CHEST Online Supplement

Lobar Emphysema Distribution Is Associated With 5-Year Radiological Disease Progression

Adel Boueiz, MD; Yale Chang, PhD; Michael H. Cho, MD, MPH; George R. Washko, MD; Raul San José Estépar, PhD; Russell P. Bowler, MD; James D. Crapo, MD; Dawn L. DeMeo, MD, MPH; Jennifer G. Dy, PhD; Edwin K. Silverman, MD, PhD; Peter J. Castaldi, MD; for the COPDGene Investigators

CHEST 2018; 153(1):65-76

 DOI: 10.1016/j.chest.2017.09.022

e-Appendix 1.

COPDGene study: Inclusion and exclusion criteria

Subjects were aged 45 to 80 years and had at least 10 pack-years of lifetime smoking history[1]. Exclusion criteria included pregnancy, history of other lung diseases except asthma, prior lobectomy or lung volume reduction surgery, active cancer undergoing treatment, or known or suspected lung cancer[1].

Spirometric data

At both visits, spirometry was performed before and after administration of 180 mcg of albuterol (ndd Easy-One spirometer, Andover, MA). Percent predicted values were calculated using Hankinson NHANES equations[2]. COPD was defined by post-bronchodilator $FEV_1/FVC<0.70$ at baseline visit per the GOLD guidelines[3]. Bronchodilator responsiveness was defined as an increase in FEV_1 or FVC by 200 ml and 12% from baseline. Disease severity was described by GOLD spirometric grade. "GOLD 0" was defined as post-bronchodilator $FEV_1/FVC\geq0.70$ at baseline visit and FEV_1 percent predicted ≥80%. Participants with $FEV_1/FVC\geq0.70$ but with $FEV_1<80\%$ predicted were considered to have Preserved Ratio Impaired Spirometry (PRISm)[4].

Genotyping and imputation quality control

Genome-wide SNP genotyping was performed using the Illumina Omni Express array (Illumina, Inc; San Diego, CA). Additional markers were imputed using MaCH and minimac with the 1,000 Genomes Phase I v3 European (EUR) and cosmopolitan reference panels for the NHW and AA subjects, respectively[5, 6]. Standard quality control steps were performed on DNA samples and single nucleotide polymorphism (SNP) data as previously described[7, 8]. SNPs with imputation $r^2<0.3$, a minor allele frequency less than 1%, or extreme Hardy-Weinberg deviation (P-value $<10^{-8}$) were excluded from further analyses. Post-imputation, a total of 5,870,826 variants in COPDGene NHW subjects and 8,151,810 variants in COPDGene AA subjects passed quality control and were analyzed.

Unsupervised learning with random forest predictors

Briefly, this method can be applied to mixed (i.e. continuous and categorical) data types, and provides a data-driven approach to feature weighting and selection. This approach leverages a supervised learning method (i.e. random forests) to discriminate between the actual data and permuted data drawn from the same data distributions (in which correlations between features have been broken via a permutation procedure). In this formulation, the actual and permuted data are labeled and combined, and the random forests procedure is used to predict the "actual"

observations from the permuted ones. This leads to a natural weighting of variables based on those that are most useful for discrimination between real and permuted data, because these variables are selected more frequently in the random forests procedure. Based on the resulting trees, a similarity matrix is constructed on the number of times pairs of observations co-occur within the same terminal node. Clustering was performed on the resulting similarity matrix using the dynamic hybrid approach of Langfelder et al. which combines hierarchical clustering with distance-based clustering[9]. The dynamic tree cut algorithm uses specific similarity criteria to identify subjects that are not sufficiently similar to other members in their assigned cluster. These subjects are then reassigned to a miscellaneous group of "poorly clustered" subjects. The minimum cluster size was set to 20 and the "deep split" parameter was set to 2.

PCA plot

We applied principal component (PCA) analysis on the residualized lobar values. We then visualized PC1 and PC2, the top-two principal components explaining the highest variances of the five residualized lobar emphysema variables (Figure 1). Their explained-variance ratios were 50.7% and 26.2%, respectively.

Statistical tests

Differences in characteristics between clusters were analyzed using omnibus tests (ANOVA, Kruskal-Wallis, and chi-square tests respectively for the normally distributed variables, non-normally distributed, and categorical variables) followed by pairwise tests (t-tests, Nemenyi tests, and chi-square tests respectively for the normally distributed, non-normally distributed variables and categorical variables). Univariate and multivariate regression models were used to relate COPD-related measures to upper versus lower lobe emphysema predominant cluster membership. Ordinal logistic regression was used for MMRC. Quantile regression with bootstrap standard error estimation was used for analysis of the medians of the other continuous variables. Subgroup analyses were also performed by GOLD grade, emphysema severity, and Visit 1 BMI. All tests were done at a significance level of 0.05 with Bonferroni adjustment for multiple comparisons.

