test.

Finally, the most statistically significant SNPs at both loci (rs4077833 and rs507211, respectively) were tested for eQTL associations, using a publicly accessible eQTL database of lymphoblastoid cell lines (28). A statistical significance threshold of $P < 0.05$ was used in both analyses.

Supplementary Table 3.5 Details of SNP genotyping, quality control (QC), imputation, and statistical analysis across the 14 cohort studies

Study	Genotyping platform	QC filters for excluding genotyped SNPs	N, genotyped autosomal SNPs passing QC	Imputation software	NCBI Build for imputation reference (HapMap CEU)	N, SNPs used for analysis (MAF≥1%)	Statistical analysis software	Genomic control factor (λ_{gc})
ARIC	Affymetrix 6.0	call rate<95%, HWE P<10^{-6}, MAF<1%, or no chromosomal location	669,450	MACH v1.0.16	build 36, release 22	2,449,419	lme4, R	1.04
B58C*	Illumina 550K (2 deposits) + 610K	call rate<95%, HWE P<10^{-4}, MAF<1%, or inconsistent (P<10^{-4}) allele frequencies across 3 genotype deposits	519,040	MACH v1.0.16	build 35, release 21	2,460,629	lme4, R	1.01
BHS	Illumina 610-Quad	call rate<95%, HWE P<5.7x10^{-7}, or MAF<1%	549,294	MACH v1.0.16	build 36, release 22	2,420,960	lme4, R	1.05
CARDIA	Affymetrix 6.0	call rate<95%, HWE P<10^{-4}, or MAF<2%	578,568	BEAGLE	build 36, release 22	2,276,434	lme4, R	1.03
CHS	Illumina HumanHap 370CNV	call rate<97%, no heterozygotes, HWE P<10^{-5}, >2 duplicate errors, Mendelian inconsistency (for HapMap CEU trios), or	306,655	BIMBAM	build 36, release 22	2,190,045	lme4, R	1.00

Study	Genotyping platform	QC filters for excluding genotyped SNPs	N, genotyped autosomal SNPs passing QC	Imputation software	NCBI Build for imputation reference (HapMap CEU)	N, SNPs used for analysis (MAF≥1%)	Statistical analysis software	Genomic control factor (λ_{gc})
		no mapping in dbSNP						
FHS[†]	Affymetrix 500K + 50K Human Gene Focused Panel	call rate<97%, HWE $P<10^{-6}$, MAF<1%, differential missingness related to genotype (mishap procedure in PLINK) with $P<10^{-9}$, Mendelian errors>100, or absence from HapMap	378,163	MACH v1.0.15	build 36, release 22	2,411,786	pedigremm, R	1.09
Health ABC	Illumina Human1M-Duo	call rate < 95%, HWE $P<10^{-6}$, or MAF > 1%	914,263	MACH	build 36, release 22	2,470,255	lme4, R	1.04
KORA	Affymetrix 6.0	call rate<93%	909,622	IMPUTE 0.4.2	build 36, release 22	2,368,243	lme4, R	0.99
LBC1921	Ilumina 610-Quadv1	call rate<98%, HWE $P<10^{-3}$, or MAF<1%	542,050	MACH v1.0.16	build 36, release 22	2,302,855	lme4, R	0.97
LBC1932	Ilumina 610-Quadv1	call rate<98%, HWE $P<10^{-3}$, or MAF<1%	542,050	MACH v1.0.16	build 36, release 22	2,304,176	lme4, R	1.01
PIVUS	Illumina OmniExpress + Metabochip	monomorphic, HWE $P<10^{-6}$, call rate <95% if MAF>=5%, or call rate<99% if MAF<0.05	738,879	IMPUTE 2.0	build 36, release 22	2,436,058	lme4, R	1.00
RS	Illumina	call rate<98%,	537,405	MACH	build 36,	2,377,064	lme4, R	0.96

Study	Genotyping platform	QC filters for excluding genotyped SNPs	N, genotyped autosomal SNPs passing QC	Imputation software	NCBI Build for imputation reference (HapMap CEU)	N, SNPs used for analysis (MAF≥1%)	Statistical analysis software	Genomic control factor (λ_{gc})
	HumanHap 550K	HWE $P<10^{-6}$, or MAF<1%		v1.0.15	release 22			
SAPALDIA	Illumina Human 610K quad	call rate<97%, HWE $P<10^{-4}$, or MAF<5%	582,892	MACH v1.0.16	build 36, release 22	2,456,064	lme4, R	1.11
SHIP	Affymetrix 6.0	none	869,224	IMPUTE v0.5.0	build 36, release 22	2,450,720	lme4, R	0.99

Definition of abbreviations: ARIC = Atherosclerosis Risk in Communities; B58C = British 1958 Birth Cohort; BHS = Busselton Health Study; CARDIA = Coronary Artery Risk Development in Young Adults; CHS = Cardiovascular Health Study; FHS = Framingham Heart Study; Health ABC = Health, Aging, and Body Composition; HWE = Hardy Weinberg equilibrium; MAF = minor allele frequency; KORA = Cooperative Health Research in the Region of Augsburg; LBC1921 = Lothian Birth Cohort 1921; LBC1936 = Lothian Birth Cohort 1936; PIVUS = Prospective Investigation of the Vasculature in Uppsala Seniors; RS = Rotterdam Study; SAPALDIA = Swiss Study on Air Pollution and Lung Diseases in Adults; SHIP = Study of Health in Pomerania; SNP = single-nucleotide polymorphism.

* Three original subsets of B58C were combined for this analysis, following a new phase of genotyping with a common platform.

† To account for relatedness among subjects, the linear mixed effects regression model implemented in FHS used the pedigreemm package that adjusts for family structure (29).

Supplementary Table 3.6 Regression results for single nucleotide polymorphisms associated with the rate of change in FEV_1 (mL/year) at $P < 1 \times 10^{-5}$ in the meta-analysis of 14 cohort studies (N=27,249)

SNP	Chr	Position	Closest Gene(s)	Coded Allele	Noncoded Allele	Frequency	β	SE	*P* Value	Heterogeneity *P* Value
rs12137475	1	44059735	*ST3GAL3*	T	C	0.11	-3.52	0.76	3.90×10^{-6}	0.48
rs766488	1	61583103	*NFIA*	A	G	0.31	1.37	0.30	6.60×10^{-6}	0.84
rs17698444	1	215483178	*ESRRG /GPATCH2*	C	G	0.89	-2.21	0.47	2.62×10^{-6}	0.69
rs12692550	2	159958017	*BAZ2B*	T	C	0.17	-1.69	0.37	5.16×10^{-6}	0.08
rs2260732	13	113235802	*TMCO3*	A	G	0.28	1.42	0.31	4.83×10^{-6}	0.25
rs2260722	13	113236292		A	G	0.72	-1.51	0.32	1.83×10^{-6}	0.27
rs2479753	13	113240886		C	G	0.28	1.42	0.31	4.77×10^{-6}	0.22
rs2259541	13	113253338		C	G	0.72	-1.43	0.31	4.69×10^{-6}	0.20
rs3935740	15	79413780	*IL16/STARD5 /TMC3*	A	G	0.09	2.23	0.49	4.54×10^{-6}	1.00
rs4077833	15	79419738		C	G	0.10	2.31	0.46	5.71×10^{-7}	0.96
rs8027498	15	89595638	*SV2B*	A	G	0.25	1.43	0.32	9.41×10^{-6}	0.32
rs8051319	16	15794449	*MYH11*	T	C	0.60	1.46	0.32	5.12×10^{-6}	0.60
rs740557	17	62451139	*CACNG4*	C	G	0.85	-2.28	0.49	3.59×10^{-6}	0.80

Definition of abbreviations: Chr = chromosome; SE = standard error; SNP = single-nucleotide polymorphism.

Supplementary Table 3.7 Regression results for single nucleotide polymorphisms associated with the rate of change in FEV_1 (mL/year) at $P < 1 \times 10^{-5}$ in the meta-analysis of the five cohort studies with ≥3 FEV_1 measurements per participant (N=10,476)

SNP	Chr	Position	Closest Gene(s)	Coded Allele	Noncoded Allele	Frequency	β	SE	*P* Value	Heterogeneity *P* Value
rs10186544	2	28536678	*FOSL2/PLB1*	T	C	0.33	1.58	0.35	7.26×10^{-6}	0.72
rs10198727	2	28536753		A	T	0.67	-1.58	0.35	7.45×10^{-6}	0.71
rs10209416	2	28536819		A	G	0.33	1.58	0.35	7.23×10^{-6}	0.71
rs10209501	2	28536881		A	G	0.33	1.58	0.35	7.09×10^{-6}	0.71
rs12692550	2	159958017	*BAZ2B*	T	C	0.18	-2.01	0.42	2.02×10^{-6}	0.02
rs1729588	3	110790025	*FLJ25363 /MIR4445*	A	G	0.30	1.60	0.36	8.38×10^{-6}	0.33
rs10764052	10	19863473	*C10orf112*	T	C	0.46	1.50	0.33	6.68×10^{-6}	0.69
rs10764053	10	19863644		T	G	0.47	1.53	0.33	4.15×10^{-6}	0.76
rs12219073	10	19870510		C	G	0.54	-1.55	0.34	4.78×10^{-6}	0.72
rs17729837	10	19900824		A	C	0.52	-1.50	0.33	6.21×10^{-6}	0.69
rs10740924	10	19907692		C	G	0.53	1.47	0.32	4.55×10^{-6}	0.39
rs7095285	10	19909284		A	G	0.53	1.47	0.32	4.23×10^{-6}	0.39
rs1409737	10	19926336		A	G	0.53	1.45	0.32	6.52×10^{-6}	0.39
rs12770750	10	19928004		T	C	0.47	-1.45	0.32	6.38×10^{-6}	0.45
rs7898799	10	19929473		T	C	0.53	1.53	0.33	4.94×10^{-6}	0.34
rs10740927	10	19930342		T	G	0.47	-1.45	0.32	6.36×10^{-6}	0.45
rs7915851	10	19933596		A	G	0.53	1.46	0.32	6.10×10^{-6}	0.46
rs627684	11	86050787	*ME3*	A	G	0.31	1.78	0.35	3.70×10^{-7}	0.60
rs601988	11	86050799		A	G	0.31	1.77	0.35	4.46×10^{-7}	0.56
rs507123	11	86054358		A	G	0.33	1.81	0.35	2.30×10^{-7}	0.45
rs507211	11	86054387		A	G	0.25	2.09	0.37	2.18×10^{-8}	0.65
rs626049	11	86054989		T	C	0.31	1.74	0.35	7.20×10^{-7}	0.50
rs594361	11	86064432		A	G	0.69	-1.78	0.35	3.85×10^{-7}	0.56

rs642245	11	86067184		A	G	0.19	2.19	0.42	1.43×10^{-7}	0.40

Definition of abbreviations: Chr = chromosome; SE = standard error; SNP = single-nucleotide polymorphism.

