NOVEL INSIGHTS INTO THE GENETIC BASIS OF CHRONIC OBSTRUCTIVE PULMONARY DISEASE

by

Josiah E. Radder

B.S. in Biochemistry, Arizona State University, 2007

B.S. in Biology, Arizona State University, 2007

B.A. in Religious Studies, Arizona State University, 2007

Submitted to the Graduate Faculty of

University of Pittsburgh School of Medicine in partial fulfillment

of the requirements for the degree of

Ph.D. in Cellular and Molecular Pathology

University of Pittsburgh

2016

UNIVERSITY OF PITTSBURGH

SCHOOL OF MEDICINE

This dissertation was presented

by

Josiah E. Radder

It was defended on

August 9, 2016

and approved by

Wendy M. Mars, PhD
Committee Chair
Associate Professor, Department of Pathology

M. Michael Barmada, PhD
Associate Professor, Department of Human Genetics

Yuri E. Nikiforov, MD, PhD
Professor, Department of Pathology

Frank C. Sciurba, MD
Associate Professor, Department of Medicine

Steven D. Shapiro, MD
Major Advisor
Distinguished Professor, Department of Medicine

Portions of this dissertation were adapted from the following publication, with permission of Future Medicine Ltd:

Radder JE, Shapiro SD, Berndt AM. Personalized Medicine for Chronic, Complex Diseases: COPD as an Example. Personalized Medicine. October 2014. Vol. 11, No. 7, Pages 669-679.

NOVEL INSIGHTS INTO THE GENETIC BASIS OF CHRONIC OBSTRUCTIVE PULMONARY DISEASE

Josiah E. Radder, Ph.D.

University of Pittsburgh, 2016

Chronic obstructive pulmonary disease (COPD), defined as irreversible airflow limitation, is caused by a complex interaction of environmental exposures, most commonly cigarette smoke, and genetic factors. Genetic studies of COPD have used tests of genome-wide linkage and association to identify loci that contribute to disease susceptibility. However, as seen in other chronic diseases, the best-replicated loci associated with COPD only account for a small portion of disease heritability. Identifying additional genetic determinants of chronic diseases offers the opportunity to better understand their biology as well as the promise of better disease prediction and patient stratification, first steps in the development of precision medicine.

The genetic architecture underlying chronic disease is complex, and it is likely that there is still common variation contributing to COPD that has been masked from association studies by phenotypic and genotypic heterogeneity. Further, there is evidence that rare variation contributes to chronic disease susceptibility, and rare variants in *SERPINA1* leading to alpha-1 antitrypsin deficiency support this in COPD. The trio of studies presented in this work aim to detect both of these types of variation. In the first, we employ an extreme-trait study design to detect rare variants in the first whole genome sequencing study of COPD. Using this approach, we identify a previously unreported non-synonymous variant associated with COPD, and two suggestively associated candidate genes, *PTPRO* and *ZNF816*. In the second and third studies, we integrate mouse and human genetic data to identify undetected common variants associated with human

disease and mouse models of disease. The first study uses a mouse model of cigarette smoke-induced emphysema and identifies the gene *ABI3BP* as a potential candidate gene. The second looks at early life determinants of chronic disease by measuring airspace size in mice at maturity, leading to identification of *IL1R2*, which plays a previously undescribed role in lung development. Finally, we demonstrate that by integrating the results of genetic studies, it is possible to gather additional information about the genetic architecture of chronic diseases like COPD.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PREFACE

ACKNOWLEDGEMENTS

There are so many people to thank for helping me finish the work presented here, I could probably have written an entire chapter just describing the contributions of my mentors, colleagues, friends and family. Thank you to all of you for the many ways you've helped, mentioned here or not, I couldn't have done any of it alone.

First, to my wife Anna, thank you for being a daily source of joy for me and June, we both love you. I'm so grateful to have had you by my side for all of our adventures up to this point and so glad that you'll be there for the adventures that will come with going back to medical school. Knowing that I have your support is always a huge source of comfort and motivation. And of course, thank you for being a good mom to June. June - thank you for always giving me something to look forward to when I get home, and for always avoiding permanent marker when you color on my laptop.

I can't thank Steve Shapiro enough for his mentorship over the last three years. From the first time we met to talk about a summer project, Steve has encouraged me to ask interesting questions and then offered me the means to pursue them, even when they haven't been easy. As a graduate student, the opportunity to work on the first whole genome sequencing study of COPD was an exceptional one and taught me a myriad of new skills. More importantly, working on these projects with a great scientist like Steve taught me how to think critically, how to manage a project involving lots of people, and how to prioritize at each step along the way. Steve also gave me many opportunities to present my science and learn about others' science at international meetings and

courses all over the world. Steve's support has also extended outside of the lab – he has been a great resource in thinking about my clinical career and was extremely flexible as a mentor to a graduate student who was also a new dad. As a clinical leader, an internationally-recognized scientist and an executive in the healthcare system, Steve is an incredible example of the many ways to not only be successful in medicine but also shape the field, and he will always be my model of a physician-scientist.

I am extremely grateful to the Shapiro lab for their friendship and their contribution to this work. Thanks to Alyssa Gregory for her essential, daily support in the lab. From her expertise with experimental design to insightful comments on manuscripts to her knowledge of how to get pretty much anything done, all of the work presented here relied on her in some way. I'd like to thank Neil Kelly for always offering advice about navigating graduate and medical school, and more importantly for always being willing to help me think through scientific problems carefully and critically. Thanks to Paula Nave for helping me get time in Steve's schedule, planning thesis committee meetings and numerous other organizational tasks that only she could have done. Thanks to Chris Burton for her hard work on the animal work presented in this study and for being willing to help even when it meant coming to the lab for experiments at all hours of the day. Thanks to Stephen Scott and Andy Metz for contributing to all of the smoking studies presented here. Veli Bakalov, Murat Kaynar, and JP Wood, thank you for always being open to an interesting conversation, whether it was about science or not. Annerose Berndt was a great resource at the beginning of my thesis work, helping me to acquire methods for association-testing in mice and stoking my interest in the genetics of chronic disease. Finally, I'd like to thank Adriana Leme for her contributions to this work, particularly measuring morphometry and lung function in mice, which much of this work depends on.

Executing the large projects described here also required a significant amount of mentorship and collaboration from outside of our lab. In addition to always being willing to share her resources and time, Yingze Zhang offered invaluable help with designing and executing experiments, editing manuscripts, and making sure that I stayed on track to get all of those experiments done on a timeline that made sense for an MD/PhD student. She always made herself available to me to talk about work or life and I appreciate her friendship and mentorship. Frank Sciurba contributed significantly to the work presented here through the curation and management of the Pittsburgh SCCOR cohort and always serving as an expert on the physiology of COPD. Frank also gave me the great chance to meet individuals with COPD in his clinic, a unique opportunity to see some of the individuals that take part in the studies that contributed to this work. I'd also like to thank Yanxia Chu in Yingze Zhang's lab, Jo Anne Phillips and Chad Karoleski in the Emphysema COPD Research Center, Jacob Yundt in UPMC's Internet Services Division, and Melissa Saul in the Department of Biomedical Informatics for their contributions to the projects presented here.

Wendy Mars was my first rotation advisor, my thesis committee chair and my constant advocate and I am extremely grateful for the many times she made time in her schedule to talk about everything from experimental design to what was going on in my life. Thanks to Mike Barmada, for being a great professor and always responding with a quick answer whenever I wasn't quite sure where to go next with my genetics work. Yuri Nikiforov gave me great insight into my projects by offering the perspective of a clinician actually implementing clinical genetics, and I really appreciate that he was always confident in the findings of this work. I'd also like to thank Steve, Frank, Wendy, Mike and Yuri together for some really great committee meetings that almost always ran over because we had so much to discuss. Thanks to Joanne Flynn, my career

advisor, who has been a huge support in helping me think through each step in my training from the first year of medical school over numerous speed bumps to the point I'm at today. I greatly appreciate not only her help but her encouragement over the last 5 years. Thanks to the current and past administration of the Pitt-CMU MSTP, including Richard Steinman, Paul Monga, Clayton Wiley, Manjit Singh and Phuong Pham.

I can't capture all of the ways in which my mom, Debora Radder, has helped me get to this point in a few short sentences so I won't try and will instead thank her specifically for instilling in me the values of hard work and patience. As a family physician raising four kids (a job she always made look easy), she taught me that it's possible to do anything, but that you should always do it the right way, even when it's not the easiest way. Thanks to my sisters Lori, Chrissy and Jeannie, my brother- and sister-in-law Evan and Annie (and my nieces Rosie and Elsie), my friends Adam, Marshall, Sam, Bari, Alyce, Taylor, and all my other friends from the MSTP and medical school – having such a great support system makes the good times so much sweeter and the hard times so much easier.

Finally, I would like to dedicate this work to my father. Although he isn't here to see me defend this dissertation, his love for me and the educational philosophy that he spent his life sharing were hugely important in helping me reach this goal.

ABBREVIATIONS

AA	African American	IT	intratracheal
AAT	alpha 1-antitrypsin	LD	linkage disequilibrium
AATD	alpha 1-antitrypsin deficiency	LOD	log of the odds
ABI3BP	Abi binding protein, type 3	MAF	minor allele frequency
BAL	bronchoalveolar lavage	MME	macrophage elastase
BEOCOPD	Boston Early-Onset COPD Study	MMP12	matrix metalloproteinase 12
BICD1	bicaudal D homolog 1	MRI	magnetic resonance imaging
BPD	bronchopulmonary dysplasia	NAS/NETT	Normative Aging Study and National Emphysema Treatment Trial
bwa	Burrows-Wheeler alignment	NCBI	National Center for Biotechnology Information
CCDC38	coiled-coil domain containing 38	NGS	next generation sequencing
CDC	Centers for Disease Control and Prevention	NHANES	National Health and Nutrition Examination Survey
CFTR	cystic fibrosis transmembrane receptor	NHGRI	National Health and Genome Research Institute
CHRNA3/5	α-nicotinic acetylcholine receptor	NHW	non-Hispanic white
CO	carbon monoxide	NIEHS	National Institute of Environmental Health Sciences
COPD	Chronic obstructive pulmonary disease	PaO_2	partial pressure of oxygen in arterial blood
CSE	cigarette smoke extract	PBS	phosphate buffered saline
CT	computed tomography	PCA	principal component analysis
CVAS	common variant association studies	PCR	polymerase chain reaction
CYP2A6	cytochrome P450 2A6	PFT	pulmonary function testing
DLCO	diffusing capacity for carbon monoxide	PTPRO	protein tyrosine phosphatase receptor, type O
ECLIPSE	Evaluation of COPD Longitudinally to Identify Predictive Surrogate Endpoints	QQ	quantile-quantile
ECM	extracellular matrix	RV	residual volume
ELANE	neutrophil elastase	RVAS	rare variant association study
EMBL-EBI	European Molecular Biology Laboratory European Bioinformatics Institute	SCCOR	Specialized Center of Clinically Oriented Research

FAM13A	family with sequence similarity 13, member A	SERPINA1	serine peptidase inhibitor A1
$FEF_{25-75\%}$	forced expiratory flow at 25% to 75% forced vital capacity	SKAT-O	optimized sequence kernel association test
FEV1	forced expiratory volume in one second	SNPs	single nucleotide polymorphisms
FRC	functional residual capacity	STR	short tandem repeat
FVC	forced vital capacity	TCGS	Transcontinental COPD Genetics Study
GATK	genome analysis toolkit	TERT	telomerase reverse transcriptase
GEMMA	genome-wide efficient mixed-model analysis	TLC	total lung capacity
GOLD	Global Initiative for Chronic Obstructive Lung Disease	TOPMed	Trans-Omics for Precision Medicine
GWAS	genome-wide association study	VEGF	vascular endothelial growth factor
HHIP	hedgehog interacting protein	WES	whole exome sequencing
HU	Hounsfield Units	WGS	whole genome sequencing
HU	fraction of voxels less than −950 Hounsfield Units	WHO	World Health Organization
IC	inspiratory capacity	ZNF816	zinc finger protein 816
IL1R2	interleukin-1 receptor, type 2		

1.0 INTRODUCTION

Chronic obstructive pulmonary disease (COPD) is a major cause of worldwide morbidity and mortality (1-3). Defined as airflow limitation that is progressive and irreversible, the World Health Organization (WHO) estimates that the disease is the world's third leading cause of mortality, responsible for over 5% of deaths (4). This number is predicted to rise in the next ten years largely due to increased cigarette smoking, the primary environmental determinant of the disease, in the developing world (1, 4). Treatment of COPD imposes a significant burden on the health care system with costs estimated at 50 billion dollars in 2010 (5-7).

While exposure to cigarette smoke is the most common environmental factor contributing to COPD, less than half of heavy smokers ever develop disease, demonstrating the complex interaction of environment and genetic susceptibility in disease pathogenesis (3, 8). Although a number of susceptibility loci have been identified in association studies, together they only account for a small portion of the heritability of COPD (9). This work employs modern genetic approaches to identify genomic loci and candidate genes contributing to susceptibility to COPD and offers support for their functional role in disease.

1.1 CHRONIC OBSTRUCTIVE PULMONARY DISEASE

1.1.1 Pathophysiology

According to the Global Initiative for Chronic Obstructive Lung Disease (GOLD), COPD "is characterized by persistent airflow limitation that is usually progressive and associated with an enhanced chronic inflammatory response in the airways and the lung to noxious particles or gases" (10). Cigarette smoking is the most common cause of exposure to noxious particles in the developed world, although biomass fuel smoke exposure has a significant impact in the developing world (11, 12). Three pathologic processes contribute to the airflow limitation characteristic of COPD: chronic bronchitis, emphysema and small airway disease and fibrosis (13). Chronic bronchitis describes hypersecretion of mucus, an abundance of inflammatory exudate, and ultimately remodeling in the upper airways, leading to both airway obstruction and a chronic productive cough that is a common symptom of COPD (14). Emphysema involves the destruction of the lung parenchyma leading to altered gas exchange and a loss of elastic recoil in the lung (14). The small conducting airways are also a major site of obstruction following fibrotic remodeling (14). Vascular changes resulting from long-term gas exchange abnormalities can result in pulmonary hypertension and right ventricular hypertrophy late in the disease process (15). Finally, systemic inflammation (see 1.1.2) contributes to numerous comorbidities including skeletal muscle wasting, cardiovascular disease, anemia, osteoporosis, depression, and lung cancer (16).

1.1.2 Pathogenesis

The GOLD definition of COPD (section 1.1.1) describes a chronic dysregulated inflammatory response that is a hallmark of the disease process. The acute response to cigarette smoke largely involves induction of innate immune cells such as macrophages and neutrophils that release pro-inflammatory chemokines, cytokines and growth factors (14). Macrophages are predominant in the process and contribute to the breakdown of elastin, a structural protein that is a primary contributor to the lung's elasticity, via secretion of metalloproteinases (see below) (17). As the response to cigarette smoke exposure becomes chronic, the adaptive immune system is also recruited, particularly CD8+ T-cells (18). T-cells and B-cells organize in lymphoid follicles in the small airways that are unique to individuals with COPD (19). Macrophage polarity is altered to a pro-inflammatory M1 state following chronic cigarette smoke exposure, and combined with the adaptive response contributes to the dysregulated immune state that persists even after smoking cessation in many individuals (20). Lung epithelial cells also play a major role in the inflammatory response through production of inflammatory mediators and altered mucin production (21, 22).

Inflammation in COPD contributes to the breakdown of essential structural proteins in the lung altering the local balance of proteases and anti-proteases. Our understanding of this mechanism was established through studies in individuals with a congenital deficiency of alpha-1 antitrypsin (AAT), a potent inhibitor of serine proteases, who also develop early-onset and severe emphysema (see section 1.2.1 for more detail on the genetic inheritance of AAT deficiency) (13, 23, 24). Animal studies show that instillation of elastolytic enzymes into the rat lung induce emphysema (25, 26). Further, gene targeted null mutant mice for major inflammatory elastases

such as *MMP12*, which encodes macrophage elastase, and *ELANE*, which encodes neutrophil elastase, are protected from cigarette smoke-induced emphysema (27, 28).

Oxidative stress in COPD is related to both exogenous and endogenous sources of oxidants. Cigarette smoke generates more than 10^{15} oxidants per inhalation (29). Immune cells, particularly macrophages, produce reactive oxygen species (ROS) and reactive nitrogen species (RNS) in response to inflammatory stimuli via the NADPH oxidase system (30). Individuals with COPD have a lower level of antioxidants to combat this oxidative stress, and this can continue chronically after cigarette smoke exposure (31). An imbalance of oxidants and antioxidants results in oxidative damage to DNA and proteins, causing disruption of numerous downstream processes; particularly significant in COPD is the inactivation of metalloproteinases (32, 33). Oxidative species can also lead to recruitment of inflammatory cells, causing a cycle of inflammation and oxidative damage (31).

Long-term inflammation and oxidative damage may result in apoptosis and senescence of both immune cells and the alveolar epithelial cells of the lung (34). Blocking vascular endothelial growth factor (VEGF) receptors induces apoptosis, or programmed cell death, of alveolar epithelial cells in mice, an animal model of emphysema that appears to be dependent on oxidative stress and independent of inflammation (35). In humans, the picture is more complicated - VEGF and VEGFR are lower in individuals with emphysema but CD8+ T-cells are also associated with markers of apoptosis (36, 37). Markers of cellular senescence are increased in individuals with COPD and, in combination with apoptosis, may contribute to cellular loss in the alveolar epithelium (38). Although it remains unclear whether apoptosis occurs in response to loss of extracellular matrix, particularly since the VEGFR model of emphysema also results in elastin damage, or whether independent mechanisms triggering apoptosis also lead to disease, it seems

probable that environmental or genetic factors that increase apoptosis or senescence exacerbate disease severity and progression (34).

Each of the pathogenetic mechanisms described to this point may also contribute to comorbidities of COPD, due to their systemic effects. Systemic inflammation occurring from chronic cigarette smoke exposure results in cardiovascular complications, metabolic dysregulation, osteoporosis and anemia (39). Systemic inflammation also increases endogenous oxidative stress throughout the body in addition to a 'spill over' of oxidative stress from the lung (30). Oxidative stress can contribute to systemic apoptosis – weight loss in COPD is correlated with apoptosis of skeletal muscle (40).

Finally, while the majority of studies looking at the pathogenesis of COPD focus on the effects of environmental exposures in adult life, developmental factors also contribute to disease (8). Studies of the natural history of COPD show that a portion of the population have reduced lung growth in adulthood prior to FEV1 decline and that this predisposes to COPD later in life (41, 42). Prenatal exposures (such as maternal smoking) as well as postnatal exposures (such as childhood respiratory infections) are associated with decreased lung function in adulthood (43, 44). Further, birth weight not only correlates with FEV1 measured in adults but also with mortality from COPD, suggesting that early life factors directly contribute to disease outcomes (45). Thus, prenatal and early life factors negatively affect lung development or growth resulting in lower lung function in adulthood which increases disease susceptibility (46).

1.1.3 Clinical characteristics

Symptoms of COPD reflect both the airflow limitation that defines the disease and the pathophysiology leading to this limitation (section 1.1.1). These include dyspnea made worse with physical exertion, chronic cough with mucus production, chest tightness and wheezing, fatigue, and susceptibility to respiratory infections (10). Exacerbations of COPD involve acute worsening of these symptoms, particularly dyspnea, sputum production, and cough (47). Patients with similar spirometric presentations may have varying symptoms and predispositions to exacerbations, with "frequent exacerbators" suffering from increased morbidity and consuming much greater portions of healthcare resources (section 1.1.4) (48, 49). The systemic effects of COPD, described in section 1.1.2, result in numerous comorbidities including musculoskeletal dysfunction, osteoporosis, diabetes mellitus, and lung cancer (50).

Airflow limitation in COPD is measured using spirometry, the primary component of pulmonary function testing (PFT). Spirometry is recommended for screening of airflow obstruction in individuals with symptoms of COPD, but is not recommended for broader screening (10, 51). A spirometer measures inhalation and exhalation volumes as a function of time, allowing for the calculation of numerous metrics of lung function. Essential spirometry measurements in COPD include the forced expiratory volume in one second (FEV1), the volume of air that can be exhaled in one second of a forced expiratory maneuver after taking a full breath, and the forced vital capacity (FVC), the total volume of air that can be exhaled in a forced maneuver after taking a full breath. When measured following treatment with a bronchodilator to detect reversibility (see section 1.1.3), a ratio of FEV1/FVC less than 70% "confirms the presence of persistent airflow limitation and thus of COPD" (10). Severity of COPD is described in terms of GOLD scores, a

metric that classifies individuals based on the extent of their obstruction as measured by FEV1 (Table 1).

Table 1. GOLD scoring of severity in individuals with COPD

GOLD 1	Mild	Percent Predicted FEV1 > 80%
GOLD 2	Moderate	50% < Percent Predicted FEV1 < 80%
GOLD 3	Severe	30% < Percent Predicted FEV1 < 50%
GOLD 4	Very Severe	Percent Predicted FEV1 < 30%

Percent predicted FEV1 calculated following bronchodilator administration. Adapted from (10).

A number of other clinical measures can be used to clarify the clinical phenotype of COPD. Body plethysmography measures lung volumes, including total lung capacity (TLC), inspiratory capacity (IC), functional residual capacity (FRC), and residual volume (RV) which can be used to determine whether a reduction in FVC is due to hyperinflation or air trapping. The diffusing capacity for carbon monoxide (DLCO) measures gas exchange abnormalities in COPD and can aid in assessing extent of lower airspace loss but is not highly sensitive (52). Computed tomography (CT) is more efficient for this purpose and may allow for early detection of emphysema, although the clinical implications of quantitative measures of lung densities using these approaches still require clarification in large clinical trials (53, 54). CT is used clinically to describe patterns of emphysema that represent known etiologies of disease (see AATD in section 1.2.1) or to identify candidates for specific types of surgical interventions (see lung reduction surgery in section 1.1.5) and it is likely that CT and magnetic resonance imaging (MRI) can provide valuable information about the pathophysiology of an individual's disease (55).

1.1.4 Burden

Given the complex clinical presentation of COPD described in section 1.1.3, it is not surprising that estimates of disease prevalence vary depending on how it is defined. In the United States, in the National Health Interview Survey conducted by the Centers for Disease Control and Prevention (CDC), approximately 12 million Americans were estimated to have COPD, representing a prevalence of approximately 3.8% (56). In the National Health and Nutrition Examination Survey (NHANES), the post-bronchodilator prevalence of COPD defined by the fixed ratio of FEV1/FVC was 14.0% (57). These findings are reflected on a global scale, where self-reported prevalence of COPD was 3.7% and prevalence of COPD as measured by spirometry was 10.1% in a large meta-analysis (58). The worldwide prevalence of emphysema as measured by radiography was 3.2% in this study (59). The absence of major studies in much of the world, particularly Africa, makes comparison of prevalence between countries difficult but there is likely a significant rate of underdiagnosis of COPD worldwide (60).

The cost of COPD to both society and patients is significant. In the United States in 2010, there were approximately 10.3 million physician office visits, 1.5 million emergency room visits, and 699,000 hospitalizations with COPD recorded as the primary diagnosis (56). This extensive utilization of the healthcare system cost approximately $30 billion dollars in direct health care costs and $20 billion dollars in indirect costs, primarily due to the COPD population's reduced ability to work (7). Exacerbations are particularly expensive and are estimated to cost an average of $7,100 for each exacerbation-related hospital admission (61). Patient quality of life is also directly related to exacerbation frequency (62). Health-related quality of life progressively

decreases as severity of COPD increases (63). Thus, individuals suffering the greatest morbidity also have the largest healthcare utilization.

In addition to the significant morbidity and economic cost of COPD, it is considered the third leading cause of death in the world by the WHO, with an estimated 3.1 million deaths attributed to the disease in 2012 (4). This represents a rate of approximately 44.2/100,000 population, with the US rate of 42 deaths/100,000 population reported in 2002 relatively representative, although there are regions of the world with significantly higher mortality due to COPD (64). Mortality has been shown to significantly correlate with cigarette consumption and poverty (65). While the best-proven approach to improving survival in COPD is smoking cessation, in patients with severe disease, oxygen therapy and lung volume reduction therapy have been shown to reduce mortality (see section 1.1.5) (64).

1.1.5 Treatments

Although no treatment has been found to reverse COPD, smoking cessation has the greatest impact on survivability (see section 1.1.4). For the same reason, pharmacologic therapies are primarily designed to manage symptoms and transiently reduce airflow limitation. Bronchodilators are the most commonly used drug for managing airflow obstruction and act by modulating airway smooth muscle tone in order to improve flow through the upper airways. A number of drugs with different mechanisms including $beta_2$-agonists, anticholinergics, methylxanthines, phosphodiesterase-4 inhibitors and corticosteroids are used alone or in combination for their bronchodilatory effects (10). GOLD recommendations include combining spirometric classification of severity, number

of exacerbations and the extent of symptoms to determine which bronchodilator or combination of drugs should be used for therapy (10).

A number of non-pharmacologic therapies can also be useful in COPD, particularly in severe disease. Pulmonary rehabilitation including exercise training for both the lower and upper extremities, nutrition counseling and education about the disease, has been shown to improve quality of life and decrease dyspnea (66). Oxygen therapy can reduce symptoms and improve exercise tolerability in individuals with hypoxemia, measured by the partial pressure of oxygen in arterial blood (PaO_2) or oxygen saturation (67). Exacerbations require the use of oxygen therapy, bronchodilators, corticosteroids, and potentially antibiotics and noninvasive mechanical ventilation (10).

In severe cases of emphysema resulting in significant amounts of air trapping and hyperinflation, patients may benefit from lung volume reduction surgery, although it carries significant risks (10). In order to reduce these risks a number of minimally invasive approaches have been developed. Recently, one of these approaches – the placement of endobronchial coils to induce volume reduction and restore lung recoil – has been shown to improve exercise capacity, although longer follow up studies will be necessary (68, 69). Finally, in very severe cases, lung transplantation may be used to improve lung function and quality of life (10).

Clearly, current treatment paradigms for COPD take very little of the clinical variability described in section 1.1.3 into account, with the majority of pharmacologic and surgical therapies aiming to reduce airflow obstruction (70). GOLD recommendations for the treatment of COPD separate patients according to a combination of spirometric classification and frequency of exacerbations for increased symptomatic relief, but also mention that “comorbid conditions may influence mortality and hospitalizations, and should be looked for routinely and treated

appropriately" (10). Thus, current approaches to treatment of COPD are generally untargeted and treat individual symptoms independently of one another, even though, as described in section 1.1.2, underlying mechanisms may contribute to more than one aspect of disease expression (71).

1.2 GENETICS OF COPD

1.2.1 Mendelian inheritance

The first known genetic determinant of COPD was described by Laurell and Eriksson in 1963 when they observed the absence of the α-1 protein in serum protein electrophoreses from five individuals, three of whom also had early-onset emphysema (24). They described four variants in the protein α-1 antitrypsin (AAT) according to their movement in gel electrophoresis, F (fast), M (medium), S (slow), and Z (very slow) (72). Population studies demonstrated that the M variant was wild type and the Z variant was significantly associated with disease and deficiency of AAT (73). Individuals with PiZZ phenotype are most susceptible to early onset disease and have approximately 10% normal levels of AAT. Individuals with other variant genotypes, most commonly PiMZ, also have moderately decreased levels of AAT, but the clinical implications of this decrease remain unclear (74). By studying the functional effects of the Z variant, AAT was found to be a serine protease inhibitor, leading to the naming of the gene: serine peptidase inhibitor A1 (*SERPINA1*) (75, 76). AAT is a particularly potent inhibitor of elastase and when it is deficient the protease is uninhibited in its destruction of elastin fibers in the lung, which are not regenerated in adulthood (77, 78). The protein also appears to play a more nuanced role in immune regulation,

which may contribute to a lung phenotype even in non-smokers with AATD (79). Augmentation of AAT may slow deterioration of lung function and correct systemic effects in AATD (80).

Mutations in elastin (*ELN*) and other genes contributing to elastin fiber structure such as *FBLN5* result in cutis laxa and some individuals with this condition develop early-onset emphysema (81). The observation that deleterious variants in telomerase reverse transcriptase (*TERT*) occur in approximately 1% of individuals with severe emphysema has suggested that this may be another monogenic cause of COPD, but these findings require replication in larger cohorts (82).

1.2.2 Early genetic epidemiology

In a foundational study of the natural history of COPD, Fletcher and Peto observed that only 30% of current or former smokers and half of heavy smokers (greater than 15 cigarettes/day) develop COPD (8). The authors suggested for the first time that this difference may be due to heritability outside of AATD : 'Is susceptibility in any way analogous to α1-antitrypsin deficiency or due to quantitative differences in leucocyte proteolytic enzymes?' (83) Familial aggregation studies supported this hypothesis, indicating pulmonary function and chronic bronchitis occur more frequently in related individuals (84, 85). Twin studies showed that pulmonary function deficits are much more common in monozygotic than dizygotic twins even after accounting for environmental exposures and differences in body habitus that could contribute to FEV1 (86, 87). Finally, complex segregation analysis suggested that COPD was most likely due to multifactorial inheritance (88, 89).

1.2.3 Linkage analyses

Genetic linkage analysis was one of the first tools used to map the location of trait-related loci on the human genome. The technique relies on the knowledge that the closer together two genetic markers are located, the more likely they are to remain together during meiosis (90). In early practice, the goal of linkage analyses was to identify the genomic location of disease-causing genes by testing for co-segregation of the disease with markers of known genomic location in families. One of the best-recognized examples of this type of study was the identification of a marker on chromosome 7 linked to cystic fibrosis, leading to the characterization of *CFTR* (91). Eventually, it became possible to test for linkage across the genome by genotyping markers spaced at regular, known genetic distances and testing for the expected number of recombinations in affected families between each marker (92). The log of the odds (LOD) score is a ratio of the expected number of recombinations given the null hypothesis and the observed number of recombinations, and is the primary measure of linkage with a trait. The majority of these studies relied on short tandem repeat (STR) polymorphisms that were identified with polymerase chain reaction (PCR) and thus the number of genotypes that could be measured per individual was technologically limited.

The first linkage analysis in COPD looked at 72 families of individuals with severe, early-onset COPD in the Boston Early-Onset COPD Study (BEOCOPD). Using non-parametric linkage analysis, the authors identified suggestive linkage with moderate airflow obstruction (FEV1 < 60% predicted, FEV1/FVC < 90% predicted) at chromosomes 12 and 19, and with mild airflow obstruction (FEV1 < 80% predicted, FEV1/FVC < 90% predicted) at chromosomes 8 and 19 (93). Restricting their findings to members of the families who smoked supported evidence for linkage

at chromosome 12p (93). A study looking at the same families but also including extended relatives, tested for linkage with spirometric traits of COPD and demonstrated evidence for linkage with FEV1/FVC on chromosomes 1, 2, and 17, with FVC on 1, and FEV1 at 12p (94). A final study of this cohort looked at additional spirometric traits and demonstrated linkage with $FEF_{25-75\%}$ and $FEF_{25-75\%}$/FVC on chromosomes 2q and 12p and confirmed linkage of FEV1 at chromosome 12p (95). The results of major linkage analyses of COPD and COPD-related traits can be seen in Table 2.

Table 2. Linkage studies of COPD

Study	Locus	LOD	Phenotype
Joost et al, *AJRCCM*, 2002 (96)	6q terminus	2.4	FEV1 general
	21p terminus	2.6	FVC general
Silverman et al, *Hum Mol Genet*, 2002 (93)	12p12	2.09	moderate airflow obstruction
	12p12		mild airflow obstruction
Silverman et al, *AJHG*, 2002 (94)	2q34	2.99	FEV1/FVC
	17q21	2.68	FEV1/FVC
	12p12	2.88	FEV1 in COPD
	12p12	2.68	FEV1/FVC
	1p36	2.05	FVC
Wilk et al, *AJRCCM*, 2003 (97)	4p15	3.5	FEV1/FVC general
	18p11	2.4	FEV1 general
	18q21	2.9	FVC
Palmer et al, *Hum Mol Genet*, 2003 (98)	8p23	3.3	pbFEV1
	1p21	2.24	pbFEV2
	8q24	2.01	pbFEV3
	2q35	4.42	pbFEV1/FVC
	1p31	2.52	pbFEV1/FVC
	17q21	2.44	pbFEV1/FVC
DeMeo et al, *AJRCCM*, 2004 (95)	2q35	5.03	FEF25-75, smokers
	12p12	3.46	FEF25-75/FVC, smokers

1.2.4 Candidate gene association studies

While linkage analysis was successful at identifying loci for many monogenic traits, in diseases with multifactorial inheritance, results were not definitive and their resolution was relatively low. For example, only a few of the results in Table 2 reached genome-wide significance (LOD score > 3), and no candidate genes in any of these regions are supported by significant evidence (see section 1.2.6) (99-101). In 1996, Risch and Merikangas compared linkage mapping to tests of association, in which measured allelic or genotypic frequencies are statistically compared between

two groups with different traits, and demonstrated that for loci with a moderate effect on a phenotype occurring frequently in the population, association is significantly more effective than linkage (102). At that time, however, the Human Genome Project was still in progress, the number of known polymorphisms in the genome was limited, and no technology existed to genotype polymorphisms on a large scale. Thus, many early association studies chose to test for association with markers in genes or groups of genes with known biological function (103).

Numerous candidate gene association studies were conducted in COPD, although many failed to identify significant associations (100, 104-106). Two studies attempting to identify candidate genes for loci at 12p and 2q, identified in linkage studies of spirometric traits of COPD (see Table 2), suggested roles for the genes *SOX5* and *XXCR5* (100, 101). Associations with the 'Z' allele in *SERPINA1*, previously identified in biochemical studies (see section 1.2.1), were observed and novel associations with *IL1RN* were identified in another study (107). In a meta-analysis of COPD candidate gene association studies, variation in *TGFB1*, *IL1RN*, *TNFA*, and *GSTP1* were significantly associated with COPD in several populations (108).

1.2.5 Common variant association studies

The first successful genome-wide association study (GWAS) was conducted in 2002, testing for association of approximately 92,000 common (minor allele frequency (MAF) > 5%) single nucleotide polymorphisms (SNPs) with occurrence of myocardial infarction (109). In 2014, the National Health and Genome Research Institute (NHGRI) published the first description of their GWAS Catalog containing the results of 1,751 studies looking at over 700 traits, indicating the rapid and massive growth of this technique in the ensuing years (110). The majority of GWAS to-

date have looked at association of common variants with traits, and this type of study will be referred to as a common variant association study (CVAS) from this point.

CVAS have been conducted to study the genetic contribution of numerous COPD related-traits. The first CVAS in COPD used a case-control study design in a Norwegian population and identified a genome-wide significant association at 15q25 near the gene *CHRNA3/5* and a genome-wide suggestive association at 4q31 near *HHIP* that was replicated in an independent cohort (111). *CHRNA3/5*, encoding the α-nicotinic acetylcholine receptor, appeared to be an obvious smoking-related gene and *HHIP* was concurrently shown to be associated with FEV1 in the general population, offering significant promise for the findings of future studies (112). In a joint study of Europeans and non-Hispanic white (NHW) Americans with COPD, Cho et al described an additional association with *FAM13A* (113). A larger GWAS looking at individuals from the Evaluation of COPD Longitudinally to Identify Predictive Surrogate Endpoints (ECLIPSE) cohort and replicated in the Normative Aging Study and National Emphysema Treatment Trial (NAS/NETT), Bergen, Norway (GenKOLS) and COPDGene studies, identified a novel locus associated with COPD at 19q13 near the genes *RAB4B*, *EGLN2*, *MIA*, and *CYP2A6* (114). Looking at the same cohorts in a meta-analysis, association with *MMP12* and *TGFB2* was observed (115).

Poorly defined phenotypes significantly decrease power in association studies, with phenotypic heterogeneity of 50% increasing the sample size needed to detect an associated variant threefold (116). A recent study looking at individuals scored as GOLD 1, frequently used in genetic studies to define ‘mild obstruction’, indicated significant phenotypic heterogeneity of this trait, reflecting the diverse clinical presentation of COPD described in section 1.1.2 (117). In an attempt to overcome this problem, CVAS have been designed to test for association with more well-defined phenotypes of COPD. The first GWAS of emphysema identified association with the gene

BICD1, a locus that had not previously been detected in studies testing for association with COPD in the same cohorts (118). Recently, a study testing for association with distinct patterns of emphysema by CT found that different loci reported in CVAS of COPD are associated with unique histological patterns of emphysema (119).

CVAS have also been conducted looking at COPD-related traits. Association with pulmonary function in the general population has been studied in extremely large populations, successfully identifying numerous loci, some of which also associate with COPD (see Table 3) (112, 120-123). CVAS looking at the genetics of smoking habits, including traits such as cigarettes smoked per day, pack-years and exhaled carbon monoxide (CO), have repeatedly shown association with the locus near *CHRNA3/5*, supporting the hypothesis that this association with COPD is due to smoking behavior (124, 125). Other studies have shown unique loci contributing to other COPD-relevant phenotypes such as response to bronchodilators and body composition (126, 127).

Table 3. CVAS in COPD and COPD-related phenotypes

Study	Phenotype	Ethnicity	Significant genes/loci	Suggestive genes/loci
Wilk JB et al, *BMC Med Genet*, 2007 (120)	FEV1/FVC	NHW		*GSTO2, IL6R*
Pillai SG et al, *PLoS Genet*, 2009 (111)	COPD	European	*CHRNA 3/5*	*HHIP*
Wilk JB et al, *PLoS Genet*, 2009 (112)	FEV1/FVC	NHW	4q31	
Repapi E et al, *Nat Genet*, 2010 (128)	FEV1/FVC	European	4q31 (*HHIP*) 2q35(*TNS1*) 4q24(*GSTCD*) 5q33(*HTR4*) 6p21(*AGER*) 15q23(*THSD4*)	
Cho MH et al, *Nat Genet*, 2010 (113)	COPD	NHW/European	4q22 (*FAM13A*) *CHRNA3/5*	
Pillai SG et al, *AJRCCM*, 2010 (129)	Pack Years, visual emphysema score, obstruction	NHW	*CHRNA3/5*	
	FEV1 in smokers		*IREB2*	
	FEV1/FVC in smokers, fat-free body mass, exacerbations		*HHIP*	
	FEV1, FEV1/FVC in smokers		*FAM13A*	
Hancock DB et al, *Nat Genet*, 2010 (123)	FEV1, FEV1/FVC	European	*HHIP, GPR126, ADAM19, AGER-PPT2, FAM13A, PTCH1, PID1, HTR4, INTS12-GSTCD-NPNT*	
Wan ES et al, *AJRCCM*, 2011 (126)	fat-free mass index in COPD	NHW/European	*FTO*	
Siedlinski et al, *Thorax*, 2011 (124)	cigarettes smoked per day in GOLD >2		*CYP2A6*	

Table 3 continued

Soler Artigas M et al, *Nat Genet*, 2011 (121)	FEV1, FEV1/FVC	European	*MFAP2, TGFB2, HDAC4, RARB, MECOM (EVI1), SPATA9, ARMC2, NCR3, ZKSCAN3, CDC123, C10orf11, LRP1, CCDC38, MMP15, CFDP1, KCNE2*	
Cho MH et al, *Hum Mol Genet*, 2012 (114)	COPD	NHW/European	*CYP2A6*	
Hardin M et al, *AJRCCM*, 2012 (130)	severe to very severe COPD	Polish	*CHRNA3/5, IREB2, ADCY2*	
Wilk JB et al, *AJRCCM*, 2012 (131)	airflow obstruction	European	*CHRNA3/5/IREB 2, HTR4*	
Hancock DB et al, *PLoS Genet*, 2012 (122)	FEV1, FEV1/FVC	multiple	*DNER, HLA-DQB/HLA-DQA2, KCNJ2/SOX9*	
Hansel NN et al, *Hum Genet*, 2013 (132)	mild to moderate COPD	European	*TMEM26, FOXA1, ANK3*	
Yao TC et al, *J Allergy Clin Immunol*, 2014 (133)	FEV1, FVC, FEV1/FVC	Hutterites	*THSD4-UACA-TLE3*	
Manichaikul A et al, *AJRCCM*, 2014 (134)	percent emphysema on CT	multiple	*SNRPF, PPT2, MAN2B1, DHX15, MGAT5B, MAN1C1*	
Tang W et al, *PLoS One*, 2014 (135)	Rate of change of FEV1	European	*IL16/STARD5/T MC3*	
Castaldi et al, *AJRCCM*, 2014 (119)	local histogram emphysema patterns	NHW	*MYO1D, VMA8, HHIP, IREB2/CHRNA3, CYP2A6, TGFB2, MMP12*	
Bloom AJ et al, *Ann Am Thorac Soc*, 2014 (125)	Exhaled CO, cigarettes smoked per day	NHW, AA	*CHRNA3/5/B4*	

Table 3 continued

Lee JH et al, *Respir Res*, 2014 (136)	total lung capacity in individuals with COPD	NHW/European	*DNAH5*	
McDonald ML et al, *AJRCMB*, 2014 (137)	resting oxygenation in COPD	AA, NHW		*SPO2*, *KIF7*
Cho MH et al, *Lancet Respir Med*, 2014 (115)	moderate to severe COPD	AA, NHW	*CHRNA3*, *HHIP*, *RIN3*, *MMP12*, *TGFB2*	
Hansel NN et al, *AJRCMB*, 2015 (138)	airway responsiveness in COPD	NHW	*LINGO2*	
Cho MH et al, *AJRCCM*, 2015 (139)	airway quantitative imaging phenotypes, gas trapping	NHW/European	*HHIP*, 15q25, *AGER*, *SERPINA10*, *DLC1*	
Hardin M et al, *Pharmacogenomics J*, 2015 (127)	response to bronchodilators	AA, NHW		*KCNK1*, *KCNJ2*
Lutz SM et al, *BMC Genet*, 2015 (140)	post-bronchodilator FEV1, FEV1/FVC in smokers	AA, NHW	15q25, *HHIP*, *TGFB2*, *DBH*, *CYP2A5*, *FAM13A*, *MMP12*, *RIN3*	
Wain LV et al, *Lancet Respir Med*, 2015 (141)	extreme traits of FEV1 in smokers and never smokers	British	*KANSL1*, *TSEN54*, *TET2*, *RBM19/TBX5*, *PNT*, *HLA-DQB1/HLA-DQA2*	

1.2.6 Identification of causal variants and genes

The purpose of identifying genomic loci that contribute to disease is two-fold. First, knowledge of susceptibility loci aids in prediction of future disease in an individual, simply through the

measurement of genotype. Second, these loci may tell us something about the biology of the disease process. Although the first purpose does not require identification of a functional effect of variation, the second does, and follow-up studies have been conducted to identify causal variants of linkage mapping and CVAS (142). In some diseases, these studies have been extremely successful. For example, IL-23 signaling was identified as a major pathway contributing to inflammatory bowel disease following a CVAS identifying associations between *IL23R* and Crohn's disease (143, 144). However, for many loci identified by linkage analyses or CVAS, similar findings are lacking (142).

In COPD, the results of follow-up studies to identify causal variants have been mixed. Using a fine-mapping approach, Hersh et al. tested for association of variants in two candidate genes at a locus in linkage with FEV1 at 8p using the BEOCOPD population but failed to offer convincing evidence for either gene (145). The gene *SERPINE2* located at the chromosome 2q locus in linkage with FEV1 was identified as a possible candidate gene and this finding was supported by a number of fine-mapping association studies (105, 106, 146). However, there is support for the gene *XCCR5* at the same locus, bringing into question the role of *SERPINE2* or suggesting that more than one gene may contribute to disease at these loci (101).

A number of follow-up studies have looked at the function of genes identified in CVAS of COPD. SNPs associated with pulmonary function in the general population alter regulatory motifs near the gene *HTR4*, and mice lacking this gene have altered baseline pulmonary function (147, 148). Similar effects on regulatory function have been shown for the gene *HHIP* and variants in LD with a commonly associated SNP have been shown to be in an enhancer region of the gene (149). *Hhip*$^{-/-}$ mice die at birth from respiratory failure while mice that are haploinsufficient for the gene have increased susceptibility to cigarette smoke-induced emphysema (150). Further,

Hhip$^{+/-}$ mice are predisposed to age-related emphysema suggesting that this gene contributes to both development and disease (151).

Despite some success at identifying the biological relevance of loci identified in genetic studies of COPD, the majority of our knowledge of the pathogenesis of the disease is due to biochemical and molecular studies (see section 1.1.2). Neither candidate gene nor genome-wide association studies have identified variation associated with COPD in most of these genes (see sections 1.2.4 and 1.2.5). A notable exception to this is *MMP12*, identified in murine studies (see section 1.1.2) nearly 20 years before a meta-analysis of CVASs identified nearby variants yielding genome-wide significant association with COPD (27, 115).

1.2.7 Genetic architecture of complex traits

In an attempt to clarify both the biological relevance of genomic loci contributing to disease pathogenesis (see section 1.2.6) and to better predict and model disease risk, linkage analyses (section 1.2.3) and CVAS (section 1.2.4) have aimed to identify susceptibility loci for chronic diseases. In COPD, these studies have successfully identified frequently replicated loci that clearly alter disease risk (Table 2 and Table 3). Yet, a recent study demonstrated that SNPs at the four best-supported loci in COPD, near the genes *HHIP*, *CHRNA3/5/IREB2*, *CYP2A6*, and *FAM13A*, account for only 8% of the heritability, or the variation in the trait due to genetic variation, of FEV1 and FEV1/FVC in individuals with COPD (9). This finding reflects what is seen in other chronic diseases – in one estimate, the median heritability explained by statistically significant CVAS results for ten traits including psychiatric, autoimmune and metabolic disorders, was 9.8% (152). Thus, despite the significant resources and effort that have gone into conducting CVAS

across thousands of complex traits, a ubiquitous 'missing heritability' problem suggests additional genetic variation remains to be found (153, 154).

The surge of CVAS in chronic disease was based on the 'common disease, common variant' hypothesis, first introduced with knowledge that linkage analyses had been mostly unsuccessful at identifying disease-contributing loci for complex traits (102). This hypothesis posits that there are likely to be several contributing loci for a given complex trait, each with a small to moderate contribution to disease risk. Since each locus individually has a relatively small effect on organismal fitness, these variants would not experience significant evolutionary selection and would be common in the population (155). So, the 'missing heritability problem' in chronic diseases is not an inability of CVASs to detect true common variants associated with disease, as much as inflated expectations that they would identify the entire heritable spectrum of complex diseases based on an over-simplified model of their underlying genetic architecture (153, 156, 157).

One hypothesized model to account for 'missing heritability' in complex traits assumes that an extremely large number of common variants, each with a miniscule effect on genetic risk, act in concert to result in disease (153, 158). Under this model, only a portion of heritability can ever be detected by CVAS, as variants with an extremely small effect size would require sample sizes larger than the entire population to reach multiple testing correction thresholds for genome-wide significance (158). This theory has been partially supported in the study of genetic contributions to height, which is highly heritable (159). Nearly 400 loci have been identified contributing to height, but together they only account for a portion of its genetic variance (160). Calculating heritability using the contribution of all SNPs in a GWAS of height simultaneously, however, explained over 60% of the variance of the trait (160, 161). Importantly, the authors of

that study conclude that there are a finite number of variants contributing to genetic variance – an important argument that the majority of meaningful trait-related variants can be identified. While it is unknown to what extent this model applies to disease-related traits, accounting for variation that doesn't reach genome-wide significance can also explain additional heritability in disease (158).

Another explanation for the 'missing heritability' in complex disease is that rare variants contribute to disease susceptibility (162). Rare variant studies have recently become technologically possible due to the reduction in cost of next generation sequencing (NGS) technologies including targeted detection of all variants in coding regions of the genome, or whole exome sequencing (WES), and detection of all variants in the genome by whole genome sequencing (WGS). Using these techniques, recent studies have identified rare variants associated with common traits such as triglyceride levels and red blood cell characteristics (163, 164). Similar studies have observed rare variants associated with numerous chronic diseases including type 2 diabetes and schizophrenia (165, 166). Rare variants may even be responsible for some signals detected in CVAS due to the generation of a 'synthetic association', unlikely according to statistical analyses, but as yet unproven experimentally (167-170).

The massively parallel sequencing techniques used in NGS studies also allow for comparisons of read depth in order to detect large structural and copy number variants in the genome. Although few studies have investigated the role of these variants in chronic disease, they are likely to contribute to their genetic architecture (171, 172). Genetic studies of psychiatric disorders have been particularly successful at identifying copy number variants and large translocations that contribute to disease (173).

In addition to the contribution of variants that cannot be identified in classic CVAS, there are likely to be underlying genetic interactions that explain the 'missing heritability' problem. Statistical and experimental approaches show that a portion of the heritability of complex diseases can be attributed to epistasis, or gene-gene interactions (174, 175). Although the role of specific environmental exposures such as cigarette smoke in COPD may be well understood, the impact of other environmental factors remains unclear, and unrecognized gene-environment interactions likely contribute to disease risk (176, 177). Epigenetic factors also play a role in the heredity of complex diseases, but the extent of this contribution and their precise mechanisms of action require further exploration (178).

Most likely, the genetic architecture underlying chronic diseases like COPD is exactly as described: complex. Numerous genes are likely to be affected by a combination of rare, common and structural variants, each potentially interacting with other genes or affected by environmental exposures. Moreover, the genetic architecture underlying each individual disease may be unique (179). These points are underlined in a recent review of psychiatric disorders – each of the nine genetic disorders reviewed has variable heritability, each receives contribution from multiple types of variants, and each appears to have a different underlying architecture (173). Thus, investigation of the entire genetic spectrum of disease will be necessary to identify additional susceptibility variants in diseases like COPD, an approach that requires the use and integration of multiple techniques for identifying susceptibility loci.

1.2.8 Rare variant association studies

Our improving understanding of the genetic architecture of complex diseases (see section 1.2.6) has supported recent studies attempting to identify rare variants contributing to COPD in addition to those in *SERPINA1* (see section 1.2.1). The first WES study in a phenotype of COPD looked at 100 individuals with significant smoking histories but without obstruction and identified a non-synonymous variant in the gene *CCDC38,* suggested to be involved in ciliary function (180). Bruse et al also used WES in individuals that are protected from COPD but compared them to individuals with very severe disease using a case-control design. Combining several variant prioritization techniques with *in vitro* functional studies they suggested a role for two genes, *TACC2* and *MYO1E (181)*. Qiao et al used WES to look at predicted deleterious variants that segregated in families with early onset, severe COPD in the BEOCOPD cohort, and although they offered suggestive evidence for the genes *DNAH8*, *ALCAM*, *RARS*, and *GBF1*, they also noted that genetic heterogeneity significantly reduced their power (182).

