Candidate-Wide Genetic Association Study of Emphysema Using the MESA-Lung Cohort and the CARe Resource

Elizabeth Oelsner, July 2009

Research question: is there a significant association between candidate genes from the CARe resource—including approximately 2,000 candidate genes, predominantly chosen for cardiovascular phenotypes—and a quantitative emphysema phenotype among the subjects in multiethnic MESA-Lung cohort?

Study Rationale:

Background: COPD and emphysema comprise a major public health challenge. Together, they are projected to overtake stroke as the 3rd leading cause of death in the United States by 2020[1]. Mortality for COPD was 120,000 in 2000, which represented a 67% increase since 1980. The greatest increases in death rates were seen among traditionally under-studied subgroups: mortality from COPD increased 87% among blacks and tripled among women[2]. Meanwhile, treatment and prevention strategies remain elusive due to a relative paucity of information regarding underlying mechanisms. To date, medical therapies may improve symptoms, but do not affect progressive decline in lung function. The only therapies proven to decrease mortality remain smoking cessation and supplemental oxygen. For emphysema, lung reduction surgery and A1AT supplementation may hold promise, but only for a small subset of patients.

Definitions: Emphysema is traditionally seen as a subtype of COPD. According to the GOLD definition, COPD is defined *physiologically* by the presence of airflow limitation that is not fully reversible[3]. By contrast, emphysema is defined *anatomically* as an abnormal, permanent enlargement of airspaces distal to the terminal bronchioles with destruction of walls but without obvious fibrosis. COPD and emphysema are often considered to coexist, although recent studies have shown only a modest correlation (r 0.4) between emphysema as defined by CT scan and low lung function. Meanwhile, there is relatively little correlation between emphysema and the alternative COPD subtype, chronic bronchitis, which is defined *symptomatically* as a productive cough for 3 or more months in 2 or more years.

Emphysema and cardiovascular disease: This study will use the CARe resource, which includes some candidate genes selected for lung disease, but is predominantly comprised of candidate genes for cardiovascular phenotypes[12]. The cardiovascular candidate genes appear promising since significant overlap exists between COPD/emphysema and cardiovascular diseases. COPD increases the risk of cardiovascular disease by two- to three-fold[4], cardiovascular disease is major cause of death among COPD patients, and low lung function has been shown to be an independent predictor of ischemic cardiovascular disease[1]. The factors responsible for this interrelationship remain unknown, however it has been posited that inflammatory, endothelial, and other microvascular factors may underlie both disorders[4-5, 11].

Unexplained variability in COPD/emphysema: Although smoking is the most important risk factor, it is neither necessary nor sufficient for development of COPD or emphysema. Only 20% of smokers develop COPD[6]. Other factors believed to be associated with COPD (pollution, infection, occupational exposure, hyperresponsive airways, A1AT) account for a small number of cases[7]. Emphysema can be found in nonsmokers upon autopsy[8] and chest CT (preliminary analyses of the MESA-Lung). This unexplained variability may suggest a genetic risk for COPD and emphysema.

Genetic variation in risk for COPD/emphysema: Studies indicate that lung function and COPD are heritable[6], but they are under only modest genetic influence. SERPINA2 has been associated with COPD, however the gene has unknown function, and there have been no other consistently replicated loci. SERPINA1 is a known cause of A1AT deficiency and resultant severe emphysema, however this mutation accounts for a very small proportion of COPD and emphysema cases. Nonetheless, a recent study of emphysema, as assessed by chest CT, demonstrated familial clustering of percent emphysema that was independent of smoking. Furthermore, the clustering of the emphysema phenotype was noted to be stronger than that of lung function, suggesting that quantitative emphysema phenotypes may be better candidates for genetic association analysis versus COPD/bronchitis phenotypes[9].

Limitations of prior studies and current aims: Previous studies have been limited by phenotypic heterogeneity; case-control design; restriction to smokers; and use of predominantly Caucasian populations. To our knowledge, this study is novel in that it will examine a robust, quantitative phenotype for emphysema with respect to cardiovascular candidate genes, using a population of both smokers and nonsmokers in a large multiethnic cohort. The aim of the study is to identify significant gene associations for emphysema, which may consequently elucidate disease pathways, set therapeutic targets, and assist in risk prediction and prevention programs.

Study design:

Study subjects: MESA-Lung is a subset of the original prospective MESA cohort, which was recruited in 2000-2002. The individuals were 45-84 years old and free of symptomatic cardiovascular disease at baseline. Subjects of multiple ethnicities were recruited at each of six study centers. 3,965 subjects from the original cohort were subsequently enrolled in MESA-Lung, and have CT-based lung data obtained from cardiac CT.

Candidate genes: 2,100 candidate genes were selected for NHLBI's CARe (Candidate Gene Association Resource) project, yielding a dataset of approximately 55,000 SNPs[12]. The CARe resource encompases nine NHLBI cohort studies including MESA, ARIC, and the Framingham Heart Study. The majority of CARe candidate genes are hypothesized or known to relate to cardiovascular phenotypes, although some candidates were also chosen specifically for lung, blood, and sleep phenotypes.

Phenotype: Percent emphysema is a validated, quantitative phenotype generated from analysis of lung CT[14]. Of note, in MESA-Lung, percent emphysema was estimated from the lung fields obtained upon cardiac CT scans, rather than via full lung scans. However, MESA-Lung investigators have found a high correlation between percent emphysema as calculated by cardiac CT versus full lung scan (0.99 for diffuse emphysema, 0.91 for apical emphysema). Scans were repeated for each subject, and 10% of CT scans were re-read for quality control; results were reproducible (ICC 0.93).