REFERENCES

1 Regan EA, Hokanson JE, Murphy JR et al. Genetic epidemiology of COPD (COPDGene) study design. *COPD*. 2010;7(1):32-43.

2 Hankinson JL, Odencrantz JR, Fedan KB. Spirometric reference values from a sample of the general U.S. population. *Am J Respir Crit Care Med*. 1999;159(1):179-187.

3 Vestbo J, Hurd SS, Agusti AG et al. Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: GOLD executive summary. *Am J Respir Crit Care Med*. 2013;187(4):347-365.

4 Wan ES, Castaldi PJ, Cho MH et al. Epidemiology, genetics, and subtyping of preserved ratio impaired spirometry (PRISm) in COPDGene. *Respir Res*. 2014;15:89-014-0089-y.

5 Li Y, Willer CJ, Ding J, Scheet P, Abecasis GR. MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genet Epidemiol*. 2010;34(8):816-834.

6 Howie B, Fuchsberger C, Stephens M, Marchini J, Abecasis GR. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nat Genet*. 2012;44(8):955-959.

7 Regan EA, Hokanson JE, Murphy JR et al. Genetic epidemiology of COPD (COPDGene) study design. *COPD*. 2010;7(1):32-43.

8 Pillai SG, Ge D, Zhu G et al. A genome-wide association study in chronic obstructive pulmonary disease (COPD): identification of two major susceptibility loci. *PLoS Genet*. 2009;5(3):e1000421.

9 Langfelder P, Zhang B, Horvath S. Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R. *Bioinformatics*. 2008;24(5):719-720.

e-Table 1. Definitions of COPD characteristics and comorbidities.

Characteristic or comorbidity	**Definition**
Physician diagnosis of asthma before age 40	Self-report
Blue bloater	Chronic bronchitis (defined below), BMI > 25, Resting oxygen saturation < 90%
Bronchodilator responsiveness	Increase in FEV_1 or FVC by 200 ml and 12% from baseline
Cerebrovascular disease	Self-report of transient ischemic attack or stroke
Chronic bronchitis	Chronic cough and phlegm for ≥ 3 months/year for at least 2 consecutive years
Coronary artery disease	Self-report of heart attack, coronary artery disease, angina, angioplasty, or coronary artery bypass graft
COPD exacerbation	Self-report of acute worsening of respiratory symptoms that required the use of antibiotics and/or systemic steroids in the previous year
Severe COPD exacerbation	Self-report of COPD exacerbation requiring an emergency department visit or hospital admission.
Diabetes mellitus	Self-report
Dyslipidemia	Self-report
Gastroesophageal reflux disease	Self-report
Hypoxemia	Resting oxygen saturation ≤ 88%
Heart failure	Self-report
Hypertension	Self-report
Metabolic syndrome	3 of 4: BMI ≥ 30 (measured), self-reported diabetes mellitus, hypertension, and high cholesterol
Obesity	BMI ≥ 30
Osteoporosis	Self-report
Peptic ulcer disease	Self-report
Peripheral vascular disease	Self-report
Pink puffer	Emphysema > 10%, BMI ≤ 20, Resting oxygen saturation ≥ 90%
Pneumothorax	Self-report
Severe, early-onset COPD	Age < 55 years, FEV_1 < 50% predicted
Sleep apnea	Self-report

e-Table 2. Characteristics of study subjects.

	COPDGene Non-Hispanic Whites	COPDGene African Americans
Sample size	6195	3015
Age (Years)	62.30 [55.10;68.80]	53.10 [49.30;58.40]
Sex (%Male)	3277 (52.90%)	1681 (55.76%)
Pack-Years	42.00 [30.05;58.70]	34.50 [22.50;46.80]
FEV_1, %predicted	78.30 [55.80;93.20]	86.00 [69.70;98.80]
Total emphysema (%LAA-950)	2.82 [0.84;8.82]	0.97 [0.35;3.04]
GOLD:		
PRISm	647	476
0	2370	1656
1	580	165
2	1351	419
3	834	211
4	413	88
COPD cases (COPD GOLD 2, 3, 4)	2598 (41.94%)	718 (23.80%)
Subjects with total emphysema greater than 5%	2265 (36.56%)	502 (16.60%)