Exome arrays are an alternative, cost-effective approach to studying rare variation. These arrays allow simultaneous genotyping of more than 600,000 variants, with a standard 240,000 coding variants specifically chosen to represent up to 97% of known non-synonymous variation in multiple human populations, the majority of it rare (141). In the first exome array study of COPD, the authors failed to identify any novel regions associating with GOLD score > 2, but did offer suggestive evidence for the genes *MOCS3*, *SERPINA12*, and *IFIT3* (183). A similar study using exome arrays on the BEOCOPD, COPDGene, International COPD Genetics Network (ICGN), and the Transcontinental COPD Genetics Study (TCGS) did not identify any rare variation

associated with COPD but did identify a genome-wide significant association with a common, non-synonymous variant in *IL27*, a gene known to play a role in other inflammatory diseases (184).

1.3 PRECISION MEDICINE FOR COMPLEX DISEASE

As the Human Genome Project approached its goal of sequencing the human genome at the beginning of the 21st century, there was significant speculation about the impact of this achievement (185). Some predicted that the results would be rapid, allowing medicine to quickly progress to the point where we would have "an intimate knowledge of the propensities of each individual for multifactorial disorders" (186). Others were more circumspect but also perceived that, given time, the genetic code would offer valuable clinical insight into a broad range of diseases (187).

It was in this context that early adopters of the term "personalized medicine" began to think about a future in which individual genetic variability, newly measurable on an unprecedented scale, would affect the way that therapeutics could be designed and applied (188). Since then, the term has been used broadly to describe medicine that utilizes individual variation – from large "'omics" data to variation in common laboratory measurements – in order to provide better application and dosing of drugs, improved clinical stratification of patients, as well as to identify biomarkers that indicate susceptibility or severity of disease. A number of terms have been used to describe this approach including individualized medicine (189), network medicine (190), stratified medicine (191), P4 Medicine: personalized, predictive, preventive and participatory (192), and precision medicine (193). Recently, the National Research Council recommended the

use of the term precision medicine, as the word ‘personalized’ could be misinterpreted and so that term is used in this work (194).

Some fields have been more successful than others in utilizing genomic and individual clinical information to advance not only research discovery but also clinical decision-making. In cancer therapeutics, the recognition that many cases of familial hereditary breast cancer can be attributed to mutations in the *BRCA1* and *2* genes was a valuable confirmation that certain subtypes of cancers should be managed as unique clinical entities (195, 196). Since then, targeted therapies have been developed to treat other subsets of breast cancer (tamoxifen, trastuzumab) as well as other cancer types including chronic myelogenous leukemia (imatinib), gastrointestinal stromal tumors (imatinib), metastatic melanoma (vemurafenib), and non-small cell lung cancer (gefitinib) (197-201). Each of these therapies relies on molecular analysis to place patients in more individualized treatment groups. Recently, a study demonstrated that a significant number of thyroid tumors identified with a unique pathologic signature are low risk and that substituting a non-cancer diagnosis for these tumors can reduce medical overtreatment and the psychological burden of such a diagnosis, indicating the significant impact of correct disease classification (202). The field of pharmacogenomics has also successfully recognized that not every patient responds to drugs in the same way. There are currently over one hundred FDA-approved pharmaceuticals which contain pharmacogenomics recommendations in their labeling (203). A clinically utilized example is HLA typing to predict common adverse drug reactions (204).

In contrast to the fields of pharmacogenomics and cancer therapeutics, relatively few precision-based therapies exist for chronic, complex diseases. This is despite WHO statistics that chronic cardiovascular, lung and metabolic disease are among the top ten killers in the world (4). Indeed, over 70% of care in the US today is for chronic disease, and this will increase as our

population ages (205). Despite this, the majority of treatments for COPD are designed for symptomatic relief and patient classification into severity of disease state is largely based on PFTs (51). As seen in sections 1.1.2 and 1.1.3, many chronic diseases are in reality heterogeneous syndromes that have a wide spectrum of clinical manifestations, with the pathogenesis of each presentation attributable to multiple interacting genetic and environmental factors. Further, as seen in section 1.2.7, the genetic architecture underlying each measurable trait is extremely complex. Improving our understanding of the genetic basis of COPD offers the potential for targeted therapies, improved risk stratification and reduction, and a better understanding of molecular mechanisms of disease. Based on recent findings supporting the complex underlying genetic architecture in these diseases (see section 1.2.7), it will be necessary to identify rare variants, structural variants, and common variants that have not been identified in current CVAS in order to achieve this goal.

2.0 EXTREME TRAIT WHOLE GENOME SEQUENCING OF EMPHYSEMA

2.1 INTRODUCTION

Although many susceptibility loci for COPD have been identified, they only account for a small portion of the heritability of disease (section 1.2.7). Rare variants are likely to contribute to the genetic architecture of COPD and rare variants in *SERPINA1* leading to AATD support this role (see sections 1.2.1 and 1.2.8).

Advances in next generation sequencing technologies have begun to make sequencing studies in order to identify rare variants contributing to disease both technologically and financially feasible, but studies on large populations are still extremely expensive (162, 206). Approaches to maximize power in smaller populations include studies limited to the extremes of a phenotype or within families (207, 208). Recent whole exome sequencing studies have taken one of these two approaches by looking at heavy smokers, severe emphysematous phenotypes or families of severe early-onset COPD, identifying candidate genes (section 1.2.8) (180-182). The extreme trait sequencing approach is based on evidence that individuals at the extreme of a phenotype are enriched for causal variants (209). This approach is particularly useful outside of the exome, where our understanding of the impact of non-coding variants on gene expression remains limited but is clearly important. However, the contribution of such loci has not been studied in a sequencing study. Here we report the findings from the first WGS study in COPD.

2.2 METHODS

2.2.1 Study Design

This study was approved by the Institutional Review Board for Human Subject Research at the University of Pittsburgh. To control for genetic differences between populations, only US non-Hispanic whites (NHW) were included in this study. Subject recruitment and clinical evaluation of subjects in the Pittsburgh COPD Specialized Center of Clinically Oriented Research (SCCOR) has been described previously (210). Briefly, participants were current or former smokers ages 40–79 with a minimum 10 pack-year cigarette smoking history. Each subject completed a chest CT scan, pre- and post-bronchodilator spirometry and plethysmographic lung volume measurement, diffusion capacity, and demographic and medical history questionnaires. Quantitative measurements of emphysema were analyzed using the percentage of low-attenuation units on density histogram analysis defined as the fraction of voxels less than −950 Hounsfield Units (HU) (F-950) as a percent of total voxels (211).

From this population, 102 subjects with severe emphysema as measured by CT scan (susceptible, F-950 > 0.05) and 86 subjects that did not develop emphysema (resistant, F-950 ≤ 0.01) were identified. An algorithm written in Python was used to match these individuals by between cohorts based on sex, age and smoking history, identifying a maximum number of most-similar individuals. To reach a total of 70 individuals in each cohort, several pairs that were not sex-matched were identified as being the most closely related in their other demographic values (Figure 1A).

Genotyping of candidate non-synonymous substitutions was performed in all NHW individuals in the Pittsburgh SCCOR cohort for whom DNA was available (Figure 1B).

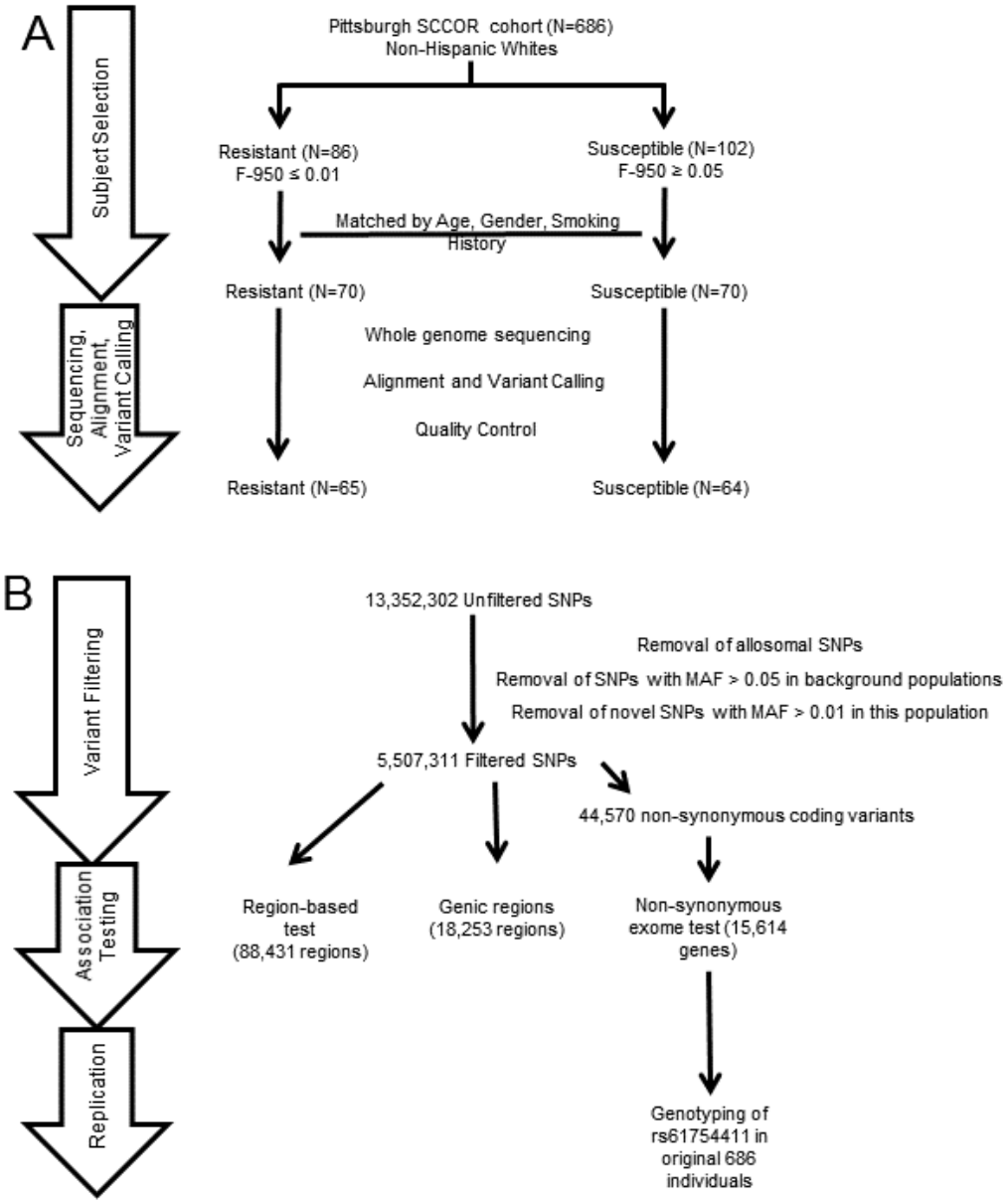

Figure 1. Extreme-trait whole genome sequencing study design. (A) Patient selection and quality control and (B) association testing using three different group-wise tests comparing rare variants or rare non-synonymous variants in 65 individuals resistant to emphysema and 64 individuals susceptible to emphysema.

2.2.2 Sequencing and genotyping

Knome Biosciences (Boston, MA) sequenced 40 subjects (20 resistant, 20 susceptible) and Hudson Alpha Biotechnology (Birmingham, AL) sequenced an additional 100 (50 resistant, 50 susceptible). All individuals were sequenced using Illumina HiSeq technology. Library preparation by both of these centers has previously been described (212, 213). Prior to sequencing, the ratio of DNA with ligated adapters was measured by fluoremetric means prior to sequencing, and preliminary reads on MiSeq technology were used to confirm equal sample distribution and to prevent contamination.

Genomic DNA was isolated from blood using the QIAamp DNA isolation kit (QIAgen, Valencia, CA). A non-synonymous substitution in the gene *PTPRO*, rs61754411, was genotyped using the Taqman platform (214) with pre-designed primer and probes and 7900 DNA analyzer (ABI, Foster City, CA).

2.2.3 Data analysis

We followed the Broad Institute's genome analysis toolkit (GATK) best practices workflow to align and call variants (215). Briefly, Burrows-Wheeler Alignment (bwa, Version 0.7.12-r1039) was used to align sequences to GRCh37 and Picard (Version 1.126) was used for de-duplication (216). Recalibrated variant calls for each individual were generated, followed by joint genotyping and SNP and indel recalibration (217). Final sequencing metrics including depth of coverage, distribution of coverage and variant calling statistics were calculated following analysis of the entire sample set using GATK (Version 3.3-0) and SAMtools (Version 1.1) (218). While

performing quality control, three individuals were removed due to genetic sex not matching recorded sex at blood draw and two individuals were removed because of inadequate sequencing depth for reliable variant calling (Figure 2, whole genome depth < 20X). One susceptible individual with PiZZ genotype in *SERPINA1* was also removed.

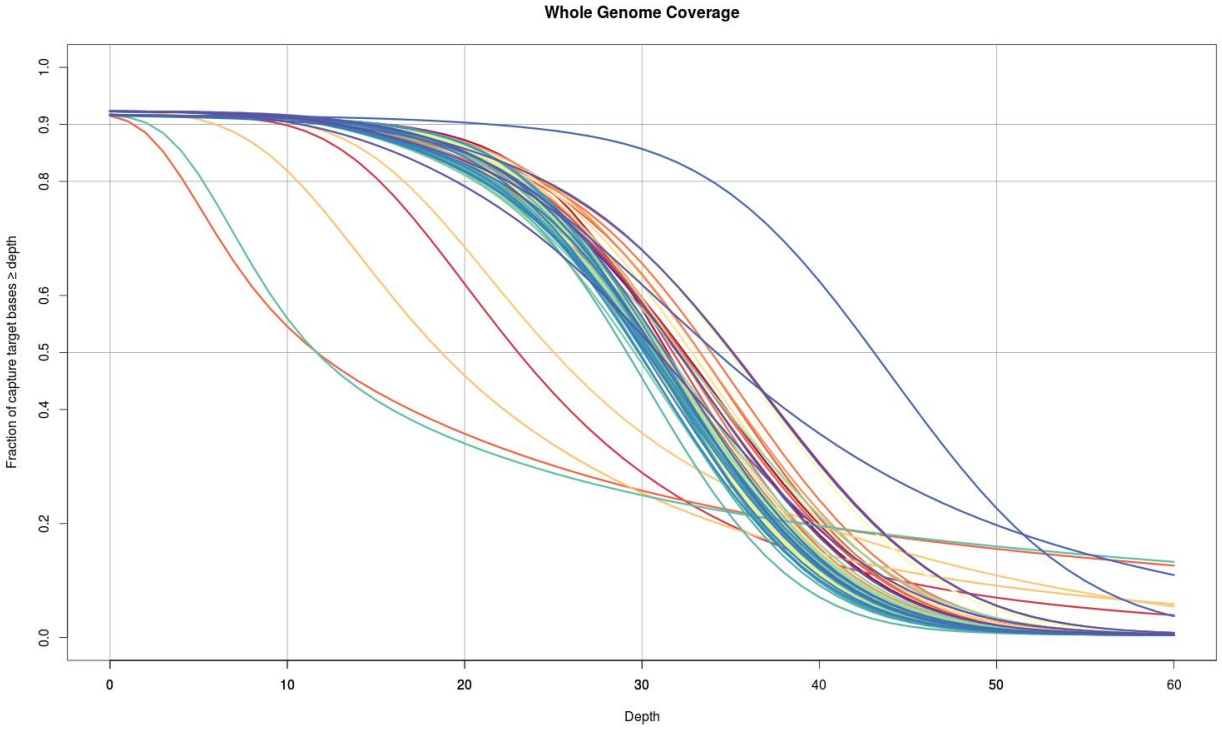

Figure 2. Genome-wide sequencing depth. Depth of sequencing compared to the fraction of the genome covered at that depth for all individuals sequenced in this study. Each individual is represented as a single line.

Population stratification was tested between the sub-cohorts by pruning all variants for which there was more than 10% missingness and that were not in Hardy-Weinberg equilibrium using PLINK (Version 1.90) (219). Principal components analysis (PCA) was used on this pruned set of SNPs and Tracy Widom statistics were calculated for each of the top principal components

using Eigenstrat software (Version 6.0.1) (Table 4) (220). Following pruning, three individuals were removed from the susceptible cohort and two from the resistant cohort due to significant variation along the most significant principal components, PC2 and PC6 (Table 4 and Figure 3).

Table 4. Eigenvectors from principal components analysis of population stratification

Eigenvector	P-value*
1	0.916943
2	0.0923547
3	0.785011
4	0.385701
5	0.463261
6	0.0156966
7	0.463955
8	0.532515
9	0.772406
10	0.526688

*As determined by Tracy Widom test

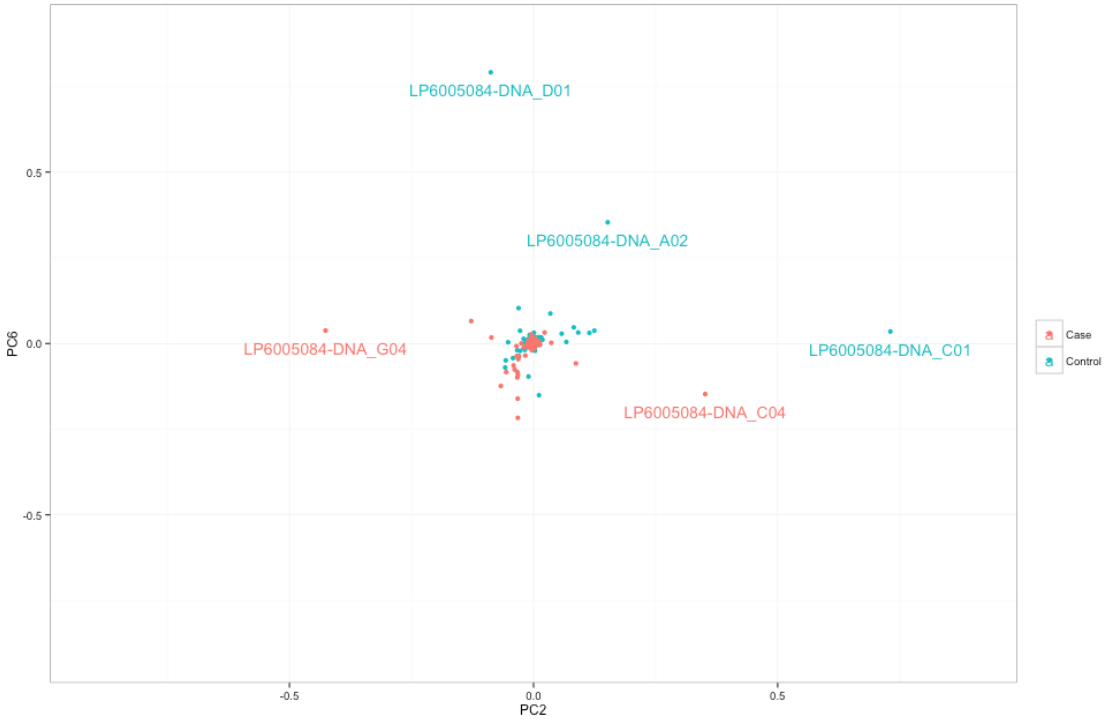

Figure 3. Population stratification of sub-cohorts along two significant principal components. Scatter plot of the two most significant principal components (PC2 and PC6) as determined by Tracy Widom statistics with 134 individuals that passed sequencing quality control. Outliers are labeled and were removed from the final analysis.

2.2.4 Rare variant filtering

Variants were annotated using Annovar and filtered with bcftools. First, variants were removed if minor allele frequency (MAF) > 0.05 in European populations in 1000 Genomes or if MAF > 0.05 in 6500 Exomes European populations. Then, variants were removed if they lacked AF in European populations but MAF in the entire 1000 Genomes populations > 0.05 and finally, remaining variants were filtered if the empiric MAF > 0.01.

2.2.5 Definition of regions and genes

We separated the autosomes into 30,000 base pair (bp, 1000 bp = kbp) regions starting at the beginning of each chromosome. A new region was generated beginning exactly 30kbp from the last base pair using Python. We identified genes as all intronic, exonic and regulatory space (gene start +/- 1kbp) of protein coding genes. We downloaded the start and end locations of all protein coding genes mapped to GRCh37 from Ensembl's Biomart and added 1kbp to start and end locations. Non-synonymous mutations (missense and nonsense) were annotated using EPACTS (version 3.2.6). EPACTS utilizes gencodeV14 for gene and variant annotation.

2.2.6 Association testing and statistics

We tested for association of single rare variants with emphysema using the efficient mixed-model association expedited (EMMAX) algorithm as implemented in EPACTS (221). Unadjusted P-values were also generated using Fisher's Exact Test. Single variants were annotated with Annovar (Version 2014-11-12) from dbSNP144 and PolyPhen and Sift predictions were downloaded from Ensembl.

We used the optimized sequence kernel association test (SKAT-O) as implemented in EPACTS to test for association of groups of rare SNPs with emphysema (222). We first tested across standardized 30kbp windows (section 2.2.5). Then, we tested for all rare variation within genic regions. Finally, we tested for association of non-synonymous variants (missense and nonsense substitutions) with our phenotype across the exome. Annotations of non-synonymous variants were made with EPACTS.

All association tests were corrected for multiple comparisons using Bonferroni correction. All association tests except for EMMAX and Fisher's Exact test were adjusted for the following covariates: age, sex, pack years, and eigenvalues from principal components 2 and 6 identified with Eigenstrat (Table 4). As EMMAX directly accounts for population structure, this test was only adjusted for age, sex, and pack years. Fisher's Exact Test is reported unadjusted. Two sided t-tests were used to compare population demographic and clinical variables. Tracy Widom statistics were used to compare the principal components generated in Eigenstrat.

2.3 RESULTS

2.3.1 Cohort characteristics

From a population of heavy smokers, we identified 102 non-Hispanic white (NHW) subjects with severe emphysema as measured by CT scan (susceptible, F-950 > 0.05) and 86 NHW subjects that did not develop emphysema (resistant, F-950 $\leq$ 0.01) despite similar smoking histories. We matched individuals in this population to identify 70 susceptible and 70 resistant individuals. Six individuals were removed following sequencing quality control and five were removed as population outliers, leaving a population of 64 resistant and 65 susceptible individuals in the following analysis. Resistant and susceptible individuals shared similar characteristics in terms of sex (% female, susceptible: 43.8, resistant: 47.7), age (median years, susceptible: 61, resistant: 63), and smoking history (mean pack years, susceptible: 54.4, resistant: 48.5) but were significantly

different in spirometric (mean percent predicted $FEV_{1,}$ susceptible: 25.3, resistant: 98.0) and radiologic attributes (mean F-950, susceptible: 0.269, resistant: 0.004) (Figure 4, Table 5).

Table 5. Clinical and demographic characteristics of sequenced individuals

	Emphysema Susceptible	Emphysema Resistant	P-value*
Number of Subjects Sequenced	64	65	
Females (%)	43.8	47.7	0.67
Median Age at PFT (Years) [IQR]	61 [57-65]	63 [60-66]	0.01
Mean Smoking History (Pack Years) [IQR]	54.4 [35.6-61.5]	48.5 [30.0-60.0]	0.25
Mean Percent Predicted FEV1 (%) [IQR]	25.3 [19.8-30.0]	98.0 [89.0 - 104.0]	7.63E-29
Mean F-950 [IQR]	0.269 [0.172 - 0.351]	0.004 [0.002 - 0.006]	2.00E-24

* Two-tailed t-test

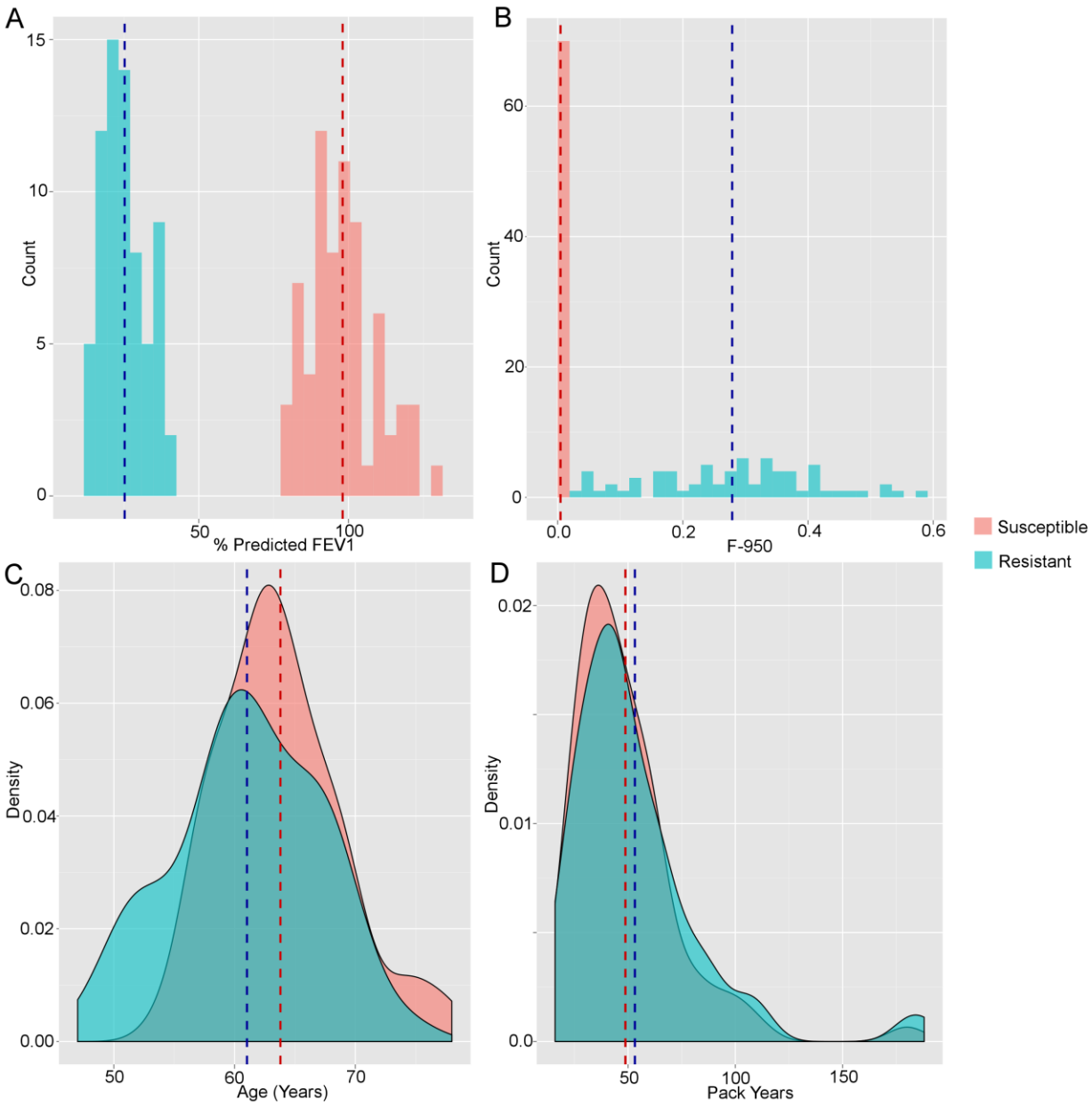

Figure 4. Extreme traits of emphysema in the Pittsburgh SCCOR cohort. Analyzed sub-cohorts are clinical extremes in terms of obstruction as measured by (A) percent predicted FEV1 and (B) F-950 despite having very similar (D) smoking histories (pack years) at a similar (C) age (years). Dotted lines represent sub-cohort means.

2.3.2 Sequencing Results

The average depth of sequencing in all analyzed samples was 30.2X across the genome, with 83.5% coverage of the genome at 20X depth of coverage (Table 6). We identified 13,352,302 SNPs, 2,682,220 of which were novel (compared to dbSNP138) in 135 individuals that passed sequencing quality control.

We used a stepwise approach to filter our variant calls for rare variation, which we defined as any variant with a MAF < 0.05 in background European populations or MAF < 0.01 empirically if no MAF existed in any of these populations (see section 2.2.4). This relatively inclusive definition of rare variation was chosen due to the small size of both the European background populations and the study cohorts. Using this approach, we identified 5,673,659 rare SNPs, with an average of 79,207 rare SNPs per subject (Table 6). We excluded allosomal SNPs from our final analysis and tested across 5,507,311 rare autosomal SNPs.

Table 6. Alignment and variant calling statistics

Mean Aligned Read Depth	30.2X
Mean Aligned at 8X	91.20%
Mean Aligned at 20X	83.50%

Raw Calls	All	Known	Novel
nSNPs	13352302	10670082	2682220
TiTvRatio	2.17	2.21	2.05
nSNPsPerSample	3257790	3227968	29822
TiTvRatioPerSample	2.13	2.13	1.99

Filtered Calls	All	Known	Novel
nSNPs	5673659	3242735	2430924
TiTvRatio	2.18	2.3	2.04
nSNPsPerSample	79207	58973	20234
TiTvRatioPerSample	2.2	2.26	2.03

2.3.3 Genome-wide single rare variant association tests

In order to test for genotype-phenotype association, we dichotomized susceptible individuals as "controls" and resistant individuals as "cases," a binary phenotype which we hereafter refer to as emphysema. We tested each rare autosomal SNP individually for its association with emphysema using the EMMAX algorithm. There were no significant associations following Bonferroni correction (5,507,311 tests, $P \leq 8.8 \times 10^{-9}$). The most significant association was with a set of four noncoding SNPs at 9p13 (rs117400947, rs80121798, rs77945177, rs75985055, $P=7.1 \times 10^{-5}$). The most significant non-synonymous substitution was rs75683534 ($P=1.4 \times 10^{-3}$), a C to A transversion resulting in a premature stop codon in the gene *PIF1* that occurred in 9 susceptible individuals and 0 resistant individuals (Appendix A, Table 19). A quantile-quantile (QQ) plot of the results of this test can be seen in Figure 5.

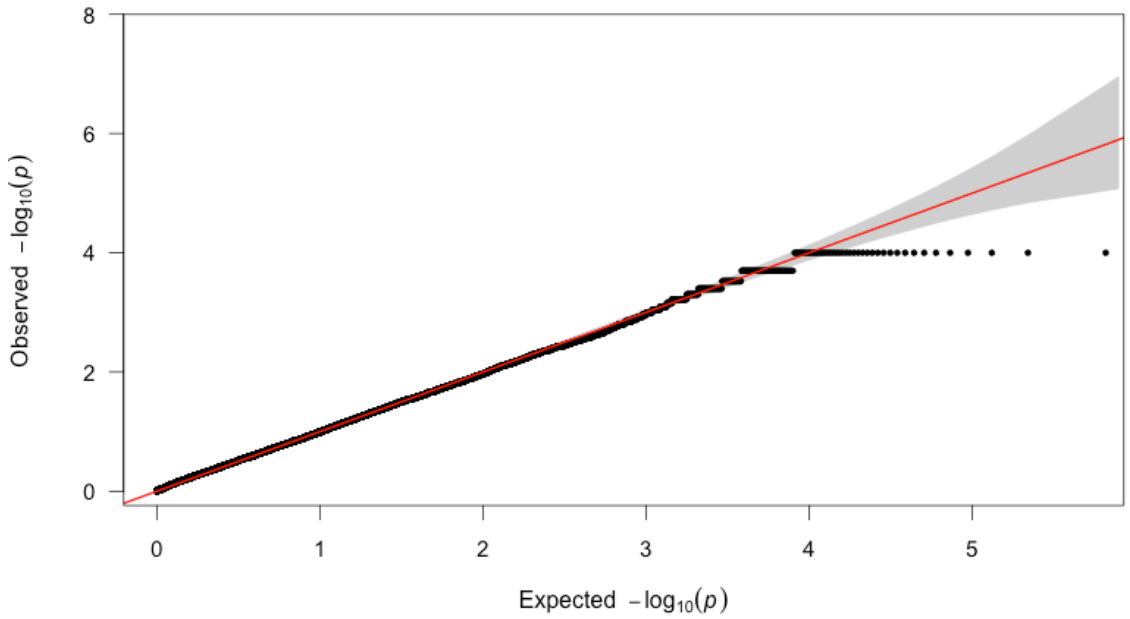

Figure 5. Quantile-quantile plot showing P-values from single variant association test. QQ plot of P-values resulting from tests for association of single variants with emphysema using EMMAX. This QQ plot shows minimal inflation, but deflation at higher expected P-values demonstrates a lack of power in this cohort at the single-variant level.

2.3.4 Genome-wide region-based rare variant association tests

We grouped rare autosomal SNPs in 30kbp windows across the genome and tested for association using SKAT-O (222). The top association in this genome-wide scan was a locus at 19q13.41 (Figure 6 and Table 7, chr19:53430000-53459999, $P = 4.5\text{x}10^{-6}$). Although not significantly associated following Bonferroni multiple-testing correction (88,431 tests, $P \leq 5.4\text{x}10^{-7}$), this locus is suggestively associated ($P \leq 1.1\text{x}10^{-5}$). Fifty-six percent of individuals in this study harbored at least one rare variant in a total of 79 different loci within this region (Appendix A, Table 20). Rare

variants were preferentially harbored by individuals with susceptibility to emphysema (Appendix A, Table 20). The 10 most significantly associated regions can be seen in Table 7, and a QQ plot can be seen in Figure 7.

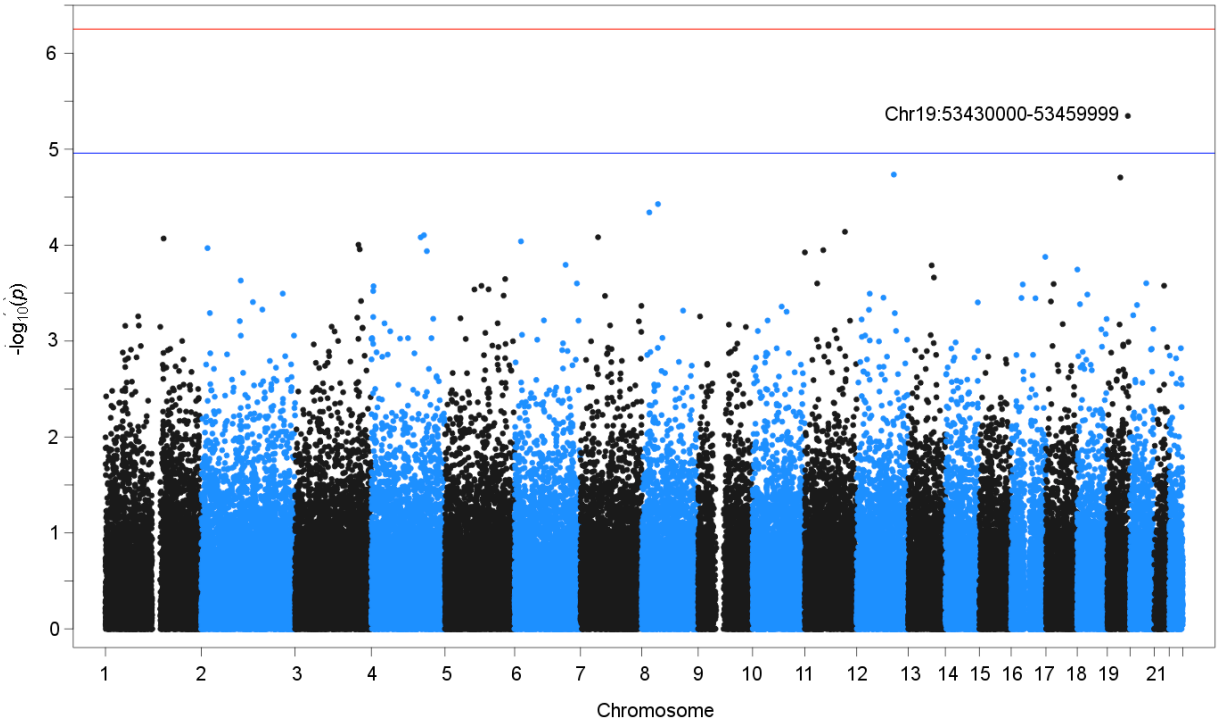

Figure 6. Manhattan plot showing results of 30kbp sliding window association test. Manhattan plot of 30kbp regions spanning the genome. Each dot represents a single 30kbp region and the strength of its association with emphysema is plotted on the y-axis. The top association at 19q31 is labeled. Other top associations are reported in Table 7. The top threshold represents genome-wide significant association, and the bottom threshold represents genome-wide suggestive association following Bonferroni correction.

Table 7. Top associations with emphysema in 30kbp sliding window association test

Chr	Start	End	Fraction with Rare*	Unique Rare#	Unique Singletons ^	P-value	Gene in Region
19	53430000	53459999	0.55814	79	62	4.51E-06	*ZNF816, ZNF321P*
12	96300000	96329999	0.37209	33	18	1.84E-05	*CCDC38*
19	33810000	33839999	0.36434	43	34	1.97E-05	-
8	42060000	42089999	0.45736	67	48	3.74E-05	*PLAT*
8	19560000	19589999	0.62791	77	51	4.56E-05	*CSGALNACT1*
11	104400000	104429999	0.51938	66	47	7.26E-05	-
4	136500000	136529999	0.57364	62	44	7.88E-05	-
7	45630000	45659999	0.51163	66	41	8.28E-05	*ADCY1*
4	128040000	128069999	0.44961	50	40	8.30E-05	-
1	151350000	151379999	0.3876	44	32	8.53E-05	*PSMB4, POGZ*

* Fraction of entire population (129 individuals) that harbors at least one rare variant, # Number of unique variants that passed filtering in this region, ^Number of unique variants that only occurred once in either sub-cohort of this population

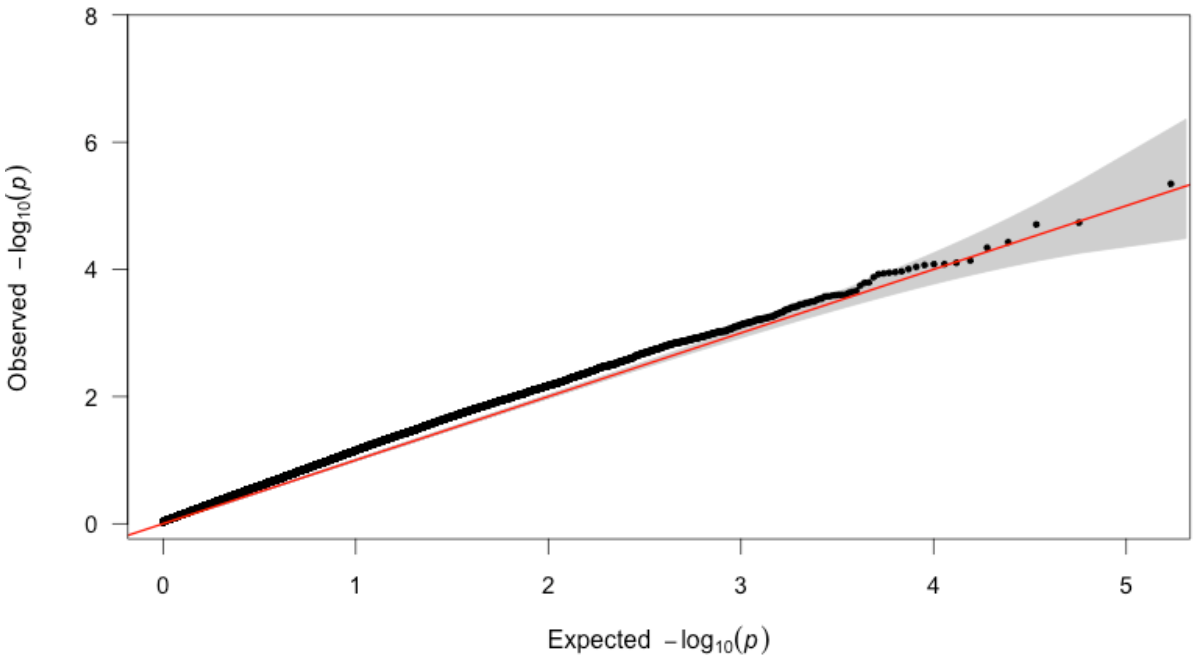

Figure 7. Quantile-quantile plot showing P-values from 30kbp region test. QQ plot of P-values resulting from tests for association of 30kbp regions using SKAT-O.

2.3.5 Gene-based and non-synonymous rare variant association tests

We tested for association of all rare variation with emphysema in coding and non-coding regions of genic regions (introns, exons, and 1kb flanking region in each direction) across the exome. The most significant association in this test was the gene *ZNF816* (P= 7.2×10^{-6}), located at 19q13 and partially covered by the top association in the region-based test (Figure 8, Appendix A, Table 20). The 129 individuals in the population harbored 120 different rare variants in this gene (Table 8). As in the region covered in the gene-based test, rare variation was more abundant in the susceptible population (Appendix A, Table 20). The majority of the rare variation located in this gene was

intronic, although there were six non-synonymous variants that also occurred preferentially in the susceptible population (Appendix A, Table 20). A QQ plot for this test can be seen in Figure 9.

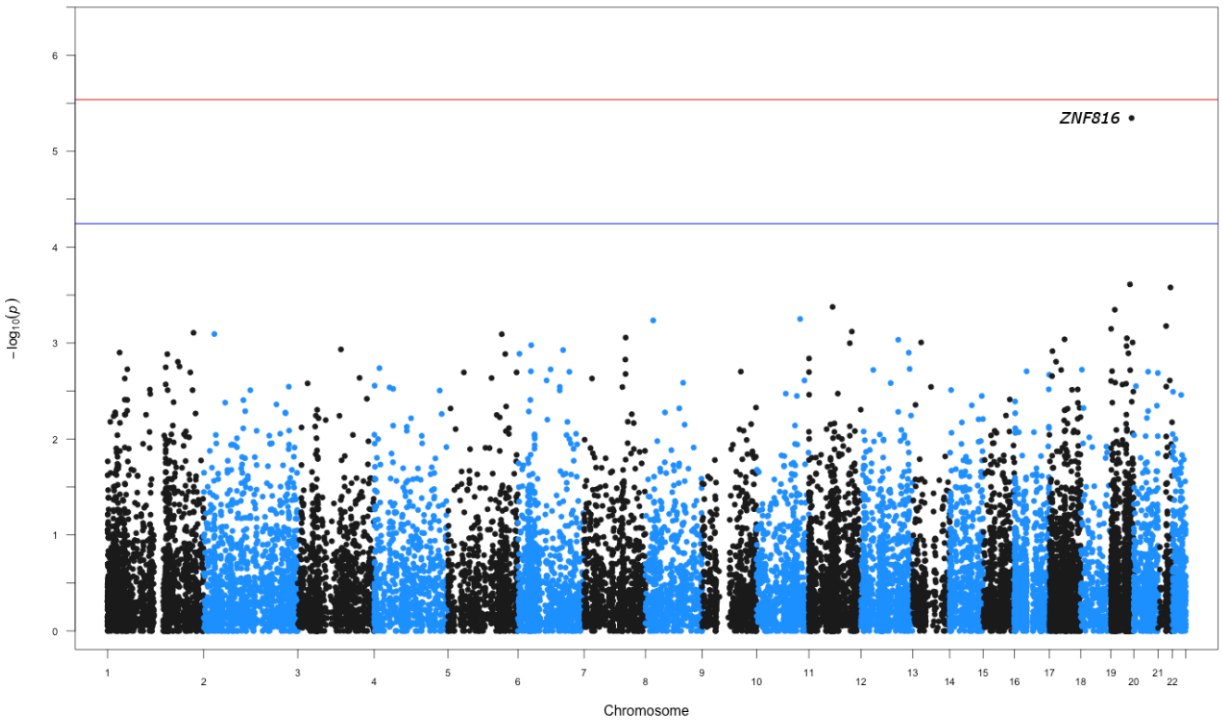

Figure 8. Manhattan plot showing results of gene-based association test. Manhattan plot of gene-based test for association with emphysema including intronic, exonic and 1kbp flanking regions for all protein-coding genes using SKAT-O. Each dot represents a single gene and the strength of its association with emphysema is plotted on the y-axis. The top association with *ZNF816* is labeled. Other top associations are reported in Table 8. The top threshold represents genome-wide significant association, and the bottom threshold represents genome-wide suggestive association following Bonferroni correction.

Table 8. Top associations with emphysema in gene-based association test

Gene	Chr	Start	End	Fraction Rare*	Unique Rare#	Unique Singletons^	P-value
ZNF816	19	53429414	53466404	0.60465	120	96	4.52E-06
BCAT2	19	49298537	49314089	0.17829	20	15	0.000245
TMPRSS2	21	42835502	42903906	0.84496	181	126	0.000263
DAGLA	11	61447355	61515326	0.74419	124	83	0.00042
COL5A3	19	10070204	10121507	0.71318	103	67	0.00045
BBIP1	10	112661005	112680031	0.30233	25	16	0.000561
ATP6V1B2	8	20054059	20084860	0.55039	63	36	0.000581
KRTAP24	21	31652835	31655692	0.062016	5	3	0.000665
CDC34	19	531546	543025	0.37209	39	20	0.000711
C11orf53	11	111125842	111157151	0.72093	56	25	0.000759

* Fraction of entire population (129 individuals) that harbors at least one rare variant, # Number of unique variants that passed filtering in this region, ^Number of unique variants that only occurred once in either sub-cohort of this population

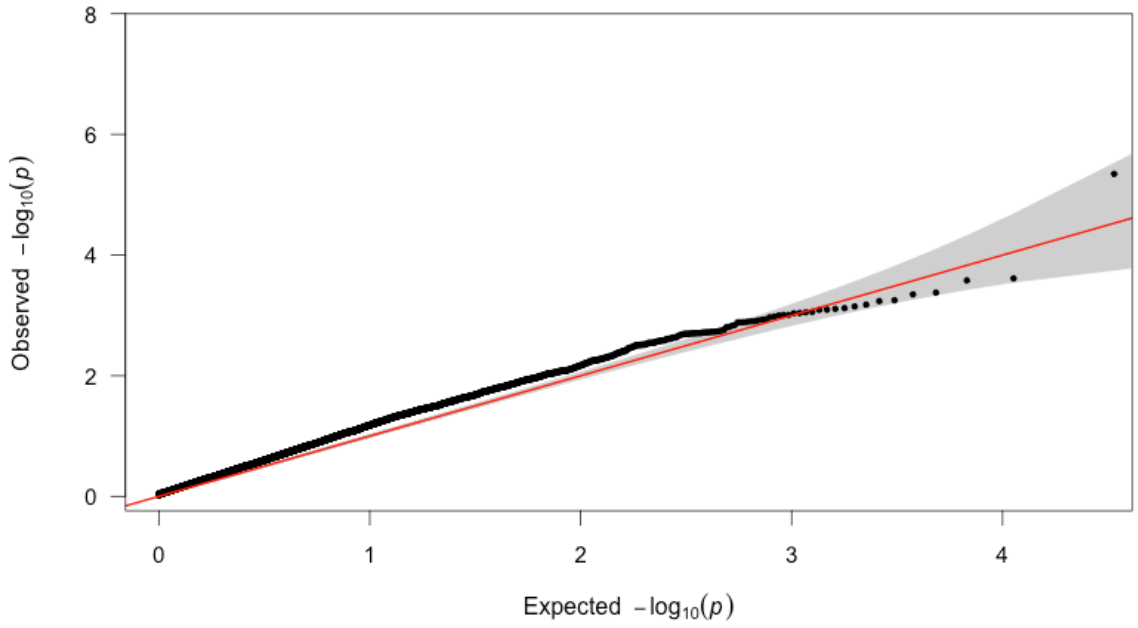

Figure 9. Quantile-quantile plot showing P-values resulting from gene-based association test. QQ plot of P-values resulting from tests for association of gene-based regions using SKAT-O. Deflation in the test statistic can be observed for results with low P-values.

Finally, we tested for association using a group-wise test for non-synonymous coding variants grouped by gene across the exome (223). The most significant association was with the gene *PTPRO* located on chromosome 12 ($P = 4.0x10^{-5}$). We identified four separate rare, heterozygous, non-synonymous substitutions in the gene, with the alternate allele of each substitution occurring only in the susceptible population. One of these substitutions, rs61754411, occurred in 8 individuals in the susceptible group but not at all in the resistant group and was predicted to be deleterious by both PolyPhen and Sift (Table 9). We genotyped this SNP in all 686 NHW individuals in the Pittsburgh SCCOR cohort and found that it was significantly associated with F-950 ($P = 0.035$) and % Predicted FEV_1 ($P = 0.009$) under a dominant model (Figure 10B and 10C, Table 9). Interestingly, we did not identify any individuals homozygous with the alternate allele, nor could we find any individual with this genotype in reported populations.

