2017

Transcriptomic alterations underlying pathogenesis and carcinogenesis in COPD

BOSTON UNIVERSITY

SCHOOL OF MEDICINE

Dissertation

TRANSCRIPTOMIC ALTERATIONS UNDERLYING PATHOGENESIS AND CARCINOGENESIS IN COPD

by

JACOB J. KANTROWITZ

B.A., Vassar College, 2009

Submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

2017

Approved by

First Reader _____

Avrum Spira, M.D., M.Sc.
Professor of Medicine

Second Reader _____

Marc Lenburg, Ph.D.
Professor of Medicine

DEDICATION

I would like to dedicate this work to my patient spouse, Sarah, without whom this thesis would never have come together.

ACKNOWLEDGMENTS

Many people – family, friends, and mentors alike – have influenced me along the way towards the completion of this dissertation. I would be remiss if I did not start by acknowledging my first mentor, Dr. Daniel Wolf, who helped me understand how a physician-scientist successfully navigates the beginning of his career, simultaneously manages and mentors research personnel, and pursues research questions inspired by an unmet clinical need. Dan helped me apply and get into BUSM's MD/PhD program and I cannot thank him enough.

Over the last six years I have been at BU and Drs. John Schwartz and Steve Borkan along with many of the other dual degree students have supported me and helped me grow in my understanding of medicine and what it means to ask probing scientific questions.

I rotated in the Spira-Lenburg lab after having met Avi during my second year in his short-lived role as advisor to the MD/PhD program. I joined the lab immediately thereafter, and though I never intended to learn about bioinformatics, I am so glad I was given the opportunity to do so. I cannot thank Drs. Avi Spira and Marc Lenburg enough for allowing me to join them in their scientific endeavors. They have both been wonderful mentors who complemented each other perfectly, both providing insights the other could not. The have taught me to question everything and to think deeply about the research of others and myself. They have taught me to focus on the details but that at the end of the day, the packaging, the story you tell, is just as important and that the details

fade away. Without a good story, your work, excellent though it may be, will have no impact, and impacting patient care is the ultimate attainable goal. The members of my thesis committee, Drs. Matt Jones, George O'Connor, and Evan Johnson, have also provided insights throughout the last few years and helped guide my own research in invaluable ways.

Net, I would like to thank all of the many members of the Spira-Lenburg lab who have answered so many of my questions about statistics, programming, biology, and medicine. These folks have given me a reason to come to lab every day over the last four years and have supported me through those typical graduate student days where we all ask, "what are we doing here?!" I have enjoyed my time and been successful in my PhD as a direct result of all the wonderful people that are part of the lab. Thank you all so much.

Lastly, I would like to thank my parents for always believing that I could do whatever I set my mind to and my extended family for bringing love and stability to my life. Most importantly, I would like to thank my wife Sarah, without whom none of this work would have been possible. She has steadied me, coached me, and pushed me to be the best version of myself both personally and professionally. I have learned so much from her and I keep learning from her every day.

I love research and have been thrilled to go to work every day but this work is impossible to do alone. Thanks to all those that helped me along the way.

TRANSCRIPTOMIC ALTERATIONS UNDERLYING PATHOGENESIS AND CARCINOGENESIS IN COPD

JACOB J. KANTROWITZ

Boston University School of Medicine, 2017

Ph.D. degree requirements completed in 2017

Dual M.D./Ph.D. degrees expected in 2019

Major Professor: Avrum Spira, M.D., M.Sc., Professor of Medicine

Chronic obstructive pulmonary disease (COPD) is a major cause of morbidity and mortality worldwide and is a risk factor for lung cancer development. COPD encompasses both emphysema and chronic bronchitis, the pathogenesis of which are unclear. In this dissertation, I leveraged genome-wide gene-expression studies of emphysema and lung cancer to investigate pathogenesis and carcinogenesis in COPD.

Tobacco smoke is the primary cause of emphysema. The most severe form is also associated with alpha1-antitrypsin deficiency (AATD) resulting from a mutation. In this study, I leveraged multiple lung samples from patients with emphysema, with or without AATD. While genes involved in tissue repair decreased with emphysema severity, the unfolded protein response (UPR) was uniquely changed in AATD lungs. AATD may play multiple roles in emphysema and UPR activation suggests AAT replacement therapy may be insufficient to treat this form of emphysema.

Emphysema is a progressive disease, and the mean linear intercept (Lm) can serve as a surrogate of progression. I evaluated whether Lm increases in non-diseased lungs may represent similar processes to those occurring in emphysema, and could offer insight into early stages of disease or homeostasis. Genes involved in tissue repair increased with Lm in controls but decreased in disease. Tissue repair processes may be active in even the non-insulted lung, suggesting their activity is necessary for lung homeostasis and their deficiency may drive emphysema progression.

Finally, COPD patients are at increased lung cancer risk, and transcriptomic changes common to both diseases could explain this risk. In both COPD and lung cancer, I discovered that H3K27Me3 regulated genes are repressed, and that the methyltransferase responsible for H3K27me3, EZH2, is induced. H3K27Me3, an oncogenic histone modification, may drive carcinogenesis and pathogenesis in COPD.

Though usual and AATD emphysema share transcriptomic signatures associated with tissue repair, which may be active in the normal homeostatic lung, the UPR changes in AATD emphysema only; successful therapeutic strategies in emphysema will need to account for this difference. In COPD, H3K27Me3 may play a role in both pathogenesis and carcinogenesis, making it an attractive target for therapeutic interventions, but one that would need further augmentation in AATD.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AAT Alpha1-antitrypsin

AATD ..... Alpha1-antitrypsin deficiency

ANOVA ..... Analysis of variance

ATF4 ..... Activating transcription factor 4

ATF6 ..... Activating transcription factor 6

BU ..... Boston University

CDF ..... Chip definition file

COPD ..... Chronic obstructive pulmonary disease

CLE ..... Centrilobular emphysema

ECM ..... Extracellular matrix

ER ..... Endoplasmic reticulum

ERAD ..... ER associated degradation

EZH2 ..... Enhancer of zeste homolog 2

GHK ..... glycyl-L-histidyl-L-lysine

GO ..... Gene ontology

GOLD ..... Global initiative for chronic obstructive lung disease

GSEA ..... Gene set enrichment analysis

H3K27Me3 ..... Trimethylation of Lysine 27 on Histone H3

HMW ..... High molecular weight

IRE1a ..... inositol-requiring protein 1α

KEGG ..... Kyoto encyclopedia of genes and genomes

LCM ..... Laser capture microdissection

Lm ..... Mean linear intercept

LME ..... Linear mixed effects

ln ..... Natural log

LRT ..... Likelihood ratio test

MSIGDB ..... Molecular signatures database

NE ..... Neutrophil elastase

PERK ..... Protein kinase RNA-like endoplasmic reticulum kinase

PFTs ..... Pulmonary function tests

PLE ..... Panlobular emphysema

PRC2 ..... Polycomb repressor 2

ROS ..... Reactive oxygen species

SERPINA1 ..... Serpin family A member 1

SFTPD ..... Surfactant protein D

TGFb ..... Transforming growth factor beta

UBC ..... University of British Columbia

UPR ..... Unfolded Protein Response

XBP1 ..... X-Box-Binding Protein 1

Chapter One: Introduction

Public health burden of COPD and lung cancer

Chronic obstructive pulmonary disease (COPD) is a major cause of morbidity and mortality in the United States and throughout the world. It is debilitating for patients and costly to the health care system and the economy. While it was the fifth most common cause of death in 1980, the Centers for Disease Control reported that COPD was the third most common cause of death in the U.S in 2014 [1]. In 2012, the total cost of COPD was estimated to be $68 billion, including $54 billion in health care expenditures and $14 billion in lost productivity. Approximately 15 million people are diagnosed with COPD (6.4% of the U.S. population) and an estimated 12 million people have COPD but are undiagnosed [2]. These individuals are more likely to be unable to work, to have an activity limitation, and to have difficulty walking or climbing stairs [3]. Patients with COPD are also more likely to develop lung cancer [4-9]. Lung cancer is the malignancy responsible for the most cancer related deaths and has a five-year survival rate of only 18.7% [10].

Just over 50 years ago, in 1964, the Surgeon General released the first report on the deleterious health effects of smoking and tobacco use. Education efforts regarding the effects of tobacco use have decreased smoking rates in the United States to their lowest in recent history, but rates differ based on socioeconomic status. More than 30% of people below the poverty line regularly smoke, while that rate drops to about 20% for those above the line [11]. Decreasing domestic smoking rates have decreased overall incidence

of tobacco related illness, but smoking rates and the incidence of tobacco related illnesses are increasing worldwide. Fifty percent of young men and 10% of young women are smoking throughout the world. Approximately 100 million deaths were caused by tobacco in the 20th century and if current smoking rates continue, almost one billion deaths will be tobacco-attributable in this century, mostly in low- and middle-income countries [12]. Furthermore, the research funding rate for these tobacco related diseases is incommensurate with the significant health impact on patients, the health care related costs, and their overall economic and societal impact [13].

Chronic obstructive pulmonary disease

The main risk factor for COPD is smoking, but only 25% of smokers will develop COPD, suggesting the accounting for interaction between exposure and genetics will be critical for understanding disease pathogenesis [14]. Most patients with COPD are current or former smokers and are not diagnosed until at least age 40, indicating cumulative exposure increases the risk of disease development. The common symptoms of COPD include chronic cough, increased sputum production, anorexia, weight loss, fatigue, and dyspnea. The COPD phenotype is defined by decreased lung function as measured by pulmonary function tests (PFTs), which quantify the degree of lung function and are used for diagnosing and monitoring COPD over time. COPD is staged according to the severity of lung function decline based on advice laid out by the Global Initiative for Chronic Obstructive Lung Disease (GOLD) [15]. Symptom severity varies greatly between patients due to the nature of the underlying lung damage, which is caused by a

combination of emphysematous destruction and chronic bronchitis. Both of these processes lead to lung obstruction, but do so through different mechanisms. Emphysema manifests as the destruction of the alveolar sacs and septal walls, as well as a decrease in the elasticity and structural integrity of the lungs. The surface area for gas exchange is limited by alveolar destruction and the lung collapses more readily with decreasing elasticity, leading to air trapping. Chronic bronchitis, on the other hand, manifests as increased sputum production and inflammatory infiltration of the airways, which lead to a decrease in airway size. Once these pathologies are initiated, they inexorably and irreversibly progress. The treatments that exist for COPD are not disease modifying, but rather are symptom managing. Patients are commonly given bronchodilators and inhaled corticosteroids to relax the muscles surrounding the airway and decrease bronchial inflammation, respectively. These treatments help with day-to-day management of COPD but offer no long-term benefit [16]. There are currently no cures for COPD, which explains why it was the third most common cause of death in the U.S. in 2014 [1].

Emphysema and alpha1-antitrypsin deficiency

Emphysema is one component of COPD and normally develops between 45 and 60 years of age. The exact molecular mechanisms underlying emphysema pathogenesis are not well understood, but a role for an imbalance between the proteinases and anti-proteinases of the lung is well accepted. Though smoking remains the primary risk factor for the development of emphysema, inherited mutations in the SERPINA1 gene, which codes for alpha1-antitrypsin (AAT), leads to AAT deficiency (AATD) and is associated with a

severe and early onset form of emphysema and liver disease. Functional alpha1-antitrypsin normally inhibits neutrophil elastase (NE), a potent serine proteinase released by activated neutrophils that can degrade the extracellular matrix of the lung. Inherited mutations in the SERPINA1 gene can lead to AAT misfolding, inactivity, and deficiency. The majority of functional AAT is normally produced and secreted by the liver into the circulation, through which it travels to the lung to protect it from excessive degradation by NE. The prevailing theory of AATD-associated emphysema pathogenesis is that the lack of functional AAT allows neutrophil elastase released from tobacco-activated and recruited neutrophils to excessively degrade the lung's extracellular matrix, from which recovery becomes impossible, and emphysema ensues. Liver disease, including hepatocellular carcinoma and cirrhosis is also common in AATD, and likely develops from the accumulation of misfolded AAT in the hepatocytes. This inherited disease is estimated to affect 1 in 3000 to 5000 people in the United States [17, 18] and up to 5% of COPD patients [19, 20], which means it is a driver of significant morbidity and mortality.

Several cross-sectional studies have identified pathways changed between the lungs of patients with and without COPD [21-24], often with little overlap. A few studies have examined specifically the emphysema portion of COPD [25-28]. One of these studies has examined the potential transcriptomic effects of AATD in emphysema and found that "usual" and AATD-associated emphysema mostly shared molecular profiles [25]. The most granular molecular study to date examined gene expression signatures of emphysema severity using repeated measures from within the same lungs [28], which

facilitated the discovery of glycyl-L-histidyl-L-lysine (GHK) as a potential treatment for emphysema. No study to date has identified transcriptomic level information to help explain the more severe emphysema phenotype observed in AATD patients.

Lung cancer

The link between smoking and lung cancer is well documented, with at least eight out of ten lung cancer deaths linked to smoking [29]. Lung cancer risk is increased in former smokers up to 30 years after smoking cessation [30]. Symptoms are similar to COPD and include chronic cough, weight loss, dyspnea, and fatigue, and can include hemoptysis (i.e. coughing up blood). Like COPD, lung cancer, once initiated, is difficult to curtail. Lung cancer is responsible for the most cancer deaths annually even though it is not the most prevalent, and the five year survival has improved only from 12% to 19% since the 1970s [10]. Usually, symptoms of lung cancer do not occur until the disease has reached an advanced stage, which make it harder to treat and leads to high mortality rates. This is complicated by the fact that COPD increases lung cancer risk and that the symptoms of COPD and lung cancer are similar. While patients diagnosed with stage I cancer have a 50% five-year survival rate, patients with stage IV disease have a 1% five-year survival rate [29].

COPD increases lung cancer risk

In 1986 one study found that while controlling for age, sex, occupation, and smoking history, COPD patients had an increased rate of lung cancer development [4]. Because of

the strong shared etiological link to tobacco, matching for smoking history was critically important for this study, and suggested for the first time that COPD increased lung cancer risk, regardless of smoking history. Several groups have affirmed the association between COPD and increased lung cancer risk [6-8, 31]. One meta-analysis of 39 studies calculated that the combined relative risk of lung cancer in COPD was 1.83 [32]. In fact, tumors in the lungs of patients with COPD were more likely to originate in regions of severe emphysema [33]. Furthermore, AATD, also carries an increased lung cancer risk [34-36], suggesting that regardless of the underlying etiology, emphysema increases lung cancer risk. Authors of one of the largest studies to date suggested that smoking habit and the timing of COPD diagnosis, rather than the presence of the disease, could account for the observed increased incidence of lung cancer among COPD patients. When correcting for smoking and including only patients with a 10 year or greater history of COPD, however, this study still found that the odds of lung cancer development was greater among the COPD population (odds ratio: 2.18) [5]. The mechanisms that increase lung cancer incidence in COPD are unclear.

Specific Aims

Aim 1: Compare and contrast gene expression effects in emphysematous lungs with or without alpha1-antitrypsin deficiency

Emphysema is a disease process characterized by unchecked, progressive, and irreversible destruction of functional lung tissue. Smoking is the primary risk factor for

emphysema, but alpha1-antitrypsin deficiency is the most important inherited disorder associated with emphysema. Approximately 5% of COPD patients are diagnosed with AATD, but this is thought to underrepresent the actual prevalence of disease. Patients with AATD present with severe emphysema at an early age, and though this AATD-associated emphysema is thought to result from alpha1-antitrypsin deficiency, which allows the potent proteinase, neutrophil elastase, to destroy the lung unchecked, only one group has examined the genome wide implications of AATD in the lung. The goal of this study is to identify transcriptomic similarities and differences between emphysematous lungs with or without AATD, by comparing "usual" and AATD-associated emphysema, by identifying gene expression profiles associated with emphysema severity across both of these primary forms of emphysema, and by identifying profiles associated with disease severity differently between the two groups.

Aim 2: Compare and contrast gene expression signatures of healthy alveolar spacing and emphysematous destruction

Though COPD is the most common form of smoking associated disease, only approximately 25% of smokers will ever develop the disease, suggesting a strong underlying genetic component to disease, that might manifest in differences measurable at the gene expression level. In fact, the estimated heritability of lung function and COPD is 37%, according to one study [37]. Many groups have already characterized the cross-sectional gene expression differences between healthy smokers and smokers with COPD, but no groups have explored the transcriptome as it relates to healthy alveolar

spacing. The goal of this study is to identify the genes and processes that change in the healthy lung with alveolar spacing and to compare these changes to those occurring with emphysema severity in diseased lungs. These changes could provide insight into early stages of emphysema or the lung homeostasis program and its relationship with disease.

Aim 3: Connect the COPD lung and lung cancer transcriptomes

COPD is the most prevalent smoking associated disease. Lung cancer, also a common smoking associated disease, is responsible for the most malignancy related deaths. Unfortunately, the presence of COPD has been demonstrated time and again to increase the incidence of lung cancer, independent of smoking history, suggesting COPD and lung cancer share a common disease process or alternatively, that the pathogenic processes of COPD drive carcinogenesis. Though several groups have reviewed potential mechanistic links between these two disease states, no study to date has explored the possibility that a transcriptomic connection may link these diseases at the molecular level. The goal of this study is to examine the transcriptomic effects of COPD and lung cancer and to identify any overlapping processes or molecular events that help explain the increased lung cancer risk observed in the COPD population.

Chapter Two: Data acquisition, processing, and profiling

Purpose

In chapters three and four I investigate gene expression changes associated with emphysema severity and alveolar spacing (i.e. the mean linear intercept, Lm) across 15 lungs collected as part of one study. These lungs were collected, processed, and profiled by our collaborators at the University of British Columbia (UBC) in the laboratory of James Hogg, M.D., Ph.D. Our laboratory received tissue samples stored in qiazol from UBC, and Gang Liu, Ph.D. and his team, isolated RNA from these samples so that the BU microarray core could hybridize them onto microarrays, scan the arrays, and generate CEL files, which I received. I then was able to normalize the CEL files and generate quality control metrics. The purpose of this brief chapter is to provide details regarding the methodologies utilized by our collaborators and the members of our lab that were necessary so that I could perform the various gene expression analyses, which are detailed in chapters three and four.

Materials and Methods

Sample acquisition and processing

Ten emphysematous lungs were collected by our collaborators at UBC after being explanted from patients with GOLD stage IV COPD, with or without alpha1-antitrypsin deficiency. Five control lungs were collected from non-emphysematous donors after they were released for research because no suitable match could be found for the lungs. All 15 lungs were inflated, frozen, and subsequently sliced at 2 cm intervals in the transverse

plane into eight slices. Two adjacent cores were taken from these slices. One core was processed on the microCT by Dragos Vasilescu, Ph.D. (UBC) to measure the mean linear intercept (Lm). The second core was processed by Daisuke Kinose, Ph.D. (UBC), by laser capture microdissection (LCM) to separate airway walls from parenchymal tissue.

Tissue laser capture microdissection (LCM)

Daisuke Kinose, Ph.D., of UBC, used hematoxylin and eosin stained sections to find airways. His definition of airway was that of a structure with airway epithelial cells. Most of the airway sections contained multiple airway structures, likely at different airway generations. He tried not to include airways with cartilage in an attempt to isolate airway generations from around the terminal bronchioles and respiratory bronchioles. He collected whole airway walls using LCM, so airway samples consisted of airway epithelial cells, smooth muscle cells, fibroblasts, inflammatory cells if they were there, and endothelial cells. Parenchymal tissue was all that remained after airway wall removal, and included alveolar tissue (type 1 and 2 epithelial cells), vessels (endothelial cells, smooth muscle cells), inflammatory cells, and fibroblasts.

Measurement of mean linear intercept (Lm)

Emphysema severity, or alveolar spacing in the case of non-emphysematous lungs, was measured by Dragos Vasilescu Ph.D. He calculated the mean linear intercept (Lm), as previously described [28].

Microarray sample processing and quality control

Gang Liu, Ph.D. and members of our lab isolated high molecular weight (HMW; mRNA-containing fraction) RNA from LCM tissue cores using the miRNeasy Mini Kit (Qiagen). The RNA integrity was assessed using an Agilent 2100 Bioanalyzer and RNA purity was assessed using a NanoDrop spectrophotometer. One ug of RNA was processed and hybridized onto the Human Exon 1.0 ST array (Affymetrix Inc.) by the BU microarray core, according to the manufacturer's protocol as previously described. Members of the core then scanned the arrays to generate CEL files.

I used the robust multi array normalization algorithm [38] and BrainArray V17 CDF to generate log2 normalized gene expression data that could be used within the R environment for statistical analysis. I used the RLE and NUSE algorithms [39] to test array quality and identify any abnormal or low quality arrays. Normalization was completed and quality control metrics were generated using R v 2.1.15.

Chapter Three: Gene expression signatures of lung tissue destruction in "usual" emphysema and alpha1-antitrypsin deficiency associated emphysema

Abstract

I identified a number of genes whose expression changed with increasing emphysema severity regardless of the alpha1-antitrypsin deficiency (AATD) status of the originating lungs. The majority of identified genes decreased with disease severity and were enriched with genes involved in various aspects of lung repair. Among the genes that increased with severity was Surfactant protein D (SFTPD), which has been demonstrated to be a biomarker of COPD and lung damage [40-43], as well as a tag for apoptotic cells, marking them for immune cell mediated phagocytosis [44]. Interestingly, I identified a set of genes that changed with increasing emphysema severity uniquely among the lungs with AATD. These genes were associated with the endoplasmic reticulum (ER) membrane, ER associated degradation (ERAD), and the unfolded protein response (UPR). Genes induced along all three arms of the UPR were enriched with genes that increased with emphysema severity specifically in the AATD lungs. I have provided evidence that AATD is associated with changes to the UPR that may drive AATD-emphysema.

Background

Chronic obstructive pulmonary disease (COPD) is a serious public health burden characterized by a combination of chronic bronchitis and emphysema. At least 25% of cigarette smokers go on to develop COPD [14], but smoking is only one of a few risk

factors for the disease. In addition to smoking, inherited mutations in the SERPINA1 gene have also been linked to the development of COPD, specifically through the development of alpha1-antitrypsin deficiency (AATD). This inherited disease has an estimated prevalence of 1 case per 3000 to 5000 people in the United States [17, 18] and up to 5% of COPD patients are thought to have AATD [19, 20]. Alpha1-antitrypsin (AAT) – the protein coded for by the gene SERPINA1 – normally inhibits neutrophil elastase (NE) and its deficiency leads to early onset and severe emphysema and liver disease. There are several SERPINA1 mutations associated with AATD, but the Z mutation, in which the glutamic acid at position 342 is replaced with a lysine (the resulting protein is referred to as ZAAT), is the most clinically relevant as it leads to severe disease. As much as 95% of the AATD population has at least one Z allele [45, 46]. The Z mutation leads not only to a loss of the functional protein, but also leads to a gain-of-toxic function. The ZAAT protein cannot properly fold, polymerizes, and accumulates in globules in the endoplasmic reticulum (ER) in some types of AAT secreting cells. In the liver, the accumulation of abnormally folded ZAAT causes cirrhosis and hepatocellular carcinoma [47, 48]. However, in the lung, exactly which cells secrete ZAAT and the potential implications of its accumulation remain less well understood.