FEV_1: Forced expiratory volume in 1 second. %LAA-950: Percentage of lung voxels with attenuation on CT densitometry lower than -950 Hounsfield units. GOLD: Global Initiative for Chronic Obstructive Lung Disease; PRISm: Preserved Ratio Impaired Spirometry. Continuous variables are expressed as median and interquartile range (25^{th} to 75^{th} percentile) and categorical variables are expressed as absolute values and proportion.

e-Table 3. Characteristics at Visit 1 and Visit 2 of the subjects who were unassigned by the clustering algorithm (N=6,197). 3,101 of these 6,197 subjects had 5-year follow up visit.

Demographics	
Age (Years)	59.60 [52.50;67.00]
Male (%)	56.95%
Non-Hispanic Whites (%)	69.95%
Smoking History (Pack-years)	39.00 [26.90;54.00]
Current smoking (%)	48.85%
CT measurements	
Total emphysema (% LAA-950)	2.40 [0.90;7.12]
Right upper lobe emphysema	1.90 [0.68;6.65]
Right middle lobe emphysema	3.37 [1.25;9.16]
Right lower lobe emphysema	1.65 [0.55;5.66]
Left upper lobe emphysema	2.96 [1.13;8.29]
Left lower lobe emphysema	2.00 [0.78;5.78]
U/L ratio	1.31 [0.94;1.86]
Airway wall thickening	61.01 [58.98;63.29]
Pi10	3.65 [3.58;3.73]
Percent gas trapping (%)	17.79 [9.00;35.27]
Physiology	
FEV_1 post-bronchodilator (Percent predicted)	82.30 [61.40;96.30]
FEV_1/FVC	0.71 [0.59;0.78]
Pre/Post-bronchodilator FEV_1 % change	4.40 [0.60;9.40]
Pre/Post-bronchodilator FVC % change	2.00 [-2.30;8.00]
Bronchodilator responsiveness	21.04%
FEF2575	1.60 [0.68;2.59]
GOLD: PRISm 0 1 2 3 4	 10.55% 44.96% 8.84% 18.81% 11.20% 5.63%
6-minute walk distance (Feet)	1407.00 [1151.00;1660.00]
BODE index	1.35 ± 1.82
Patient-reported outcomes	
MMRC dyspnea score	1.27 ± 1.43
SGRQ score	31.21 ± 25.64

Chronic bronchitis (%)	18.25%
Exacerbation frequency (%)	20.00%
History of severe exacerbation(s) (%)	10.44%
Family history of COPD	26.54%
Severe, early-onset COPD	2.44%
Pink puffer	1.13%
Blue Bloater	0.54%
Hypoxia	2.65%
Comorbidities	
Coronary artery disease	6.42%
Heart failure	2.90%
Stroke	2.48%
Cerebrovascular disease	4.37%
Peripheral vascular disease	2.31%
Hypertension	43.01%
Diabetes	12.50%
Dyslipidemia	38.63%
Metabolic syndrome	16.62%
Obesity	36.01%
Sleep apnea	14.33%
GERD	25.30%
Osteoporosis	8.65%
Pneumothorax	3.10%
Physician diagnosis of asthma before age 40	8.12%
Longitudinal changes in 5-year follow up	
Change of FEV_1 between Visit 1 and Visit 2 (mL/year)	-37.00 [-67.00;-8.00]
Change of FEV_1 between Visit 1 and Visit 2 (percent predicted/year)	-0.37 [-1.52;0.75]
Change of 6-minute walk distance between Visit 1 and Visit 2 (Feet)	-123.00 [-311.00;65.00]
Change of MMRC dyspnea score between Visit 1 and Visit 2	0.06 ± 1.20
Change of SGRQ between Visit 1 and Visit 2	0.65 ± 14.95
Percent change of total emphysema between Visit 1 and Visit 2	-6.00% [-57.83;45.46]
Percent change of the percent gas trapping between Visit 1 and Visit 2	3.21% [-29.43;40.33]
Percent change of CT-total lung capacity between Visit 1 and Visit 2	-0.86% [-5.51;4.14]