Supplementary Table 3.8 Association of the 14 sentinel SNPs from the meta-analyses in the AGES-Reykjavík study (AGES) and the Lung Health Study (LHS) for the rate of change in FEV_1 (mL/year)

SNP	Chr	Closest Gene(s)	Coded Allele	Freq	Meta-analysis*			AGES (n = 1,494)			LHS (n = 4,048)		
					β	SE	*P* Value	β	SE	*P* Value	β	SE	*P* Value
Meta-analysis of 14 cohort studies (n = 27,249)													
rs12137475	1	*ST3GAL3*	T	0.11	-3.5	0.8	3.90×10^{-6}	1.2	1.5	0.42	2.6	2.5	0.30
rs766488	1	*NFIA*	A	0.31	1.4	0.3	6.60×10^{-6}	0.1	0.7	0.94	-2.0	1.3	0.12
rs17698444	1	*ESRRG /GPATCH2*	C	0.89	-2.2	0.5	2.62×10^{-6}	-0.4	1.0	0.72	-0.2	1.9	0.93
rs12692550	2	*BAZ2B*	T	0.17	-1.7	0.4	5.16×10^{-6}	0.4	0.9	0.66	-0.6	1.6	0.72
rs2259541	13	*TMCO3*	C	0.72	-1.5	0.3	1.83×10^{-6}	0.8	0.7	0.29	0.0	1.3	0.99
rs4077833	15	*IL16/STARD5 /TMC3*	C	0.10	2.3	0.5	5.71×10^{-7}	-0.3	1.1	0.79	-0.5	2.0	0.80
rs8027498	15	*SV2B*	A	0.25	1.4	0.3	9.41×10^{-6}	0.1	0.8	0.87	-1.7	1.4	0.22
rs8051319	16	*MYH11*	T	0.60	1.7	0.3	5.12×10^{-6}	-0.6	0.7	0.35	0.4	1.3	0.76
rs740557	17	*CACNG4*	C	0.85	-2.3	0.5	3.59×10^{-6}	-2.2	1.2	0.08	2.7	2.5	0.27
Meta-analysis of the five cohort studies with ≥3 more FEV_1 measurements per participant (n = 10,476)													
rs10209501	2	*FOSL2/PLB1*	A	0.33	1.6	0.4	7.09×10^{-6}	-0.3	0.7	0.62	-1.3	1.3	0.31
rs12692550	2	*BAZ2B*	T	0.18	-2.0	0.4	2.02×10^{-6}	0.4	0.9	0.66	-0.6	1.6	0.72
rs1729588	3	*FLJ25363 /MIR4445*	A	0.30	1.6	0.4	8.38×10^{-6}	0.9	0.7	0.20	0.9	1.3	0.48
rs10764053	10	*C10orf112*	A	0.47	1.5	0.3	4.15×10^{-6}	0.5	0.6	0.40	-0.6	1.2	0.62
rs507211	11	*ME3*	A	0.25	2.1	0.4	2.18×10^{-8}	1.0	0.7	0.15	0.3	1.4	0.85

Definition of abbreviations: Chr = chromosome; Freq = frequency; SE = standard error; SNP = single-nucleotide polymorphism.

* Association results of the 14 sentinel SNPs with the rate of change in FEV_1 (mL/year) in the corresponding discovery meta-analysis are shown for comparison.

Supplementary Table 3.9 Association of previously reported loci in GWAS of cross-sectional lung function (1, 30, 31) with the rate of change in FEV_1 (mL/year) in the meta-analysis of 14 cohort studies (N=27,249)

SNP	Chr	Coded Allele	Frequency	β	SE	*P* Value*	Gene	Reference
Loci associated with cross-sectional FEV_1								
rs2571445	2	A	0.39	-0.18	0.29	0.54	*TNS1*	30
rs1344555	3	T	0.20	0.24	0.36	0.52	*MECOM*	1
rs17035960	4	T	0.07	0.05	0.57	0.93	*GSTCD*	31
rs17036052	4	T	0.06	0.32	0.69	0.64	*GSTCD*	31
rs17036090	4	T	0.93	0.09	0.57	0.88	*GSTCD*	31
rs11727189	4	T	0.07	-0.09	0.58	0.88	*GSTCD*	31
rs10516526	4	A	0.93	0.07	0.55	0.90	*GSTCD*	30
rs11097901	4	T	0.07	-0.03	0.56	0.95	*GSTCD*	31
rs11728716	4	A	0.07	0.05	0.55	0.92	*GSTCD*	31
rs17036341	4	C	0.93	-0.01	0.55	0.99	*GSTCD*	31
rs17331332	4	A	0.08	0.07	0.55	0.90	*GSTCD*	31
rs3995090	5	A	0.61	0.36	0.29	0.21	*HTR4*	30
rs6889822	5	A	0.62	0.32	0.29	0.27	*HTR4*	30
rs6903823	6	A	0.77	0.32	0.34	0.35	*ZKSCAN3*	1
rs7068966	10	T	0.53	-0.35	0.28	0.22	*CDC123*	1
rs11001819	10	A	0.48	-0.32	0.29	0.26	*C10orf11*	1
Loci associated with cross-sectional FEV_1/FVC								
rs2284746	1	C	0.48	-0.46	0.29	0.11	*MFAP2*	1
rs993925	1	T	0.34	0.16	0.31	0.60	*TGFB2*	1
rs10498230	2	T	0.07	1.59	0.57	**0.005**	*PID1*	31
rs1435867	2	T	0.93	-1.52	0.56	**0.006**	*PID1*	31
rs12477314	2	T	0.21	0.04	0.36	0.90	*HDAC4*	1
rs1529672	3	A	0.18	0.64	0.40	0.11	*RARB*	1
rs6830970	4	A	0.65	0.16	0.30	0.60	*FAM13A*	31
rs2869967	4	T	0.61	0.10	0.29	0.72	*FAM13A*	31
rs1032295	4	T	0.59	-0.55	0.30	0.07	*HHIP*	30
rs12504628	4	T	0.60	-0.48	0.29	0.10	*HHIP*	31
rs1980057	4	T	0.40	0.57	0.29	**0.048**	*HHIP*	31
rs153916	5	T	0.53	-0.10	0.28	0.71	*SPATA9*	1
rs11168048	5	T	0.58	0.26	0.29	0.38	*HTR4*	31
rs7735184	5	T	0.40	-0.30	0.29	0.29	*HTR4*	31
rs2277027	5	A	0.65	-0.37	0.29	0.20	*ADAM19*	31
rs1422795	5	T	0.65	-0.34	0.29	0.25	*ADAM19*	31
rs2857595	6	A	0.19	0.37	0.38	0.33	*NCR3*	1

SNP	Chr	Coded Allele	Frequency	β	SE	*P* Value*	Gene	Reference
rs2070600	6	T	0.05	0.64	0.68	0.34	*AGER*	30, 31
rs2798641	6	T	0.19	0.13	0.37	0.72	*ARMC2*	1
rs11155242	6	A	0.79	-0.56	0.35	0.11	*GPR126*	31
rs6937121	6	T	0.70	-0.79	0.31	**0.010**	*GPR126*	31
rs3817928	6	A	0.79	-0.59	0.35	0.09	*GPR126*	31
rs7776375	6	A	0.71	-0.78	0.32	**0.015**	*GPR126*	31
rs16909898	9	A	0.91	-0.82	0.52	0.11	*PTCH1*	31
rs10512249	9	A	0.09	0.73	0.50	0.14	*PTCH1*	31
rs7068966	10	T	0.53	-0.35	0.28	0.22	*CDC123*	1
rs11172113	12	T	0.62	-0.18	0.30	0.55	*LRP1*	1
rs1036429	12	T	0.23	-0.14	0.34	0.68	*CCDC38*	1
rs12899618	15	A	0.15	0.61	0.40	0.13	*THSD4*	30
rs12447804	16	T	0.20	0.02	0.36	0.95	*MMP15*	1
rs2865531	16	A	0.59	-0.58	0.29	**0.042**	*CFDP1*	1
rs9978142	21	A	0.85	-0.42	0.41	0.31	*KCNE2*	1

Definition of abbreviations: Chr = chromosome; SE = standard error; SNP = single-nucleotide polymorphism.

* *P* values below the statistically significant threshold of 0.05 are shown in bold.

Supplementary Table 3.10 mRNA expression profiling of the implicated genes at the two novel loci in human lung and control tissues*

Gene	Human Tissues/Cells				
	Lung	Human Airway Smooth Muscle Cells	Human Bronchial Epithelial Cells	Peripheral Blood Mononuclear Cells	Brain
IL16	+	+	-	+	+
STARD5	+	+	+	+	+
TMC3	-	-	-	-	+
ME3	+	+	+	-	+

* Primer sequences are provided in Table E7. A "+" sign indicates the presence of the transcript, and a "-" sign indicates its absence. All products were sequence verified.

Supplementary Table 3.11 Primers for mRNA expression profiling

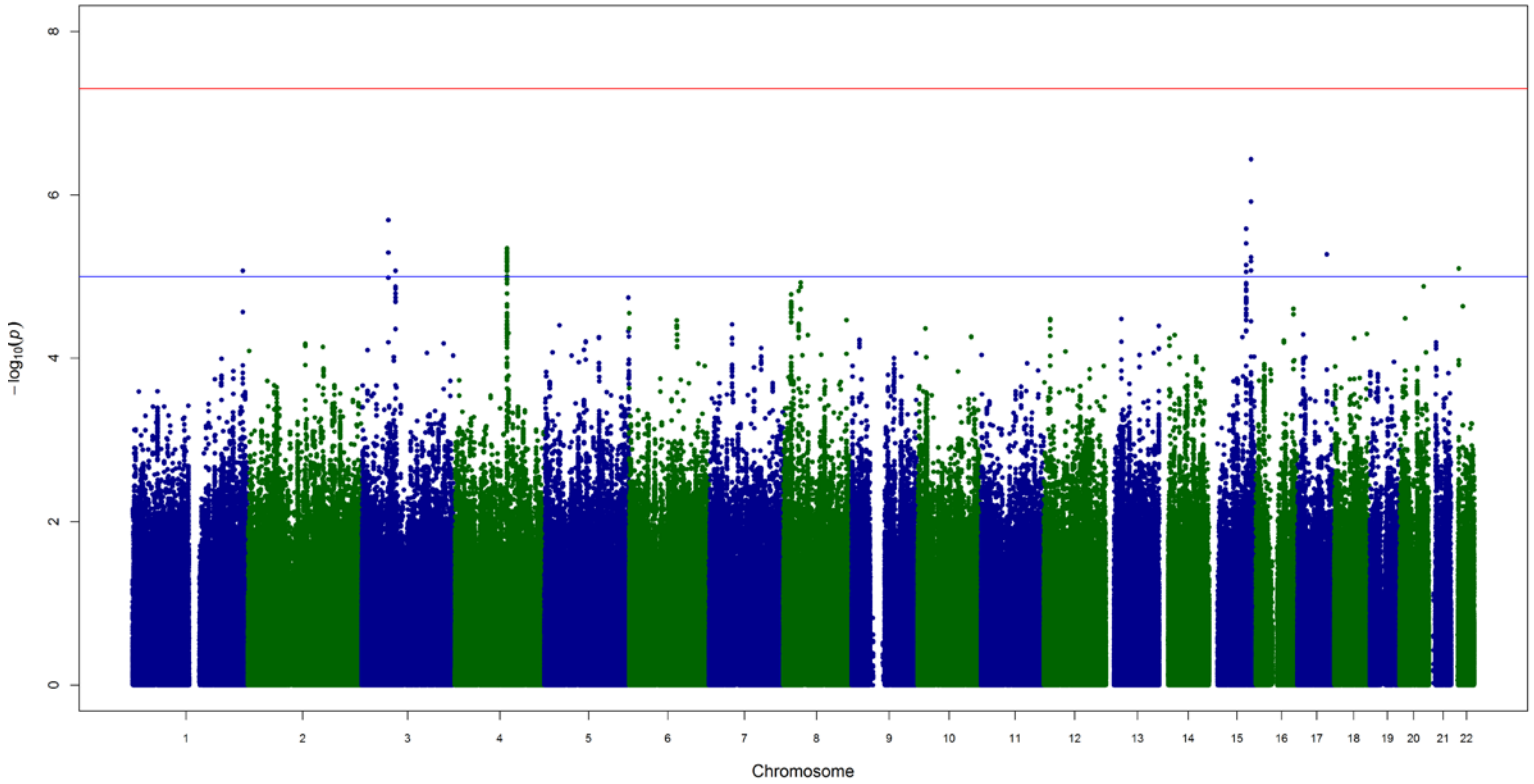
Gene	Sequence
IL16	Forward primer 5'-CTCGCTCAACCTTTCAGAGC-3'
	Reverse primer 5'-TCTGTGAACCGTAATCACCTTG-3'
STARD5	Forward primer 5'-AGGGAACCTGTACCGAGGAG-3'
	Reverse primer 5'-GGTGGGTTCCCCTGGAAG-3'
TMC3	Forward primer 5'-CATTCCAGAGCTGATTGCAG-3'
	Reverse primer 5'-GGTAGCCATTTCCTCAATGC-3'
ME3	Forward primer 5'-GACCTGGACAAGTACATCATTCTC-3'
	Reverse primer 5'-TGGCAGCAACACCTATGATG-3'