The traditional paradigm of disease pathogenesis in AATD implicates an imbalance between proteinases (released from neutrophils) and anti-proteinases, resulting in emphysema. ZAAT polymerization and retention in the liver decreases circulating AAT,

and insufficient functional AAT allows neutrophil elastase to destroy healthy, functional tissue in the lung. However, recent studies have shown that AAT is produced in the lung at least by bronchial epithelial cells [49-51], if not also by alveolar cells [52], suggesting the possibility that AATD-related emphysema does not result from liver dysfunction alone; the lung may play a role in its own destruction. Moreover, the ZAAT protein acts as a powerful neutrophil chemoattractant, suggesting its role in emphysema pathogenesis is multifaceted [49], as neutrophils are increased in the lungs of AATD patients [53]. Lastly, the ZAAT protein accumulates within the ER of hepatocytes and has been demonstrated to cause protein immobility [54], sensitizing these cells to secondary sources of ER stress, which leads to increased rates of apoptosis. What relevance this finding in particular has in the lungs has not yet been evaluated.

Several groups have profiled the transcriptional changes associated with COPD or emphysema cross-sectionally, relying mostly on pulmonary function tests (PFTs) to dichotomize their patient samples [21, 25, 26, 55-58]. Only one study to date has examined the transcriptomic effects in the lung associated with alpha1-antitrypsin deficiency related emphysema. Golpon and colleagues [25] showed that gene expression profiles are mostly shared between “usual” emphysema and AATD-related emphysema. Our group has previously improved on the methodology of these cross-sectional studies in order to gain insights into the mechanisms of emphysema progression within a patient. In six lungs from patients with COPD (1 with AATD) and two donor lungs, our lab captured paired cores from regularly spaced regions from the apex to the base of

explanted lungs, profiled gene expression from one core, and measured the mean linear intercept with microCT in the other. In this way, Campbell and colleagues were able to identify a decrease in activity of the TGFb signaling pathway with increasing emphysema severity and identify GHK as a potential disease modifying agent in emphysema that acts on the TGFb pathway [28].

In the present study our collaborators have further improved this paradigm by laser capture microdissecting (LCM) whole airway walls away from the functional parenchymal tissue, which has enabled us to profile gene expression in both tissue compartments separately. Furthermore, we have included 5 SERPINA1-MM normal lungs with centrilobular emphysema (CLE) and 5 SERPINA1-ZZ lungs with panlobular emphysema (PLE), enabling a more detailed examination of gene expression changes associated with the AATD-associated emphysema subtype.

I identified a number of genes whose expression changed with increasing emphysema severity regardless of the SERPINA1 mutation status of the originating lungs. The majority of identified genes decreased with emphysema severity and were enriched with genes involved in lung repair. Though I did not identify any genes differentially expressed between emphysema subtypes, I did identify a set of genes associated with the unfolded protein response (UPR) that increased with emphysema severity in AATD lungs only. Overall, “usual” and AATD emphysema share grossly similar gene expression

profiles, but importantly differ in that AATD emphysema is additionally associated with the UPR.

Materials and Methods

Microarray gene expression data analysis strategy

Statistical analysis was completed in a manner similar to Campbell's publication [28]. In the expression data from the parenchymal tissue I utilized linear mixed effects (LME) modeling in conjunction with the likelihood ratio test (LRT) to identify differentially expressed genes associated with disease status (i.e. emphysematous tissue v control tissue) or emphysema severity (i.e. ln(Lm)). I undertook three separate analyses to identify (1) genes differentially expressed between emphysema subtypes, (2) genes changing with increasing emphysema severity across both subtypes, and (3) genes changing with emphysema severity differently between the emphysema subtypes. In each analysis, I attempted to identify a significant signature, compared the signature to previous studies, and performed biological enrichment analysis. In the analyses examining gene expression effects associated with Lm, I also tested whether the signatures changed concordantly with Lm after standardizing Lm within each patient lung to a mean of zero and standard deviation of one (Z-Score). This helps guard against outlier patients from disproportionately driving the makeup of the signature. For analysis 2 I further tested the consistency of the signature by similarly standardizing Lm within each group, which guards against the Lm differences between the groups from driving the signature makeup.

All statistical analysis was done in R v3.1.1. The R package *nlme* v3.1-128 was used for all LME modeling and nested LME models were compared by analysis of variance

(ANOVA) using LRT. Comparisons to external studies were done by gene set enrichment analysis, GSEA v2.2.3 [59]. Signatures were hierarchically clustered prior to biological enrichment analysis, which was done in Enrichr [60] with the search space limited to KEGG, Biocarta, Reactome, and gene ontology (GO) databases.

Analysis 1. Gene expression differences between "usual" and AATD emphysema

To identify genes differentially expressed between emphysema subtypes I generated two LME models and compared them by ANOVA:

(1) Gene ~ slice + Lm + random(patient) + error

(2) Gene ~ slice + Lm + group + random(patient) + error

In these and the following equations, slice represents the height from which a sample was taken from within a lung, Lm represents emphysema severity – and we have used the natural log of Lm – and group represents the emphysema subtype of the sample. Note throughout this dissertation I refer to Lm to mean the natural log of Lm, for ease of reading.

Analysis 2. Gene expression changes associated with emphysema severity

I next sought to identify genes that change with emphysema severity consistently across both "usual" and AATD emphysema and used the two following equations, which I compared by ANOVA:

(3) Gene ~ slice + group + random(patient) + error

(4) Gene ~ slice + group + Lm + random(patient) + error

Analysis 3. Gene expression changes associated with emphysema severity, dependent on subtype

Lastly, I wanted to identify any gene expression profiles that change with emphysema severity in an emphysema subtype-dependent manner and used the following equations, again compared by ANOVA:

(5) Gene ~ slice + group + Lm + random(patient) + error

(6) Gene ~ slice + group + Lm + group:Lm + random(patient) + error

In this last set of equations, group:Lm represents the interaction between Lm and group.

Results

Study Population

Lm was measured using micro-CT scans taken from regularly spaced regions from within the lungs of ten patients with emphysema, five of whom also had AATD. As expected, the patients with AATD emphysema had a higher mean Lm than the patients with "usual" emphysema (Figure 1). Table 1 shows demographic, clinical, and genotype information for the ten lungs. The five lungs with AATD had panlobular emphysema (PLE) and were ZZ homozygotes for SERPINA1, while the five lungs without AATD had centrilobular emphysema (CLE) and were MM homozygotes (i.e. normal).

Table 1. Demographic information for the 10 emphysema lungs.

Patient	Group	Sex	Age	Pack Years	Smoker	LM Summary	LM Range
6970	AATD	M	55	6	former	1337 (604)	494-2322
7026	AATD	M	55	9	former	1490 (696)	780-2802
7031	AATD	M	48	25	former	1201 (167)	955-1390
7034	AATD	M	51	25	former	1665 (466)	901-2364
7263	AATD	F	39	18	former	1209 (132)	1098-1498
6996	CLE	F	77	45	former	512 (176)	340-903
7305	CLE	F	58	30	former	823 (99)	715-967
7307	CLE	M	55	80	former	959 (718)	475-2635
7336	CLE	F	59	40	former	572 (232)	371-1102
7337	CLE	F	53	24	former	1397 (536)	839-2363

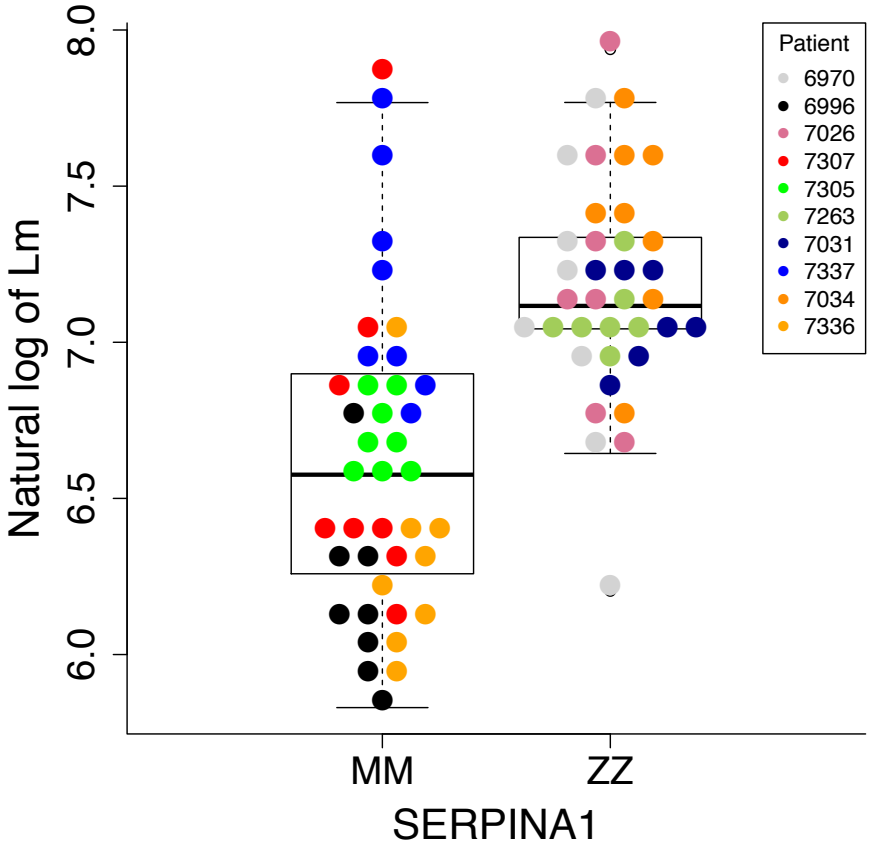

Figure 1. Distribution of natural log of Lm measurements by SERPINA1 mutation status. Each patient is represented by a different color and has up to 8 samples. The samples from the ZZ lungs have significantly higher values than the MM lungs ($p < 0.0001$).

No genes differentially expressed between emphysema subtypes

Due to the unique presentation normally associated with AATD emphysema, I first sought to identify any gene expression differences between the two types of emphysema. I identified only 668 genes differentially expressed between emphysema subtypes (uncorrected $p < 0.05$), yet I would expect 985 genes to be differentially expressed at that p-value threshold by chance (data not shown).

Pathways associated with regional emphysema severity independent of genotype

The lack of gene expression differences between the two emphysema subtypes lead me to the hypothesis that the two subtypes are grossly similar, so I set out to expand on these similarities. Using linear mixed effects modeling, I identified 571 genes associated with Lm consistently across "usual" and AATD emphysema ($q < 0.05$) (Figure 2 panel A). Hierarchical clustering identified two clusters with 126 and 445 genes. The 126 genes were positively associated with Lm and the 445 genes were negatively associated with Lm. Using Enrichr [60], I found the positively associated genes were enriched with genes involved in eukaryotic translation elongation ($q < 0.001$), and also included Surfactant protein D (Figure 2 panel B), a known biomarker of lung damage [42], emphysema [41], and COPD [40]. The negatively associated genes were enriched with genes involved in axon guidance, focal adhesion, ECM and lamellipodium organization, angiogenesis, and the regulation of epithelial cell migration and of the actin cytoskeleton ($q < 0.001$). SRGAP2 is a gene involved in axon guidance and is provided as an exemplar gene that decreased with Lm (Figure 2 panel C).

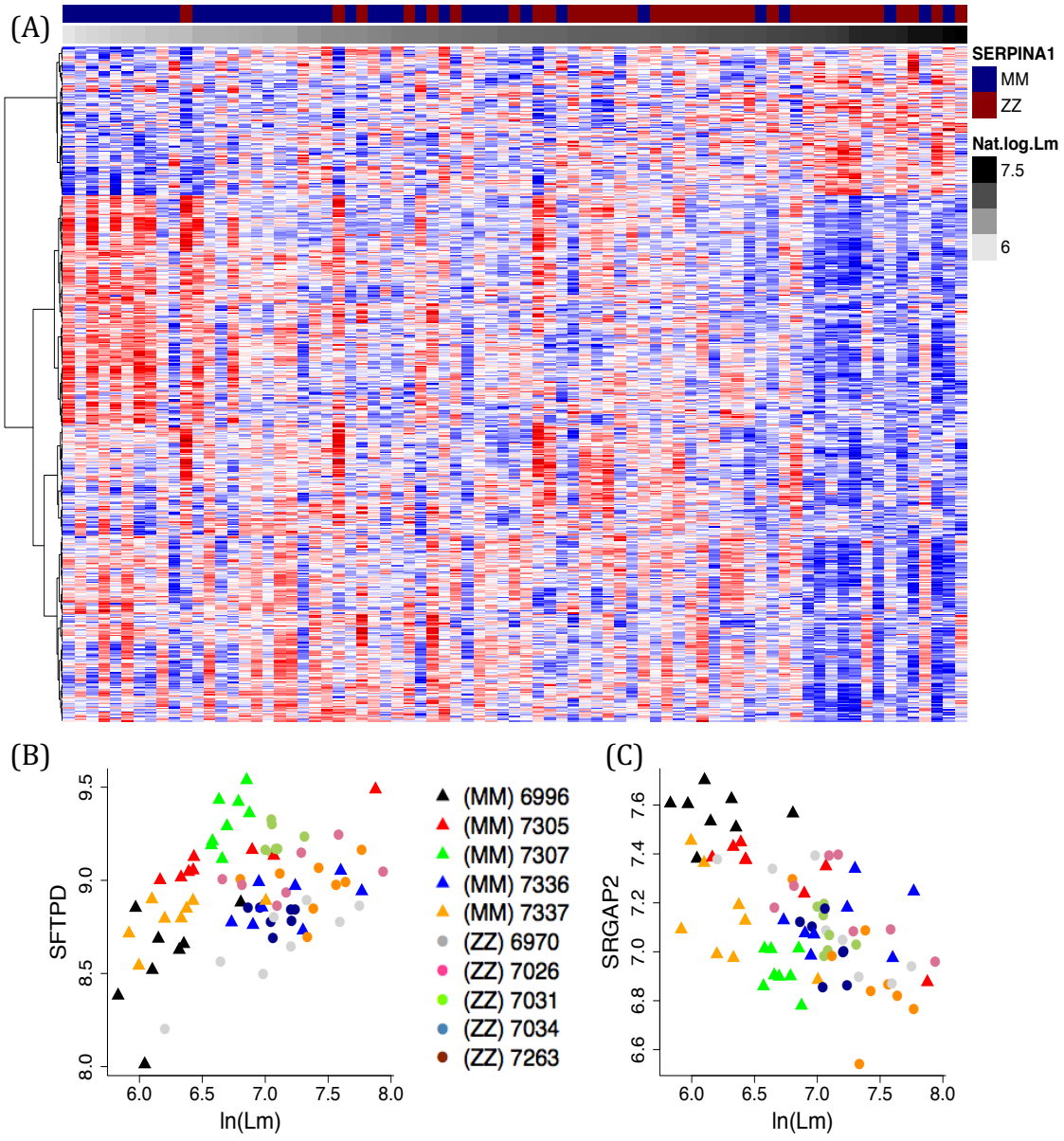

Figure 2. Gene expression signature of regional emphysema severity independent of SERPINA1 status. (A) Supervised heat map of 571 genes (q < 0.05) associated with Lm. Samples (columns) are organized from mild to severe emphysema from left to right. Rows represent genes and are hierarchically clustered, resulting in two clusters of genes increasing or decreasing with disease severity. (B) Expression of SFTPD and (C) SRGAP2 are plotted against the natural log of Lm. The color of each dot represents the patient from which the sample was derived and the shape represents the SERPINA1 status of the patient (triangle = MM, circle = ZZ).

Table 2. Enrichment results for genes decreasing with Lm consistently across "usual" and AATD emphysema.

Category	FDR q value
Axon guidance (Homo sapiens)	1.47E-13
Developmental Biology (Homo sapiens)	6.21E-11
Focal adhesion (Homo sapiens)	3.21E-08
Angiogenesis	1.23E-07
Signal transduction (Homo sapiens)	1.31E-07
Regulation of cell morphogenesis	1.44E-07
Lamellipodium assembly	9.77E-07
GTPase regulator activity	9.88E-07
Extracellular matrix organization	1.18E-06
Regulation of epithelial cell migration	7.61E-06
Regulation of actin cytoskeleton (Homo sapiens)	2.35E-05

To test whether samples with outlier Lm measurements disproportionately affected the makeup of this gene set, I standardized (Z-score normalized) the Lm measurements within each lung and repeated the same statistical analysis. I ranked all genes by the t-statistic of the standardized Lm measurement and tested whether the 126-gene set and 445-gene set were concordantly enriched, with GSEA. The 126 and 445 gene-sets were positively and negatively enriched, respectively (GSEA $q < 0.001$). I also wanted to demonstrate that the emphysema-severity associated genes were not identified due to group differences in Lm between MM and ZZ (AATD) lungs. To test this possibility, I standardized the Lm measurements within each group, ranked the genes by the t-statistic, and used GSEA to test for concordant enrichment. The 126 and 445 gene-sets were again concordantly enriched (GSEA $q < 0.001$) (Figure 3).

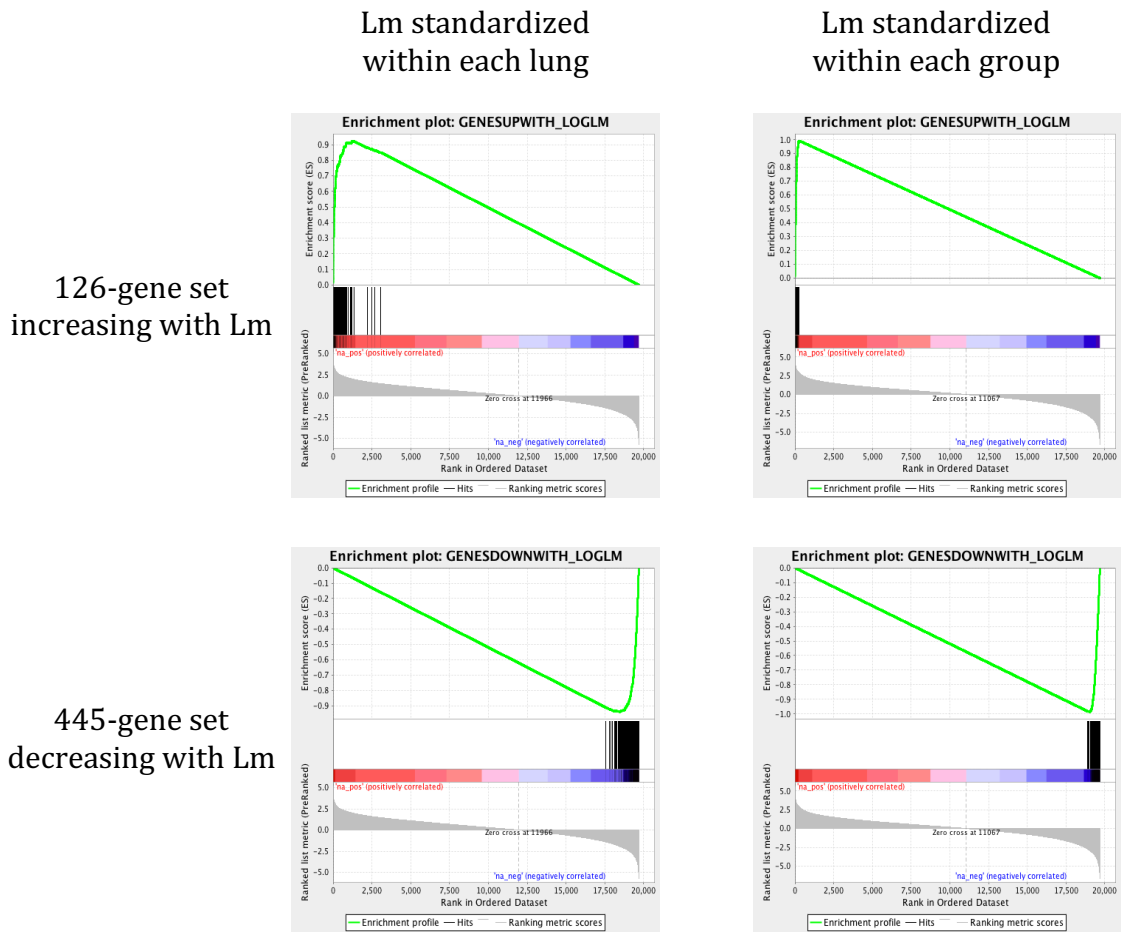

Figure 3. Genes changing with Lm across "usual" and AATD emphysema are enriched after standardizing Lm measurements. In the left column, genes are ranked by the t-statistic of the Lm measurement standardized within each lung. In the right column, genes are ranked by the t-statistic of the Lm measurement standardized within each group. In the first row, the set of 126 genes that increase with Lm is positively enriched in both ranked lists ($q < 0.001$). In the second row, the set of 445 genes that decrease with Lm is negatively enriched in both ranked lists ($q < 0.001$). The red to blue color bar represents the ranked list of all genes, with red indicating a positive t-statistic and blue a negative t-statistic. Each black tick represents the location of one of the genes from the tested gene set.

Connection to previous gene expression signature of emphysema severity

Campbell et al. previously published a 127-gene signature of emphysema severity in a cohort of eight lungs, six of which were from COPD patients (one with AATD), and two of which were from non-COPD donors [28]. Given the similarities between the two emphysema subtypes thus far presented, I wanted to test whether this previous signature

was consistent in these ten new emphysema lungs and shared across both MM and ZZ lungs. I ranked genes by the t-statistic for Lm from equation (4), above, and tested the enrichment of the 127-gene signature, using GSEA. The signature was concordantly enriched in this new set of lungs ($q < 0.001$) (Figure 4).

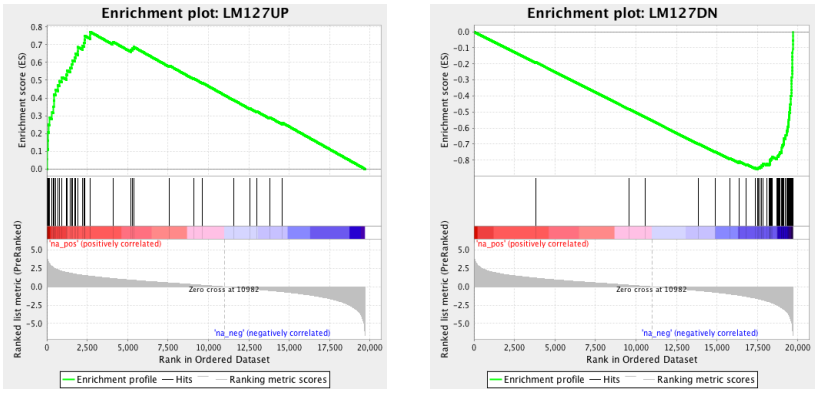

Figure 4. The previously published 127- gene signature of emphysema severity is concordantly enriched in this new set of ten emphysematous lungs. The left panel shows the set of genes increasing with Lm from the 127-gene signature tested against the new lungs, and the right panel shows the set of genes decreasing with Lm. Both sets of genes were concordantly enriched ($q < 0.001$).

Pathways associated with regional emphysema severity dependent on genotype

Thus far, I have characterized genomic similarities between the two emphysema subtypes, but the distinct clinical picture associated with AATD emphysema, that of a younger smoker with severe disease, begs the questions of whether there are gene expression alterations specific to AATD that could explain these differences. Using LME, I identified 616 genes associated with emphysema severity and dependent on genotype (i.e. the interaction effect) among MM and ZZ lungs ($p < 0.01$; 197 expected by chance). Hierarchical clustering identified three clusters of (1) 342, (2), 52, and (3) 222 genes (Figure 5, panel A). None of the clusters was strongly associated with Lm in the MM lungs. In the ZZ lungs, however, cluster 1 decreased with Lm, while clusters 2 and 3

increased with Lm (Figure 5, panel B). Cluster 1 was enriched with genes involved in ATP binding, regulation of leukocyte degranulation, and regulation of GTPase activity ($q < 0.05$). Cluster 2 was enriched with genes involved in tissue homeostasis and multi-drug resistance, but each of these results depended on one gene each ($q < 0.05$). Cluster 3 was enriched with genes involved in the unfolded protein response (UPR) and protein export and processing, processes known to be disrupted in AATD ($q < 0.05$, Table 3).