Emphysema is defined as percent of CT low attenuation area below -950 Hounsfield units at end-inspiration (% LAA-950); Percent gas trapping is measured at end-exhalation and defined as the percentage of lung voxels with density less than -856 Hounsfield units; U/L ratio: Ratio of %LAA-950 in both upper lobes to that in both lower lobes; Airway wall area percent is the percentage of the wall area compared with the total bronchial area for segmental airways; Pi10: Root of the wall area of a

hypothetical airway of 10-mm internal perimeter. "Percent change between Visit 1 and Visit 2" variables are defined as (Value at Visit 2 - Value at Visit 1) / Value at Visit 1.
FEV_1: Forced expiratory volume in 1 second; FVC: Forced vital capacity; Bronchodilator responsiveness: Increase in FEV_1 or FVC by 200 ml and 12% from baseline; GOLD: Global Initiative for Chronic Obstructive Lung Disease; PRISm: Preserved Ratio Impaired Spirometry; BODE index: Body mass index, measure of airflow obstruction, dyspnea score and exercise capacity; MMRC: Modified Medical Research Council; SGRQ: St. George's Respiratory Questionnaire; Exacerbation frequency: Percent of subjects reporting at least one COPD exacerbation in the previous year; Severe exacerbation is defined as COPD exacerbation requiring an emergency department visit or hospital admission; Hypoxia: Resting oxygen saturation ≤ 88%; Severe, early-onset COPD: Age < 55 years, FEV_1 < 50% predicted; Pink puffer: Emphysema > 10%, BMI ≤ 20, O_2 sat ≥ 90%; Blue bloater: Chronic bronchitis, BMI > 25, O_2 sat < 90%; Cerebrovascular disease: Self-report of transient ischemic attack or stroke; GERD: Gastro-esophageal reflux disease; Metabolic syndrome: 3 of 4: BMI ≥ 30 (measured), diabetes mellitus, hypertension, and high cholesterol (all self-report); Obesity: BMI≥30.
Variables are expressed as mean and standard deviation for continuous normally distributed variables, median and interquartile range (25^{th} to 75^{th} percentile) for continuous non-normally distributed variables, and percentages for categorical variables.

e-Table 4. Other imaging and clinical characteristics at Visit 1 of the subjects in each of the emphysema distribution clusters

	Minimal emphysema cluster (N=1,312)	**Upper-lobe predominant (ULP) cluster (N=905)**	**Lower-lobe predominant (LLP) cluster (N=796)**	**P-value *(All groups omnibus test)***	**P-value *(ULP vs LLP)***
Lobar Emphysema (%LAA-950)					
Right upper lobe emphysema	0.21 [0.11;0.36]	10.49 [4.35;23.93]	2.63 [0.79;10.38]	***<0.001***	***<0.001***
Right middle lobe emphysema	0.26 [0.14;0.43]	2.83 [0.93;7.04]	5.86 [2.14;14.49]	***<0.001***	***<0.001***
Right lower lobe emphysema	0.19 [0.10;0.30]	1.62 [0.62;4.54]	6.45 [2.46;19.11]	***<0.001***	***<0.001***
Left upper lobe emphysema	0.24 [0.14;0.44]	6.80 [3.03;16.73]	4.10 [1.53;11.98]	***<0.001***	***<0.001***
Left lower lobe emphysema	0.23 [0.13;0.42]	1.58 [0.68;4.12]	7.22 [3.29;20.55]	***<0.001***	***<0.001***
Physiology					
Pre/Post-bronchodilator FEV_1 % change	3.00 [-0.90;7.50]	5.10 [0.30;10.30]	5.85 [1.33;11.68]	***<0.001***	0.06
Pre/Post-bronchodilator FVC % change	0.40 [-3.80;6.20]	2.70 [-1.70;9.20]	4.00 [-0.70;11.08]	***<0.001***	***0.01***
FEF2575	2.25 [1.52;3.07]	0.94 [0.48;1.70]	0.72 [0.32;1.92]	***<0.001***	0.6
GOLD: PRISm 0 1 2 3 4	 25.08% 61.28% 3.28% 8.31% 1.83% 0.23%	 8.95% 24.75% 11.60% 33.15% 17.35% 4.20%	 7.41% 26.63% 6.16% 24.50% 21.36% 13.94%	 ***<0.001***	 ***<0.001***
BODE index	0.92 ± 1.24	1.83 ± 1.84	2.12 ± 2.18	***<0.001***	***0.001***
Hypoxia	1.07%	4.10%	4.40%	***<0.001***	1.00
6-minute walk distance (Feet)	1400.00 [1164.00;1590.00]	1280.50 [1010.00;1495.00]	1329.00 [1062.50;1598.00]	***<0.001***	***<0.001***