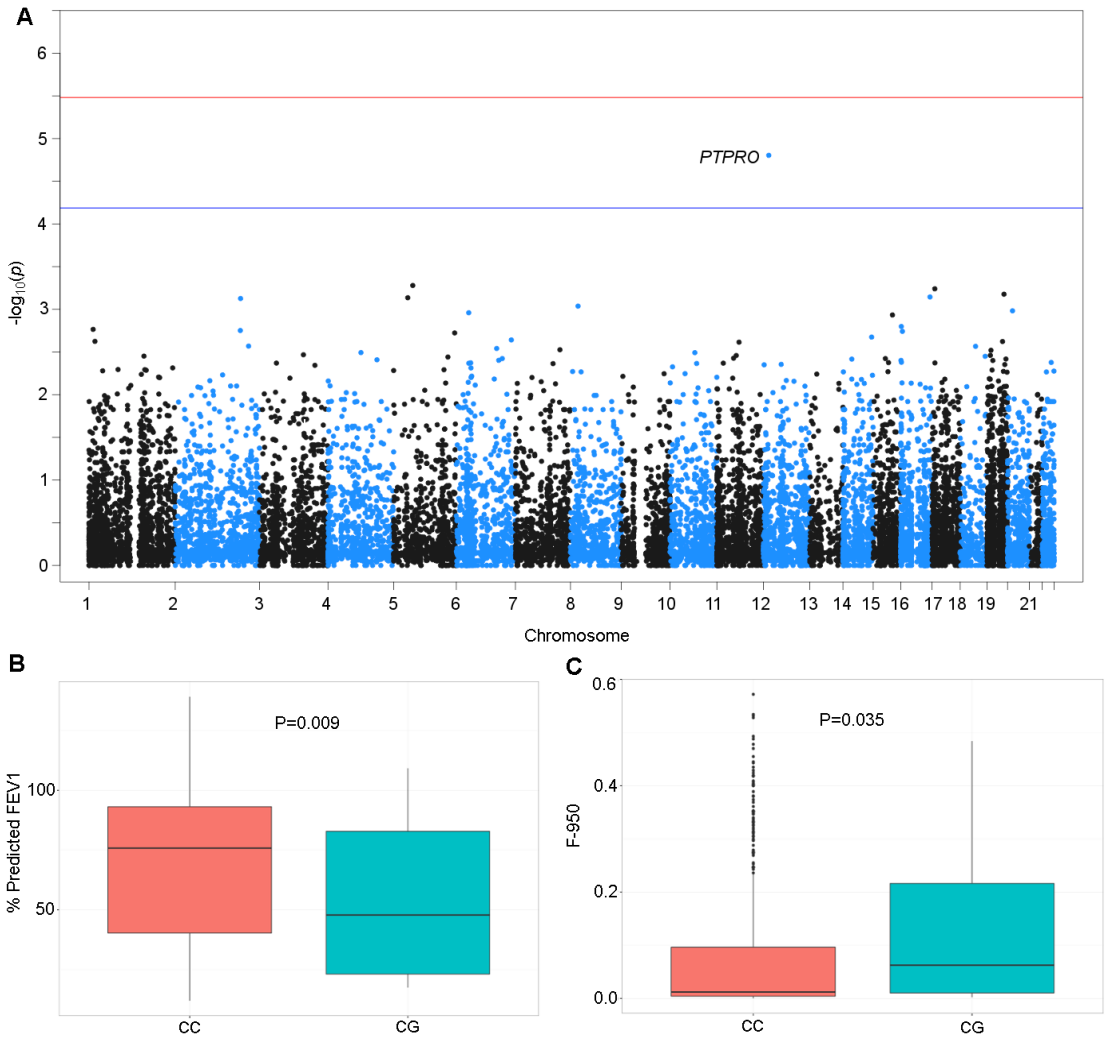

Figure 10. Manhattan plot showing results of association with non-synonymous variants grouped by gene and validation in Pittsburgh SCCOR cohort. Non-synonymous mutations in *PTPRO* are associated with emphysema. A) Manhattan plot of exome-wide test taking into consideration only non-synonymous variation. The top association is the gene *PTPRO* on chromosome 12. The top threshold represents genome-wide significant association, and the bottom threshold represents genome-wide suggestive association following Bonferroni correction. The most prevalent non-synonymous SNP in this gene was genotyped in the entire SCCOR cohort, where it was significantly associated with B) % Predicted FEV1 and C) F-950.

Table 9. Top associations with emphysema for non-synonymous variants grouped by gene

Gene	Chr	Start	End	Fraction of Population with Rare*	Unique Rare NS#	Unique NS Singletons^	P-value
PTPRO	12	15654574	15656846	0.085271	4	2	1.57E-05
IL6ST	5	55237014	55272085	0.054264	4	3	5.26E-04
ALOX15B	17	7942928	7950394	0.062016	6	3	5.75E-04
TMEM143	19	48845944	48866796	0.062016	4	1	6.65E-04
ADAD2	16	84227605	84230332	0.069767	6	4	7.17E-04
CARD6	5	40841579	40860097	0.14729	7	2	7.32E-04
TFPI	2	188331704	188348943	0.085271	2	0	7.47E-04
EPB49	8	21917015	21929924	0.069767	5	2	9.18E-04
FLRT3	20	14306868	14307107	0.062016	4	3	1.04E-03
ETV7	6	36336764	36343720	0.14729	8	5	1.10E-03

* Fraction of entire population (129 individuals) that harbors at least one rare variant, # Number of unique variants that passed filtering in this region, ^Number of unique variants that only occurred once in either sub-cohort of this population

Table 10. Measurement of rs61754411 genotype in the Pittsburgh SCCOR cohort

rs61754411 Genotype	n	Median Age [IQR]	Females (%)	Smoking History (Pack Years) [IQR]	F-950 [IQR]	FEV1 (percent predicted) [IQR]
CC	652	64 [60-69]	46.6	54.8 [30 -60]	0.076 [0.004-0.096]	0.692 [0.403-0.931]
CG	34	64 [59-66]	41.2	56.1 [35-70]	0.126 [0.010-0.216]	0.542 [0.232-0.828]
P-value*		0.217	0.539	0.843	0.035	0.009

*Two-tailed t-test

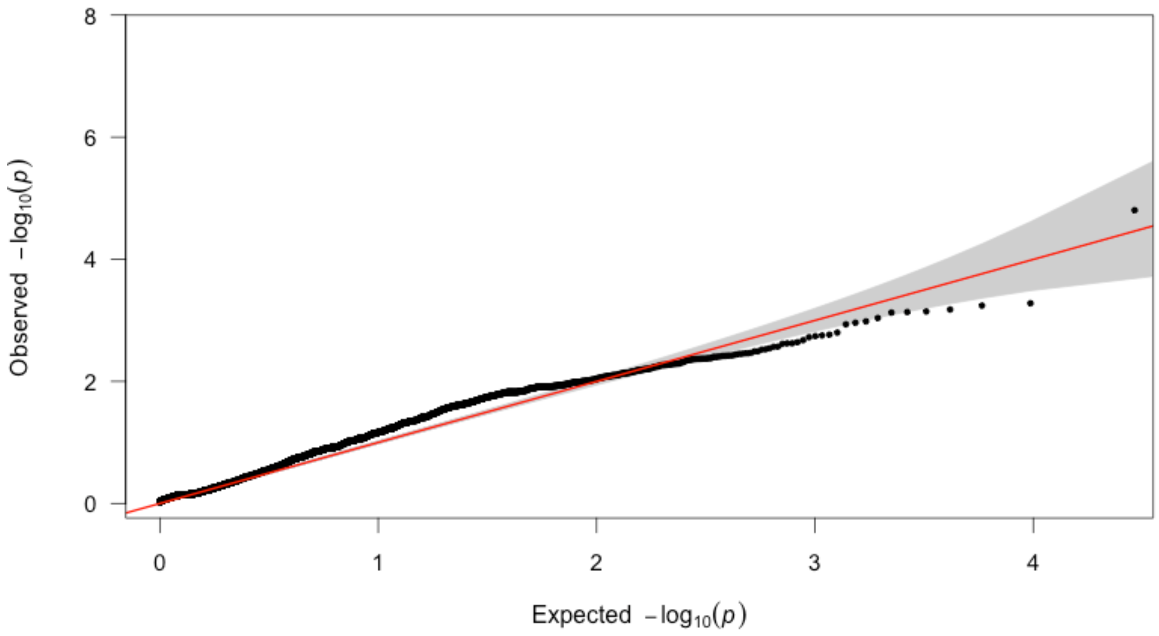

Figure 11. Quantile-quantile plot showing P-values from group-wise testing of non-synonymous variants by gene. QQ plot of P-values resulting from tests for association of gene-based regions collapsing non-synonymous variation by gene across the exome using SKAT-O.

2.4 DISCUSSION

We hypothesized that by selecting two populations of heavy smokers with extreme phenotypes of emphysema we would be able to identify associations with disease using group-wise rare variant tests. The selection of individuals with extreme traits has been shown to enrich for causal variants and this approach has been used successfully to identify candidate genes in sequencing studies (209, 224). Our cohort included two groups of heavy smokers that had extremely different outcomes in terms of both obstruction and emphysema (Figure 4, Table 5).

A primary goal of this study was to test for association of non-genic variants with emphysema. However, our knowledge of the effect of variants in non-coding regions remains limited, making it extremely difficult to prioritize variants as is possible in coding regions (225). Based on this, we chose to test for association with all filtered rare variants in standardized windows across the genome. Using physical distance rather than genetic distance to define the windows tested with SKAT-O was a simple approach to identify clusters of rare variation between two populations. Although more complex approaches have been described, SKAT remains an ideal approach when it is unknown whether variants in a region will have similar or different directions of effect (226).

We identified a single suggestive association using this approach, located on 19q13.41. This region is enriched for rare variants within the gene *ZNF816* in the emphysema susceptible sub-cohort. Notably, when we tested for association of all rare variation with emphysema across gene-based regions, *ZNF816* was also our most significant result ($P= 7.2x10^{-6}$, Table 8). Based on its sequence, this gene, which encodes zinc finger protein 816 (ZNF816), contains 15 zinc finger domains and is likely to be a DNA binding protein, but otherwise its function remains undescribed. It is interesting to note however, that early linkage studies of COPD identified linkage between 19q and pre-bronchodilator FEV_1 in smokers (93, 99).

Although falling below conservative Bonferroni multiple testing correction thresholds, a number of other top associations from our whole-genome scan are of interest. The second most significant association in this test included intronic and exonic variants of the coiled-coil domain containing 38 (*CCDC38)* gene ($P=1.84x10^{-5}$). This gene was identified as a candidate in the first WES study of heavy smokers, looking at resistance to airflow obstruction (180). In that study, the authors identified the non-synonymous SNP rs10859974 as nominally associated with their

phenotype. This SNP was filtered due to MAF in our study suggesting this association is independent of that finding. A common variant association study also identified an intronic variant associated with lung function that is located within our associated region (121). Although the function of *CCDC38*, is not well-understood, it has been suggested to play a role in ciliary function (180).

A region on 8p11.21 was the fourth highest association with emphysema ($P = 3.74x10^{-5}$) and contains the regulatory region for the gene *PLAT*, which encodes the protein tissue plasminogen activator (PLAT, also known as t-PA). PLAT is a serine protease that catalyzes the cleavage of plasminogen to plasmin, activating the fibrinolytic system as well as activating matrix metalloproteinases that have been implicated in lung destruction leading to emphysema (section 1.1.2) (27, 227). The plasminogen system is altered in COPD, with levels of plasminogen and soluble urokinase plasminogen activator receptor (*SPLAUR*, also known as suPAR) correlating with both stable COPD and acute exacerbations of disease (228, 229). Serine protease inhibitor E1 (*SERPINE1*, also known as PAI), the primary inhibitor of plasminogen activation has also been shown to be increased in smoking (230).

A number of top regions in this study that do not contain any coding material are in close proximity to previously identified common variant associations with a phenotype of COPD. One of these is at 11q22.3, 1.6Mbp from an association with severe COPD between the genes *MMP3* and *MMP12* (115). Another is located 8.9Mbp from a locus on 4q31 near the gene *HHIP* that has repeatedly been shown to be associated with phenotypes of pulmonary function and COPD, including FEV_1 in smokers (115).

While WGS allows interrogation of the non-coding regions of the genome, it also offers more efficient coverage of exonic SNPs than WES (223). By testing for association with

emphysema in a gene-based test collapsing only non-synonymous variants, we identified a suggestive association with the gene *PTPRO* on 12p12.3. We replicated the association between the most prevalent rare non-synonymous SNP in this gene (rs61754411) and emphysema and obstruction in a larger cohort. Similar to the top association from our region-based test, the location of this gene is notable as it is located less than 1Mbp from the short tandem repeat D12S1715, which was shown to demonstrate the most significant linkage with moderate airflow obstruction in smokers as well as FEV1 in smokers in two early linkage studies of COPD (93, 94). While *SOX5* was suggested as a candidate gene in this region, it is relatively distant from the top LOD score in that study (100). In this study, only individuals susceptible to emphysema harbored any of the four deleterious substitutions in this gene, which encodes the protein tyrosine phosphatase, receptor type O (PTPRO). PTPRO, a tumor suppressor, is regulated by methylation in lung and liver tumors, with increased methylation leading to decreased growth (231, 232). Prenatal tobacco smoke exposure leads to a significant increase in methylation of *PTPRO,* suggesting that there may be a functional role for the protein following cigarette smoke exposure, perhaps inhibiting tissue repair (233).

Hence, we have identified an association of rare variation with a gene previously associated with protection from heavy smoking in an exome sequencing study, two associated regions at loci previously shown to be in linkage with phenotypes of COPD, and two regions near associations of common variation with numerous phenotypes of COPD. Replication in larger cohorts will be necessary to provide stronger evidence for each of these loci, but the implications are intriguing. It is likely that coding regions can be affected by both common and rare variation, as seen with *CCDC38*. However, regions that are in relatively close proximity but too distant to be plausible regulatory regions for the same genes, such as the association in this study identified near 4q31,

may represent larger 'disease susceptibility regions' in which multiple nearby loci may contribute to COPD susceptibility independently, as has been seen in other complex diseases (234).

Our study has limitations. As the first whole genome sequencing study of emphysema, our sample sizes were limited - even with the extremes of a well-defined phenotype, we were unable to demonstrate significant association at the genome-wide level. Our filtering approach employed an arbitrary MAF, which is more inclusive than has been used in some sequencing studies but still lower than what is measured in a common variant study where markers are specifically chosen for MAF > 0.05. We did this in part because we relied on allele frequencies from relatively small NHW populations from public databases as a background. Finally, the gene-based tests reported here do not correct for gene size and thus favor the detection of larger genes.

2.5 CONCLUSIONS

This is the first study in which individuals with a phenotype of COPD were whole genome sequenced and the first investigation of the contribution of genome-wide rare variation to a phenotype of COPD. Suggestive associations of emphysema with the genes *ZNF816* and *PTPRO* will require further study to determine their biological relevance in this disease. The replication of an association with the gene *CCDC38* using an alternate rare variant analysis approach gives further evidence that rare variation in this gene contributes to emphysema susceptibility. The suggestive or nominal association of numerous loci with previous human genetic evidence for linkage or association with COPD supports the hypothesis that rare variation contributes to this disease in a complex manner. However, as seen in recent exome studies, the effect is likely to be

due to contributions from multiple genes, and potentially as seen here, from multiple genes in nearby ‘disease susceptibility regions.’

3.0 INTEGRATED MURINE AND HUMAN GENOME-WIDE SCANS OF SUSCEPTIBILITY TO CIGARETTE SMOKE-INDUCED EMPHYSEMA

3.1 INTRODUCTION

One hypothesis to explain the 'missing heritability' of chronic diseases is that common variants that do not reach genome-wide significance contribute to disease risk. A number of approaches to identify this type of variant have been proposed, and many of them leverage additional genetic evidence from other sources. For example, Andreassen et al used suspected pleiotropy in two psychiatric disorders to identify previously unrecognized loci that have overlapping genetic mechanisms (235). An alternative to using additional human studies to generate this supportive information is to compare genetic studies in mice to human genetic results. Inbred mice are readily available, their short life span allows studies of diseases that progress for decades in humans to occur in significantly shorter time frames, and their environments can be carefully controlled. These advantages have frequently been employed in molecular studies, and many mechanisms delineated in mouse models have also been found to play a role in the heritability of human disease (236).

Like human CVAS, murine CVAS test the statistical probability of an association given each individual's genotype and phenotype, in this case using inbred strains as individuals. Using the same technologies that have revolutionized human studies such as microarray genotyping and NGS, since inbred mice are homozygous at all loci it is possible to genotype one mouse and test for association with quantitative phenotypes *in silico*. This approach was first demonstrated by

Pletcher et al. in a study of high-density lipoprotein phenotypes with 10,990 SNPs in 48 inbred mouse strains to replicate previous findings from human linkage analyses (237). Since then, genotyping of inbred strains has improved, with whole genome sequences now available for 17 strains of mice, and dense genotyping (approximately 100,000 SNPs, with imputation up to 4 million SNPs) for 94 strains of mice (238).

One of the major concerns with the classic mouse GWAS has been its lack of power to differentiate between spurious associations and positive results (239). These false positives were found to be largely attributable to the complex population structure of the highly related strains. Although linear mixed model corrections (EMMA, EMMAX, GEMMA) or principal component analysis (PCA) can be used to take account of this structure, these corrections often reduce the power of a study (221, 239-241). However, coupled with other techniques like linkage mapping or comparison to genetic studies in humans, mouse GWAS do reveal genetic contributors to the disease they model (242, 243). Since proposed alternatives such as the Mouse Hybrid Diversity Panel often include phenotyping of over 100 strains of mice, for the long periods of exposure that must be conducted to model chronic disease, mouse GWAS using inbred strains still offer significant potential for discoveries in diseases like COPD if their results are supported in additional studies.

In this strain survey, we chose to use a mouse model of pulmonary emphysema in which environmental exposure can be tightly controlled and the extent of emphysema can be directly measured in the lung. This model accounts for difficulties with studying emphysema in humans, notably accurate quantification of the extent of emphysema, accurately accounting for factors such as body size and degree of inflation, and heterogeneous environmental effects. Mice develop emphysema following chronic cigarette smoke exposure and this model has been successfully used

to elucidate the mechanisms of human COPD (27, 244). While variable susceptibility to cigarette smoke-induced emphysema has been reported in five inbred strains of mice, the susceptibility of other commonly used inbred strains remains undescribed as do genetic contributions to that variability (245). To date, no genome-wide assessments using cross-strain comparisons of mouse models of COPD have been conducted, although they have effectively identified candidate genes in other lung diseases, including asthma and pulmonary hypertension (243, 246).

In this study we confirm that mice have variable susceptibility to cigarette smoke-induced emphysema as measured by alveolar chord length (CL) and show that this trait is continuous when measured in 34 inbred strains. By testing for association of this quantitative trait across the genome and integrating our findings with a human genome-wide scan of COPD, we identify a candidate gene for future study.

3.2 METHODS

3.2.1 Animal experiments

All animal experiments were performed in accordance with the Institutional Animal Care and Use Committee (IACUC) of the University Of Pittsburgh School Of Medicine. Mice were housed in a pathogen-free barrier facility that maintains a 12-hour light/dark cycle in Plexiglas cages (one to four mice per cage) with free access to autoclaved water and irradiated pellet food. Animal health, weight, and overall behavior were monitored throughout the experiments.

Female mice of 34 inbred strains (129S1/SvImJ, A/J, AKR/J, BALB/cByJ, BALB/cJ, BPN/3J, BTBRT+tf/J, BUB/BnJ, C3H/HeJ,C57BL/10J, C57BL/6J, C57BLKS/J, C57BR/cdJ, C57L/J, CBA/J, CE/J, DBA/1J, DBA/2J, FVB/NJ, I/LnJ, KK/HlJ, LG/J, LP/J, MRL/MpJ, NOD/ShiLtJ, NON/ShiLtJ, NZO/HlLtJ, NZW/LacJ, P/J, PL/J, RIIIS/J, SJL/J, SM/J, SWR/J) were exposed to either room air (NS, n=4-7/strain) or cigarette smoke (SM, 4 cigarettes/day, 5 days/week for 6 months, n=3-7/strain) as previously described (12). Kentucky Reference Cigarettes (1R5F) were obtained from the Tobacco and Health Research Institute of the University of Kentucky (Lexington, KY). Following smoke exposure, mice were sacrificed, tracheostomized, and lungs were removed and inflated with 10% buffered formalin to a constant pressure of 25 cm water for 10 minutes. Lungs were fixed for 24 hours in formalin before embedding in paraffin. Serial midsagittal sections were obtained and stained with modified Gill's stain. Using Scion Image software (Version 4.0.2, Scion Corp., Frederick, MD), mean alveolar chord length (CL) was calculated on 10 randomly selected 200X fields per slide. Airway and vascular structures were masked from the analysis and the images were manually thresholded. CL was determined in both a horizontal and vertical plane, allowing for the calculation of alveolar airspace areas. Finally, the quantitative phenotype we used in genetic association testing (SMCL) was determined by the following equation: log(SM-NS). This log-transformed value was used in order to ensure phenotype normality.

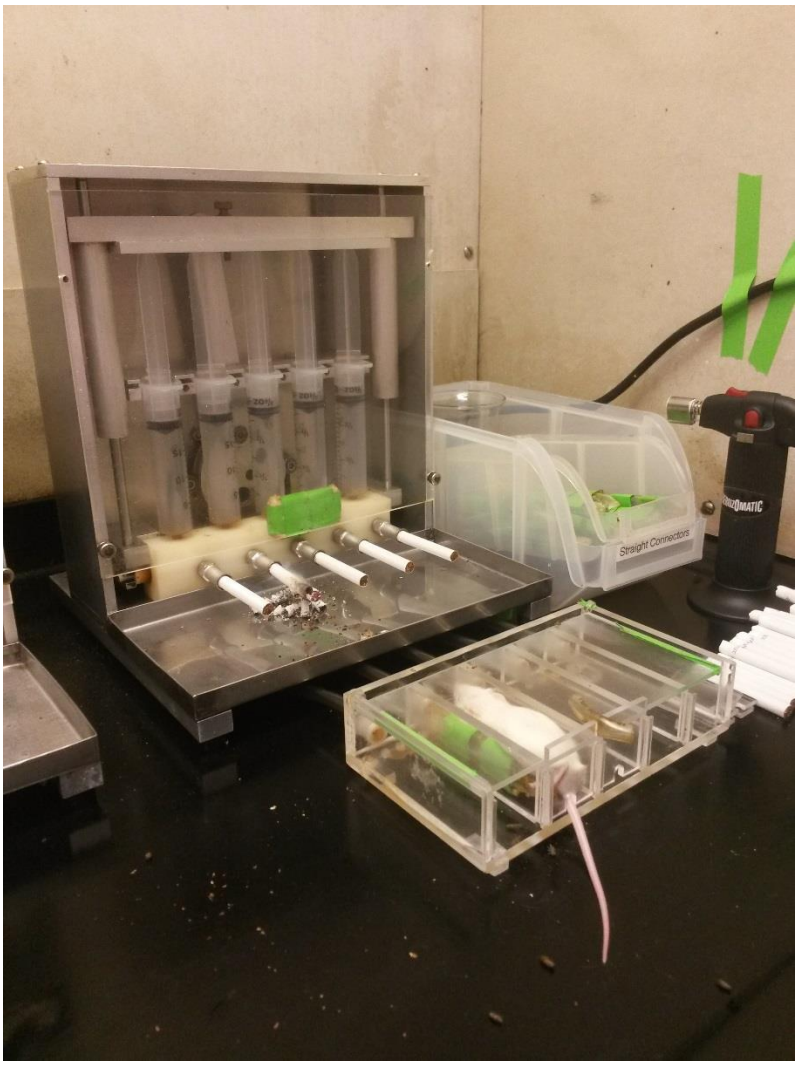

Figure 12. Setup for exposing mice to cigarette smoke.

3.2.2 Genetic association testing

Testing for association between SMCL and 4 million SNPs published by the NIEHS was performed using the genome-wide efficient mixed model algorithm (GEMMA) (238, 241). These SNPs were aligned to NCBI37 and all genomic locations reported in this study reflect that unless otherwise noted. To account for non-variant sites and sites in complete LD between all 34 strains,

simpleM was employed to determine the effective number of tests which was then used to calculate Bonferroni corrected thresholds in this study (247, 248).

3.2.3 Identification of genomic regions enriched for nominal associations

A candidate region was defined as all SNPs yielding $P < 10^{-3}$ (nominal associations) that were within 1Mbp of another nominal SNP. The number of nominal associations in each region was recorded and 1Mbp was added in each direction to the location of the boundary SNPs in order to account for the significant linkage disequilibrium that can occur between inbred mouse strains (249). A hypergeometric enrichment test implemented in R was then used to identify regions enriched for nominal associations compared to the average number of nominal associations in an average region size.

3.2.4 Identification of nominal associations in human GWAS data

Regions in the human genome that are homologous to candidate regions identified in the mouse strain survey were identified using mouse-human homology maps, publically available from the NCBI (250). Human homologs to all genes in the candidate regions were identified. Human homologous regions were identified as continuous blocks of these genes and all separating non-coding regions. Manhattan plots of searched regions were generated using LocusZoom (Version 1.1, http://locuszoom.sph.umich.edu/locuszoom/) (251).

3.2.5 Searching for nominal SNPs in meta-analysis of human COPD

Homologous regions in the human genome were searched for nominal associations ($P < 10^{-4}$) in a previously published meta-analysis of individuals with moderate to severe COPD from the COPDGene, ECLIPSE, NETT/NAS and Norway GenKOLS studies (115). Nominal associations were recorded and the closest genes to these associations were identified as genes of interest.

3.2.6 Identification of candidate genes

All variants resulting in a change to the amino acid sequence of the encoded protein (missense variants, stop gained variants, and frameshift variants) with a predicted deleterious SIFT score (SIFT < 0.05) were identified in Ensembl (dbSNP142) in each of the mouse homologs of the genes of interest in this study. The genotype of each of these non-synonymous variants in CBA/J and A/J mice was downloaded from Ensembl (GRCm38.p4) and the genotype of any variant which differed between these two extreme strains was downloaded for all strains in this study if available.

3.2.7 mRNA expression

Lung tissue was homogenized in Trizol solution and total RNA was isolated according to the manufacturer's instructions (Thermo Fisher, Grand Island, NY). RNA was quantified and used for reverse transcription with reverse transcriptase (Applied Biosystems, Grand Island, NY). Total cDNA was used for real time PCR using primer and probe sets specific for the target genes

(Applied Biosystems, Grand Island, NY). Relative fold change was calculated by comparing abundance normalized to GAPDH between samples with the ΔΔCT method.

3.2.8 Statistics

Chord lengths are reported as mean ± standard deviation. Error bars in figures represent SEM unless otherwise noted. P-values represent between-group comparisons using Welch's two sample t-test unless otherwise noted. Bonferroni corrected genome-wide thresholds were calculated as alpha/effective number of tests determined by simpleM, using an alpha of 0.05 for genome-wide significance and 1 for genome-wide suggestive significance. Statistics were calculated using R (http://cran.r-project.org/).

3.3 RESULTS:

3.3.1 Susceptibility to cigarette smoke-induced emphysema varies continuously in inbred mouse strains

We measured CL in 34 inbred mouse strains after exposure to either six months of cigarette smoke exposure (SM) or room air (NS) (Table 11, Figure 13). There was significant variability in response to SM compared to NS (Table 11, Figure 14). This trait was continuous across 34 inbred strains with CBA/J mice having essentially no response to cigarette smoke (Δ CL -0.3 +/- 1.2 μm) and A/J mice being most susceptible to cigarette smoke (Δ CL 7.0 +/- 2.2 μm) (Figure 13, Table 11). The mean change in chord length across all strains was 3.7 +/- 1.6 μm, and mice of the

129S1/SvImJ strain most closely reflected this change. While the majority of the strains demonstrated a significant change in CL following cigarette smoke exposure, seven strains did not (BPN/3J, CBA/J, ILn/J, KK/HlJ, NOD/ShiLtJ, NZO/HILtJ, and RIIIS/J, Table 11).

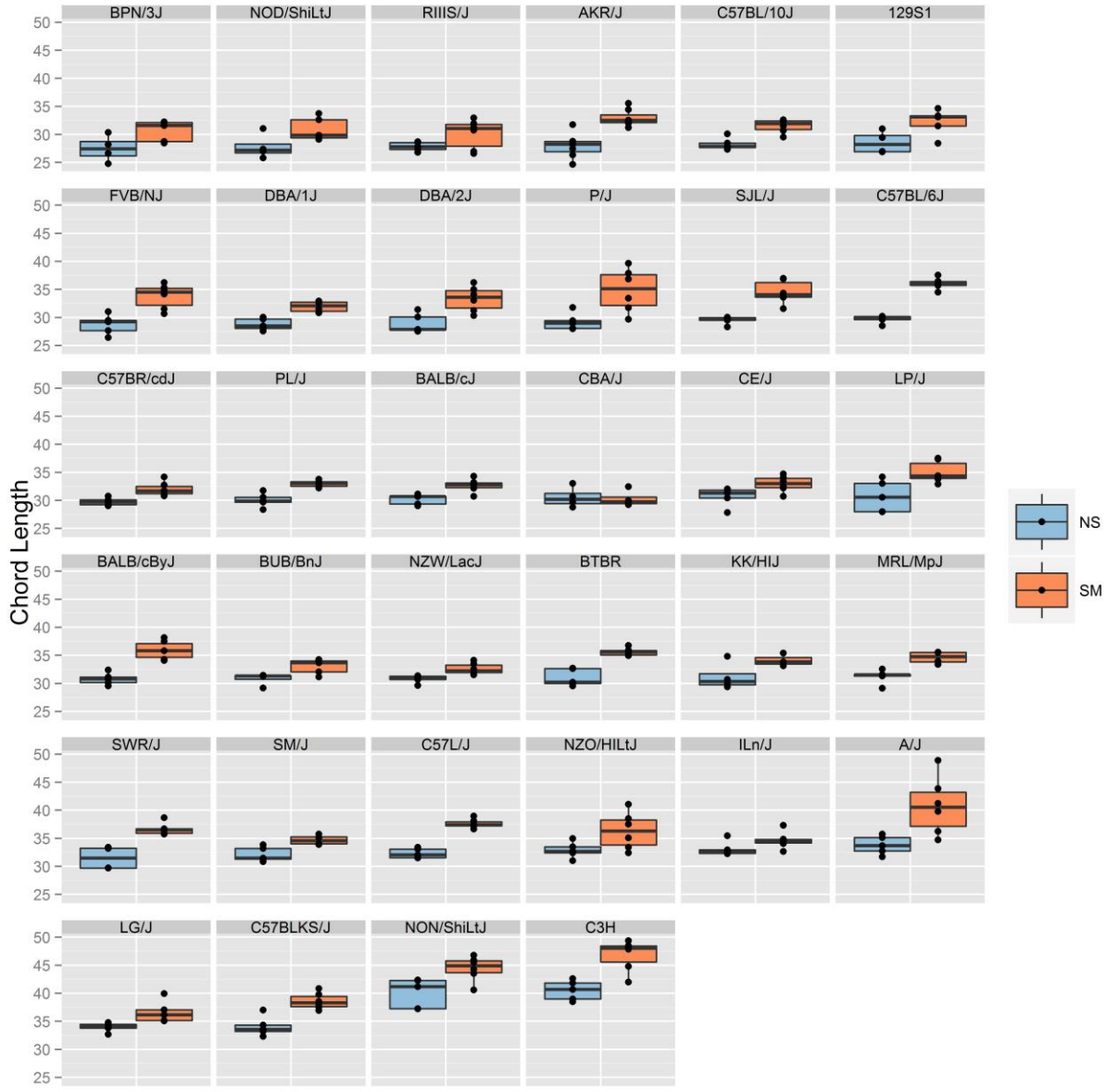

Figure 13. Alveolar chord lengths of 34 inbred strains of mice exposed to cigarette smoke. Alveolar chord length of 34 inbred strains of mice exposed to either room air (NS) or cigarette smoke (SM) for six months.

Table 11. Alveolar chord length in 34 inbred strains of mice exposed to cigarette smoke

	NS			SM			P-value*	Δ CL[#]
	n	Mean CL	SEM	n	Mean CL	SEM		
129S1/SvImJ	4	28.55	0.48	5	32.18	0.67	4.58E-02	3.63
A/J	5	33.78	0.74	6	40.77	2.11	1.83E-02	6.99
AKR/J	7	27.98	0.31	7	32.91	0.41	3.82E-04	4.93
BALB/cByJ	5	30.79	0.77	6	35.93	0.92	2.12E-04	5.14
BALB/cJ	5	30.16	0.79	6	32.64	1.15	4.10E-03	2.48
BPN/3J	4	27.49	1.00	5	30.62	1.07	6.13E-02	3.13
BTBRT+tf/J	5	31.00	0.83	6	35.59	0.57	9.79E-05	4.59
BUB/BnJ	4	30.79	0.91	5	33.01	0.73	3.28E-02	2.22
C3H/HeJ	5	40.49	0.41	6	46.76	0.48	1.95E-03	6.27
C57BL/10J	5	28.27	0.49	6	31.50	0.50	1.34E-03	3.23
C57BL/6J	5	29.67	0.79	6	36.02	0.90	8.21E-07	6.35
C57BLKS/J	5	34.05	0.32	6	38.57	0.85	1.27E-03	4.52
C57BR/cdJ	5	29.74	0.32	6	32.02	0.41	5.64E-03	2.27
C57L/J	5	32.29	0.56	6	37.60	0.25	2.35E-06	5.30
CBA/J	4	30.52	0.48	4	30.27	0.38	8.39E-01	-0.25
CE/J	5	30.67	0.80	6	32.94	0.60	4.09E-02	2.26
DBA/1J	5	28.75	1.04	6	31.92	0.53	5.23E-04	3.17
DBA/2J	5	28.92	0.59	6	33.31	0.44	5.96E-03	4.38
FVB/NJ	5	28.73	0.69	6	33.77	0.28	2.63E-03	5.04
I/LnJ	5	33.12	1.28	6	34.65	0.78	1.13E-01	1.53
KK/HlJ	4	31.19	1.17	3	34.10	0.90	1.22E-01	2.92
LG/J	5	33.92	1.24	5	36.64	0.68	2.25E-02	2.72
LP/J	5	30.72	0.39	6	35.03	0.33	1.52E-02	4.31
MRL/MpJ	5	31.19	0.36	4	34.59	0.89	3.91E-03	3.39
NOD/ShiLtJ	4	27.80	0.36	5	30.93	1.09	6.80E-02	3.13
NON/ShiLtJ	5	40.03	0.59	6	44.40	0.62	1.49E-02	4.37
NZO/HILtJ	5	32.88	0.57	6	36.31	0.55	5.97E-02	3.43
NZW/LacJ	5	30.79	0.68	6	32.57	1.58	9.09E-03	1.78
P/J	5	29.23	0.76	6	34.85	0.59	1.41E-02	5.62
PL/J	5	30.02	1.13	6	32.96	0.93	6.19E-04	2.93
RIIIS/J	5	27.82	0.65	6	30.07	1.34	1.04E-01	2.25
SJL/J	5	29.50	0.31	6	34.50	0.51	6.58E-04	4.99
SM/J	5	32.11	0.55	4	34.67	0.60	1.27E-02	2.56
SWR/J	4	31.48	1.18	5	36.68	0.84	2.02E-03	5.20

* P-values are the result of two-tailed t-tests between SM and NS groups by strain, # Difference between mean SM CL and mean NS CL

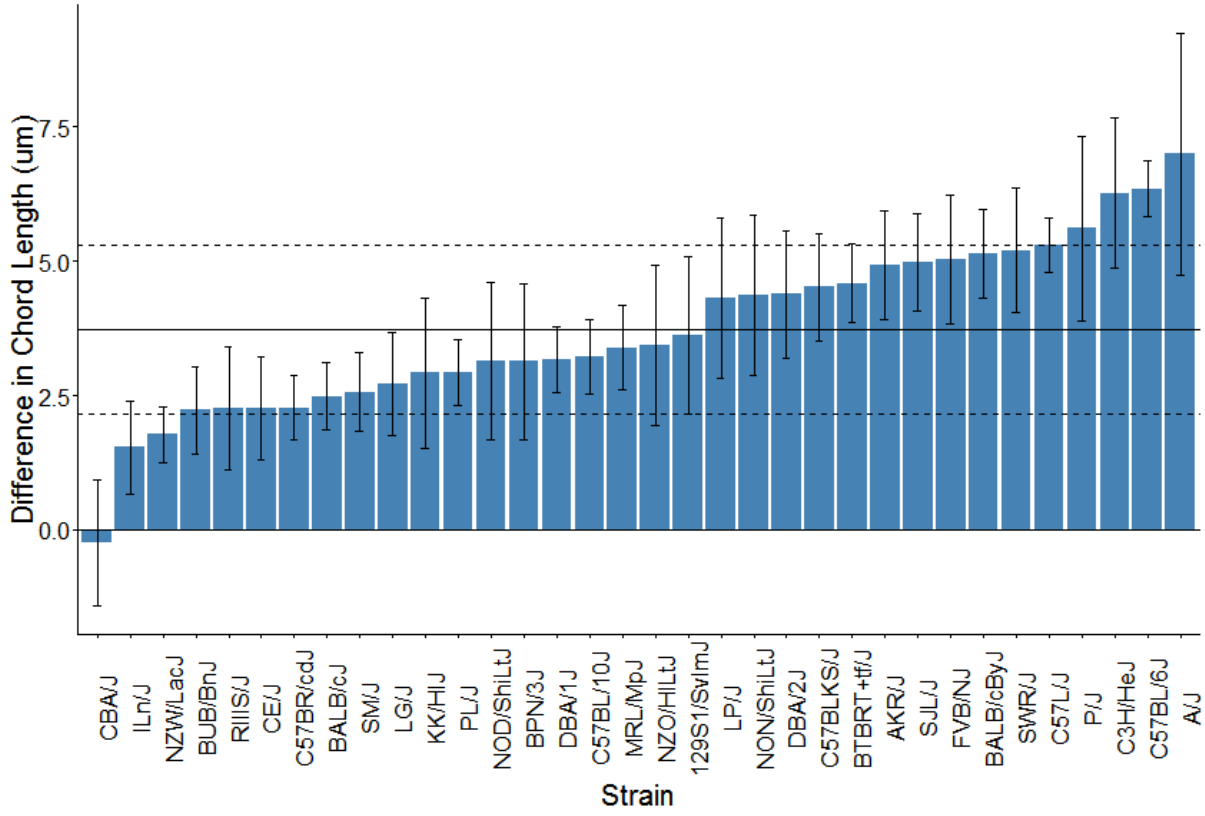

Figure 14. Response to cigarette smoke exposure in 34 inbred strains of mice. Inbred strains of mice have variable susceptibility to cigarette smoke-induced emphysema. Absolute difference in chord length between mice exposed to cigarette smoke and those exposed to room air with SEM. Solid line represents the mean difference in all strains and the dotted lines represent the mean +/- SD.

3.3.2 Genetic contribution to cigarette smoke-induced emphysema is complex in the mouse

We tested for association of SMCL with known genotypes of 34 inbred strains of mice using GEMMA (241). No association with a single variant reached genome-wide ($P < 6.3x10^{-7}$) or genome-wide suggestive significance ($P < 1.\ 3x10^{-5}$), thresholds determined by Bonferroni correction of the 79,492 effective tests determined by simpleM (247). The most significant

association was with rs46114044 located at chr16:54194946 ($P = 7.21x10^{-5}$, Table 12, Figure 15). The top fifty most significant associations can be seen in Table 12.

Table 12. Top 50 most significant SNPs following testing for association with SMCL

SNP	Chr	BP	P-value
NES15713391	16	57704187	7.21E-05
NES15713394	16	57704330	7.21E-05
NES15713400	16	57704563	7.21E-05
NES15713403	16	57704669	7.21E-05
NES15713405	16	57704898	7.21E-05
NES16507860	16	70196382	2.47E-04
NES16507828	16	70199737	2.47E-04
NES16507800	16	70202182	2.47E-04
NES16507715	16	70207149	2.47E-04
NES16461519	16	58415266	2.72E-04
NES16461432	16	58422830	2.72E-04
NES16461435	16	58423286	2.72E-04
NES16461436	16	58423335	2.72E-04
NES16461443	16	58423860	2.72E-04
NES16461447	16	58424113	2.72E-04
NES16461450	16	58424245	2.72E-04
NES16461451	16	58424368	2.72E-04
NES16461452	16	58424487	2.72E-04
NES16461454	16	58424564	2.72E-04
NES16461455	16	58424578	2.72E-04
NES16461456	16	58425007	2.72E-04
NES16461404	16	58425793	2.72E-04
NES16461372	16	58431523	2.72E-04
NES16461287	16	58437863	2.72E-04
NES16461272	16	58439253	2.72E-04
NES16461277	16	58439830	2.72E-04
NES16461278	16	58439865	2.72E-04
NES16461259	16	58440085	2.72E-04
NES16461261	16	58440161	2.72E-04
NES16461256	16	58441448	2.72E-04
NES16461237	16	58442890	2.72E-04
NES16461239	16	58443143	2.72E-04
NES16461247	16	58443682	2.72E-04
NES16461219	16	58444119	2.72E-04

NES16461202	16	58446668	2.72E-04
NES16461203	16	58446800	2.72E-04
NES16461166	16	58447932	2.72E-04
NES16461186	16	58448798	2.72E-04
NES16461133	16	58449357	2.72E-04
NES16461135	16	58449410	2.72E-04
NES16461137	16	58449471	2.72E-04
NES16461139	16	58449543	2.72E-04
NES16461141	16	58449642	2.72E-04
NES16461145	16	58450042	2.72E-04
NES16461100	16	58450973	2.72E-04
NES16461014	16	58456238	2.72E-04
NES16461019	16	58456527	2.72E-04
NES16461020	16	58456665	2.72E-04
NES16461031	16	58457356	2.72E-04
NES16460911	16	58461084	2.72E-04

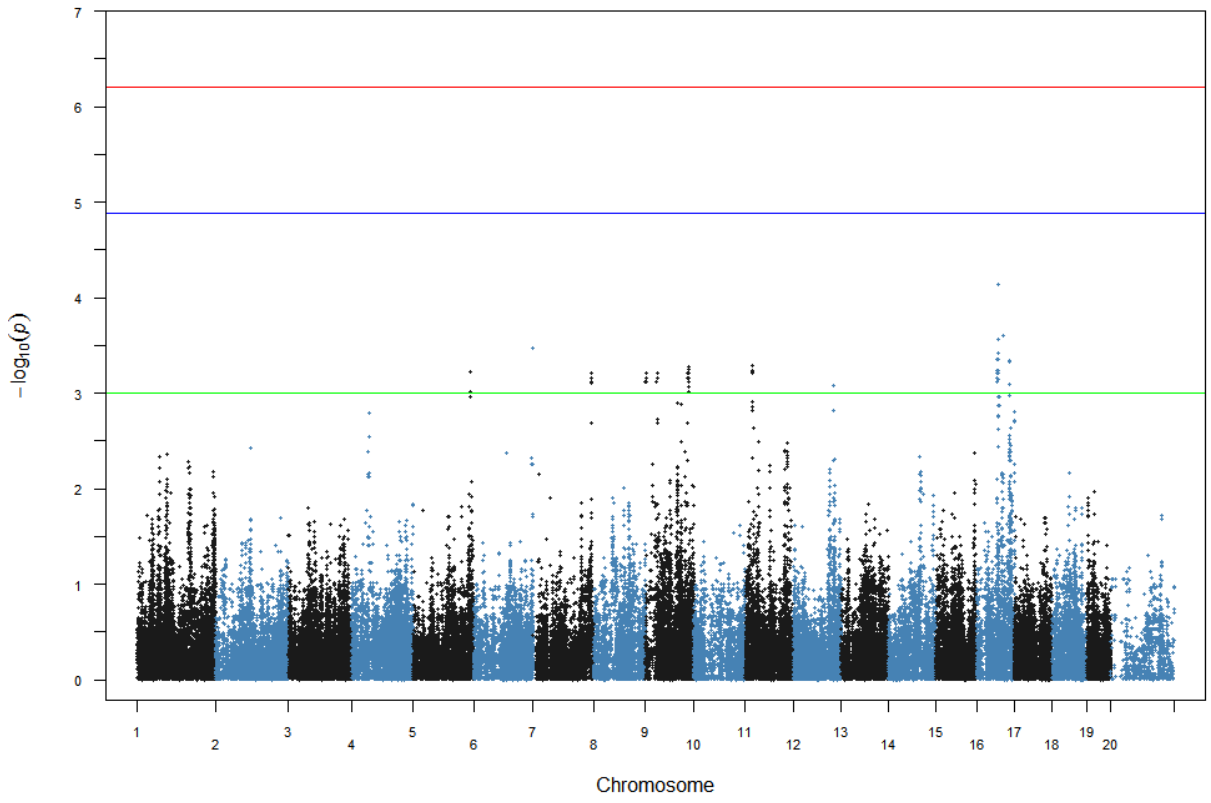

Figure 15. Manhattan plot showing association with SMCL. Manhattan plot of associations with SMCL. Each dot represents a single nucleotide polymorphism (SNP) with its genomic location represented on the x-axis and the negative log of P-values on the y-axis. The top threshold represents genome-wide significance following Bonferroni correction of the effective number of tests, the middle threshold represents suggestive significance and the bottom threshold is drawn at $P=10^{-3}$ to demonstrate nominally associated SNPs were used to identify candidate regions.

In order to identify loci that may be of further interest despite falling below strict multiple testing correction thresholds, we identified regions of variation that had a statistically improbable number of associations below a nominal threshold of significance (P<=0.001) compared to random regions. Using this approach we identified two candidate regions, chr16:54194946-58510508 (region-based P = 0.03) and chr9:107615932-111617795 (region-based P=0.005) (Table 13). Importantly, the region on chromosome 16 also contains the most significantly associated SNP in this study, indicating a region with nominal and robust association.

Table 13. Candidate regions and region-associated P-values

Chr	Start	End	Lowest P-value in Region	Number of nominal SNPs	Region-associated P-value
5	144132865	144522414	5.90E-04	13	1
6	147568652	147572360	3.37E-04	2	1
7	146250946	146777905	6.08E-04	208	0.5943
9	3878327	6240922	6.08E-04	683	0.1385
9	29543038	29543038	7.56E-04	1	1
9	32890118	33020069	6.08E-04	2	1
9	107615932	111617795	5.32E-04	1134	0.0054
11	19232959	20754394	5.10E-04	121	0.8403
12	102164703	102164703	8.35E-04	1	1
16	54194946	58510508	7.21E-05	903	0.0301
16	70196382	70207149	2.47E-04	4	1
16	84379042	84426219	4.55E-04	9	1

3.3.3 Identification of candidate associations from human GWAS

We mapped our candidate regions from the mouse to homologous regions in the human genome using a gene-based approach. Both of these regions mapped to human chromosome 3 (Table 14). We then searched the results of a previously conducted meta-analysis of COPD GWAS in these

regions for any association yielding $P<1x10^{-4}$ (115). Only one association met this threshold, located at 3q12.2 (Figure 16). The most significantly associated variant at this locus is rs62280585 ($P = 7.72x10^{-5}$) which is located between the genes *IMPG2* and *ABI3BP*.

Table 14. Genomic location of homologous candidate regions in the mouse and human genomes

Human			Mouse		
Chr	Start	End	Chr	Start	End
16	54194946	58510508	3	96533425	102198685
9	107615932	111617795	3	35683849	50400230

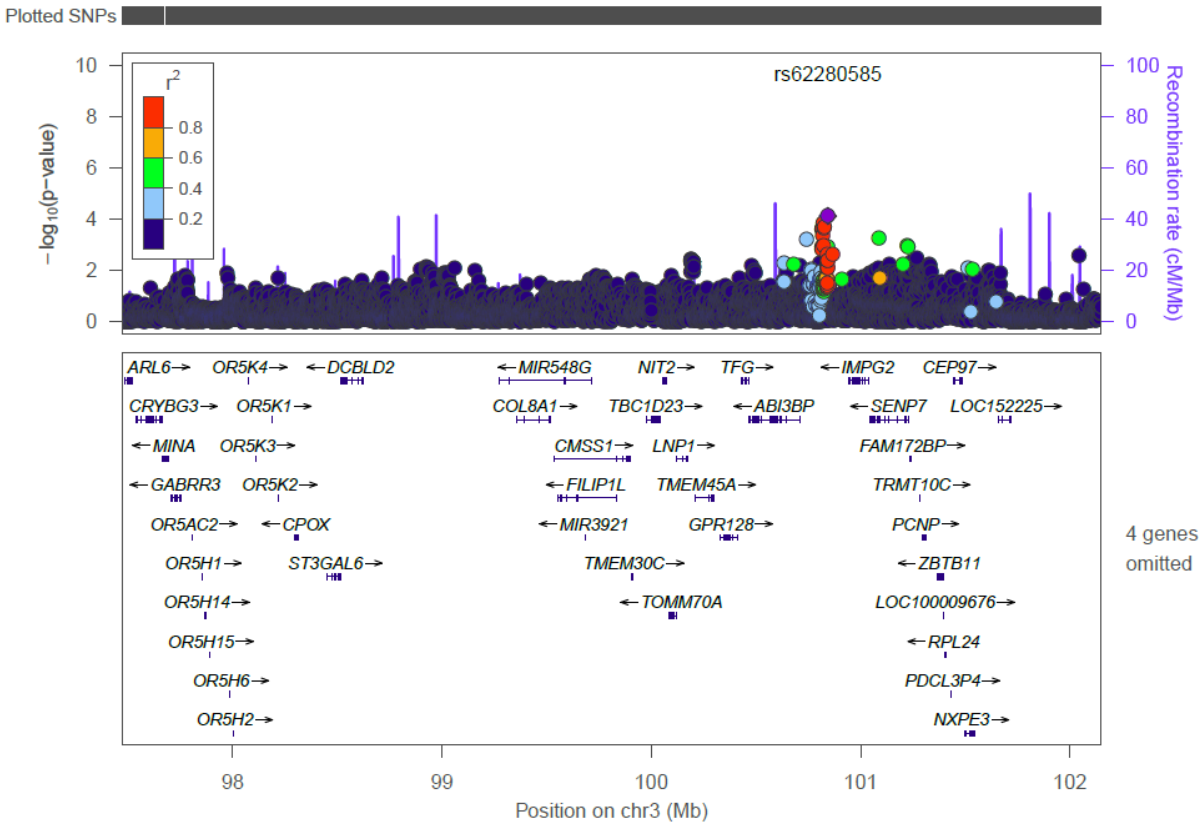

Figure 16. Manhattan plot showing association at human syntenic candidate region. Manhattan plot showing location on chromosome 3 and gene annotations on the x-axis and negative log of P-values on the y-axis. Linkage disequilibrium with the most significant SNP as determined from the CEU cohort of 1000 Genomes is illustrated by different colored SNPs. Plot created with LocusZoom (Version 1.1).

3.3.4 *Abi3bp* contains non-synonymous mutations and is differentially expressed between extreme strains

We attempted to identify coding non-synonymous SNPs (cnSNPs) in the murine homologs of *ABI3BP* and *IMPG2*, in order to identify potentially causal variants. We compared the most susceptible strain, A/J, to the most resistant strain, CBA/J, and determined whether either strain

held the alternate allele of a cnSNP predicted to be deleterious by SIFT. Two non-synonymous mutations in *Abi3bp*, rs253717433 and rs45694439 are predicted to be deleterious to the gene and the alternate allele of each of these variants occurs in CBA/J mice but not in A/J mice (Figure 17). Of note, one of these variants, rs253717433, occurs in exon 5 of *Abi3bp* which is shared among all known transcripts. The only other strain in this study harboring this allele is SM/J which has a small but significant response to cigarette smoke (ΔL_m = 2.6 +/- 0.7 μm, Table 11). The alternate allele of rs45694439 is harbored by several susceptible strains, but this mutation only occurs in one of the four known transcripts of this gene (Figure 17A). Neither of the extreme strains harbor the alternate allele of any of the non-synonymous mutations in *Impg2* with predicted deleterious effect (Figure 17B).