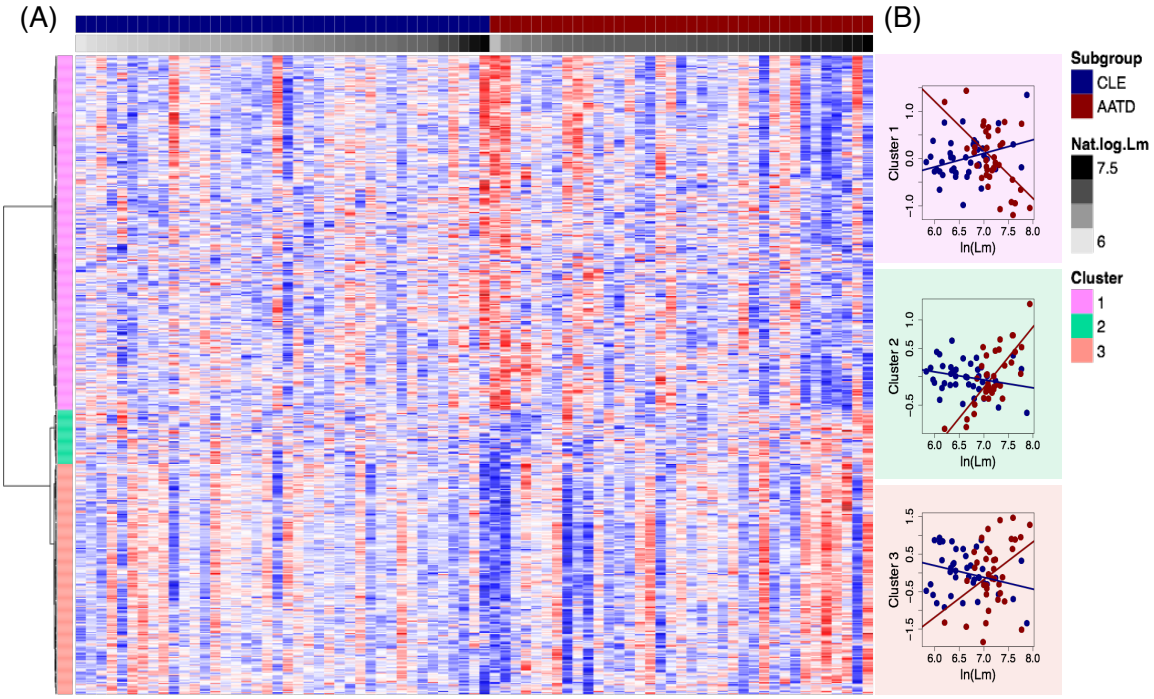

Figure 5. Gene expression profiles changing with emphysema severity dependent on SERPINA1 status. (A) Supervised heat map of 616 genes (uncorrected $p < 0.01$; 197 expected by chance) associated with the interaction between the natural log of Lm and emphysema subtype. Samples (columns) are supervised first by group and then within each group from mild to severe emphysema (left to right). Rows represent genes, are z-score normalized and hierarchically clustered, resulting in three columns (1 - pink, 2 - green, and 3 - orange). (B) The z-scored mean of each of these three clusters is plotted against the natural log of Lm, colored by group, and a best fit line is added.

In order to demonstrate that the 616 identified genes were associated with regional emphysema severity dependent on genotype and were not identified based on differences in Lm measurements between lungs, I again standardized the Lm measurements within

each lung and reran the statistical analysis. This analysis revealed 231 genes (uncorrected $p < 0.01$) associated with standardized emphysema severity and dependent on genotype, 124 of which were included in the original 616 genes ($p < 0.0001$ by Fisher's exact test).

Relationship to expression profiles induced by activation of the UPR

To further test the relationship between the 222 genes from cluster 3 (orange cluster, Figure 5) and the unfolded protein response, which is known to be disrupted in the liver in AATD, I used GSEA to test whether the genes induced by activation of the UPR are enriched with genes from cluster 3. There are three arms of the UPR, each regulated by a different transcription factor, namely ATF4, ATF6, and XBP1. Cluster 3 genes were enriched amongst the genes most induced by overexpression of ATF4 compared to empty vector control in BEAS2B cells ($p < 0.001$) [21]. This cluster was similarly enriched amongst genes induced by orthogonal, small-molecule-mediated activation of ATF6 (trimethoprim treatment) or XBP1 (dox treatment) in HEK293 cells ($p < 0.001$) [61]. Tunicamycin is a classic inducer of the UPR that leads to the accumulation of proteins in the ER, similar to what occurs in the liver of AATD patients. Cluster 3 genes were enriched amongst the genes most induced by tunicamycin treatment of immortalized B cells compared to DMSO treatment ($p < 0.001$) (Figure 6) [62].

Table 3. Enrichment results for genes from cluster 3, which have no relation to Lm in MM lungs yet increase with Lm in ZZ lungs (q < 0.05).

Category	FDR q value
Unfolded protein response (Homo sapiens)	0.003
N-glycan trimming and elongation in the cis-Golgi (Homo sapiens)	0.004
Nucleolus	0.008
Protein export (Homo sapiens)	0.017
Protein processing endoplasmic reticulum (Homo sapiens)	0.039
Spliceosome (Homo sapiens)	0.039
Extracellular vesicular exosome	0.044

ATF4 Overexpression ATF6 Activation XBP1 Activation Tunicamycin

Figure 6. Genes increasing with emphysema severity in AATD are induced with activation of the unfolded protein response. Cluster 3, with genes increasing specifically in AATD emphysema, is enriched amongst genes most induced by ATF4 over-expression in BEAS2B cells, ATF6 or XBP1 activation in HEK293 cells, or treatment with tunicamycin in immortalized B cells (GSEA p < 0.001). In each of the figures, the red to blue color bar represents the set of all measured genes, with red representing genes induced by UPR activation, and blue representing genes reduced by UPR activation.

Discussion

The goals of this study were to identify the gene expression changes associated with emphysema severity among lungs with and without alpha1-antitrypsin deficiency, and to identify any effects associated with one subtype or the other. By analyzing the transcriptome throughout ten explanted emphysematous lungs, I have identified gene expression changes common to both forms of emphysema and changes unique among the AATD lungs related to the UPR.

Campbell and colleagues previously identified a 127-gene signature of emphysema severity that was enriched with genes involved in tissue remodeling and wound repair. Specifically, they found that genes modulated as part of the TGFb signaling pathway were down regulated with emphysema severity, and were regulating collagen contraction. Their reversal by treatment with TGFb1 rescued a collagen contraction deficit in fibroblasts taken from GOLD stage IV COPD patients with severe emphysema [28]. In the present study, this gene expression signature was enriched with genes that changed with emphysema severity among MM and ZZ lungs in the parenchymal tissue. This finding suggests that there are mechanisms of disease pathogenesis and progression that are shared between MM and ZZ emphysematous lungs, even though the emphysematous destruction and clinical picture manifest differently in each group. One previous study has examined transcriptomic differences between “usual” and AATD-associated emphysema [25] and the authors reported that these two emphysema subtypes shared many transcriptomic changes. The concordant enrichment of the 127-gene signature in

both subtypes provides further evidence for this observation. Furthermore, I was unable to identify any genes significantly differentially expressed between the two subtypes. In the rest of this study, I set out to further characterize the similarities across the subtypes and to identify any existing transcriptomic differences associated with worsening emphysema that do depend on the ZZ mutation.

The 127-gene signature was reversed not only by treatment with TGFb1, but also by an endogenous, copper binding tripeptide, GHK [28]. As this transcriptionally identified compound and others similarly identified make their way into routine clinical care, characterization of the transcriptomic similarities and differences between emphysema and AATD-emphysema – the primary inherited form of the disease – will become increasingly important in order to predict the patients for whom these compounds will be effective. Identification of shared pathways may not only provide insight into pathogenesis, but may highlight potential additional therapeutic avenues going forward. Here, I identified 571 genes related to epithelial repair that decreased with emphysema severity, regardless of genotype.

Axonal guidance cues are molecules that are known to guide axonal growth, but more recently have been shown to play a critical role in lung growth, angiogenesis, and repair. Vadivel and colleagues [63] demonstrated that airspace enlargement in newborn rats was associated with decreased lung EphrinB2 (EFNB2), an axonal guidance cue that was among the genes that decreased with emphysema severity. More broadly, we know that

whatever the cause, epithelial repair in the lung is deficient in emphysema. In the normal lung, epithelial restitution involves epithelial cell spreading and migration to cover the denuded area. These cellular dynamics require actin regulation for the establishment of cellular polarity and the extension of lamellipodia in the direction of migration. During migration, the actin cytoskeleton interacts with focal adhesions, and integrins associate with components of the extracellular matrix for traction [64]. Among the genes that decreased with emphysema severity were genes associated with focal adhesions, ECM and lamellipodium organization, migration of epithelial cells, and regulation of the actin cytoskeleton, all processes critical to lung re-epithelialization and repair [64].

There are many clinical challenges encountered in the treatment of emphysema. Among them are a lack of understanding of pathogenesis, a lack of treatments, and a lack of available biomarkers. Most of the gene expression profiles that changed with Lm decreased with increasing severity. However, among the 126 genes that increased with severity was SFTPD, the gene coding for surfactant protein D, a marker of lung damage [42] and potential biomarker of COPD and emphysema severity [40, 41, 43, 65, 66]. Surfactant protein D is a protein involved in marking apoptotic cells for immune-cell mediated clearance [44] and its increasing with damage in emphysema supports evidence pointing towards increased apoptosis as one of the pathogenic mechanisms in this disease (REF). Others have observed a relationship between serum levels of SFTPD and emphysema before [41] but none have measured it repeatedly within a single lung nor measured it at the level of gene expression. That the expression of SFTPD increased with

emphysema severity within a single lung, regardless of the SERPINA1 genotype, suggests it may be a dynamic biomarker of disease severity that could be used to track disease progression and not only facilitate diagnosis.

Alpha1-antitrypsin deficiency associated emphysema may not represent the same public health burden as "usual" emphysema, but it leads to more severe and earlier onset disease that places a larger disease burden on those diagnosed. Accumulation of the ZAAT protein in the liver leads to cirrhosis and hepatocellular carcinoma and decreases the amount of functional AAT in circulation. The loss of AAT allows unchecked destruction of the lung tissue by neutrophil elastase, at least partly explaining the emphysema associated with AATD. Recently, the ZZ protein has been shown to accumulate not only in hepatocytes, but also in bronchial epithelial cells [51], suggesting that the lung may play a role in its own destruction. One study showed that accumulation of the ZZ protein hypersensitized hepatocytes to ER stress, which, upon secondary ER insult, more readily lead to induction of the unfolded protein response (UPR) and apoptosis [54]. The UPR primarily supports cell survival and its activation has many downstream effects, mediated through three arms. One arm, regulated by PERK and ATF4, transiently decreases the translation rate to decrease a cell's protein folding burden [67-69]. A second arm, regulated by ATF6, increases the amount of protein folding chaperones [70, 71]. The last arm, regulated by IRE1a and XBP1, increases ER associated degradation [72, 73], and activates autophagy for better energy homeostasis [74, 75]. In the face of chronic, unresolvable ER stress, however, the UPR can instead lead to cell death [76]. Among the

AATD dependent genes, cluster 3 was enriched with genes involved in the UPR (Table 3), including genes involved in protein processing, protein export, and ER associated degradation (e.g. SERP1, DNAJB9, DNAJC1, DNAJC21, SEC62, SEC63, and HSP90B1). I further demonstrated that cluster 3 genes were enriched with genes induced by each of the three arms of the UPR or by treatment with tunicamycin (Figure 6). This is all to suggest that genes involved in the UPR are increasingly activated with worsening emphysema severity uniquely among AATD lungs. Given the role that accumulation of the ZZ protein plays in the manifestation of liver disease in the setting of AATD, increasing UPR activity associated with worsening emphysema severity in the lungs of AATD patients should not be surprising, but it may have important implications for AATD emphysema pathogenesis. Specifically, if the ZAAT protein is accumulating in the cells of the lung like it accumulates in the liver and sensitizing pulmonary cells to secondary sources of ER stress (e.g. tobacco smoke), then providing AAT replacement therapy will be necessary but insufficient to treat this particular form of emphysema. The findings presented in this study suggest the lack of functional AAT is not the only problem in disease, but rather that the ZAAT protein has gained a toxic function that affects the pathogenesis of both liver and lung disease in AATD.

Conclusions

In the current study, I have uncovered genomic changes associated with emphysema progression in the lungs of patients with “usual” or AATD-associated emphysema. There are many shared changes between these two emphysema subtypes. Genes involved in

various aspects of wound repair decrease in expression with increasing emphysema severity, while SFTPD increases with disease severity, suggesting it could act as a biomarker of disease progression. Importantly, I was able to provide evidence that the unfolded protein response is increasingly activated with emphysematous destruction specifically in AATD. Not only do the lungs of these AATD patients have a decrease in genes involved in re-epithelialization, and a decrease in the amount of circulating, functional AAT, but the increasing activity of the UPR in these lungs may further exacerbate emphysema by favoring cell death in the face of chronic, unresolvable ER stress.

Chapter Four: Gene expression profiles associated with healthy alveolar spacing and emphysematous destruction

Introduction

Emphysema, or destruction of the functional alveolar tissue compartments of the lung, represents one pathologic aspect of chronic obstructive pulmonary disease (COPD), which is predicted to be the third leading cause of death worldwide by 2020 [77]. Tobacco smoke is the leading cause of emphysema and COPD, but not nearly all smokers develop pulmonary disease; approximately 25% of smokers develop COPD [14]. Though chronic cough, sputum production, and frequent infections are common symptoms of emphysema, dyspnea, or shortness of breath, is the most debilitating factor. Unfortunately, no emphysema modifying therapies exist and the disease inexorably progresses, even after smoking cessation. That the majority of long term, continuous smokers do not develop emphysema implicates genetic susceptibility as an important factor in disease pathogenesis, and also suggests the possibility that protective and repair mechanisms are sufficiently active in the majority of lungs. Cells of the normal, undamaged lung may turn over slowly, but rapid repair is possible. Epithelial cells migrate to cover denuded spaces, undifferentiated cells proliferate to replace the full epithelial layer, and tissue homeostasis can be restored as soon as 14 days post pulmonary injury [64, 78]. The exact mechanisms that regulate either emphysema pathogenesis or physiologically appropriate repair remain poorly characterized, but by improving our understanding of these processes we may learn to prevent emphysema

from occurring in the first place, or to augment the normal repair process in emphysema to reverse damage that accumulates over a lifetime of smoking.

A process like emphysema is undoubtedly multi-faceted, but one widely accepted mechanism by which emphysema arises is an imbalance between the proteinases and anti-proteinases within the lung. Inhaled tobacco smoke recruits neutrophils [79], which are known to be increased in the lungs of patients with COPD, even after smoking cessation [80]. Neutrophils release a powerful elastase (i.e. neutrophil elastase, NE) capable of degrading the extracellular matrix and destroying the parenchymal tissue of the lung[81]. Patients with AATD, an inherited deficiency of the primary inhibitor of NE, have insufficient circulating alpha1-antitrypsin and so neutrophil elastase destroys the parenchyma, unchecked, leading to severe emphysematous destruction at an early age. Neutrophil recruitment to, and accumulation in the lung is not sufficient to cause emphysema, however, as acute inflammatory processes, like those seen in pneumonia, develop and resolve without permanent loss of functional tissue. Importantly, not all smokers develop appreciable disease, pointing towards the presence of additional mechanisms relevant for emphysema pathogenesis.

Inflammation and proteolysis alone cannot explain emphysema and so other processes must be involved. Tobacco smoke not only recruits immune cells, but also causes damage through the production of reactive oxygen species (ROS) [82, 83]. The oxidative stress response and the unfolded protein response (UPR) normally help protect against ROS-

induced damage. NFE2L2 (aka NRF2), the primary transcription factor that induces the expression of more than 100 anti-oxidant genes in response to ROS, has been shown to be disrupted in COPD [24, 84, 85]. Aspects of the UPR are also disrupted in disease [21, 86, 87].

Not only protective mechanisms, but also repair mechanisms are known to be disrupted in the lungs of patients with COPD and emphysema. Fibroblasts are normally responsible for deposition of extracellular matrix, a critical aspect of post-injury re-epithelialization, and have diminished repair capacity to produce ECM in COPD patients [88]. This deficiency may be regulated by the TGFβ signaling pathway [28], which is well known to be altered in COPD [25, 89-91]. Notch signaling is also important in alveolar morphogenesis and repair and is likewise disrupted in COPD [23, 26, 92-94]. Lastly, at least one study has demonstrated that epithelial cells in the lungs of patients from COPD have a diminished ability to migrate and cover a denuded space, thought to be the first step in repair of a damaged epithelium [64, 95]. Cell signaling, especially the Notch and TGFβ pathways, is disrupted in COPD and leads to observable phenotypic deficiencies in epithelial cell and fibroblast repair functions.

Few studies have focused on profiling the gene expression programs active within a healthy lung. Cross-sectional transcriptomic studies include a non-diseased control group to be used as a baseline for comparison to a disease state, and any identified differentially expressed genes are interpreted in the context of the disease of interest. These kinds of

studies thus offer little insight into the activity of gene expression repair or protection programs in the normal lung and how they maintain homeostasis of the lung . One study to date has investigated *in vivo* the gene expression changes associated with normal repair [78]. These authors reported an increase in proliferating dedifferentiated cells and a tightly coordinated set of genes involved in cell cycle were present seven days post pulmonary injury, and that by 14 days post injury the airway epithelium appeared normal and gene expression alterations had reverted to baseline.

The cells of the lung have been observed to be capable of proliferation, migration, and repair [63, 64, 78, 92, 94, 96, 97]. However, the mechanisms underlying these processes remain poorly characterized, and only in rare instances are otherwise healthy lungs collected for research, which is necessary for the elucidation of the normal processes occurring in the lung. In this study, I have leveraged data from five lungs from donors for whom no match could be found. These lungs were released for research and our collaborators at UBC subjected eight sections from each of these lungs to laser capture microdissection, separating airway and parenchymal tissue, and microCT to measure Lm. Members of our lab completed gene expression profiling on these sections. I have used these data to identify the genes and pathways that change with increasing alveolar spacing and to determine their relation to similar changes associated with emphysema. Campbell et al. previously completed a similar study including lungs mostly from patients with emphysema and identified TGFβ signaling as a pathway with genes decreasing with emphysema progression [28]. The goal of the present study is to identify

genes changing similarly and differently with alveolar spacing in healthy lungs and emphysema severity in diseased lungs. I report for the first time that there are genes associated with normal increases in alveolar spacing within healthy lungs, as measured by the mean linear intercept (Lm). Surprisingly, the majority of expression profiles positively associated with Lm in controls were negatively associated with Lm in diseased lungs, suggesting similar processes may be active in both groups but behaving differently. Macrophage M2 marker genes (e.g. MSR1, MARCO) were among the few genes that increased with Lm in both control and emphysematous lungs, while genes involved in TGFb signaling and its regulation (e.g. NEDD4, TGFbR1) increased in control and decreased with Lm in diseased lungs.

Materials and Methods

Microarray gene expression data analysis strategy

I completed the following statistical analyses in a manner similar to a previous publication from our lab group [28]. In the expression data from the parenchymal tissue I utilized linear mixed effects modeling in conjunction with the likelihood ratio test (LRT) to identify differentially expressed genes associated with disease status (i.e. emphysematous tissue v control tissue) or emphysema severity and alveolar spacing (i.e. the natural log of Lm. Throughout the rest of this chapter I will simply refer to Lm, for easier reading). I undertook three separate analyses to identify (1) genes associated with group status, (2) genes changing with increasing alveolar spacing and emphysema severity across control and emphysematous lungs, and (3) genes changing with increasing

alveolar spacing differently than they change with emphysema severity. In each analysis, I identified a signature, compared the signature to a published data set or gene expression signature, and performed biological enrichment analysis. In the analyses examining gene expression effects associated with Lm, I also tested whether the signatures changed concordantly with Lm after standardizing Lm within each patient lung to a mean of zero and standard deviation of one (Z-Score). This helps to guard against patient outliers from disproportionately driving the makeup of the signature. For analysis 2, I further tested the consistency of the signature by similarly standardizing Lm within each group. This can help guard against a difference in the mean Lm between groups from affecting the Lm signature.

I completed all statistical analysis in R v3.1.1. I used the R package *nlme* v3.1-128 for all LME modeling and I compared nested LME models by analysis of variance (ANOVA) using LRT. In some cases I used gene set enrichment analysis, GSEA v2.2.3 [59] to perform comparative genomic analysis. Any signatures I identified I also hierarchically clustered prior to biological enrichment analysis, which I completed with Enrichr [60], with the search space limited to KEGG, Biocarta, Reactome, and gene ontology (GO) databases.

Analysis 1. Gene expression differences between emphysema and control lungs

To identify genes associated with disease status I generated two models and compared them by the LRT:

(1) *Gene ~ slice + ln(Lm) + random(patient) + error*

(2) *Gene ~ slice + ln(Lm) + disease + random(patient) + error*

Here, slice represents the height from which a sample was taken from within a lung, ln(Lm) represents alveolar spacing or emphysema severity, and disease represents a binary variable denoting whether the tissue is from an emphysematous or from a non-emphysematous lung.

Analysis 2. Gene expression changes associated with alveolar spacing and emphysema severity

I next wanted to identify genes changing with increasing alveolar spacing and emphysema severity – i.e. Lm – consistently across control and emphysematous lungs and utilized the following modeling schema, again comparing models with the LRT:

(3) Gene ~ slice + disease + random(patient) + error

(4) Gene ~ slice + disease + ln(Lm) + random(patient) + error

Analysis 3. Gene expression changes associated with alveolar spacing or emphysema severity

Finally, I wanted to identify any gene expression profiles associated with alveolar spacing differently than they are associated with emphysema severity. I did so by identifying gene expression profiles better fit to a model that includes the interaction between Lm and group than to a model that does not include the interaction:

(5) Gene ~ slice + disease + ln(Lm) + random(patient) + error

(6) Gene ~ slice + disease + ln(Lm) + disease:ln(Lm) + random(patient) + error

Results

Study Population

Our collaborators at the University of British Columbia (UBC) measured Lm using micro-CT scans taken from regularly spaced regions from within the lungs of five patients with emphysema and five organ donors with intact lungs released for research. As expected, the lungs from patients with emphysema had a higher mean Lm than the lungs from the organ donors with no appreciable emphysema ($p < 0.0001$) (Figure 7). Table 4 shows demographic, clinical, and imaging information for the 10 lungs. The five diseased lungs were characterized by centrilobular emphysema (CLE).

Table 4. Demographic information for the 10 lungs, five from emphysematous smokers, and five from donors for whom an appropriate match could not be found.