COPD characteristics					
History of severe exacerbation(s) (%)	8.84%	18.23%	15.58%	***<0.001***	0.49
Family history of COPD	21.19%	29.65%	31.53%	***<0.001***	1.00
Severe, early-onset COPD	0.76%	4.09%	3.77%	***<0.001***	1.00
Pink puffer	0%	0.57%	2.83%	***<0.001***	***0.002***
Blue Bloater	0.23%	0.88%	0.88%	0.07	1.00
Comorbidities					
Physician diagnosis of asthma before age 40	11.52%	7.51%	9.17%	***0.006***	0.75
Heart failure	2.90%	3.98%	3.39%	0.38	1.00
Cerebrovascular disease	3.66%	5.64%	4.02%	0.07	0.46
Peripheral vascular disease	1.83%	2.21%	2.76%	0.36	1.00
Hypertension	40.70%	45.08%	49.37%	***<0.001***	0.26
GERD	21.27%	26.08%	29.02%	***<0.001***	0.58
Osteoporosis	5.87%	11.93%	11.06%	***<0.001***	1.00
Pneumothorax	1.30%	5.64%	3.64%	***<0.001***	0.21

%LAA-950 is defined as percent of CT densitometry low attenuation area below -950 Hounsfield units at end-inspiration using Slicer software; Percent gas trapping is measured at end-exhalation and defined as the percentage of lung voxels with density less than -856 Hounsfield units. FEV_1: Forced expiratory volume in 1 second; FVC: Forced vital capacity; Bronchodilator responsiveness: Increase in FEV_1 or FVC by 200 ml and 12% from baseline; GOLD: Global Initiative for Chronic Obstructive Lung Disease; PRISm: Preserved Ratio Impaired Spirometry; BODE index: Body mass index, measure of airflow obstruction, dyspnea score and exercise capacity; Hypoxia: Resting oxygen saturation ≤ 88%; Severe exacerbation is defined as COPD exacerbation requiring an emergency department visit or hospital admission; Severe, early-onset COPD: Age < 55 years, FEV_1 < 50% predicted; Pink puffer: Emphysema > 10%, BMI ≤ 20, O_2 sat ≥ 90%; Blue bloater: Chronic bronchitis, BMI > 25, O_2 sat < 90%; Cerebrovascular disease: Self-report of transient ischemic attack or stroke; GERD: Gastro-esophageal reflux disease.
Variables are expressed as mean and standard deviation for continuous normally distributed variables, median and interquartile range (25^{th} to 75^{th} percentile) for continuous non-normally distributed variables, and percentages for categorical variables. Omnibus P-values are obtained using ANOVA for the continuous normally distributed variables, Kruskal-Wallis test for the continuous non-normally distributed variables, and chi-square test for the proportions. P-values of ULP *vs* LLP are from pairwise t-tests, Nemenyi tests, and chi-square post hoc tests for pairwise comparisons between ULP and LLP clusters for the continuous normally distributed, non-normally distributed variables and categorical variables, respectively. P-values < 0.05 are bolded and italicized.

e-Table 5. Five-year changes in imaging characteristics in upper versus lower-lobe predominant emphysema clusters: Percent changes between Visit 1 and Visit 2 relative to Visit 1.

	Upper-lobe predominant (ULP) cluster	**Lower-lobe predominant (LLP) cluster**	**P-value (ULP *vs* LLP)**
Percent change of total emphysema (%LAA-950) between Visit 1 and Visit 2 relative to Visit 1			
All	23.26% [-11.93;62.26]	-0.12% [-55.29;46.54]	***<0.001***
Current smokers at Visit 1 and Visit 2	31.77% [-10.17;77.85]	-13.38% [-77.77;74.72]	***0.02***
Former smokers at Visit 1 and Visit 2	11.64% [-13.18;38.83]	-5.07% [-49.99;26.08]	***0.01***
Percent change of gas trapping between Visit 1 and Visit 2 relative to Visit 1			
All	17.49% [-12.98;53.97]	4.13% [-28.55;24.63]	***<0.001***
Current smokers at Visit 1 and Visit 2	30.41% [0.32;74.42]	19.40% [-31.30;43.38]	0.11
Former smokers at Visit 1 and Visit 2	6.19% [-17.44;29.95]	-2.03% [-29.79;13.08]	***0.02***
Percent change of CT-total lung capacity between Visit 1 and Visit 2 relative to Visit 1			
All	0.99% [-5.14;6.14]	-1.46% [-5.27;2.34]	***0.001***
Current smokers at Visit 1 and Visit 2	1.94% [-4.57;7.03]	-0.41% [-4.77;3.78]	0.17
Former smokers at Visit 1 and Visit 2	-0.10% [-5.00;4.98]	-1.91% [-5.61;1.37]	***0.02***

Total emphysema: Percent of CT low attenuation area below -950 Hounsfield units at end-inspiration using Thirona software (% LAA-950); Percent gas trapping is measured at end-exhalation and defined as the percentage of lung voxels with density less than -856 Hounsfield units.

