ONLINE SUPPLEMENTAL MATERIALS
Association between Acute Respiratory Disease Events and the *MUC5B* Promoter Polymorphism in Smokers – Online Supplement

Samuel Y. Ash, MD, MPH (1); Rola Harmouche, PhD (2); Rachel K. Putman MD, MPH (1); James C. Ross, PhD (2); Fernando J. Martinez, MD, MS (3); Augustine M. Choi, MD (3); Russell P. Bowler, MD, PhD (4); Elizabeth A. Regan, MD, PhD (5); Jeffrey L. Curtis, MD (6); MeiLan K. Han, MD, MS (7); Richard C. Boucher, MD (8); Wanda O'Neal, PhD (8); Hiroto Hatabu, MD, PhD (9); David A. Lynch, MD (10); Ivan O. Rosas, MD (1); Gary M. Hunninghake, MD, MPH (1); Raul San Jose Estepar, PhD (2) and George R. Washko, MD (1) for the COPDGene Investigators

(1) Division of Pulmonary and Critical Care Medicine, Department of Medicine, Brigham and Women's Hospital, 75 Francis St, Boston, MA; (2) Laboratory of Mathematics in Imaging, Department of Radiology, Brigham and Women's Hospital, 75 Francis St, Boston, MA; (3) Department of Medicine, Weil Cornell Medical College, 425 East 61st St, New York, NY; (4) Division of Pulmonary, Critical Care and Sleep Medicine, Department of Medicine, National Jewish Health, 1400 Jackson St, Denver, CO; (5) Division of Rheumatology, Department of Medicine, National Jewish Health, 1400 Jackson St, Denver, CO; (6) Division of Pulmonary and Critical Care Medicine, Department of Internal Medicine, University of Michigan and VA Ann Arbor Healthcare System, 2215 Fuller Road, Ann Arbor, MI; (7) Division of Pulmonary and Critical Care Medicine, Department of Internal Medicine, University of Michigan, 1500 E Medical Center Dr, Ann Arbor, MI; (8) Marisco Lung Institute, University of North Carolina, 7008 Marisco Hall, Chapel Hill, NC; (9) Department of Radiology, Brigham and Women's Hospital, 75 Francis St, Boston, MA; (10) Department of Radiology, National Jewish Health, 1400 Jackson St, Denver, CO.

SUPPLEMENTAL METHODS

Cohort

COPDGene is an ongoing longitudinal observational study of 10,300 individuals, predominately ever smokers with or without COPD, plus never-smoker control individuals.[1] Participants with active lung diseases other than COPD and asthma were excluded; all participants underwent baseline testing, including an extensive in-person interview, volumetric high -resolution CT scan of the chest, and spirometry. Prospective data are collected through the Longitudinal Follow-up (LFU) portion of the study. As part of the LFU, participants are contacted using email, an automated telephone system or a manual telephone system either two or four times per year and asked to report on intercurrent hospitalizations, emergency room visits and other medical events such as ARD events.[2]

Our study was restricted to ever-smokers without or with COPD who had available complete imaging, genetic, clinical (including all covariates) and longitudinal data. We utilized the August 31, 2016 version of the baseline COPDGene dataset for all of the clinical, imaging and spirometric data, and the March 31, 2017 version of the LFU COPDGene dataset for the ARD event data.

Objective CT Analysis

The automated method of parenchymal characterization and CT analysis has been described previously.[3] [4] Briefly, two pulmonologists (SYA and GRW) trained the classification tool by placing 33,865 fiducials identifying radiologic tissue subtypes in the chest CT scans of 138 randomly-selected individuals. These training points were used to develop tissue classification vectors for each disease subtype; next, de novo regions of lung were classified based on their similarity to these vectors. This method assigned labels

[normal, emphysematous (centrilobular and paraseptal),[1] or interstitial (reticular, honeycombing, centrilobular nodule, linear scar, nodular, subpleural line, and ground glass)] to every portion of the lung parenchyma. The radiologic tissue-type volumes were then expressed as a percent of total lung volume (i.e., percent normal, percent emphysema, and percent interstitial). We defined participants as having interstitial or emphysematous features if those features occupied >10% of their lung volume. We selected this threshold based on previous work identifying visually-defined interstitial lung abnormalities and emphysema using this and other objective methods, and because it resulted in a prevalence rate of interstitial features roughly similar to, albeit somewhat higher than, that described in visual analyses of smokers.[4 5] We also performed sensitivity analyses using a 5% threshold for each feature type. Participants who had both interstitial features and emphysematous features were included in both the interstitial and emphysema analyses.