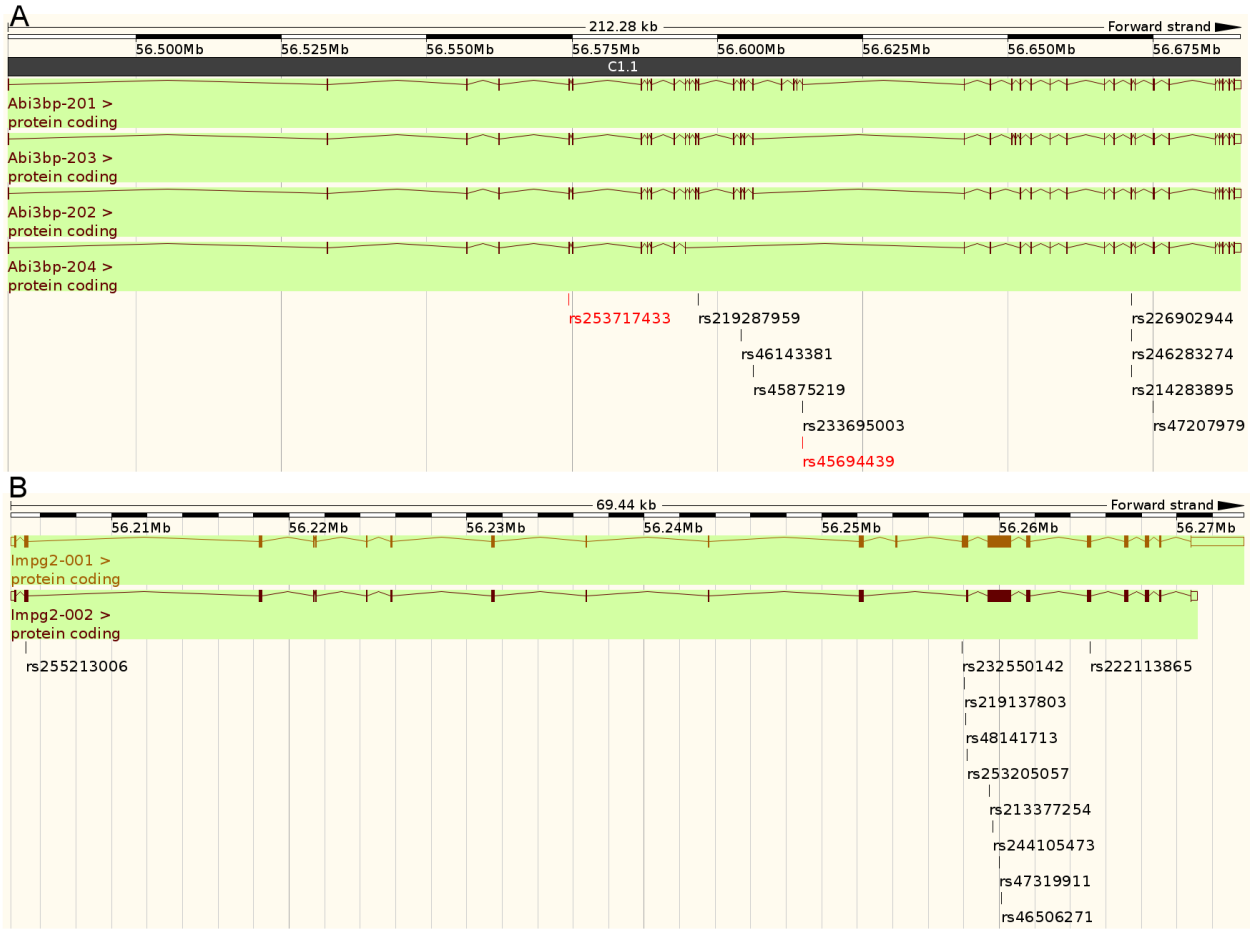

Figure 17. Non-synonymous, deleterious variants in *Abi3bp* and *Impg2*. Non-synonymous variants predicted to be deleterious by SIFT in (A) *Abi3bp* and (B) *Impg2* on mouse chromosome 3. A/J mice harbor the reference allele for all labeled variants, CBA/J mice harbor the alternate allele of those labeled in red. Figure generated with UCSC Genome Browser (252).

Finally, we looked at the RNA expression of Abi3bp in the two extreme strains of this study to identify potential functional effects. There is a significant difference in *Abi3bp* gene expression between the extreme strains, with an approximately 100 fold greater expression in A/J mice that are susceptible to cigarette-smoke induced emphysema than in CBA/J mice that are resistant. (Figure 18).

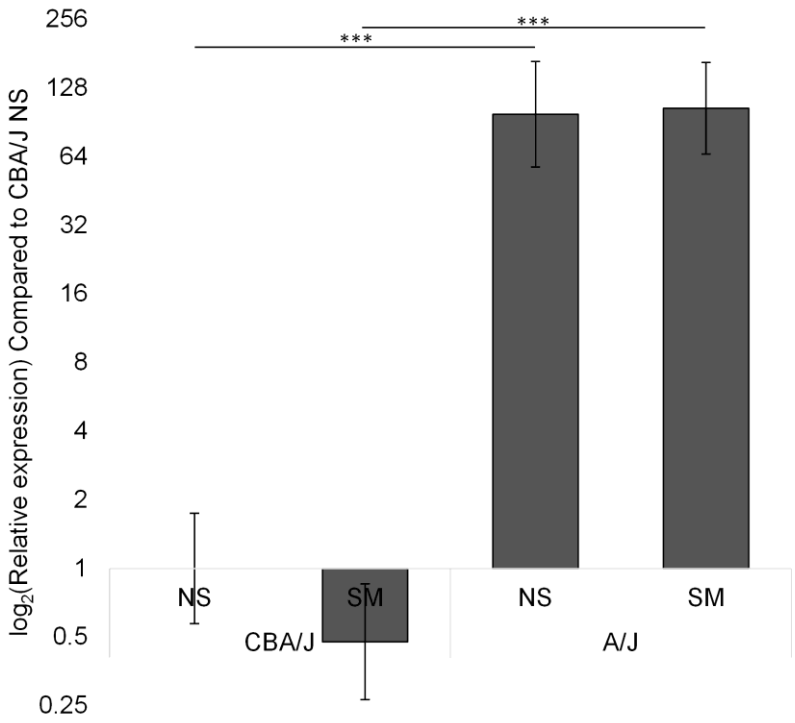

Figure 18. RNA expression of Abi3bp in A/J and CBA/J mice. RNA expression of Abi3bp in A/J and CBA/J mice exposed to cigarette smoke (SM, n=5/strain) or room air (NS, n=5/strain). RNA was isolated from whole lungs after 6 months of exposure to room air (NS) or cigarette smoke (SM); P-values determined by two-tailed t-test of between-group comparisons of ΔCt (***P<0.001).

3.4 DISCUSSION

In this study, we measured alveolar chord length, a surrogate for airspace size and an important measure of emphysema, following chronic cigarette smoke-exposure in 34 inbred strains of mice. This study adds to our understanding of the variable susceptibility to cigarette smoke-induced emphysema in mice, originally described in five inbred strains (NZWLac/J, C57BL6/J, A/J, SJ/L, and AKR/J) (245). Through the measurement of this trait in 29 additional strains, we showed that

there is a continuous range of response to chronic cigarette smoke exposure, from mice that have essentially no response (CBA/J) to mice that are extremely susceptible (A/J). Importantly, we report several common inbred strains in addition to the previously reported NZWLac/J that are extremely resistant to cigarette smoke, demonstrating no significant change in chord length following chronic smoke exposure. At the same time, commonly used strains are comparatively more susceptible to cigarette smoke-induced emphysema than previously reported – for example, C57BL/6J mice are the second most susceptible strain in our study.

While the genetic evidence offered by human GWAS results offer improved resolution and additional confidence to the findings of an inbred strain GWAS, the findings of murine genome-wide scans may also inform our interpretation of the results of future human COPD GWAS. Significant and replicated variants in human studies explain only a small portion of the heritability of COPD and other complex diseases (9, 152). Although some of this 'missing' heritability is likely to be explained by rare and structural variation, it has been demonstrated that for some traits, a significant portion can be explained by common variation failing to reach genome-wide significance due to lack of power or phenotypic heterogeneity (161). Genetic evidence at homologous loci in both mice and humans can thus be helpful in identifying associations that fail to meet significance thresholds but may still represent susceptibility loci.

By integrating our mouse and human genome-wide scans in this way, we identified *Abi3bp* as a candidate gene in emphysema in a relatively large (4.3Mb) region originally identified in our murine inbred strain association study. ABI3BP (Also known as TARSH and eratin) is an extracellular matrix binding protein that was first identified in a yeast two-hybrid screen as binding to c-Abl binding protein ABI3 (253). Additional studies have shown that differential expression or splicing of the gene occurs in lung and thyroid cancers and specifically that reduced expression

of the gene increases tumor growth (254-256). ABI3BP has also been shown to play a significant role in mesenchymal stem cell and cardiac progenitor cell growth; cells lacking *ABI3BP* demonstrate significant increases in proliferation and differentiation (257, 258). In our study we showed that CBA/J mice, that are highly resistant to cigarette smoke-induced emphysema, carry predicted non-synonymous deleterious mutations in *Abi3bp* and have significantly lower expression of *Abi3bp* transcript. It is intriguing to consider that a functional decrease of this gene, which controls growth and differentiation of stem and tumor cells, may also protect from cigarette smoke-induced emphysema potentially by promoting growth and survival of the lung epithelium.

This study has limitations. Although we measured the response to cigarette smoke-induced emphysema in 34 inbred strains, we were still unable to detect any regions that met strict multiple testing correction thresholds. As mentioned, this is partially due to the genetic complexity of the trait and partially because of the power of inbred strain surveys, and a larger survey may be able to overcome this. Since we did not have any regions at genome-wide significance, we chose to only use the most significant association to identify candidate genes. While this was useful for identifying *Abi3bp*, it remains possible that other nominal associations could be of interest with additional supporting evidence.

3.5 CONCLUSIONS

We have demonstrated that susceptibility to cigarette smoke-induced emphysema is a variable trait in 34 inbred strains of mice. Although we were not able to identify any single variants associated with this phenotype at a genome-wide level, we integrated our mouse genome-wide scan with the

results of a human GWAS and identified a novel candidate gene, *Abi3bp*. As the absence of *Abi3bp* has been shown to increase cellular growth in both tumor and stem cells, its potential role in pulmonary emphysema is intriguing and requires further investigation.

4.0 INTEGRATED MURINE AND HUMAN GENOME-WIDE SCANS OF EARLY LIFE PREDISPOSITION TO COPD

4.1 INTRODUCTION

Significant evidence supports the role of a developmental predisposition to chronic respiratory disease (section 1.1.2) but the genetic factors contributing to prenatal and early life predisposition to chronic respiratory disease have not been directly assessed. CVAS have identified numerous susceptibility loci for pulmonary function in the general population (128, 259). Interestingly, many of the loci associated with pulmonary function differ from those associated with COPD. One notable exception to this is a robustly replicated locus at 4q31 near the gene *HHIP* (115, 259). *HHIP* is a regulator of the sonic hedgehog signaling pathway which plays a known role in lung development, indicating that its association with both pulmonary function and COPD may be at least in part due to this role (see section 1.2.6) (260).

A major reason that human genetic studies of early life contributors to COPD have not been conducted is the lack of well-characterized longitudinal cohorts in which pulmonary function is measured in early life. As an alternative to human studies, we look at alveolar size at maturity in a panel of inbred mice to approximate lung growth and development. We test for genetic association and integrate our findings with a genome-wide scan of human disease in order to suggest candidate genes that play a role in both development and disease. This is one of the first genetic studies testing for association with early life predisposition to decreased lung growth and the first time this trait has been systematically studied in the mouse.

4.2 METHODS

4.2.1 Animal experiments

The University of Pittsburgh Institutional Animal Care and Use Committee (IACUC) approved all animal studies described here. Mice were anesthetized using sodium pentobarbital and were sacrificed by carbon dioxide narcosis.

Mice of 36 strains (129S1/SvImJ, A/J, AKR/J, BALB/cByJ, BALB/cJ, BPN/3J, BTBRT+tf/J, BUB/BnJ, C3H/HeJ, C57BL/10J, C57BL/6J, C57BLKS/J, C57BR/cdJ, C57L/J, CBA/J, CE/J, DBA/1J, DBA/2J, FVB/NJ, I/LnJ, KK/HlJ, LG/J, LP/J, MRL/MpJ, NOD/ShiLtJ, NON/ShiLtJ, NZO/HlLtJ, NZW/LacJ, P/J, PL/J, RIIIS/J, SJL/J, SM/J, SWR/J) were used in the following animal studies. Unless otherwise described, mice were sacrificed at 10 weeks of age.

4.2.1.1 Morphometric analysis

Animals used for morphometry were tracheostomized, lungs were removed and inflated with 10% buffered formalin to a constant pressure of 25 cm H2O for 10 minutes. Lungs were fixed for 24 hours in formalin before embedding in paraffin. Serial midsagittal sections were obtained and stained with modified Gill's stain. Using Scion Image software (Version 4.0.2, Scion Corp., Frederick, MD), mean alveolar chord length (CL) was calculated on 10 randomly selected 200X fields per slide. Airway and vascular structures were masked from the analysis and the images were manually thresholded. CL was determined in both a horizontal and vertical plane, allowing for the calculation of alveolar airspace areas.

4.2.1.2 Cigarette smoke exposure:

Kentucky Reference Cigarettes (1R5F) were obtained from the Tobacco and Health Research Institute of the University of Kentucky (Lexington, KY) and cigarette smoke exposure was performed as previously described (28). Female C3H/HeJ or NOD/ShiLtJ mice received exposures equivalent to a single cigarette, two weeks of smoke, or six months cigarette smoke. For single cigarette smoke exposure, mice were exposed to one whole cigarette at 10 weeks of age. Thirty minutes later, animals were sacrificed and lungs were immediately harvested and frozen in liquid nitrogen for downstream mRNA or protein analysis. For two week exposures, mice were exposed to either room air or cigarette smoke (two cigarettes a day, 5 days a week for two weeks) beginning at 10 weeks of age. Lungs were harvested the afternoon of the last exposure day and frozen in liquid nitrogen. For six month chronic cigarette smoke exposures, female mice of C3H/HeJ, NOD/ShiLtJ, or *Il1r1*$^{-/-}$ strains were exposed to either room air (NS) or cigarette smoke (SM, two cigarettes a day, 5 days/week for 6 months). Following smoke exposure, mice were sacrificed, and lungs were either frozen in liquid nitrogen or processed for morphometry as described above.

4.2.1.3 Lipopolysaccharide exposure

Ten-week-old C3H/HeJ or NOD/ShiLtJ mice were exposed to 5 ug of LPS (E. coli strain 11-B4, Sigma, St. Louis, MO) diluted in phosphate buffered saline (PBS) or saline alone by intratracheal (IT) administration. Thirty minutes after exposure, mice were sacrificed and lungs were harvested and immediately frozen in liquid nitrogen.

4.2.1.4 *In utero* anakinra exposure

Ten-week-old female C3H/HeJ mice were injected with 50mg/kg anakinra in saline (Swedish Orphan Biovitrum AB, Stockholm, Sweden) or saline subcutaneously twice daily beginning the day after being paired with male C3H/HeJ mice. If determined to be pregnant, mice were treated through the duration of their pregnancy. Pups were allowed to age to ten weeks and morphometry was measured as described above.

4.2.2 Genetic association testing

Genome-wide scans were performed using the GEMMA algorithm and 4 million SNPs made publicly available by the NIEHS and published in the Mouse HapMap project (238, 241). simpleM was used to determine the effective number of tests using a principal components based approach to account for non-variant loci and loci in total linkage disequilibrium between these 36 strains. The effective number of tests was then used for Bonferroni multiple testing correction of the results (247).

4.2.3 Identification of homologous regions in the human genome

A candidate region was defined as all SNPs yielding $P < 10^{-3}$ (nominal associations) that were within 1Mb (Mb = 1 million base pairs) of another nominally significant SNP. Regions in the human genome homologous to the candidate regions identified in the mouse strain survey were identified using mouse-human homology maps, publically available from the NCBI (250). Human

homologs to all genes in the candidate regions were identified. Human homologous regions were identified as continuous blocks of these genes and separating non-coding regions.

4.2.4 Identification of nominal SNPs in meta-analysis of human COPD

Homologous regions in the human genome were searched for association in a previously published meta-analysis of individuals with moderate to severe COPD and control smokers from the COPDGene, ECLIPSE, NETT/NAS and Norway GenKOLS studies (115). Top associations in these regions were recorded and the closest genes to any association of nominal significance ($P < 10^{-3}$) was recorded as a candidate gene.

4.2.5 Identification of potentially causal SNPs

All non-synonymous variants (missense variants, stop gained variants, and frameshift variants) with a predicted deleterious SIFT score (SIFT < 0.05) were identified in Ensembl (Release 84, dbSNP142) in each of the mouse homologs of the genes of interest in this study (261, 262).

4.2.6 mRNA Expression

Total RNA was collected from whole lung tissue with an RNA collection kit (Qiagen, Germantown, MD) and stored at -80C. Using a NanoDrop spectrophotometer samples were quantified and cDNA was obtained by reverse transcription using an RT kit (Life Sciences, St. Petersburg, FL). Quantitative real-time PCR was conducted on an Applied Biosystems 3700 system with gene-specific TaqMan primers (Life Sciences, St. Petersburg, FL).

4.2.7 Antibodies

Primary antibodies against IL-1R1 (AF771, R & D Systems, Minneapolis, MN), IL-1R2 (AF563, R & D Systems, Minneapolis, MN), AQP5 (sc-9891, Santa Cruz Biotechnology, Santa Cruz, CA), SP-C (sc-7706, Santa Cruz Biotechnology, Santa Cruz, CA), and F4/80 (sc-59171, Santa Cruz Biotechnology, Santa Cruz, CA) were used in the following protein detection methods. CD 115 PE (AFS98, eBioscience, San Diego, CA) and CD11c PE (N418, eBioscience, San Diego, CA) were used for cell selection. Secondary antibodies for immunofluorescence included species-relevant Alexa Fluor antibodies (Thermo Fisher Scientific, Pittsburgh, PA).

4.2.8 Isolation of perfused, lavaged lung tissue, alveolar macrophages and blood monocytes

C3H/HeJ or NOD/ShiLtJ mice were sacrificed, blood was drawn from the left ventricle, and the systemic circulation was perfused with PBS. Following tracheostomy, bronchoalveolar lavage using PBS was performed and BAL fluid was obtained. Whole blood was treated with RBC lysis buffer (Sigma, St. Louis, MO), centrifuged and following resuspension, the pellet was used for blood monocyte isolation using CD115 PE as the primary antibody, anti-PE MACS beads as a secondary antibody (Miltenyi Biotec, Bergisch Gladbach, Germany), and MACS separation columns (Miltenyi Biotec, Bergisch Gladbach, Germany) according to the recommended protocol. BAL fluid was centrifuged, and this pellet was used with CD11c PE as a primary antibody antibody in the same way, for isolation of alveolar macrophages.

4.2.9 Immunofluorescence

Immunofluorescence was performed as previously described. Briefly, frozen slices of whole lung tissue were permeabilized, blocked with bovine serum albumin, exposed to primary antibodies, secondary antibodies and stained with Hoechst to visualize the nucleus. Stained sections were imaged using a Fluoview 1000 upright confocal microscope (Olympus, Center Valley, PA).

4.2.10 Western Blotting

Detection of proteins by immunoblotting was performed as previously described. Briefly, whole lung tissue was homogenized mechanically and resuspended in RIPA buffer. Equalized protein samples were separated on 4-15% Tris-HCl gels, transferred to PVDF membranes and probed with protein-specific antibodies. HRP-conjugated antibodies targeted against the primary antibody were, detected with chemiluminescent substrates (Pierce) and visualized by x-ray film. Quantification was performed with ImageJ.

4.3 RESULTS

4.3.1 Alveolar chord length varies between inbred strains at 10 weeks of age

We attempted to model structural differences resulting from development and early growth in the lung by looking at mean alveolar chord length, an approximation of alveolar size, at ten weeks of age (10wkCL) in a panel of inbred mice. In mice of 36 inbred strains (n=6-9 mice/group) we

observed a significant inter-strain variation of this quantitative trait (Figure 19, Table 15). NOD/ShiLtJ mice demonstrated the smallest 10wkCL (25.8 ± 1.5 μm) and C3H/HeJ mice demonstrated the largest 10wkCL (38.6 ± 3.9 μm). The mean 10wkCL for all mice in this study was 30.7 ± 3.4 μm. In addition to C3H/HeJ mice, NON/ShiLtJ (37.7 ± 1.5 μm) and A/J mice (35.0 ± 2.3 μm) had relatively large airspaces at ten weeks of age while DBA1/J (27.0 ± 1.6 μm), DBA2/J (27.3 ± 1.3 μm) and the wild-derived PWD/PhJ (27.8 ± 1.0 μm) had relatively small airspaces (Table 15).

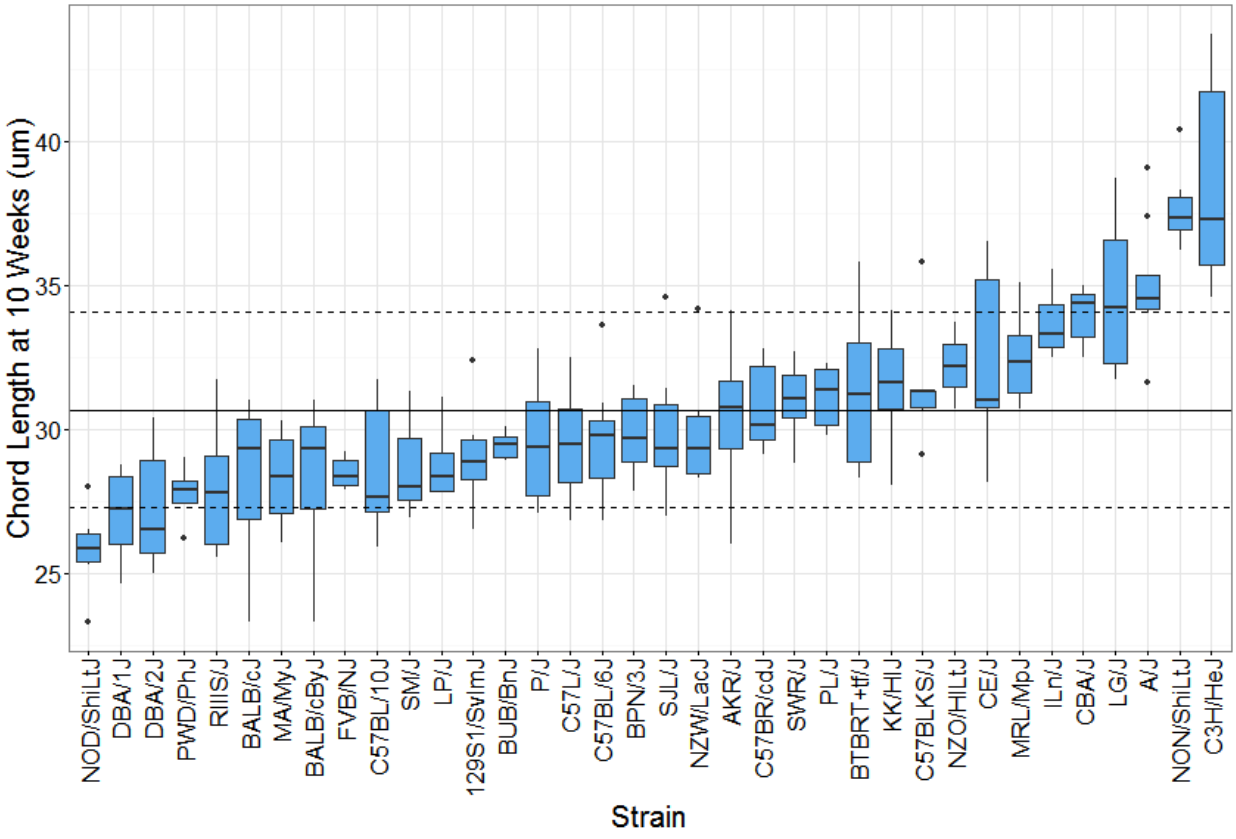

Figure 19. Chord length at 10 weeks of age for 36 inbred strains of mice. Mice of 36 inbred strains (n=6-9/strain) have variable chord lengths at 10 weeks of age. Solid horizontal line represents the mean difference in all strains and the dotted horizontal lines represent the mean of all strains ± SD.

Table 15. Alveolar chord length at 10 weeks of age in 36 inbred strains of mice

Strain	n	Mean 10wkCL	Stdev
129S1/SvImJ	6	29.1	2.0
A/J	8	35.0	2.3
AKR/J	6	30.4	2.8
BALB/cByJ	8	28.4	2.6
BALB/cJ	6	28.3	3.0
BPN/3J	8	29.8	1.3
BTBRT+tf/J	6	31.4	3.0
BUB/BnJ	6	29.5	0.5
C3H/HeJ	6	38.6	3.9
C57BL/10J	6	28.6	2.4
C57BL/6J	8	29.7	2.1
C57BLKS/J	6	31.6	2.2
C57BR/cdJ	6	30.8	1.6
C57L/J	6	29.5	2.1
CBA/J	7	34.0	1.0
CE/J	9	32.3	2.9
DBA/1J	6	27.0	1.6
DBA/2J	6	27.3	2.3
FVB/NJ	6	28.5	0.6
ILn/J	7	33.7	1.2
KK/HlJ	6	31.5	2.1
LG/J	8	34.6	2.6
LP/J	6	28.8	1.3
MA/MyJ	6	28.3	1.7
MRL/MpJ	6	32.5	1.7
NOD/ShiLtJ	6	25.8	1.5
NON/ShiLtJ	6	37.7	1.5
NZO/HlLtJ	6	32.2	1.1
NZW/LacJ	6	30.0	2.2
P/J	7	29.5	2.1
PL/J	6	31.2	1.1
PWD/PhJ	6	27.8	1.0
RIIIS/J	6	28.0	2.4
SJL/J	6	30.0	2.7
SM/J	6	28.6	1.7
SWR/J	6	31.0	1.4

4.3.2 Genetic association testing identifies variants suggestively associated with 10wkCL

We conducted a genome-wide test for association between 4 million SNPs published by the NIEHS and 10wkCL (238). The most significant association was with rs33154484, an intronic variant of *Nr3c2* at 8qC1 (Figure 20, Table 16, $P= 3.64x10^{-6}$). This association was suggestive but not significant (suggestive – $P<8.4x10^{-6}$, significant – $P<4.0x10^{-7}$) at the genome-wide level following Bonferroni correction of the effective number of tests (118,393) as determined by simpleM. Several other loci at 1qC1, 8qC1, and 1qB contained at least one suggestively associated SNP (Table 16).

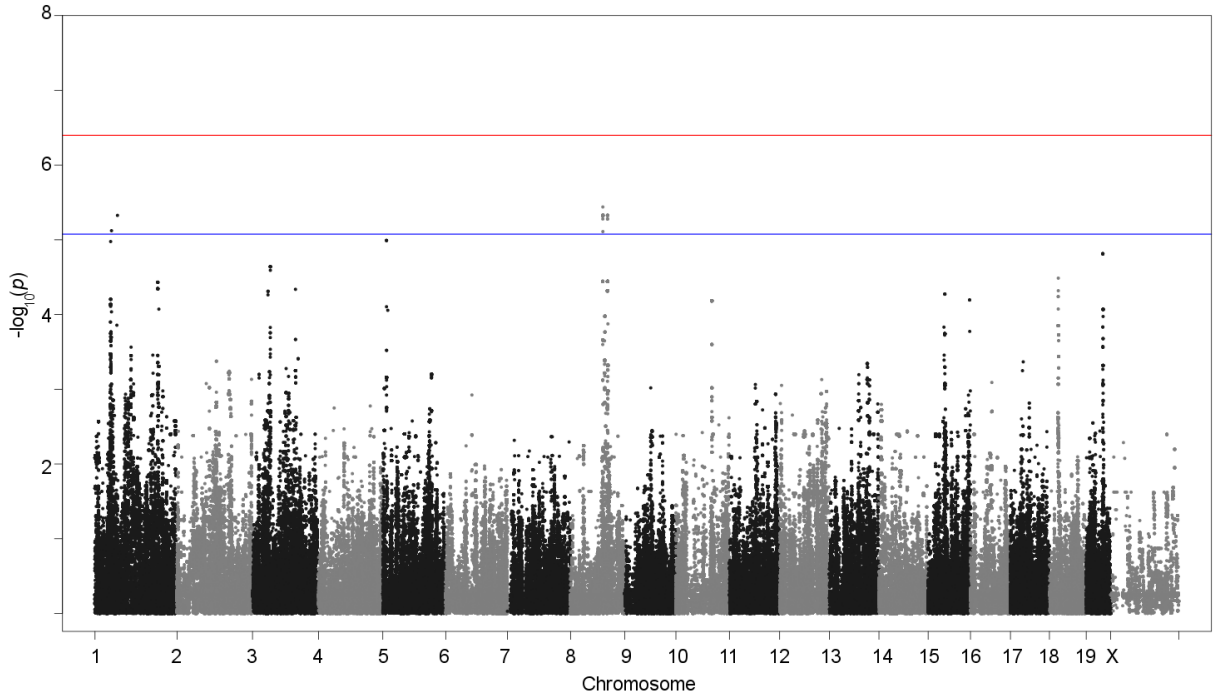

Figure 20. Manhattan plot showing association with 10wkCL. Manhattan plot of associations with 10wkCL. Each dot represents a single nucleotide polymorphism (SNP) with its genomic location represented on the x-axis and the negative log of P-values on the y-axis. The top threshold (red) represents genome-wide significance following Bonferroni correction of the effective number of tests, and the lower threshold (blue) represents genome-wide suggestive significance.

Table 16. Variants suggestively associated with 10wkCL

Chr	BP	P-value	SNP
8	79564934	3.64E-06	mm37-8-79564934:NES14066236
1	56225427	4.71E-06	rs13475877:MAG14282659:NES14282659
8	79142238	4.71E-06	NES16284294
8	79205459	4.71E-06	NES16283916
8	79205519	4.71E-06	NES16283917
8	79208276	4.71E-06	NES16283899
8	79210220	4.71E-06	NES16283843
8	79213262	4.71E-06	NES16283782
8	79214330	4.71E-06	NES16283796
8	79214563	4.71E-06	NES16283729
8	79279933	4.71E-06	NES14067935
8	79348872	4.71E-06	NES14066836
8	79361156	4.71E-06	mm37-8-79361156:rs3676644:NES14066739
8	79368817	4.71E-06	NES14066697
8	79369183	4.71E-06	NES14066705
8	79372040	4.71E-06	NES14057139
8	79372107	4.71E-06	NES14057141
8	79372176	4.71E-06	NES14057142
8	79372644	4.71E-06	NES14057148
8	79373422	4.71E-06	NES14057099
8	79400600	4.71E-06	mm37-8-79400600:NES14056554
8	79400764	4.71E-06	mm37-8-79400764:NES14056557
8	79427254	4.71E-06	NES14056247
8	79427487	4.71E-06	NES14056131
8	79428278	4.71E-06	NES14056135
8	79428474	4.71E-06	NES14056137
8	79429503	4.71E-06	NES14056086
8	79430344	4.71E-06	NES14056049
8	79431489	4.71E-06	NES14055981
8	79432710	4.71E-06	NES14055882
8	79432761	4.71E-06	NES14055883
8	79434009	4.71E-06	mm37-8-79434009:NES14055884
8	79434052	4.71E-06	NES14055885
8	79434147	4.71E-06	NES14055886
8	79438813	4.71E-06	NES14055735
8	79439597	4.71E-06	NES14055745
8	79440171	4.71E-06	NES14066638
8	79511863	4.71E-06	NES14066332
8	79532925	4.71E-06	mm37-8-79532925

Table 16 continued

8	79541874	4.71E-06	mm37-8-79541874
8	79568840	4.71E-06	NES14066217
8	79569211	4.71E-06	NES14066220
8	79570189	4.71E-06	NES14066212
8	79854044	4.71E-06	NES14982349
8	79869287	4.71E-06	NES14982146
8	79870669	4.71E-06	NES14982161
8	79872908	4.71E-06	mm37-8-79872908:NES14982089
8	79872943	4.71E-06	mm37-8-79872943
8	90590642	4.71E-06	NES14061147
8	90593471	4.71E-06	NES14061089
8	90681391	4.71E-06	NES14063803
8	90746561	4.71E-06	NES16265493
8	90778159	4.71E-06	NES14062796
8	90807840	4.71E-06	NES14062258
8	90814685	4.71E-06	NES14062068
8	90815448	4.71E-06	NES14062071
8	79408593	5.27E-06	mm37-8-79408593
8	90623652	5.27E-06	mm37-8-90623652:rs6235292:NES14060695
8	90725209	5.27E-06	mm37-8-90725209:rs6210223
1	42093961	7.55E-06	NES11101025
8	79584455	7.77E-06	mm37-8-79584455:NES14066168

4.3.3 Integration of mouse and human genome-wide scans

To prioritize top loci and detect candidate genes that may contribute both to early life predisposition and to COPD, we integrated the findings of our mouse genome-wide scan with the results from a meta-analysis of human COPD CVAS.

We began by looking for candidate regions in the mouse that were enriched for nominal associations ($P<10^{-3}$). Six regions at 1qB, 1qG, 2qF1, 8qC1, 8qC2 and 8qC3 were identified for integration with human data (Table 17). Importantly, these top regions included the single variants with the lowest P-values in this study (Table 17). Using mouse-human homology maps, publicly

available from NCBI, we used a gene-based approach to identify human regions homologous to these candidate regions on human chromosomes 1, 2, 4, 15, 16 and 17 (Table 17, Figure 21).

Table 17. Mouse candidate regions and syntenic regions in the human genome

Mouse						
#^	Chr	Start	End	Most Significant P-value	SNP	Nearest genes
1	1	38054855	40912112	1.06E-05	rs32434308	
2	1	151808414	155066052	3.70E-05	rs33656443	
3	2	124893863	127988283	5.73E-04	mm37-2-126837474	
4	8	79426407	80918225	3.64E-06	rs33154484	
5	8	82495773	85608713	1.05E-04	rs33218702	
6	8	89365037	91566338	4.71E-06	rs38483583	
Human						
1	2	99953834	103434138	2.23E-04	rs6729857	*IL1R1/IL1R2**
2	1	182992595	186958113	3.89E-02	rs78343664	*OCLM/PDC*
	17	57187308	57232800	5.03E-03	rs12945875	*SKA2*
3	15	48413169	51298097	4.42E-05	rs1843147	*DTWD1/ATP8B4**
	7	43798272	43846941	4.23E-02	rs849185	*BLVRA*
	2	95691479	97041274	2.80E-03	rs1006021	*PROM2*
	2	110841447	111926022	4.60E-04	rs4494791	*ACOXL*
4	4	147628179	149363643	4.01E-04	rs79824712	*TTC29/EDNRA**
5	4	141786725	145659881	1.57E-12	rs13141641	*HHIP**
6	16	48572637	51185183	2.95E-02	rs74017995	*N4BP1*

^Region numbers identify homologous genomic regions in the mouse and human, *Candidate genes based on being nearest a nominal SNP (P<10-3) with lowest P-value in each region

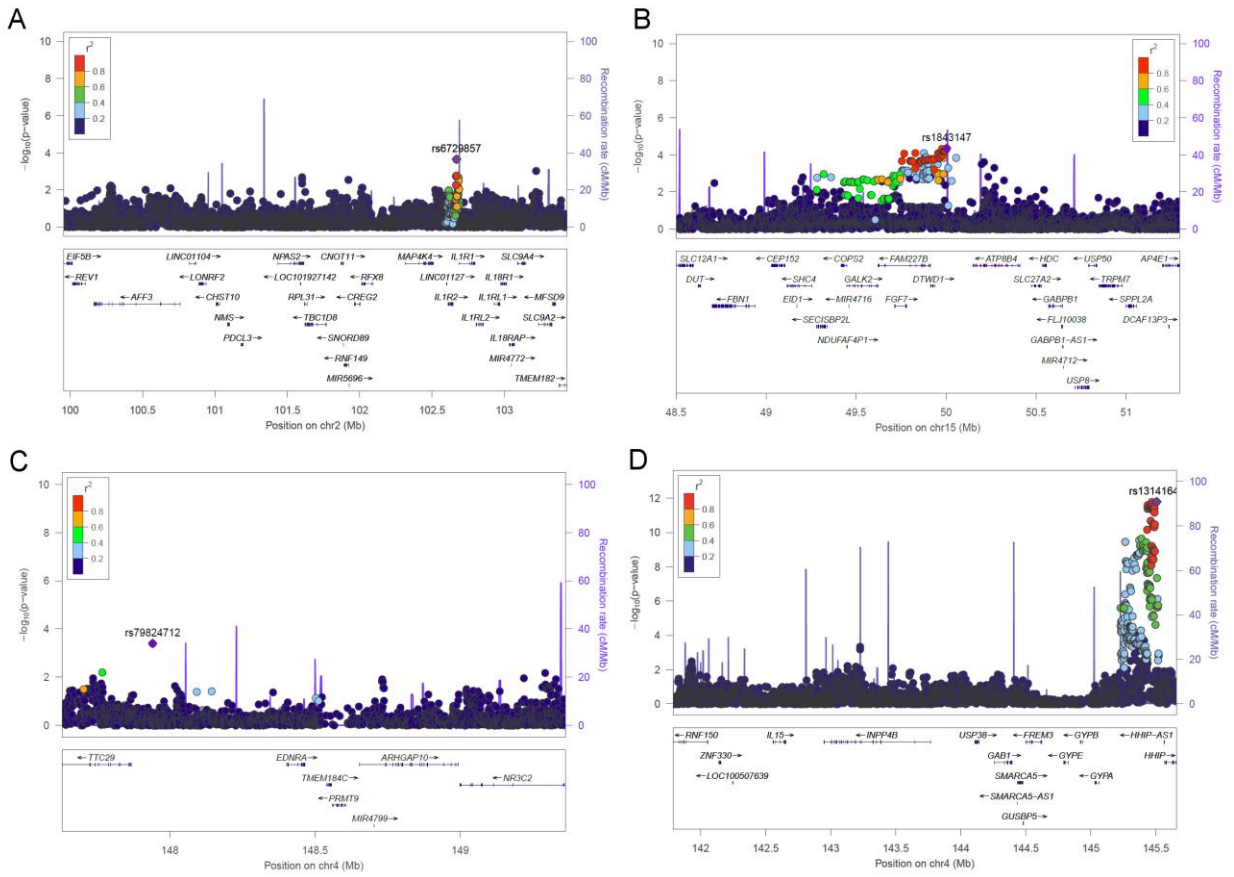

Figure 21. Manhattan plots of results from human GWAS with nominal results. Manhattan plots of associations results in human syntenic candidate regions. Four candidate regions on chromosomes 1, 4 and 15 contain nominal associations with COPD in humans. Manhattan plots showing genomic location and gene annotations on the x-axis and negative log of P-values on the y-axis. The nearest genes to these regions are (A) *IL1R1*/*IL1R2*, (B) *DTWD1*/*ATPB84*, (C) *TTC29*/*EDNRA*, and (D) *HHIP*. Linkage disequilibrium with the most significant SNP as determined from the CEU cohort of 1000 Genomes is illustrated by SNP color. Plot created with LocusZoom (Version 1.1).

4.3.4 Expression of candidate genes varies between extreme strains at 10 weeks of age

Genes containing or surrounding nominal human SNPs from human studies were chosen as candidate genes for further study. These genes were *IL1R1* (mouse homolog:*Il1r1*), *IL1R2* (*Il1r2*), *DTWD1* (*Dtwd1*), *ATP8B4* (*Atp8b4*), *TTC29*, (*Ttc29*), *EDNRA* (*Ednra*), and *HHIP* (*Hhip*). We looked for coding non-synonymous variants (cnSNPs) of functional effect that varied between the extreme strains of our study, C3H/HeJ and NOD/ShiLtJ. None of the seven candidate genes harbor any cnSNPs with SIFT score < 0.05. As variation in regulatory regions of these genes may also be causal, we measured RNA expression of the candidate genes in lung tissue from these extreme strains. We found significant differences in the expression of Il1r2, Dtwd1, Atpb8b4, and Ttc29 between the two strains (Figure 22).

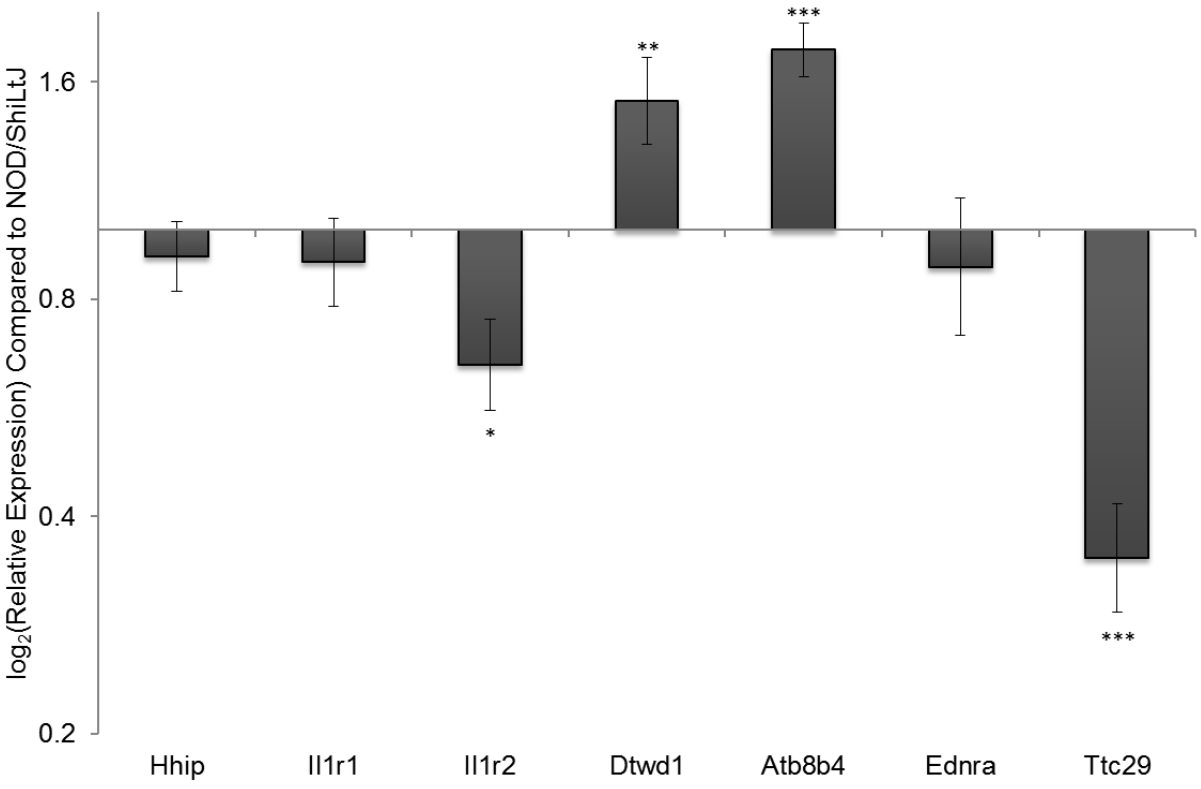

Figure 22. Relative expression of candidate genes in extreme strains. Relative expression of candidate genes in C3H/HeJ mice compared to NOD/ShiLtJ mice (n=10/strain). P-values are determined by two-tailed t-test (* $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$)

4.3.5 C3H/HeJ mice have lower relative expression of Il1r2 in the lung epithelium

We chose to follow up on *Il1r2*, as the IL-1 pathway has been shown to play a role in both prenatal and chronic lung disease but has not previously been implicated in predisposition to adult disease in the absence of prenatal exposures. Since our screening experiment was in whole lung tissue and *Il1r1* and *Il1r2* are highly expressed by innate immune cells, we compared the expression of these genes in perfused, lavaged lung tissue to isolated blood monocytes and alveolar macrophages and found that the only significant change in expression occurred in lung

tissue (Figure 23A). Further, using immunofluorescence, we found IL-1R2 is expressed on alveolar type 1 (AT1) and 2 (AT2) cells as well as alveolar macrophages (Figure 23B, C, and D). However, while C3H/HeJ mice have the same levels of IL-1R2 on AT2 and alveolar macrophages, co-localization with aquaporin 5 (AQP5), a marker of alveolar type 1 cells, was significantly lower than in NOD/ShiLtJ (Figure 23E).

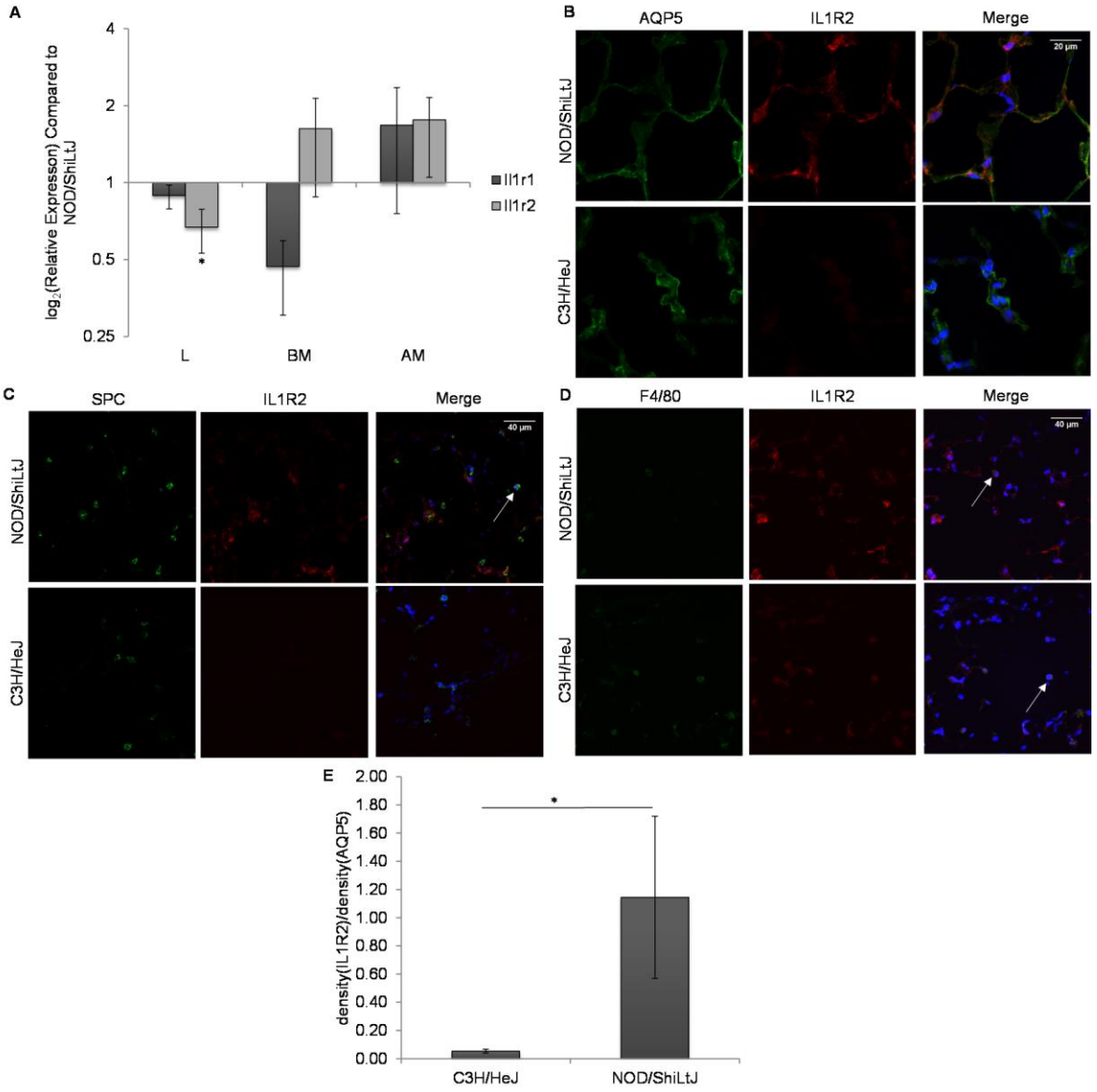

Figure 23. Levels of IL1R2 in the lung epithelium of extreme strains. (A) Lower expression of *Il1r2* in C3H/HeJ mice (n=10/group) compared to NOD/ShiLtJ mice (n=10/group) occurs in lung tissue that has been perfused and lavaged (L) but not in blood monocytes (BM) or alveolar macrophages (AM). Levels of IL1R2 are lower in cells from C3H/HeJ mice (n=8/group) compared to NOD/ShiLtJ mice (n=8/group) co-expressing (B) AQP5, a marker of lung alveolar type I cells, but not (C) SPC, a marker of alveolar type II cells or (D) F4/80, a marker of mature macrophages. (E) Quantification of density of IL1R2 compared to density of AQP5 staining is significantly different between the two strains. All p-values represent two-tailed t-tests between groups (* $P<0.05$)

4.3.6 C3H/HeJ mice have an adequate *Il1r2* response following injury

By integrating our mouse genome-wide scan of predisposition to COPD with a human genome-wide scan of COPD we attempted to identify genes contributing to developmental predisposition to disease, so we were interested in determining whether the regulation of *Il1r2* in extreme strains was also altered during lung injury. Following exposure to lipopolysaccharide (LPS) and exposure to a single cigarette, we observed a significantly higher expression of *Il1r2* in C3H/HeJ mice compared to control-treated mice (saline or room air respectively) (Figure 24A). There was no significant change in NOD/ShiLtJ mice (Figure 24A). When exposed to two weeks or six months of cigarette smoke, C3H/HeJ mice did not maintain this acute upregulation of *Il1r2* (Figure 24A). Following chronic cigarette smoke exposure over the course of six months (a mouse model of emphysema) C3H/HeJ mice, which have larger 10wkCL, also developed significantly greater airspace enlargement (Figure 24B and C).