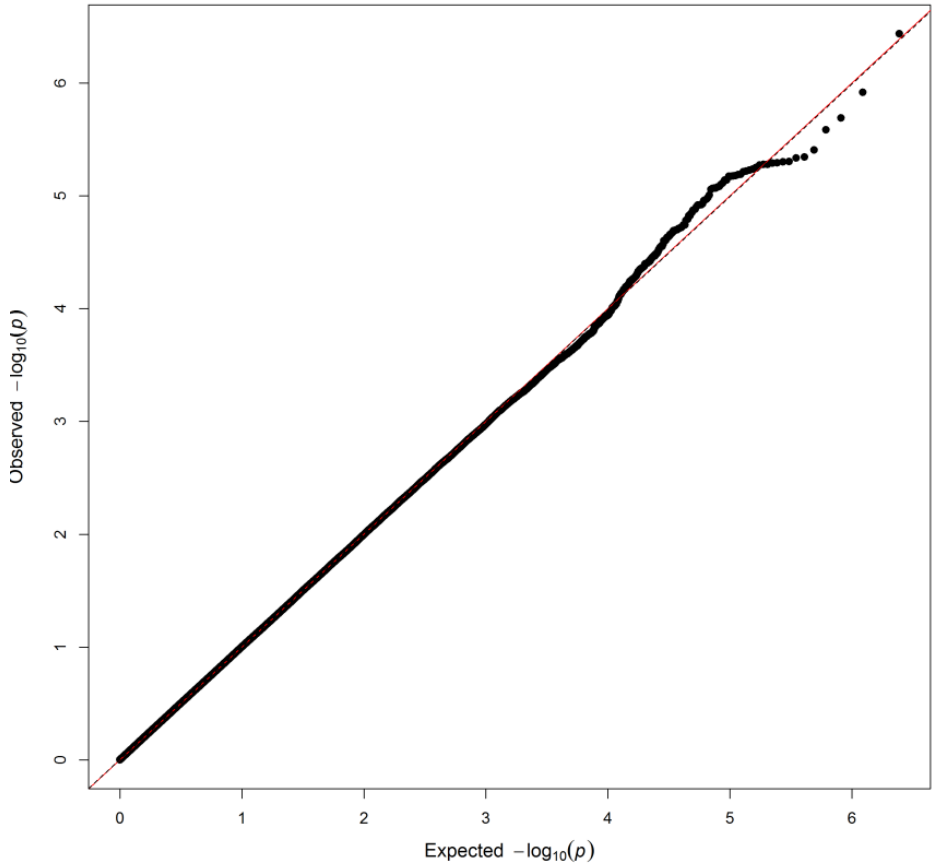
Supplementary Table 3.12 Summary of eQTL look-up for the most significant SNPs at the novel chromosome 11 and 15 loci

SNP (location)	eQTL Type	Associated Gene	Chr	*P* Value	Gene Function
rs4077833 (*IL16/STARD5/ TMC3* locus on chr 15)	Trans	*NR1I2*	3	6.84×10^{-4}	Nuclear receptor subfamily 1, group I, member 2 (*NR1I2*) encodes a pleiotropic nuclear transcription factor with a key role in the regulation of *CYP3A4*, a cytochrome P450 enzyme that metabolizes more than 50% of human clinical drugs (32). *NR1I2* is activated by a range of endogenous and xenobiotic compounds and binds to response elements in the promoter regions of many target genes in complex with the retinoic acid receptor RXR.
rs507211 (*ME3* locus on chr 11)	Trans	*KIAA1109*	4	5.20×10^{-4}	*KIAA1109* is part of a linkage disequilibrium block (*KIAA1109-TENR-IL2-IL21* gene cluster) associated with susceptibility to celiac disease (33); this region encodes the interleukins IL2 and IL21. In mice, IL21 levels are higher in blood and lung tissues of animals exposed to cigarette smoke vs. air-exposed (34). Murine *in vitro* work shows that IL21 promotes T-cell induced apoptosis and cell damage, suggesting a possible link between gene expression in this region and lung outcomes (34). Additionally, in mammals KIAA1109 is thought to function in the regulation of epithelial growth and differentiation, and in tumor development (35).

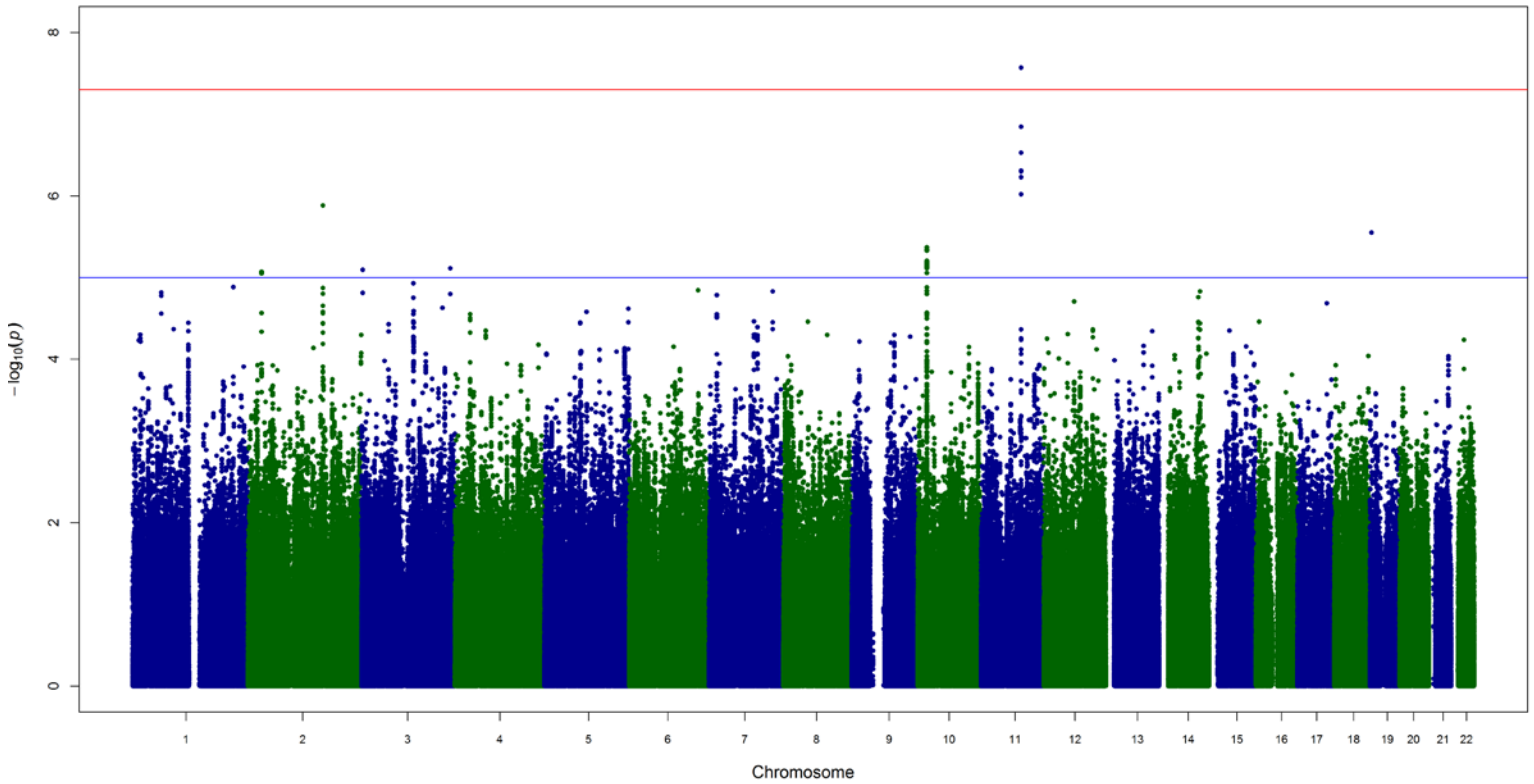
Definition of abbreviations: Chr = chromosome; eQTL = expression quantitative trait loci; SNP = single-nucleotide polymorphism.

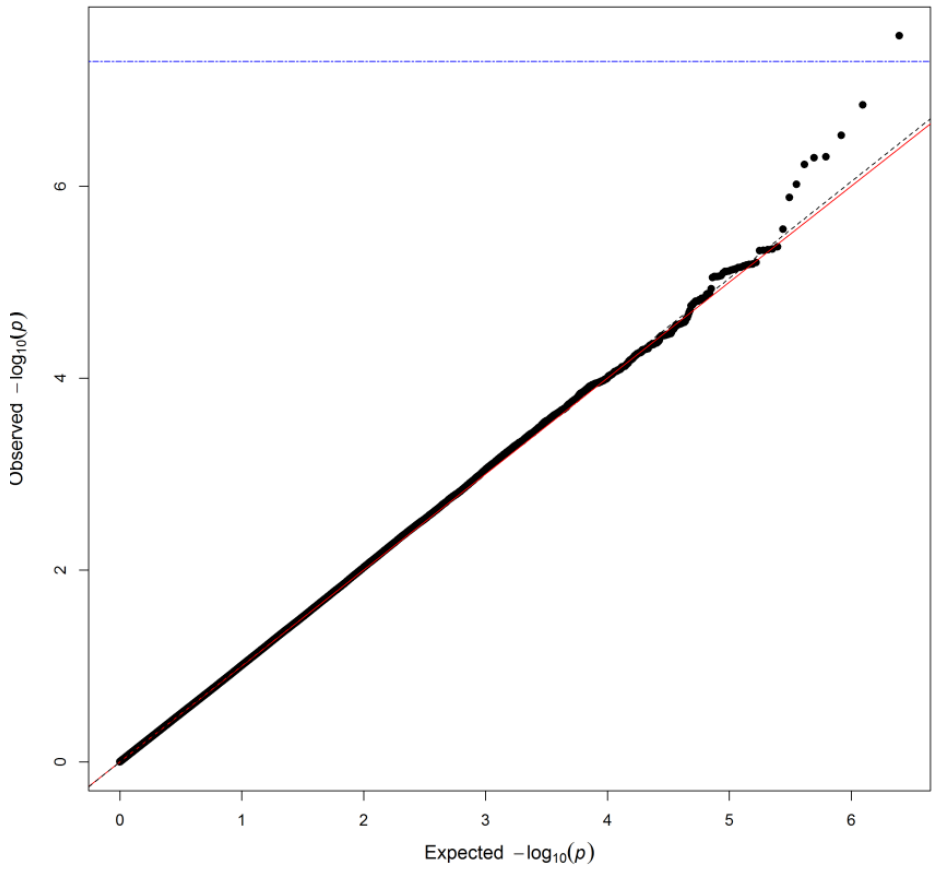
Supplementary Figure 3.3 Manhattan and QQ plots for the meta-analysis of the rate of change in FEV_1 in 14 cohort studies