Genetic and Clinical Analyses

Genotyping of the *MUC5B* promoter polymorphism (rs35705950) was performed using TaqMan Genotyping Assays (Applied Biosystems). Due to our sample size, evidence that *MUC5B* expression in the lung tissue of patients with interstitial disease does not vary between homozygotes and heterozygotes, and for ease of exposition, participants were considered to have the polymorphism if they had at least one copy of the minor allele (genotypes GT and TT).[6 7] Secondary analyses of the primary outcomes assuming an additive effect of the number of copies of the minor allele were also performed.

We defined ARD events as symptomatic respiratory episodes requiring a change in treatment including antibiotics, steroids, emergency room visit or hospitalization.[8] ARD events were assessed prospectively as part of the COPDGene long term follow up cohort as described above, and participants were considered to have had an ARD event if they reported any during the course of follow-up.[8] We also performed secondary analyses of severe ARD events, which were defined as those events requiring either an emergency room visit or hospitalization.[8]

Statistical Analyses

Summary statistics are given as the mean and standard deviation where appropriate. We used multivariable linear regression to assess the relationship between specific radiologic feature types and the *MUC5B* promoter polymorphism. Multivariable logistic, Cox and negative binomial regression were used to analyze the odds of having an ARD event, the time-to-first ARD event, and the number of ARD events during follow-up respectively. To account for varying durations of follow-up, a scale parameter was used in the negative binomial regression analyses. We formally tested for the presence of an interaction between the presence versus absence of the promoter polymorphism and the percentage of lung occupied by each parenchymal feature type (expressed as a continuous measure) using multivariable logistic regression. In order to assess for the specific interaction between each type of parenchymal features and genotype, we performed secondary analyses of each feature type compared to those with primarily "normal" parenchyma. For instance, we evaluated the interaction between interstitial features and genotype in those without emphysema, and we evaluated the interaction between emphysema and genotype in those without interstitial features. For the secondary analyses of severe ARD event, multivariable logistic regression was used to analyze the odds of reporting a severe ARD event during follow-up.

[1] Note that panlobular emphysema was not identified in the training cases likely because patients with alpha 1 antitrypsin disease were not represented in the cohort.

We adjusted multivariable analyses for the values of the following covariates at the time of the baseline interview: age, sex, body-mass index, current smoking status, pack-years of smoking, percent predicted forced expiratory volume in one second, a reported history of gastroesophageal reflux disease, the Saint George's Respiratory Questionnaire total score, clinical center, and a reported history of an ARD in the year before study enrollment.[8] We tested associations between the MUC5B promoter polymorphism and ARD events in the entire cohort and in subgroups of individuals with each feature type (interstitial and emphysematous features), based on the thresholds described above. Similarly, we analyzed the associations between each feature type and ARD events in the entire cohort and in subgroups with and without the MUC5B promoter polymorphism. Due to racial differences in the prevalence of the MUC5B promoter polymorphism, all analyses were stratified by race and the primary results given are for non-Hispanic whites (NHW) as this is the population in which the minor allele is found at an appreciable frequency.[9] Secondary analyses of the entire cohort and of African-Americans are also shown. For the Cox regression analyses, all covariates were evaluated using the Schoenfeld residuals method and none were found to violate the proportional hazards assumption.[10] To evaluate for the presence of bias due to conditioning on a collider we performed secondary analyses of the primary outcomes using models only adjusted for factors not likely to be related to genotype: age, sex, smoking status, pack year and clinical center.[11] Where given, confidence intervals (CI) are 95% CI; p values are two-sided and those <0.05 were considered to indicate statistical significance. Because of the preliminary nature of these analyses, no adjustments were made for multiple testing. All analyses were performed using SAS 9.4 (Cary, NC).