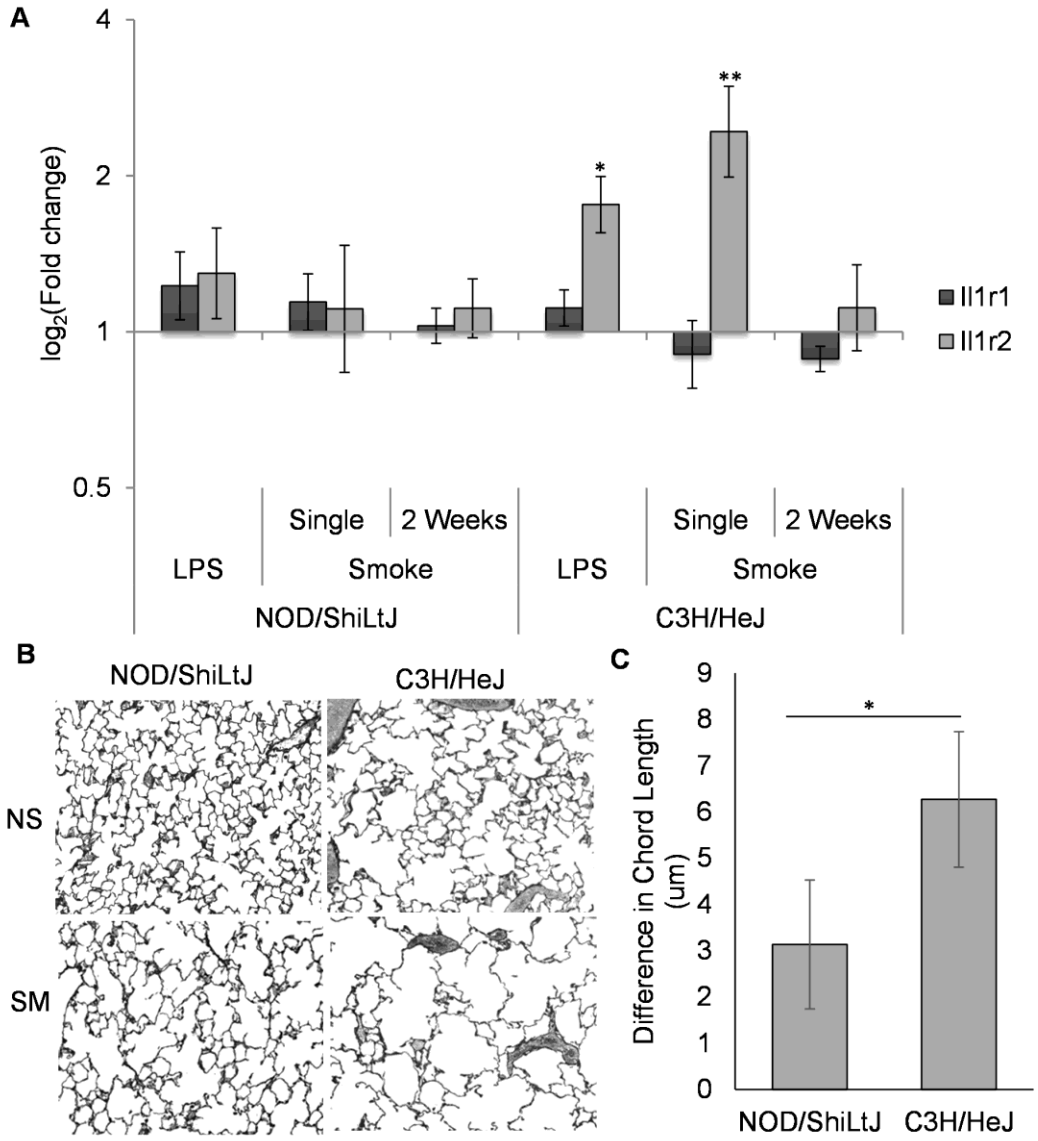

Figure 24. Extreme strain response to inflammatory stimuli. C3H/HeJ mice respond to acute stimuli with upregulation of *Il1r2* but are more susceptible to chronic cigarette smoke exposure. (A) C3H/HeJ mice exposed to lipopolysaccharide (LPS) or a single cigarette significantly upregulate *Il1r2* while NOD/ShiLtJ mice do not (n=5/strain/group). Neither strain shows a significant difference in expression of Il1r2 after two weeks of cigarette smoke exposure When exposed to six months of chronic cigarette smoke (SM), C3H/HeJ mice have a significantly higher difference in chord length between SM and room air (NS) compared to NOD/ShiLtJ mice as seen in (B) Gill staining and (C) quantified by morphometry (n=6/strain/group).

4.3.7 *Il1r1*$^{-/-}$ mice have the same 10wkCL as C57BL6/J controls

To determine whether an absence of IL-1 signaling was responsible for the differences we observed in our extreme strains, we compared the alveolar chord lengths of IL1R1$^{-/-}$ mice to their background controls (C57BL/6J) at ten weeks of age and saw no significant difference in this phenotype (Figure 25).

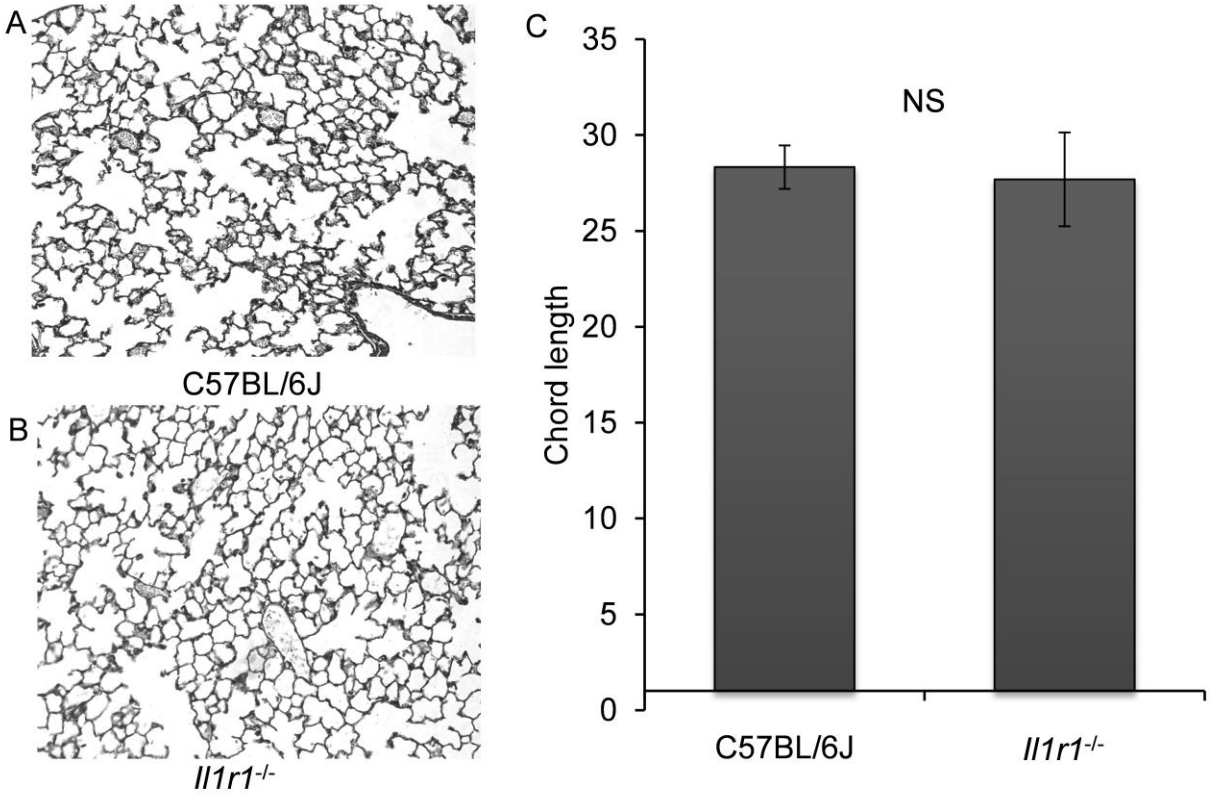

Figure 25. Chord length of IL1R1$^{-/-}$ mice at 10 weeks of age. Chord length of *Il1r1*$^{-/-}$ mice at 10 weeks of age is not significantly different than C57BL/6J control mice. Gill stained example images of lung parenchyma of (A) C57BL/6J (n=5) and (B) IL1R1$^{-/-}$ mice (n=6). (C) When measured by morphometry, chord length is not significantly different between these strains (by two-tailed t-test).

4.3.8 C3H/HeJ mice exposed to anakinra *in utero* have significantly smaller 10wkCL than saline-exposed controls

We exposed pregnant C3H/HeJ mice to anakinra, a clinically utilized IL-1 antagonist, during the course of their pregnancy and measured the alveolar chord length of their pups at 10 weeks of age (Figure 26A). C3H/HeJ mice exposed to anakinra *in utero* had a significantly smaller 10wkCL (33.1 ± 1.2 μm) than C3H/HeJ mice exposed to saline *in utero* (38.2 ± 1.0 μm, Figure 26B and C).

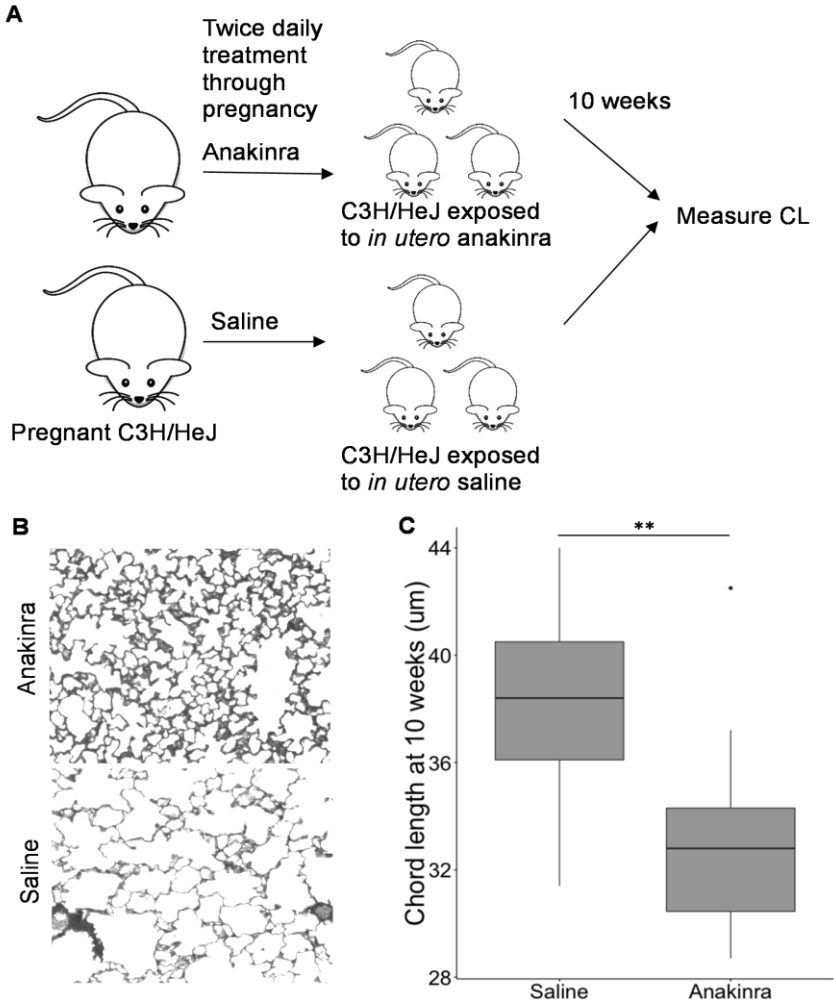

Figure 26. 10wkCL in C3H/HeJ mice exposed to in utero anakinra. (A) C3H/HeJ mice were exposed to *in utero* anakinra or saline. (B) Representative Gill stained images of C3H/HeJ mice exposed to either *in utero* anakinra or saline. (C) C3H/HeJ mice have a lower chord length following in utero anakinra exposure (n=11) compared to saline exposed controls (n=11). All p-values represent two-tailed t-tests compared between experimental group and relevant controls (* $P<0.05$, ** $P<0.01$).

4.4 DISCUSSION

In this study, we sought to leverage phenotypic differences between inbred strains of mice to identify genetic loci associated with lung development and growth and to determine whether these genetic loci were likely to be related to human COPD. To do this, we measured mean linear intercept (CL) in 36 inbred strains of mice at 10 weeks of age. CL is a stereological measurement that grossly estimates airspace size and architecture, and has been successfully used to model lung structural changes following cigarette smoke exposure (27, 263). Although alveolarization continues at a low level throughout life in mice, by far the majority of lung development and growth has finished by ten weeks of age (264, 265). Thus, we hypothesized that genetic factors contributing to this phenotype were likely to be major regulators of lung growth and development.

We conducted a genome-wide scan testing for association with 10wkCL and integrated these findings with the results of a previously published CVAS of COPD in humans. In so doing, we aimed to identify loci that contributed to both lung development and risk of COPD. The identification of one of our candidate genes, *Hhip*, strongly supported this approach. A locus near HHIP was identified in the first GWAS of COPD and pulmonary function and has been replicated in numerous studies (111, 259). The protein product of this gene, hedgehog-interacting protein (HHIP) is known to directly regulate the signaling molecule sonic hedgehog (SHH), an essential component of early lung development (260). In fact, HHIP is essential for murine lung development, as homozygous deletion of the gene causes respiratory failure at birth and haploinsufficiency of the gene predisposes to both cigarette smoke-induced emphysema and age-related emphysema (150, 151). Protein and transcript levels of HHIP have been shown to be altered

in COPD compared to non-COPD controls, and are correlated with variants in an upstream enhancer (149).

Another candidate region identified in our study contains a nominal association with rs6729857, a SNP located between *IL1R1* and *IL1R2*. There is a notable recombination peak separating this association and *IL1R1* in the human, suggesting that if this is not a directly causal SNP, nearby tagged SNPs are more likely to be affecting *IL1R2*. This was further supported by the observation of a significant difference in *Il1r2* but not *Il1r1* expression between the two extreme strains. IL-1R2 is a classic decoy receptor for the cytokine interleukin 1 (IL-1), a primary mediator of both infectious and sterile inflammation (266). The protein is structurally similar to IL-1R1 which mediates signaling of both IL-1β and IL-1α, but lacks an intracellular signaling domain and thus serves as a potent antagonist to the pathway (266).

Evidence exists for a role of IL-1 signaling in both lung development and disease. Mice exposed to human IL-1β *in utero* have a significantly increased chord length at seven days of age due to disrupted alveolar septation (267). Similarly, in chronic disease, transgenic mice induced to express human IL-1β postnatally develop an immune phenotype similar to COPD and ultimately emphysema as measured by an increase in chord length (265). Mice lacking IL-1R1, the primary signaling receptor for both IL-1β and IL-1α, are largely protected from cigarette smoke-induced emphysema (268). Support for the role of IL-1 has also been shown in human disease – peripheral neutrophils from individuals with COPD have been shown to have increased expression of *IL1B*, *IL1R2* and *IL1RA*, and such expression correlates with disease severity (269). In development, excess IL-1 has been seen to result in both early maturation and acute damage to the lung leading to simplification of lung structure and ultimately resulting in respiratory distress system or bronchopulmonary dysplasia (270).

Given the significant evidence supporting the role of IL-1 signaling in disease, we were intrigued to identify *Il1r2* as a candidate gene in this study since all of the candidate loci were identified using a gross measurement of growth and development in mice raised in sterile conditions. We found lower levels of the transcript and protein in C3H/HeJ mice that have larger 10wkCL compared to NOD/ShiLtJ mice, and observed that this difference is due to decreased IL-1R2 levels in the lung ATI cell. When exposed to stimuli, C3H/HeJ mice significantly upregulate *Il1r2* acutely while NOD/ShiLtJ mice do not, although they do not appear to be able to maintain this increased regulation chronically. Finally, C3H/HeJ mice exposed to cigarette smoke have increased airspace enlargement compared to NOD/ShiLtJ mice, indicating this predisposition to low IL1R2 levels contributes both to predisposition to large airspaces and is worsened in disease states.

Based on these findings, we hypothesized that genetic variants resulting in an insufficient level of IL-1 antagonism, as seen in the low levels of expression of *Il1r2* in the C3H/HeJ mouse, could result in the same increased CL observed in mice exposed to *in utero* inflammation or IL-1 beta upregulation. After exposing pregnant C3H/HeJ mice to anakinra, a recombinant therapeutic based on IL-1Ra, we observed that pups exposed *in utero* to the drug had a significantly smaller chord length compared to C3H/HeJ mice exposed to saline *in utero*. We did not observe any other gross morphological differences between these mice, indicating a lung-specific effect, most likely due to the large difference in expression of the protein in the lung epithelium of C3H/HeJ mice. This is the first time lung development of an inbred strain has been modulated therapeutically and implicates IL-1 signaling in lung development in the absence of disease.

The implications of these findings for human populations are intriguing. Although a number of variants predicted to be deleterious by PolyPhen or SIFT in *IL1R2* have been identified

in human populations, they are rare (See Table E2). Rare variants could potentially produce a synthetic association at this locus, which could also explain why association tests did not reach genome-wide significance (167). More likely, variants in regulatory regions of *IL1R2* affect lung development or growth in humans but this effect is difficult to detect for reasons affecting all CVAS, including varied environmental effects and phenotypic heterogeneity (116, 117). Of note, *IL1RN*, encoding the soluble IL-1 antagonist IL-1Ra that anakinra is based on, has been associated with COPD risk in several populations and variants in the same gene have been associated with lung function decline in smokers and asthmatics (108, 271-273). One possibility, then, is that variants in one or more genes function to antagonize IL-1 signaling and may jointly contribute to this phenotype.

While the literature strongly supports the role of *HHIP* in lung development and disease, and evidence presented here supports a similar function for *IL1R2*, additional function for the other candidate genes identified in this study cannot be ruled out. Several genes were differentially expressed between the two extreme strains. Further, while we chose to use the extreme strains of our study to identify genomic differences with the most dramatic effect on phenotype, it remains possible that variation in any of the candidate genes could affect other strains with large or small chord 10wkCL. *EDNRA*, the receptor for endothelin signaling, is an intriguing candidate, as variants in another endothelin receptor, *EDNRB*, have been shown to associate with obstruction in asthma (274). Moreover, although we chose to focus on the genes nearest nominal SNPs from human studies of COPD, some of these SNPs are in LD with variants in or near other potential candidate genes. An excellent example of this in our study occurs on human chromosome 15, where a nominal association with rs1843147, between the genes *DTWD1* and *ATP8B4*, is in high LD with variants in *FGF7*, a well-recognized mediator of alveolar growth and differentiation (275,

276). Supporting this finding, in another approach intended to identify candidate genes falling below strict multiple testing correction, *FGF7* was identified as a susceptibility locus for COPD (277). Interestingly, there was significantly higher relative expression of both *Dtwd1* and *Atp8b4* in C3H/HeJ mice compared to NOD/ShiLtJ mice, suggesting potential chromatin-level regulation of genes surrounding this locus.

This study has limitations. The phenotype chosen to represent lung growth and development is a relatively gross measure of these effects, and so this approach is only likely to identify genes with a significant effect on this quantitative phenotype. Similarly, combining these results with the results of human GWAS of COPD reduces the ability to detect regions that may be involved in lung growth but not with the disease itself. While in combination, using traits with significant phenotypic heterogeneity likely allowed for the detection of well-known regions near *HHIP* and *FGF7* in this study, it is also the likely reason that individual studies were underpowered to detect other candidate genes at a genome-wide significance in humans or mice.

4.5 CONCLUSIONS

In this study, we used a stepwise approach to identify candidate genes associated with a developmental predisposition to emphysema in the mouse and COPD in humans. One of our candidate genes, *Hhip*, has previously been shown to be associated with phenotypes of pulmonary function in healthy and diseased individuals. We further investigated another candidate gene, *Il1r2*, and showed that correction of a genetic predisposition to decreased expression of this IL-1 antagonist in C3H/HeJ mice also decreases that strain's 10wkCL. This study suggests the need for

further investigation into early-life genetic factors contributing to lung disease and for an improved understanding of the role of the IL-1 axis in the relationship between lung growth and disease.

5.0 INTEGRATED REVIEW OF GENETIC STUDIES OF COPD

5.1 INTRODUCTION

The genetic architecture of COPD is complex, likely involving multiple genes, a balance between susceptibility and protective variants, and gene-gene interactions all potentially affected by environmental exposures (see section 1.2.6). As outlined in section 1.2, in the last 20 years, our knowledge of genomic loci that contribute to the heritability of chronic diseases like COPD has increased significantly but remains incomplete, in part due to this complexity. In this work, we have already explored some of the architectural features of the genetics of COPD, including loci affected by rare variation (section 2) and by common variation not detected by current CVAS (section 3 and 4). In the latter approach, we integrated mouse and human genome-wide scans in order to identify loci that are associated with disease in humans and models of disease in mice. This approach is supported by studies that demonstrate that there is overlap between genes that contribute to both common human diseases and murine genetic models of disease (236, 243). Similarly, common variants, rare variants and copy number variants can contribute to human disease at the same disease-relevant loci (278). Thus, as illustrated in section 2.4, the significant store of knowledge of susceptibility loci that has been established over the last 20 years can be a useful tool in analyzing novel studies. Looking at the results of these studies in an integrated way can identify candidate genes at genomic loci of previously unclear significance, and can aid in prioritizing targets for future studies.

To demonstrate this approach in COPD, I conducted a literature review of the results of all published linkage analyses, CVAS, and RVAS for COPD-related traits and integrated them with the results of the studies reported in sections 2, 3, and 4 in a visual and tabular format.

5.2 METHODS

5.2.1 Identification of loci demonstrating linkage with COPD

The US National Library of Medicine service PubMed was searched for studies using the search term: '(genome-wide OR genomewide) AND linkage AND (copd OR emphysema OR chronic bronchitis)' in May, 2016 returning 58 results. Studies were included if they involved tests for genome-wide linkage of a phenotype of COPD measured by spirometry, clinical definition or imaging.

Polymorphic markers with suggestive linkage to a COPD phenotype (LOD > 2.0) from each of these studies were recorded and the physical locations of markers in GRCh37 were identified using the UCSC genome browser (252). When possible, the genomic region for further analysis was recorded as the physical distance between all suggestive markers. When only a single marker was available, the region was estimated as the physical location of the marker and a 3.85Mbp flanking region in either direction, based on gross estimates of sex average recombination rates in the human genome of 1.3 cM/Mb and the average genetic distance of 10cM between markers in these studies (279).

5.2.2 Retrieval of results of COPD CVAS

The results of the NHGRI-EBI GWAS catalog (version e84_r2016-06-12) were downloaded from the European Molecular Biology Laboratory-European Bioinformatics Institute's (EMBL-EBI) website (https://www.ebi.ac.uk/gwas/home). This catalog includes all SNP-trait associations with $P < 1 \times 10^{-5}$ (110). SNPs yielding suggestive association with the following traits were recorded: 'Pulmonary function', 'Chronic obstructive pulmonary disease', 'Emphysema-related traits', 'Age at smoking initiation in chronic obstructive pulmonary disease', 'Pulmonary function decline', 'Airflow obstruction', 'Smoking cessation in chronic obstructive pulmonary disease', 'Lifetime average cigarettes per day in chronic obstructive pulmonary disease', 'Current cigarettes per day in chronic obstructive pulmonary disease', 'Lung function (forced expiratory volume in 1 second)', 'Lung function (forced vital capacity)', 'Lung function (forced expiratory volume in 1 second to forced vital capacity ratio)', 'Lung function (forced expiratory flow between 25% and 75% of forced vital capacity)', 'Resting oxygen saturation in chronic osbtructive pulmonary disease (pulse oxymetry)', 'Pulmonary emphysema', 'Chronic obstructive pulmonary disease (severe)', 'Chronic obstructive pulmonary disease (moderate to severe)', 'Forced expiratory volume in 1 second', 'Asthma or chronic obstructive pulmonary disease', 'Chronic bronchitis in chronic obstructive pulmonary disease', 'Chronic bronchitis and chronic obstructive pulmonary disease', 'Airway responsiveness in chronic obstructive pulmonary disease', 'Emphysema imaging phenotypes', and 'Airway imaging phenotypes'. The location of each of these variants was identified in GRCh37 using Ensembl's archived Biomart (Ensembl Variation 75) (261).

5.2.3 Literature review to identify results of COPD RVAS

PubMed was searched for studies using the search term: ‘rare variant association study copd OR exome sequencing copd OR whole genome sequencing copd’ in May, 2016 returning 23 results. Studies were included if they involved tests for genome-wide association of rare variation using exome sequencing (3 studies), exome array analysis (2 studies) or whole genome sequencing (0 studies). Suggestive or significant associations with single variants were recorded as the physical location of the variant in GRCh37, suggestive or significant associations with genes were recorded as the start and end of the gene in GRCh37.

5.2.4 Identification of regions that are in close physical proximity

Using Python, all genomic loci identified using the methodologies described in 5.2.1, 5.2.2, 5.2.3, as well as the top ten most significant results from each of the three group-wise tests for association with rare variation described in section 2, the human genomic region homologous to the most significant result from section 3, and all human genomic regions homologous to suggestively associated results from section 4, that were within 1Mbp from each other were identified. Groups of variants reported in Table 17 only include those with multiple types of genetic evidence.

All genomic loci were also plotted using Circos (version 0.69-2) (280).

5.3 RESULTS

From six linkage studies, 24 CVAS and five RVAS that met the criteria described in the methods, I identified 20 regions from linkage analyses, 374 loci yielding at least suggestive association from CVAS, and 18 loci yielding suggestive association from RVAS. In order to identify loci with robust evidence for a role in COPD, I combined these loci with the most significant results of the studies in sections 2, 3, and 4 (see Methods) by identifying all loci that were within 1Mbp of another locus. In total, I identified 57 regions that have at least two loci within 1Mbp of another. The majority of these are replication of SNPs in multiple CVAS, so the 19 loci at which multiple types of evidence were present are shown in Table 18. These results can also be seen visually in Figure 27.

Table 18. Genomic loci with multiple pieces of evidence for contribution to COPD susceptibility

Region	Evidence	SNP/Gene	Chr	Start	End	Phenotype
1p	linkage		1	82895173	98685573	Post-bronchodilator FEV1/FVC (98)
	common variant	rs3843306	1	91288130	91288130	Pulmonary function decline (281)
	linkage		1	98685301	106281721	Post-bronchodilator FEV1 (98)
	common variant	rs17121403	1	100335977	100335977	Pulmonary function decline (281)
2q	linkage		2	214579787	232206614	FEV1/FVC (94)
	common variant	rs10932600	2	216114257	216114257	Airway imaging phenotypes (139)
	linkage		2	218246863	238078675	Post-bronchodilator FEV1/FVC (98)
	linkage		2	218576474	227029884	FEF25-75, smokers (95)
	common variant	rs2571445	2	218683154	218683154	Pulmonary function (128)
4p16	linkage		4	11267900	18968275	FEV1/FVC in the general population (282)
	common variant	rs3893377	4	15742129	15742129	Current cigarettes per day in COPD (124)
4q24	common variant	rs1541374	4	106048360	106048360	Pulmonary function (121)
	rare variant array	*TET2*	4	106067842	106200960	FEV1 (141)
	common variant	rs1982346	4	106578754	106578754	Lung function (FEV1) (283)
	common variant	rs11727189	4	106619140	106619140	Pulmonary function (123)
	common variant	rs10516526	4	106688904	106688904	Pulmonary function (128)
	rare variant array	*NPNT*	4	106816597	106892828	FEV1 (141)
	common variant	rs11097912	4	107000462	107000462	Airflow obstruction (131)
4q28	common variant	rs142200419	4	127323308	127323308	Airway imaging phenotypes (139)
	sliding window		4	128040000	128069999	Emphysema

Table 17 (continued)

4q31	10wkCL		4	141786725	145659881	Predisposition to lung disease
	common variant	rs13105210	4	145262927	145262927	Chronic obstructive pulmonary disease (284)
	common variant	rs12510916	4	145372248	145372248	Chronic obstructive pulmonary disease (284)
	common variant	rs12504628	4	145436324	145436324	Pulmonary function (128)
	common variant	rs138641402	4	145445779	145445779	Chronic obstructive pulmonary disease (284)
	common variant	rs13147758	4	145460230	145460230	Pulmonary function (112)
	common variant	rs1828591	4	145480780	145480780	Chronic obstructive pulmonary disease (285)
	common variant	rs1980057	4	145485738	145485738	Pulmonary function (123)
	common variant	rs13118928	4	145486389	145486389	Chronic obstructive pulmonary disease (113)
6p21	rare variant array	*HLA-DQB1*	6	32627241	32634466	FEV1 (141)
	common variant	rs2647044	6	32667910	32667910	Pulmonary function (121)
	common variant	rs2647050	6	32669767	32669767	Emphysema imaging phenotypes (139)
	rare variant array	*HLA-DQA2*	6	32709163	32714664	FEV1(141)
	common variant	rs9394152	6	33465482	33465482	Age at smoking initiation in COPD (124)
	common variant	rs7747216	6	33476718	33476718	Age at smoking initiation in COPD (124)
	common variant	rs9296092	6	33478496	33478496	Age at smoking initiation in COPD (124)
6p21	WES	*DNAH8*	6	38683117	38998574	Severe, early onset COPD (182)
	common variant	rs2395730	6	39784365	39784365	Pulmonary function (128)
6q	linkage		6	150000000	171115067	FEV1 (97)
	common variant	rs3734729	6	150570867	150570867	Pulmonary function (121)

Table 17 (continued)

8p22	common variant	rs2638663	8	18673731	18673731	FEV1 (283)
	sliding window	*CSGALNACT1*	8	19560000	19589999	Emphysema
	gene-based	*ATP6V1B2*	8	20054059	20084860	Emphysema
8p11	common variant	rs7006290	8	41615138	41615138	Pulmonary function decline (281)
	sliding window		8	42060000	42089999	Emphysema
10q26	WES	*TACC2*	10	123748689	124014057	Severe COPD (181)
	common variant	rs10794613	10	124514138	124514138	Smoking cessation in COPD (124)
	common variant	rs1896376	10	125515829	125515829	Smoking cessation in COPD (124)
12p	linkage		12	8157605	17015211	FEF25-75, smokers (95)
	common variant	rs2856329	12	11942720	11942720	Lung function (forced expiratory volume in 1 second) (283)
	linkage		12	12846869	20547246	FEV1 (94)
	linkage		12	13164928	20865211	Moderate airflow obstruction (93)
	linkage		12	13606455	16696869	Mild airflow obstruction (93)
	non-synonymous	*PTPRO*	12	15654574	15656846	Emphysema
	linkage		12	17889218	25589539	FEV1/FVC (94)
	common variant	rs11044734	12	19795906	19795906	Lifetime average cigarettes per day in COPD (124)
	common variant	rs11044737	12	19801753	19801753	Lifetime average cigarettes per day in COPD (124)
	common variant	rs4762767	12	19866129	19866129	Pulmonary function (121)
12q23	common variant	*CCDC38*	12	96260826	96336428	Healthy individuals with significant smoking histories (180)
	common variant	rs1036429	12	96271428	96271428	Pulmonary function (121)
	sliding window	*CCDC38*	12	96300000	96329999	Emphysema

Table 17 (continued)

12q24	rare variant array	*RBM19*	12	114251543	114404176	FEV1 (141)
	common variant	rs4767234	12	114637848	114637848	Airflow obstruction (97)
	rare variant array	*TBX5*	12	114846000	114850637	FEV1 (141)
15q25	non-synonymous	*TBC1D2B*	15	78316850	78317844	Emphysema
	common variant	rs8042238	15	78774271	78774271	Chronic obstructive pulmonary disease (136)
	common variant	rs11858836	15	78783277	78783277	Chronic obstructive pulmonary disease (114)
	common variant	rs13180	15	78789488	78789488	Chronic obstructive pulmonary disease (115)
	common variant	rs8034191	15	78806023	78806023	Chronic obstructive pulmonary disease (285)
	common variant	rs8031948	15	78816057	78816057	Airflow obstruction (131)
	common variant	rs28675338	15	78827631	78827631	Lifetime average cigarettes per day in COPD (124)
	common variant	rs2036527	15	78851615	78851615	Pulmonary function (121)
	common variant	rs503464	15	78857896	78857896	Emphysema imaging phenotypes (139)
	common variant	rs17486278	15	78867482	78867482	Airflow obstruction (131)
	common variant	rs55676755	15	78898932	78898932	Emphysema imaging phenotypes (139)
17q21	linkage		17	42823055	50523346	Post-bronchodilator FEV1/FVC (98)
	rare variant array	*KANSL1*	17	44107282	44119280	FEV1 (141)
19q13	common variant	rs11668505	19	48348363	48348363	Lung function (forced vital capacity) (283)
	common variant	*TMEM143*	19	48845944	48866796	Pulmonary function decline (281)
	gene-based	*BCAT2*	19	49298537	49314089	Emphysema
19q13	gene-based	*ZNF816*	19	53429414	53466404	Emphysema
	sliding window	*ZNF816*	19	53430000	53459999	Emphysema

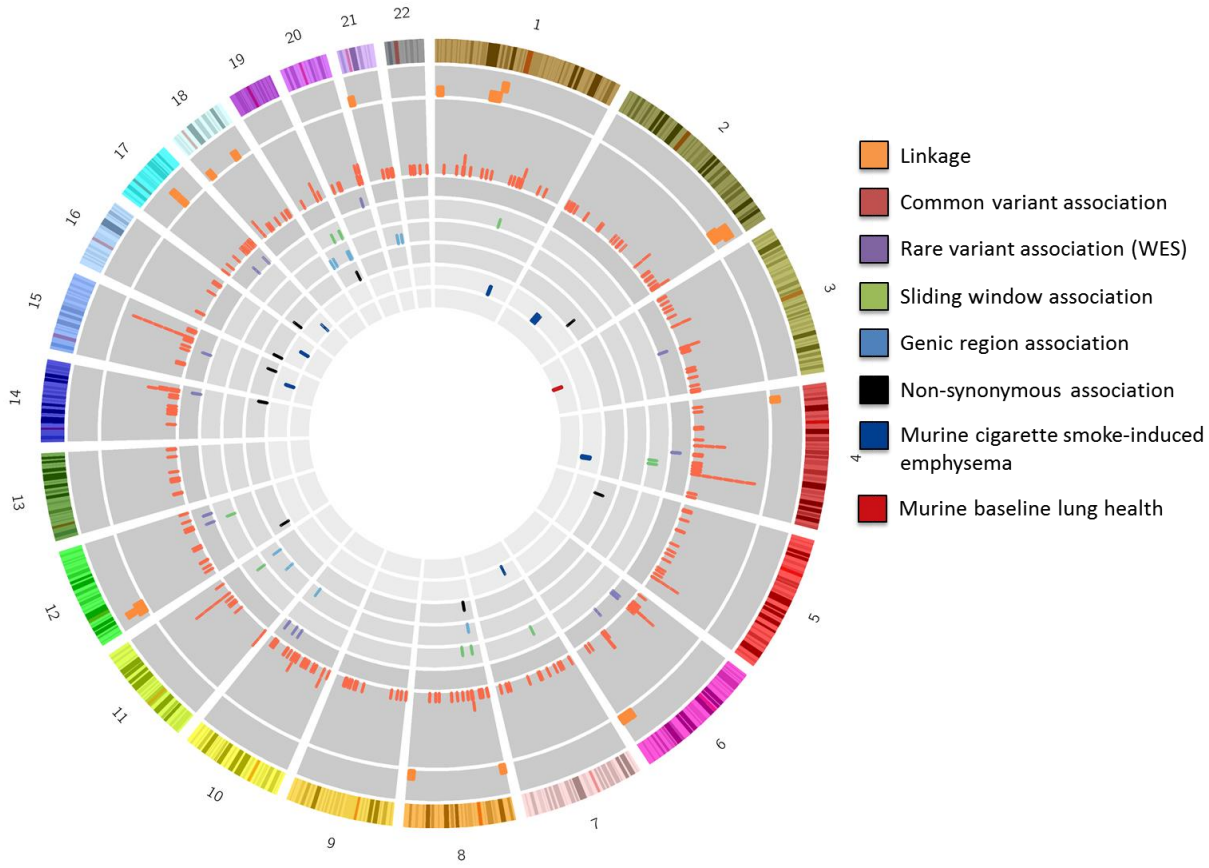

Figure 27. Circos plot of genomic locations of results of genetic studies of COPD. Circos plot demonstrating the genomic location of genome-wide suggestive results from several types of genetic studies. The outer track identifies genomic location and chromosomal banding patterns. The inner tracks represent different types of evidence mapped to their genomic location.

5.4 DISCUSSION

The analysis of all three studies in this work relies on the results of previously published genetic data to offer support for associations that do not meet strict multiple testing correction thresholds. To demonstrate how this approach can be used to prioritize loci, I conducted a literature review

and knowledge-based integration of previously published suggestive evidence from studies looking at genetic susceptibility to COPD, aiming to identify loci supported by more than one type of genetic evidence. There are 19 loci that have suggestive evidence from multiple different techniques and that contribute to either pulmonary function or COPD phenotypes within 1Mbp of another such region. For example, the well-studied region at 4q31 near the gene *HHIP*, was identified in our mouse study of predisposition to lung disease (section 4). Similarly, a frequently-replicated locus at 15q25 near the genes *CHRNA3/5* and *IREB2*, was also near one of the top results of our non-synonymous exome-wide test for association with emphysema (section 2). Almost all of the results of linkage analyses are supported by common variant associations, and some are supported by additional evidence such as 12p, which also contains the gene *PTPRO*, the top association in our non-synonymous exome-wide test in section 2. *PTPRO* is less than 1Mbp from D12S1715, the STR with the highest LOD score in multiple COPD-related traits (Table 2) (94).

Of interest, several regions identified in this approach occur in close proximity to one another at 4q24, 4q28, and 4q31. Based on the phenotypes connected to each of these regions, individual loci appear to uniquely associate with different COPD-related traits: pulmonary function, emphysema, and COPD respectively. This region may represent a larger 'disease susceptibility region' (section 4.4) where multiple susceptibility loci act independently in affecting different disease mechanism. Genomic regions with complex contribution to susceptibility like this one have been identified in other complex diseases (234).

This was a simple approach to demonstrate that there are numerous genomic locations that are enriched for COPD susceptibility variants. I only accounted for physical distance between regions, an imprecise metric of the likelihood that two loci may be contributing to the same trait. A better approach would be to use an LD-based approach in specific populations. This would be

confounded by the variable populations in which literature results are reported and the possibility of synthetic associations, in which common variants tag rare variants outside of the range of LD between common variants (167). Finally, the physical distance used here is arbitrary and affected the reported results. For example, associations near *MMP12* at 11q22, which were discussed in section 2 were not identified since the sliding-window based region result at 11q22 was greater than 1 Mbp away from variants between *MMP12* and *MMP3* identified in CVAS.

5.5 CONCLUSIONS

Using a literature-curated set of suggestively significant results from human genetic studies of COPD, I highlighted a number of genomic loci with strong supportive evidence for their contribution to disease susceptibility. In addition to offering support for the hypothesis presented in sections 1.2.7, 2, 3, and 4, that the genetic architecture underlying COPD is complex, this approach also prioritizes genomic loci for future studies. Few studies have been conducted to identify the biological function of susceptibility loci, due in part to difficulties in identifying causative variants or genes (section 1.2.6). Prioritizing loci using a combination of more than one genetic technique may be a useful approach to narrow specific causative genes. In addition, while candidate gene studies (see section 1.2.4) are less practical now that NGS techniques are widely accessible, one goal of precision medicine for COPD is early stratification of disease risk using genomics. Candidate gene panels, like those used in cancer diagnosis, could be used for this purpose but will require a limited number of genomic loci for testing and this approach could aid in prioritizing susceptibility loci for further testing.

6.0 IMPACT AND FUTURE DIRECTIONS

Our current understanding of the genetic basis of COPD, like other complex diseases, is incomplete (see section 1.2). This work employs techniques not previously used in COPD to implicate susceptibility loci and candidate genes, and offers support for the hypothesis that both rare variants and common variants not detected by current CVAS contribute to 'missing heritability' in COPD. These results will require replication and biological validation. Some possible approaches to future studies are outlined below.

6.1 REPLICATION STUDIES

In section 2, we identified suggestive associations with emphysema using region-based, gene-based and non-synonymous gene-based group-wise tests. None of these loci reached genome-wide significance – a probable explanation is a lack of power based on the small number of individuals included in the study. Although these findings are supported by previous genetic studies, replication in an independent cohort will offer the strongest support for their contribution to the susceptibility of COPD. Targeted sequencing of a panel of the top loci from this study is possible, but recent decreases in WGS cost suggests that a more effective approach is a separate WGS study on an independent population of individuals with COPD (286). Currently, the National Heart, Lung and Blood Institute (NHLBI) at the National Institutes of Health (NIH) is conducting a large WGS study, sequencing individuals with atrial fibrillation, asthma, COPD, obesity, and sleep

apnea through the Trans-Omics for Precision Medicine (TOPMed) Program. Nearly 10,000 participants with COPD will be sequenced through TOPMed. This large WGS study of COPD will offer the opportunity to test for replication of the results presented here and in the previously reported WES studies of COPD. While it will also have better statistical power to detect rare variant associations than any study conducted to-date, even such a large study may be underpowered to detect rare variants of moderate effect (208). One approach to solving this problem will be to use the extremely large population studies currently being conducted in countries like England and Iceland, which include sequencing of over 100,000 individuals each (287). When properly combined with electronic health record data, sequencing repositories will allow for rare variant association testing of numerous traits of complex diseases in extremely large populations.

Association of rare, non-synonymous mutations identified in this study should also be replicated in independent cohorts, which can be done without sequencing. As described in section 1.2, several large cohorts of individuals with COPD exist and would be ideal for this purpose.

6.2 CANDIDATE GENES

Candidate genes identified in genetic studies have successfully informed our understanding of the biological underpinnings of many diseases. In COPD, follow-up studies on candidate genes have begun to give hints to the roles of genes such as *HHIP* and *HTR4* that are near loci implicated in CVAS (see section 0) (149, 150). This work identifies a number of candidate genes in COPD, and some suggested experiments to further investigate their roles are described below.

6.2.1 *ABI3BP*

ABI3BP is an extracellular matrix-binding protein that contributes to major cellular processes, including proliferation and cellular differentiation (257, 258). Tumor cells decrease expression of the gene, in order to decrease senescence and escape cell death (256). In section 3, we demonstrated a significantly higher relative expression of *Abi3bp* in mice that are susceptible to cigarette smoke (A/J) than in mice that are resistant (CBA/J) (Figure 18). Connecting these two findings, a reasonable hypothesis is that A/J mice are more susceptible to cigarette smoke because of increased expression of *Abi3bp* and decreased protection from cell death. As mentioned in 1.1.2, senescence and apoptosis are likely to contribute to the severity and progression of COPD.

Several experiments could be conducted to test this hypothesis. Cell lines with variable *ABI3BP* gene expression could be treated with cigarette smoke extract (CSE) and apoptosis measured. Correlation between *ABI3BP* expression and apoptosis would give preliminary support for the gene's contribution to apoptosis in a COPD model. Similarly, knockdown of *ABI3BP* using siRNA *in vitro* would be expected to reduce apoptosis. *In vivo* analysis could be conducted by producing mice that transgenically upregulate *Abi3bp*. Comparing chord lengths of these mice to background controls following chronic cigarette smoke exposure would support the findings shown here in A/J and CBA/J mice but removes potential confounders related to the complex genetic background of inbred strains.

6.2.2 *PTPRO*

We identified the gene *PTPRO* using a group-wise test of non-synonymous rare variants grouped by gene across the exome in our extreme-trait study of emphysema (section 2). Protein tyrosine phosphatase, receptor-type O (PTPRO) is a known tumor suppressor and its inactivation via DNA methylation occurs in numerous tumors (231, 232). In breast cancer, PTPRO dephosphorylates ERBB2 (*HER2*) leading to its degradation and decreased tumor growth (288). Interestingly, similar enzymatic specificity has been observed in liver inflammation, where decreased *PTPRO* results in decreased NF-κB signaling via an ErbB2/Akt/GSK-3β/β-catenin cascade (289). Although little is known about the role of the gene in the lung, global *PTPRO* DNA methylation patterns are significantly altered following prenatal cigarette smoke exposure, indicating a role for epigenetic alterations of the gene in smoke exposure (233, 290). We observed several predicted deleterious non-synonymous variants in *PTPRO* that occur only in the susceptible population of our cohort, variation that is likely to have similar functional effects as inactivation by DNA methylation. There are two plausible hypotheses about the role of *PTPRO* in emphysema, based on these previous studies. First, like *ABI3BP*, the gene's role in tumor growth could suggest that alteration of *PTPRO* function can affect tissue repair or cellular survival in the lung. Second, decreased *PTPRO* function may contribute to the dysregulated inflammatory response in COPD (section 1.1.2).

Since we identified non-synonymous mutations in *PTPRO* that are predicted to be deleterious by *in silico* methods, a useful initial experiment would be to confirm that these mutations alter protein function. PTPRO is a protein phosphatase with known targets, so the introduction of each of these variants into cell lines using site-directed mutagenesis would allow for enzymatic assays testing the efficiency of wild type enzyme compared to mutant enzyme.

Dephosphorylation of a downstream target such as ERBB2 would be an ideal readout for this assay. *In vitro* studies combining CSE and siRNA silencing of *PTPRO* could test the hypothesis that *PTPRO* plays a role in alteration of cellular proliferation. *PTPRO*$^{-/-}$ mice have been described, and although they do not have a reported lung phenotype, chronic cigarette smoke exposure would model individuals with *PTPRO* mutations that develop emphysema. This would also be a useful model to see whether *PTPRO* knockdown reduces NF-κB signaling following cigarette smoke exposure in an ERBB2-dependent mechanism similar to that seen in liver inflammation.

6.2.3 *ZNF816*

The top association of both the sliding-window based approach and the gene-based approach of the WGS study described in section 2 is with the gene *ZNF816*. Searching PubMed for 'ZNF816' gives only one result (on 7/4/2016), a study identifying *ZNF816* as a candidate gene in Mayer-Rokitansky-Kuster-Hauser syndrome, a form of Mullerian agenesis (291). Structurally, the zinc finger motifs encoded by this gene suggest a DNA binding role. However, since our knowledge of this gene is extremely limited, basic exploratory studies will be necessary in order to hypothesize how it may contribute to the pathogenesis of COPD. Sequence analysis and molecular modeling of the protein may offer some insight into potential binding partners of the protein (292). Cellular and tissue localization experiments, including qPCR and immunofluorescence similar to those described in section 4 would be useful for clarifying where the gene is expressed in the lung. In addition, our study includes at least one individual with a rare, predicted deleterious, non-synonymous variant in the gene and replication of this variant using site-directed mutagenesis and comparison to wild-type protein may reveal functional alterations in relevant cell types.

6.3 IL-1 SIGNALING IN HUMAN LUNG DEVELOPMENT AND DISEASE

In section 4, we integrated mouse and human genome-wide scans to identify candidate genes likely to contribute to susceptibility to altered lung growth and development and COPD. Although we identified several candidate genes, we focused on the role of *Il1r2* in this phenotype, as IL-1 signaling is known to be involved in both prenatal and chronic disease, but had not previously been implicated in development. We demonstrated that anakinra, an IL-1 antagonist, can be used to reduce the alveolar chord length at 10 weeks of age in C3H/HeJ mice compared to mice treated with saline. This finding has important implications for human populations. It suggests that individuals with variants that reduce expression of IL-1 antagonists, such as known variants in *IL1RN*, are at increased risk of altered pulmonary development and growth (271, 273). In addition, variants in multiple genes encoding IL-1 antagonists or agonists acting in concert may contribute to susceptibility to decreased lung function. Finally, it suggests that there may be a therapeutic role for IL-1 antagonism in COPD.

Additional animal studies would be useful to support the findings presented here. No *Il1r2*$^{-/-}$ mouse has been reported in the literature but it could be used to support the findings in the C3H/HeJ mouse. An *Il1rn*$^{-/-}$ mouse is commercially available and measurement of 10wkCL in this strain could also offer further support for an altered balance of IL-1 agonists and antagonists in development and disease. Finally, crossing these two transgenic strains to produce a mouse deficient in both genes would be a powerful tool for testing the hypothesis that variation in multiple IL-1 antagonists has an additive effect on disease.

Additional genetic studies will be necessary to fully support the relevance of the results of section 4 to human populations. Although we included the findings of a meta-analysis of four

COPD cohorts in our studies, replication of the association near *IL1R2* in an independent cohort would be worthwhile. To test the hypothesis that variants in multiple agonists and antagonists may jointly contribute, looking at common variants previously associated with COPD in individuals with disease in large studies like the COPD meta-analysis would be useful. Variants in *IL1B* and *IL1RN* have been associated with COPD in some studies, and variants in *IL1A* and *IL1R1* have been associated with other inflammatory diseases (271, 272). Using a risk-scoring approach it would be possible to test whether having the minor allele of increasing numbers of these genotypes correlates with increasing FEV1.

Therapeutic uses for IL-1 antagonists face several challenges. Safety data for anakinra in pregnancy are limited, so widespread use of the drug to protect from even a moderate increase in risk of adult disease seems improbable. However, in specific situations, such as severe infections known to significantly increase the risk of bronchopulmonary dysplasia (BPD), anakinra may be a useful tool to prevent both neonatal disease and associated long-term chronic health effects (293-295). In COPD, clinical trials using IL-1β blocking antibodies have failed to show reduced decline in FEV1 or protection from exacerbations (296). Anakinra has not been used for treatment of COPD, but a phase I clinical study suggests that it may be useful for specific individuals with neutrophil-predominant asthma (297). Similarly, IL-1 antagonism in COPD may only be applicable to individuals with a specific phenotype or genetic predisposition (as suggested in section 4), which may explain the failure of IL-1β blocking antibodies. Importantly, soluble IL-1R2 has therapeutic potential given its unique pharmacologic profile but it is not currently FDA approved, unlike anakinra (266).

6.4 INTEGRATED GENETICS OF COPD

Genome-wide linkage analyses and CVAS of complex diseases have improved our understanding of the genetic basis of COPD. Despite this, these studies are affected by a lack of power due to inadequate sample sizes and phenotypic heterogeneity (see section 3.1). Today, significant resources are being put into large RVAS, as described in section 6.1. Although it may one day be possible to conduct appropriately powered studies to detect both common variants and rare variants, it is useful to prioritize future directed studies based on our current understanding of the genetic architecture underlying complex disease. One important demonstration of all three studies presented in this work is that data from one project can inform the results of another in order to offer additional evidence to support associations that do not reach genome-wide significance thresholds. The integrated literature review approach described in section 5 is a simple first step in prioritizing loci in this way, but more complex multidimensional approaches have been proposed (298). Similarly, network and pathway approaches use curated datasets to identify underlying interactions that would otherwise be missed (299, 300).