Observed $-\log_{10}(p)$
Expected $-\log_{10}(p)$
0
1
2
3
4
5
6

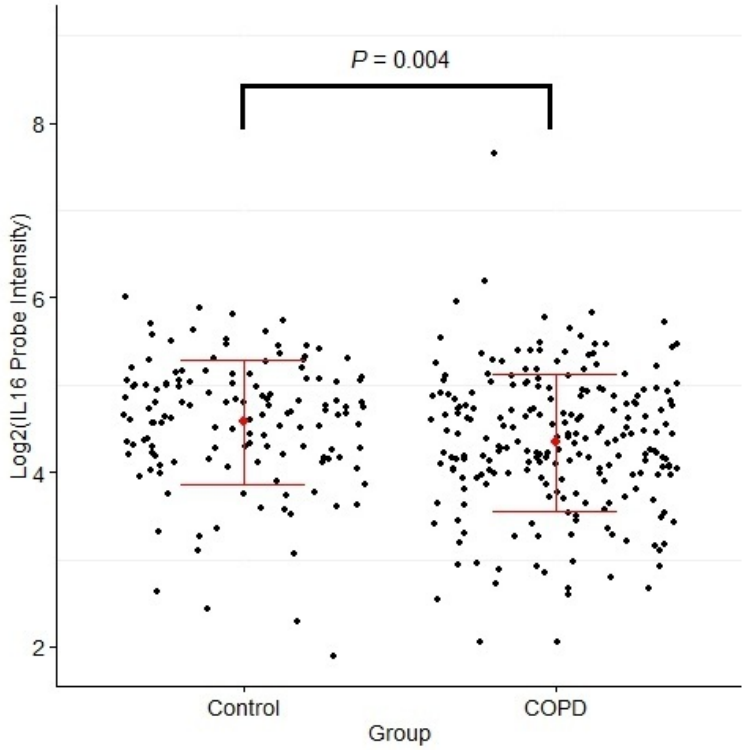
Supplementary Figure 3.4 Manhattan and QQ plots for the meta-analysis of the rate of change in FEV_1 in the five cohort studies with ≥3 more FEV_1 measurements per participant



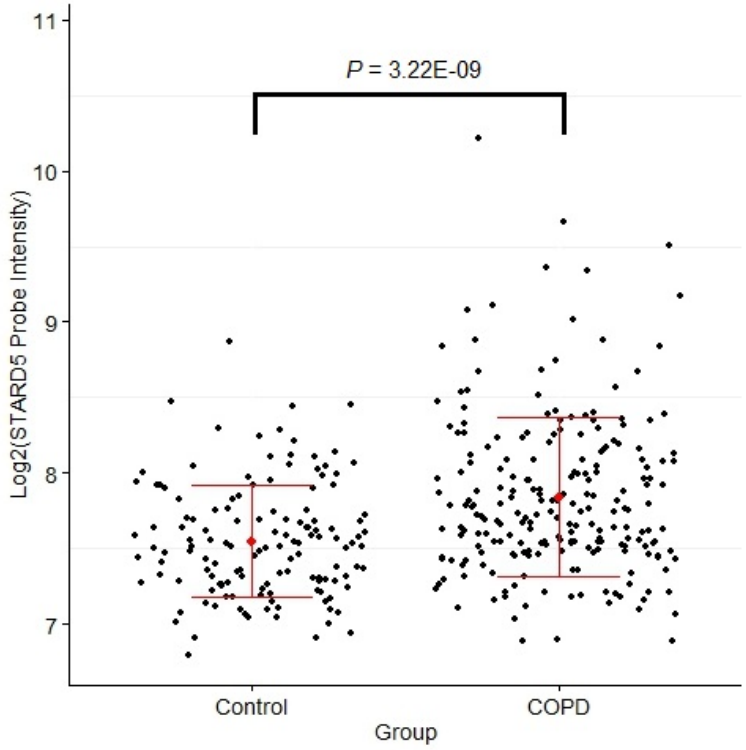

Observed $-\log_{10}(p)$
Expected $-\log_{10}(p)$
0
2
4
6
0
1
2
3
4
5
6

Supplementary Figure 3.5 mRNA expression profiling in human lung samples from 219 COPD patients and 137 controls for A) *IL16*, B) *STARD5*, and C) *ME3*, using publicly available microarray data from the Lung Genomics Research Consortium site (http://www.lung-genomics.org/). The y-axes reflect the probe intensities of each gene transcript in the binary logarithm form, with the red dots indicating the average probe intensities and the red bars indicating standard deviation. The *P* values were calculated using the two-sample t-test.

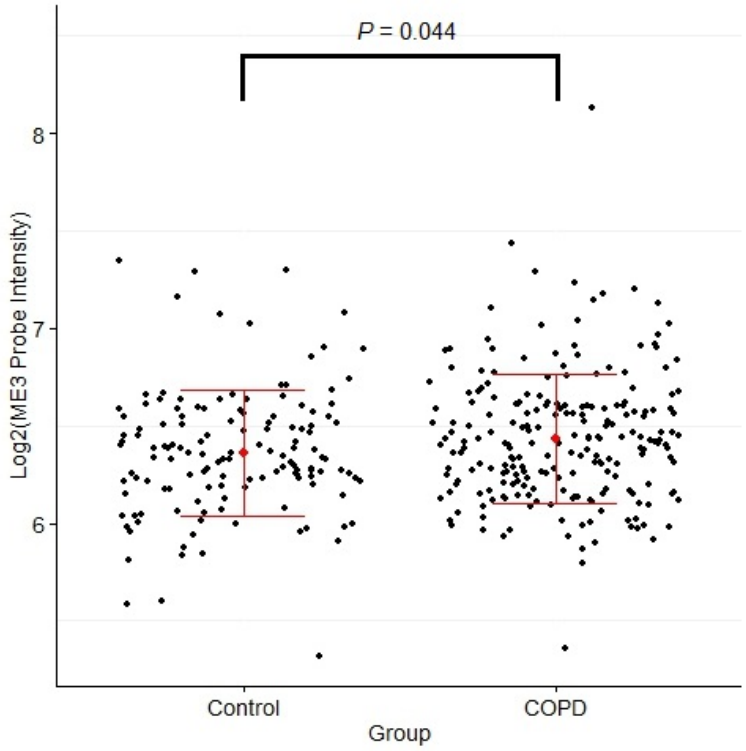
A. *IL16*



B. *STARD5*


P = 3.22E-09
11
10
9
8
7
Log2(STARD5 Probe Intensity)
Control
COPD
Group

C. *ME3*



REFERENCES

1. Soler Artigas M, Loth DW, Wain LV, et al. Genome-wide association and large-scale follow up identifies 16 new loci influencing lung function. *Nat Genet*. 2011;43(11):1082-90.

2. The Atherosclerosis Risk in Communities (ARIC) Study: design and objectives. The ARIC investigators. *Am J Epidemiol*. 1989;129(4):687-702.

3. Strachan DP, Griffiths JM, Johnston ID, et al. Ventilatory function in British adults after asthma or wheezing illness at ages 0-35. *Am J Respir Crit Care Med*. 1996;154(6 Pt 1):1629-35.

4. Strachan DP, Rudnicka AR, Power C, et al. Lifecourse influences on health among British adults: effects of region of residence in childhood and adulthood. *Int J Epidemiol*. 2007;36(3):522-31.

5. Marossy AE, Strachan DP, Rudnicka AR, et al. 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(4):355-9.

6. James AL, Palmer LJ, Kicic E, et al. Decline in lung function in the Busselton Health Study: the effects of asthma and cigarette smoking. *Am J Respir Crit Care Med*. 2005;171(2):109-14.

7. Ryan G, Knuiman MW, Divitini ML, et al. Decline in lung function and mortality: the Busselton Health Study. *J Epidemiol Community Health*. 1999;53(4):230-4.

8. ATS statement--Snowbird workshop on standardization of spirometry. *Am Rev Respir Dis*. 1979;119(5):831-8.

9. Standardization of Spirometry, 1994 Update. American Thoracic Society. *Am J Respir Crit Care Med*. 1995;152(3):1107-36.

10. Enright PL, Kronmal RA, Higgins M, et al. Spirometry reference values for women and men 65 to 85 years of age. Cardiovascular health study. *Am Rev Respir Dis*. 1993;147(1):125-33.

11. Enright PL, Kronmal RA, Higgins MW, et al. Prevalence and correlates of respiratory symptoms and disease in the elderly. Cardiovascular Health Study. *Chest*. 1994;106(3):827-34.

12. Standardization of spirometry--1987 update. Statement of the American Thoracic Society. *Am Rev Respir Dis*. 1987;136(5):1285-98.

13. Miller MR, Hankinson J, Brusasco V, et al. Standardisation of spirometry. *Eur Respir J*. 2005;26(2):319-38.

14. Waterer GW, Wan JY, Kritchevsky SB, et al. Airflow limitation is underrecognized in well-functioning older people. *J Am Geriatr Soc*. 2001;49(8):1032-8.

15. Holle R, Happich M, Lowel H, et al. KORA--a research platform for population based health research. *Gesundheitswesen*. 2005;67 Suppl 1:S19-25.

16. Wichmann HE, Gieger C, Illig T. KORA-gen--resource for population genetics, controls and a broad spectrum of disease phenotypes. *Gesundheitswesen*. 2005;67 Suppl 1:S26-30.

17. Karrasch S, Flexeder C, Behr J, et al. Spirometric reference values for advanced age from a South german population. *Respiration; international review of thoracic diseases*. 2013;85(3):210-9.

18. Lind L, Fors N, Hall J, et al. A comparison of three different methods to evaluate endothelium-dependent vasodilation in the elderly: the Prospective Investigation of the Vasculature in Uppsala Seniors (PIVUS) study. *Arteriosclerosis, thrombosis, and vascular biology*. 2005;25(11):2368-75.

19. Kunzli N, Ackermann-Liebrich U, Keller R, et al. Variability of FVC and FEV1 due to technician, team, device and subject in an eight centre study: three quality control studies in SAPALDIA. Swiss Study on Air Pollution and Lung Disease in Adults. *Eur Respir J*. 1995;8(3):371-6.

20. Kunzli N, Kuna-Dibbert B, Keidel D, et al. Longitudinal validity of spirometers--a challenge in longitudinal studies. *Swiss medical weekly*. 2005;135(33-34):503-8.