SUPPLEMENTAL RESULTS

Complete imaging, clinical (including all covariates) and longitudinal data were available for analysis on 6863 participants; they were a middle-aged group with significant representation of women, African-Americans, and current-smokers (**Table E1**). Over the total duration of follow-up (mean ± SD, 5.6±2.3 years), at least one ARD event was reported by 40.4% of participants, with 2.1±4.3 ARD events per participant. At least one copy of the minor allele of the *MUC5B* promoter polymorphism (GT or TT) was present in 15% of individuals. Interstitial features occupied 5.8 ±4.4% of lung tissue (**Table E1**). As shown previously, in NHW participants, presence of the *MUC5B* promoter polymorphism was associated with a greater percentage of lung occupied by interstitial features (additional percentage of lung occupied by interstitial features = 0.50, CI: 0.24, 0.75, $p<0.001$ in the adjusted analysis).[3]

SUPPLEMENTAL Tables:

Table E1: Summary of General Characteristics of the Cohort

General Characteristics	Entire Cohort	Those with Objective Interstitial Features	Those with Objective Emphysematous Features
Number of Participants, n available (% of entire study group)	6863 (100%)	856 (12.5%)	1907 (27.8%)
Age in Years, mean (SD)	60.7 (9.0)	61.6 (9.4)	64.8 (8.0)
Female, n (%)	3239 (47.2%)	488 (57.0%)	811 (42.5%)
Black, n (%)	1864 (27.2%)	402 (47.0%)	356 (18.7%)
Body Mass Index, mean (SD)	28.5 (5.9)	29.7 (6.2)	27.5 (6.3)
Pack Years, mean (SD)	44.6 (24.8)	45.0 (26.4)	54.1 (27.7)
Currently Smoking, n (%)	3219 (46.9%)	454 (53.0%)	510 (26.7%)
Percent Predicted Forced Expiratory Volume in One Second, mean (SD)	77.0 (26.5)	77.0 (23.3)	51.7 (24.6%)
Years Followed, mean (SD)	5.6 (2.3)	5.2 (2.3)	5.5 (2.3)
Number with ARD Event, n (%)	2773 (40.4%)	346 (40.4%)	1175 (61.6%)
Total Number of ARD Events, mean (SD)	2.1 (4.3)	2.0 (4.1)	4.0 (5.9)
Number with Severe ARD Events, n (%)	1609 (22.6%)	208 (23.2%)	753 (39.5%)
Total Number of Severe ARD Events, mean (SD)	0.7 (1.9)	0.8 (2.4)	1.3 (2.5)
Genetic Characteristics	Entire Cohort	Those with Objective Interstitial Features	Those with Objective Emphysematous Features
MUC5B GG, n (%)*	5833 (85.0%)	736 (86.0%)	1542 (83.9%)
MUC5B GT, n (%)*	977 (14.2%)	117 (13.7%)	275 (15.0%)
MUC5B TT, n (%)*	53 (0.8%)	3 (0.4%)	21 (1.1%)
Imaging Characteristics	Entire Cohort	Those with Objective Interstitial Features	Those with Objective Emphysematous Features
% Lung Occupied by Interstitial Features, mean (SD)	5.8 (4.4)	14.7 (5.1)	5.5 (3.7)
% Lung Occupied by Emphysematous Features, mean (SD)	10.6 (17.3)	7.2 (12.4)	33.8 (19.2)

* GG are individuals homozygous for the major allele, GT are those individuals heterozygous for the minor allele, and TT are those individuals homozygous for the minor allele. Abbreviations: acute respiratory disease (ARD), standard deviation (SD)

Table E2: Summary of General Characteristics of non-Hispanic Whites in the Cohort