Patient	Group	Sex	Age	PackYears	Smoker	LMSummary	LMRange
6996	Emphysema	F	77	45	former	512 (176)	340-903
7305	Emphysema	F	58	30	former	823 (99)	715-967
7307	Emphysema	M	55	80	former	959 (718)	475-2635
7336	Emphysema	F	59	40	former	572 (232)	371-1102
7337	Emphysema	F	53	24	former	1397 (536)	839-2363
6991	Control	F	65	unknown	former	400 (44)	348-491
6994	Control	M	64	15	current	389 (38)	346-429
7008	Control	M	42	15	current	353 (33)	305-398
7300	Control	M	53	0	never	381 (16)	360-405
7309	Control	M	77	0	never	339 (36)	296-389

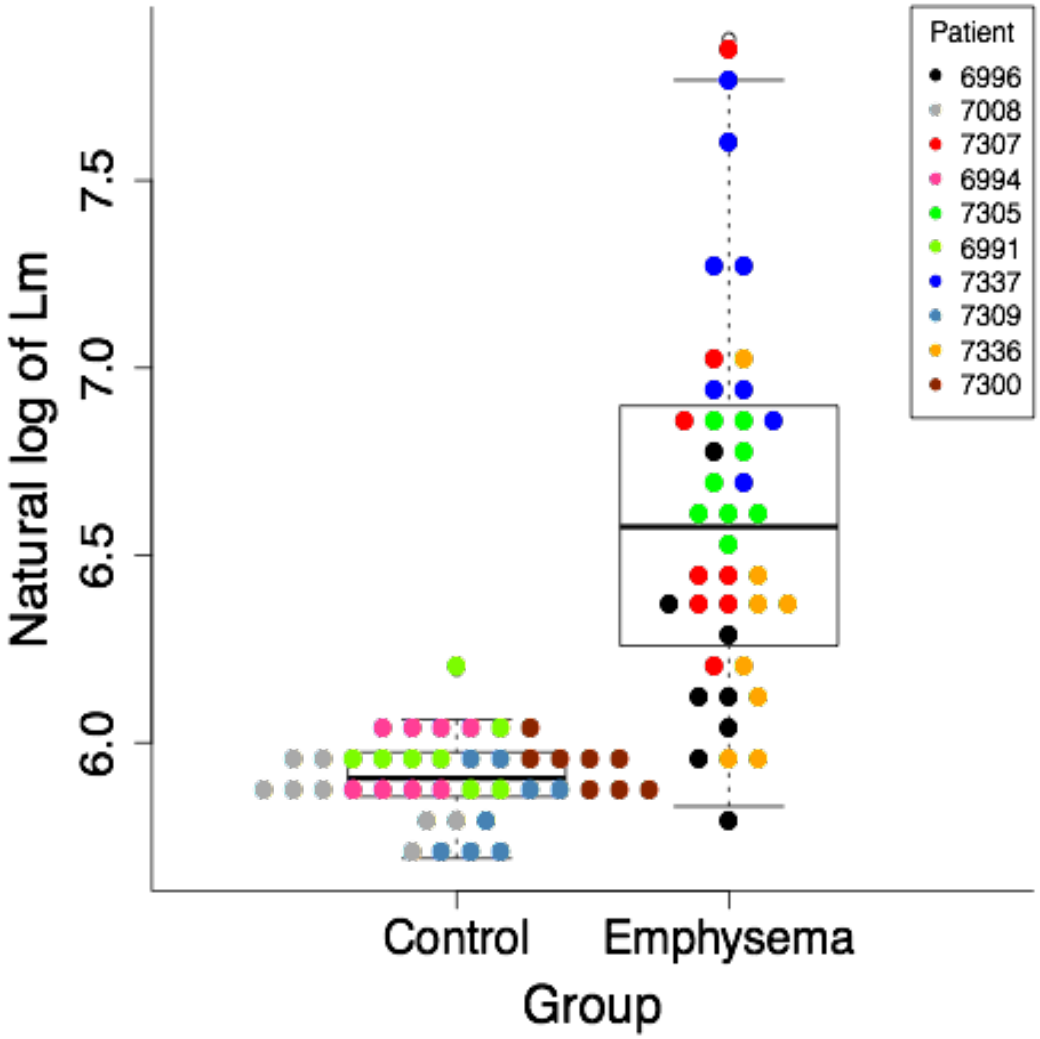

Figure 7. Distribution of natural log of Lm measurements by disease status. Each patient is represented by a different color and has up to eight samples. The samples from the emphysema patients have a significantly higher distribution than the controls ($p < 0.0001$).

Genes differentially expressed between emphysematous and control lungs

Before investigating the gene expression effects of Lm in these 10 lungs, I wanted to identify any genes that differentiated the two groups, characterize the biology of these

changes, and compare these lungs to similar, previously published samples. With linear mixed effects models controlling for height within the lung (i.e. slice), alveolar spacing (i.e. Lm), and a random patient effect, I identified 112 genes that were differentially expressed between emphysematous and control lungs ($q < 0.01$) (Figure 8A). Hierarchical clustering generated two clusters of genes, one with 60 genes increased in emphysema, and one with 52 genes decreased in emphysema. With a list of genes ranked by their association with group (i.e. the t-statistic for the coefficient of group from the model), I performed an enrichment analysis with GSEA [59] to identify pathways changing with disease status. Pathways involved in artery development were enriched with genes increased in emphysematous tissue, while pathways involved in ribosome biogenesis, mitochondrion, and energy metabolism, were all enriched with genes that

decreased ($q < 0.001$).

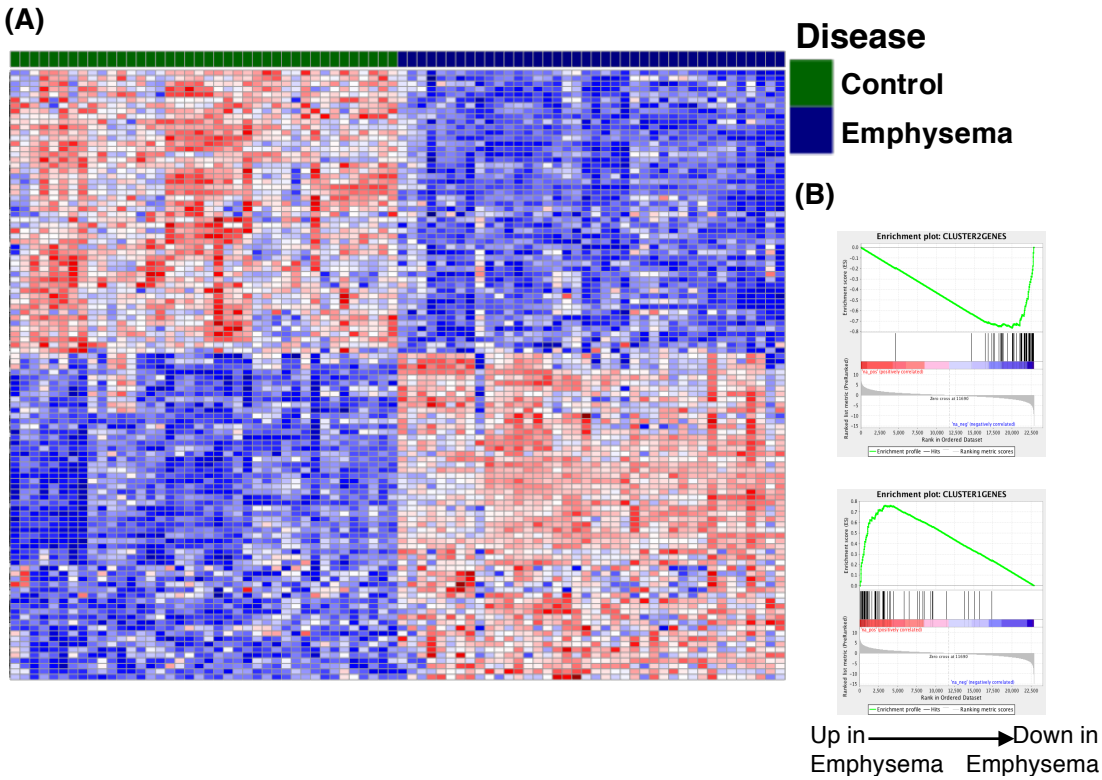

Figure 8. (A) 112 genes are differentially expressed between emphysematous and healthy, donor lungs. These genes are broken into two clusters by hierarchical clustering, one set of genes repressed (top cluster) and one induced in emphysema (bottom cluster). In the heat map, each column represents a sample and each row a gene. (B) Each of these two clusters is significantly and concordantly enriched with genes that change between emphysematous (n=5) and control lungs (n=2) from Campbell et al.'s cohort ($p < 0.001$).

Connection to independent emphysema dataset

To test whether the gene expression differences observed here were conserved in other emphysematous tissue, I ranked genes by their association (i.e. t-statistic) with emphysema in Campbell et al.'s [28] cohort of emphysematous (n=5) and control lungs (n=2). The two clusters of 60 up and 52 down genes were concordantly enriched in this independent dataset ($q < 0.001$) (Figure 8B).

Genes associated with alveolar spacing and emphysema severity

The true novelty of this project is not in characterizing gene expression profiles differentially expressed cross-sectionally across groups, but in its potential to identify gene expression profiles associated with increasing alveolar spacing in non-diseased, control lungs and emphysema severity in diseased lungs. Using linear mixed effects models, I identified 131 genes significantly associated with Lm across these 10 lungs ($q < 0.05$) (Figure 9). Hierarchical clustering generated five clusters within these 131 genes. Due to the strong differences between the groups of lungs, coupled with the uncertain nature of Lm within non-emphysematous lungs, I wanted to confirm that these Lm-associated genes were not being driven by any particular patient or by either group. To test whether any outlier patients were driving the results, I standardized the Lm measurements within each patient to a mean of zero and standard deviation of one (z-score). I reran the analysis and tested with GSEA whether the 131 genes were enriched with genes that also changed with this standardized Lm measurement. To test whether the strong group difference in Lm (Figure 8) was driving the results, I next repeated this analysis after standardizing the Lm measurements within each group. In both patient and group-standardized Lm analyses, the five original clusters were concordantly enriched ($q < 0.001$) (Figure 9B, second column shows patient-standardized results; group-standardized not shown). Clusters one and five increased with Lm, while clusters two, three, and four all decreased with Lm. Lastly, I wanted to test whether this Lm signal was detectable in each group separately or if it was being driven by one group in particular. I ranked genes by their association with Lm in each group and tested whether the five

clusters were enriched in the control lungs or in the emphysematous lungs. Cluster one is the only cluster enriched with genes that changed with Lm in controls ($q < 0.001$), while all five clusters were strongly enriched in emphysematous lungs ($q < 0.001$). Cluster one contained 42 genes, 26 of which were part of the core enrichment in both emphysematous and control lungs. Enrichr [60] revealed that these 26 genes were associated with the phagosome ($q < 0.005$ – MARCO, MSR1, NCF2, HLA-DRB1), monocyte differentiation ($q < 0.05$ – PDE1B, PPARG), activation of the immune response ($q < 0.05$ – MARCO, C1QA, SYK, MNDA, HLA-DRB1), and CD14+ Monocytes ($q < 0.001$ – DOK2, NCF2, SLC11A1, ALOX5, EMR1, C1ORF162, HK2).

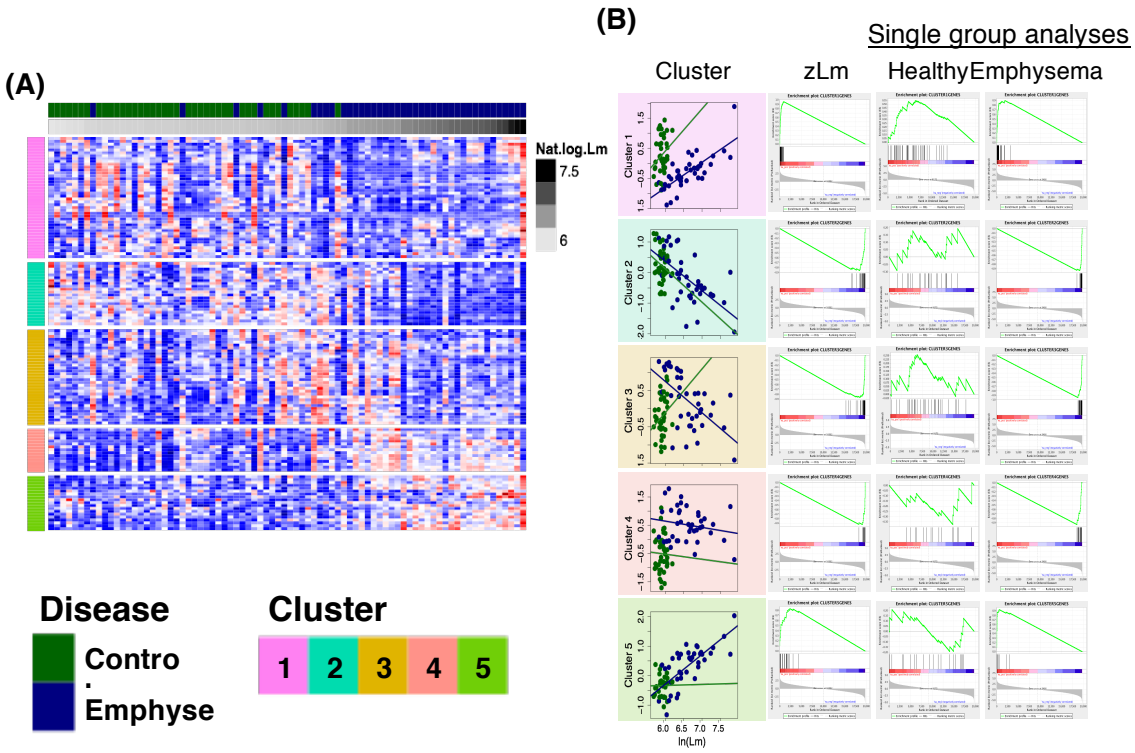

Figure 9. 131 genes are significantly associated with Lm across emphysematous (CLE) and healthy, donor lungs (q < 0.05). (A) These 131 genes are presented in a heat map, supervised by Lm from no emphysema to most severely emphysematous sample (left to right). Each column represents a sample and each row a gene expression profile. Hierarchical clustering identifies five clusters among these genes. (B) The z-scored mean of each gene cluster is plotted against the log of Lm and colored by group in the leftmost column and a best fit line is added. The second column shows that these clusters are enriched with genes that change even after standardizing Lm measurements within each lung to a mean of zero and standard deviation of one (GSEA q < 0.001). In the third and fourth columns, the five clusters are tested for enrichment with genes changing with Lm within either control or emphysematous lungs alone. Cluster one is significantly enriched in the same direction in both groups (p < 0.001). Clusters two through five are not significantly enriched in control lungs (p > 0.05). These clusters are strongly significantly enriched in emphysematous lungs (p < 0.001).

Pathways and signatures changing with Lm

The 131 genes were identified as genes associated with Lm across all 10 lungs. However, further interrogation of the five clusters they formed demonstrated that the majority of these genes (clusters two through four) were associated with Lm in emphysematous lungs only, and not in control lungs. To determine whether this lack of enrichment was due to a

lack of appreciable Lm signal within the control lungs, or rather, due to Lm associated changes that could not be identified in a model including only a main effect of Lm, I tested whether other signatures associated with Lm were enriched in the controls and the emphysematous lungs separately. Campbell and colleagues [28] published a 127-gene signature of emphysema severity, and in the prior chapter I discovered a 571-gene signature of emphysema severity. I tested with GSEA the association of both of these signatures with Lm in controls and used the five emphysematous lungs as a positive control for comparison. Both gene signatures were concordantly enriched in the emphysematous lungs ($p < 0.001$). Campbell's signature was significantly enriched in the opposite direction in controls ($q < 0.001$). Similarly, the genes decreasing with emphysema severity (chapter three signature) were enriched with genes that increased with Lm in controls ($q < 0.001$) (Figure 10C), changing in the opposite direction of that expected.

The observation that two previously identified gene expression signatures of emphysema severity were enriched in control lungs in the direction opposite of that expected begs the question, what other genes or pathways change with Lm in controls and how do these changes compare to those occurring in emphysema? In each group of lungs, I ranked genes according to the t-statistic of the coefficient for Lm and tested with GSEA whether gene sets from KEGG, Reactome, Biocarta, and Gene Ontology were enriched with Lm-associated genes. 1464 and 402 gene sets were significantly enriched with genes associated with Lm in control and emphysematous lungs, respectively ($q < 0.05$). 195 of

these gene sets were enriched with genes that increased with Lm in controls and that decreased with Lm in disease. Many of these 195 gene sets overlap, so I collapsed them into 12 gene sets involved in various processes including endothelial cell development, response to nutrient levels, proteasome mediated degradation, epithelial cell migration, and TGFβ signaling. Five of the gene sets were enriched with genes that increased in both control and diseased lungs, and I collapsed these into one gene set associated with peptide chain elongation (Figure 10A, 4B).

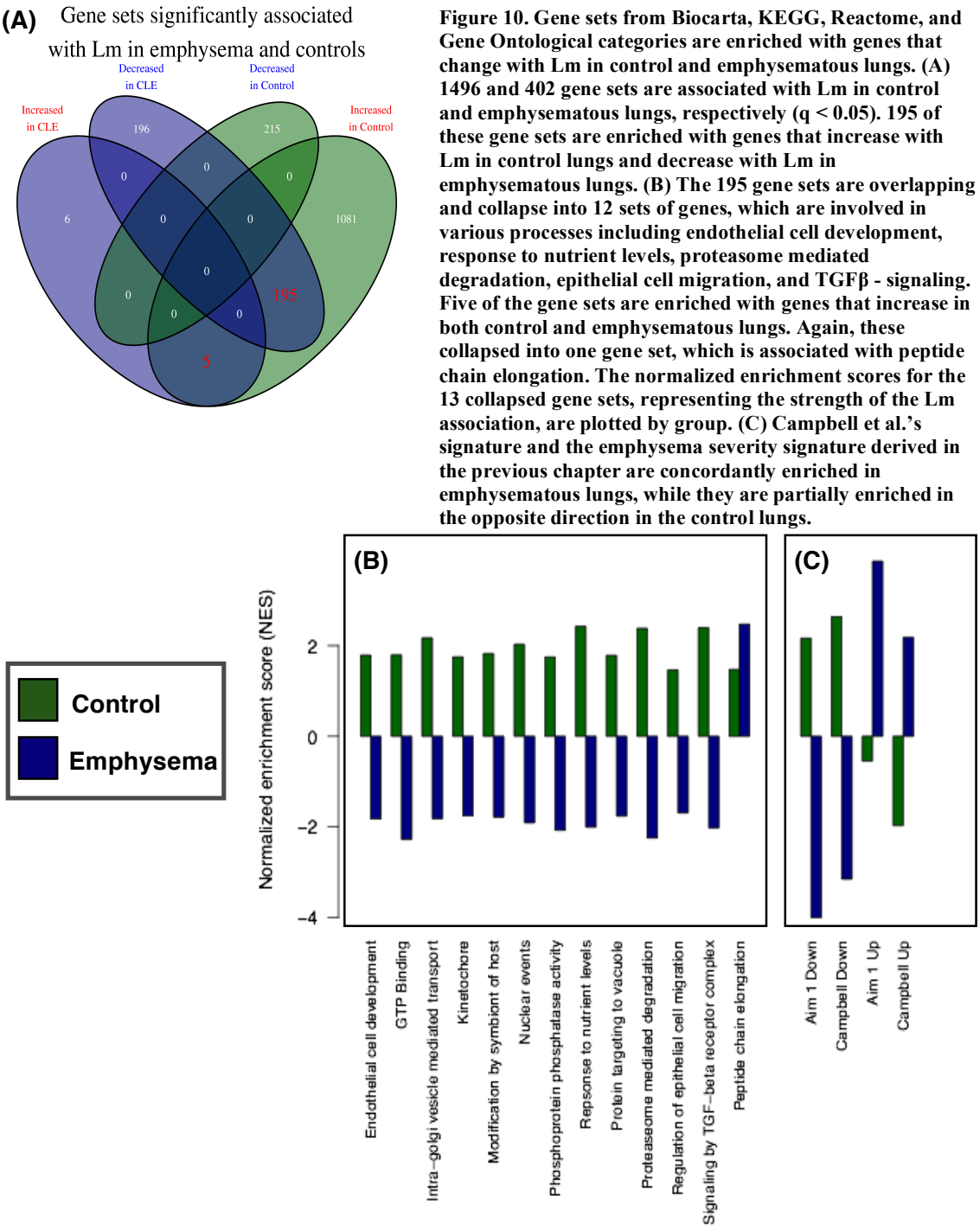

Figure 10. Gene sets from Biocarta, KEGG, Reactome, and Gene Ontological categories are enriched with genes that change with Lm in control and emphysematous lungs. (A) 1496 and 402 gene sets are associated with Lm in control and emphysematous lungs, respectively ($q < 0.05$). 195 of these gene sets are enriched with genes that increase with Lm in control lungs and decrease with Lm in emphysematous lungs. (B) The 195 gene sets are overlapping and collapse into 12 sets of genes, which are involved in various processes including endothelial cell development, response to nutrient levels, proteasome mediated degradation, epithelial cell migration, and TGFβ - signaling. Five of the gene sets are enriched with genes that increase in both control and emphysematous lungs. Again, these collapsed into one gene set, which is associated with peptide chain elongation. The normalized enrichment scores for the 13 collapsed gene sets, representing the strength of the Lm association, are plotted by group. (C) Campbell et al.'s signature and the emphysema severity signature derived in the previous chapter are concordantly enriched in emphysematous lungs, while they are partially enriched in the opposite direction in the control lungs.

Genes and pathways changing with alveolar spacing differently than they change with emphysema severity

That 195 gene sets were enriched with genes that increased with Lm in controls and decreased with Lm in emphysematous lungs, raises the possibility that there are gene expression profiles that significantly that change with Lm in a group dependent manner (i.e. they associated with the interaction between Lm and group). I identified 883 gene expression profiles significantly associated with the interaction between Lm and disease ($q < 0.1$). These genes changed with alveolar spacing in control lungs differently than they changed with emphysema severity in diseased lungs, both as measured by Lm. Hierarchical clustering of these genes broke them into two clusters of 431 and 452 genes (Figure 11A). Cluster 1 (purple) decreased with Lm in controls and increased with Lm in emphysema, while cluster 2 (green) increased with Lm in controls and decreased with Lm in emphysema (Figure 11B). Cluster 2 was enriched with genes involved in TGFβ signaling, genes regulated by the vitamin D receptor, immune system genes, and genes involved in the DNA damage response ($q < 0.005$).

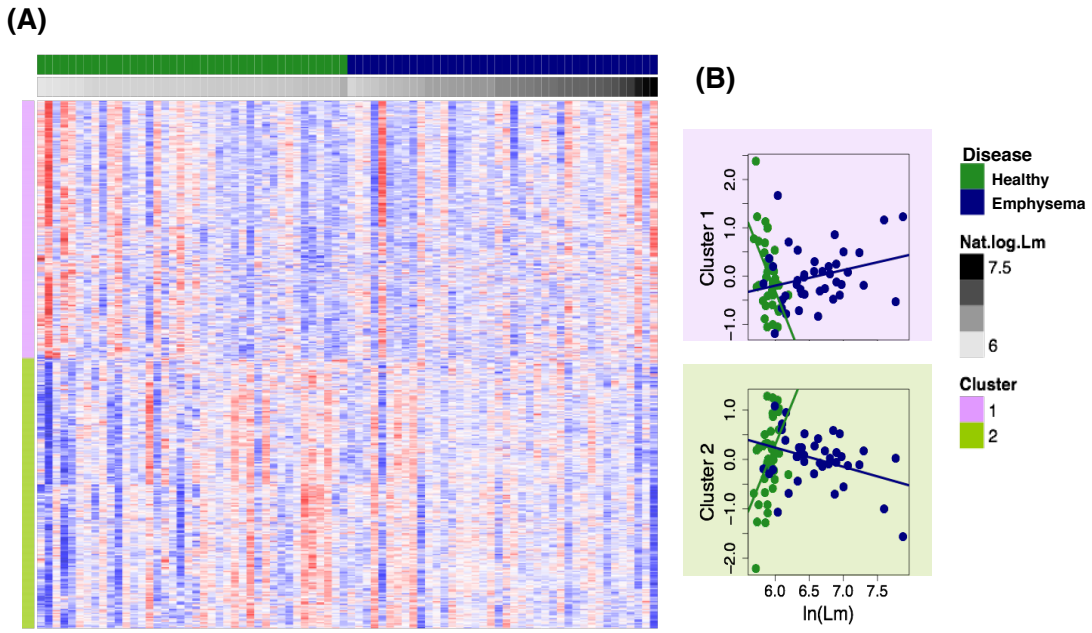

Figure 11. 883 genes are significantly associated with the interaction between Lm and disease status (q < 0.1). (A) These genes are presented in a heat map, which is supervised first by group, and then by increasing Lm from left to right. Each column represents a sample and each row a gene expression profile. Red represents relatively high expression and blue relatively low expression. Hierarchical clustering separates these into two sets of genes: cluster 1 (purple) with 431 genes and cluster 2 (green) with 452 genes. (B) The z-scored mean of these clusters is presented and a best fit line is added.