21. Nelson SB, Gardner RM, Crapo RO, et al. Performance evaluation of contemporary spirometers. *Chest*. 1990;97(2):288-97.

22. Wanger J, Clausen JL, Coates A, et al. Standardisation of the measurement of lung volumes. *Eur Respir J*. 2005;26(3):511-22.

23. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics*. 2010;26(17):2190-1.

24. Devlin B, Roeder K. Genomic control for association studies. *Biometrics*. 1999;55(4):997-1004.

25. Johnson AD, Handsaker RE, Pulit SL, et al. SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap. *Bioinformatics*. 2008;24(24):2938-9.

26. Daykin K, Widdop S, Hall IP. Control of histamine induced inositol phospholipid hydrolysis in cultured human tracheal smooth muscle cells. *Eur J Pharmacol*. 1993;246(2):135-40.

27. Danahay H, Atherton H, Jones G, et al. Interleukin-13 induces a hypersecretory ion transport phenotype in human bronchial epithelial cells. *Am J Physiol Lung Cell Mol Physiol*. 2002;282(2):L226-36.

28. Dixon AL, Liang L, Moffatt MF, et al. A genome-wide association study of global gene expression. *Nat Genet*. 2007;39(10):1202-7.

29. Vazquez AI, Bates DM, Rosa GJ, et al. Technical note: an R package for fitting generalized linear mixed models in animal breeding. *Journal of animal science*. 2010;88(2):497-504.

30. Repapi E, Sayers I, Wain LV, et al. Genome-wide association study identifies five loci associated with lung function. *Nat Genet*. 2009;42(1):36-44.

31. Hancock DB, Eijgelsheim M, Wilk JB, et al. Meta-analyses of genome-wide association studies identify multiple loci associated with pulmonary function. *Nat Genet*. 2009;42(1):45-52.

32. Orans J, Teotico DG, Redinbo MR. The nuclear xenobiotic receptor pregnane X receptor: recent insights and new challenges. *Mol Endocrinol*. 2005;19(12):2891-900.

33. van Heel DA, Franke L, Hunt KA, et al. A genome-wide association study for celiac disease identifies risk variants in the region harboring IL2 and IL21. *Nat Genet*. 2007;39(7):827-9.

34. Duan MC, Huang Y, Zhong XN, et al. Th17 cell enhances CD8 T-cell cytotoxicity via IL-21 production in emphysema mice. *Mediators of inflammation*. 2012;2012:898053.

35. Kuo MT, Wei Y, Yang X, et al. Association of fragile site-associated (FSA) gene expression with epithelial differentiation and tumor development. *Biochem Biophys Res Commun*. 2006;340(3):887-93.

CHAPTER 4

DIETARY PATTERNS, CIGARETTE SMOKING, AND LONGITUDINAL CHANGE IN LUNG FUNCTION

Wenbo Tang[1], Kristin A. Guertin[1], Kathryn B. Arnold[2], JoAnn A. Hartline[2], Lori M. Minasian[3], Scott M. Lippman[4], Eric Klein[5] and Patricia A. Cassano[1]

[1]Division of Nutritional Sciences, Cornell University, Ithaca, NY

[2]Southwest Oncology Group Statistical Center, Seattle, WA

[3]Division of Cancer Prevention, National Cancer Institute, Bethesda, MD

[4] Divisions of Cancer Medicine and Cancer Prevention and Population Sciences, University of Texas M.D. Anderson Cancer Center, Houston, TX

[5]Cleveland Clinic, Cleveland, OH

Corresponding Author: Patricia A Cassano, Division of Nutritional Sciences, Cornell University, 209 Savage Hall, Ithaca, NY 14853, pac6@cornell.edu, (607) 255-7551 (phone), (607) 255-2691 (fax).

Funding: This study was supported by grant R01HL071022 from the National Heart, Lung, and Blood Institute (to PAC), and, in part, by grant CA37429 from the National Cancer Institute/Division of Cancer Prevention, and by the National Center for Complementary and Alternative Medicine.

ABSTRACT

Background: Most previous studies of diet and lung function focus on individual nutrients or foods. Dietary pattern analysis is a strong alternative approach, but little is known about the relation between dietary patterns and longitudinal change in lung function.

Objective: We assessed the relation between dietary patterns and the rate of change in forced expiratory volume in 1 s (FEV_1) in a longitudinal cohort of North American men.

Design: Data were collected from 2,560 male adults in the Respiratory Ancillary Study. Dietary patterns were derived from food frequency questionnaire data using principal component analysis, and linear mixed-effects models were used to examine these patterns with the rate of change in FEV_1 over ~3 years, adjusting for potential confounders. Given the importance of cigarette smoking, interaction between dietary patterns and smoking was examined.

Results: Two dietary patterns were identified: a prudent pattern (high intakes of fruits, vegetables, fish, and poultry) and a Western pattern (high intakes of red, fried, and processed meats, desserts, and fries). No overall association was observed for either dietary pattern with the rate of change in FEV_1. When the interaction with smoking was considered, a more prudent diet was significantly associated with a slower rate of decline in FEV_1 in smokers (highest vs. lowest tertile: 38.8 mL per year; 95% CI: 14.4, 63.1), but not in non-smokers (highest vs. lowest tertile: -5.2 mL per year; 95% CI: -15.1, 4.7; $P_{interaction} = 0.001$).

Conclusions: In a population of North American men, a diet rich in fruits, vegetables, fish, and poultry was associated with a slower decline in lung function in smokers, whereas a diet rich in red, fried, and processed meats, desserts, and fries was not associated with change in lung function.

KEY WORDS Dietary pattern, cigarette smoking, lung function, longitudinal change, interaction, principal component analysis, chronic obstructive pulmonary disease, COPD

INTRODUCTION

Lung function is an important predictor of morbidity and mortality in the general population (1). Forced expiratory volume in 1 s (FEV_1), an easily measured spirometric parameter, is a reliable indicator of the physiological state of the lungs and airways and provides the basis for diagnosing and staging chronic obstructive pulmonary disease (COPD), a leading cause of death globally (2). Decline in lung function occurs naturally with aging, but it is sometimes accelerated by exposure to cigarette smoke, leading to reduced lung function that characterizes COPD (3, 4). Therefore, longitudinal change in lung function is an informative predictor of COPD risk, and studies of the rate of change in FEV_1 provide important insights for understanding disease pathogenesis (5, 6).

While smoking is widely recognized as the key risk factor for accelerated loss in lung function, as much as 90% of the overall variation in FEV_1 decline remained unexplained after accounting for the effects of age, height and smoking (7-10). This observation strongly suggests the importance of other modifiable risk factors, such as diet, in contributing to changes in lung function and COPD risk. In support of this hypothesis, epidemiologic studies have confirmed associations between the intake of numerous individual nutrients and foods and lung function (11-13). However, most studies do not account for the complexity in dietary intake either by directly considering the correlations among nutrients, or by considering combinations of foods and/or nutrients that are likely to be interactive or synergistic (14).

An alternative to studying single nutrients is a dietary pattern approach, which characterizes the overall diet by identifying patterns of intake (15). Several studies investigated the association of dietary patterns with lung function (16, 17) and COPD (18, 19). A diet rich in fruits, vegetables, fish, and whole-grain products was associated with higher

cross-sectional lung function levels and lower COPD risk, whereas a diet rich in red meats, refined grains, desserts, and fries was associated with lower cross-sectional lung function levels and higher COPD risk. These studies are limited by their cross-sectional design, thus longitudinal studies are needed to support stronger causal inferences.

The purpose of the current study was to determine major dietary patterns using principal component analysis, and then examine the association of the identified dietary patterns with the rate of change in FEV_1 over three years in a large, prospective population-based cohort study of North American men aged 50 and over. Our specific hypothesis was that dietary patterns would modify the effect of cigarette smoking on the rate of change in lung function, given postulated effects of dietary constituents to mitigate oxidative stress and inflammatory burden.

SUBJECTS AND METHODS

Study Population

The Respiratory Ancillary Study (RAS) is a post-randomization ancillary study to the Selenium and Vitamin E Cancer Prevention Trial (SELECT) and comprises 2,920 male adults from 16 study sites that were enrolled between July 2004 and April 2007. Details of SELECT have been previously reported (20). Eligibility criteria for SELECT included age (≥55 years; ≥50 years for African Americans), and participants in RAS comprise a subgroup of men in SELECT with selective oversampling of cigarette smokers. Local and Cornell institutional review boards approved the RAS and SELECT protocols, and written consent was obtained from all participants.

Assessment of Dietary Patterns

Dietary intake information was collected by a 120-item food frequency questionnaire (FFQ) administered at study baseline to assess average food intake during the preceding 12 months. For each food item, participants reported their average frequency of consumption during the past year in terms of a specified serving size; nine frequency categories ranging from “never or less than once per month” to “2 or more times per day” and three portion sizes were provided. These 120 food items were subsequently consolidated on the basis of similar nutrient composition and culinary use into 54 distinct food groups as inputs to the factor analysis, similar to previous studies of dietary patterns and pulmonary outcomes (16-19).

In the principal component analysis, an orthogonal transformation (Varimax rotation; SAS Institute, Cary, NC) was used to obtain uncorrelated factors, which have simpler structure and greater interpretability. The number of factors to retain was determined by evaluating the diagram of eigenvalues, the Scree plot, and the interpretability of the factors. For each retained factor (i.e., dietary pattern), an individual-level score was constructed by summing the observed intake of the component food groups weighted by their respective factor loadings; the score quantifies the degree to which an individual conformed to that dietary pattern.

Assessment of Lung Function

Lung function was measured by spirometry at three of four annual visits over a three-year follow-up period. Spirometry testing followed standardized guidelines of the American Thoracic Society (ATS) and used the EasyOne handheld, flow-sensing spirometer, which has excellent validity, reliability, and significantly simpler field implementation in comparison to

traditional desktop devices (21, 22). All spirometry tests were performed by clinical research nurses that were centrally trained in spirometry techniques, including bi-annual refresher sessions; all test results were centrally reviewed for quality with weekly feedback to each site including as-needed requests for retests in order to continually improve test quality. All FEV_1 measures that met the ATS criteria for acceptability and reliability were used; participants with a minimum follow-up duration of 24 months between the first and last FEV_1 measures were included.

Statistical Analysis

Linear mixed-effects models were used to investigate the association between dietary patterns and the rate of change in FEV_1. All models included a random intercept and a random time effect at the individual level to differentiate between- and within-individual variation, and were adjusted for the following covariates as fixed effects: time (a continuous variable quantifying the time distance between each FEV_1 measure and the study baseline), race/ethnicity, baseline age, standing height, weight, smoking status during follow-up and its interaction with time, baseline smoking pack-years, educational attainment, marital status, residence status, treatment group in SELECT (selenium as 200 μg/d *L*-selenomethionine, vitamin E as 400 IU/d *all rac*-α-tocopheryl acetate, and matching placebos, 4-arm design, thus 3 treatment dummy variables with double placebo as reference group), multivitamin use, and energy intake. Given the relatively short follow-up duration of three years, we showed that modeling the longitudinal change in FEV_1 using the linear effect of the time variable was sufficient. Smoking status during follow-up was defined as a categorical variable with four classes: never smokers at all time points (reference group), persistent smokers who were

current smokers at all time points, former smokers at all time points, and intermittent smokers who changed smoking status across time points. Prior to the final analyses, residual diagnosis was performed based on the above preliminary model to exclude FEV_1 measurements detected as outliers (|standardized residual| > 3). The main effect of each dietary pattern on the rate of change in FEV_1 was tested by including the dietary pattern × time product term in the model, where dietary pattern scores were divided into quintiles. The interaction between each dietary pattern and smoking status during follow-up was subsequently examined by including the dietary pattern × smoking status × time product term, where dietary pattern scores were divided into tertiles and smoking status was collapsed into two categories: smokers (persistent and intermittent) and non-smokers (former and never) during follow-up to maximize power to detect interactions. A test for trend across the quintiles or tertiles of each pattern was calculated by treating the categories as an ordinal variable in the mixed-effects models. All analyses were conducted using SAS software version 9.2 (SAS Institute, Cary, NC).