Statistical methods: The association analysis between percent emphysema and 55,000 SNPs will be performed by standard linear modeling using PLINK software. Analyses will be adjusted for confounders such as age, sex, height, and weight, CT scan parameters, and ancestry (principal components analysis). Effect modification by smoking will be explored in additional analyses using data regarding smoking status and pack-year history. Analyses will be repeated restricted to ever-smoking and never-smoking samples.

Power calculations: Using the Bonferroni correction for multiple testing, the alpha for this analysis will be set at 3x10-7. Power was calculated using the QUANTO program assuming a continuous variable (mean 25% empysema, SD 10) and approximately 4,000 independent subjects. Power estimates are provided based on a range of effect sizes and minor allele frequencies.

Power calculation, MESA-Lung cohort
N=4000

Allele freq	0.05	0.10	0.25
Effect size			
1	0.0008	0.0074	0.1062
2	0.1110	0.5998	**0.9960**
3	0.7695	**0.9985**	**0.9999**
4	**0.9966**	**0.9999**	**0.9999**
5	**0.9999**	**0.9999**	**0.9999**

As shown above, the study has excellent power (assuming beta = 0.8) to detect an effect size of 4% of SD or greater. It is underpowered for effect sizes of less than 2%. Power clearly improves with minor allele frequency, yielding variable power for effect sizes of 2-3%.

Power calculations were repeated for N=2000 to approximate power in a sample stratified by smoking status (approximately 50% of MESA-Lung were ever-smokers at baseline). Clearly, power is significantly diminished, however it remains excellent for effects sizes of 4-5%.

Power calculation, smokers only, nonsmokers only
N=2000

Allele freq	0.05	0.10	0.25
Effect size			
1	0.0001	0.0006	0.0086
2	0.0008	0.0931	0.6460
3	0.1637	0.7230	**0.9992**
4	0.6596	**0.9942**	**0.9999**
5	**0.9649**	**0.9999**	**0.9999**

Limitations: The study demonstrates several strengths versus prior studies, most notably a large multiethnic cohort and a validated quantitative phenotype. Nonetheless, certain limitations must be acknowledged. One potential weakness may be the use of partial CT scans versus full lung CT for quantitation of percent emphysema. Nonetheless, as noted above, the correlation between readings for partial and full lung CT has been found to be high in the MESA-Lung sample. Another concern regards the treatment of smoking history, which is known to be the most important risk factor for development of COPD/emphysema. Effect modification will be assessed, yet models stratified for smoking may be underpowered given smaller sample sizes. Of note, the risk of misspecification and thus confounding of smoking status is believed to be minimal based on multiple measures of smoking status which were found to be concordant with objective cotinine levels (discordance <2%), however this risk cannot be eliminated entirely. As discussed, the study is also underpowered to detect effect sizes less than 2% of the standard deviation. Nonetheless, the use of a quantitative phenotype and large cohort yields considerable power at moderate to large effect sizes. Lastly, use of a multiethnic cohort is both a strength and a weakness, given that stratification by ancestry further compromises statistical power. However,

previous work using the MESA cohort suggests that principal components analysis can yield reliable estimates in the overall cohort.

References:

1. Huiart L, Ernst P, Suissa S. Cardiovascular Morbidity and Mortality in COPD. Chest. October 2005; 128 (4): 2640-2646.
2. Mannino DM, Homa DM, Akinbami LJ, Ford ES, Redd SC. COPD surveillance – US, 1971-2000. In: Surveillance Summaries, August 2, 2002. MMWR 2002; 51: 1-8.
3. Pauwels RA, Buist AS, Calverley PMA, et al. Global Strategy for the diagnosis, management, and prevention of COPD. NHLBI/WHO Global Initiative for COPD, Am J Crit Care Med 2001; 163:1256-76.
4. Sin DD, Man SF. Why are patients with chronic obstructive pulmonary disease at increased risk of cardiovascular diseases? The potential role of systemic inflammation in chronic obstructive pulmonary disease. Circulation. 2003 Mar 25;107(11):1514-9.
5. Cosio MG, Saetta M, Agusti A. Immunologic Aspects of COPD. NEJM. 2009 June; 360(23): 2445-2454.
6. Wilk JB, Chen T, Gottlieb DJ, Walter RE, Nagle MW, Brandler BJ, Myers RH, Borecki IB, Silverman EK, Weiss ST, O'Connor GT. A Genome-Wide Association Study of Pulmonary Function Measures in the Framingham Heart Study. PloS Genetics 2009; 5:e1000429.
7. Snider G. COPD: Risk Factors, Pathophysiology, and Pathogenesis. Annual Review Medicine 1989; 40: 411-429.
8. Anderson AE Jr, Hernandez JA, Eckert P, Foraker AG. Emphysema in lung macrosections correlated with smoking habits. Science 1964; 144:1025-1026.
9. Patel BD, Coxson HO, Pillai SG, et al. Airway wall thickening and emphysema show independent familial aggregation in chronic obstructive pulmonary disease. Am J Respir Crit Care Med. 2008 Sep 1;178(5):500-5.
10. Molfino, NA. Genetics of COPD. Chest. 2004 May;125(5):1929-40.
11. Ning W, Li CJ, Kaminski N, et al. Comprehensive gene expression profiles reveal pathways related to the pathogenesis of chronic obstructive pulmonary disease. Proc Natl Acad Sci U S A. 2004 Oct 12;101(41):14895-900.
12. NHLBI CARe Center. http://www.broadinstitute.org/gen_analysis/care/index.php/Main_Page. Accessed July 16 2009.
13. MESA-Lung. www.mesa-nhlbi.org. Accessed July 16 2009.
14. Hoffman EA, Jiang R, Baumhauer HA et al. Reproducibility and validation of lung density measures from cardiac CT scans in the MESA Lung Study. Acad Radiology 2009; in press.