Discussion

The goal of this study was to compare and contrast gene expression changes associated with normal alveolar spacing to gene expression changes associated with emphysema severity in order to identify aspects of the lung homeostasis and repair program and determine its connection to emphysema. In the long term, the identification of any connection between the homeostasis repair program and emphysema can improve our understanding of the deficiencies of repair present emphysema and guide us towards safer strategies for developing disease modifying therapies that are similar to the naturally occurring repair system. By collecting both control and emphysematous lungs, measuring gene expression at various points throughout each lung, and quantifying alveolar spacing at these points, our collaborative effort has identified stark differences in gene expression patterns occurring in control compared to emphysematous lungs.

I compared group differences discovered here to previously profiled lungs and compared biological pathways enriched here to those previously reported. The genes decreased in emphysematous lungs are involved in energy maintenance and metabolism, consistent with what was previously observed when comparing emphysematous and control lungs [25]. Furthermore, there is a strong and clear connection to a previous gene expression dataset of emphysematous and control tissue [28] (Figure 8B). These two observations suggest that the lungs included in this study are similar to those previously profiled.

Across all 10 lungs, I identified 131 genes associated with Lm (Figure 9A). After thorough follow up, however, I discovered that these genes' relationship with Lm was driven mostly by changes in gene expression with emphysema severity, and they bore little relationship to alveolar spacing in control lungs (Figure 9B). Yet, cluster one was associated with Lm in both emphysematous and control lungs, showing that some genes change with Lm regardless of disease. Cluster one included phagosome genes, monocyte markers, and genes involved in monocyte differentiation. MSR1 (a.k.a. CD204) is expressed by anti-inflammatory, alternatively activated (M2), alveolar macrophages, and these CD204+ macrophages have been shown to be increased in the lungs of patients with COPD at GOLD stages III and IV compared to stages I and II, and to non-smokers. The number of CD204+ macrophages was also shown to be negatively correlated with the percent predicted forced expiratory volume [98]. A coding single nucleotide polymorphism (SNP) in MSR1, SNP P275A, is also associated with susceptibility to COPD and a lower forced expiratory volume to forced vital capacity ratio (FEV1/FVC) [99]. M2 macrophages serve many purposes. Using phagocytosis, they clear bacteria and cellular debris from damaged tissue; they also secrete TGFβ, promote angiogenesis, and support wound healing [100-102]. In this study, I have shown that macrophage associated genes increased with Lm in both control and emphysematous lungs. This suggests there are more M2 macrophages present in pulmonary sections with higher Lm values, regardless of disease status. In emphysematous lungs, the Lm measurement clearly indicates degree of damage, yet the biological interpretation of Lm in controls is not so readily apparent. The suggested increase in macrophages with Lm in controls can be

further extrapolated to suggest that increasing Lm in control lungs – while it does not represent damage to the degree seen in emphysema – may indicate minor amounts of damage undergoing inflammation and repair.

Though I was able to identify one cluster of macrophage associated genes increasing with Lm in both control and emphysematous lungs, the remaining four clusters were not enriched with Lm-associated genes in controls. This observation led me to question the nature of the Lm-associated gene expression changes in control compared to emphysematous lungs. Campbell et al.'s [28] 127-gene signature of emphysema severity (i.e. Lm) and my 571-gene set associated with emphysema severity were both concordantly enriched in the emphysematous lungs, but were reversed in the control lungs. Campbell's signature is associated with TGFβ signaling, inflammation, and tissue repair. The emphysema severity signature (chapter 3) is enriched with genes involved in angiogenesis and the regulation of epithelial cell migration. Broadly, these are all part of the epithelial repair system, and their reversal in controls suggests that similar genomic epithelial repair programs are active in both groups but that they act differently between diseased and non-diseased lungs. I expanded this exploration to include pathway based gene sets from KEGG, Reactome, Biocarta, and the gene ontology categories, and discovered that 195 gene sets were enriched with genes that increased and decreased with Lm in control and emphysematous lungs, respectively (Figure 10A). These gene sets are involved in TGFβ signaling, among other things. TGFβ signaling plays an important role in branching morphogenesis and epithelial differentiation in the developing lung [103],

and so this finding may represent one aspect of the developmentally conserved pathway active in lung maintenance and repair. This pathway has repeatedly been shown to be dysregulated in emphysema at both the gene expression and SNP level [28, 89, 91, 104, 105], further implicating the pathway and highlighting possible drivers of dysfunction.

Looking beyond the pathway based analysis, I wanted to determine if any genes were significantly associated with the interaction between Lm and disease-state in these 10 lungs. That is, do any gene expression profiles change with alveolar spacing in control lungs differently than they do with emphysema severity in diseased lungs. The differences between alveolar spacing and emphysema severity are emphasized by the discovery of 883 such genes (Figure 11A). Again, these genes, which are involved in TGFβ signaling (e.g. SMAD4, TGFBR1), decreased with emphysema severity and increased with Lm in control lungs. SMADs, key regulators of TGFβ signaling, are themselves regulated by the ubiquitination-proteasome pathway, specifically genes like SMURF2 and NEDD4, which are included in these 883 genes (cluster 2) [106]. NEDD4 has also previously been associated with emphysema severity [56].

Conclusions

In this study, I have identified gene expression profiles changing with alveolar spacing in control lungs and determined that these changes sit in stark contrast to those observed in emphysematous lungs. Though macrophage markers like MSR1 and MARCO increase with both healthy alveolar spacing and emphysema severity, the majority of gene

expression profiles associated with Lm are positively associated with Lm in controls and negatively associated with Lm in emphysematous lungs. In emphysematous lungs, Lm is a well-established marker of alveolar damage, but the interpretation of Lm in non-diseased lungs is less clear. The increase with Lm of the expression of macrophage marker genes, as well as genes involved in TGFβ signaling and its regulation, suggest that Lm measures normal alveolar damage undergoing healthy inflammation, resolution, and repair. The reversal of the relationship between these changes and Lm in emphysema implicates a deficiency of these normal processes as an important mediator of disease progression. One of the longstanding hypotheses of emphysema pathogenesis is that an imbalance between proteinases and anti-proteinases – brought about by increased pulmonary inflammation – leads to unchecked destruction of the parenchymal tissue. The observations presented in this study can push the field towards closer consideration of the aspects of emphysema that occur post tissue damage, namely a lack of functional tissue repair. In fact, inflammation occurs frequently in the lung without leading to emphysema, as is the case in pneumonia, and occurs as part of normal wound repair [107-110]. Therefore, the inflammation so frequently cited as a critical driver of emphysema may be secondary to the lack of repair, which cyclically leads to inflammation, more irreparable damage, and eventually the inexorable progress of emphysematous destruction.

Chapter Five: Transcriptomic connection between COPD and lung cancer implicates role for H3K27Me3 in COPD-associated pulmonary carcinogenesis

Background

Chronic obstructive pulmonary disease (COPD) is a major cause of morbidity and mortality in the United States. It is burdensome both to patients and to the health care system, and while it was the fifth most common cause of death in 1980, the Centers for Disease Control reported that it was the third most common cause of death in the U.S in 2014 [1]. Approximately 15 million people are diagnosed with COPD (6.4% of the U.S. population) and these individuals are more likely to be unable to work, to have an activity limitation, and to have difficulty walking or climbing stairs [3]. Patients with COPD are also more likely to develop lung cancer.

Skillrud and colleagues [4] observed for the first time in 1986 that COPD patients compared to healthy smokers had an increased rate of lung cancer development. These researchers prospectively compared COPD cases to controls matched for age, sex, occupation, and smoking history, and found that lung cancer developed more frequently in the patients with COPD. Matching by smoking history was critically important for this study because both COPD and lung cancer are both known to be caused by tobacco smoke. Recently, several groups have repeatedly to rediscovered the association between COPD and lung cancer [6-8, 31]. One meta-analysis of 39 studies calculated that the combined relative risk of lung cancer in COPD was 1.83 and demonstrated that diagnoses of COPD, chronic bronchitis, and emphysema all carried increased lung cancer risk [32].

One important addition to these observations came from a study in 2014 that showed that tumors in the lungs of patients with COPD were more likely to originate in regions of severe emphysematous destruction [33]. Furthermore, alpha1-antitrypsin deficiency, which is associated with the most severe form of emphysema, also carries an increased lung cancer risk [34-36]. Whether COPD is an independent risk factor for lung cancer is still debated, and as recently as 2013, Powell et al. [5] suggested that smoking habit and the timing of COPD diagnosis could account for the observed increased incidence of lung cancer among COPD patients. When correcting for smoking and including only patients with a 10 year or greater history of COPD, however, this study still found that the odds of lung cancer development was greater among the COPD population (odds ratio: 2.18). The general consensus is that COPD increases lung cancer risk.

Proposed mechanistic links between COPD and carcinogenesis

At this point, the mechanisms that prime the lungs of COPD patients for malignant growth remain unclear, but the pathogenic processes of each individual disease are at least partially understood. The hallmarks of cancer include evading apoptosis, tissue invasion, and sustained angiogenesis. Interestingly the primary processes of COPD appear unrelated to or directly at odds with those of cancer: *increased* apoptosis, *limited* angiogenesis, ineffective tissue repair, and an intense immune response [111]. Chronic inflammation has been linked to cancer in a number of organs: hepatitis B and C, *H. pylori*, and ulcerative colitis have been shown to lead to cancers of the liver, stomach, and colon, respectively [112]. The chronic inflammation of COPD could similarly be

increasing the risk of lung cancer in these patients [111, 113]. Separately, the airway obstruction defining COPD may force the retention of airborne carcinogens, increasing the likelihood of cancer development. The retention of carcinogens could lead to errors in DNA repair, activation of oncogenes, repression of tumor suppressors, or oncogenic epigenetic alterations [9]. Apoptosis plays a role in both diseases, and at first glance the evasion of apoptosis in cancer appears at odds with the increased apoptosis observed in COPD. On the other hand, increased apoptosis leading to emphysematous destruction could select for a mutation-derived apoptosis resistance, which could increase the population of cancer stem cells in damaged regions of the lung. This could also explain the increased odds of a lung cancer developing in a highly emphysematous region [33].

While several links have been suggested as the key drivers of COPD-associated carcinogenesis, no consensus has yet been reached. The role of proto-oncogene c-Src has been established in lung cancer, and it plays a role in COPD pathogenesis as well. Research showed that the inhibition of c-Src decreased the activation of matrix metalloproteinases-9 and -12 in alveolar macrophages, and also decreased the lung expression of some proteins known to play a role in COPD pathogenesis, namely, cathepsin K, IL-17, TNF-alpha, CCL2, and CXCL1 [114]. Further research has implicated the oxidative stress pathway, inflammation, the epithelial-to-mesenchymal transition (EMT), altered DNA repair, cellular proliferation and NF-kappaB [115]. Muscarinic receptor 3 (M3R) may be important in both non-small-cell lung cancer (NSCLC) and COPD, as one study showed that its expression was increased with tumor

grade and in COPD. Additionally, M3R expression was able to promote migration and invasion in NSCLC cell lines [116]. The miRNA let-7 may provide another molecular link between COPD and lung cancer; let-7 was reduced in the sputum of COPD patients, and was associated with an increase in the expression of soluble tumor necrosis factor receptor 2 (TNFR2), a let-7 target. Let-7 expression was correlated with FEV1 and is a known tumor suppressor in lung cancer [117]. A number of reviews and original articles investigating the potential molecular links between COPD and lung cancer have been published in recent years [113-115, 118-125]. The list of genes and pathways currently implicated in the pathogenic relationship between COPD and lung cancer is quite extensive. Due to the heterogenous nature of each of the two diseases, the relative importance of any single gene or pathway has been difficult to evaluate and so no consensus has emerged.

The goal of the present study is to provide a look at the genome wide connections between COPD and lung cancer that help explain the increased lung cancer risk in COPD. Hereafter, I provide evidence that genes regulated by H3K27Me3 are repressed in both COPD and lung cancer, and that these changes can help explain the inter-disease connection. Several known pulmonary tumor suppressors – SLIT2, WIF1, HPGD, and RAMP2 – are repressed in COPD, with emphysema severity, and in non-small cell lung cancer tumor tissue. These genes are regulated by the epigenetic histone mark, H3K27Me3, which is induced by the methyltransferase EZH2, the protein coding gene of which is up-regulated in the COPD lung. Clinical trials of EZH2 inhibitors for various

cancers are currently underway, and if safe and successful, could prove a disease modifying therapy in COPD and one that could decrease lung cancer risk among this debilitated patient population.

Methods

Microarray data acquisition and normalization

Many groups have profiled the gene expression patterns of lung tumors and adjacent normal lung tissue in an effort to further our understanding of the molecular events driving lung carcinogenesis. I have leveraged the publicly available gene expression data from four such studies [126-129] to connect transcriptomic changes seen in COPD and emphysema to changes seen in lung tumors. Adam Gower, Ph.D. (Boston University), obtained CEL files from the GEO repository for GSE10072, GSE27262, GSE44077, and GSE7670. He used the Robust Multiarray Average (RMA) [38] algorithm with an Entrez gene-specific probeset mapping (17.0.0) from the Molecular and Behavioral Neuroscience Institute (Brainarray) at the University of Michigan to normalize CEL files and produce gene-level expression values.

Several groups have also undertaken genome wide gene expression studies of never, former, and current smokers, with and without COPD, to help elucidate molecular alterations defining the COPD disease state. In this study, I have leveraged data from four of these studies as well [21, 22, 24, 130]. The CEL files for GSE30063, GSE37147, GSE56341, and GSE11784 were obtained and normalized by Adam Gower in the same manner as the four cancer datasets. I also leveraged all ten emphysematous lungs from Aim 1.

Testing COPD and emphysema gene expression signatures in lung tumors

Members of our lab previously described a 98-gene expression signature of COPD in the airway [21] and a 127-gene signature of emphysema severity in lung tissue [28]. In aim 1 of this thesis, I also discovered a 571-gene set associated with emphysema severity across both "usual" and alpha1-antitrypsin deficiency associated emphysema. I hypothesized that these disease-associated transcriptomic alterations would be altered in lung tumors as well as in COPD and emphysema, and that any overlapping changes between the two disease states could help explain the increased risk of lung cancer development observed in patients with COPD. I tested with gene set enrichment analysis (GSEA) [59] whether these three gene sets were enriched with genes that changed in tumors compared to adjacent normal tissue. To determine the relative strength of the enrichments I also tested whether gene sets from MSIGDB were enriched with genes changing between tumor and adjacent normal tissue.

Differential expression analysis in tumors and adjacent normal tissue

Tumor tissue is different from adjacent normal tissue both in terms of the patterns of observable gene expression and in the ratios of the extant cell populations. Therefore, any enrichment of COPD related gene sets observed between tumor and adjacent normal tissue may simply be due to the large scale transcriptomic differences observed between tumors and adjacent normal tissue. To guard against this possibility I derived four gene expression signatures, one for each of the four cancer datasets. I used the *limma* package [131] in R to identify genes differentially expressed between tumors and adjacent normal

tissue with an absolute log fold change greater than or equal to 1.5 and FDR less than 0.01. I then tested with GSEA, whether these tumor-associated signatures were enriched with genes that changed with COPD or with emphysema severity. The enrichment of tumor signatures in COPD lungs would suggest that any enrichment of COPD signatures in tumor data was not a false positive driven by the large scale transcriptomic alterations present in tumors compared to adjacent normal tissue.

Functional enrichment analysis

I used Enrichr [60] to identify functionally enriched biological categories present in the COPD, emphysema, and lung tumor signatures. I limited the search to databases of genes associated with transcription factors, histone modifications, canonical pathways from KEGG, Reactome, Biocarta, and gene ontology categories.

Results

GEO Microarray Datasets

The four lung cancer studies were comprised of samples from 148 individual patients (Table 5). Most but not all patients had paired tumor and adjacent normal tissue available. All tumor data came from studies of non-small cell lung cancer (NSCLC), specifically adenocarcinomas. The tumors profiled were from stages IA to stage IV (Table 5).

	Landi (2008)	Wei (2012)	Kadara (2014)	Su (2007)
	GSE10072	GSE27262	GSE44077	GSE7670
Patients	74	25	23	26
Gender				
Female	27	-	-	21
Male	47	-	-	5
Age	66.43 (7.45)	58.12 (12.63)	-	-
Smoking Status				
Current	28	-	-	-
Former	26	-	-	-
Never	20	-	-	-
Tumor type	Adenocarcinoma	Adenocarcinoma	NSCLC*	Adenocarcinoma
Stage				
IA	8	7	-	-
IB	22	18	-	-
IIA	4	0	-	-
IIB	22	0	-	-
IIIA	11	0	-	-
IIIB	4	0	-	-
IV	3	0	-	-

Table 5. Demographics for four publicly available lung cancer data sets that include paired tumor and adjacent normal data from 148 patients. Not all studies had all demographics included, denoted as "-".

*** NSCLC – not further specified**

The four COPD studies are comprised of 523 never and ever smokers, 153 of whom had COPD (Table 6). One study included 60 never smokers [22] (Table 6). Demographics information for the 10 emphysematous lungs can be found in Aim 1.

	Tilley (2011)	Steiling (2013)	Vucic (2014)	Wang (2012)
	GSE11784	GSE37147	GSE56341	GSE30063
Patients	(n = 94)	(n = 238)	(n = 22)	(n = 169)
COPD				
no	72	151	14	133
yes	22	87	8	36
Gender				
Female	24	103	8	-
Male	70	135	14	-
Age	44.6 (8.7)	64.53 (5.88)	63.95 (4.56)	-
Smoking				
Never	0	0	0	60
Ever	-	-	-	109
Current	94	99	0	-
Former	0	139	22	-

Table 6. Demographics for four publicly available COPD microarray data sets that include airway brushings from 523 never and ever smokers with or without COPD. Not all studies had all demographic information included, denoted as "-".

Enrichment of COPD and emphysema signatures in lung tumors

I tested whether three signatures of COPD and emphysema [21, 28] (and Aim 1) were enriched with genes that changed between lung tumors and adjacent normal tissue. I similarly tested whether 1077 gene sets from MSIGDB (KEGG, Reactome, and Biocarta

canonical pathway gene sets) were enriched, to determine the relative strength of the enrichments of the COPD and emphysema signatures. For these enrichment analyses I broke the three gene sets into sets of genes up or down regulated with disease. I generated one ranked list of genes per cancer dataset and ordered the genes by the t-statistic of the coefficient for tissue (i.e. tumor v adjacent normal). The three sets of genes down regulated in COPD or with emphysema severity were strongly negatively enriched in all four tumor datasets ($q < 0.05$). These down-regulated gene sets were the most negatively enriched among all of the gene sets tested. The three sets of genes up regulated in COPD or with emphysema were positively enriched in two of the four tumor datasets ($q < 0.05$) but were not among the most positively enriched gene sets (Figure 12A). The decreased genes from the emphysema and COPD signatures were strongly negatively enriched and clearly decreased in tumor compared to adjacent normal tissue (Figure 12B: Aim 1 genes projected into tumor data as exemplar [129]).

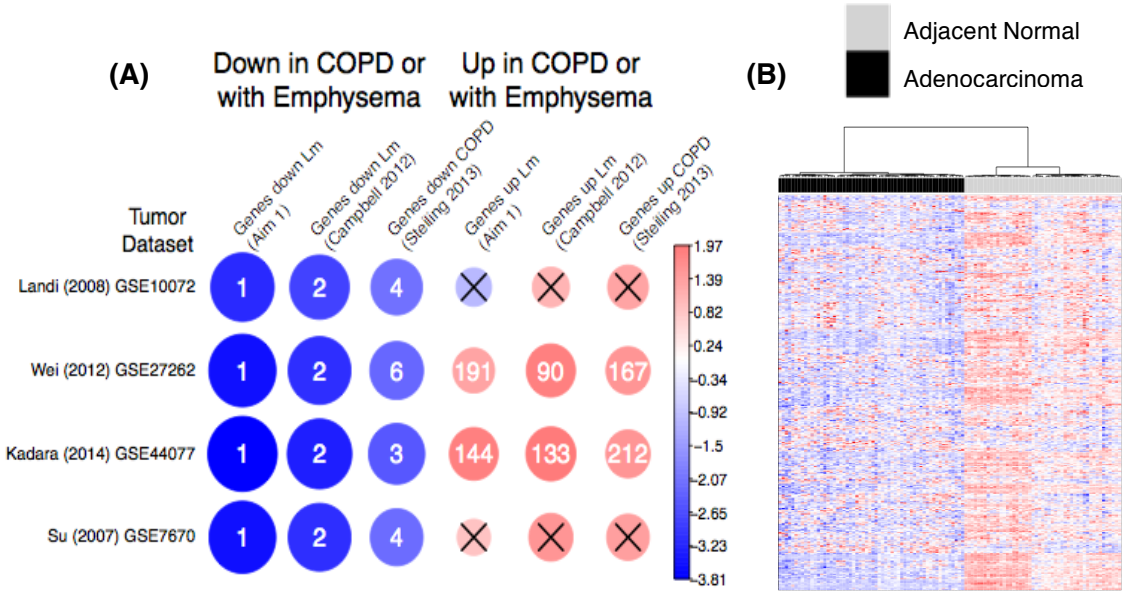

Figure 12. Gene sets decreased with emphysema severity and in COPD are decreased in lung tumor tissue. Our lab has previously published a 127-gene signature of emphysema severity (Campbell et al.) and a 98-gene signature of COPD (Steiling et al.). In Aim 1 of this document I also discovered a 571-gene set associated with emphysema severity across both "usual" and alpha1-antitrypsin deficiency associated emphysema. (A) From these three signatures, the genes that were decreased with emphysema severity or decreased in COPD lungs were consistently negatively enriched in tumors compared to adjacent normal tissue across four independent gene expression profiling experiments (Landi et al., Wei et al., Kadara et al., and Su et al.), by GSEA ($q < 0.05$). In this figure, each row represents a ranked list of genes from a microarray dataset of tumor and adjacent normal tissue. Genes were ranked according to the t-statistic of the coefficient for tissue (i.e. tumor v adjacent normal tissue). Each column represents the up or down regulated portion of each of the three signatures associated with COPD or emphysema. The blue-red color bar, as well as the size of the colored circle, represents the strength of the enrichment of the signature in tumor compared to adjacent normal tissue, i.e. the normalized enrichment score (NES) for each set of genes from GSEA. Blue indicates a negative enrichment score (genes decreased in tumors) and red a positive enrichment score (genes increased in tumors). An "X" denotes that the gene set was not significantly enriched ($q > 0.05$) I also tested whether 1077 gene sets from MSIGDB (KEGG, Reactome, and Biocarta) are enriched in the 4 cancer datasets. The numbers in each circle represent where in the list of negatively or positively enriched gene sets each of the COPD and emphysema signatures fall. E.g. Among gene sets down regulated in Landi's 2008 study, the set of genes down-regulated from Aim 1 is the top-most down-regulated gene set and Campbell's set of down genes is second. (B) In Landi et al.'s tumor and adjacent normal tissue, the expression profiles of the genes decreasing with emphysema severity (Aim 1) are visualized; they display a clear down-regulation of genes decreased with emphysema severity in lung adenocarcinoma tissue compared to adjacent normal tissue.