RESULTS

Population Characteristics

The present study excluded all participants with missing dietary data from the baseline FFQ or with unreasonably high (>4500 kcal/d) or low (<800 kcal/d) energy intakes and those that had left >70 food items blank were excluded from the analysis (23). Participants were also excluded if they had missing lung function or covariate data. The final study population included 2,590 of the 2,920 men from the original cohort (89%; **Table 4.1**). The average age was 62.7 years and about 80% of the men were White. Given the oversampling for cigarette smokers, 10.7% and 7.1% of the study participants were persistent and intermittent smokers

during follow-up, respectively. The majority of men were married and not living alone, and over 40% reported multivitamin use at study baseline. The average baseline FEV_1 was slightly under 3 L, the average number of FEV_1 measures per participant was 3.1, and the average follow-up duration was 3.0 years. Based on the mixed effects model that was fully adjusted for all confounding variables and did not include the dietary pattern variable, the estimated rate of change in FEV_1 in never smokers was a decline of 34.6 mL/year. In comparison, the rate of decline was 11.4 ± 6.5 mL/year steeper in persistent smokers, 4.2 ± 7.4 mL/year steeper in intermittent smokers, and 0.5 ± 4.1 mL/year steeper in former smokers.

Dietary Patterns

Using principal component analysis, two distinct dietary patterns were identified in RAS men. The factor loadings of food groups in the two identified patterns are shown in **Table 4.2**. Each individual factor loading value can be approximately interpreted as the correlation coefficient between the intake of a food group and a dietary pattern; food groups with the largest positive loading values contribute the most to variability in the dietary pattern, and those with the most negative values contribute the least. The first pattern, which explained the greatest amount of variation in the dietary data (8.6%), was characterized by a high intake of fruits, vegetables, fish, and poultry. The second pattern, which explained 5.5% of the total variation, was characterized by a high intake of red, fried and processed meats, desserts and sweets, butter and margarine, fries, snacks, pizza, refined grains, and potatoes. Previous studies of dietary patterns in similar populations have labeled similar patterns as the "prudent" and the "Western" patterns, respectively. Together, the "prudent" and "Western" patterns explained 14.1% of the variation in the consumption of the 54 food groups.

The characteristics of the study participants by dietary pattern group (**Table 4.1)** show that men with a more prudent diet were less likely to be smokers during the study follow-up, more likely to have lower smoking pack-years at study baseline, more likely to be married and less likely to be living alone, of higher educational status, and more often users of multivitamin supplements. In contrast, men with a more Western diet were more likely to be smokers during the study follow-up, more likely to have higher smoking pack-years, less likely to be married, more likely to have lower educational attainment, and less likely to take multivitamin supplements.

The two dietary patterns were further examined in relation to nutrient intakes using Pearson's correlation coefficients (**Table 4.3**). In particular, the prudent pattern had a strong positive correlation with intakes of dietary fiber, vitamins C, β-carotene, and magnesium and a strong negative correlation with intakes of total, saturated, and trans fat. In contrast, the Western pattern had a strong positive correlation with intakes of total, saturated, and trans fat and a strong negative correlation with intakes of dietary fiber, vitamin C, β-carotene, and magnesium.

Dietary Patterns and the Rate of Change in FEV_1

After adjusting models for potential confounding variables, neither the prudent nor the Western pattern was statistically significantly associated with the rate of change in FEV_1 over the three-year follow-up (**Table 4.4**). The estimated difference in the rate of change in FEV_1 was 1.2 mL per year (95% CI: -11.0, 13.4; *P* for trend = 0.80) between men in the highest and lowest quintiles of the prudent pattern, and 5.2 mL per year (95% CI: -11.5, 21.9, *P* for trend = 0.63) between men in the highest and lowest quintiles of the Western pattern.

Given the importance of cigarette smoking as a major risk factor for accelerated lung function loss and COPD, and the hypothesis that effects of smoking can be mitigated by exogenous sources of antioxidants including dietary sources, the interaction between dietary patterns and smoking status was examined in relation to the rate of change in FEV_1. After adjusting for potentially confounding variables, a more prudent diet was statistically significantly associated with a slower rate of decline in FEV_1 in smokers, whereas no association was observed in non-smokers (interaction *P* for trend = 0.001; **Table 4.5** and **Figure 4.1**). The estimated difference in the rate of change in FEV_1 between men in the highest and lowest tertiles of the prudent pattern was 38.8 mL per year in smokers (95% CI: 14.4, 63.1) and -5.2 mL per year in non-smokers (95% CI: -15.1, 4.7). In contrast, no interaction was observed between the Western pattern and smoking (data not shown).

DISCUSSION

In this study, we assessed the relation between dietary patterns and the longitudinal change in lung function, measured as the rate of change in FEV_1, in a large prospective cohort of North American men, and explored interaction between dietary patterns and cigarette smoking. With the use of principal component analysis, two distinct dietary patterns were identified: a prudent pattern characterized by high intakes of fruits, vegetables, poultry, and fish, and a Western pattern characterized by high intakes of red, fried and processed meats, desserts and sweets, butter and margarine, fries, snacks, pizza, refined grains, and potatoes. Neither pattern was associated with the rate of change in FEV_1 in the overall population. When the interaction with smoking was considered, a more prudent diet was significantly associated with a slower rate of decline in FEV_1 in smokers, thus counteracting the

detrimental effect of cigarette smoking, but had no association with the rate of decline in FEV_1 in non-smokers.

The use of dietary patterns to characterize overall dietary intake provides an additional dimension to the epidemiologic research of diet and lung function and COPD. Findings based on foods and food groups are relevant to public health, as it is more feasible to give population-level dietary advice to modify the intake of a certain food or food group versus a particular nutrient. To date, four studies have investigated the relation of dietary patterns with lung function and COPD outcomes. In three of them, a prudent dietary pattern rich in fruits, vegetables, fish, and whole-grain products was associated with better lung function at one time point (17) or reduced risk of incident COPD (18, 19). One study further examined the prudent pattern in relation to lung function decline over five years in a subpopulation, but found no overall association or interaction with smoking (13). Several factors could have affected the detection of an interaction between the prudent pattern and smoking in the above study. First, lung function was only measured twice (compared with three measurements in the current study), thus the statistical analysis of factors affecting the longitudinal change in lung function was likely affected by additional noise, or measurement error, in the outcome variable. Another important difference arises in the parameterization of cigarette smoke exposure, which was characterized and modeled as the cross-sectional status at study baseline. In contrast, we studied the cumulative smoking pattern during follow-up, which better represents concurrent oxidative stress due to smoking. Lastly, the population used in the above study included both males and females (compared with only males in the current study), and it is possible that the relation between diet, smoking, and lung function differs by gender. In summary, our findings are consistent with the protective associations reported for the

prudent pattern in prior research, and add important novel insights to the evidence base in the identification of diet—smoking interactions in relation to the rate of change in lung function.

Individual foods and nutrients that contribute to the prudent pattern have been studied in relation to longitudinal lung function outcomes. In a population of healthy British adults, reduced fresh fruit consumption over time was associated with a faster rate of decline in FEV_1 over seven years, with the strongest effects observed in cigarette smokers (24). In older US adults, higher intake of fruits and vegetables and higher intakes of individual nutrients with antioxidant properties were associated with a slower decline in FEV_1 over four years in smokers alone (12). Two additional studies reported overall associations of higher apple consumption and higher vitamin C intake in relation to slower decline in lung function (13, 25), regardless of smoking status. This past research on single foods and/or nutrients supports the findings reported herein.

The current study has several strengths. First, we characterized diet in terms of overall dietary patterns rather than studying individual nutrients to better reflect the consumption of meals consisting of a variety of foods with complex combinations of nutrients that are likely interactive or synergistic. Second, we focused on the longitudinal change in lung function as the primary outcome, which is understudied as an informative predictor of COPD risk due to the scarcity of cohort studies with high-quality longitudinal lung function data and the increased complexity in statistical analysis. Third, this study used a large, nationally representative population of North American men (22% African American), which was selectively enriched for cigarette smokers, allowing for the investigation and detection of meaningful interaction between dietary patterns and smoking. Fourth, the longitudinal lung function data used in this study was acquired through high-quality spirometry tests performed

by trained clinical research nurses and following standardized protocol. Finally, extensive data were available on important risk and confounding factors, such as smoking behaviors and multivitamin use, allowing the careful adjustment for these factors in statistical models.

A few limitations should be considered when evaluating the study findings. First, the use of principal component analysis to derive dietary patterns involves several arbitrary decisions that could affect the results, such as the consolidation of the initial 120 food items into 54 food groups. However, the grouping of food items was guided by knowledge of their similarity in nutrient profile and culinary use, and was similar to those used in previous studies of dietary patterns, allowing direct comparisons to be made to prior work. The resulting patterns were interpretable, and consistent with previous studies in similar populations. Second, while the labeling of the dietary pattern is subjective, labels were assigned according to existing knowledge based on extensive research on dietary patterns in Western populations (the factor loading matrix for each pattern and correlation coefficients between pattern scores and individual nutrients are provided for further evaluation). Lastly, the dietary patterns were derived based on a single FFQ administered at study baseline, thus measurement error in the classification may exist. This concern is mitigated by the understanding that FFQs are preferred measures of long-term, habitual dietary intake, with high validity in assigning the relative ranking of intakes among individuals within a single population.

In conclusion, in this prospective study of dietary patterns in relation to the longitudinal change in lung function in a large, representative cohort of North American men, we observed an interaction between the prudent dietary pattern and cigarette smoking in relation to the rate of change in FEV_1. This finding lends further support to the hypothesis

that a diet rich in fruits and vegetables is beneficial to lung function. This work extends the current evidence base by studying longitudinal outcomes, and finds that a diet rich in foods that may contribute to antioxidant capacity and/or to a reduction in inflammation is protective in cigarette smokers. Confirmation of this finding in other populations, particularly among women who are actively smoking, is a priority for future research.

Acknowledgements: We acknowledge the generous contributions of the men in the Respiratory Ancillary Study, for without their involvement this research project would not have been possible. We also acknowledge the contributions of the site leaders and the site clinical research associates, who agreed to collect data for the Respiratory Ancillary Study to SELECT, and whose collaboration and hard work were essential to the study.

Authors' Contributions: PAC, SML, and EK designed research; PAC, KAG, and JAH collected data; WT, KAG, and PAC analyzed and interpreted data; WT and PAC wrote the paper; PAC had primary responsibility for final content. All authors reviewed and approved the final manuscript.

Conflicts of Interests: None

REFERENCES

1. Schunemann HJ, Dorn J, Grant BJ, et al. Pulmonary function is a long-term predictor of mortality in the general population: 29-year follow-up of the Buffalo Health Study. *Chest*. 2000;118(3):656-64.