Such approaches can be expanded further with the addition of other 'big' datasets from transcriptomics, epigenetics, and proteomics studies, particularly useful given that some broad-sense heritability is certainly attributable to non-germline variation (see section 1.2.7) (301). A recent study integrating CVAS and gene expression data in asthma-COPD overlap syndrome identified several loci that would not have been identified in a traditional CVAS of the trait (302). In other diseases, systems biology approaches have been proposed that include the integration of multiple 'omics datasets with large clinical datasets in order to give a better perspective on disease processes (298, 303).

7.0 FINAL THOUGHTS

The last fifteen years have seen incredible advances in human genetics, from the publication of the first human genome to sequencing projects including hundreds of thousands of individuals (185, 287). This progress has largely been driven by improvements in technology that have increased the speed and reduced the cost of genotyping - even in the three years since this work began, the cost of obtaining a whole genome sequence has fallen by 300% (286). With the ability to detect the majority of genetic variation in large populations of individuals, the genetic architecture of complex diseases will be illuminated and many more susceptibility loci will be identified in diseases like COPD. As we account for this 'missing heritability', genetic testing to predict an individual's disease progression or response to therapy may become possible. It will also be necessary to identify the biological relevance of these loci, and biochemical and molecular studies *in vitro*, in animals, and in human populations will remain necessary to achieve an understanding of chronic disease that allows for individualized approaches to medicine. In this work, we used techniques that are novel to COPD to identify susceptibility loci, prioritizing those loci for further study, and giving suggestions of their potential mechanisms in disease. Precision medicine will be achieved by the integration of large genetic data like those presented here and the systematic analysis of their relevance to both the genetic architecture and the pathogenesis of complex disease.

APPENDIX A

EXTREME TRAIT WHOLE GENOME SEQUENCING TABULAR RESULTS

This section contains two large tables: results from tests for association between single rare variants and emphysema described in Chapter 2 and descriptions of single variants in the most significantly associated region at 19q31 described in Chapter 2.

Table 19. Top 100 single rare variant associations with emphysema

CH	BP	rsID	R	A	MAF	RES CNT	SUSC CNT	Fisher's Exact P	EMMA X P
9	37075550	rs117400947	G	A	0.03101	65/0/0	56/8/0	2.90E-03	7.13E-05
9	37124176	rs80121798	G	A	0.03101	65/0/0	56/8/0	2.90E-03	7.13E-05
9	37134601	rs77945177	C	T	0.03101	65/0/0	56/8/0	2.90E-03	7.13E-05
9	37314057	rs75985055	A	G	0.03101	65/0/0	56/8/0	2.90E-03	7.13E-05
15	90189675	rs148642238	G	A	0.03876	65/0/0	54/10/0	6.16E-04	1.34E-04
18	4471111	rs111462267	G	A	0.03488	65/0/0	55/9/0	1.34E-03	1.38E-04
3	173744839	rs13098521	G	T	0.05814	64/1/0	50/14/0	2.29E-04	1.93E-04
15	61872199	rs72745729	A	G	0.05426	64/1/0	52/11/1	1.91E-03	2.40E-04
15	61889602	rs17270990	G	C	0.05426	64/1/0	52/11/1	1.91E-03	2.40E-04
15	61895853	rs72745743	A	G	0.05426	64/1/0	52/11/1	1.91E-03	2.40E-04
15	61904639	rs72745747	T	A	0.05426	64/1/0	52/11/1	1.91E-03	2.40E-04
15	61909066	rs77410935	C	T	0.05426	64/1/0	52/11/1	1.91E-03	2.40E-04
8	81289897	rs11776280	A	G	0.03488	65/0/0	55/9/0	1.34E-03	2.86E-04
19	31940634	rs138069014	C	T	0.02326	65/0/0	58/6/0	1.32E-02	3.28E-04
11	2588398	rs61871513	T	G	0.03101	65/0/0	56/8/0	2.90E-03	3.99E-04
14	85438864	rs111859478	T	C	0.04264	64/1/0	54/10/0	4.26E-03	4.39E-04
7	78185077	rs192210418	G	T	0.02326	65/0/0	58/6/0	1.32E-02	4.60E-04
7	78191156	rs112837300	A	G	0.02326	65/0/0	58/6/0	1.32E-02	4.60E-04
10	46112395	rs116889933	C	T	0.03101	65/0/0	56/8/0	2.90E-03	4.69E-04
5	78475623	rs192753231	C	T	0.03876	65/0/0	54/10/0	6.16E-04	4.74E-04
5	78479584	rs78476112	T	C	0.03876	65/0/0	54/10/0	6.16E-04	4.74E-04

2	179873383	rs115470223	G	A	0.03101	65/0/0	56/8/0	2.90E-03	4.81E-04
7	79334839	rs74979129	G	C	0.03488	65/0/0	55/9/0	1.34E-03	4.87E-04
7	139341566	rs113455073	C	T	0.01938	65/0/0	59/5/0	2.77E-02	4.91E-04
21	37146881	rs4816515	C	A	0.02326	65/0/0	58/6/0	1.32E-02	5.15E-04
6	99594258	rs140042863	A	G	0.02713	65/0/0	57/7/0	6.22E-03	5.25E-04
6	164321549	rs148204975	C	T	0.05426	64/1/0	51/13/0	4.87E-04	5.26E-04
14	94555818	rs117299001	A	G	0.03101	65/0/0	56/8/0	2.90E-03	5.28E-04
15	34124930	rs72716807	G	A	0.02713	65/0/0	57/7/0	6.22E-03	5.30E-04
15	34173857	rs72716841	C	T	0.02713	65/0/0	57/7/0	6.22E-03	5.30E-04
7	158417916	rs62478283	C	G	0.02326	65/0/0	58/6/0	1.32E-02	5.50E-04
4	8538084	rs116146113	A	T	0.04651	53/12/0	64/0/0	5.00E-01	5.52E-04
6	16616251	rs143110024	G	A	0.03876	55/10/0	64/0/0	5.00E-01	5.53E-04
6	16618885	rs142770327	G	A	0.03876	55/10/0	64/0/0	5.00E-01	5.53E-04
10	8658326	rs142243250	T	C	0.04651	65/0/0	52/12/0	1.26E-04	5.57E-04
18	34999616	rs74848094	C	G	0.03488	64/1/0	56/8/0	1.67E-02	5.61E-04
8	22983423	rs75164982	G	T	0.03488	65/0/0	55/9/0	1.34E-03	5.94E-04
16	8851223	rs11647952	C	G	0.05426	63/2/0	52/12/0	4.43E-03	6.15E-04
19	33812605	rs79285836	G	T	0.02326	65/0/0	58/6/0	1.32E-02	6.18E-04
18	34924293	rs111773668	C	T	0.02713	65/0/0	57/7/0	6.22E-03	6.20E-04
4	96344416	rs17380396	C	T	0.02713	65/0/0	57/7/0	6.22E-03	6.37E-04
4	124006098	rs77202835	A	T	0.04264	55/10/0	63/1/0	5.01E-01	6.40E-04
3	142971115	rs183482652	G	A	0.02713	65/0/0	57/7/0	6.22E-03	6.42E-04
11	125706441	rs79456491	C	T	0.02713	65/0/0	57/7/0	6.22E-03	6.85E-04
11	125706973	rs75243465	A	T	0.02713	65/0/0	57/7/0	6.22E-03	6.85E-04
11	125771340	rs76032729	T	C	0.02713	65/0/0	57/7/0	6.22E-03	6.85E-04
20	60671377	rs138999600	A	G	0.03101	65/0/0	56/8/0	2.90E-03	7.02E-04
4	70114376	rs182747166	C	T	0.02326	65/0/0	58/6/0	1.32E-02	7.64E-04
4	70292860	rs141048116	T	C	0.02326	65/0/0	58/6/0	1.32E-02	7.64E-04
4	70311770	rs148015248	G	A	0.02326	65/0/0	58/6/0	1.32E-02	7.64E-04
6	103722012	rs118164141	G	T	0.03101	65/0/0	56/8/0	2.90E-03	7.79E-04
11	125380076	rs144720627	G	A	0.03101	65/0/0	56/8/0	2.90E-03	8.01E-04
14	40422078	rs117435965	G	A	0.05039	64/1/0	52/12/0	1.02E-03	8.11E-04
9	136404884	rs147342065	G	A	0.02713	65/0/0	57/7/0	6.22E-03	8.15E-04
2	83816815	rs11674200	A	C	0.04264	64/1/0	54/10/0	4.26E-03	8.24E-04
5	95202969	rs143135605	C	T	0.03101	58/7/0	63/1/0	5.09E-01	8.29E-04
7	14622069	rs77314012	A	G	0.04264	55/10/0	63/1/0	5.01E-01	8.35E-04
7	14628074	rs78555834	T	C	0.04264	55/10/0	63/1/0	5.01E-01	8.35E-04

7	14628400	rs62443559	A	C	0.04264	55/10/0	63/1/0	5.01E-01	8.35E-04
7	14634332	rs62445562	T	G	0.04264	55/10/0	63/1/0	5.01E-01	8.35E-04
7	14635257	rs117728549	T	C	0.04264	55/10/0	63/1/0	5.01E-01	8.35E-04
7	14637904	rs62445565	T	C	0.04264	55/10/0	63/1/0	5.01E-01	8.35E-04
2	12516191	rs149393532	T	G	0.02326	65/0/0	58/6/0	1.32E-02	8.40E-04
5	78514519	rs141529932	C	T	0.03488	65/0/0	55/9/0	1.34E-03	8.40E-04
5	78589856	rs77378764	T	A	0.03488	65/0/0	55/9/0	1.34E-03	8.40E-04
3	169970698	rs138920376	C	G	0.02713	65/0/0	57/7/0	6.22E-03	9.05E-04
3	169991792	rs145849596	G	T	0.02713	65/0/0	57/7/0	6.22E-03	9.05E-04
3	169996200	rs115345596	G	A	0.02713	65/0/0	57/7/0	6.22E-03	9.05E-04
3	173594773	rs35842459	A	G	0.06202	63/2/0	50/14/0	1.14E-03	9.13E-04
3	173692407	rs13100934	A	G	0.06202	63/2/0	50/14/0	1.14E-03	9.13E-04
4	69945014	rs78748694	T	G	0.02326	65/0/0	58/6/0	1.32E-02	9.13E-04
8	80363918	rs56059927	C	T	0.02713	65/0/0	57/7/0	6.22E-03	9.20E-04
7	30647900	rs71530517	G	A	0.03876	64/1/0	55/9/0	8.51E-03	9.51E-04
2	222364436		A	T	0.05195	40/0/0	31/4/2	4.31E-02	9.65E-04
12	39280387	rs113363481	A	C	0.01938	65/0/0	59/5/0	2.77E-02	9.85E-04
12	39606170	rs189149634	T	A	0.01938	65/0/0	59/5/0	2.77E-02	9.85E-04
9	37051488	rs77572427	C	G	0.02326	65/0/0	58/6/0	1.32E-02	9.91E-04
9	37172516	rs140817625	T	C	0.02326	65/0/0	58/6/0	1.32E-02	9.91E-04
9	37172676	rs144676436	G	T	0.02326	65/0/0	58/6/0	1.32E-02	9.91E-04
9	37176133	rs186530114	C	T	0.02326	65/0/0	58/6/0	1.32E-02	9.91E-04
14	68912098	rs117086110	C	T	0.05039	64/1/0	54/8/2	1.47E-02	1.01E-03
8	14698506	rs140046273	C	A	0.03488	56/9/0	64/0/0	5.00E-01	1.01E-03
8	14701780	rs117495771	T	C	0.03488	56/9/0	64/0/0	5.00E-01	1.01E-03
8	14751370	rs146299088	G	T	0.03488	56/9/0	64/0/0	5.00E-01	1.01E-03
8	14754589	rs188849037	C	G	0.03488	56/9/0	64/0/0	5.00E-01	1.01E-03
8	14754764	rs143351817	C	T	0.03488	56/9/0	64/0/0	5.00E-01	1.01E-03
6	164116241	rs76377186	G	A	0.03101	65/0/0	56/8/0	2.90E-03	1.01E-03
6	164157762	rs78273041	G	A	0.03101	65/0/0	56/8/0	2.90E-03	1.01E-03
3	54065004	rs147514340	T	C	0.03876	55/10/0	64/0/0	5.00E-01	1.06E-03
9	127199876	rs139831782	T	C	0.04651	64/1/0	53/11/0	2.10E-03	1.06E-03
11	21143884	rs74343549	T	C	0.02713	65/0/0	57/7/0	6.22E-03	1.07E-03
12	33964475	rs111393860	T	A	0.01938	65/0/0	59/5/0	2.77E-02	1.10E-03
12	33979286	rs77690974	C	T	0.01938	65/0/0	59/5/0	2.77E-02	1.10E-03
12	34243833	rs113312838	G	A	0.01938	65/0/0	59/5/0	2.77E-02	1.10E-03
12	34310738	rs112841437	T	A	0.01938	65/0/0	59/5/0	2.77E-02	1.10E-03

12	34697903	rs140355302	C	A	0.01938	65/0/0	59/5/0	2.77E-02	1.10E-03
12	34755876	rs111647128	C	A	0.01938	65/0/0	59/5/0	2.77E-02	1.10E-03
12	38049795		C	T	0.01938	65/0/0	59/5/0	2.77E-02	1.10E-03
12	38052440		G	A	0.01938	65/0/0	59/5/0	2.77E-02	1.10E-03
12	38070260	rs139708476	G	A	0.01938	65/0/0	59/5/0	2.77E-02	1.10E-03

Table 20. Rare variants in suggestively associated 30kbp region at 19q31

Chr	BP	rsID	Ref	Alt	MAF	RESCNT	SUSCNT	P-value EMMAX	P-value Fisher's Exact	TR	NS
19	53429414	rs188376139	G	A	0.00775	65/0/0	62/2/0	NA	0.24		
19	53429531	rs138869751	G	A	0.00388	65/0/0	63/1/0	NA	0.50		
19	53430152	rs144927208	C	T	0.00775	65/0/0	62/2/0	NA	0.24	*	
19	53430463		C	T	0.00775	63/2/0	64/0/0	NA	0.57	*	
19	53430708		G	A	0.00388	65/0/0	63/1/0	NA	0.50	*	
19	53430787	rs113303237	G	A	0.00388	65/0/0	63/1/0	NA	0.50	*	
19	53430883		A	G	0.00388	64/1/0	64/0/0	NA	0.67	*	
19	53430966	rs145727688	G	A	0.00388	65/0/0	63/1/0	NA	0.50	*	
19	53430994		A	C	0.00388	65/0/0	63/1/0	NA	0.50	*	
19	53432433	rs201431366	T	C	0.00388	65/0/0	63/1/0	NA	0.50	*	*
19	53432434	rs368707615	G	A	0.00388	65/0/0	63/1/0	NA	0.50	*	*
19	53432449	rs191740949	C	G	0.00388	65/0/0	63/1/0	NA	0.50	*	*
19	53432603	rs11880173	G	T	0.00388	65/0/0	63/1/0	NA	0.50	*	
19	53432854		A	G	0.00388	65/0/0	63/1/0	NA	0.50	*	
19	53433186	rs73936446	A	G	0.00388	64/1/0	64/0/0	NA	0.67	*	
19	53433347	rs184201997	A	C	0.00388	65/0/0	63/1/0	NA	0.50	*	
19	53433779	rs117992573	T	C	0.02326	65/0/0	58/6/0	0.02	0.01	*	
19	53434172		G	A	0.00388	65/0/0	63/1/0	NA	0.50	*	
19	53435491	rs142908218	A	T	0.00775	64/1/0	63/1/0	NA	0.99	*	
19	53435806	rs184509400	C	T	0.00388	64/1/0	64/0/0	NA	0.67	*	
19	53436018		C	T	0.00388	65/0/0	63/1/0	NA	0.50	*	
19	53436384		C	T	0.00459	52/1/0	56/0/0	NA	0.66	*	
19	53436881		T	C	0.00388	64/1/0	64/0/0	NA	0.67	*	
19	53437356		C	A	0.00388	65/0/0	63/1/0	NA	0.50	*	
19	53437589		G	A	0	65/0/0	64/0/0	NA	1.00	*	

19	53437638	rs182735408	T	A	0.00388	65/0/0	63/1/0	NA	0.50	*	
19	53438497		C	G	0.00388	65/0/0	63/1/0	NA	0.50	*	
19	53438548	rs368064459	G	A	0.00388	65/0/0	63/1/0	NA	0.50	*	
19	53438661	rs145781156	G	C	0.02713	62/3/0	60/4/0	0.64	0.70	*	
19	53439162	rs115986175	C	T	0.0155	64/1/0	61/3/0	0.72	0.36	*	
19	53439198	rs73055039	G	A	0.00388	65/0/0	63/1/0	NA	0.50	*	
19	53439524	rs145982953	T	C	0.00388	65/0/0	63/1/0	NA	0.50	*	
19	53440178	rs78149901	G	A	0.0155	64/1/0	61/3/0	0.24	0.36		
19	53440653	rs10402011	C	T	0.00388	65/0/0	63/1/0	NA	0.50		
19	53441295	rs10408640	C	T	0.00388	65/0/0	63/1/0	NA	0.50		
19	53441741	rs10411282	T	C	0.00388	65/0/0	63/1/0	NA	0.50		
19	53441962	rs111305521	T	C	0.00388	65/0/0	63/1/0	NA	0.50		
19	53441978	rs181118223	G	A	0.00388	65/0/0	63/1/0	NA	0.50		
19	53442258	rs73055051	T	C	0.00388	65/0/0	63/1/0	NA	0.50		
19	53442548		C	T	0.00388	65/0/0	63/1/0	NA	0.50		
19	53442969	rs77289965	C	T	0.01938	63/2/0	61/3/0	0.33	0.67		
19	53443767	rs149342050	C	T	0.00775	65/0/0	62/2/0	NA	0.24		
19	53444153	rs117035844	A	C	0.00388	65/0/0	63/1/0	NA	0.50		
19	53444251	rs112878696	C	T	0.00388	65/0/0	63/1/0	NA	0.50		
19	53444258	rs111714424	G	A	0.00388	65/0/0	63/1/0	NA	0.50		
19	53444490	rs190552009	G	A	0.00388	65/0/0	63/1/0	NA	0.50		
19	53444664	rs140251724	C	G	0.01163	62/3/0	64/0/0	0.08	0.53		
19	53444797		C	T	0.00388	64/1/0	64/0/0	NA	0.67		
19	53444888	rs181429793	C	G	0.00775	64/1/0	63/1/0	NA	0.99		
19	53444945		C	T	0.00388	65/0/0	63/1/0	NA	0.50		
19	53445603	rs148671377	C	A	0.00388	64/1/0	64/0/0	NA	0.67		
19	53447139	rs138542567	G	C	0.00388	64/1/0	64/0/0	NA	0.67		
19	53447288	rs147249472	G	A	0.00775	65/0/0	62/2/0	NA	0.24		
19	53447695	rs184291701	C	T	0.00775	64/1/0	63/1/0	NA	0.99		
19	53448477		A	G	0.00388	65/0/0	63/1/0	NA	0.50		
19	53448528		C	T	0.00388	65/0/0	63/1/0	NA	0.50		
19	53449470	rs2011866	T	A	0.00775	65/0/0	62/2/0	NA	0.24		
19	53449521	rs187405014	A	G	0.00388	65/0/0	63/1/0	NA	0.50		
19	53449645		G	A	0.00388	64/1/0	64/0/0	NA	0.67		
19	53451295		C	T	0.00388	64/1/0	64/0/0	NA	0.67		
19	53451336	rs142649507	A	C	0.00388	65/0/0	63/1/0	NA	0.50		
19	53451911		C	T	0	65/0/0	64/0/0	NA	1.00		

19	53452544		C	T	0.00388	65/0/0	63/1/0	NA	0.50		
19	53453131	rs146901011	G	A	0.00388	64/1/0	64/0/0	NA	0.67		
19	53453403	rs149695139	C	A	0.00388	64/1/0	64/0/0	NA	0.67		*
19	53453530	rs138196750	G	C	0.00388	65/0/0	63/1/0	NA	0.50		*
19	53453880	rs138017999	G	A	0.02326	61/4/0	62/2/0	0.67	0.65		*
19	53454301		G	A	0.00388	65/0/0	63/1/0	NA	0.50		
19	53454654	rs139245005	T	C	0.00388	65/0/0	63/1/0	NA	0.50		
19	53455229	rs145113439	T	C	0.00388	65/0/0	63/1/0	NA	0.50		
19	53456179	rs185038404	T	C	0.00388	65/0/0	63/1/0	NA	0.50		
19	53456397		C	T	0.00775	65/0/0	62/2/0	NA	0.24		
19	53457415	rs189344696	T	C	0.00388	65/0/0	63/1/0	NA	0.50		
19	53458122	rs150144819	C	T	0.00388	65/0/0	63/1/0	NA	0.50		
19	53458855	rs113242585	T	C	0.00388	65/0/0	63/1/0	NA	0.50		
19	53458891	rs137897568	G	T	0.00388	65/0/0	63/1/0	NA	0.50		
19	53459155	rs182706082	A	G	0.00388	65/0/0	63/1/0	NA	0.50		
19	53459265	rs140408256	A	G	0.00388	64/1/0	64/0/0	NA	0.67		
19	53459540	rs193096949	G	A	0.00388	65/0/0	63/1/0	NA	0.50		
19	53459573	rs185460031	C	T	0.00388	65/0/0	63/1/0	NA	0.50		
19	53459581	rs139377623	C	T	0.00775	64/1/0	63/1/0	NA	0.99		
19	53459792	rs146649631	C	T	0.00388	65/0/0	63/1/0	NA	0.50		
19	53459946	rs185565621	A	G	0.00388	65/0/0	63/1/0	NA	0.50		
19	53460267	rs146806388	A	G	0.00388	65/0/0	63/1/0	NA	0.50		
19	53460290	rs184996223	T	C	0.00388	65/0/0	63/1/0	NA	0.50		
19	53460534	rs187216641	C	T	0.00388	65/0/0	63/1/0	NA	0.50		
19	53460590	rs191584336	G	C	0.00388	65/0/0	63/1/0	NA	0.50		
19	53460814	rs191714893	G	A	0.00388	65/0/0	63/1/0	NA	0.50		
19	53460911	rs182493458	C	T	0.00388	65/0/0	63/1/0	NA	0.50		
19	53460976	rs187574014	C	G	0.00388	65/0/0	63/1/0	NA	0.50		
19	53461003	rs192566620	A	G	0.00388	65/0/0	63/1/0	NA	0.50		
19	53461079	rs185574162	C	T	0.00388	65/0/0	63/1/0	NA	0.50		
19	53461170	rs192162626	A	T	0.00388	65/0/0	63/1/0	NA	0.50		
19	53461333	rs184282140	T	C	0.00388	65/0/0	63/1/0	NA	0.50		
19	53461447	rs188656697	C	A	0.00388	65/0/0	63/1/0	NA	0.50		
19	53461495	rs181188954	A	T	0.00388	65/0/0	63/1/0	NA	0.50		
19	53461541	rs189205014	G	A	0.0155	64/1/0	61/3/0	0.23	0.36		
19	53461623	rs145781671	C	G	0.00388	65/0/0	63/1/0	NA	0.50		
19	53461691	rs180676933	G	C	0.0155	62/3/0	63/1/0	0.19	0.62		

19	53461850		G	C	0.00388	64/1/0	64/0/0	NA	0.67		
19	53462136	rs148559635	G	A	0.00388	65/0/0	63/1/0	NA	0.50		
19	53462242	rs141948083	C	T	0.00775	64/1/0	63/1/0	NA	0.99		
19	53462269	rs146141405	C	T	0.00388	65/0/0	63/1/0	NA	0.50		
19	53462279		G	A	0.00388	65/0/0	63/1/0	NA	0.50		
19	53462294		T	G	0.00388	65/0/0	63/1/0	NA	0.50		
19	53462507	rs140190373	T	C	0.00388	65/0/0	63/1/0	NA	0.50		
19	53462548	rs185996096	A	G	0.00388	65/0/0	63/1/0	NA	0.50		
19	53462587	rs188084686	C	A	0.00388	65/0/0	63/1/0	NA	0.50		
19	53463043	rs146234264	T	C	0.00388	65/0/0	63/1/0	NA	0.50		
19	53463114	rs138652725	C	G	0.00388	65/0/0	63/1/0	NA	0.50		
19	53463449	rs150311813	G	A	0.00388	65/0/0	63/1/0	NA	0.50		
19	53463718		C	G	0.00388	64/1/0	64/0/0	NA	0.67		
19	53463964	rs146129847	C	G	0.00388	65/0/0	63/1/0	NA	0.50		
19	53464050	rs117695884	T	C	0.0155	63/2/0	62/2/0	0.97	0.99		
19	53464136	rs146404860	G	A	0.00388	65/0/0	63/1/0	NA	0.50		
19	53464169	rs115102058	C	T	0.00388	65/0/0	63/1/0	NA	0.50		
19	53464452	rs184779820	T	C	0.00388	65/0/0	63/1/0	NA	0.50		
19	53464513	rs180913203	A	G	0.00775	64/1/0	63/1/0	NA	0.99		
19	53465148	rs145204396	C	T	0.01163	65/0/0	61/3/0	0.10	0.12		
19	53465793	rs186408317	G	A	0.00388	65/0/0	63/1/0	NA	0.50		
19	53466208	rs117179994	G	A	0.00388	65/0/0	63/1/0	NA	0.50		
19	53466404	rs187857443	G	A	0.00388	65/0/0	63/1/0	NA	0.50		

APPENDIX B

MURINE GWAS TABULAR RESULTS

Table 21. Known predicted deleterious variants in IL1R2

Variant*	Chr	BP^	Ref	Alleles	Global MAF	SIFT	PolyPhen
rs371567278	2	102009585	T	C/T	<0.001	0.04	0.824
rs563050165	2	102009597	C	T/C	<0.001	0.04	0.89
rs779320040	2	102009627	T	C/T	<0.001	0	0.998
rs772776365	2	102009631	C	T/C	<0.001	0.03	0.898
rs747233440	2	102009641	C	G/C	<0.001	0	0.97
rs771059282	2	102009642	G	T/G	<0.001	0	1
rs749271077	2	102009658	G	A/G	<0.001	0.05	0.771
rs781298497	2	102009714	G	T/G	<0.001	0	0.999
rs139061430	2	102009747	T	C/T	<0.001	0	1
rs145651311	2	102009748	A	G/A	0.003	0.02	0.904
rs765986681	2	102009774	C	T/C	<0.001	0	0.982
rs776289279	2	102009775	T	G/T	<0.001	0	0.999
rs751909955	2	102009810	G	T/G	<0.001	0	1
rs548206510	2	102015885	A	G/A	<0.001	0	1
rs746839384	2	102015897	G	C/G	<0.001	0	0.991
rs770663605	2	102015905	G	C/G	<0.001	0	0.982
rs757210661	2	102015906	C	T/C	<0.001	0	1
rs762159296	2	102015947	C	T/C	<0.001	0	0.993
rs766781084	2	102015957	G	A/G	<0.001	0	1
rs567036098	2	102015980	C	G/C	<0.001	0	0.999
rs371661087	2	102016051	T	G/T	<0.001	0	0.998
rs781638818	2	102019687	T	G/T	<0.001	0	0.776
rs780347152	2	102019699	C	T/C	<0.001	0	0.998
rs753591681	2	102019735	G	A/G	<0.001	0	0.973
rs764905517	2	102019740	T	C/T	<0.001	0.02	0.963
rs138667044	2	102019741	A	G/A/T	<0.001	0.02	0.855
rs538224389	2	102019744	A	G/A	<0.001	0	0.999
rs370953333	2	102019762	G	A/G	<0.001	0.05	0.937

rs760605159	2	102019789	A	G/A	<0.001	0.03	0.977
rs111549937	2	102019803	T	C/T	<0.001	0	0.974
rs369886122	2	102022222	T	C/T	<0.001	0.01	0.998
rs758651036	2	102022223	G	C/G	<0.001	0	0.999
rs183091210	2	102022236	G	A/G	<0.001	0	0.962
rs199970902	2	102022244	T	C/T	<0.001	0	0.985
rs751969491	2	102024533	T	G/T	<0.001	0	1
rs142547344	2	102024551	G	C/G/T	<0.001	0	0.999
rs142547344	2	102024551	T	C/G/T	<0.001	0	0.994
rs756165966	2	102024556	G	A/G	<0.001	0	0.974
rs753708553	2	102024566	A	T/A	<0.001	0	0.997
rs369088604	2	102024574	C	G/A/C	<0.001	0	0.973
rs770557266	2	102024611	G	A/G	<0.001	0	0.998
rs550245933	2	102024646	A	C/A/T	<0.001	0	0.993
rs550245933	2	102024646	T	C/A/T	<0.001	0	0.999
rs374536258	2	102024647	A	G/A	<0.001	0	0.991
rs780802660	2	102024649	A	G/A	<0.001	0	0.991
rs3218976	2	102024655	A	G/A/T	<0.001	0.01	0.962
rs772421571	2	102024658	A	G/A	<0.001	0	0.98
rs776757109	2	102024664	T	C/T	<0.001	0.01	0.951
rs764160421	2	102026113	G	A/G	<0.001	0	0.868
rs761679495	2	102026139	T	A/T	<0.001	0	0.953
rs201032704	2	102026140	C	T/C	<0.001	0	0.957
rs139885028	2	102026149	T	C/T	<0.001	0	0.882
rs144482163	2	102026155	C	T/C	<0.001	0.01	0.983
rs747569608	2	102026199	C	T/C	<0.001	0	0.999
rs766020542	2	102028235	T	C/T	<0.001	0.05	0.961
rs756425037	2	102028261	T	C/T	<0.001	0	1
rs780634571	2	102028262	T	C/T	<0.001	0	1
rs748144607	2	102028288	C	G/C	<0.001	0	0.955
rs772243120	2	102028292	A	G/A	<0.001	0	0.991
rs773007739	2	102028295	G	T/G	<0.001	0	0.794
rs368078017	2	102028385	A	C/A/G	<0.001	0	0.997
rs368078017	2	102028385	G	C/A/G	<0.001	0	0.997
rs371567278	2	102009585	T	C/T	<0.001	0.04	0.824
rs563050165	2	102009597	C	T/C	<0.001	0.04	0.89
rs779320040	2	102009627	T	C/T	<0.001	0	0.998
rs772776365	2	102009631	C	T/C	<0.001	0.03	0.898
rs747233440	2	102009641	C	G/C	<0.001	0	0.97
rs771059282	2	102009642	G	T/G	<0.001	0	1
rs749271077	2	102009658	G	A/G	<0.001	0.05	0.771
rs781298497	2	102009714	G	T/G	<0.001	0	0.999

rs139061430	2	102009747	T	C/T	<0.001	0	1
rs145651311	2	102009748	A	G/A	0.003	0.02	0.904
rs765986681	2	102009774	C	T/C	<0.001	0	0.982
rs776289279	2	102009775	T	G/T	<0.001	0	0.999
rs751909955	2	102009810	G	T/G	<0.001	0	1
rs548206510	2	102015885	A	G/A	<0.001	0	1
rs746839384	2	102015897	G	C/G	<0.001	0	0.991
rs770663605	2	102015905	G	C/G	<0.001	0	0.982
rs757210661	2	102015906	C	T/C	<0.001	0	1
rs762159296	2	102015947	C	T/C	<0.001	0	0.993
rs766781084	2	102015957	G	A/G	<0.001	0	1
rs567036098	2	102015980	C	G/C	<0.001	0	0.999
rs371661087	2	102016051	T	G/T	<0.001	0	0.998
rs781638818	2	102019687	T	G/T	<0.001	0	0.776
rs780347152	2	102019699	C	T/C	<0.001	0	0.998
rs753591681	2	102019735	G	A/G	<0.001	0	0.973
rs764905517	2	102019740	T	C/T	<0.001	0.02	0.963
rs138667044	2	102019741	A	G/A/T	<0.001	0.02	0.855
rs538224389	2	102019744	A	G/A	<0.001	0	0.999
rs370953333	2	102019762	G	A/G	<0.001	0.05	0.937
rs760605159	2	102019789	A	G/A	<0.001	0.03	0.977
rs111549937	2	102019803	T	C/T	<0.001	0	0.974
rs369886122	2	102022222	T	C/T	<0.001	0.01	0.998
rs758651036	2	102022223	G	C/G	<0.001	0	0.999
rs183091210	2	102022236	G	A/G	<0.001	0	0.962
rs199970902	2	102022244	T	C/T	<0.001	0	0.985
rs751969491	2	102024533	T	G/T	<0.001	0	1
rs142547344	2	102024551	G	C/G/T	<0.001	0	0.999
rs142547344	2	102024551	T	C/G/T	<0.001	0	0.994
rs756165966	2	102024556	G	A/G	<0.001	0	0.974
rs753708553	2	102024566	A	T/A	<0.001	0	0.997
rs369088604	2	102024574	C	G/A/C	<0.001	0	0.973
rs770557266	2	102024611	G	A/G	<0.001	0	0.998
rs550245933	2	102024646	A	C/A/T	<0.001	0	0.993
rs550245933	2	102024646	T	C/A/T	<0.001	0	0.999
rs374536258	2	102024647	A	G/A	<0.001	0	0.991
rs780802660	2	102024649	A	G/A	<0.001	0	0.991
rs3218976	2	102024655	A	G/A/T	<0.001	0.01	0.962
rs772421571	2	102024658	A	G/A	<0.001	0	0.98
rs776757109	2	102024664	T	C/T	<0.001	0.01	0.951
rs764160421	2	102026113	G	A/G	<0.001	0	0.868
rs761679495	2	102026139	T	A/T	<0.001	0	0.953

rs201032704	2	102026140	C	T/C	<0.001	0	0.957
rs139885028	2	102026149	T	C/T	<0.001	0	0.882
rs144482163	2	102026155	C	T/C	<0.001	0.01	0.983
rs747569608	2	102026199	C	T/C	<0.001	0	0.999
rs766020542	2	102028235	T	C/T	<0.001	0.05	0.961
rs756425037	2	102028261	T	C/T	<0.001	0	1
rs780634571	2	102028262	T	C/T	<0.001	0	1
rs748144607	2	102028288	C	G/C	<0.001	0	0.955
rs772243120	2	102028292	A	G/A	<0.001	0	0.991
rs773007739	2	102028295	G	T/G	<0.001	0	0.794
rs368078017	2	102028385	A	C/A/G	<0.001	0	0.997
rs368078017	2	102028385	G	C/A/G	<0.001	0	0.997
rs371567278	2	102009585	T	C/T	<0.001	0.04	0.824
rs563050165	2	102009597	C	T/C	<0.001	0.04	0.89
rs779320040	2	102009627	T	C/T	<0.001	0	0.998
rs772776365	2	102009631	C	T/C	<0.001	0.03	0.898
rs747233440	2	102009641	C	G/C	<0.001	0	0.97
rs771059282	2	102009642	G	T/G	<0.001	0	1
rs749271077	2	102009658	G	A/G	<0.001	0.04	0.771
rs781298497	2	102009714	G	T/G	<0.001	0	0.999
rs139061430	2	102009747	T	C/T	<0.001	0	1
rs145651311	2	102009748	A	G/A	<0.001	0.02	0.904
rs765986681	2	102009774	C	T/C	<0.001	0	0.982
rs776289279	2	102009775	T	G/T	<0.001	0	0.999
rs751909955	2	102009810	G	T/G	<0.001	0	1
rs548206510	2	102015885	A	G/A	<0.001	0	1
rs746839384	2	102015897	G	C/G	<0.001	0	0.991
rs770663605	2	102015905	G	C/G	<0.001	0	0.982
rs757210661	2	102015906	C	T/C	<0.001	0	1
rs762159296	2	102015947	C	T/C	<0.001	0	0.993
rs766781084	2	102015957	G	A/G	<0.001	0	1
rs567036098	2	102015980	C	G/C	<0.001	0	0.999
rs371661087	2	102016051	T	G/T	<0.001	0	0.998
rs781638818	2	102019687	T	G/T	<0.001	0	0.776
rs780347152	2	102019699	C	T/C	<0.001	0	0.998
rs753591681	2	102019735	G	A/G	<0.001	0	0.973
rs764905517	2	102019740	T	C/T	<0.001	0.01	0.963
rs138667044	2	102019741	A	G/A/T	<0.001	0.02	0.855
rs538224389	2	102019744	A	G/A	<0.001	0	0.999
rs370953333	2	102019762	G	A/G	<0.001	0.05	0.937
rs760605159	2	102019789	A	G/A	<0.001	0.04	0.977
rs111549937	2	102019803	T	C/T	<0.001	0	0.974

rs369886122	2	102022222	T	C/T	<0.001	0.01	0.998
rs758651036	2	102022223	G	C/G	<0.001	0	0.999
rs183091210	2	102022236	G	A/G	<0.001	0	0.962
rs199970902	2	102022244	T	C/T	<0.001	0	0.985
rs751969491	2	102024533	T	G/T	<0.001	0	1
rs142547344	2	102024551	G	C/G/T	<0.001	0	0.999
rs142547344	2	102024551	T	C/G/T	<0.001	0	0.994
rs756165966	2	102024556	G	A/G	<0.001	0	0.974
rs753708553	2	102024566	A	T/A	<0.001	0	0.997
rs369088604	2	102024574	C	G/A/C	<0.001	0	0.973
rs770557266	2	102024611	G	A/G	<0.001	0	0.998
rs550245933	2	102024646	A	C/A/T	<0.001	0	0.993
rs550245933	2	102024646	T	C/A/T	<0.001	0	0.999
rs374536258	2	102024647	A	G/A	<0.001	0	0.991
rs780802660	2	102024649	A	G/A	<0.001	0	0.991
rs3218976	2	102024655	A	G/A/T	<0.001	0.01	0.962
rs772421571	2	102024658	A	G/A	<0.001	0	0.98
rs776757109	2	102024664	T	C/T	<0.001	0	0.951
rs371567278	2	102009585	T	C/T	<0.001	0.05	0.824
rs779320040	2	102009627	T	C/T	<0.001	0.01	0.998
rs772776365	2	102009631	C	T/C	<0.001	0.03	0.898
rs747233440	2	102009641	C	G/C	<0.001	0	0.97
rs771059282	2	102009642	G	T/G	<0.001	0.01	1
rs781298497	2	102009714	G	T/G	<0.001	0	0.999
rs139061430	2	102009747	T	C/T	<0.001	0	1
rs145651311	2	102009748	A	G/A	<0.001	0.02	0.904
rs765986681	2	102009774	C	T/C	<0.001	0	0.982
rs776289279	2	102009775	T	G/T	<0.001	0	0.999
rs751909955	2	102009810	G	T/G	<0.001	0	1
rs548206510	2	102015885	A	G/A	<0.001	0	1
rs746839384	2	102015897	G	C/G	<0.001	0	0.991
rs770663605	2	102015905	G	C/G	<0.001	0	0.982
rs757210661	2	102015906	C	T/C	<0.001	0	1
rs762159296	2	102015947	C	T/C	<0.001	0	0.993
rs766781084	2	102015957	G	A/G	<0.001	0	1
rs567036098	2	102015980	C	G/C	<0.001	0	0.999
rs367683772	2	102016016	T	C/T	<0.001	0.04	0.922
rs371661087	2	102016051	T	G/T	<0.001	0	0.998
rs781638818	2	102019687	T	G/T	<0.001	0	0.776
rs780347152	2	102019699	C	T/C	<0.001	0	0.998
rs753591681	2	102019735	G	A/G	<0.001	0	0.973
rs764905517	2	102019740	T	C/T	<0.001	0.03	0.963

rs138667044	2	102019741	A	G/A/T	<0.001	0.05	0.855
rs538224389	2	102019744	A	G/A	<0.001	0	0.999
rs370953333	2	102019762	G	A/G	<0.001	0.03	0.937
rs760605159	2	102019789	A	G/A	<0.001	0	0.977
rs111549937	2	102019803	T	C/T	<0.001	0	0.974

* SNP IDs were identified in dbSNP146, ^Chromosomal location mapped to GRCh38

BIBLIOGRAPHY

1. Buist AS, Vollmer WM, McBurnie MA. Worldwide burden of COPD in high- and low-income countries. Part I. The burden of obstructive lung disease (BOLD) initiative. *Int J Tuberc Lung Dis* 2008; 12: 703-708.
2. Chapman KR, Mannino DM, Soriano JB, Vermeire PA, Buist AS, Thun MJ, Connell C, Jemal A, Lee TA, Miravitlles M, Aldington S, Beasley R. Epidemiology and costs of chronic obstructive pulmonary disease. *The European respiratory journal* 2006; 27: 188-207.
3. Decramer M, Janssens W, Miravitlles M. Chronic obstructive pulmonary disease. *Lancet* 2012; 379: 1341-1351.
4. The global burden of disease: 2004 update: World Health Organization; 2008.
5. Menzin J, Boulanger L, Marton J, Guadagno L, Dastani H, Dirani R, Phillips A, Shah H. The economic burden of chronic obstructive pulmonary disease (COPD) in a U.S. Medicare population. *Respir Med* 2008; 102: 1248-1256.
6. Morbidity and Mortality: 2012 Chartbook on Cardiovascular, Lung and Blood Diseases: U.S. Department of Health and Human Services. National Institutes of Health. National Heart Lung and Blood Institute; 2012.
7. Guarascio AJ, Ray SM, Finch CK, Self TH. The clinical and economic burden of chronic obstructive pulmonary disease in the USA. *ClinicoEconomics and Outcomes Research: CEOR* 2013; 5: 235-245.
8. Fletcher C, Peto R. The natural history of chronic airflow obstruction. *Br Med J* 1977; 1: 1645-1648.
9. Zhou JJ, Cho MH, Castaldi PJ, Hersh CP, Silverman EK, Laird NM. Heritability of chronic obstructive pulmonary disease and related phenotypes in smokers. *American journal of respiratory and critical care medicine* 2013; 188: 941-947.
10. From the Global Strategy for the Diagnosis MaPoC, Global Initiative for Chronic Obstructive Lung Disease (GOLD) 2016. Available from: http://www.goldcopd.org/.
11. Raherison C, Girodet PO. Epidemiology of COPD. *European respiratory review : an official journal of the European Respiratory Society* 2009; 18: 213-221.
12. Torres-Duque C, Maldonado D, Pérez-Padilla R, Ezzati M, Viegi G. Biomass Fuels and Respiratory Diseases. *Proceedings of the American Thoracic Society* 2008; 5: 577-590.
13. Macnee W. Pathogenesis of chronic obstructive pulmonary disease. *Clinics in chest medicine* 2007; 28: 479-513, v.
14. MacNee W. Pathology, pathogenesis, and pathophysiology. *BMJ : British Medical Journal* 2006; 332: 1202-1204.
15. Naeije R. Pulmonary Hypertension and Right Heart Failure in Chronic Obstructive Pulmonary Disease. *Proceedings of the American Thoracic Society* 2005; 2: 20-22.
16. Choudhury G, Rabinovich R, MacNee W. Comorbidities and systemic effects of chronic obstructive pulmonary disease. *Clinics in chest medicine* 2014; 35.
17. Russell RE, Thorley A, Culpitt SV, Dodd S, Donnelly LE, Demattos C, Fitzgerald M, Barnes PJ. Alveolar macrophage-mediated elastolysis: roles of matrix metalloproteinases,

cysteine, and serine proteases. *American journal of physiology Lung cellular and molecular physiology* 2002; 283: L867-873.
18. King PT. Inflammation in chronic obstructive pulmonary disease and its role in cardiovascular disease and lung cancer. *Clinical and Translational Medicine* 2015; 4: 1-13.
19. Cosio MG, Saetta M, Agusti A. Immunologic aspects of chronic obstructive pulmonary disease. *The New England journal of medicine* 2009; 360.
20. Vlahos R, Bozinovski S. Role of alveolar macrophages in chronic obstructive pulmonary disease. *Frontiers Immunol* 2014; 5.
21. Hiemstra PS, McCray PB, Jr., Bals R. The innate immune function of airway epithelial cells in inflammatory lung disease. *The European respiratory journal* 2015; 45: 1150-1162.
22. Cheng DS, Han W, Chen SM, Sherrill TP, Chont M, Park GY, Sheller JR, Polosukhin VV, Christman JW, Yull FE, Blackwell TS. Airway epithelium controls lung inflammation and injury through the NF-kappa B pathway. *Journal of immunology (Baltimore, Md : 1950)* 2007; 178: 6504-6513.
23. Wood AM, Stockley RA. Alpha one antitrypsin deficiency: from gene to treatment. *Respiration; international review of thoracic diseases* 2007; 74: 481-492.
24. Laurell CB. ELECTROPHORETIC MICROHETEROGENEITY OF SERUM ALPHA-1-ANTITRYPSIN. *Scandinavian journal of clinical and laboratory investigation* 1965; 17: 271-274.
25. Gross P, Pfitzer EA, Tolker E, Babyak MA, Kaschak M. Experimental Emphysema: Its Production with Papain in Normal and Silicotic Rats. *Arch Environ Health* 1965; 11: 50-58.
26. Janoff A, Sloan B, Weinbaum G, Damiano V, Sandhaus RA, Elias J, Kimbel P. Experimental emphysema induced with purified human neutrophil elastase: tissue localization of the instilled protease. *Am Rev Respir Dis* 1977; 115: 461-478.
27. Hautamaki RD, Kobayashi DK, Senior RM, Shapiro SD. Requirement for macrophage elastase for cigarette smoke-induced emphysema in mice. *Science (New York, NY)* 1997; 277: 2002-2004.
28. Shapiro SD, Goldstein NM, Houghton AM, Kobayashi DK, Kelley D, Belaaouaj A. Neutrophil elastase contributes to cigarette smoke-induced emphysema in mice. *The American journal of pathology* 2003; 163: 2329-2335.
29. Pryor WA, Stone K. Oxidants in cigarette smoke. Radicals, hydrogen peroxide, peroxynitrate, and peroxynitrite. *Annals of the New York Academy of Sciences* 1993; 686: 12-27; discussion 27-18.
30. Van Eeden SF, Sin DD. Oxidative stress in chronic obstructive pulmonary disease: A lung and systemic process. *Canadian respiratory journal : journal of the Canadian Thoracic Society* 2013; 20: 27-29.
31. Kirkham PA, Barnes PJ. Oxidative stress in COPD. *Chest* 2013; 144: 266-273.
32. Tzortzaki EG, Dimakou K, Neofytou E, Tsikritsaki K, Samara K, Avgousti M, Amargianitakis V, Gousiou A, Menikou S, Siafakas NM. Oxidative DNA damage and somatic mutations: a link to the molecular pathogenesis of chronic inflammatory airway diseases. *Chest* 2012; 141: 1243-1250.

33. Domej W, Oettl K, Renner W. Oxidative stress and free radicals in COPD – implications and relevance for treatment. *International journal of chronic obstructive pulmonary disease* 2014; 9: 1207-1224.
34. Aoshiba K, Nagai A. Senescence hypothesis for the pathogenetic mechanism of chronic obstructive pulmonary disease. *Proc Am Thorac Soc* 2009; 6: 596-601.
35. Kasahara Y, Tuder RM, Taraseviciene-Stewart L, Le Cras TD, Abman S, Hirth PK, Waltenberger J, Voelkel NF. Inhibition of VEGF receptors causes lung cell apoptosis and emphysema. *J Clin Invest* 2000; 106: 1311-1319.
36. Majo J, Ghezzo H, Cosio MG. Lymphocyte population and apoptosis in the lungs of smokers and their relation to emphysema. *The European respiratory journal* 2001; 17: 946-953.
37. Kasahara Y, Tuder RM, Cool CD, Lynch DA, Flores SC, Voelkel NF. Endothelial cell death and decreased expression of vascular endothelial growth factor and vascular endothelial growth factor receptor 2 in emphysema. *American journal of respiratory and critical care medicine* 2001; 163: 737-744.
38. Tsuji T, Aoshiba K, Nagai A. Alveolar cell senescence exacerbates pulmonary inflammation in patients with chronic obstructive pulmonary disease. *Respiration; international review of thoracic diseases* 2010; 80: 59-70.
39. Cavaillès A, Brinchault-Rabin G, Dixmier A, Goupil F, Gut-Gobert C, Marchand-Adam S, Meurice J-C, Morel H, Person-Tacnet C, Leroyer C, Diot P. Comorbidities of COPD. *European Respiratory Review* 2013; 22: 454-475.
40. Agusti AG, Sauleda J, Miralles C, Gomez C, Togores B, Sala E, Batle S, Busquets X. Skeletal muscle apoptosis and weight loss in chronic obstructive pulmonary disease. *American journal of respiratory and critical care medicine* 2002; 166: 485-489.
41. McGeachie MJ, Yates KP, Zhou X, Guo F, Sternberg AL, Van Natta ML, Wise RA, Szefler SJ, Sharma S, Kho AT, Cho MH, Croteau-Chonka DC, Castaldi PJ, Jain G, Sanyal A, Zhan Y, Lajoie BR, Dekker J, Stamatoyannopoulos J, Covar RA, Zeiger RS, Adkinson NF, Williams PV, Kelly HW, Grasemann H, Vonk JM, Koppelman GH, Postma DS, Raby BA, Houston I, Lu Q, Fuhlbrigge AL, Tantisira KG, Silverman EK, Tonascia J, Weiss ST, Strunk RC. Patterns of Growth and Decline in Lung Function in Persistent Childhood Asthma. *New England Journal of Medicine* 2016; 374: 1842-1852.
42. Tager IB, Segal MR, Speizer FE, Weiss ST. The natural history of forced expiratory volumes. Effect of cigarette smoking and respiratory symptoms. *Am Rev Respir Dis* 1988; 138: 837-849.
43. Britton J, Martinez FD. The relationship of childhood respiratory infection to growth and decline in lung function. *American journal of respiratory and critical care medicine* 1996; 154: S240-245.
44. Dratva J, Zemp E, Dharmage SC, Accordini S, Burdet L, Gislason T, Heinrich J, Janson C, Jarvis D, de Marco R, Norback D, Pons M, Real FG, Sunyer J, Villani S, Probst-Hensch N, Svanes C. Early Life Origins of Lung Ageing: Early Life Exposures and Lung Function Decline in Adulthood in Two European Cohorts Aged 28-73 Years. *PLoS One* 2016; 11: e0145127.
45. Barker DJ, Godfrey KM, Fall C, Osmond C, Winter PD, Shaheen SO. Relation of birth weight and childhood respiratory infection to adult lung function and death from chronic obstructive airways disease. *BMJ : British Medical Journal* 1991; 303: 671-675.