"Percent change between Visit 1 and Visit 2" variables are defined as (Value at Visit 2 - Value at Visit 1) / Value at Visit 1. Negative values indicate decrease of the analyzed variable at Visit 2. Variables are expressed as median and interquartile range (25^{th} to 75^{th} percentile). P-values of ULP *vs* LLP are from Nemenyi tests for pairwise comparisons between ULP and LLP clusters following omnibus Kruskal-Wallis tests. Sample sizes in each subgroup: All (n=437 *(ULP)*; n=394 *(LLP)*); Current smokers at Visit 1 and Visit 2 (n=185 *(ULP)*; n=102 *(LLP)*); Former smokers at Visit 1 and Visit 2 (n=192 *(ULP)*; n=222 *(LLP)*). P-values < 0.05 are bolded and italicized.

e-Table 6: Five-year changes in PRM in upper versus lower lobe predominant emphysema clusters.

	Upper lobe predominant (ULP) cluster	Lower lobe predominant (LLP) cluster	P-value (ULP *vs* LLP)
Change of PRM-percent emphysema between Visit 1 and Visit 2			
All	0.92 [-0.20;4.10]	0.04 [-0.53;2.73]	***<0.001***
Current smokers at Visit 1 and Visit 2	0.94 [0.08;3.42]	-0.01 [-0.20;2.74]	***<0.001***
Former smokers at Visit 1 and Visit 2	0.82 [-0.46;3.97]	0.003 [-0.87;1.73]	***0.01***
Percent change of PRM-percent emphysema between Visit 1 and Visit 2 relative to Visit 1			
All	31.72% [-6.71;85.64]	3.59% [-57.05;64.28]	***<0.001***
Current smokers at Visit 1 and Visit 2	51.35% [8.48;125.36]	-4.96% [-70.05;108.27]	0.08
Former smokers at Visit 1 and Visit 2	20.01% [-14.74;52.20]	0.53% [-55.74;40.09]	***0.02***
Change of PRM-functional small airways disease (PRM-fSAD) between Visit 1 and Visit 2			
All	1.56 [-2.49;6.46]	0.11 [-3.31;4.21]	***0.01***
Current smokers at Visit 1 and Visit 2	2.92 [-0.36;7.58]	1.30 [-1.70;7.09]	0.28
Former smokers at Visit 1 and Visit 2	0.25 [-3.64;4.78]	-0.93 [-4.76;2.59]	0.20
Percent change of PRM-functional small airways disease (PRM-fSAD) between Visit 1 and Visit 2 relative to Visit 1			
All	12.43% [-13.73;43.24]	0.24% [-20.93;24.15]	***0.01***
Current smokers at Visit 1 and Visit 2	20.58% [-4.59;65.25]	11.88% [-18.24;40.86]	0.32
Former smokers at Visit 1 and Visit 2	1.39 [-16.51;26.87]	-7.61 [-22.55;13.66]	0.12

PRM: Parametric Response Mapping.
"Change between Visit 1 and Visit 2" variables are defined as (Value at Visit 2 - Value at Visit 1); "Percent change between Visit 1 and Visit 2" variables are defined as (Value at Visit 2 - Value at Visit 1) / Value at Visit 1. Negative values indicate decrease of the analyzed variable at Visit 2. Variables are expressed as median and interquartile range (25^{th} to 75^{th} percentile). P-values of ULP *vs* LLP are from Nemenyi tests for pairwise comparisons between ULP and LLP clusters following omnibus Kruskal-Wallis tests.
Sample sizes in each subgroup: All (n=437 *(ULP)*; n=394 *(LLP)*); Current smokers at Visit 1 and Visit 2 (n=185 *(ULP)*; n=102 *(LLP)*); Former smokers at Visit 1 and Visit 2 (n=192 *(ULP)*; n=222 *(LLP)*).

e-Table 7. 5-year changes in total emphysema and U/L ratio stratified by Visit 1 BMI:

		Upper-lobe predominant (ULP) cluster	Lower-lobe predominant (LLP) cluster	P-value (ULP *vs* LLP)
Change of total emphysema between Visit 1 and Visit 2				
All		1.05 [-0.50;3.90]	-0.01 [-2.19;2.99]	***<0.001***
BMI: 18.50 - 24.99		1.47 [-0.41;4.47]	1.97 [-0.14;7.28]	0.89
BMI: 25.00 - 29.99		1.07 [-0.68;4.19]	-0.14 [-2.70;2.59]	***0.01***
BMI: 30.00 - 34.99		0.80 [-0.26;4.12]	-0.55 [-3.27;2.12]	***<0.001***
BMI: 35.00 - 39.99		0.65 [0.19;1.98]	-0.51 [-1.79;1.32]	***0.004***
BMI ≥ 40.00		0.12 [-0.63;2.10]	-1.00 [-3.25;0.27]	0.19
Percent change of total emphysema between Visit 1 and Visit 2 relative to Visit 1				
All		23.26% [-11.93;62.26]	-0.12% [-55.29;46.54]	***<0.001***
BMI: 18.50 - 24.99		34.71 [-1.54;78.91]	18.06 [-6.86;58.03]	0.79
BMI: 25.00 - 29.99		15.21 [-19.10;52.67]	-8.54 [-57.01;37.49]	***0.04***
BMI: 30.00 - 34.99		16.65 [-8.01;42.14]	-23.25 [-75.16;46.87]	***0.04***
BMI: 35.00 - 39.99		24.64 [9.12;43.85]	-35.54 [-64.96;35.96]	0.06
BMI ≥ 40.00		7.12 [-54.20;34.64]	-40.80 [-72.71;14.58]	0.52
U/L ratio				
All	Visit 1	4.10 [2.98;7.17]	0.65 [0.49;0.81]	***<0.001***
	Visit 2	3.80 [2.35;6.78]	0.66 [0.46;0.90]	***<0.001***
BMI: 18.50 - 24.99	Visit 1	4.14 [3.23;7.15]	0.58 [0.45;0.75]	***<0.001***
	Visit 2	3.86 [2.30;6.48]	0.59 [0.38;0.73]	***<0.001***
BMI: 25.00 - 29.99	Visit 1	3.79 [2.83;6.17]	0.67 [0.51;0.82]	***<0.001***
	Visit 2	3.14 [2.13;6.06]	0.66 [0.48;0.87]	***<0.001***
BMI: 30.00 - 34.99	Visit 1	5.10 [3.09;6.61]	0.71 [0.55;0.85]	***<0.001***
	Visit 2	4.19 [2.91;8.68]	0.76 [0.49;0.99]	***<0.001***
BMI: 35.00 - 39.99	Visit 1	5.50 [2.93;8.85]	0.62 [0.49;0.72]	***<0.001***
	Visit 2	4.92 [2.80;7.10]	0.78 [0.48;1.03]	***<0.001***
BMI ≥ 40.00	Visit 1	4.63 [2.90;7.55]	0.73 [0.61;0.83]	***<0.001***
	Visit 2	9.12 [2.67;20.91]	0.71 [0.46;1.02]	***<0.001***

Total emphysema: Percent of CT densitometry low attenuation area below -950 Hounsfield units at end-inspiration using Thirona software (% LAA-950); U/L ratio: Ratio of %LAA-950 in both upper lobes to that in both lower lobes.
The subgroups are based on the World Health Organization classification of body mass index (BMI) and obesity.
"Change of total emphysema between Visit 1 and Visit 2" is defined as (Value at Visit 2 - Value at Visit 1). "Percent change between Visit 1 and Visit 2" variable is defined as (Value at Visit 2 - Value

at Visit 1) / Value at Visit 1. Variables are expressed as median and interquartile range (25th to 75th percentile). Negative values indicate decrease of the analyzed variable at Visit 2. P-values of ULP *vs* LLP are from Nemenyi tests for pairwise comparisons between ULP and LLP clusters following omnibus Kruskal-Wallis tests.
Sample sizes in each subgroup: BMI: 18.5-24.99 (n= 157 *(ULP)*; n= 96 *(LLP)*); BMI: 25.00-29.99 (n= 152 *(ULP)*; n= 144 *(LLP)*); BMI: 30.00-34.99 (n= 76 *(ULP)*; n= 82 *(LLP)*); BMI 35.00-39.99 (n= 37 *(ULP)*; n= 39 *(LLP)*); BMI≥40.00 (n= 10 *(ULP)*; n= 27 *(LLP)*). P-values < 0.05 are bolded and italicized.

e-Table 8. Five-year prospective changes in spirometric features, functional measures, and imaging characteristics in upper versus lower-lobe predominant emphysema clusters: Subgroup analyses by GOLD grade.