Identification of differentially expressed genes in tumors compared to adjacent normal tissue

Tumors and adjacent normal tissue are grossly different. In each of the four datasets I identified a set of genes with an absolute log fold change greater than or equal to 1.5 and

FDR < 0.01 (Table 7) and subsequently tested the enrichment of these gene sets in the four COPD studies.

First Author	GEO Access #	Up	Down
Landi	GSE10072	65	131
Wei	GSE27262	264	480
Kadara	GSE44077	143	273
Su	GSE7670	41	155

Table 7. Number of genes with an absolute fold change greater than or equal to 1.5 and FDR < 0.01 in each of the four tumor and adjacent normal studies.

COPD and emphysematous lungs enriched with tumor associated gene expression changes

I tested with GSEA the enrichment of the sets of genes differentially expressed between NSCLCs and adjacent normal tissue (Table 7) in COPD and emphysematous lungs. The sets of genes increased in tumors were consistently enriched with genes increased in the lungs of COPD patients or with genes increasing with emphysema severity ($q < 0.05$ in all five data sets). The sets of genes decreased in tumors were enriched with genes decreased in COPD or with emphysema severity ($q < 0.05$ in four of five data sets) (Figure 13).

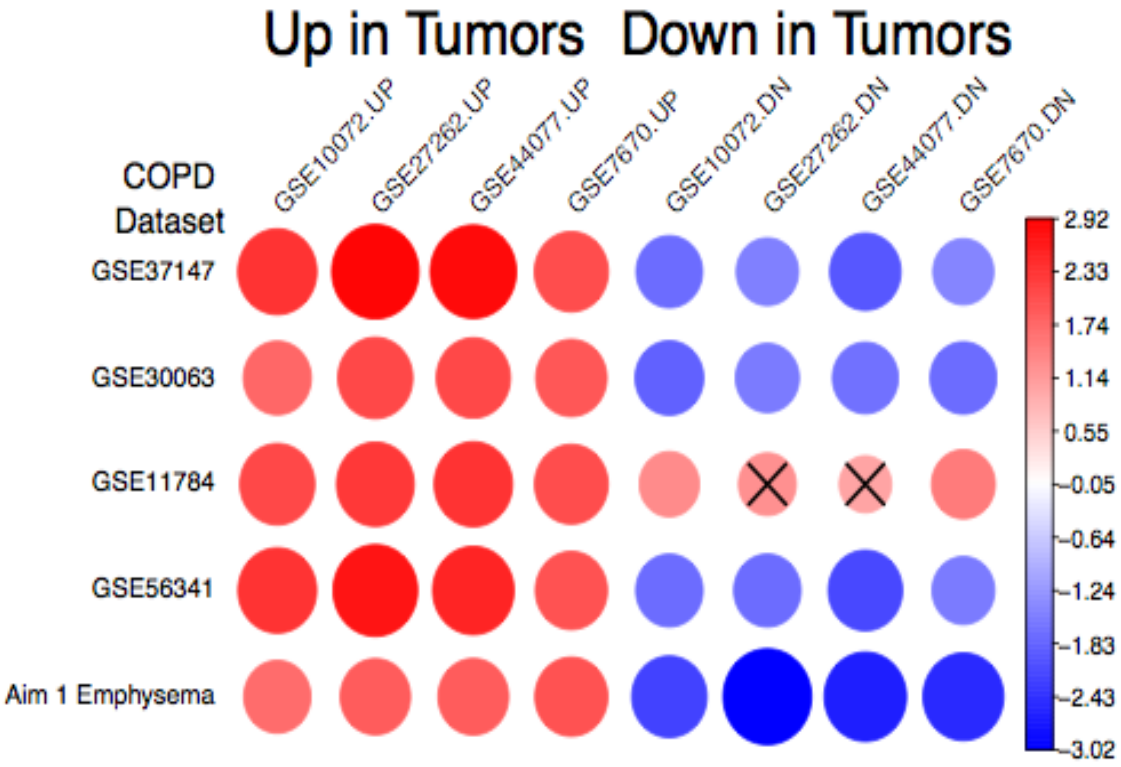

Figure 13. Genes with an absolute fold change greater than 1.5 ($q < 0.01$) in tumor compared to adjacent normal tissue are enriched in COPD and emphysema datasets. Columns 1-4 and columns 5-8 represent sets of genes increased and decreased in tumor compared to adjacent normal tissue, respectively, from GSE10072, GSE27262, GSE44077, and GSE7670. Each row represents a ranked list of genes from a microarray dataset of COPD or emphysema. In the COPD datasets, genes were ranked according to the t-statistic of the coefficient for disease (e.g. smoker with COPD v healthy smoker), and in the emphysema dataset (10 lungs from aim 1), genes were ranked according to the t-statistic of the coefficient for emphysema severity. The blue-red color bar, as well as the size of the colored circle, represents the strength of the enrichment of the sets of tumor-associated genes in the COPD and emphysema datasets, i.e. the normalized enrichment score (NES) for each set of genes from GSEA. Blue indicates a negative enrichment score (genes decreased in COPD or with emphysema) and red a positive enrichment score (genes increased in COPD or with emphysema). An "X" denotes that the gene set was not significantly enriched ($q > 0.05$).

COPD, emphysema, and lung cancer signatures associated with H3K27Me3

Using Enrichr [60], I discovered that the gene sets decreased in COPD, with emphysema severity, or in lung tumors, were all enriched with genes associated with H3K27Me3 and SUZ12 ($q < 0.05$) (Figure 14A). SUZ12, EED, and EZH2 are the three genes that code for the subunits of the polycomb repressor complex 2 (PRC2), which is responsible for

trimethylating the 27th lysine of histone 3 (H3K27Me3), a histone modification responsible for repressing many genes involved in developmental pathways. EZH2 codes for the methyltransferase that actually generates H3K27Me3. Using Fisher's method for combining independent tests baring on the same question, I tested whether each of these three genes were differentially expressed by COPD status within and across the four COPD data sets. EZH2 was significantly up-regulated in three of the four datasets ($p < 0.05$) and was significantly up-regulated when combining all data sets ($p < 0.0001$). Neither SUZ12 nor EED were differentially expressed by COPD status. I similarly tested the three genes between tumors and adjacent normal tissue. Each of the genes was differentially up-regulated in tumors compared to adjacent normal tissue, both in and across the four tumor datasets ($p < 0.001$). (Figure 14B).

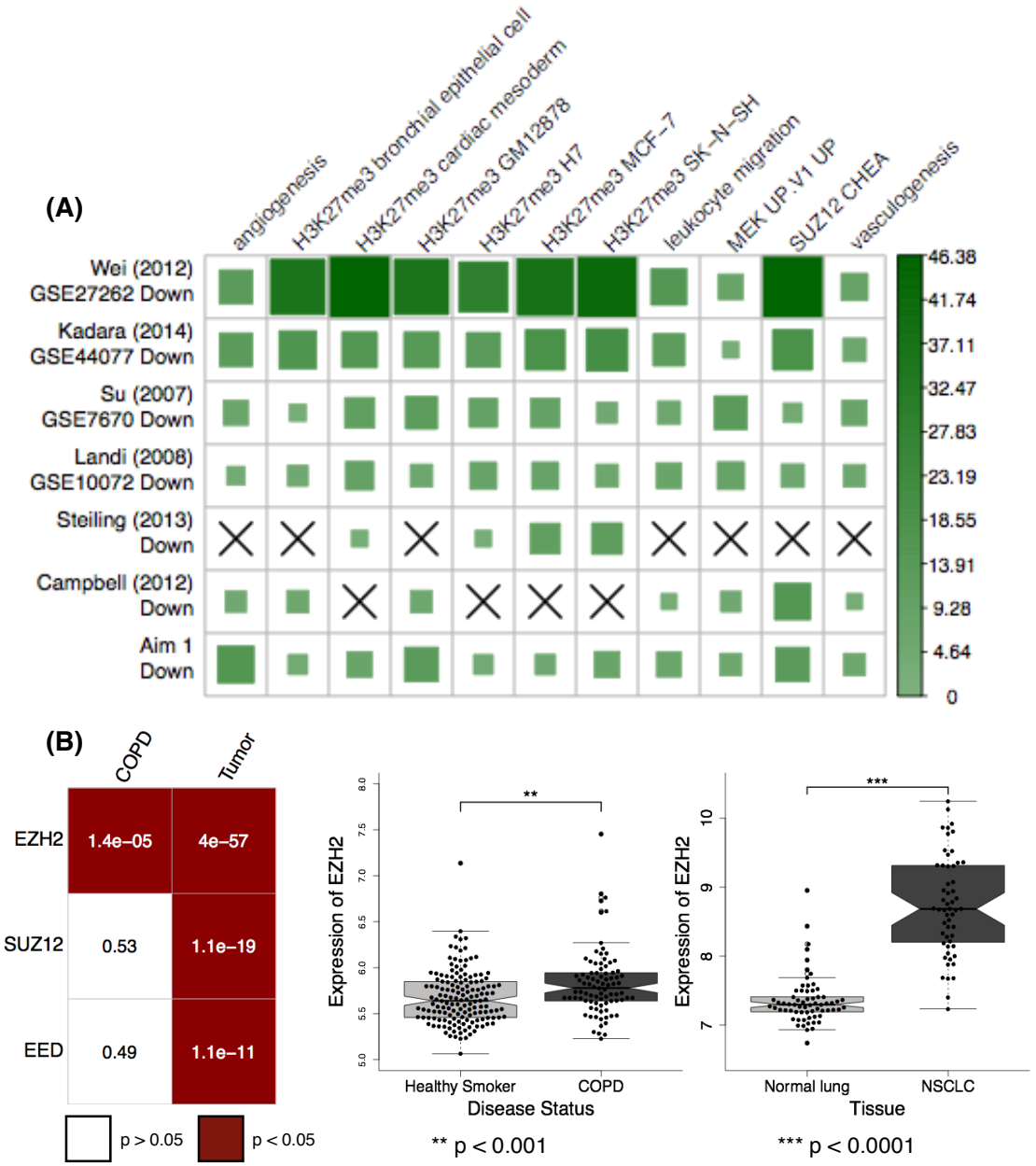

Figure 14. Genes decreased in tumors, in COPD or with emphysema severity are associated with H3K27Me3 and angiogenesis. (A) Enrichment results from each set of genes down-regulated in tumors, in COPD, or with emphysema are overlapping. Here, each row represents one of the disease associated gene sets tested for enrichment in any biological categories (e.g. Aim 1 down genes). Each column represents an enrichment term significantly associated with at least six of the seven tested signatures. The green color and size of each square represents the -log(q) value for the enrichment and each green square represents a significant enrichment result (q < 0.05). Genes regulated by H3K27Me3 are enriched among all of the gene signatures. SUZ12 codes for a subunit of the polycomb repressive complex 2 (PRC2), which is responsible for the trimethylation mark H3K27Me3. (B) The three subunits of PRC2 are EED, SUZ12, and EZH2. Fisher's combined probability test allows meta-analysis of disparate data sources by combining p-values for independent tests baring on the same hypothesis. Combining the results across the cancer datasets shows that each of these three subunits is over-expressed in tumors compared to adjacent normal tissue. EZH2 is also over-expressed in the lungs of COPD patients compared to healthy smokers (B: left most panel). EZH2 expression is increased in COPD lungs compared to healthy smoker lungs in GSE37147 (B: middle panel) and in tumors compared to adjacent normal tissue in GSE44077 (B: right most panel).

The association of the COPD, emphysema, and tumor genes sets with H3K27Me3 was due to 75 genes from these gene sets. These genes were either decreased in tumors and in COPD (e.g. SLIT2, WIF1) or decreased in tumors and with emphysema severity (e.g. EPAS1). No genes were decreased in all gene sets (Figure 15).

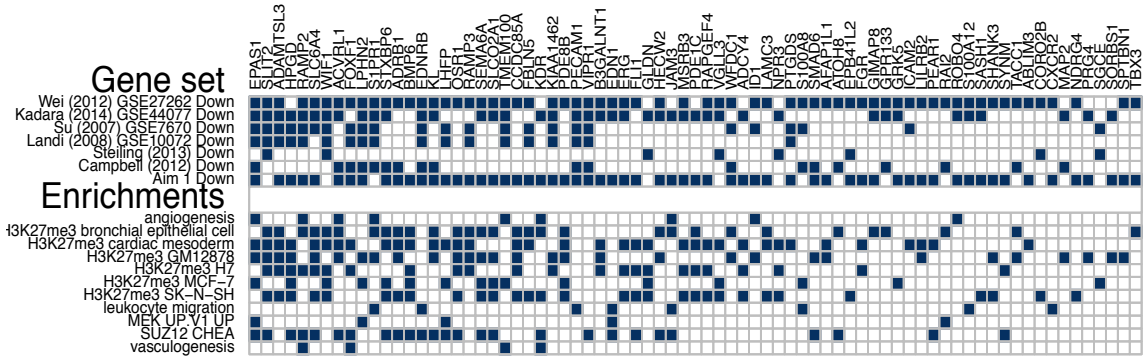

Figure 15. The sets of genes decreased in tumors, COPD, and emphysema are enriched with genes involved in angiogenesis and regulated by H3K27Me3. These enrichments are associated with 75 genes that overlap across gene sets and enrichment terms. Each row represents either a disease gene set or enrichment term. Each column represents a single gene. A blue box means that the gene is a member of the gene set or enrichment term. Genes are sorted such that genes that are members of the most signatures and enrichment terms are to the left, and those that are members of the fewest signatures or enrichment terms are to the right.

Discussion

Patients with chronic obstructive pulmonary disease are at increased risk of developing lung cancer. The goal of this study was to identify shared transcriptomic alterations in lung cancer and COPD that could drive carcinogenesis in COPD. I have identified H3K27Me3 (trimethylation of histone 3 on lysine 27) and EZH2 – the gene coding for the methyltransferase responsible for the histone modification – as a potential epigenetic regulator of COPD and lung cancer, and a shared regulatory event that could drive the increased lung cancer risk observed among COPD patients. I demonstrated that sets of genes decreased in COPD and with increasing emphysema severity were strongly enriched with genes decreased in tumors (Figure 12), even more so than gene sets associated with canonical pathways. Interestingly, the genes decreasing with emphysema severity are at their lowest expression levels in the most damaged regions of emphysematous lungs, and are also decreased in tumors. To an extent then, cells surviving in the damaged regions of the lung are transcriptomically similar to tumors, which is in line with the observation that tumors are more likely to arise in these damaged regions [33].

H3K27Me3 and EZH2

Gene expression signatures of COPD and emphysema were enriched in tumor tissue and gene sets differentially expressed between tumor and adjacent normal tissue were enriched in COPD and emphysematous lung (Figure 13). These observations suggest that these disease states are transcriptomically similar, if pathologically different. Moreover,

all of these interconnected gene sets were enriched with genes regulated by H3K27Me3 and SUZ12 (Figure 14), which is part of the polycomb repressor complex 2 (PRC2) responsible for generating H3K27Me3 [132]. PRC2 is made up of SUZ12, EED, and EZH2, the enzymatic methyltransferase actually responsible for the action of PRC2, and EZH2 was significantly overexpressed in the lungs of COPD patients (Figure 14). EZH2 overexpression has been observed in gastric, ovarian, breast, bladder, and lung cancers and is often associated with a worse prognosis [133-142]. Overexpression of this methyltransferase has been demonstrated to be oncogenic in mice but insufficient to drive carcinogenesis 100% of the time. Approximately 45% of mice with overexpressed pulmonary EZH2 developed lung adenocarcinomas, and long term treatment with an EZH2 inhibitor lead to decreased tumor burden in mice. Nearly all adenocarcinomas from TCGA overexpressed EZH2 compared to normal tissue, and tumors with the highest EZH2 expression were less likely (25% v 43%) to harbor canonical driver mutations in EGFR or KRAS [142], suggesting EZH2 overexpression is oncogenic on its own. EZH2 overexpression also affects response to treatment, as its knockdown enhanced cisplatin sensitivity of cisplatin-resistant cells [141]. A meta-analysis of 10 studies, including 1695 patients, demonstrated that EZH2 overexpression was associated with decreased overall survival in Asian lung cancer patients and in patients with stage I lung adenocarcinoma [143]. That EZH2 was overexpressed in the lungs of COPD patients and that genes regulated by H3K27Me3 were down regulated, suggests that EZH2 overexpression could be one mechanism that increases lung cancer risk in COPD patients.

Among the genes associated with H3K27me3 were SLIT2, WIF1, and HPGD (Figure 15), which are known tumor suppressors in the lung [144-147]. The repression of tumor suppressors is thought to be the primary mechanism by which EZH2 is itself oncogenic. EZH2 has previously been shown to repress SLIT2 in prostate cancer [148] and repress WIF1 in response to bacterial infection [149]. In addition, EPAS1 was among the genes decreased in tumors and with emphysema severity. EPAS1 codes for hypoxia inducible factor 2-alpha (HIF-2a), a suspected tumor suppressor in soft tissue sarcomas [150] and a key regulator of COPD [151]. RAMP2 was among the down regulated genes and has also been implicated as a tumor suppressor in the lung. It is frequently hyper-methylated in lung tumors, its expression is negatively correlated with tumor grade, and ectopic expression of RAMP2 can inhibit lung cancer cell growth and induce apoptosis [152]. Several of the genes overlapping between COPD and lung cancer and associated with H3K27Me3 are tumor suppressors in the lung, suggesting EZH2 overexpression increases lung cancer risk in COPD through H3K27Me3 regulated repression of these tumor suppressors.

Epigenetics in smoking, COPD and lung cancer

Epigenetic modifications are thought to mediate pulmonary effects of smoking and to play a role in the pathogenesis of COPD and lung cancer. The present study lends further credence to this hypothesis. Cigarette smoke induces distinct histone marks on histones 3 and 4 [153] as well as the expression of multiple chromatin modification enzymes,

including methyltransferases and acetyltransferases, histone kinases, and ubiquitinases in human bronchial epithelial cells [154]. Differential methylation sites identified between smokers and non-smokers could also distinguish normal tissue from tumor tissue with a high degree of accuracy, implicating these kinds of methylation changes in both the response to cigarettes and pulmonary carcinogenesis. Some methylation changes were reversible upon smoking cessation, but some changes remained even after 22 years post smoking cessation [155]. This is further reflected in the observation that smoking affects gene expression and that some expression alterations revert to baseline post cessation and others remain irreversibly altered [156]. Furthermore, differentially methylated sites have been discovered between COPD and non-COPD smokers and the affected genes are involved in immune and inflammatory system pathways, the responses to stress and external stimuli, as well as wound healing and coagulation cascades [157]. Aberrant DNA methylation is a genome wide phenomenon in the small airways of COPD patients and regulates genes that are also part of the anti-oxidant NRF2 related pathway [24]. DNA methylation of genes involved in inflammatory pathways and lung development processes were also associated with COPD among an African-American cohort [158]. Lastly, some regulatory T-cell associated immune genes are differentially expressed and methylated in the NSCLC tumors from COPD patients compared to the tumors from non-COPD NSCLC patients [159]. Smoking, COPD, and lung cancer in COPD are all associated with various methylation changes that regulate genes involved in inflammation, protective pathways, and developmental processes and wound healing.

Clinical implications of EZH2 driven carcinogenesis in COPD

EZH2 inhibitors have been shown to decrease tumor burden in mice and to prevent growth among lung cancer cells. A clinical trial for Tazemetostat, an EZH2 inhibitor, is currently underway and testing its efficacy against soft tissue sarcomas (https://clinicaltrials.gov/ct2/show/NCT02601950). If this trial shows that Tazemetostat is effective, its use could be expanded and tested for efficacy against lung cancer in COPD patients. Furthermore, depending on the safety and side effect profile of the drug, it could potentially be given to treat COPD as a disease modifying agent in a manner that could also decrease long term lung cancer risk in this patient population. EZH2 inhibitors are currently an area of intense research interest for treating various cancers [160].

Conclusions

In this study, I demonstrated that there are shared transcriptomic alterations associated with COPD, emphysema, and lung cancer. Several tumor suppressors are among the shared down-regulated genes, namely SLIT2, WIF1, HPGD, and RAMP2. These genes, and the others shared between these disease states, may be regulated by H3K27Me3, the effector of which, EZH2, is increased in the lungs of patients with COPD. As EZH2 is a known oncogene in the lung, EZH2 overexpression in the lungs of COPD patients and the down-regulation of genes repressed by H3K27Me3, is one mechanism that can increase the lung cancer risk among COPD patients. The lungs of patients with COPD and the most damaged emphysematous regions share transcriptomic changes with tumors and may thus be primed for carcinogenesis. If EZH2 inhibitors prove effective and safe

therapeutics for cancer, these kinds of interventions could be used both to potentially modify disease progression and decrease cancer risk in COPD.

Chapter Six: Conclusions, caveats, and future directions

In this dissertation, I leveraged genome-wide gene-expression studies of emphysema and lung cancer to investigate pathogenesis and carcinogenesis in COPD. The results presented in these studies present important

Like "usual" emphysema, AATD associated emphysema is associated with decreased expression of genes critical to the re-epithelialization process, which helps explain why there is deficient repair in emphysematous lungs. While there were no obvious changes between the groups, the unfolded protein response appeared to play a unique role in AATD. The UPR affects disease progression in the livers of AATD patients, but had not previously been demonstrated in the lungs, and has important implications for the pathogenesis of emphysema in this especially susceptible population. The accepted mechanism of emphysema development in these patients has long been that the deficiency of alpha1-antitrypsin causes an imbalance between the proteinases and anti-proteinases of the lung, which leads to excessive and irreversible degradation of the extracellular matrix and alveolar septal walls, i.e. emphysema. Unique activity of the unfolded protein response in the lungs of AATD patients suggests that misfolded AAT, resulting from the SERPINA1 ZZ mutation, accumulates in the cells of the lung and may sensitize those cells to environmental stressors, like cigarette smoke. These observations require further testing. Our collaborators at the University of British Columbia are in the process of extracting proteins from lung cores with little damage and cores with extensive damage to test whether classic markers of the UPR (e.g. PERK) increase with damage in

AATD. These findings can impact the assumptions about mechanisms of emphysema pathogenesis in AATD and may also impact clinical care of these patients, especially because replacement AAT therapy is under investigation. The UPR findings suggest that AAT replacement will be necessary but insufficient to treat this severe form of emphysema because the disease is driven by not just loss of the functional protein, but also by the gain of toxic function that prevents AAT from being secreted.

Emphysema is a progressive disease, and the mean linear intercept (Lm) can serve as a surrogate of progression, but its biological relevance in non-emphysematous lungs had not previously been evaluated nor compared to Lm in emphysema. I wondered whether emphysematous transcriptomic changes would be associated with Lm in controls, and they were, just not in the way I originally anticipated. I discovered that genes that decreased with emphysema severity similarly increased with alveolar spacing (i.e. Lm) in non-diseased control lungs. These genes were related to TGF-beta signaling, endothelial cell development, epithelial cell migration, and response to nutrient levels, and may be part of the larger homeostatic process that maintains the non-diseased lung. Importantly, these findings are based on only five lungs and must be reproduced. As I demonstrated, these lungs were similar to those previously profiled, but this does not mean that these findings will hold true in larger cohorts. Overall, these findings point towards tissue repair processes being active in even the non-insulted lung and that their deficiency in emphysema may drive disease progression.