46. Vestbo J, Lange P. Natural history of COPD: Focusing on change in FEV1. *Respirology (Carlton, Vic)* 2016; 21: 34-43.
47. Wedzicha JA, Donaldson GC. Exacerbations of chronic obstructive pulmonary disease. *Respiratory care* 2003; 48: 1204-1213; discussion 1213-1205.
48. Segreti A, Stirpe E, Rogliani P, Cazzola M. Defining phenotypes in COPD: an aid to personalized healthcare. *Molecular diagnosis & therapy* 2014; 18: 381-388.
49. Agusti A, Calverley PM, Celli B, Coxson HO, Edwards LD, Lomas DA, MacNee W, Miller BE, Rennard S, Silverman EK, Tal-Singer R, Wouters E, Yates JC, Vestbo J. Characterisation of COPD heterogeneity in the ECLIPSE cohort. *Respiratory research* 2010; 11: 122.
50. Smith MC, Wrobel JP. Epidemiology and clinical impact of major comorbidities in patients with COPD. *International journal of chronic obstructive pulmonary disease* 2014; 9: 871-888.
51. Qaseem A, Wilt TJ, Weinberger SE, Hanania NA, Criner G, van der Molen T, Marciniuk DD, Denberg T, Schunemann H, Wedzicha W, MacDonald R, Shekelle P. Diagnosis and management of stable chronic obstructive pulmonary disease: a clinical practice guideline update from the American College of Physicians, American College of Chest Physicians, American Thoracic Society, and European Respiratory Society. *Annals of internal medicine* 2011; 155: 179-191.
52. Clausen JL. The diagnosis of emphysema, chronic bronchitis, and asthma. *Clinics in chest medicine* 1990; 11: 405-416.
53. Lynch DA, Austin JH, Hogg JC, Grenier PA, Kauczor HU, Bankier AA, Barr RG, Colby TV, Galvin JR, Gevenois PA, Coxson HO, Hoffman EA, Newell JD, Jr., Pistolesi M, Silverman EK, Crapo JD. CT-Definable Subtypes of Chronic Obstructive Pulmonary Disease: A Statement of the Fleischner Society. *Radiology* 2015; 277: 192-205.
54. Grydeland TB, Dirksen A, Coxson HO, Pillai SG, Sharma S, Eide GE, Gulsvik A, Bakke PS. Quantitative computed tomography: emphysema and airway wall thickness by sex, age and smoking. *The European respiratory journal* 2009; 34: 858-865.
55. Coxson HO, Leipsic J, Parraga G, Sin DD. Using pulmonary imaging to move chronic obstructive pulmonary disease beyond FEV1. *American journal of respiratory and critical care medicine* 2014; 190: 135-144.
56. Ford ES, Croft JB, Mannino DM, Wheaton AG, Zhang X, Giles WH. COPD Surveillance—United States, 1999-2011. *Chest* 2013; 144: 284-305.
57. Tilert T, Dillon C, Paulose-Ram R, Hnizdo E, Doney B. Estimating the U.S. prevalence of chronic obstructive pulmonary disease using pre- and post-bronchodilator spirometry: the National Health and Nutrition Examination Survey (NHANES) 2007–2010. *Respiratory research* 2013; 14: 1-11.
58. Halbert RJ, Natoli JL, Gano A, Badamgarav E, Buist AS, Mannino DM. Global burden of COPD: systematic review and meta-analysis. *The European respiratory journal* 2006; 28.
59. Mannino DM, Buist AS. Global burden of COPD: risk factors, prevalence, and future trends. *Lancet* 2007; 370: 765-773.
60. Bernd L, Joan BS, Michael S, Bernhard K, Lowie EV, Louisa G, Peter B, Marc M, Francisco G-R, Kaveh A, Julio A, Ana MM, Rogelio P-P, Maria MdO, Carlos AT-D, Andres C, Mauricio G-G, Sonia B. Determinants of Underdiagnosis of COPD in National and International Surveys. *Chest* 2015; 148: 971-985.

61. Connors AF, Jr., Dawson NV, Thomas C, Harrell FE, Jr., Desbiens N, Fulkerson WJ, Kussin P, Bellamy P, Goldman L, Knaus WA. Outcomes following acute exacerbation of severe chronic obstructive lung disease. The SUPPORT investigators (Study to Understand Prognoses and Preferences for Outcomes and Risks of Treatments). *American journal of respiratory and critical care medicine* 1996; 154: 959-967.
62. Seemungal TA, Donaldson GC, Paul EA, Bestall JC, Jeffries DJ, Wedzicha JA. Effect of exacerbation on quality of life in patients with chronic obstructive pulmonary disease. *American journal of respiratory and critical care medicine* 1998; 157: 1418-1422.
63. Sin DD, Anthonisen NR, Soriano JB, Agusti AG. Mortality in COPD: Role of comorbidities. *The European respiratory journal* 2006; 28: 1245-1257.
64. Berry CE, Wise RA. Mortality in COPD: causes, risk factors, and prevention. *Copd* 2010; 7: 375-382.
65. Burney P, Jithoo A, Kato B, Janson C, Mannino D, Niżankowska-Mogilnicka E, Studnicka M, Tan W, Bateman E, Koçabas A, Vollmer WM, Gislason T, Marks G, Koul PA, Harrabi I, Gnatiuc L, Buist S, Study ftBoOLD. Chronic obstructive pulmonary disease mortality and prevalence: the associations with smoking and poverty—a BOLD analysis. *Thorax* 2014; 69: 465-473.
66. Ries AL, Bauldoff GS, Carlin BW, Casaburi R, Emery CF, Mahler DA, Make B, Rochester CL, Zuwallack R, Herrerias C. Pulmonary Rehabilitation: Joint ACCP/AACVPR Evidence-Based Clinical Practice Guidelines. *Chest* 2007; 131: 4s-42s.
67. Stoller JK, Panos RJ, Krachman S, Doherty DE, Make B. Oxygen therapy for patients with COPD: current evidence and the long-term oxygen treatment trial. *Chest* 2010; 138: 179-187.
68. Sciurba FC, Chandra D, Bon J. Bronchoscopic Lung Volume Reduction in COPD: Lessons in Implementing Clinically Based Precision Medicine. *JAMA : the journal of the American Medical Association* 2016; 315: 139-141.
69. Deslee G, Mal H, Dutau H, Bourdin A, Vergnon JM, Pison C, Kessler R, Jounieaux V, Thiberville L, Leroy S, Marceau A, Laroumagne S, Mallet JP, Dukic S, Barbe C, Bulsei J, Jolly D, Durand-Zaleski I, Marquette CH. Lung Volume Reduction Coil Treatment vs Usual Care in Patients With Severe Emphysema: The REVOLENS Randomized Clinical Trial. *JAMA : the journal of the American Medical Association* 2016; 315: 175-184.
70. Cazzola M, Segreti A, Rogliani P. Comparative effectiveness of drugs for chronic obstructive pulmonary disease. *Drugs of today (Barcelona, Spain : 1998)* 2012; 48: 785-794.
71. Fabbri LM, Rabe KF. From COPD to chronic systemic inflammatory syndrome? *Lancet* 2007; 370: 797-799.
72. Fagerhol MK, Laurell CB. The Pi system-inherited variants of serum alpha 1-antitrypsin. *Progress in medical genetics* 1970; 7: 96-111.
73. Eriksson S. Studies in alpha 1-antitrypsin deficiency. *Acta medica Scandinavica Supplementum* 1965; 432: 1-85.
74. Needham M, Stockley RA. Alpha 1-antitrypsin deficiency. 3: Clinical manifestations and natural history. *Thorax* 2004; 59: 441-445.
75. Lai EC, Kao FT, Law ML, Woo SL. Assignment of the alpha 1-antitrypsin gene and a sequence-related gene to human chromosome 14 by molecular hybridization. *American journal of human genetics* 1983; 35: 385-392.

76. Long GL, Chandra T, Woo SL, Davie EW, Kurachi K. Complete sequence of the cDNA for human alpha 1-antitrypsin and the gene for the S variant. *Biochemistry* 1984; 23: 4828-4837.
77. Jeppsson JO, Laurell CB, Fagerhol M. Properties of isolated human alpha1-antitrypsins of Pi types M, S and Z. *Eur J Biochem* 1978; 83: 143-153.
78. Shapiro SD, Endicott SK, Province MA, Pierce JA, Campbell EJ. Marked longevity of human lung parenchymal elastic fibers deduced from prevalence of D-aspartate and nuclear weapons-related radiocarbon. *Journal of Clinical Investigation* 1991; 87: 1828-1834.
79. Bergin DA, Reeves EP, Meleady P, Henry M, McElvaney OJ, Carroll TP, Condron C, Chotirmall SH, Clynes M, O'Neill SJ, McElvaney NG. alpha-1 Antitrypsin regulates human neutrophil chemotaxis induced by soluble immune complexes and IL-8. *J Clin Invest* 2010; 120: 4236-4250.
80. Campos MA, Lascano J. alpha1 Antitrypsin deficiency: current best practice in testing and augmentation therapy. *Therapeutic advances in respiratory disease* 2014; 8: 150-161.
81. Hu Q, Shifren A, Sens C, Choi J, Szabo Z, Starcher BC, Knutsen RH, Shipley JM, Davis EC, Mecham RP, Urban Z. Mechanisms of emphysema in autosomal dominant cutis laxa. *Matrix Biology* 2010; 29: 621-628.
82. Stanley SE, Chen JJL, Podlevsky JD, Alder JK, Hansel NN, Mathias RA, Qi X, Rafaels NM, Wise RA, Silverman EK, Barnes KC, Armanios M. Telomerase mutations in smokers with severe emphysema. *The Journal of Clinical Investigation*; 125: 563-570.
83. Fletcher CM. Letter: Natural history of chronic bronchitis. *Br Med J* 1976; 1: 1592-1593.
84. Larson RK, Barman ML, Kueppers F, Fudenberg HH. Genetic and environmental determinants of chronic obstructive pulmonary disease. *Annals of internal medicine* 1970; 72: 627-632.
85. Tager IB, Rosner B, Tishler PV, Speizer FE, Kass EH. Household aggregation of pulmonary function and chronic bronchitis. *Am Rev Respir Dis* 1976; 114: 485-492.
86. Redline S, Tishler PV, Lewitter FI, Tager IB, Munoz A, Speizer FE. Assessment of genetic and nongenetic influences on pulmonary function. A twin study. *Am Rev Respir Dis* 1987; 135: 217-222.
87. Hubert HB, Fabsitz RR, Feinleib M, Gwinn C. Genetic and environmental influences on pulmonary function in adult twins. *Am Rev Respir Dis* 1982; 125: 409-415.
88. Chen Y, Horne SL, Rennie DC, Dosman JA. Segregation analysis of two lung function indices in a random sample of young families: the Humboldt Family Study. *Genetic epidemiology* 1996; 13: 35-47.
89. Givelber RJ, Couropmitree NN, Gottlieb DJ, Evans JC, Levy D, Myers RH, O'Connor GT. Segregation analysis of pulmonary function among families in the Framingham Study. *American journal of respiratory and critical care medicine* 1998; 157: 1445-1451.
90. Morton NE. Sequential tests for the detection of linkage. *American journal of human genetics* 1955; 7: 277-318.
91. Tsui L, Buchwald M, Barker D, Braman J, Knowlton R, Schumm J, Eiberg H, Mohr J, Kennedy D, Plavsic N, et a. Cystic fibrosis locus defined by a genetically linked polymorphic DNA marker. *Science (New York, NY)* 1985; 230: 1054-1057.
92. Murray J, Buetow K, Weber J, Ludwigsen S, Scherpbier-Heddema T, Manion F, Quillen J, Sheffield V, Sunden S, Duyk G, et a. A comprehensive human linkage map with

centimorgan density. Cooperative Human Linkage Center (CHLC). *Science (New York, NY)* 1994; 265: 2049-2054.
93. Silverman EK, Mosley JD, Palmer LJ, Barth M, Senter JM, Brown A, Drazen JM, Kwiatkowski DJ, Chapman HA, Campbell EJ, Province MA, Rao DC, Reilly JJ, Ginns LC, Speizer FE, Weiss ST. Genome-wide linkage analysis of severe, early-onset chronic obstructive pulmonary disease: airflow obstruction and chronic bronchitis phenotypes. *Human molecular genetics* 2002; 11: 623-632.
94. Silverman EK, Palmer LJ, Mosley JD, Barth M, Senter JM, Brown A, Drazen JM, Kwiatkowski DJ, Chapman HA, Campbell EJ, Province MA, Rao DC, Reilly JJ, Ginns LC, Speizer FE, Weiss ST. Genomewide linkage analysis of quantitative spirometric phenotypes in severe early-onset chronic obstructive pulmonary disease. *American journal of human genetics* 2002; 70: 1229-1239.
95. DeMeo DL, Celedón JC, Lange C, Reilly JJ, Chapman HA, Sylvia JS, Speizer FE, Weiss ST, Silverman EK. Genome-wide Linkage of Forced Mid-expiratory Flow in Chronic Obstructive Pulmonary Disease. *American journal of respiratory and critical care medicine* 2004; 170: 1294-1301.
96. Joost O, Wilk JB, Cupples LA, Harmon M, Shearman AM, Baldwin CT, O'Connor GT, Myers RH, Gottlieb DJ. Genetic loci influencing lung function: a genome-wide scan in the Framingham Study. *American journal of respiratory and critical care medicine* 2002; 165: 795-799.
97. Wilk JB, DeStefano AL, Arnett DK, Rich SS, Djousse L, Crapo RO, Leppert MF, Province MA, Cupples LA, Gottlieb DJ, Myers RH. A genome-wide scan of pulmonary function measures in the National Heart, Lung, and Blood Institute Family Heart Study. *American journal of respiratory and critical care medicine* 2003; 167: 1528-1533.
98. Palmer LJ, Celedon JC, Chapman HA, Speizer FE, Weiss ST, Silverman EK. Genome-wide linkage analysis of bronchodilator responsiveness and post-bronchodilator spirometric phenotypes in chronic obstructive pulmonary disease. *Human molecular genetics* 2003; 12: 1199-1210.
99. Celedón JC, Lange C, Raby BA, Litonjua AA, Palmer LJ, DeMeo DL, Reilly JJ, Kwiatkowski DJ, Chapman HA, Laird N, Sylvia JS, Hernandez M, Speizer FE, Weiss ST, Silverman EK. The transforming growth factor-β1 (TGFB1) gene is associated with chronic obstructive pulmonary disease (COPD). *Human molecular genetics* 2004; 13: 1649-1656.
100. Hersh CP, Silverman EK, Gascon J, Bhattacharya S, Klanderman BJ, Litonjua AA, Lefebvre V, Sparrow D, Reilly JJ, Anderson WH, Lomas DA, Mariani TJ. SOX5 is a candidate gene for chronic obstructive pulmonary disease susceptibility and is necessary for lung development. *American journal of respiratory and critical care medicine* 2011; 183: 1482-1489.
101. Hersh CP, Pillai SG, Zhu G, Lomas DA, Bakke P, Gulsvik A, DeMeo DL, Klanderman BJ, Lazarus R, Litonjua AA, Sparrow D, Reilly JJ, Agusti A, Calverley PMA, Donner CF, Levy RD, Make BJ, Paré PD, Rennard SI, Vestbo J, Wouters EFM, Scholand MB, Coon H, Hoidal J, Silverman EK. Multistudy Fine Mapping of Chromosome 2q Identifies XRCC5 as a Chronic Obstructive Pulmonary Disease Susceptibility Gene. *American journal of respiratory and critical care medicine* 2010; 182: 605-613.

102. Risch N, Merikangas K. The future of genetic studies of complex human diseases. *Science (New York, NY)* 1996; 273: 1516-1517.
103. Wilkening S, Chen B, Bermejo JL, Canzian F. Is there still a need for candidate gene approaches in the era of genome-wide association studies? *Genomics* 2009; 93: 415-419.
104. Wang X, Li L, Xiao J, Jin C, Huang K, Kang X, Wu X, Lv F. Association of ADAM33 gene polymorphisms with COPD in a northeastern Chinese population. *BMC Med Genet* 2009; 10: 132.
105. Kukkonen M, Tiili E, Vehmas T, Oksa P, Piirilä P, Hirvonen A. Association of SERPINE2 and other COPD candidate genes to different emphysema subtypes and lung function. *Respiratory Medicine*; 107: S6.
106. DeMeo DL, Mariani TJ, Lange C, Srisuma S, Litonjua AA, Celedon JC, Lake SL, Reilly JJ, Chapman HA, Mecham BH, Haley KJ, Sylvia JS, Sparrow D, Spira AE, Beane J, Pinto-Plata V, Speizer FE, Shapiro SD, Weiss ST, Silverman EK. The SERPINE2 gene is associated with chronic obstructive pulmonary disease. *American journal of human genetics* 2006; 78: 253-264.
107. Silverman EK, Spira A, Paré PD. Genetics and Genomics of Chronic Obstructive Pulmonary Disease. *Proceedings of the American Thoracic Society* 2009; 6: 539-542.
108. Smolonska J, Wijmenga C, Postma DS, Boezen HM. Meta-analyses on suspected chronic obstructive pulmonary disease genes: a summary of 20 years' research. *American journal of respiratory and critical care medicine* 2009; 180: 618-631.
109. Ozaki K, Ohnishi Y, Iida A, Sekine A, Yamada R, Tsunoda T, Sato H, Sato H, Hori M, Nakamura Y, Tanaka T. Functional SNPs in the lymphotoxin-alpha gene that are associated with susceptibility to myocardial infarction. *Nat Genet* 2002; 32: 650-654.
110. Welter D, MacArthur J, Morales J, Burdett T, Hall P, Junkins H, Klemm A, Flicek P, Manolio T, Hindorff L, Parkinson H. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic acids research* 2014; 42: D1001-D1006.
111. Pillai SG, Ge D, Zhu G, Kong X, Shianna KV, Need AC, Feng S, Hersh CP, Bakke P, Gulsvik A, Ruppert A, Lodrup Carlsen KC, Roses A, Anderson W, Rennard SI, Lomas DA, Silverman EK, Goldstein DB. A genome-wide association study in chronic obstructive pulmonary disease (COPD): identification of two major susceptibility loci. *PLoS Genet* 2009; 5: e1000421.
112. Wilk JB, Chen TH, Gottlieb DJ, Walter RE, Nagle MW, Brandler BJ, Myers RH, Borecki IB, Silverman EK, Weiss ST, O'Connor GT. A genome-wide association study of pulmonary function measures in the Framingham Heart Study. *PLoS Genet* 2009; 5: e1000429.
113. Cho MH, Boutaoui N, Klanderman BJ, Sylvia JS, Ziniti JP, Hersh CP, DeMeo DL, Hunninghake GM, Litonjua AA, Sparrow D, Lange C, Won S, Murphy JR, Beaty TH, Regan EA, Make BJ, Hokanson JE, Crapo JD, Kong X, Anderson WH, Tal-Singer R, Lomas DA, Bakke P, Gulsvik A, Pillai SG, Silverman EK. Variants in FAM13A are associated with chronic obstructive pulmonary disease. *Nat Genet* 2010; 42: 200-202.
114. Cho MH, Castaldi PJ, Wan ES, Siedlinski M, Hersh CP, Demeo DL, Himes BE, Sylvia JS, Klanderman BJ, Ziniti JP, Lange C, Litonjua AA, Sparrow D, Regan EA, Make BJ, Hokanson JE, Murray T, Hetmanski JB, Pillai SG, Kong X, Anderson WH, Tal-Singer R, Lomas DA, Coxson HO, Edwards LD, MacNee W, Vestbo J, Yates JC, Agusti A, Calverley PM, Celli B, Crim C, Rennard S, Wouters E, Bakke P, Gulsvik A, Crapo JD,

Beaty TH, Silverman EK. A genome-wide association study of COPD identifies a susceptibility locus on chromosome 19q13. *Human molecular genetics* 2012; 21: 947-957.
115. Cho MH, McDonald ML, Zhou X, Mattheisen M, Castaldi PJ, Hersh CP, Demeo DL, Sylvia JS, Ziniti J, Laird NM, Lange C, Litonjua AA, Sparrow D, Casaburi R, Barr RG, Regan EA, Make BJ, Hokanson JE, Lutz S, Dudenkov TM, Farzadegan H, Hetmanski JB, Tal-Singer R, Lomas DA, Bakke P, Gulsvik A, Crapo JD, Silverman EK, Beaty TH. Risk loci for chronic obstructive pulmonary disease: a genome-wide association study and meta-analysis. *The lancet Respiratory medicine* 2014; 2: 214-225.
116. Manchia M, Cullis J, Turecki G, Rouleau GA, Uher R, Alda M. The Impact of Phenotypic and Genetic Heterogeneity on Results of Genome Wide Association Studies of Complex Diseases. *PLoS ONE* 2013; 8: e76295.
117. Lee JH, Cho MH, McDonald ML, Hersh CP, Castaldi PJ, Crapo JD, Wan ES, Dy JG, Chang Y, Regan EA, Hardin M, DeMeo DL, Silverman EK. Phenotypic and genetic heterogeneity among subjects with mild airflow obstruction in COPDGene. *Respir Med* 2014; 108: 1469-1480.
118. Kong X, Cho MH, Anderson W, Coxson HO, Muller N, Washko G, Hoffman EA, Bakke P, Gulsvik A, Lomas DA, Silverman EK, Pillai SG. Genome-wide Association Study Identifies BICD1 as a Susceptibility Gene for Emphysema. *American journal of respiratory and critical care medicine* 2011; 183: 43-49.
119. Castaldi PJ, Cho MH, San Jose Estepar R, McDonald ML, Laird N, Beaty TH, Washko G, Crapo JD, Silverman EK. Genome-wide association identifies regulatory Loci associated with distinct local histogram emphysema patterns. *American journal of respiratory and critical care medicine* 2014; 190: 399-409.
120. Wilk JB, Walter RE, Laramie JM, Gottlieb DJ, O'Connor GT. Framingham Heart Study genome-wide association: results for pulmonary function measures. *BMC Med Genet* 2007; 8 Suppl 1: S8.
121. Soler Artigas M, Loth DW, Wain LV, Gharib SA, Obeidat M, Tang W, Zhai G, Zhao JH, Smith AV, Huffman JE, Albrecht E, Jackson CM, Evans DM, Cadby G, Fornage M, Manichaikul A, Lopez LM, Johnson T, Aldrich MC, Aspelund T, Barroso I, Campbell H, Cassano PA, Couper DJ, Eiriksdottir G, Franceschini N, Garcia M, Gieger C, Gislason GK, Grkovic I, Hammond CJ, Hancock DB, Harris TB, Ramasamy A, Heckbert SR, Heliovaara M, Homuth G, Hysi PG, James AL, Jankovic S, Joubert BR, Karrasch S, Klopp N, Koch B, Kritchevsky SB, Launer LJ, Liu Y, Loehr LR, Lohman K, Loos RJ, Lumley T, Al Balushi KA, Ang WQ, Barr RG, Beilby J, Blakey JD, Boban M, Boraska V, Brisman J, Britton JR, Brusselle GG, Cooper C, Curjuric I, Dahgam S, Deary IJ, Ebrahim S, Eijgelsheim M, Francks C, Gaysina D, Granell R, Gu X, Hankinson JL, Hardy R, Harris SE, Henderson J, Henry A, Hingorani AD, Hofman A, Holt PG, Hui J, Hunter ML, Imboden M, Jameson KA, Kerr SM, Kolcic I, Kronenberg F, Liu JZ, Marchini J, McKeever T, Morris AD, Olin AC, Porteous DJ, Postma DS, Rich SS, Ring SM, Rivadeneira F, Rochat T, Sayer AA, Sayers I, Sly PD, Smith GD, Sood A, Starr JM, Uitterlinden AG, Vonk JM, Wannamethee SG, Whincup PH, Wijmenga C, Williams OD, Wong A, Mangino M, Marciante KD, McArdle WL, Meibohm B, Morrison AC, North KE, Omenaas E, Palmer LJ, Pietilainen KH, Pin I, Pola Sbreve Ek O, Pouta A, Psaty BM, Hartikainen AL, Rantanen T, Ripatti S, Rotter JI, Rudan I, Rudnicka AR, Schulz H, Shin

SY, Spector TD, Surakka I, Vitart V, Volzke H, Wareham NJ, Warrington NM, Wichmann HE, Wild SH, Wilk JB, Wjst M, Wright AF, Zgaga L, Zemunik T, Pennell CE, Nyberg F, Kuh D, Holloway JW, Boezen HM, Lawlor DA, Morris RW, Probst-Hensch N, Kaprio J, Wilson JF, Hayward C, Kahonen M, Heinrich J, Musk AW, Jarvis DL, Glaser S, Jarvelin MR, Ch Stricker BH, Elliott P, O'Connor GT, Strachan DP, London SJ, Hall IP, Gudnason V, Tobin MD. Genome-wide association and large-scale follow up identifies 16 new loci influencing lung function. *Nat Genet* 2011; 43: 1082-1090.

122. Hancock DB, Soler Artigas M, Gharib SA, Henry A, Manichaikul A, Ramasamy A, Loth DW, Imboden M, Koch B, McArdle WL, Smith AV, Smolonska J, Sood A, Tang W, Wilk JB, Zhai G, Zhao JH, Aschard H, Burkart KM, Curjuric I, Eijgelsheim M, Elliott P, Gu X, Harris TB, Janson C, Homuth G, Hysi PG, Liu JZ, Loehr LR, Lohman K, Loos RJ, Manning AK, Marciante KD, Obeidat M, Postma DS, Aldrich MC, Brusselle GG, Chen TH, Eiriksdottir G, Franceschini N, Heinrich J, Rotter JI, Wijmenga C, Williams OD, Bentley AR, Hofman A, Laurie CC, Lumley T, Morrison AC, Joubert BR, Rivadeneira F, Couper DJ, Kritchevsky SB, Liu Y, Wjst M, Wain LV, Vonk JM, Uitterlinden AG, Rochat T, Rich SS, Psaty BM, O'Connor GT, North KE, Mirel DB, Meibohm B, Launer LJ, Khaw KT, Hartikainen AL, Hammond CJ, Glaser S, Marchini J, Kraft P, Wareham NJ, Volzke H, Stricker BH, Spector TD, Probst-Hensch NM, Jarvis D, Jarvelin MR, Heckbert SR, Gudnason V, Boezen HM, Barr RG, Cassano PA, Strachan DP, Fornage M, Hall IP, Dupuis J, Tobin MD, London SJ. Genome-wide joint meta-analysis of SNP and SNP-by-smoking interaction identifies novel loci for pulmonary function. *PLoS Genet* 2012; 8: e1003098.
123. Hancock DB, Eijgelsheim M, Wilk JB, Gharib SA, Loehr LR, Marciante KD, Franceschini N, van Durme YM, Chen TH, Barr RG, Schabath MB, Couper DJ, Brusselle GG, Psaty BM, van Duijn CM, Rotter JI, Uitterlinden AG, Hofman A, Punjabi NM, Rivadeneira F, Morrison AC, Enright PL, North KE, Heckbert SR, Lumley T, Stricker BH, O'Connor GT, London SJ. Meta-analyses of genome-wide association studies identify multiple loci associated with pulmonary function. *Nat Genet* 2010; 42: 45-52.
124. Siedlinski M, Cho MH, Bakke P, Gulsvik A, Lomas DA, Anderson W, Kong X, Rennard SI, Beaty TH, Hokanson JE, Crapo JD, Silverman EK. Genome-wide association study of smoking behaviours in patients with COPD. *Thorax* 2011; 66: 894-902.
125. Bloom AJ, Hartz SM, Baker TB, Chen LS, Piper ME, Fox L, Martinez M, Hatsukami D, Johnson EO, Laurie CC, Saccone NL, Goate A, Bierut LJ. Beyond cigarettes per day. A genome-wide association study of the biomarker carbon monoxide. *Annals of the American Thoracic Society* 2014; 11: 1003-1010.
126. Wan ES, Cho MH, Boutaoui N, Klanderman BJ, Sylvia JS, Ziniti JP, Won S, Lange C, Pillai SG, Anderson WH, Kong X, Lomas DA, Bakke PS, Gulsvik A, Regan EA, Murphy JR, Make BJ, Crapo JD, Wouters EF, Celli BR, Silverman EK, DeMeo DL. Genome-wide association analysis of body mass in chronic obstructive pulmonary disease. *Am J Respir Cell Mol Biol* 2011; 45: 304-310.
127. Hardin M, Cho MH, McDonald ML, Wan E, Lomas DA, Coxson HO, MacNee W, Vestbo J, Yates JC, Agusti A, Calverley PM, Celli B, Crim C, Rennard S, Wouters E, Bakke P, Bhatt SP, Kim V, Ramsdell J, Regan EA, Make BJ, Hokanson JE, Crapo JD, Beaty TH,

Hersh CP. A genome-wide analysis of the response to inhaled beta-agonists in chronic obstructive pulmonary disease. *The pharmacogenomics journal* 2015.
128. Repapi E, Sayers I, Wain LV, Burton PR, Johnson T, Obeidat M, Zhao JH, Ramasamy A, Zhai G, Vitart V, Huffman JE, Igl W, Albrecht E, Deloukas P, Henderson J, Granell R, McArdle WL, Rudnicka AR, Barroso I, Loos RJ, Wareham NJ, Mustelin L, Rantanen T, Surakka I, Imboden M, Wichmann HE, Grkovic I, Jankovic S, Zgaga L, Hartikainen AL, Peltonen L, Gyllensten U, Johansson A, Zaboli G, Campbell H, Wild SH, Wilson JF, Glaser S, Homuth G, Volzke H, Mangino M, Soranzo N, Spector TD, Polasek O, Rudan I, Wright AF, Heliovaara M, Ripatti S, Pouta A, Naluai AT, Olin AC, Toren K, Cooper MN, James AL, Palmer LJ, Hingorani AD, Wannamethee SG, Whincup PH, Smith GD, Ebrahim S, McKeever TM, Pavord ID, MacLeod AK, Morris AD, Porteous DJ, Cooper C, Dennison E, Shaheen S, Karrasch S, Schnabel E, Schulz H, Grallert H, Bouatia-Naji N, Delplanque J, Froguel P, Blakey JD, Britton JR, Morris RW, Holloway JW, Lawlor DA, Hui J, Nyberg F, Jarvelin MR, Jackson C, Kahonen M, Kaprio J, Probst-Hensch NM, Koch B, Hayward C, Evans DM, Elliott P, Strachan DP, Hall IP, Tobin MD. Genome-wide association study identifies five loci associated with lung function. *Nat Genet* 2010; 42: 36-44.
129. Pillai SG, Kong X, Edwards LD, Cho MH, Anderson WH, Coxson HO, Lomas DA, Silverman EK. Loci identified by genome-wide association studies influence different disease-related phenotypes in chronic obstructive pulmonary disease. *American journal of respiratory and critical care medicine* 2010; 182: 1498-1505.
130. Hardin M, Zielinski J, Wan ES, Hersh CP, Castaldi PJ, Schwinder E, Hawrylkiewicz I, Sliwinski P, Cho MH, Silverman EK. CHRNA3/5, IREB2, and ADCY2 are associated with severe chronic obstructive pulmonary disease in Poland. *Am J Respir Cell Mol Biol* 2012; 47: 203-208.
131. Wilk JB, Shrine NR, Loehr LR, Zhao JH, Manichaikul A, Lopez LM, Smith AV, Heckbert SR, Smolonska J, Tang W, Loth DW, Curjuric I, Hui J, Cho MH, Latourelle JC, Henry AP, Aldrich M, Bakke P, Beaty TH, Bentley AR, Borecki IB, Brusselle GG, Burkart KM, Chen TH, Couper D, Crapo JD, Davies G, Dupuis J, Franceschini N, Gulsvik A, Hancock DB, Harris TB, Hofman A, Imboden M, James AL, Khaw KT, Lahousse L, Launer LJ, Litonjua A, Liu Y, Lohman KK, Lomas DA, Lumley T, Marciante KD, McArdle WL, Meibohm B, Morrison AC, Musk AW, Myers RH, North KE, Postma DS, Psaty BM, Rich SS, Rivadeneira F, Rochat T, Rotter JI, Soler Artigas M, Starr JM, Uitterlinden AG, Wareham NJ, Wijmenga C, Zanen P, Province MA, Silverman EK, Deary IJ, Palmer LJ, Cassano PA, Gudnason V, Barr RG, Loos RJ, Strachan DP, London SJ, Boezen HM, Probst-Hensch N, Gharib SA, Hall IP, O'Connor GT, Tobin MD, Stricker BH. Genome-wide association studies identify CHRNA5/3 and HTR4 in the development of airflow obstruction. *American journal of respiratory and critical care medicine* 2012; 186: 622-632.
132. Hansel NN, Ruczinski I, Rafaels N, Sin DD, Daley D, Malinina A, Huang L, Sandford A, Murray T, Kim Y, Vergara C, Heckbert SR, Psaty BM, Li G, Elliott WM, Aminuddin F, Dupuis J, O'Connor GT, Doheny K, Scott AF, Boezen HM, Postma DS, Smolonska J, Zanen P, Mohamed Hoesein FA, de Koning HJ, Crystal RG, Tanaka T, Ferrucci L, Silverman E, Wan E, Vestbo J, Lomas DA, Connett J, Wise RA, Neptune ER, Mathias

RA, Pare PD, Beaty TH, Barnes KC. Genome-wide study identifies two loci associated with lung function decline in mild to moderate COPD. *Human genetics* 2013; 132: 79-90.
133. Yao TC, Du G, Han L, Sun Y, Hu D, Yang JJ, Mathias R, Roth LA, Rafaels N, Thompson EE, Loisel DA, Anderson R, Eng C, Arruabarrena Orbegozo M, Young M, Klocksieben JM, Anderson E, Shanovich K, Lester LA, Williams LK, Barnes KC, Burchard EG, Nicolae DL, Abney M, Ober C. Genome-wide association study of lung function phenotypes in a founder population. *The Journal of allergy and clinical immunology* 2014; 133: 248-255.e241-210.
134. Manichaikul A, Hoffman EA, Smolonska J, Gao W, Cho MH, Baumhauer H, Budoff M, Austin JH, Washko GR, Carr JJ, Kaufman JD, Pottinger T, Powell CA, Wijmenga C, Zanen P, Groen HJ, Postma DS, Wanner A, Rouhani FN, Brantly ML, Powell R, Smith BM, Rabinowitz D, Raffel LJ, Hinckley Stukovsky KD, Crapo JD, Beaty TH, Hokanson JE, Silverman EK, Dupuis J, O'Connor GT, Boezen HM, Rich SS, Barr RG. Genome-wide study of percent emphysema on computed tomography in the general population. The Multi-Ethnic Study of Atherosclerosis Lung/SNP Health Association Resource Study. *American journal of respiratory and critical care medicine* 2014; 189: 408-418.
135. Tang W, Kowgier M, Loth DW, Soler Artigas M, Joubert BR, Hodge E, Gharib SA, Smith AV, Ruczinski I, Gudnason V, Mathias RA, Harris TB, Hansel NN, Launer LJ, Barnes KC, Hansen JG, Albrecht E, Aldrich MC, Allerhand M, Barr RG, Brusselle GG, Couper DJ, Curjuric I, Davies G, Deary IJ, Dupuis J, Fall T, Foy M, Franceschini N, Gao W, Glaser S, Gu X, Hancock DB, Heinrich J, Hofman A, Imboden M, Ingelsson E, James A, Karrasch S, Koch B, Kritchevsky SB, Kumar A, Lahousse L, Li G, Lind L, Lindgren C, Liu Y, Lohman K, Lumley T, McArdle WL, Meibohm B, Morris AP, Morrison AC, Musk B, North KE, Palmer LJ, Probst-Hensch NM, Psaty BM, Rivadeneira F, Rotter JI, Schulz H, Smith LJ, Sood A, Starr JM, Strachan DP, Teumer A, Uitterlinden AG, Volzke H, Voorman A, Wain LV, Wells MT, Wilk JB, Williams OD, Heckbert SR, Stricker BH, London SJ, Fornage M, Tobin MD, O'Connor GT, Hall IP, Cassano PA. Large-scale genome-wide association studies and meta-analyses of longitudinal change in adult lung function. *PLoS One* 2014; 9: e100776.
136. Lee JH, McDonald ML, Cho MH, Wan ES, Castaldi PJ, Hunninghake GM, Marchetti N, Lynch DA, Crapo JD, Lomas DA, Coxson HO, Bakke PS, Silverman EK, Hersh CP. DNAH5 is associated with total lung capacity in chronic obstructive pulmonary disease. *Respiratory research* 2014; 15: 97.
137. McDonald ML, Cho MH, Sorheim IC, Lutz SM, Castaldi PJ, Lomas DA, Coxson HO, Edwards LD, MacNee W, Vestbo J, Yates JC, Agusti A, Calverley PM, Celli B, Crim C, Rennard SI, Wouters EF, Bakke P, Tal-Singer R, Miller BE, Gulsvik A, Casaburi R, Wells JM, Regan EA, Make BJ, Hokanson JE, Lange C, Crapo JD, Beaty TH, Silverman EK, Hersh CP. Common genetic variants associated with resting oxygenation in chronic obstructive pulmonary disease. *Am J Respir Cell Mol Biol* 2014; 51: 678-687.
138. Hansel NN, Pare PD, Rafaels N, Sin DD, Sandford A, Daley D, Vergara C, Huang L, Elliott WM, Pascoe CD, Arsenault BA, Postma DS, Boezen HM, Bosse Y, van den Berge M, Hiemstra PS, Cho MH, Litonjua AA, Sparrow D, Ober C, Wise RA, Connett J, Neptune ER, Beaty TH, Ruczinski I, Mathias RA, Barnes KC. Genome-Wide Association Study Identification of Novel Loci Associated with Airway Responsiveness in Chronic Obstructive Pulmonary Disease. *Am J Respir Cell Mol Biol* 2015; 53: 226-234.

139. Cho MH, Castaldi PJ, Hersh CP, Hobbs BD, Barr RG, Tal-Singer R, Bakke P, Gulsvik A, San Jose Estepar R, Van Beek EJ, Coxson HO, Lynch DA, Washko GR, Laird NM, Crapo JD, Beaty TH, Silverman EK. A Genome-Wide Association Study of Emphysema and Airway Quantitative Imaging Phenotypes. *American journal of respiratory and critical care medicine* 2015; 192: 559-569.
140. Lutz SM, Cho MH, Young K, Hersh CP, Castaldi PJ, McDonald ML, Regan E, Mattheisen M, DeMeo DL, Parker M, Foreman M, Make BJ, Jensen RL, Casaburi R, Lomas DA, Bhatt SP, Bakke P, Gulsvik A, Crapo JD, Beaty TH, Laird NM, Lange C, Hokanson JE, Silverman EK. A genome-wide association study identifies risk loci for spirometric measures among smokers of European and African ancestry. *BMC genetics* 2015; 16: 138.
141. Wain LV, Shrine N, Miller S, Jackson VE, Ntalla I, Soler Artigas M, Billington CK, Kheirallah AK, Allen R, Cook JP, Probert K, Obeidat M, Bosse Y, Hao K, Postma DS, Pare PD, Ramasamy A, Magi R, Mihailov E, Reinmaa E, Melen E, O'Connell J, Frangou E, Delaneau O, Freeman C, Petkova D, McCarthy M, Sayers I, Deloukas P, Hubbard R, Pavord I, Hansell AL, Thomson NC, Zeggini E, Morris AP, Marchini J, Strachan DP, Tobin MD, Hall IP. Novel insights into the genetics of smoking behaviour, lung function, and chronic obstructive pulmonary disease (UK BiLEVE): a genetic association study in UK Biobank. *The lancet Respiratory medicine* 2015; 3: 769-781.
142. Hormozdiari F, Kostem E, Kang EY, Pasaniuc B, Eskin E. Identifying Causal Variants at Loci with Multiple Signals of Association. *Genetics* 2014; 198: 497-508.
143. Duerr RH, Taylor KD, Brant SR, Rioux JD, Silverberg MS, Daly MJ, Steinhart AH, Abraham C, Regueiro M, Griffiths A, Dassopoulos T, Bitton A, Yang H, Targan S, Datta LW, Kistner EO, Schumm LP, Lee AT, Gregersen PK, Barmada MM, Rotter JI, Nicolae DL, Cho JH. A genome-wide association study identifies IL23R as an inflammatory bowel disease gene. *Science (New York, NY)* 2006; 314: 1461-1463.
144. Siakavellas SI, Bamias G. Role of the IL-23/IL-17 axis in Crohn's disease. *Discovery medicine* 2012; 14: 253-262.
145. Hersh CP, DeMeo DL, Raby BA, Litonjua AA, Sylvia JS, Sparrow D, Reilly JJ, Silverman EK. Genetic linkage and association analysis of COPD-related traits on chromosome 8p. *Copd* 2006; 3: 189-194.
146. Zhu G, Warren L, Aponte J, Gulsvik A, Bakke P, Anderson WH, Lomas DA, Silverman EK, Pillai SG. The SERPINE2 Gene Is Associated with Chronic Obstructive Pulmonary Disease in Two Large Populations. *American journal of respiratory and critical care medicine* 2007; 176: 167-173.
147. Hodge E, Nelson CP, Miller S, Billington CK, Stewart CE, Swan C, Malarstig A, Henry AP, Gowland C, Melen E, Hall IP, Sayers I. HTR4 gene structure and altered expression in the developing lung. *Respiratory research* 2013; 14: 77.
148. House JS, Li H, DeGraff LM, Flake G, Zeldin DC, London SJ. Genetic variation in HTR4 and lung function: GWAS follow-up in mouse. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology* 2015; 29: 323-335.
149. Zhou X, Baron RM, Hardin M, Cho MH, Zielinski J, Hawrylkiewicz I, Sliwinski P, Hersh CP, Mancini JD, Lu K, Thibault D, Donahue AL, Klanderman BJ, Rosner B, Raby BA, Lu Q, Geldart AM, Layne MD, Perrella MA, Weiss ST, Choi AM, Silverman EK.

Identification of a chronic obstructive pulmonary disease genetic determinant that regulates HHIP. *Human molecular genetics* 2012; 21: 1325-1335.
150. Lao T, Glass K, Qiu W, Polverino F, Gupta K, Morrow J, Mancini JD, Vuong L, Perrella MA, Hersh CP, Owen CA, Quackenbush J, Yuan GC, Silverman EK, Zhou X. Haploinsufficiency of Hedgehog interacting protein causes increased emphysema induced by cigarette smoke through network rewiring. *Genome medicine* 2015; 7: 12.
151. Lao T, Jiang Z, Yun J, Qiu W, Guo F, Huang C, Mancini JD, Gupta K, Laucho-Contreras ME, Naing ZZ, Zhang L, Perrella MA, Owen CA, Silverman EK, Zhou X. Hhip haploinsufficiency sensitizes mice to age-related emphysema. *Proc Natl Acad Sci U S A* 2016.
152. So HC, Gui AH, Cherny SS, Sham PC. Evaluating the heritability explained by known susceptibility variants: a survey of ten complex diseases. *Genetic epidemiology* 2011; 35: 310-317.
153. Gibson G. Rare and common variants: twenty arguments. *Nature reviews Genetics* 2011; 13: 135-145.
154. Spencer CC, Su Z, Donnelly P, Marchini J. Designing genome-wide association studies: sample size, power, imputation, and the choice of genotyping chip. *PLoS Genet* 2009; 5: e1000477.
155. Pritchard JK, Cox NJ. The allelic architecture of human disease genes: common disease–common variant…or not? *Human molecular genetics* 2002; 11: 2417-2423.
156. Visscher PM, Brown MA, McCarthy MI, Yang J. Five years of GWAS discovery. *American journal of human genetics* 2012; 90: 7-24.
157. Wjst M, Sargurupremraj M, Arnold M. Genome-wide association studies in asthma: what they really told us about pathogenesis. *Current opinion in allergy and clinical immunology* 2013; 13: 112-118.
158. Gibson G. Hints of hidden heritability in GWAS. *Nat Genet* 2010; 42: 558-560.
159. Visscher PM. Sizing up human height variation. *Nat Genet* 2008; 40: 489-490.
160. Wood AR, Esko T, Yang J, Vedantam S, Pers TH, Gustafsson S, Chu AY, Estrada K, Luan Ja, Kutalik Z, Amin N, Buchkovich ML, Croteau-Chonka DC, Day FR, Duan Y, Fall T, Fehrmann R, Ferreira T, Jackson AU, Karjalainen J, Lo KS, Locke AE, Magi R, Mihailov E, Porcu E, Randall JC, Scherag A, Vinkhuyzen AAE, Westra H-J, Winkler TW, Workalemahu T, Zhao JH, Absher D, Albrecht E, Anderson D, Baron J, Beekman M, Demirkan A, Ehret GB, Feenstra B, Feitosa MF, Fischer K, Fraser RM, Goel A, Gong J, Justice AE, Kanoni S, Kleber ME, Kristiansson K, Lim U, Lotay V, Lui JC, Mangino M, Leach IM, Medina-Gomez C, Nalls MA, Nyholt DR, Palmer CD, Pasko D, Pechlivanis S, Prokopenko I, Ried JS, Ripke S, Shungin D, Stancakova A, Strawbridge RJ, Sung YJ, Tanaka T, Teumer A, Trompet S, van der Laan SW, van Setten J, Van Vliet-Ostaptchouk JV, Wang Z, Yengo L, Zhang W, Afzal U, Arnlov J, Arscott GM, Bandinelli S, Barrett A, Bellis C, Bennett AJ, Berne C, Bluher M, Bolton JL, Bottcher Y, Boyd HA, Bruinenberg M, Buckley BM, Buyske S, Caspersen IH, Chines PS, Clarke R, Claudi-Boehm S, Cooper M, Daw EW, De Jong PA, Deelen J, Delgado G, Denny JC, Dhonukshe-Rutten R, Dimitriou M, Doney ASF, Dorr M, Eklund N, Eury E, Folkersen L, Garcia ME, Geller F, Giedraitis V, Go AS, Grallert H, Grammer TB, Graszler J, Gronberg H, de Groot LCPGM, Groves CJ, Haessler J, Hall P, Haller T, Hallmans G, Hannemann A, Hartman CA, Hassinen M, Hayward C, Heard-Costa NL, Helmer Q,

Hemani G, Henders AK, Hillege HL, Hlatky MA, Hoffmann W, Hoffmann P, Holmen O, Houwing-Duistermaat JJ, Illig T, Isaacs A, James AL, Jeff J, Johansen B, Johansson A, Jolley J, Juliusdottir T, Junttila J, Kho AN, Kinnunen L, Klopp N, Kocher T, Kratzer W, Lichtner P, Lind L, Lindstrom J, Lobbens S, Lorentzon M, Lu Y, Lyssenko V, Magnusson PKE, Mahajan A, Maillard M, McArdle WL, McKenzie CA, McLachlan S, McLaren PJ, Menni C, Merger S, Milani L, Moayyeri A, Monda KL, Morken MA, Muller G, Muller-Nurasyid M, Musk AW, Narisu N, Nauck M, Nolte IM, Nothen MM, Oozageer L, Pilz S, Rayner NW, Renstrom F, Robertson NR, Rose LM, Roussel R, Sanna S, Scharnagl H, Scholtens S, Schumacher FR, Schunkert H, Scott RA, Sehmi J, Seufferlein T, Shi J, Silventoinen K, Smit JH, Smith AV, Smolonska J, Stanton AV, Stirrups K, Stott DJ, Stringham HM, Sundstrom J, Swertz MA, Syvanen A-C, Tayo BO, Thorleifsson G, Tyrer JP, van Dijk S, van Schoor NM, van der Velde N, van Heemst D, van Oort FVA, Vermeulen SH, Verweij N, Vonk JM, Waite LL, Waldenberger M, Wennauer R, Wilkens LR, Willenborg C, Wilsgaard T, Wojczynski MK, Wong A, Wright AF, Zhang Q, Arveiler D, Bakker SJL, Beilby J, Bergman RN, Bergmann S, Biffar R, Blangero J, Boomsma DI, Bornstein SR, Bovet P, Brambilla P, Brown MJ, Campbell H, Caulfield MJ, Chakravarti A, Collins R, Collins FS, Crawford DC, Cupples LA, Danesh J, de Faire U, den Ruijter HM, Erbel R, Erdmann J, Eriksson JG, Farrall M, Ferrannini E, Ferrieres J, Ford I, Forouhi NG, Forrester T, Gansevoort RT, Gejman PV, Gieger C, Golay A, Gottesman O, Gudnason V, Gyllensten U, Haas DW, Hall AS, Harris TB, Hattersley AT, Heath AC, Hengstenberg C, Hicks AA, Hindorff LA, Hingorani AD, Hofman A, Hovingh GK, Humphries SE, Hunt SC, Hypponen E, Jacobs KB, Jarvelin M-R, Jousilahti P, Jula AM, Kaprio J, Kastelein JJP, Kayser M, Kee F, Keinanen-Kiukaanniemi SM, Kiemeney LA, Kooner JS, Kooperberg C, Koskinen S, Kovacs P, Kraja AT, Kumari M, Kuusisto J, Lakka TA, Langenberg C, Le Marchand L, Lehtimaki T, Lupoli S, Madden PAF, Mannisto S, Manunta P, Marette A, Matise TC, McKnight B, Meitinger T, Moll FL, Montgomery GW, Morris AD, Morris AP, Murray JC, Nelis M, Ohlsson C, Oldehinkel AJ, Ong KK, Ouwehand WH, Pasterkamp G, Peters A, Pramstaller PP, Price JF, Qi L, Raitakari OT, Rankinen T, Rao DC, Rice TK, Ritchie M, Rudan I, Salomaa V, Samani NJ, Saramies J, Sarzynski MA, Schwarz PEH, Sebert S, Sever P, Shuldiner AR, Sinisalo J, Steinthorsdottir V, Stolk RP, Tardif J-C, Tonjes A, Tremblay A, Tremoli E, Virtamo J, Vohl M-C, The Electronic Medical R, Genomics C, The MC, The PC, The LifeLines Cohort S, Amouyel P, Asselbergs FW, Assimes TL, Bochud M, Boehm BO, Boerwinkle E, Bottinger EP, Bouchard C, Cauchi S, Chambers JC, Chanock SJ, Cooper RS, de Bakker PIW, Dedoussis G, Ferrucci L, Franks PW, Froguel P, Groop LC, Haiman CA, Hamsten A, Hayes MG, Hui J, Hunter DJ, Hveem K, Jukema JW, Kaplan RC, Kivimaki M, Kuh D, Laakso M, Liu Y, Martin NG, Marz W, Melbye M, Moebus S, Munroe PB, Njolstad I, Oostra BA, Palmer CNA, Pedersen NL, Perola M, Perusse L, Peters U, Powell JE, Power C, Quertermous T, Rauramaa R, Reinmaa E, Ridker PM, Rivadeneira F, Rotter JI, Saaristo TE, Saleheen D, Schlessinger D, Slagboom PE, Snieder H, Spector TD, Strauch K, Stumvoll M, Tuomilehto J, Uusitupa M, van der Harst P, Volzke H, Walker M, Wareham NJ, Watkins H, Wichmann HE, Wilson JF, Zanen P, Deloukas P, Heid IM, Lindgren CM, Mohlke KL, Speliotes EK, Thorsteinsdottir U, Barroso I, Fox CS, North KE, Strachan DP, Beckmann JS, Berndt SI, Boehnke M, Borecki IB, McCarthy MI, Metspalu A, Stefansson K, Uitterlinden AG, van