2. Rabe KF, Hurd S, Anzueto A, et al. Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: GOLD executive summary. *Am J Respir Crit Care Med*. 2007;176(6):532-55.

3. Fletcher C. The natural history of chronic bronchitis and emphysema: an eight-year study of early chronic obstructive lung disease in working men in London. Oxford University Press. 1976.

4. Rennard SI, Vestbo J. Natural Histories of Chronic Obstructive Pulmonary Disease. *Proceedings of the American Thoracic Society*. 2008;5(9):878-83.

5. Han MK, Agusti A, Calverley PM, et al. Chronic obstructive pulmonary disease phenotypes: The future of COPD. *Am J Respir Crit Care Med*. 2010;182(5):598-604.

6. Wise R. The value of forced expiratory volume in 1 second decline in the assessment of chronic obstructive pulmonary disease progression. *Am J Med*. 2006;119(10):4-11.

7. Beaty TH, Menkes HA, Cohen BH, et al. Risk factors associated with longitudinal change in pulmonary function. *Am Rev Respir Dis*. 1984;129(5):660-7.

8. Beck GJ, Doyle CA, Schachter EN. Smoking and lung function. *Am Rev Respir Dis*. 1981;123(2):149-55.

9. Kauffmann F, Drouet D, Lellouch J, et al. Twelve years spirometric changes among Paris area workers. *Int J Epidemiol*. 1979;8(3):201-12.

10. Peat JK, Woolcock AJ, Cullen K. Decline of lung function and development of chronic airflow limitation: a longitudinal study of non-smokers and smokers in Busselton, Western Australia. *Thorax*. 1990;45(1):32-7.

11. Romieu I, Trenga C. Diet and obstructive lung diseases. *Epidemiologic reviews*. 2001;23(2):268-87.

12. Bentley AR, Kritchevsky SB, Harris TB, et al. Dietary antioxidants and forced expiratory volume in 1 s decline: the Health, Aging and Body Composition study. *Eur Respir J*. 2012;39(4):979-84.

13. McKeever TM, Scrivener S, Broadfield E, et al. Prospective study of diet and decline in lung function in a general population. *Am J Respir Crit Care Med*. 2002;165(9):1299-303.

14. National Research Council (U.S.). Committee on Diet and Health. Diet and health : implications for reducing chronic disease risk. Washington, D.C.: National Academy Press; 1989. xiv, 749 p. p.

15. Hu FB. Dietary pattern analysis: a new direction in nutritional epidemiology. *Current opinion in lipidology*. 2002;13(1):3-9.

16. McKeever TM, Lewis SA, Cassano PA, et al. Patterns of dietary intake and relation to respiratory disease, forced expiratory volume in 1 s, and decline in 5-y forced expiratory volume. *Am J Clin Nutr*. 2010;92(2):408-15.

17. Shaheen SO, Jameson KA, Syddall HE, et al. The relationship of dietary patterns with adult lung function and COPD. *Eur Respir J*. 2010;36(2):277-84.

18. Varraso R, Fung TT, Barr RG, et al. Prospective study of dietary patterns and chronic obstructive pulmonary disease among US women. *Am J Clin Nutr*. 2007;86(2):488-95.

19. Varraso R, Fung TT, Hu FB, et al. Prospective study of dietary patterns and chronic obstructive pulmonary disease among US men. *Thorax*. 2007;62(9):786-91.

20. Lippman SM, Klein EA, Goodman PJ, et al. Effect of selenium and vitamin E on risk of prostate cancer and other cancers: the Selenium and Vitamin E Cancer Prevention Trial (SELECT). *JAMA*. 2009;301(1):39-51.

21. Barr RG, Stemple KJ, Mesia-Vela S, et al. Reproducibility and validity of a handheld spirometer. *Respir Care*. 2008;53(4):433-41.

22. Standardization of Spirometry, 1994 Update. American Thoracic Society. *Am J Respir Crit Care Med*. 1995;152(3):1107-36.

23. Hu FB, Rimm E, Smith-Warner SA, et al. Reproducibility and validity of dietary patterns assessed with a food-frequency questionnaire. *Am J Clin Nutr*. 1999;69(2):243-9.

24. Carey IM, Strachan DP, Cook DG. Effects of changes in fresh fruit consumption on ventilatory function in healthy British adults. *Am J Respir Crit Care Med*. 1998;158(3):728-33.

25. Butland BK, Fehily AM, Elwood PC. Diet, lung function, and lung function decline in a cohort of 2512 middle aged men. *Thorax*. 2000;55(2):102-8.

26. Willett W, Stampfer MJ. Total energy intake: implications for epidemiologic analyses. *Am J Epidemiol*. 1986;124(1):17-27.

TABLE 4.1 Characteristics of study participants, overall and by quintile (Q) of dietary pattern score*

	Overall Population (n = 2,590)	Prudent pattern			Western pattern		
		Q1 (n = 524)	Q3 (n = 515)	Q5 (n = 508)	Q1 (n = 507)	Q3 (n = 521)	Q5 (n = 521)
Age (y)	62.7 ± 6.4	62.5 ± 6.8	62.9 ± 6.0	62.6 ± 6.4	63.3 ± 6.1	62.8 ± 6.7	61.6 ± 6.1
White (%)	77.9	70.0	80.6	80.0	79.7	77.7	72.9
Height (cm)	176.4 ± 7.1	176.5 ± 7.2	176.3 ± 6.8	176.7 ± 7.1	175.6 ± 7.2	176.5 ± 7.2	177.4 ± 6.8
Weight (kg)	89.4 ± 17.4	88.6 ± 18.0	89.7 ± 17.0	90.5 ± 17.7	86.1 ± 16.5	89.7 ± 16.6	93.6 ± 18.8
Smoking status during follow-up (%)							
Never smokers	35.3	28.4	36.7	36.6	42.0	35.9	26.1
Former smokers	46.8	43.7	48.4	51.0	48.3	47.0	47.0
Intermittent smokers	7.1	9.4	6.6	5.1	3.2	8.1	9.8
Persistent smokers	10.7	18.5	8.4	7.3	6.5	9.0	17.1
Pack-years+	21.4 (7 – 36)	23.9 (12 – 43)	21.1 (7 – 34)	16.4 (7 – 34)	14.1 (7 – 34)	19.2 (7 – 36)	23.9 (14 – 43)
Married (%)	77.0	71.0	79.0	80.1	78.3	75.8	72.4
Living alone (%)	14.4	19.5	15.0	12.0	14.8	15.9	15.9
Educational attainment (%)							
High school	28.4	40.5	25.8	20.9	21.5	30.1	37.0
College	41.3	42.8	41.9	43.3	41.6	40.5	39.7
Graduate school	30.3	16.8	32.2	35.8	36.9	29.4	23.2
Multivitamin use (%)	40.6	35.3	39.4	47.2	45.6	38.8	35.9
Total energy (kcal)	2303.7 ± 806.3	1898.2 ± 735.2	2227.6 ± 714.7	2806.2 ± 760.5	1582.0 ± 575.8	2132.9 ± 488.4	3285.6 ± 573.9
FEV_1 (mL)	2973.9 ± 678.0	2895.5 ± 666.1	3005.3 ± 673.4	3006.1 ± 672.0	2983.0 ± 627.6	3043.4 ± 688.8	2923.0 ± 684.6

[*] Data presented are from the study baseline, unless otherwise stated. Quintile 1 represents the lowest dietary pattern intake and quintile 5 the highest. Data are presented as %, mean ± SD or median (interquartile range).

[+] Determined as no. of packs smoked per day × no. of years smoked among ever smokers.

TABLE 4.2 Factor loadings for the prudent and Western dietary patterns at study baseline from the principal component analysis*

Food group	Prudent pattern	Western pattern
Leafy vegetables	0.69	-
Carrots, squashes and yams	0.66	-
Cruciferous vegetables	0.63	-
Tomatoes	0.60	-
Peppers	0.55	-
Fruits	0.55	-
Legumes	0.54	-
Onions and garlic	0.50	-
Fish	0.44	-
Avocado	0.34	-
Dressing (low-fat)	0.32	-
Corn	0.32	-
Poultry	0.32	-
Mixed dishes with meat	-	0.50
Desserts/sweets	-	0.47
Fries	-	0.47
Red meats	-	0.46
Butter and margarine	-	0.46
Condiments	-	0.43
Fried meats	-	0.42
Processed meats	-	0.41
Snacks	-	0.39
Pizza	-	0.34
Refined Grains	-	0.32
Cream soups	-	0.31
Mayo (high-fat)	-	0.31
Potatoes	-	0.31

* Factor loadings represent the correlations between the intakes of food groups and dietary pattern scores. Absolute values < 0.30 were not shown for simplicity. Factor loadings presented are those computed from the orthogonal rotation.

TABLE 4.3 Pearson's correlation coefficients (*r*) between the prudent and Western dietary pattern scores and nutrient intakes[*]

	Prudent pattern	Western pattern
Total fat (g/d)	-0.21	0.24
Saturated fat (g/d)	-0.34	0.27
Monounsaturated fat (g/d)	-0.16	0.19
Polyunsaturated fat (g/d)	0.01[+]	0.13
Trans fat (g/d)	-0.38	0.38
Cholesterol (mg/d)	-0.05	0.09
Carbohydrates (g/d)	0.15	-0.07
Proteins (g/d)	0.16	-0.04
Dietary Fiber (g/d)	0.50	-0.21
Vitamin A (μg/d)	-0.08	0.06
Vitamin C (mg/d)	0.46	-0.22
Vitamin D (μg/d)	0.12	-0.14
Vitamin E (mg/d)	0.18	-0.09
β-Carotene (μg/d)	0.60	-0.25
Magnesium (mg/d)	0.45	-0.34
Selenium (μg/d)	0.18	-0.05

[*] Intakes of presented nutrients were computed from the food frequency questionnaire administered at study baseline and were adjusted for energy intake prior to calculating the Pearson's correlation coefficients using the residual method (26). All correlations were significant ($P < 0.05$), unless otherwise indicated.

[+] $P > 0.05$.

TABLE 4.4 Association between quintile (Q) of dietary patterns and the rate of change in forced expiratory volume in 1s (FEV_1)[*]

	Prudent pattern			Western pattern		
	β (mL/y)	95% CI	*P* for trend[+]	β (mL/y)	95% CI	*P* for trend[+]
Q1 (lowest)	Reference			Reference		
Q2	6.7	-4.7, 18.2		1.0	-10.9, 12.9	
Q3	4.9	-6.3, 16.2	0.80	-6.5	-18.8, 5.9	0.63
Q4	0.1	-11.3, 11.4		3.8	-10.0, 17.7	
Q5 (highest)	1.2	-11.0, 13.4		5.2	-11.5, 21.9	

[*] Data presented are effect estimates from the linear mixed effects models adjusted for race/ethnicity, age, standing height, weight, smoking status during follow-up and its interaction with time, smoking pack-years, educational attainment, marital status, residence status, treatment group in SELECT, multivitamin use, and energy intake (all at study baseline, unless otherwise stated).

[+] Dietary pattern quintiles were modeled as a continuous variable in the mixed effects models.