General Characteristics	All non-Hispanic Whites	non-Hispanic Whites with Objective Interstitial Features	non-Hispanic Whites with Objective Emphysematous Features
Number of Participants, n available (% of entire cohort)	4999 (72.8%)	454 (6.6%)	1488 (21.7%)
Age in Years, mean (SD)	62.6 (8.7)	65.6 (8.8)	65.6 (7.8)
Female, n (% of NHW)	2370 (47.4%)	253 (55.7%)	624 (41.9%)
Body Mass Index, mean (SD)	28.4 (5.7)	29.5 (6.0)	27.5 (6.0)
Pack Years, mean (SD)	46.9 (25.7)	49.5 (29.3)	56.4 (28.0)
Currently Smoking, n (% of NHW)	1795 (35.9%)	152 (33.5%)	331 (22.2%)
Percent Predicted Forced Expiratory Volume in One Second, mean (SD)	74.7 (26.6)	78.8 (2.2)	51.1 (24.1)
Years Followed, mean (SD)	6.0 (2.1)	5.7 (2.1)	5.6 (2.2)
Number with Exacerbations, n (% of NHW)	2167 (43.4%)	204 (44.9%)	934 (62.8%)
Total Number of Exacerbations, mean (SD)	2.3 (4.6)	2.2 (4.0)	4.2 (6.0)
Number with Severe Exacerbations, n (% of NHW)	1104 (22.1%)	95 (20.9%)	574 (38.6%)
Total Number of Severe Exacerbations, mean (SD)	0.6 (1.7)	0.6 (1.8)	1.2 (2.3)
Genetic Characteristics	All non-Hispanic Whites	non-Hispanic Whites with Objective Interstitial Features	non-Hispanic Whites with Objective Emphysematous Features
MUC5B GG, n (% of NHW)*	4051 (81.0%)	357 (78.6%)	1207 (81.1%)
MUC5B GT, n (% of NHW)*	896 (17.9%)	94 (20.7%)	261 (17.5%)
MUC5B TT, n (% of NHW)*	52 (1.0%)	3 (0.7%)	20 (1.3%)
Imaging Characteristics	All non-Hispanic Whites	non-Hispanic Whites with Objective Interstitial Features	non-Hispanic Whites with Objective Emphysematous Features
% Lung Occupied by Interstitial Features, mean (SD)	5.3 (3.8)	14.3 (4.5)	5.2 (3.5)
% Lung Occupied by Emphysematous Features, mean (SD)	11.8 (18.2)	8.9 (13.9)	34.3 (19.3)

Table E3: Effect of *MUC5B* Genotype on Acute Respiratory Disease Events in Non-Hispanic Whites: results stratified by Parenchymal Feature Type and Expressed per Copy of the Minor Allele (Additive Model)

	Odds of Reporting Exacerbation			Time to First Exacerbation			Number of Exacerbations		
Non-Hispanic Whites	Odds Ratio	Confidence Interval	p	Hazard Ratio	Confidence Interval	p	Incident Rate Ratio	Confidence Interval	p
Entire Cohort	0.93	0.80, 1.08	0.310	0.95	0.86, 1.05	0.276	0.97	0.86, 1.08	0.571
Those with Interstitial Features	0.39	0.22, 0.69	0.001	0.56	0.38, 0.84	0.005	0.59	0.39, 0.89	0.011
Those with Emphysematous Features	1.13	0.86, 1.48	0.402	1.06	0.91, 1.23	0.488	1.03	0.88, 1.21	0.684

Notes:

1) All effects expressed per copy of the minor allele (T).
2) Odds of reporting an ARD event assessed using logistic regression, time-to-first ARD event assessed using Cox regression and number of ARD events assessed using negative binomial regression.
3) All analyses presented are adjusted for age, sex, body mass index, current smoking status, pack years of smoking, percent predicted forced expiratory volume in one second, a reported history of gastroesophageal reflux disease, the total Saint George's Respiratory Questionnaire Score, the clinical center, and a reported history of an ARD in the year prior to study enrollment.
4) Confidence intervals are 95% confidence intervals.
5) Abbreviations: acute respiratory disease (ARD)

Table E4: Effect of *MUC5B* Genotype on Severe Acute Respiratory Disease Events: results stratified by Parenchymal Feature Type

Entire Cohort	Odds Ratio	Confidence Interval	p
All Races	0.91	0.77, 1.09	0.305
White	0.98	0.82, 1.19	0.884
Black	0.89	0.51, 1.56	0.687
Those with Interstitial Features	Odds Ratio	Confidence Interval	p
All Races	0.66	0.38, 1.14	0.134
White	0.48	0.23, 0.97	0.042
Black	1.48	0.55, 4.01	0.441
Those with Emphysematous Features	Odds Ratio	Confidence Interval	p
All Races	1.20	0.91, 1.58	0.190
White	1.26	0.94, 1.68	0.120
Black	0.92	0.27, 3.13	0.889

Notes:

1) Odds ratios are expressed as the odds of those with at least one copy of the *MUC5B* promoter polymorphism compared to those without the *MUC5B* promoter polymorphism.
2) All analyses presented are adjusted for age, sex, body mass index, current smoking status, pack years of smoking, percent predicted forced expiratory volume in one second, a reported history of gastroesophageal reflux disease, the total Saint George's Respiratory Questionnaire Score, the clinical center, and a reported history of an ARD in the year prior to study enrollment.
3) Odds of reporting an ARD event assessed using logistic regression
4) Confidence intervals are 95% confidence intervals.
5) Abbreviations: acute respiratory disease (ARD)

Table E5A: Effect of Interstitial Features on Acute Respiratory Disease Events: results stratified by *MUC5B* Genotype