Finally, COPD and emphysema patients are at increased lung cancer risk, and I discovered that H3K27Me3 is one mechanism that could explain this. This epigenetic methylation mark is known to repress the expression of developmentally necessary genes, including tumor suppressors. Overexpression of the methyltransferase, EZH2, responsible for generating the histone mark, has been demonstrated to be oncogenic and to play a role in multiple cancer types. Genes normally repressed by EZH2 and H3K27Me3 were repressed in COPD and emphysema, and EZH2 was overexpressed in COPD. EZH2 overexpression could therefore be oncogenic in COPD patients and could be a good target for lung cancer interception and possibly disease modification in this debilitated patient population. Critically, whether H3K27Me3 is actually increased in the tissue of COPD patients has not been explored as part of this dissertation and should be investigated before any attempts to modify this epigenetic event are made. The EZH2 inhibitor, Tazemetostat, is currently being investigated as a targeted therapy patients with soft tissue sarcomas with particular EZH2 alterations. If successful, this treatment could be extended into various cancers, including lung cancer. More importantly, if well-tolerated, Tazemetostat could be repurposed to treat COPD patients as a preventive lung cancer measure that might affect COPD and emphysema pathogenesis as well.

Though usual and AATD emphysema share transcriptomic signatures associated with tissue repair, which may be active in the normal homeostatic lung, the UPR changes in AATD emphysema only; successful therapeutic strategies in emphysema will need to account for this difference. In COPD, H3K27Me3 may play a role in both pathogenesis

and carcinogenesis, making it an attractive target for therapeutic interventions, but one that would need further augmentation in AATD.

Journal Abbreviations

AJP.....*American Journal of Physiology*

Am J Pathol.....*American Journal of Pathology*

Am J Respir Cell Mol Biol . American Journal of Respiratory Cell and Molecular Biology

Am J Respir Crit Care Med ..American Journal of Respiratory and Critical Care Medicine

Am Rev Respir Dis.....*American Review of Respiratory Disease*

Ann Intern Med.....*Annals of Internal Medicine*

Annals ATS.....*Annals of the American Thoracic Society*

Arch Intern Med.....*Archives of Internal Medicine*

Biochem J.....*Biochemical Journal*

Biom rep.....*Biomedical Reports*

Cancer Sci.....*Cancer Science*

CELREP.....*Cell Reports*

Cell Res.....*Cell Research*

Clin Sci.....*Clinical Science*

Curr Mol Med.....*Current Molecular Medicine*

Eur Respir J.....*European Respiratory Journal*

Eur J Cancer Care.....*European Journal of Cancer Care*

FEBS Lett.....*Federation of European Biochemical Societies Letters*

Gen Thorac Cardiovasc Surg.....*General Thoracic Cardiovascular Surgery*

IJERPH.....*International Journal of Environmental Research and Public Health*

IJMS.....*International Journal of Molecular Sciences*

J Clin Invest ..... *Journal of Clinical Investigation*

JCO ..... *Journal of Clinical Oncology*

J Thorac Oncol ..... *Journal of Thoracic Oncology*

Nat Cell Biol ..... *Nature Cell Biology*

Nat Rev Cancer ..... *Nature Reviews Cancer*

N Engl J Med ..... *New England Journal of Medicine*

Physiol Genomics ..... *Physiological Genomics*

PLOS ..... *Public Library of Science*

Respir Res ..... *Respiratory Research*

TACG ..... *The Application of Clinical Genetics*

WJCO ..... *World Journal of Clinical Oncology*

WJGPT ..... *World Journal of Gastrointestinal Pharmacology and Therapeutics*

BIBLIOGRAPHY

1. National Center for Health Statistics: **Health, United States, 2015 (4/2016)**. 2016:1–461.

2. National Heart, Lung, and Blood Institute: *Morbidity & Mortality*. 2012:1–116.

3. **Employment and Activity Limitations Among Adults with Chronic Obstructive Pulmonary Disease — United States, 2013**

4. Skillrud DM, Offord KP, Miller RD: **Higher risk of lung cancer in chronic obstructive pulmonary disease. A prospective, matched, controlled study.** *Ann Intern Med* 1986, **105**:503–507.

5. Powell HA, Iyen-Omofoman B, Baldwin DR, Hubbard RB, Tata LJ: **Chronic obstructive pulmonary disease and risk of lung cancer: the importance of smoking and timing of diagnosis.** *J Thorac Oncol* 2013, **8**:6–11.

6. Young RP, Hopkins RJ, Christmas T, Black PN, Metcalf P, Gamble GD: **COPD prevalence is increased in lung cancer, independent of age, sex and smoking history**. *Eur Respir J* 2009, **34**:380–386.

7. Papi A, Casoni G, Caramori G, Guzzinati I, Boschetto P, Ravenna F, Calia N, Petruzzelli S, Corbetta L, Cavallesco G, Forini E, Saetta M, Ciaccia A, Fabbri LM: **COPD increases the risk of squamous histological subtype in smokers who develop non-small cell lung carcinoma.** *Thorax* 2004, **59**:679–681.

8. Kiri VA, Soriano JB, Visick G, Fabbri LM: **Recent trends in lung cancer and its association with COPD: an analysis using the UK GP Research Database**. *Primary Care Respiratory Journal* 2009, **19**:57–61.

9. Takiguchi Y: **Chronic obstructive pulmonary disease as a risk factor for lung cancer**. *WJCO* 2014, **5**:660–8.

10. Jemal A, Ward EM, Johnson CJ, Cronin KA, Ma J, Ryerson AB, Mariotto A, Lake AJ, Wilson R, Sherman RL, Anderson RN, Henley SJ, Kohler BA, Penberthy L, Feuer EJ, Weir HK: **Annual Report to the Nation on the Status of Cancer, 1975–2014, Featuring Survival**. *Journal of the National Cancer Institute* 2017, **109**:1–22.

11. Hiscock R, Bauld L, Amos A, Fidler JA, Munafò M: **Socioeconomic status and smoking: a review**. *Annals of the New York Academy of Sciences* 2011, **1248**:107–123.

12. Jha P, Peto R: **Global Effects of Smoking, of Quitting, and of Taxing Tobacco**. *N Engl J Med* 2014, **370**:60–68.

13. Peters-Golden M, Klinger JR, Carson SS, ATS Research Advocacy Committee: **The case for increased funding for research in pulmonary and critical care.** *Am J Respir Crit Care Med* 2012, **186**:213–215.

14. Lokke A, Lange P, Scharling H, Fabricius P, Vestbo J: **Developing COPD: a 25 year follow up study of the general population**. *Thorax* 2006, **61**:935–939.

15. **Global Strategy for the Diagnosis, Management and Prevention of Chronic Obstructive Lung Disease 2017 Report**. 2017.

16. **What is COPD?** [https://www.nhlbi.nih.gov/health/health-topics/topics/copd]

17. Silverman EK, Miletich JP, Pierce JA, Sherman LA, Endicott SK, Broze GJ, Campbell EJ: **Alpha-1-Antitrypsin Deficiency High prevalence in the St. Louis area**. *Am Rev Respir Dis* 1989, **140**:961–966.

18. Silverman EK, Sandhaus RA: **Clinical practice. Alpha1-antitrypsin deficiency.** *N Engl J Med* 2009, **360**:2749–2757.

19. Brode SK, Ling SC, Chapman KR: **Alpha-1 antitrypsin deficiency: a commonly overlooked cause of lung disease**. *Canadian Medical Association Journal* 2012, **184**:1365–1371.

20. Lieberman J, Winter B, Sastre A: **Alpha 1-antitrypsin Pi-types in 965 COPD patients.** *CHEST* 1986, **89**:370–373.

21. Steiling K, van den Berge M, Hijazi K, Florido R, Campbell J, Liu G, Xiao J, Zhang X, Duclos G, Drizik E, Si H, Perdomo C, Dumont C, Coxson HO, Alekseyev YO, Sin D, Pare P, Hogg JC, McWilliams A, Hiemstra PS, Sterk PJ, Timens W, Chang JT, Sebastiani P, O'Connor GT, Bild AH, Postma DS, Lam S, Spira A, Lenburg ME: **A Dynamic Bronchial Airway Gene Expression Signature of Chronic Obstructive Pulmonary Disease and Lung Function Impairment**. *Am J Respir Crit Care Med* 2013, **187**:933–942.

22. Wang R, Ahmed J, Wang G, Hassan I, Strulovici-Barel Y, Salit J, Mezey JG, Crystal RG: **Airway Epithelial Expression of TLR5 Is Downregulated in Healthy Smokers and Smokers with Chronic Obstructive Pulmonary Disease**. *The Journal of Immunology* 2012, **189**:2217–2225.

23. Tilley AE, Harvey B-G, Heguy A, Hackett NR, Wang R, O'Connor TP, Crystal RG: **Down-regulation of the Notch Pathway in Human Airway Epithelium in Association with Smoking and Chronic Obstructive Pulmonary Disease**. *Am J Respir Crit Care Med* 2009, **179**:457–466.

24. Vucic EA, Chari R, Thu KL, Wilson IM, Cotton AM, Kennett JY, Zhang M,

Lonergan KM, Steiling K, Brown CJ, McWilliams A, Ohtani K, Lenburg ME, Sin DD, Spira A, MacAulay CE, Lam S, Lam WL: **DNA Methylation Is Globally Disrupted and Associated with Expression Changes in Chronic Obstructive Pulmonary Disease Small Airways**. *Am J Respir Cell Mol Biol* 2014, **50**:912–922.

25. Golpon HA, Coldren CD, Zamora MR, Cosgrove GP, Moore MD, Tuder RM, Geraci MW, Voelkel NF: **Emphysema Lung Tissue Gene Expression Profiling**. *Am J Respir Cell Mol Biol* 2004, **31**:595–600.

26. Spira A, Beane J, Pinto-Plata V, Kadar A, Liu G, Shah V, Celli B, Brody JS: **Gene Expression Profiling of Human Lung Tissue from Smokers with Severe Emphysema**. *Am J Respir Cell Mol Biol* 2004, **31**:601–610.

27. Stepaniants S, Wang I-M, Boie Y, Mortimer J, Kennedy B, Elliott M, Hayashi S, Luo H, Wong J, Loy L, Coulter S, Roberts CJ, Hogg JC, Sin DD, O'Neill G, Crackower M, Morris M, Paré PD, Obeidat M: **Genes related to emphysema are enriched for ubiquitination pathways**. *BMC Pulmonary Medicine* 2014, **14**:113–11.

28. Campbell JD, McDonough JE, Zeskind JE, Hackett TL, Pechkovsky DV, Brandsma C-A, Suzuki M, Gosselink JV, Liu G, Alekseyev YO, Xiao J, Zhang X, Hayashi S, Cooper JD, Timens W, Postma DS, Knight DA, Lenburg ME, Hogg JC, Spira A: **A gene expression signature of emphysema- related lung destruction and its reversal by the tripeptide GHK**. *Genome Medicine* 2012, **4**:67.

29. **Non-Small Cell Lung Cancer** [https://www.cancer.org/cancer/non-small-cell-lung-cancer/causes-risks-prevention/risk-factors.html]

30. Ebbert JO, Yang P, Vachon CM, Vierkant RA, Cerhan JR, Folsom AR, Sellers TA: **Lung Cancer Risk Reduction After Smoking Cessation: Observations From a Prospective Cohort of Women**. *JCO* 2003, **21**:921–926.

31. Wang H, Yang L, Zou L, Huang D, Guo Y, Pan M, Tan Y, Zhong H, Ji W, Ran P, Zhong N, Lu J: **Association between Chronic Obstructive Pulmonary Disease and Lung Cancer: A Case-Control Study in Southern Chinese and a Meta-Analysis**. *PLoS ONE* 2012, **7**:e46144–9.

32. Brenner DR, McLaughlin JR, Hung RJ: **Previous Lung Diseases and Lung Cancer Risk: A Systematic Review and Meta-Analysis**. *PLoS ONE* 2011, **6**:e17479–10.

33. Hohberger LA, Schroeder DR, Bartholmai BJ, Yang P, Wendt CH, Bitterman PB, Larsson O, Limper AH: **Correlation of Regional Emphysema and Lung Cancer: A Lung Tissue Research Consortium-Based Study**. *Journal of Thoracic Oncology* 2014, **9**:639–645.

34. MarÃ a Torres-DurÃ n MD P, PhD AR-R, MD IP-L, MD JA-A, Virginia Leiro-

FernÃndez MD P, Carmen Montero-MartÃ nez MD P, MD CP, MD OC-AÃN, MD AG-GM, Francisco J GonzÃ lez-Barcala MD P, Cristina MartÃ nez MD P, MD RGN-T, Mariano Provencio MD P, MD MAJM-M, Alberto FernÃndez-Villar MD P, Juan Miguel Barros-Dios MD P: **Alpha-1 Antitrypsin Deficiency and Lung Cancer Risk**. *Journal of Thoracic Oncology* 2015, **10**:1279–1284.

35. Yang P, Sun Z, Krowka MJ, Aubry M-C, Bamlet WR, Wampfler JA, Thibodeau SN, Katzmann JA, Allen MS, Midthun DE, Marks RS, de Andrade M: **Alpha1-antitrypsin deficiency carriers, tobacco smoke, chronic obstructive pulmonary disease, and lung cancer risk.** *Arch Intern Med* 2008, **168**:1097–1103.

36. Topic AS, Jelic-Ivanovic ZD, Spasojevic-Kalimanovska VV, Spasic SM: **Association of Moderate Alpha-1 Antitrypsin Deficiency with Lung Cancer in the Serbian Population**. *Archives of Medical Research* 2006, **37**:866–870.

37. Zhou JJ, Cho MH, Castaldi PJ, Hersh CP, Silverman EK, Laird NM: **Heritability of Chronic Obstructive Pulmonary Disease and Related Phenotypes in Smokers**. *Am J Respir Crit Care Med* 2013, **188**:941–947.

38. Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, Speed TP: **Exploration, normalization, and summaries of high density oligonucleotide array probe level data.** *Biostatistics* 2003, **4**:249–264.

39. Brettschneider J, Collin F, Bolstad BM, Speed TP: **Quality Assessment for Short Oligonucleotide Microarray Data**. *Technometrics* 2008, **50**:241–264.

40. Winkler C, Atochina-Vasserman EN, Holz O, Beers MF, Erpenbeck VJ, Krug N, Roepcke S, Lauer G, Elmlinger M, Hohlfeld JM: **Comprehensive characterisation of pulmonary and serum surfactant protein D in COPD.** *Respir Res* 2011, **12**:29.

41. Ito E, Oka R, Ishii T, Korekane H, Kurimoto A, Kizuka Y, Kitazume S, Ariki S, Takahashi M, Kuroki Y, Kida K, Taniguchi N: **Fucosylated surfactant protein-D is a biomarker candidate for the development of chronic obstructive pulmonary disease**. *Journal of Proteomics* 2015, **127**(PB):386–394.

42. Johansson SL, Tan Q, Holst R, Christiansen L, Hansen NCG, Hojland AT, Wulf-Johansson H, Schlosser A, Titlestad IL, Vestbo J, Holmskov U, Kyvik KO, Sorensen GL: **Surfactant protein D is a candidate biomarker for subclinical tobacco smoke-induced lung damage**. *AJP: Lung Cellular and Molecular Physiology* 2014, **306**:L887–L895.

43. Lomas DA, Silverman EK, Edwards LD, Locantore NW, Miller BE, Horstman DH, Tal-Singer R, on behalf of the Evaluation of COPD Longitudinally to Identify Predictive Surrogate Endpoints study investigators: **Serum surfactant protein D is steroid sensitive and associated with exacerbations of COPD**. *Eur Respir J* 2009, **34**:95–102.

44. Vandivier RW, Ogden CA, Fadok VA, Hoffmann PR, Brown KK, Botto M, Walport MJ, Fisher JH, Henson PM, Greene KE: **Role of Surfactant Proteins A, D, and C1q in the Clearance of Apoptotic Cells In Vivo and In Vitro: Calreticulin and CD91 as a Common Collectin Receptor Complex**. *The Journal of Immunology* 2002, **169**:3978–3986.

45. Greene CM: **Z α-1 antitrypsin deficiency and the endoplasmic reticulum stress response**. *WJGPT* 2010, **1**:94–8.

46. Blanco I, Bueno P, Diego I, Pérez-Holanda S, Casas-Maldonado F, Esquinas C, Miravitlles M: **Alpha-1 antitrypsin Pi*Z gene frequency and Pi*ZZ genotype numbers worldwide: an update.** *COPD* 2017, **12**:561–569.

47. Perlmutter DH, Brodsky JL, Balistreri WF, Trapnell BC: **Molecular pathogenesis of alpha-1-antitrypsin deficiency-associated liver disease: A meeting review**. *Hepatology* 2007, **45**:1313–1323.

48. Eriksson S, Carlson J, Velez R: **Risk of Cirrhosis and primary liver cancer in alpha**. *N Engl J Med* 1986, **314**:736–739.

49. Mulgrew AT, Taggart CC, Lawless MW, Greene CM, Brantly ML, O'Neill SJ, McElvaney NG: **Z α1-Antitrypsin Polymerizes in the Lung and Acts as a Neutrophil Chemoattractant**. *CHEST* 2004, **125**:1952–1957.

50. Greene CM, Miller SDW, Carroll TP, Oglesby IK, Ahmed F, O'Mahony M, Taggart CC, McElvaney NG, O'Neill SJ: **Anti-apoptotic effects of Z 1-antitrypsin in human bronchial epithelial cells**. *Eur Respir J* 2010, **35**:1155–1163.

51. Pini L, Tiberio L, Venkatesan N, Bezzi M, Corda L, Luisetti M, Ferrarotti I, Malerba M, Lomas DA, Janciauskiene S, Vizzardi E, Modina D, Schiaffonati L, Tantucci C: **The role of bronchial epithelial cells in the pathogenesis of COPD in Z-alpha-1 antitrypsin deficiency**. *Respir Res* 2014, **15**:1–10.

52. Venembre P, Boutten A, Seta N, Dehoux MS, Crestani B, Aubier M, Durand G: **Secretion of alpha 1-antitrypsin by alveolar epithelial cells.** *FEBS Lett* 1994, **346**:171–174.

53. **The Role of Neutrophils in Alpha-1 Antitrypsin Deficiency.** *Annals ATS* 2016, **13 Suppl 4**:S297–304.

54. Ordóñez A, Snapp EL, Tan L, Miranda E, Marciniak SJ, Lomas DA: **Endoplasmic reticulum polymers impair luminal protein mobility and sensitize to cellular stress in alpha 1-antitrypsin deficiency**. *Hepatology* 2013, **57**:2049–2060.

55. Bhattacharya S, Srisuma S, DeMeo DL, Shapiro SD, Bueno R, Silverman EK, Reilly

JJ, Mariani TJ: **Molecular Biomarkers for Quantitative and Discrete COPD Phenotypes**. *Am J Respir Cell Mol Biol* 2009, **40**:359–367.

56. Francis SMS, Larsen JE, Pavey SJ, Bowman RV, Hayward NK, Fong KM, Yang IA: **Expression profiling identifies genes involved in emphysema severity**. *Respir Res* 2009, **10**:613–10.

57. Wang I-M, Stepaniants S, Boie Y, Mortimer JR, Kennedy B, Elliott M, Hayashi S, Loy L, Coulter S, Cervino S, Harris J, Thornton M, Raubertas R, Roberts C, Hogg JC, Crackower M, O'Neill G, Paré PD: **Gene Expression Profiling in Patients with Chronic Obstructive Pulmonary Disease and Lung Cancer**. *Am J Respir Crit Care Med* 2008, **177**:402–411.

58. Ning W, Li C-J, Kaminski N, Feghali-Bostwick CA, Alber SM, Di YP, Otterbein SL, Song R, Hayashi S, Zhou Z, Pinsky DJ, Watkins SC, Pilewski JM, Sciurba FC, Peters DG, Hogg JC, Choi AMK: **Comprehensive gene expression profiles reveal pathways related to the pathogenesis of chronic obstructive pulmonary disease.** *Proceedings of the National Academy of Sciences* 2004, **101**:14895–14900.

59. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP: **Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles.** *Proceedings of the National Academy of Sciences* 2005, **102**:15545–15550.

60. Chen EY, Tan CM, Kou Y, Duan Q, Wang Z, Meirelles GV, Clark NR, Ma'ayan A: **Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool.** *BMC Bioinformatics* 2013, **14**:128.

61. Shoulders MD, Ryno LM, Genereux JC, Moresco JJ, Tu PG, Wu C, Yates JR III, Su AI, Kelly JW, Wiseman RL: **Stress-Independent Activation of XBP1s and/or ATF6 Reveals Three Functionally Diverse ER Proteostasis Environments**. *CELREP* 2013, **3**:1279–1292.

62. Dombroski BA, Nayak RR, Ewens KG, Ankener W, Cheung VG, Spielman RS: **Gene Expression and Genetic Variation in Response to Endoplasmic Reticulum Stress in Human Cells**. *The American Journal of Human Genetics* 2010, **86**:719–729.

63. Vadivel A, van Haaften T, Alphonse RS, Rey-Parra G-J, Ionescu L, Haromy A, Eaton F, Michelakis E, Thébaud B: **Critical role of the axonal guidance cue EphrinB2 in lung growth, angiogenesis, and repair.** *Am J Respir Crit Care Med* 2012, **185**:564–574.

64. Crosby LM, Waters CM: **Epithelial repair mechanisms in the lung**. *AJP: Lung Cellular and Molecular Physiology* 2010, **298**:L715–L731.

65. Zien Alaabden A, Mohammad Y, Fahoum S: **The role of serum surfactant protein**

D as a biomarker of exacerbation of chronic obstructive pulmonary disease. *Qatar Medical Journal* 2015, **2015**:18–8.

66. Puchelle E, Zahm JM, Tournier JM, Coraux C: **Airway Epithelial Repair, Regeneration, and Remodeling after Injury in Chronic Obstructive Pulmonary Disease**. *Proceedings of the American Thoracic Society* 2006, **3**:726–733.

67. Sood R, Porter AC, Ma K, Quilliam LA, Wek RC: **Pancreatic eukaryotic initiation factor-2alpha kinase (PEK) homologues in humans, Drosophila melanogaster and Caenorhabditis elegans that mediate translational control in response to endoplasmic reticulum stress.** *Biochem J* 2000, **346 Pt 2**:281–293.

68. Harding HP, Zhang Y, Ron D: **Protein translation and folding are coupled by an endoplasmic-reticulum-resident kinase.** *Nature* 1999, **397**:271–274.

69. Shi Y, Vattem KM, Sood R, An J, Liang J, Stramm L, Wek RC: **Identification and characterization of pancreatic eukaryotic initiation factor 2 alpha-subunit kinase, PEK, involved in translational control.** *Molecular and Cellular Biology* 1998, **18**:7499–7509.

70. Yoshida H, Okada T, Haze K, Yanagi H, Yura T, Negishi M, Mori K: **ATF6 activated by proteolysis binds in the presence of NF-Y (CBF) directly to the cis-acting element responsible for the mammalian unfolded protein response.** *Molecular and Cellular Biology* 2000, **20**:6755–6767.

71. Haze K, Yoshida H, Yanagi H, Yura T, Mori K: **Mammalian transcription factor ATF6 is synthesized as a transmembrane protein and activated by proteolysis in response to endoplasmic reticulum stress.** *Molecular Biology of the Cell* 1999, **10**:3787–3799.

72. Lee A-H, Iwakoshi NN, Glimcher LH: **XBP-1 regulates a subset of endoplasmic reticulum resident chaperone genes in the unfolded protein response.** *Molecular and Cellular Biology* 2003, **23**:7448–7459.

73. Calfon M, Zeng H, Urano F, Till JH, Hubbard SR: **IRE1 couples endoplasmic reticulum load to secretory capacity by processing the XBP-1 mRNA**. *Nature* 2002, **420**:202–202.

74. Dey S, Sayers CM, Verginadis II, Lehman SL, Cheng Y, Cerniglia GJ, Tuttle SW, Feldman MD, Zhang PJL, Fuchs SY, Diehl JA, Koumenis C: **ATF4-dependent induction of heme oxygenase 1 prevents anoikis and promotes metastasis**. *J Clin Invest* 2015, **125**:2592–2608.