TABLE 4.5 Association between tertile (T) of the prudent pattern and the rate of change in forced expiratory volume in 1s (FEV_1) according to smoking status during follow-up[*]

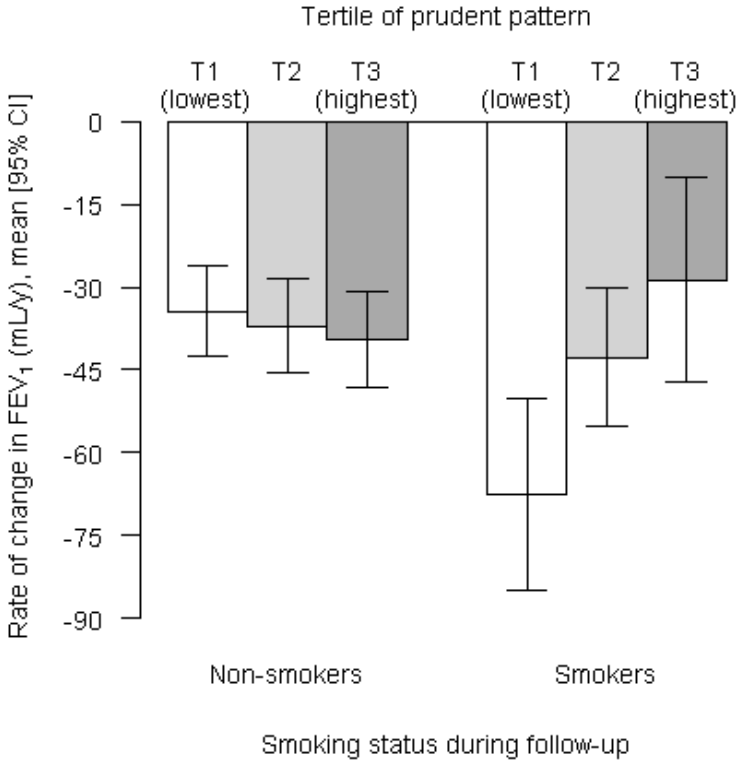
	Non-smokers			Smokers		
	β (mL/y)	95% CI	*P* for trend[+]	β (mL/y)	95% CI	*P* for trend[+]
T1 (lowest)	Reference			Reference		
T2	-2.7	-12.5, 7.0	0.31	24.8	4.1, 45.4	0.002
T3 (highest)	-5.2	-15.1, 4.7		38.8	14.4, 63.1	
Interaction *P* for trend[+, ǂ]			0.001			

[*] Data presented are effect estimates from the linear mixed effects models adjusted for race/ethnicity, age, standing height, weight, smoking status during follow-up and its interaction with time, smoking pack-years, educational attainment , marital status, residence status, treatment group in SELECT, multivitamin use, and energy intake (all at study baseline, unless otherwise stated).

[+] Dietary pattern tertiles were modeled as a continuous variable in the mixed effects models.

[ǂ] Interaction was modeled using the three-way product term of prudent status × smoking status × time in the mixed effects model.

FIGURE 4.1 Estimated rate of change in forced expiratory volume in 1s (FEV_1) according to tertile of prudent pattern intake and smoking status during follow-up. Open bars represent the lowest intake tertile (T1), lightshade bars represent the medium intake tertile (T2), and dark shaded bars represent the highest intake tertile (T3). The estimates were computed from the linear mixed effects model that was adjusted for all covariates and included the prudent pattern × smoking status × time product term (interaction *P* for trend = 0.001). CI = confidence interval.



CHAPTER 5

CONCLUSION

Reduced lung function is associated with increased morbidity and mortality at the population level (1), and is a primary characteristic of chronic obstructive pulmonary disease (COPD) (2). COPD is currently the third leading cause of death in the United States and a significant public health burden globally (3-5). Longitudinal change in lung function is an informative indicator of COPD risk and serves as a valuable research outcome for gaining insights into COPD pathogenesis and for developing effective preventive and curative strategies (6-9).

Cigarette smoking is the most important risk factor for accelerated loss in lung function and COPD (10, 11), but other factors, including genetics and nutrition, are postulated to also play important roles in explaining the variability in lung function decline and COPD susceptibility. Existing population-level research investigating the association of genetic and nutritional factors with lung outcomes has generated extensive evidence supporting their involvement, but longitudinal studies based on large, representative populations and with careful consideration of the effect of smoking are lacking.

In this context, the three projects comprising this dissertation focus on the role of genetics, nutrition, and cigarette smoking in the determination of longitudinal change in lung function over the adult life course and in the etiology of chronic obstructive pulmonary disease (COPD). Specifically, the overarching goal of this research is to better understand whether common genetic variants, both within the antioxidant enzyme network and

throughout the human genome, as well as overall patterns in dietary intake contribute to individuals' susceptibility to COPD by influencing the rate of change in lung function beyond the natural aging process, particularly in persons experiencing elevated levels of oxidative stress from exposure to cigarette smoke.

These projects were designed to be distinct in terms of objectives, study design, population, and statistical methods, but are complementary in addressing major gaps in the published literature. A brief summary of each project is presented below.

Genetic Variation in Antioxidant Enzymes, Cigarette Smoking and Longitudinal Change in Lung Function (Chapter 2)

Most published candidate gene association studies of lung function focus on a limited set of candidate genes and cross-sectional outcomes. The first project addressed this gap by exploring the association of 384 single-nucleotide polymorphisms (SNPs), selected systematically from 56 antioxidant enzyme genes, with the rate of change in two lung function measures, forced expiratory volume in the first second (FEV_1) and the ratio of FEV_1/forced vital capacity (FEV_1/FVC). The analysis was conducted separately in African and European American elderly adults from the Health, Aging, and Body Composition (Health ABC) study. In European Americans, single-marker analyses confirmed a prior finding involving the *GCLC* SNP rs17883901 in relation to the rate of change in FEV_1/FVC. In addition, we identified statistically significant gene-by-smoking interactions for two novel candidate genes (*mGST3* and *IHD3B*) with the rate of change in FEV_1/FVC in European Americans, and observed gene-level replications for both genes in African Americans at nominal significance. Overall, these findings support the hypothesis that genetic variation in

genes contributing to the endogenous antioxidant defense affects longitudinal change in lung function, particularly in cigarette smokers who are more susceptible to an imbalance between oxidative stress and antioxidant protection.

Genome-Wide Association Studies of Longitudinal Change in Adult Lung Function (Chapter 3)

To explore genetic variation beyond the scope of the antioxidant enzyme network, and to address the lack of genome-wide association studies (GWAS) of longitudinal lung function outcomes, the second project investigated the association of ~2.5M common SNPs throughout the human genome with the rate of change in FEV_1, in a combined sample of over 27,000 adults of European ancestry from 14 prospective cohort studies. The meta-analysis incorporating results from all 14 cohorts produced suggestive evidence for association at the novel *IL16/STARD5/TMC3* locus on chromosome 15. A second meta-analysis using the five cohorts with optimal longitudinal lung function data identified the novel *ME3* locus on chromosome 11 at genome-wide significance. We confirmed the expression of *IL16*, *STARD5*, and *ME3* in multiple lung tissues, and used publicly available microarray data to demonstrate differential expression of all three genes in lung samples from COPD patients compared with controls. These results together suggest biologically plausible functional links between the identified genes and lung function, which are promising candidates for further studies.

Dietary Patterns, Cigarette Smoking, and Longitudinal Change in Lung Function (Chapter 4)

The third project investigated the relation between overall dietary patterns and the rate

of change in FEV_1 in 2,560 male adults from the Respiratory Ancillary Study. Two distinct dietary patterns were derived using principal component analysis based on food frequency questionnaire data collected at study baseline: a prudent pattern characterized by high intakes of fruits, vegetables, fish, and poultry and a Western pattern characterized by high intakes of red, fried, and processed meats, desserts, and fries. Although neither pattern was associated with the rate of change in FEV_1 in the overall population, when the interaction between dietary pattern and smoking was considered, the prudent diet was associated with a slower rate of decline in FEV_1 only in cigarette smokers. These findings support the hypothesis that a diet rich in foods and nutrients that may augment antioxidant and/or anti-inflammatory capacity is protective against accelerated lung function loss and COPD in smokers.

Future Directions

While findings from this dissertation research contribute to a more comprehensive understanding of the role of genetics, nutrition, and cigarette smoking in the determination of longitudinal change in lung function and COPD susceptibility, it also suggests directions for future research. First, the identification of strong gene-by-smoking and diet-by-smoking interactions, as reported in the first and third projects, clearly highlights the importance for future studies to more carefully characterize and capture the interrelatedness and interaction between these three distinct exposures. Although not explicitly pursued in this dissertation research, the presence of extensive interaction between genes and nutrients is becoming increasingly well-recognized and future population-level studies designed to address such questions hold great promise in helping to identify simple nutritional interventions targeting specific population subgroups for the prevention and treatment of chronic diseases including

COPD. An important observation from the GWAS of longitudinal change in lung function is the considerable phenotypic heterogeneity among the 14 cohort studies in the meta-analysis, and the challenge it presents to the identification of modest genetic effects on a longitudinal outcome at stringent statistical thresholds. Large-scale GWAS meta-analyses involving tens of thousands of study participants have so far been successful in the study of cross-sectional complex traits and diseases. However, this strategy is unlikely to enjoy the same success when applied to longitudinal outcomes, given the considerable increase of noise present in the phenotypic data and the possibly smaller genetic contribution as measured by heritability. Therefore, future GWAS of longitudinal outcomes face the dilemma of needing to further increase sample size by enrolling more individual studies with different characteristics, while at the same time reduce phenotypic heterogeneity in order to detect true genetic associations with more modest effect magnitude.

REFERENCES

1. Schunemann HJ, Dorn J, Grant BJ, et al. Pulmonary function is a long-term predictor of mortality in the general population: 29-year follow-up of the Buffalo Health Study. *Chest*. 2000;118(3):656-64.

2. Rabe KF, Hurd S, Anzueto A, et al. Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: GOLD executive summary. *Am J Respir Crit Care Med*. 2007;176(6):532-55.

3. Lopez AD, Murray CC. The global burden of disease, 1990-2020. *Nature medicine*. 1998;4(11):1241-3.

4. Michaud CM, Murray CJ, Bloom BR. Burden of disease--implications for future research. *JAMA*. 2001;285(5):535-9.

5. Minino AM, Murphy SL. Death in the United States, 2010. *NCHS data brief*. 2012(99):1-8.

6. Fletcher C. The natural history of chronic bronchitis and emphysema: an eight-year study of early chronic obstructive lung disease in working men in London. Oxford University Press. 1976.

7. Wise R. The value of forced expiratory volume in 1 second decline in the assessment of chronic obstructive pulmonary disease progression. *Am J Med*. 2006;119(10):4-11.

8. Han MK, Agusti A, Calverley PM, et al. Chronic obstructive pulmonary disease phenotypes: The future of COPD. *Am J Respir Crit Care Med*. 2010;182(5):598-604.

9. Rennard SI, Vestbo J. Natural Histories of Chronic Obstructive Pulmonary Disease. *Proceedings of the American Thoracic Society*. 2008;5(9):878-83.

10. Lindberg A, Bjerg A, Ronmark E, et al. Prevalence and underdiagnosis of COPD by disease severity and the attributable fraction of smoking Report from the Obstructive Lung Disease in Northern Sweden Studies. *Respir Med*. 2006;100(2):264-72.

11. Mannino DM, Buist AS, Petty TL, et al. Lung function and mortality in the United States: data from the First National Health and Nutrition Examination Survey follow up study. *Thorax*. 2003;58(5):388-93.