	Odds of Reporting ARD Event			Time-to-first ARD Event			Number of ARD Events		
Entire Cohort	Odds Ratio	Confidence Interval	p	Hazard Ratio	Confidence Interval	p	Incident Rate Ratio	Confidence Interval	p
All Races	0.91	0.77, 1.08	0.285	0.96	0.86, 1.08	0.506	1.04	0.91, 1.19	0.575
White	0.96	0.77, 1.20	0.736	1.00	0.86, 1.16	0.992	1.01	0.85, 1.19	0.938
Black	0.90	0.70, 1.17	0.441	0.96	0.79, 1.17	0.694	1.09	0.86, 1.39	0.475
Those with *MUC5B* Promoter Polymorphism (GT or TT)	Odds Ratio	Confidence Interval	p	Hazard Ratio	Confidence Interval	p	Incident Rate Ratio	Confidence Interval	p
All Races	0.66	0.42, 1.05	0.080	0.79	0.57, 1.11	0.179	0.82	0.57, 1.18	0.278
White	0.57	0.34, 0.97	0.037	0.74	0.50, 1.09	0.123	0.78	0.53, 1.16	0.226
Black	*	*	*	*	*	*	*	*	*
Those without *MUC5B* Promoter Polymorphism (GG)	Odds Ratio	Confidence Interval	p	Hazard Ratio	Confidence Interval	p	Incident Rate Ratio	Confidence Interval	p
All Races	1.00	0.83, 1.20	0.977	1.02	0.90, 1.15	0.806	1.11	0.95, 1.29	0.178
White	1.18	0.91, 1.51	0.210	1.10	0.94, 1.30	0.232	1.12	0.92, 1.36	0.244
Black	0.87	0.66, 1.14	0.314	0.96	0.79, 1.18	0.701	1.10	0.85, 1.41	0.465

Notes:

1) All effects expressed as those with interstitial features compared to those without interstitial features.
2) Odds of reporting an ARD event assessed using logistic regression, time-to-first ARD event assessed using Cox regression and number of ARD events assessed using negative binomial regression.
3) All analyses presented are adjusted for age, sex, body mass index, current smoking status, pack years of smoking, percent predicted forced expiratory volume in one second, a reported history of gastroesophageal reflux disease, the total Saint George's Respiratory Questionnaire Score, the clinical center, and a reported history of an ARD in the year prior to study enrollment.
4) Confidence intervals are 95% confidence intervals.
5) *n=82. Insufficient to perform multivariable model.
6) Abbreviations: acute respiratory disease (ARD)

Table E5B: Effect of Emphysematous Features on Acute Respiratory Disease Events: results stratified by *MUC5B* Genotype

	Odds of Reporting ARD Event			Time-to-first ARD Event			Number of ARD Event		
Entire Cohort	Odds Ratio	Confidence Interval	p	Hazard Ratio	Confidence Interval	p	Incident Rate Ratio	Confidence Interval	p
All Races	1.41	1.21, 1.65	<0.001	1.23	1.11, 1.36	<0.001	1.30	1.15, 1.47	<0.001
White	1.34	1.13, 1.60	<0.001	1.23	1.10, 1.38	<0.001	1.23	1.08, 1.41	0.002
Black	1.69	1.21, 2.36	<0.001	1.21	0.96, 1.54	0.111	1.48	1.08, 2.03	0.150
Those with *MUC5B* Promoter Polymorphism (GT or TT)	Odds Ratio	Confidence Interval	p	Hazard Ratio	Confidence Interval	p	Incident Rate Ratio	Confidence Interval	p
All Races	2.06	1.38, 3.06	<0.001	1.48	1.15, 1.90	0.002	1.33	0.99, 1.80	0.060
White	1.81	1.20, 2.75	0.005	1.41	1.09, 1.84	0.010	1.29	0.95, 1.75	0.100
Black	*	*	*	*	*	*	*	*	*
Those without *MUC5B* Promoter Polymorphism (GG)	Odds Ratio	Confidence Interval	p	Hazard Ratio	Confidence Interval	p	Incident Rate Ratio	Confidence Interval	p
All Races	1.33	1.12, 1.58	0.001	1.20	1.07, 1.34	0.002	1.32	1.15, 1.51	<0.001
White	1.25	1.02, 1.53	0.030	1.20	1.05, 1.36	0.006	1.23	1.06, 1.43	0.008
Black	1.64	1.16, 2.33	0.006	1.23	0.96, 1.58	0.100	1.49	1.07, 2.07	0.018