Duijn CM, Franke L, Willer CJ, Price AL, Lettre G, Loos RJF, Weedon MN, Ingelsson E, O'Connell JR, Abecasis GR, Chasman DI, Goddard ME, Visscher PM, Hirschhorn JN, Frayling TM. Defining the role of common variation in the genomic and biological architecture of adult human height. *Nat Genet* 2014; 46: 1173-1186.
161. Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, Nyholt DR, Madden PA, Heath AC, Martin NG, Montgomery GW, Goddard ME, Visscher PM. Common SNPs explain a large proportion of heritability for human height. *Nature genetics* 2010; 42: 565-569.
162. Zuk O, Schaffner SF, Samocha K, Do R, Hechter E, Kathiresan S, Daly MJ, Neale BM, Sunyaev SR, Lander ES. Searching for missing heritability: designing rare variant association studies. *Proc Natl Acad Sci U S A* 2014; 111: E455-464.
163. van Leeuwen EM, Sabo A, Bis JC, Huffman JE, Manichaikul A, Smith AV, Feitosa MF, Demissie S, Joshi PK, Duan Q, Marten J, van Klinken JB, Surakka I, Nolte IM, Zhang W, Mbarek H, Li-Gao R, Trompet S, Verweij N, Evangelou E, Lyytikainen LP, Tayo BO, Deelen J, van der Most PJ, van der Laan SW, Arking DE, Morrison A, Dehghan A, Franco OH, Hofman A, Rivadeneira F, Sijbrands EJ, Uitterlinden AG, Mychaleckyj JC, Campbell A, Hocking LJ, Padmanabhan S, Brody JA, Rice KM, White CC, Harris T, Isaacs A, Campbell H, Lange LA, Rudan I, Kolcic I, Navarro P, Zemunik T, Salomaa V, Kooner AS, Kooner JS, Lehne B, Scott WR, Tan ST, de Geus EJ, Milaneschi Y, Penninx BW, Willemsen G, de Mutsert R, Ford I, Gansevoort RT, Segura-Lepe MP, Raitakari OT, Viikari JS, Nikus K, Forrester T, McKenzie CA, de Craen AJ, de Ruijter HM, Pasterkamp G, Snieder H, Oldehinkel AJ, Slagboom PE, Cooper RS, Kahonen M, Lehtimaki T, Elliott P, van der Harst P, Jukema JW, Mook-Kanamori DO, Boomsma DI, Chambers JC, Swertz M, Ripatti S, Willems van Dijk K, Vitart V, Polasek O, Hayward C, Wilson JG, Wilson JF, Gudnason V, Rich SS, Psaty BM, Borecki IB, Boerwinkle E, Rotter JI, Cupples LA, van Duijn CM. Meta-analysis of 49 549 individuals imputed with the 1000 Genomes Project reveals an exonic damaging variant in ANGPTL4 determining fasting TG levels. *Journal of medical genetics* 2016; 53: 441-449.
164. Chami N, Chen MH, Slater AJ, Eicher JD, Evangelou E, Tajuddin SM, Love-Gregory L, Kacprowski T, Schick UM, Nomura A, Giri A, Lessard S, Brody JA, Schurmann C, Pankratz N, Yanek LR, Manichaikul A, Pazoki R, Mihailov E, Hill WD, Raffield LM, Burt A, Bartz TM, Becker DM, Becker LC, Boerwinkle E, Bork-Jensen J, Bottinger EP, O'Donoghue ML, Crosslin DR, de Denus S, Dube MP, Elliott P, Engstrom G, Evans MK, Floyd JS, Fornage M, Gao H, Greinacher A, Gudnason V, Hansen T, Harris TB, Hayward C, Hernesniemi J, Highland HM, Hirschhorn JN, Hofman A, Irvin MR, Kahonen M, Lange E, Launer LJ, Lehtimaki T, Li J, Liewald DC, Linneberg A, Liu Y, Lu Y, Lyytikainen LP, Magi R, Mathias RA, Melander O, Metspalu A, Mononen N, Nalls MA, Nickerson DA, Nikus K, O'Donnell CJ, Orho-Melander M, Pedersen O, Petersmann A, Polfus L, Psaty BM, Raitakari OT, Raitoharju E, Richard M, Rice KM, Rivadeneira F, Rotter JI, Schmidt F, Smith AV, Starr JM, Taylor KD, Teumer A, Thuesen BH, Torstenson ES, Tracy RP, Tzoulaki I, Zakai NA, Vacchi-Suzzi C, van Duijn CM, van Rooij FJ, Cushman M, Deary IJ, Velez Edwards DR, Vergnaud AC, Wallentin L, Waterworth DM, White HD, Wilson JG, Zonderman AB, Kathiresan S, Grarup N, Esko T, Loos RJ, Lange LA, Faraday N, Abumrad NA, Edwards TL, Ganesh SK, Auer PL, Johnson AD, Reiner AP, Lettre G. Exome Genotyping Identifies

Pleiotropic Variants Associated with Red Blood Cell Traits. *American journal of human genetics* 2016; 99: 8-21.
165. Steinthorsdottir V, Thorleifsson G, Sulem P. Identification of low-frequency and rare sequence variants associated with elevated or reduced risk of type 2 diabetes. 2014; 46: 294-298.
166. Singh T, Kurki MI, Curtis D, Purcell SM, Crooks L, McRae J, Suvisaari J, Chheda H, Blackwood D, Breen G, Pietilainen O, Gerety SS, Ayub M, Blyth M, Cole T, Collier D, Coomber EL, Craddock N, Daly MJ, Danesh J, DiForti M, Foster A, Freimer NB, Geschwind D, Johnstone M, Joss S, Kirov G, Korkko J, Kuismin O, Holmans P, Hultman CM, Iyegbe C, Lonnqvist J, Mannikko M, McCarroll SA, McGuffin P, McIntosh AM, McQuillin A, Moilanen JS, Moore C, Murray RM, Newbury-Ecob R, Ouwehand W, Paunio T, Prigmore E, Rees E, Roberts D, Sambrook J, Sklar P, Clair DS, Veijola J, Walters JT, Williams H, Sullivan PF, Hurles ME, O'Donovan MC, Palotie A, Owen MJ, Barrett JC. Rare loss-of-function variants in SETD1A are associated with schizophrenia and developmental disorders. *Nature neuroscience* 2016; 19: 571-577.
167. Dickson SP, Wang K, Krantz I, Hakonarson H, Goldstein DB. Rare Variants Create Synthetic Genome-Wide Associations. *PLoS biology* 2010; 8: e1000294.
168. Anderson CA, Soranzo N, Zeggini E, Barrett JC. Synthetic associations are unlikely to account for many common disease genome-wide association signals. *PLoS biology* 2011; 9: e1000580.
169. Wray NR, Purcell SM, Visscher PM. Synthetic associations created by rare variants do not explain most GWAS results. *PLoS biology* 2011; 9: e1000579.
170. Orozco G, Barrett JC, Zeggini E. Synthetic associations in the context of genome-wide association scan signals. *Human molecular genetics* 2010; 19: R137-144.
171. Mills RE, Walter K, Stewart C, Handsaker RE, Chen K, Alkan C, Abyzov A, Yoon SC, Ye K, Cheetham RK, Chinwalla A, Conrad DF, Fu Y, Grubert F, Hajirasouliha I, Hormozdiari F, Iakoucheva LM, Iqbal Z, Kang S, Kidd JM, Konkel MK, Korn J, Khurana E, Kural D, Lam HY, Leng J, Li R, Li Y, Lin CY, Luo R, Mu XJ, Nemesh J, Peckham HE, Rausch T, Scally A, Shi X, Stromberg MP, Stutz AM, Urban AE, Walker JA, Wu J, Zhang Y, Zhang ZD, Batzer MA, Ding L, Marth GT, McVean G, Sebat J, Snyder M, Wang J, Ye K, Eichler EE, Gerstein MB, Hurles ME, Lee C, McCarroll SA, Korbel JO. Mapping copy number variation by population-scale genome sequencing. *Nature* 2011; 470: 59-65.
172. Raphael BJ. Chapter 6: Structural variation and medical genomics. *PLoS computational biology* 2012; 8: e1002821.
173. Sullivan PF, Daly MJ, O'Donovan M. Genetic architectures of psychiatric disorders: the emerging picture and its implications. *Nature reviews Genetics* 2012; 13: 537-551.
174. Zuk O, Hechter E, Sunyaev SR, Lander ES. The mystery of missing heritability: Genetic interactions create phantom heritability. *Proceedings of the National Academy of Sciences* 2012; 109: 1193-1198.
175. Jiang X, Barmada MM, Visweswaran S. Identifying genetic interactions in genome-wide data using Bayesian networks. *Genetic epidemiology* 2010; 34: 575-581.
176. Kaprio J. Twins and the mystery of missing heritability: the contribution of gene-environment interactions. *Journal of internal medicine* 2012; 272: 440-448.

177. Rappaport SM. Genetic Factors Are Not the Major Causes of Chronic Diseases. *PLoS ONE* 2016; 11: e0154387.
178. Trerotola M, Relli V, Simeone P, Alberti S. Epigenetic inheritance and the missing heritability. *Human genomics* 2015; 9: 17.
179. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, McCarthy MI, Ramos EM, Cardon LR, Chakravarti A, Cho JH, Guttmacher AE, Kong A, Kruglyak L, Mardis E, Rotimi CN, Slatkin M, Valle D, Whittemore AS, Boehnke M, Clark AG, Eichler EE, Gibson G, Haines JL, Mackay TF, McCarroll SA, Visscher PM. Finding the missing heritability of complex diseases. *Nature* 2009; 461: 747-753.
180. Wain LV, Sayers I, Soler Artigas M, Portelli MA, Zeggini E, Obeidat Me, Sin DD, Bossé Y, Nickle D, Brandsma C-A, Malarstig A, Vangjeli C, Jelinsky SA, John S, Kilty I, McKeever T, Shrine NRG, Cook JP, Patel S, Spector TD, Hollox EJ, Hall IP, Tobin MD. Whole Exome Re-Sequencing Implicates <italic>CCDC38</italic> and Cilia Structure and Function in Resistance to Smoking Related Airflow Obstruction. *PLoS Genet* 2014; 10: e1004314.
181. Bruse S, Moreau M, Bromberg Y, Jang JH, Wang N, Ha H, Picchi M, Lin Y, Langley RJ, Qualls C, Klensney-Tait J, Zabner J, Leng S, Mao J, Belinsky SA, Xing J, Nyunoya T. Whole exome sequencing identifies novel candidate genes that modify chronic obstructive pulmonary disease susceptibility. *Human genomics* 2016; 10: 1.
182. Qiao D, Lange C, Beaty TH, Crapo JD, Barnes KC, Bamshad M, Hersh CP, Morrow J, Pinto-Plata VM, Marchetti N, Bueno R, Celli BR, Criner GJ, Silverman EK, Cho MH. Exome Sequencing Analysis in Severe, Early-Onset Chronic Obstructive Pulmonary Disease. *American journal of respiratory and critical care medicine* 2016.
183. Jackson VE, Ntalla I, Sayers I, Morris R, Whincup P, Casas J-P, Amuzu A, Choi M, Dale C, Kumari M, Engmann J, Kalsheker N, Chappell S, Guetta-Baranes T, McKeever TM, Palmer CNA, Tavendale R, Holloway JW, Sayer AA, Dennison EM, Cooper C, Bafadhel M, Barker B, Brightling C, Bolton CE, John ME, Parker SG, Moffat MF, Wardlaw AJ, Connolly MJ, Porteous DJ, Smith BH, Padmanabhan S, Hocking L, Stirrups KE, Deloukas P, Strachan DP, Hall IP, Tobin MD, Wain LV. Exome-wide analysis of rare coding variation identifies novel associations with COPD and airflow limitation in MOCS3, IFIT3 and SERPINA12. *Thorax* 2016; 71: 501-509.
184. Hobbs BD, Parker MM, Chen H, Lao T, Hardin M, Qiao D, Hawrylkiewicz I, Sliwinski P, Yim JJ, Kim WJ, Kim DK, Castaldi PJ, Hersh CP, Morrow J, Celli BR, Pinto-Plata VM, Criner GJ, Marchetti N, Bueno R, Agusti A, Make BJ, Crapo JD, Calverley PM, Donner CF, Lomas DA, Wouters EF, Vestbo J, Pare PD, Levy RD, Rennard SI, Zhou X, Laird NM, Lin X, Beaty TH, Silverman EK, Cho MH. Exome Array Analysis Identifies A Common Variant in IL27 Associated with Chronic Obstructive Pulmonary Disease. *American journal of respiratory and critical care medicine* 2016.
185. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, Funke R, Gage D, Harris K, Heaford A, Howland J, Kann L, Lehoczky J, LeVine R, McEwan P, McKernan K, Meldrim J, Mesirov JP, Miranda C, Morris W, Naylor J, Raymond C, Rosetti M, Santos R, Sheridan A, Sougnez C, Stange-Thomann N, Stojanovic N, Subramanian A, Wyman D, Rogers J, Sulston J, Ainscough R, Beck S, Bentley D, Burton J, Clee C, Carter N, Coulson A, Deadman R, Deloukas P, Dunham A, Dunham I, Durbin R, French L, Grafham D, Gregory S, Hubbard T,

Humphray S, Hunt A, Jones M, Lloyd C, McMurray A, Matthews L, Mercer S, Milne S, Mullikin JC, Mungall A, Plumb R, Ross M, Shownkeen R, Sims S, Waterston RH, Wilson RK, Hillier LW, McPherson JD, Marra MA, Mardis ER, Fulton LA, Chinwalla AT, Pepin KH, Gish WR, Chissoe SL, Wendl MC, Delehaunty KD, Miner TL, Delehaunty A, Kramer JB, Cook LL, Fulton RS, Johnson DL, Minx PJ, Clifton SW, Hawkins T, Branscomb E, Predki P, Richardson P, Wenning S, Slezak T, Doggett N, Cheng JF, Olsen A, Lucas S, Elkin C, Uberbacher E, Frazier M, Gibbs RA, Muzny DM, Scherer SE, Bouck JB, Sodergren EJ, Worley KC, Rives CM, Gorrell JH, Metzker ML, Naylor SL, Kucherlapati RS, Nelson DL, Weinstock GM, Sakaki Y, Fujiyama A, Hattori M, Yada T, Toyoda A, Itoh T, Kawagoe C, Watanabe H, Totoki Y, Taylor T, Weissenbach J, Heilig R, Saurin W, Artiguenave F, Brottier P, Bruls T, Pelletier E, Robert C, Wincker P, Smith DR, Doucette-Stamm L, Rubenfield M, Weinstock K, Lee HM, Dubois J, Rosenthal A, Platzer M, Nyakatura G, Taudien S, Rump A, Yang H, Yu J, Wang J, Huang G, Gu J, Hood L, Rowen L, Madan A, Qin S, Davis RW, Federspiel NA, Abola AP, Proctor MJ, Myers RM, Schmutz J, Dickson M, Grimwood J, Cox DR, Olson MV, Kaul R, Raymond C, Shimizu N, Kawasaki K, Minoshima S, Evans GA, Athanasiou M, Schultz R, Roe BA, Chen F, Pan H, Ramser J, Lehrach H, Reinhardt R, McCombie WR, de la Bastide M, Dedhia N, Blocker H, Hornischer K, Nordsiek G, Agarwala R, Aravind L, Bailey JA, Bateman A, Batzoglou S, Birney E, Bork P, Brown DG, Burge CB, Cerutti L, Chen HC, Church D, Clamp M, Copley RR, Doerks T, Eddy SR, Eichler EE, Furey TS, Galagan J, Gilbert JG, Harmon C, Hayashizaki Y, Haussler D, Hermjakob H, Hokamp K, Jang W, Johnson LS, Jones TA, Kasif S, Kaspryzk A, Kennedy S, Kent WJ, Kitts P, Koonin EV, Korf I, Kulp D, Lancet D, Lowe TM, McLysaght A, Mikkelsen T, Moran JV, Mulder N, Pollara VJ, Ponting CP, Schuler G, Schultz J, Slater G, Smit AF, Stupka E, Szustakowski J, Thierry-Mieg D, Thierry-Mieg J, Wagner L, Wallis J, Wheeler R, Williams A, Wolf YI, Wolfe KH, Yang SP, Yeh RF, Collins F, Guyer MS, Peterson J, Felsenfeld A, Wetterstrand KA, Patrinos A, Morgan MJ, de Jong P, Catanese JJ, Osoegawa K, Shizuya H, Choi S, Chen YJ. Initial sequencing and analysis of the human genome. *Nature* 2001; 409: 860-921.

186. Hoffman EP. The Evolving Genome Project: current and future impact. *American journal of human genetics* 1994; 54: 129-136.
187. Green ED, Waterston RH. The human genome project. Prospects and implications for clinical medicine. *JAMA : the journal of the American Medical Association* 1991; 266: 1966-1975.
188. Langreth R, Waldholz M. New era of personalized medicine: targeting drugs for each unique genetic profile. *The oncologist* 1999; 4: 426-427.
189. Shastry BS. Pharmacogenetics and the concept of individualized medicine. *The pharmacogenomics journal* 2006; 6: 16-21.
190. Silverman EK, Loscalzo J. Developing new drug treatments in the era of network medicine. *Clinical pharmacology and therapeutics* 2013; 93: 26-28.
191. Hall IP. Stratified medicine: drugs meet genetics. *European respiratory review : an official journal of the European Respiratory Society* 2013; 22: 53-57.
192. Tian Q, Price ND, Hood L. Systems cancer medicine: towards realization of predictive, preventive, personalized and participatory (P4) medicine. *Journal of internal medicine* 2012; 271: 111-121.

193. Robinson PN. Deep phenotyping for precision medicine. *Human mutation* 2012; 33: 777-780.
194. National Research Council Committee on AFfDaNToD. The National Academies Collection: Reports funded by National Institutes of Health. Toward Precision Medicine: Building a Knowledge Network for Biomedical Research and a New Taxonomy of Disease. Washington (DC): National Academies Press (US)
National Academy of Sciences.; 2011.
195. Hall JM, Lee MK, Newman B, Morrow JE, Anderson LA, Huey B, King MC. Linkage of early-onset familial breast cancer to chromosome 17q21. *Science (New York, NY)* 1990; 250: 1684-1689.
196. King MC, Marks JH, Mandell JB. Breast and ovarian cancer risks due to inherited mutations in BRCA1 and BRCA2. *Science (New York, NY)* 2003; 302: 643-646.
197. Balachandran VP, Dematteo RP. Targeted therapy for cancer: the gastrointestinal stromal tumor model. *Surgical oncology clinics of North America* 2013; 22: 805-821.
198. Cuzick J, Sestak I, Bonanni B, Costantino JP, Cummings S, DeCensi A, Dowsett M, Forbes JF, Ford L, LaCroix AZ, Mershon J, Mitlak BH, Powles T, Veronesi U, Vogel V, Wickerham DL. Selective oestrogen receptor modulators in prevention of breast cancer: an updated meta-analysis of individual participant data. *Lancet* 2013; 381: 1827-1834.
199. Jelovac D, Emens LA. HER2-directed therapy for metastatic breast cancer. *Oncology (Williston Park, NY)* 2013; 27: 166-175.
200. Chapman PB, Hauschild A, Robert C, Haanen JB, Ascierto P, Larkin J, Dummer R, Garbe C, Testori A, Maio M, Hogg D, Lorigan P, Lebbe C, Jouary T, Schadendorf D, Ribas A, O'Day SJ, Sosman JA, Kirkwood JM, Eggermont AM, Dreno B, Nolop K, Li J, Nelson B, Hou J, Lee RJ, Flaherty KT, McArthur GA. Improved survival with vemurafenib in melanoma with BRAF V600E mutation. *The New England journal of medicine* 2011; 364: 2507-2516.
201. Paez JG, Janne PA, Lee JC, Tracy S, Greulich H, Gabriel S, Herman P, Kaye FJ, Lindeman N, Boggon TJ, Naoki K, Sasaki H, Fujii Y, Eck MJ, Sellers WR, Johnson BE, Meyerson M. EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. *Science (New York, NY)* 2004; 304: 1497-1500.
202. Nikiforov YE, Seethala RR, Tallini G, Baloch ZW, Basolo F, Thompson LD, Barletta JA, Wenig BM, Al Ghuzlan A, Kakudo K, Giordano TJ, Alves VA, Khanafshar E, Asa SL, El-Naggar AK, Gooding WE, Hodak SP, Lloyd RV, Maytal G, Mete O, Nikiforova MN, Nose V, Papotti M, Poller DN, Sadow PM, Tischler AS, Tuttle RM, Wall KB, LiVolsi VA, Randolph GW, Ghossein RA. Nomenclature Revision for Encapsulated Follicular Variant of Papillary Thyroid Carcinoma: A Paradigm Shift to Reduce Overtreatment of Indolent Tumors. *JAMA oncology* 2016.
203. Ritchie MD. The success of pharmacogenomics in moving genetic association studies from bench to bedside: study design and implementation of precision medicine in the post-GWAS era. *Human genetics* 2012; 131: 1615-1626.
204. Becquemont L. HLA: a pharmacogenomics success story. *Pharmacogenomics* 2010; 11: 277-281.
205. Agusti A, MacNee W. The COPD control panel: towards personalised medicine in COPD. *Thorax* 2013; 68: 687-690.

206. Mardis ER. The impact of next-generation sequencing technology on genetics. *Trends in Genetics* 2008; 24: 133-141.
207. Cirulli ET, Goldstein DB. Uncovering the roles of rare variants in common disease through whole-genome sequencing. *Nature reviews Genetics* 2010; 11: 415-425.
208. Zuk O, Schaffner SF, Samocha K, Do R, Hechter E, Kathiresan S, Daly MJ, Neale BM, Sunyaev SR, Lander ES. Searching for missing heritability: Designing rare variant association studies. *Proceedings of the National Academy of Sciences* 2014; 111: E455-E464.
209. Barnett IJ, Lee S, Lin X. Detecting Rare Variant Effects Using Extreme Phenotype Sampling in Sequencing Association Studies. *Genetic epidemiology* 2013; 37: 142-151.
210. Bon J, Fuhrman CR, Weissfeld JL, Duncan SR, Branch RA, Chang CC, Zhang Y, Leader JK, Gur D, Greenspan SL, Sciurba FC. Radiographic emphysema predicts low bone mineral density in a tobacco-exposed cohort. *American journal of respiratory and critical care medicine* 2011; 183: 885-890.
211. Muller NL, Staples CA, Miller RR, Abboud RT. "Density mask". An objective method to quantitate emphysema using computed tomography. *Chest* 1988; 94: 782-787.
212. Allen EG, Grus WE, Narayan S, Espinel W, Sherman SL. Approaches to identify genetic variants that influence the risk for onset of fragile X-associated primary ovarian insufficiency (FXPOI): a preliminary study. *Frontiers in genetics* 2014; 5: 260.
213. Yuen RK, Thiruvahindrapuram B, Merico D, Walker S, Tammimies K, Hoang N, Chrysler C, Nalpathamkalam T, Pellecchia G, Liu Y, Gazzellone MJ, D'Abate L, Deneault E, Howe JL, Liu RS, Thompson A, Zarrei M, Uddin M, Marshall CR, Ring RH, Zwaigenbaum L, Ray PN, Weksberg R, Carter MT, Fernandez BA, Roberts W, Szatmari P, Scherer SW. Whole-genome sequencing of quartet families with autism spectrum disorder. *Nature medicine* 2015; 21: 185-191.
214. De Gregori M, Diatchenko L, Belfer I, Allegri M. OPRM1 receptor as new biomarker to help the prediction of post mastectomy pain and recurrence in breast cance. *Minerva anestesiologica* 2014.
215. Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, Del Angel G, Levy-Moonshine A, Jordan T, Shakir K, Roazen D, Thibault J, Banks E, Garimella KV, Altshuler D, Gabriel S, DePristo MA. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Current protocols in bioinformatics / editoral board, Andreas D Baxevanis [et al]* 2013; 11: 11.10.11-11.10.33.
216. Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics (Oxford, England)* 2009; 25: 1754-1760.
217. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome research* 2010; 20: 1297-1303.
218. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. The Sequence Alignment/Map format and SAMtools. *Bioinformatics (Oxford, England)* 2009; 25: 2078-2079.
219. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC. PLINK: a tool set for whole-genome association and

population-based linkage analyses. *American journal of human genetics* 2007; 81: 559-575.
220. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* 2006; 38: 904-909.
221. Kang HM, Sul JH, Service SK, Zaitlen NA, Kong SY, Freimer NB, Sabatti C, Eskin E. Variance component model to account for sample structure in genome-wide association studies. *Nat Genet* 2010; 42: 348-354.
222. Lee S, Emond MJ, Bamshad MJ, Barnes KC, Rieder MJ, Nickerson DA, Christiani DC, Wurfel MM, Lin X. Optimal unified approach for rare-variant association testing with application to small-sample case-control whole-exome sequencing studies. *American journal of human genetics* 2012; 91: 224-237.
223. Belkadi A, Bolze A, Itan Y, Cobat A, Vincent QB, Antipenko A, Shang L, Boisson B, Casanova JL, Abel L. Whole-genome sequencing is more powerful than whole-exome sequencing for detecting exome variants. *Proc Natl Acad Sci U S A* 2015; 112: 5473-5478.
224. Emond MJ, Louie T, Emerson J, Zhao W, Mathias RA, Knowles MR, Wright FA, Rieder MJ, Tabor HK, Nickerson DA, Barnes KC, Gibson RL, Bamshad MJ. Exome sequencing of extreme phenotypes identifies DCTN4 as a modifier of chronic Pseudomonas aeruginosa infection in cystic fibrosis. *Nat Genet* 2012; 44: 886-889.
225. Zhang F, Lupski JR. Non-coding genetic variants in human disease. *Human molecular genetics* 2015; 24: R102-110.
226. Lin W-Y. Association Testing of Clustered Rare Causal Variants in Case-Control Studies. *PLoS ONE* 2014; 9: e94337.
227. Schuliga M, Westall G, Xia Y, Stewart AG. The plasminogen activation system: new targets in lung inflammation and remodeling. *Current Opinion in Pharmacology* 2013; 13: 386-393.
228. Can U, Guzelant A, Yerlikaya FH, Yosunkaya S. The role of serum soluble urokinase-type plasminogen activator receptor in stable chronic obstructive pulmonary disease. *Journal of investigative medicine : the official publication of the American Federation for Clinical Research* 2014; 62: 938-943.
229. Gumus A, Altintas N, Cinarka H, Kirbas A, Haziroglu M, Karatas M, Sahin U. Soluble urokinase-type plasminogen activator receptor is a novel biomarker predicting acute exacerbation in COPD. *International journal of chronic obstructive pulmonary disease* 2015; 10: 357-365.
230. Cho S, Kang J, Lyttle C, Harris K, Daley B, Grammer L, Avila P, Kumar R, Schleimer R. Association of elevated plasminogen activator inhibitor 1 levels with diminished lung function in patients with asthma. *Annals of Allergy, Asthma and Immunology* 2011; 106: 371-377.
231. Motiwala T, Ghoshal K, Das A, Majumder S, Weichenhan D, Wu YZ, Holman K, James SJ, Jacob ST, Plass C. Suppression of the protein tyrosine phosphatase receptor type O gene (PTPRO) by methylation in hepatocellular carcinomas. *Oncogene* 2003; 22: 6319-6331.
232. Motiwala T, Kutay H, Ghoshal K, Bai S, Seimiya H, Tsuruo T, Suster S, Morrison C, Jacob ST. Protein tyrosine phosphatase receptor-type O (PTPRO) exhibits characteristics of a

candidate tumor suppressor in human lung cancer. *Proc Natl Acad Sci U S A* 2004; 101: 13844-13849.
233. Breton CV, Byun HM, Wenten M, Pan F, Yang A, Gilliland FD. Prenatal tobacco smoke exposure affects global and gene-specific DNA methylation. *American journal of respiratory and critical care medicine* 2009; 180: 462-467.
234. Hughes T, Coit P, Adler A, Yilmaz V, Aksu K, Duzgun N, Keser G, Cefle A, Yazici A, Ergen A, Alpsoy E, Salvarani C, Casali B, Kotter I, Gutierrez-Achury J, Wijmenga C, Direskeneli H, Saruhan-Direskeneli G, Sawalha AH. Identification of multiple independent susceptibility loci in the HLA region in Behcet's disease. *Nat Genet* 2013; 45: 319-324.
235. Andreassen OA, Thompson WK, Schork AJ, Ripke S, Mattingsdal M, Kelsoe JR, Kendler KS, O'Donovan MC, Rujescu D, Werge T, Sklar P, Roddey JC, Chen C-H, McEvoy L, Desikan RS, Djurovic S, Dale AM, The Psychiatric Genomics C, Bipolar D, Schizophrenia Working G. Improved Detection of Common Variants Associated with Schizophrenia and Bipolar Disorder Using Pleiotropy-Informed Conditional False Discovery Rate. *PLoS Genet* 2013; 9: e1003455.
236. Zhang Y, De S, Garner JR, Smith K, Wang SA, Becker KG. Systematic analysis, comparison, and integration of disease based human genetic association data and mouse genetic phenotypic information. *BMC Medical Genomics* 2010; 3: 1-22.
237. Pletcher MT, McClurg P, Batalov S, Su AI, Barnes SW, Lagler E, Korstanje R, Wang X, Nusskern D, Bogue MA, Mural RJ, Paigen B, Wiltshire T. Use of a dense single nucleotide polymorphism map for in silico mapping in the mouse. *PLoS biology* 2004; 2: e393.
238. Kirby A, Kang HM, Wade CM, Cotsapas C, Kostem E, Han B, Furlotte N, Kang EY, Rivas M, Bogue MA, Frazer KA, Johnson FM, Beilharz EJ, Cox DR, Eskin E, Daly MJ. Fine mapping in 94 inbred mouse strains using a high-density haplotype resource. *Genetics* 2010; 185: 1081-1095.
239. Flint J, Eskin E. Genome-wide association studies in mice. *Nature reviews Genetics* 2012; 13: 807-817.
240. Kang HM, Zaitlen NA, Wade CM, Kirby A, Heckerman D, Daly MJ, Eskin E. Efficient control of population structure in model organism association mapping. *Genetics* 2008; 178: 1709-1723.
241. Zhou X, Stephens M. Genome-wide efficient mixed-model analysis for association studies. *Nat Genet* 2012; 44: 821-824.
242. Manenti G, Galvan A, Pettinicchio A, Trincucci G, Spada E, Zolin A, Milani S, Gonzalez-Neira A, Dragani TA. Mouse genome-wide association mapping needs linkage analysis to avoid false-positive Loci. *PLoS Genet* 2009; 5: e1000331.
243. Himes BE, Sheppard K, Berndt A, Leme AS, Myers RA, Gignoux CR, Levin AM, Gauderman WJ, Yang JJ, Mathias RA, Romieu I, Torgerson DG, Roth LA, Huntsman S, Eng C, Klanderman B, Ziniti J, Senter-Sylvia J, Szefler SJ, Lemanske RF, Jr., Zeiger RS, Strunk RC, Martinez FD, Boushey H, Chinchilli VM, Israel E, Mauger D, Koppelman GH, Postma DS, Nieuwenhuis MA, Vonk JM, Lima JJ, Irvin CG, Peters SP, Kubo M, Tamari M, Nakamura Y, Litonjua AA, Tantisira KG, Raby BA, Bleecker ER, Meyers DA, London SJ, Barnes KC, Gilliland FD, Williams LK, Burchard EG, Nicolae DL, Ober C, DeMeo DL, Silverman EK, Paigen B, Churchill G, Shapiro SD, Weiss ST. Integration

of mouse and human genome-wide association data identifies KCNIP4 as an asthma gene. *PLoS One* 2013; 8: e56179.
244. Mahadeva R, Shapiro SD. Chronic obstructive pulmonary disease • 3: Experimental animal models of pulmonary emphysema. *Thorax* 2002; 57: 908-914.
245. Guerassimov A, Hoshino Y, Takubo Y, Turcotte A, Yamamoto M, Ghezzo H, Triantafillopoulos A, Whittaker K, Hoidal JR, Cosio MG. The development of emphysema in cigarette smoke-exposed mice is strain dependent. *American journal of respiratory and critical care medicine* 2004; 170: 974-980.
246. Kelly NJ, Radder JE, Baust JJ, Burton CL, Lai YC, Potoka KC, Agostini BA, Wood JP, Bachman TN, Vanderpool RR, Dandachi N, Leme AS, Gregory AD, Morris A, Mora AL, Gladwin MT, Shapiro SD. Genetic association of guanylate cyclase subunit Gucy1b3 in a murine metabolic model of pulmonary hypertension. *Under Review* 2016.
247. Gao X, Becker LC, Becker DM, Starmer JD, Province MA. Avoiding the high Bonferroni penalty in genome-wide association studies. *Genetic epidemiology* 2010; 34: 100-105.
248. Gao X. Multiple testing corrections for imputed SNPs. *Genetic epidemiology* 2011; 35: 154-158.
249. Laurie CC, Nickerson DA, Anderson AD, Weir BS, Livingston RJ, Dean MD, Smith KL, Schadt EE, Nachman MW. Linkage Disequilibrium in Wild Mice. *PLoS Genet* 2007; 3: e144.
250. NCBI. Homology Maps. 2014 May 6, 2014 [cited 2014 September 1]. Available from: http://www.ncbi.nlm.nih.gov/projects/homology/maps/.
251. Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Gliedt TP, Boehnke M, Abecasis GR, Willer CJ. LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics (Oxford, England)* 2010; 26: 2336-2337.
252. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D. The human genome browser at UCSC. *Genome research* 2002; 12: 996-1006.
253. Matsuda S, Iriyama C, Yokozaki S, Ichigotani Y, Shirafuji N, Yamaki K, Hayakawa T, Hamaguchi M. Cloning and sequencing of a novel human gene that encodes a putative target protein of Nesh-SH3. *Journal of human genetics* 2001; 46: 483-486.
254. Terauchi K, Shimada J, Uekawa N, Yaoi T, Maruyama M, Fushiki S. Cancer-associated loss of TARSH gene expression in human primary lung cancer. *Journal of cancer research and clinical oncology* 2006; 132: 28-34.
255. Guimaraes GS, Latini FR, Camacho CP, Maciel RM, Dias-Neto E, Cerutti JM. Identification of candidates for tumor-specific alternative splicing in the thyroid. *Genes, chromosomes & cancer* 2006; 45: 540-553.
256. Latini FR, Hemerly JP, Oler G, Riggins GJ, Cerutti JM. Re-expression of ABI3-binding protein suppresses thyroid tumor growth by promoting senescence and inhibiting invasion. *Endocrine-related cancer* 2008; 15: 787-799.
257. Hodgkinson CP, Naidoo V, Patti KG, Gomez JA, Schmeckpeper J, Zhang Z, Davis B, Pratt RE, Mirotsou M, Dzau VJ. Abi3bp is a multifunctional autocrine/paracrine factor that regulates mesenchymal stem cell biology. *Stem cells (Dayton, Ohio)* 2013; 31: 1669-1682.
258. Hodgkinson CP, Gomez JA, Payne AJ, Zhang L, Wang X, Dal-Pra S, Pratt RE, Dzau VJ. Abi3bp regulates cardiac progenitor cell proliferation and differentiation. *Circulation research* 2014; 115: 1007-1016.

259. Wilk JB, Chen T-h, Gottlieb DJ, Walter RE, Nagle MW, Brandler BJ, Myers RH, Borecki IB, Silverman EK, Weiss ST, O'Connor GT. A Genome-Wide Association Study of Pulmonary Function Measures in the Framingham Heart Study. *PLoS Genet* 2009; 5: e1000429.
260. Kugler MC, Joyner AL, Loomis CA, Munger JS. Sonic hedgehog signaling in the lung. From development to disease. *Am J Respir Cell Mol Biol* 2015; 52: 1-13.
261. Flicek P, Amode MR, Barrell D, Beal K, Billis K, Brent S, Carvalho-Silva D, Clapham P, Coates G, Fitzgerald S, Gil L, Girón CG, Gordon L, Hourlier T, Hunt S, Johnson N, Juettemann T, Kähäri AK, Keenan S, Kulesha E, Martin FJ, Maurel T, McLaren WM, Murphy DN, Nag R, Overduin B, Pignatelli M, Pritchard B, Pritchard E, Riat HS, Ruffier M, Sheppard D, Taylor K, Thormann A, Trevanion SJ, Vullo A, Wilder SP, Wilson M, Zadissa A, Aken BL, Birney E, Cunningham F, Harrow J, Herrero J, Hubbard TJP, Kinsella R, Muffato M, Parker A, Spudich G, Yates A, Zerbino DR, Searle SMJ. Ensembl 2014. *Nucleic acids research* 2014; 42: D749-D755.
262. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protocols* 2009; 4: 1073-1081.
263. Knudsen L, Weibel ER, Gundersen HJG, Weinstein FV, Ochs M. Assessment of air space size characteristics by intercept (chord) measurement: an accurate and efficient stereological approach. *Journal of Applied Physiology* 2010; 108: 412-421.
264. Warburton D, El-Hashash A, Carraro G, Tiozzo C, Sala F, Rogers O, De Langhe S, Kemp PJ, Riccardi D, Torday J, Bellusci S, Shi W, Lubkin SR, Jesudason E. Lung Organogenesis. *Current Topics in Developmental Biology* 2010; 90: 73-158.
265. Lappalainen U, Whitsett JA, Wert SE, Tichelaar JW, Bry K. Interleukin-1beta causes pulmonary inflammation, emphysema, and airway remodeling in the adult murine lung. *Am J Respir Cell Mol Biol* 2005; 32: 311-318.
266. Peters VA, Joesting JJ, Freund GG. IL-1 receptor 2 (IL-1R2) and its role in immune regulation. *Brain, behavior, and immunity* 2013; 32: 1-8.
267. Bry K, Whitsett JA, Lappalainen U. IL-1β Disrupts Postnatal Lung Morphogenesis in the Mouse. *American journal of respiratory cell and molecular biology* 2007; 36: 32-42.
268. Churg A, Zhou S, Wang X, Wang R, Wright JL. The role of interleukin-1beta in murine cigarette smoke-induced emphysema and small airway remodeling. *Am J Respir Cell Mol Biol* 2009; 40: 482-490.
269. Oudijk EJ, Nijhuis EH, Zwank MD, van de Graaf EA, Mager HJ, Coffer PJ, Lammers JW, Koenderman L. Systemic inflammation in COPD visualised by gene profiling in peripheral blood neutrophils. *Thorax* 2005; 60: 538-544.
270. Kramer BW, Kallapur S, Newnham J, Jobe AH. Prenatal inflammation and lung development. *Seminars in fetal & neonatal medicine* 2009; 14: 2-7.
271. Xie ZK, Huang QP, Huang J, Xie ZF. Association between the IL1B, IL1RN polymorphisms and COPD risk: a meta-analysis. *Scientific reports* 2014; 4: 6202.
272. Gohlke H, Illig T, Bahnweg M, Klopp N, Andre E, Altmuller J, Herbon N, Werner M, Knapp M, Pescollderungg L, Boner A, Malerba G, Pignatti PF, Wjst M. Association of the interleukin-1 receptor antagonist gene with asthma. *American journal of respiratory and critical care medicine* 2004; 169: 1217-1223.

273. Joos L, McIntyre L, Ruan J, Connett JE, Anthonisen NR, Weir TD, Paré PD, Sandford AJ. Association of IL-1β and IL-1 receptor antagonist haplotypes with rate of decline in lung function in smokers. *Thorax* 2001; 56: 863-866.
274. Taillé C, Guénégou A, Almolki A, Piperaud M, Leynaert B, Vuillaumier S, Neukirch F, Boczkowski J, Aubier M, Benessiano J, Crestani B. ET(B)receptor polymorphism is associated with airway obstruction. *BMC Pulmonary Medicine* 2007; 7: 5-5.
275. Padela S, Yi M, Cabacungan J, Shek S, Belcastro R, Masood A, Jankov RP, Tanswell AK. A Critical Role for Fibroblast Growth Factor-7 during Early Alveolar Formation in the Neonatal Rat. *Pediatric research* 2008; 63: 232-238.
276. Cardoso WV, Itoh A, Nogawa H, Mason I, Brody JS. FGF-1 and FGF-7 induce distinct patterns of growth and differentiation in embryonic lung epithelium. *Developmental dynamics : an official publication of the American Association of Anatomists* 1997; 208: 398-405.
277. Brehm JM, Hagiwara K, Tesfaigzi Y, Bruse S, Mariani TJ, Bhattacharya S, Boutaoui N, Ziniti JP, Soto-Quiros ME, Avila L, Cho MH, Himes B, Litonjua AA, Jacobson F, Bakke P, Gulsvik A, Anderson WH, Lomas DA, Forno E, Datta S, Silverman EK, Celedon JC. Identification of FGF7 as a novel susceptibility locus for chronic obstructive pulmonary disease. *Thorax* 2011; 66: 1085-1090.
278. Bettens K, Brouwers N, Engelborghs S, Lambert J-C, Rogaeva E, Vandenberghe R, Le Bastard N, Pasquier F, Vermeulen S, Van Dongen J, Mattheijssens M, Peeters K, Mayeux R, St George-Hyslop P, Amouyel P, De Deyn PP, Sleegers K, Van Broeckhoven C. Both common variations and rare non-synonymous substitutions and small insertion/deletions in CLU are associated with increased Alzheimer risk. *Molecular Neurodegeneration* 2012; 7: 1-12.
279. Yu A, Zhao C, Fan Y, Jang W, Mungall AJ, Deloukas P, Olsen A, Doggett NA, Ghebranious N, Broman KW, Weber JL. Comparison of human genetic and sequence-based physical maps. *Nature* 2001; 409: 951-953.
280. Krzywinski MI, Schein JE, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA. Circos: An information aesthetic for comparative genomics. *Genome research* 2009.
281. Imboden M, Bouzigon E, Curjuric I, Ramasamy A, Kumar A, Hancock DB, Wilk JB, Vonk JM, Thun GA, Siroux V, Nadif R, Monier F, Gonzalez JR, Wjst M, Heinrich J, Loehr LR, Franceschini N, North KE, Altmuller J, Koppelman GH, Guerra S, Kronenberg F, Lathrop M, Moffatt MF, O'Connor GT, Strachan DP, Postma DS, London SJ, Schindler C, Kogevinas M, Kauffmann F, Jarvis DL, Demenais F, Probst-Hensch NM. Genome-wide association study of lung function decline in adults with and without asthma. *The Journal of allergy and clinical immunology* 2012; 129: 1218-1228.
282. Wilk JB, DeStefano AL, Joost O, Myers RH, Cupples LA, Slater K, Atwood LD, Heard-Costa NL, Herbert A, O'Connor GT, Gottlieb DJ. Linkage and association with pulmonary function measures on chromosome 6q27 in the Framingham Heart Study. *Human molecular genetics* 2003; 12: 2745-2751.
283. Ong BA, Li J, McDonough JM, Wei Z, Kim C, Chiavacci R, Mentch F, Caboot JB, Spergel J, Allen JL, Sleiman PM, Hakonarson H. Gene network analysis in a pediatric cohort identifies novel lung function genes. *PLoS One* 2013; 8: e72899.
284. Lee JH, Cho MH, Hersh CP, McDonald ML, Wells JM, Dransfield MT, Bowler RP, Lynch DA, Lomas DA, Crapo JD, Silverman EK. IREB2 and GALC are associated with

pulmonary artery enlargement in chronic obstructive pulmonary disease. *Am J Respir Cell Mol Biol* 2015; 52: 365-376.
285. Pillai SG, Ge D, Zhu G, Kong X, Shianna KV, Need AC. A genome-wide association study in chronic obstructive pulmonary disease (COPD): identification of two major susceptibility loci. *PLoS Genet* 2009; 5.
286. Wetterstrand K. DNA Sequencing Costs: Data from the NHGRI Genome Sequencing Program (GSP). [cited 2016 7/06/2016]. Available from: Available at: www.genome.gov/sequencingcosts.
287. Gudbjartsson DF, Helgason H, Gudjonsson SA, Zink F, Oddson A, Gylfason A, Besenbacher S, Magnusson G, Halldorsson BV, Hjartarson E, Sigurdsson GT, Stacey SN, Frigge ML, Holm H, Saemundsdottir J, Helgadottir HT, Johannsdottir H, Sigfusson G, Thorgeirsson G, Sverrisson JT, Gretarsdottir S, Walters GB, Rafnar T, Thjodleifsson B, Bjornsson ES, Olafsson S, Thorarinsdottir H, Steingrimsdottir T, Gudmundsdottir TS, Theodors A, Jonasson JG, Sigurdsson A, Bjornsdottir G, Jonsson JJ, Thorarensen O, Ludvigsson P, Gudbjartsson H, Eyjolfsson GI, Sigurdardottir O, Olafsson I, Arnar DO, Magnusson OT, Kong A, Masson G, Thorsteinsdottir U, Helgason A, Sulem P, Stefansson K. Large-scale whole-genome sequencing of the Icelandic population. *Nat Genet* 2015; 47: 435-444.
288. Dong H, Ma L, Gan J, Lin W, Chen C, Yao Z, Du L, Zheng L, Ke C, Huang X, Song H, Kumar R, Yeung SC, Zhang H. PTPRO represses ERBB2-driven breast oncogenesis by dephosphorylation and endosomal internalization of ERBB2. *Oncogene* 2016.
289. Zhang X, Tan Z, Wang Y, Tang J, Jiang R, Hou J, Zhuo H, Wang X, Ji J, Qin X, Sun B. PTPRO-associated hepatic stellate cell activation plays a critical role in liver fibrosis. *Cellular physiology and biochemistry : international journal of experimental cellular physiology, biochemistry, and pharmacology* 2015; 35: 885-898.
290. Nielsen CH, Larsen A, Nielsen AL. DNA methylation alterations in response to prenatal exposure of maternal cigarette smoking: A persistent epigenetic impact on health from maternal lifestyle? *Archives of toxicology* 2016; 90: 231-245.
291. Chen MJ, Wei SY, Yang WS, Wu TT, Li HY, Ho HN, Yang YS, Chen PL. Concurrent exome-targeted next-generation sequencing and single nucleotide polymorphism array to identify the causative genetic aberrations of isolated Mayer-Rokitansky-Kuster-Hauser syndrome. *Human reproduction (Oxford, England)* 2015; 30: 1732-1742.
292. Krishna SS, Majumdar I, Grishin NV. Structural classification of zinc fingers: SURVEY AND SUMMARY. *Nucleic acids research* 2003; 31: 532-550.
293. Berger CT, Recher M, Steiner U, Hauser TM. A patient's wish: anakinra in pregnancy. *Annals of the rheumatic diseases* 2009; 68: 1794-1795.
294. Chang Z, Spong CY, Jesus AA, Davis MA, Plass N, Stone DL, Chapelle D, Hoffmann P, Kastner DL, Barron K, Goldbach-Mansky RT, Stratton P. Anakinra use during pregnancy in patients with cryopyrin-associated periodic syndromes (CAPS). *Arthritis & rheumatology (Hoboken, NJ)* 2014; 66: 3227-3232.
295. Nold MF, Mangan NE, Rudloff I, Cho SX, Shariatian N, Samarasinghe TD, Skuza EM, Pedersen J, Veldman A, Berger PJ, Nold-Petry CA. Interleukin-1 receptor antagonist prevents murine bronchopulmonary dysplasia induced by perinatal inflammation and hyperoxia. *Proc Natl Acad Sci U S A* 2013; 110: 14384-14389.

296. Caramori G, Adcock IM, Di Stefano A, Chung KF. Cytokine inhibition in the treatment of COPD. *International journal of chronic obstructive pulmonary disease* 2014; 9: 397-412.
297. Hernandez ML, Mills K, Almond M, Todoric K, Aleman MM, Zhang H, Zhou H, Peden DB. Interleukin-1 receptor antagonist reduces endotoxin-induced airway inflammation in healthy volunteers. *The Journal of allergy and clinical immunology* 2015; 135: 379-385.
298. Ritchie MD, Holzinger ER, Li R, Pendergrass SA, Kim D. Methods of integrating data to uncover genotype-phenotype interactions. *Nature reviews Genetics* 2015; 16: 85-97.
299. Goh KI, Cusick ME, Valle D, Childs B, Vidal M, Barabasi AL. The human disease network. *Proc Natl Acad Sci USA* 2007; 104.
300. Wang K, Li M, Bucan M. Pathway-based approaches for analysis of genomewide association studies. *American journal of human genetics* 2007; 81: 1278-1283.
301. Meng C, Kuster B, Culhane AC, Gholami AM. A multivariate approach to the integration of multi-omics datasets. *BMC bioinformatics* 2014; 15: 1-13.
302. Megan EH, Michael HC, Merry-Lynn M, Peter JC, Ruth T-S, Diego JM, Barry JM, James DC, Russell PB, Edwin KS, Craig PH. A Multi-Omics Approach Identifies Inflammatory Patterns In The Asthma-COPD Overlap Syndrome. D91 GENETICS AND GENOMICS OF COPD: American Thoracic Society; 2016. p. A7473-A7473.
303. Knox SS. From 'omics' to complex disease: a systems biology approach to gene-environment interactions in cancer. *Cancer Cell International* 2010; 10: 11-11.