75. Rzymski T, Milani M, Pike L, Buffa F, Mellor HR, Winchester L, Pires I, Hammond E, Ragoussis I, Harris AL: **Regulation of autophagy by ATF4 in response to severe**

hypoxia. *Oncogene* 2010, **29**:4424–4435.

76. Tabas I, Ron D: **Integrating the mechanisms of apoptosis induced by endoplasmic reticulum stress**. *Nat Cell Biol* 2011, **13**:184–190.

77. Murray CJ, Lopez AD: **Alternative projections of mortality and disability by cause 1990–2020: Global Burden of Disease Study**. *The Lancet* 1997, **349**:1498–1504.

78. Heguy A, Harvey B-G, Leopold PL, Dolgalev I, Raman T, Crystal RG: **Responses of the human airway epithelium transcriptome to in vivo injury**. *Physiol Genomics* 2007, **29**:139–148.

79. John G, Kohse K, Orasche J, Reda A, Schnelle-Kreis J, Zimmermann R, Schmid O, Eickelberg O, Yildirim AÖ: **The composition of cigarette smoke determines inflammatory cell recruitment to the lung in COPD mouse models.** *Clin Sci* 2014, **126**:207–221.

80. Rutgers SR, Postma DS, Hacken ten NH, Kauffman HF, van Der Mark TW, Koëter GH, Timens W: **Ongoing airway inflammation in patients with COPD who do not currently smoke.** *Thorax* 2000, **55**:12–18.

81. Shapiro SD, Goldstein NM, Houghton AM, Kobayashi DK, Kelley D, Belaaouaj A: **Neutrophil elastase contributes to cigarette smoke-induced emphysema in mice.** *Am J Pathol* 2003, **163**:2329–2335.

82. Valavanidis A, Vlachogianni T, Fiotakis K: **Tobacco Smoke: Involvement of Reactive Oxygen Species and Stable Free Radicals in Mechanisms of Oxidative Damage, Carcinogenesis and Synergistic Effects with Other Respirable Particles**. *IJERPH* 2009, **6**:445–462.

83. Huang MF, Lin WL, Ma YC: **A study of reactive oxygen species in mainstream of cigarette.** *Indoor Air* 2005, **15**:135–140.

84. Pierrou S, Broberg P, O'Donnell RA, Pawłowski K, Virtala R, Lindqvist E, Richter A, Wilson SJ, Angco G, Möller S, Bergstrand H, Koopmann W, Wieslander E, Strömstedt P-E, Holgate ST, Davies DE, Lund J, Djukanovic R: **Expression of Genes Involved in Oxidative Stress Responses in Airway Epithelial Cells of Smokers with Chronic Obstructive Pulmonary Disease**. *Am J Respir Crit Care Med* 2007, **175**:577–586.

85. Sasaki H: **Polymorphisms of NRF2 gene correlated with decreased FEV1 in lung cancers of smokers**. *biom rep* 2013:1–5.

86. Kelsen SG: **The Unfolded Protein Response in Chronic Obstructive Pulmonary Disease.** *Annals ATS* 2016, **13 Suppl 2**:S138–45.

87. Ribeiro CMP, O'Neal WK: **Endoplasmic reticulum stress in chronic obstructive lung diseases.** *Curr Mol Med* 2012, **12**:872–882.

88. Togo S, Holz O, Liu X, Sugiura H, Kamio K, Wang X, Kawasaki S, Ahn Y, Fredriksson K, Skold CM, Mueller KC, Branscheid D, Welker L, Watz H, Magnussen H, Rennard SI: **Lung Fibroblast Repair Functions in Patients with Chronic Obstructive Pulmonary Disease Are Altered by Multiple Mechanisms**. *Am J Respir Crit Care Med* 2008, **178**:248–260.

89. Königshoff M KNEO: **TGF-β signalling in COPD: deciphering genetic and cellular susceptibilities for future therapeutic regimens**. 2009:1–10.

90. Hobbs BD, Hersh CP: **Integrative genomics of chronic obstructive pulmonary disease**. *Biochemical and Biophysical Research Communications* 2014, **452**:276–286.

91. Hersh CP, Hansel NN, Barnes KC, Lomas DA, Pillai SG, Coxson HO, Mathias RA, Rafaels NM, Wise RA, Connett JE, Klanderman BJ, Jacobson FL, Gill R, Litonjua AA, Sparrow D, Reilly JJ, Silverman EK, the ICGN Investigators: **Transforming Growth Factor-β Receptor-3 Is Associated with Pulmonary Emphysema**. *Am J Respir Cell Mol Biol* 2009, **41**:324–331.

92. Xing Y, Li A, Borok Z, Li C, Minoo P: **NOTCH1 Is Required for Regeneration of Clara Cells During Repair of Airway Injury**. *STEM CELLS* 2012, **30**:946–955.

93. Tsao P-N, Matsuoka C, Wei S-C, Sato A, Sato S, Hasegawa K, Chen H-K, Ling T-Y, Mori M, Cardoso WV, Morimoto M: **Epithelial Notch signaling regulates lung alveolar morphogenesis and airway epithelial integrity**. *Proceedings of the National Academy of Sciences* 2016, **113**:8242–8247.

94. Chambers RC, Mercer PF: **Mechanisms of Alveolar Epithelial Injury, Repair, and Fibrosis**. *Annals ATS* 2015, **12**:S16–S20.

95. Perotin J-M, Adam D, Vella-Boucaud J, Delepine G, Sandu S, Jonvel A-C, Prevost A, Berthiot GR, Pison C, Lebargy FO, Birembaut P, Coraux C, Deslee GT: **Delay of airway epithelial wound repair in COPD is associated with airflow obstruction severity**. *Respir Res* 2014, **15**:1–9.

96. Beers MF, Morrisey EE: **The three R's of lung health and disease: repair, remodeling, and regeneration**. *J Clin Invest* 2011, **121**:2065–2073.

97. Hogan BLM, Barkauskas CE, Chapman HA, Epstein JA, Jain R, Hsia CCW, Niklason L, Calle E, Le A, Randell SH, Rock J, Snitow M, Krummel M, Stripp BR, Vu T, White ES, Whitsett JA, Morrisey EE: **Repair and regeneration of the respiratory system: complexity, plasticity, and mechanisms of lung stem cell function.** *Cell Stem Cell* 2014, **15**:123–138.

98. Kaku Y, Imaoka H, Morimatsu Y, Komohara Y, Ohnishi K, Oda H, Takenaka S, Matsuoka M, Kawayama T, Takeya M, Hoshino T: **Overexpression of CD163, CD204 and CD206 on Alveolar Macrophages in the Lungs of Patients with Severe Chronic Obstructive Pulmonary Disease**. *PLoS ONE* 2014, **9**:e87400–8.

99. Ohar JA, Hamilton RF Jr, Zheng S, Sadeghnejad A, Sterling DA, Xu J, Meyers DA, Bleecker ER, Holian A: **COPD Is Associated With a Macrophage Scavenger Receptor-1 Gene Sequence Variation**. *CHEST* 2010, **137**:1098–1107.

100. Herold S: **Acute lung injury: how macrophages orchestrate resolution of inflammation and tissue repair**. 2011:1–13.

101. Gordon S, Martinez FO: **Alternative Activation of Macrophages: Mechanism and Functions**. *Immunity* 2010, **32**:593–604.

102. Mosser DM, Edwards JP: **Exploring the full spectrum of macrophage activation**. *Nature Publishing Group* 2008, **8**:958–969.

103. Bartram U, Speer CP: **The Role of Transforming Growth Factor β in Lung Development and Disease**. *CHEST* 2004, **125**:754–765.

104. Wu X, Yuan B, López E, Bai C, Wang X: **Gene polymorphisms and chronic obstructive pulmonary disease**. *Journal of Cellular and Molecular Medicine* 2013, **18**:15–26.

105. Xu J, Lamouille S, Derynck R: **TGF-β-induced epithelial to mesenchymal transition**. *Cell Res* 2009, **19**:156–172.

106. Liu T, Feng X-H: **Regulation of TGF-β signalling by protein phosphatases**. *Biochem J* 2010, **430**:191–198.

107. Guo S, DiPietro LA: **Factors Affecting Wound Healing**. *Journal of Dental Research* 2010, **89**:219–229.

108. Eming SA, Krieg T, Davidson JM: **Inflammation in Wound Repair: Molecular and Cellular Mechanisms**. *Journal of Investigative Dermatology* 2007, **127**:514–525.

109. Eming SA, Martin P, Tomic-Canic M: **Wound repair and regeneration: mechanisms, signaling, and translation.** *Science Translational Medicine* 2014, **6**:265sr6–265sr6.

110. Sohal SS, Soltani A, Reid D, Ward C, Wills K, Muller H, Walters E: **A randomized controlled trial of inhaled corticosteroids (ICS) on markers of epithelial–mesenchymal transition (EMT) in large airway samples in COPD: an exploratory proof of concept study**. *COPD* 2014:533–10.

111. Brody JS: **State of the Art. Chronic Obstructive Pulmonary Disease, Inflammation, and Lung Cancer**. *Proceedings of the American Thoracic Society* 2006, **3**:535–537.

112. Karnold: **Feeding the Flame: New Research Adds to Role of Inflammation in Cancer Development**. 2005:1–3.

113. Houghton AM: **Mechanistic links between COPD and lung cancer**. *Nat Rev Cancer* 2013, **13**:1–13.

114. Geraghty P, Hardigan A, Foronjy RF: **Cigarette Smoke Activates the Proto-Oncogene c-Src to Promote Airway Inflammation and Lung Tissue Destruction**. *Am J Respir Cell Mol Biol* 2014, **50**:559–570.

115. Pastor M, Nogal A, Molina-Pinelo S, Meléndez R, Romero-Romero B, Mediano M, López-Campos J, García-Carbonero R, Sanchez-Gastaldo A, Carnero A, Paz-Ares L: **Identification of Oxidative Stress Related Proteins as Biomarkers for Lung Cancer and Chronic Obstructive Pulmonary Disease in Bronchoalveolar Lavage**. *IJMS* 2013, **14**:3440–3455.

116. Lin G, Sun L, Wang R, Guo Y, Xie C: **Overexpression of muscarinic receptor 3 promotes metastasis and predicts poor prognosis in non-small-cell lung cancer.** *J Thorac Oncol* 2014, **9**:170–178.

117. Shin Il J, Brusselle GG: **Mechanistic links between COPD and lung cancer: a role of microRNA let-7?** *Nature Publishing Group* 2014, **14**:1–1.

118. Goldkorn T, Filosto S, Chung S: **Lung Injury and Lung Cancer Caused by Cigarette Smoke-Induced Oxidative Stress: Molecular Mechanisms and Therapeutic Opportunities Involving the Ceramide-Generating Machinery and Epidermal Growth Factor Receptor**. *Antioxidants & Redox Signaling* 2014, **21**:2149–2174.

119. Sekine Y, Hata A, Koh E, Hiroshima K: **Lung carcinogenesis from chronic obstructive pulmonary disease: characteristics of lung cancer from COPD and contribution of signal transducers and lung stem cells in the inflammatory microenvironment**. *Gen Thorac Cardiovasc Surg* 2014, **62**:415–421.

120. Rooney C, Sethi T: **The Epithelial Cell and Lung Cancer: The Link between Chronic Obstructive Pulmonary Disease and Lung Cancer**. *Respiration* 2011, **81**:89–104.

121. Vaguliene N, Zemaitis M, Lavinskiene S, Miliauskas S, Sakalauskas R: **Local and systemic neutrophilic inflammation in patients with lung cancer and chronic obstructive pulmonary disease**. *BMC Immunology* 2013, **14**:1–1.

122. YOUNG RP, HOPKINS RJ: **How the genetics of lung cancer may overlap with COPD**. *Respirology* 2011, **16**:1047–1055.

123. Young R: **Genetic evidence linking lung cancer and COPD: a new perspective**. *TACG* 2011:99–13.

124. El-Zein RA, Young RP, Hopkins RJ, Etzel CJ: **Genetic Predisposition to Chronic Obstructive Pulmonary Disease and/or Lung Cancer: Important Considerations When Evaluating Risk**. *Cancer Prevention Research* 2012, **5**:522–527.

125. Adcock IM, Caramori G, Barnes PJ: **Chronic Obstructive Pulmonary Disease and Lung Cancer: New Molecular Insights**. *Respiration* 2011, **81**:265–284.

126. Wei T-YW, Juan C-C, Hisa J-Y, Su L-J, Lee Y-CG, Chou H-Y, Chen J-MM, Wu Y-C, Chiu S-C, Hsu C-P, Liu K-L, Yu C-TR: **Protein arginine methyltransferase 5 is a potential oncoprotein that upregulates G1 cyclins/cyclin-dependent kinases and the phosphoinositide 3-kinase/AKT signaling cascade**. *Cancer Sci* 2012, **103**:1640–1650.

127. Su L-J, Chang C-W, Wu Y-C, Chen K-C, Lin C-J, Liang S-C, Lin C-H, Whang-Peng J, Hsu S-L, Chen C-H, Huang C-YF: **Selection of DDX5 as a novel internal control for Q-RT-PCR from microarray data using a block bootstrap re-sampling scheme**. *BMC Genomics* 2007, **8**:140–12.

128. Kadara H, Fujimoto J, Yoo SY, Maki Y, Gower AC, Kabbout M, Garcia MM, Chow CW, Chu Z, Mendoza G, Shen L, Kalhor N, Hong WK, Moran C, Wang J, Spira A, Coombes KR, Wistuba II: **Transcriptomic Architecture of the Adjacent Airway Field Cancerization in Non-Small Cell Lung Cancer**. *JNCI Journal of the National Cancer Institute* 2014, **106**:dju004–dju004.

129. Landi MT, Dracheva T, Rotunno M, Figueroa JD, Liu H, Dasgupta A, Mann FE, Fukuoka J, Hames M, Bergen AW, Murphy SE, Yang P, Pesatori AC, Consonni D, Bertazzi PA, Wacholder S, Shih JH, Caporaso NE, Jen J: **Gene Expression Signature of Cigarette Smoking and Its Role in Lung Adenocarcinoma Development and Survival**. *PLoS ONE* 2008, **3**:e1651–8.

130. Tilley AE, O'Connor TP, Hackett NR, Strulovici-Barel Y, Salit J, Amoroso N, Zhou XK, Raman T, Omberg L, Clark A, Mezey J, Crystal RG: **Biologic Phenotyping of the Human Small Airway Epithelial Response to Cigarette Smoking**. *PLoS ONE* 2011, **6**:e22798–11.

131. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK: **limma powers differential expression analyses for RNA-sequencing and microarray studies**. *Nucleic Acids Research* 2015, **43**:e47–e47.

132. Margueron R, Reinberg D: **The Polycomb complex PRC2 and its mark in life**.

Nature 2011, **469**:343–349.

133. He L-J, Cai M-Y, Xu G-L, Li J-J, Weng Z-J, Xu D-Z, Luo G-Y, Zhu S-L, Xie D: **Prognostic significance of overexpression of EZH2 and H3k27me3 proteins in gastric cancer.** *Asian Pacific Journal of Cancer Prevention* 2012, **13**:3173–3178.

134. Hu S, Yu L, Li Z, Shen Y, Wang J, Cai J, Xiao L, Wang Z: **Overexpression of EZH2 contributes to acquired cisplatin resistance in ovarian cancer cells in vitro and in vivo.** *Cancer Biology & Therapy* 2010, **10**:788–795.

135. Fujii S, Tokita K, Wada N, Ito K, Yamauchi C, Ito Y, Ochiai A: **MEK-ERK pathway regulates EZH2 overexpression in association with aggressive breast cancer subtypes.** *Oncogene* 2011, **30**:4118–4128.

136. ZHANG YB, NIU HT, CHANG JW, DONG GL, MA XB: **EZH2 silencing by RNA interference inhibits proliferation in bladder cancer cell lines.** *Eur J Cancer Care (Engl)* 2011, **20**:106–112.

137. Takashina T, Kinoshita I, Kikuchi J, Shimizu Y, Sakakibara-Konishi J, Oizumi S, Nishimura M, Dosaka-Akita H: **Combined inhibition of EZH2 and histone deacetylases as a potential epigenetic therapy for non-small-cell lung cancer cells**. *Cancer Science* 2016, **107**:955–962.

138. LaFave LM, Béguelin W, Koche R, Teater M, Spitzer B, Chramiec A, Papalexi E, Keller MD, Hricik T, Konstantinoff K, Micol J-B, Durham B, Knutson SK, Campbell JE, Blum G, Shi X, Doud EH, Krivtsov AV, Chung YR, Khodos I, de Stanchina E, Ouerfelli O, Adusumilli PS, Thomas PM, Kelleher NL, Luo M, Keilhack H, Abdel-Wahab O, Melnick A, Armstrong SA, et al.: **Loss of BAP1 function leads to EZH2-dependent transformation**. *Nature Medicine* 2015, **21**:1344–1349.

139. Zhang Y, Yu X, Chen L, Zhang Z, Feng S: **EZH2 overexpression is associated with poor prognosis in patients with glioma**. *Oncotarget* 2016.

140. Wang Y, Chen Y, Geng H, Qi C, Liu Y, Yue D: **Overexpression of YB1 and EZH2 are associated with cancer metastasis and poor prognosis in renal cell carcinomas.** *Tumour Biology* 2015, **36**:7159–7166.

141. Zhang J: **The expression and significance of the enhancer of zeste homolog 2 in lung adenocarcinoma**. *Oncologyl Reports* 2012:1–8.

142. Zhang H, Qi J, Reyes JM, Li L, Rao PK, Li F, Lin CY, Perry JA, Lawlor MA, Federation A, De Raedt T, Li YY, Liu Y, Duarte MA, Zhang Y, Herter-Sprie GS, Kikuchi E, Carretero J, Perou CM, Reibel JB, Paulk J, Bronson RT, Watanabe H, Brainson CF, Kim CF, Hammerman PS, Brown M, Cichowski K, Long H, Bradner JE, et al.: **Oncogenic Deregulation of EZH2 as an Opportunity for Targeted Therapy in**

Lung Cancer. *Cancer Discovery* 2016, **6**:1006–1021.

143. Wang X, Zhao H, Lv L, Bao L, Wang X, Han S: **Prognostic Significance of EZH2 Expression in Non-Small Cell Lung Cancer: A Meta-analysis**. *Science Reports* 2015:1–6.

144. Tseng RC, Lee SH, Hsu HS, Chen BH, Tsai WC, Tzao C, Wang YC: **SLIT2 Attenuation during Lung Cancer Progression Deregulates -Catenin and E-Cadherin and Associates with Poor Prognosis**. *Cancer Research* 2010, **70**:543–551.

145. Kim J, You L, Xu Z, Kuchenbecker K, Raz D, He B, Jablons D: **Wnt inhibitory factor inhibits lung cancer cell growth**. *The Journal of Thoracic and Cardiovascular Surgery* 2007, **133**:733–737.

146. Huang G, Eisenberg R, Yan M, Monti S, Lawrence E, Fu P, Walbroehl J, Lowenberg E, Golub T, Merchan J, Tenen DG, Markowitz SD, Halmos B: **15-Hydroxyprostaglandin Dehydrogenase is a Target of Hepatocyte Nuclear Factor 3 and a Tumor Suppressor in Lung Cancer**. *Cancer Research* 2008, **68**:5040–5048.

147. Ding Y: **NAD+-linked 15-hydroxyprostaglandin dehydrogenase (15-PGDH) behaves as a tumor suppressor in lung cancer**. *Carcinogenesis* 2004, **26**:65–72.

148. Yu J, Cao Q, Yu J, Wu L, Dallol A, Li J, Chen G, Grasso C, Cao X, Lonigro RJ, Varambally S, Mehra R, Palanisamy N, Wu JY, Latif F, Chinnaiyan AM: **The neuronal repellent SLIT2 is a target for repression by EZH2 in prostate cancer**. *Oncogene* 2010, **29**:5370–5380.

149. Roy BC, Subramaniam D, Ahmed I, Jala VR, Hester CM, Greiner KA, Haribabu B, Anant S, Umar S: **Role of bacterial infection in the epigenetic regulation of Wnt antagonist WIF1 by PRC2 protein EZH2**. 2014, **34**:4519–4530.

150. Nakazawa MS, Eisinger-Mathason TSK, Sadri N, Ochocki JD, Gade TPF, Amin RK, Simon MC: **Epigenetic re-expression of HIF-2α suppresses soft tissue sarcoma growth.** *Nature Communications* 2016, **7**:10539.

151. Yoo S, Takikawa S, Geraghty P, Argmann C, Campbell J, Lin L, Huang T, Tu Z, Feronjy R, Spira A, Schadt EE, Powell CA, Zhu J: **Integrative Analysis of DNA Methylation and Gene Expression Data Identifies EPAS1 as a Key Regulator of COPD**. *PLoS Genetics* 2015, **11**:e1004898–16.

152. Yue W, Dacic S, Sun Q, Landreneau R, Guo M, Zhou W, Siegfried JM, Yu J, Zhang L: **Frequent Inactivation of RAMP2, EFEMP1 and Dutt1 in Lung Cancer by Promoter Hypermethylation**. *Clinical Cancer Research* 2007, **13**:4336–4344.

153. Sundar IK, Nevid MZ, Friedman AE, Rahman I: **Cigarette Smoke Induces Distinct**

Histone Modifications in Lung Cells: Implications for the Pathogenesis of COPD and Lung Cancer. *Journal of Proteome Research* 2014, **13**:982–996.

154. Sundar IK, Rahman I: **Gene expression profiling of epigenetic chromatin modification enzymes and histone marks by cigarette smoke: implications for COPD and lung cancer.** *Am J Physiol Lung Cell Mol Physiol* 2016, **311**:L1245–L1258.

155. Ambatipudi S, Cuenin C, Hernandez-Vargas H, Ghantous A, Le Calvez-Kelm F, Kaaks R, Barrdahl M, Boeing H, Aleksandrova K, Trichopoulou A, Lagiou P, Naska A, Palli D, Krogh V, Polidoro S, Tumino R, Panico S, Bueno-de-Mesquita B, Peeters PH, Quirós JR, Navarro C, Ardanaz E, Dorronsoro M, Key T, Vineis P, Murphy N, Riboli E, Romieu I, Herceg Z: **Tobacco smoking-associated genome-wide DNA methylation changes in the EPIC study**. *Epigenomics* 2016, **8**:599–618.

156. Beane J, Sebastiani P, Liu G, Brody JS, Lenburg ME, Spira A: **Reversible and permanent effects of tobacco smoke exposure on airway epithelial gene expression.** *Genome Biology* 2007, **8**:R201.

157. Qiu W, Baccarelli A, Carey VJ, Boutaoui N, Bacherman H, Klanderman B, Rennard S, Agusti A, Anderson W, Lomas DA, DeMeo DL: **Variable DNA Methylation Is Associated with Chronic Obstructive Pulmonary Disease and Lung Function**. *Am J Respir Crit Care Med* 2012, **185**:373–381.

158. Busch R, Qiu W, Lasky-Su J, Morrow J, Criner G, DeMeo D: **Differential DNA methylation marks and gene comethylation of COPD in African- Americans with COPD exacerbations**. *Respir Res* 2016:1–15.

159. Wauters E, Janssens W, Vansteenkiste J, Decaluwé H, Heulens N, Thienpont B, Zhao H, Smeets D, Sagaert X, Coolen J, Decramer M, Liston A, De Leyn P, Moisse M, Lambrechts D: **DNA methylation profiling of non-small cell lung cancer reveals a COPD-driven immune-related signature.** *Thorax* 2015, **70**:1113–1122.

160. Morera L, Lübbert M, Jung M: **Targeting histone methyltransferases and demethylases in clinical trials for cancer therapy**. *Clinical Epigenetics* 2016:1–16.

CURRICULUM VITAE