Notes:

1) All effects expressed as those with emphysematous features compared to those without emphysematous features.
2) Odds of reporting an ARD event assessed using logistic regression, time-to-first ARD event assessed using Cox regression and number of ARD events assessed using negative binomial regression.
3) All analyses presented are adjusted for age, sex, body mass index, current smoking status, pack years of smoking, percent predicted forced expiratory volume in one second, a reported history of gastroesophageal reflux disease, the total Saint George's Respiratory Questionnaire Score, the clinical center, and a reported history of an ARD in the year prior to study enrollment.
4) Confidence intervals are 95% confidence intervals.
5) *n=82. Insufficient to perform multivariable model.
6) Abbreviations: acute respiratory disease (ARD)

Table E6A: Effect of Interstitial Features on Acute Respiratory Disease Events in Non-Hispanic Whites without Emphysematous Features: results stratified by *MUC5B* Genotype

Non-Hispanic Whites with interstitial features versus those without interstitial features (dichotomized at 10%) and ***excluding those with Emphysema > 10%***	Odds of Reporting Exacerbation		
	Odds Ratio	Confidence Interval	p
Entire Cohort	0.89	0.68, 1.15	0.360
Those with *MUC5B* Promoter Polymorphism (GT or TT)	0.44	0.23, 0.85	0.014
Those without *MUC5B* Promoter Polymorphism (GG)	1.14	0.84, 1.55	0.398
Non-Hispanic Whites with interstitial features versus those without interstitial features (dichotomized at 10%) and ***excluding those with Emphysema > 5%***	Odds of Reporting Exacerbation		
	Odds Ratio	Confidence Interval	p
Entire Cohort	0.99	0.75, 1.31	0.943
Those with *MUC5B* Promoter Polymorphism (GT or TT)	0.49	0.24, 0.98	0.043
Those without *MUC5B* Promoter Polymorphism (GG)	1.28	0.92, 1.78	0.144

Notes:

1) All effects expressed as those with interstitial features compared to those without interstitial features.
2) Odds of reporting an ARD event assessed using logistic regression.
3) All analyses presented are adjusted for age, sex, body mass index, current smoking status, pack years of smoking, percent predicted forced expiratory volume in one second, a reported history of gastroesophageal reflux disease, the total Saint George's Respiratory Questionnaire Score, the clinical center, and a reported history of an ARD in the year prior to study enrollment.
4) Confidence intervals are 95% confidence intervals.
5) Abbreviations: acute respiratory disease (ARD)

Table E6B: Effect of Emphysematous features on Acute Respiratory Disease Events in Non-Hispanic Whites without Interstitial Features: results stratified by *MUC5B* Genotype

Non-Hispanic Whites with emphysematous features versus those without emphysematous features (dichotomized at 10%) ***excluding those with interstitial > 10%***	Odds of Reporting Exacerbation		
	Odds Ratio	Confidence Interval	p
Entire Cohort	1.34	1.11, 1.62	0.022
Those with *MUC5B* Promoter Polymorphism (GT or TT)	1.67	1.07, 2.60	0.025
Those without *MUC5B* Promoter Polymorphism (GG)	1.29	1.04, 1.60	0.020
Non-Hispanic Whites with emphysematous features versus those without emphysematous features (dichotomized at 10%) ***excluding those with interstitial > 5%***	Odds of Reporting Exacerbation		
	Odds Ratio	Confidence Interval	p
Entire Cohort	1.36	1.07, 1.73	0.011
Those with *MUC5B* Promoter Polymorphism (GT or TT)	1.70	0.93, 3.11	0.084
Those without *MUC5B* Promoter Polymorphism (GG)	1.30	0.99, 1.71	0.058

Notes:

1) All effects expressed as those with emphysema compared to those without emphysema.
2) Odds of reporting an ARD event assessed using logistic regression.
3) All analyses presented are adjusted for age, sex, body mass index, current smoking status, pack years of smoking, percent predicted forced expiratory volume in one second, a reported history of gastroesophageal reflux disease, the total Saint George's Respiratory Questionnaire Score, the clinical center, and a reported history of an ARD in the year prior to study enrollment.
4) Confidence intervals are 95% confidence intervals.
5) Abbreviations: acute respiratory disease (ARD)

Table E7: Effect of the Interaction between *MUC5B* Genotype and Continuously Measured Parenchymal Feature Types on the Odds of Reporting an Acute Respiratory Disease Event

Interstitial Features	Odds Ratio for Interaction Term	Confidence Interval	p interaction
Entire Cohort	0.963	0.926, 1.000	0.051
Whites	0.948	0.907, 0.990	0.016
Blacks	1.011	0.917, 1.115	0.820
Emphysematous Features	Odds Ratio for Interaction Term	Confidence Interval	p interaction
Entire Cohort	1.004	0.995, 1.013	0.401
Whites	1.005	0.995, 1.015	0.315
Blacks	0.982	0.982, 1.015	0.285

Notes:

1) Interaction between parenchymal feature types measured as continuous variable and presence of *MUC5B* promoter polymorphism (GT/TT vs. GG).
2) Odds of reporting an ARD event assessed using logistic regression
3) Multivariable analyses adjusted for age, sex, body mass index, current smoking status, pack years of smoking, percent predicted forced expiratory volume in one second, a reported history of gastroesophageal reflux disease, the total Saint George's Respiratory Questionnaire Score, the clinical center, and a reported history of an ARD in the year prior to study enrollment.
4) Abbreviations: acute respiratory disease (ARD)

Table E8: Effect of the Interaction between *MUC5B* Genotype and Continuously Measured Parenchymal Feature Types on the Odds of Reporting an Acute Respiratory Disease Event in Non-Hispanic Whites – with the Exclusion of Those with the "Other" Feature Type as Noted

Non-Hispanic Whites	Odds Ratio for Interaction Term	Confidence Interval	p interaction
Interaction between Interstitial Changes and *MUC5b* Genotype – ***Excluding those with Emphysema > 10%***	0.94	0.89, 0.99	0.015
Interaction between Emphysematous Changes and *MUC5b* Genotype – ***Excluding those with Interstitial > 10%***	1.00	0.99, 1.01	0.411

Notes:

1) Interaction between parenchymal feature types measured as continuous variable and presence of *MUC5B* promoter polymorphism (GT/TT vs. GG).
2) Odds of reporting an ARD event assessed using logistic regression
3) Multivariable analyses adjusted for age, sex, body mass index, current smoking status, pack years of smoking, percent predicted forced expiratory volume in one second, a reported history of gastroesophageal reflux disease, the total Saint George's Respiratory Questionnaire Score, the clinical center, and a reported history of an ARD in the year prior to study enrollment.
4) Abbreviations: acute respiratory disease (ARD)

Table E9: Effect of *MUC5B* Genotype on Acute Respiratory Disease Events in those with Parenchymal Disease defined as Greater than 5% of Lung occupied by each Feature Type: results stratified by Parenchymal Feature Type

	Odds of Reporting ARD Event			Time-to-First ARD Event			Number of ARD Events		
Those with Interstitial Features (>5% of lung occupied)	Odds Ratio	Confidence Interval	p	Hazard Ratio	Confidence Interval	p	Incident Rate Ratio	Confidence Interval	p
All Races	0.85	0.68, 1.07	0.164	0.90	0.77, 1.04	0.150	0.97	0.82, 1.16	0.769
White	0.79	0.62, 1.02	0.067	0.84	0.72, 0.99	0.039	0.98	0.82, 1.17	0.808
Black	1.05	0.55, 1.99	0.891	1.24	0.77, 1.99	0.370	0.80	0.44, 1.46	0.475
Those with Emphysematous Features (>5% of lung occupied)	Odds Ratio	Confidence Interval	p	Hazard Ratio	Confidence Interval	p	Incident Rate Ratio	Confidence Interval	p
All Races	0.96	0.76, 1.22	0.743	1.00	0.86, 1.15	0.940	1.06	0.91, 1.24	0.465
White	0.97	0.75, 1.25	0.804	0.98	0.85, 1.14	0.810	1.10	0.94, 1.29	0.236
Black	0.70	0.28, 1.72	0.434	0.87	0.45, 1.66	0.670	0.48	0.21, 1.08	0.075

Notes:

1) All effects expressed as those with *MUC5B* promoter polymorphism (GT and TT) compared to those without the polymorphism (GG).
2) Odds of reporting an ARD event assessed using logistic regression, time-to-first ARD event assessed using Cox regression and number of ARD events assessed using negative binomial regression.
3) All analyses presented are adjusted for age, sex, body mass index, current smoking status, pack years of smoking, percent predicted forced expiratory volume in one second, a reported history of gastroesophageal reflux disease, the total Saint George's Respiratory Questionnaire Score, the clinical center, and a reported history of an ARD in the year prior to study enrollment.
4) Confidence intervals are 95% confidence intervals.
5) Abbreviations: acute respiratory disease (ARD)

Table E10: Effect of *MUC5B* Genotype on Acute Respiratory Disease Events in Non-Hispanic Whites with Limited Adjustments to Evaluate for Collider Bias: results stratified by Parenchymal Feature Type

	Odds of Reporting Exacerbation			Time to First Exacerbation			Number of Exacerbations		
Non-Hispanic Whites	Odds Ratio	Confidence Interval	p	Hazard Ratio	Confidence Interval	p	Incident Rate Ratio	Confidence Interval	p
Entire Cohort	0.98	0.85, 1.14	0.783	0.97	0.87, 1.08	0.600	1.06	0.92, 1.23	0.422
Those with Interstitial Features	0.42	0.25, 0.70	0.001	0.55	0.37, 0.82	0.003	0.69	0.42, 1.13	0.138
Those with Emphysematous Features	1.12	0.84, 1.48	0.444	1.01	0.85, 1.18	0.947	1.11	0.91, 1.35	0.315

Notes:

1) All effects expressed as those with *MUC5B* promoter polymorphism (GT and TT) compared to those without the polymorphism (GG).
2) Odds of reporting an ARD event assessed using logistic regression, time-to-first ARD event assessed using Cox regression and number of ARD events assessed using negative binomial regression.
3) In order to determine if collider bias exists in the setting of clinical variables potentially related to genotype, the analyses presented are only adjusted for age, sex, current smoking status, pack years of smoking, and the clinical center.
4) Confidence intervals are 95% confidence intervals.
5) Abbreviations: acute respiratory disease (ARD)

SUPPLEMENTAL REFERENCES

1. Regan EA, Hokanson JE, Murphy JR, et al. Genetic epidemiology of COPD (COPDGene) study design. *Copd* 2010;7(1):32-43. doi: 10.3109/15412550903499522 [published Online First: 2010/03/11]
2. Stewart JI, Moyle S, Criner GJ, et al. Automated telecommunication to obtain longitudinal follow-up in a multicenter cross-sectional COPD study. *Copd* 2012;9(5):466-72. doi: 10.3109/15412555.2012.690010
3. Ash SY, Harmouche R, Putman RK, et al. Clinical and Genetic Associations of Objectively Identified Interstitial Changes in Smokers. *CHEST Journal* 2017;In press
4. Ash SY, Harmouche R, Ross JC, et al. The Objective Identification and Quantification of Interstitial Lung Abnormalities in Smokers. *Academic radiology* 2016 doi: 10.1016/j.acra.2016.08.023
5. Washko GR, Lynch DA, Matsuoka S, et al. Identification of early interstitial lung disease in smokers from the COPDGene Study. *Academic radiology* 2010;17(1):48-53. doi: 10.1016/j.acra.2009.07.016
6. Seibold MA, Wise AL, Speer MC, et al. A common MUC5B promoter polymorphism and pulmonary fibrosis. *The New England journal of medicine* 2011;364(16):1503-12. doi: 10.1056/NEJMoa1013660
7. Chung JH, Peljto AL, Chawla A, et al. CT Imaging Phenotypes of Pulmonary Fibrosis in the MUC5B Promoter Site Polymorphism. *Chest* 2016;149(5):1215-22. doi: 10.1016/j.chest.2015.11.009
8. Bowler RP, Kim V, Regan E, et al. Prediction of acute respiratory disease in current and former smokers with and without COPD. *Chest* 2014;146(4):941-50. doi: 10.1378/chest.13-2946
9. Variant: rs35705950 Population Genetics e!Ensembl [Available from: http://www.ensembl.org/Homo_sapiens/Variation/Population?db=core;r=11:1219491-1220491;v=rs35705950;vdb=variation;vf=10484009 accessed August 1, 2017 2017.
10. Schoenfeld D. Partial residuals for the proportional hazards regression model. *Biometrika* 1982;69(1):239-41.
11. Cole SR, Platt RW, Schisterman EF, et al. Illustrating bias due to conditioning on a collider. *International journal of epidemiology* 2010;39(2):417-20. doi: 10.1093/ije/dyp334