Upper-lobe predominant cluster *vs.* Lower-lobe predominant cluster	**Sample size**	**Univariate model**		
		Beta-coefficient	**Standard error**	**P-value**
Change of total emphysema between Visit 1 and Visit 2 (% LAA-950 per year)				
Overall	1413	0.19	0.05	***0.001***
PRISm	198	0.24	0.10	***0.02***
GOLD 0	614	0.25	0.06	***<0.001***
GOLD 1-2	409	0.15	0.13	0.23
GOLD 3-4	192	0.19	0.19	0.32
Change of percent gas trapping between Visit 1 and Visit 2 (% per year)				
Overall	1413	0.48	0.21	***0.02***
PRISm	198	0.32	0.49	0.52
GOLD 0	614	0.53	0.20	***0.01***
GOLD 1-2	409	0.35	0.24	0.14
GOLD 3-4	192	0.34	0.31	0.27
Change of CT-total lung capacity between Visit 1 and Visit 2 (% per year)				
Overall	1413	0.03	0.01	***0.001***
PRISm	198	0.04	0.02	0.03
GOLD 0	614	0.01	0.02	0.37
GOLD 1-2	409	0.02	0.01	***0.02***
GOLD 3-4	192	0.02	0.02	0.15
Change of FEV_1 between Visit 1 and Visit 2 (mL/year)				
Overall	1413	-9.00	4.74	0.06
PRISm	198	-4.00	17.14	0.82
GOLD 0	614	2.00	6.73	0.77
GOLD 1-2	409	-17.00	5.21	***0.001***
GOLD 3-4	192	-2.00	8.49	0.81
Change of FEV_1 between Visit 1 and Visit 2 (percent predicted/year)				
Overall	1413	-0.16	0.13	0.22
PRISm	198	0.06	0.63	0.93
GOLD 0	614	-0.09	0.20	0.64
GOLD 1-2	409	-0.44	0.27	0.11
GOLD 3-4	192	0.25	0.29	0.39

GOLD: Global Initiative for Chronic Obstructive Lung Disease; PRISm: Preserved Ratio Impaired Spirometry; FEV_1: Forced expiratory volume in 1 second; Total emphysema: Percent of CT densitometry low attenuation area below -950 Hounsfield units at end-inspiration using Thirona software (% LAA-950); Percent gas trapping is measured at end-exhalation and defined as the percentage of lung voxels with density less than -856 Hounsfield units.

The longitudinal outcomes of change per year for each individual are defined as (Value at Visit 2 - Value at Visit 1) / (Time between Visit 1 and Visit 2). Negative values represent a lower value of the outcome measured at Visit 2. Median quantile regression models were used for all the variables except MMRC where ordinal logistic regression model was used. Lower-lobe predominant cluster is the reference cluster. P-values < 0.05 are bolded and italicized.

e-Table 9. Five-year prospective changes in spirometric features, functional measures, and imaging characteristics in upper versus lower-lobe predominant emphysema clusters: Subgroup analyses by total emphysema severity.

Upper-lobe predominant cluster ***vs.*** **Lower-lobe predominant cluster**	**Sample size**	**Univariate model**		
		Beta-coefficient	**Standard error**	**P-value**
Change of total emphysema between Visit 1 and Visit 2 (% LAA-950 per year)				
Overall	1413	0.19	0.05	***0.001***
Total emphysema ≤5%	991	0.11	0.03	***0.001***
Total emphysema > 5%	422	0.40	0.15	***0.01***
Change of percent gas trapping between Visit 1 and Visit 2 (% per year)				
Overall	1413	0.48	0.21	***0.02***
Total emphysema ≤5%	991	0.27	0.20	0.19
Total emphysema > 5%	422	0.78	0.28	***0.01***
Change of CT-total lung capacity between Visit 1 and Visit 2 (% per year)				
Overall	1413	0.03	0.01	***0.001***
Total emphysema ≤5%	991	0.03	0.01	***0.004***
Total emphysema > 5%	422	0.02	0.01	***0.01***
Change of FEV_1 between Visit 1 and Visit 2 (mL/year)				
Overall	1413	-9.00	4.74	0.06
Total emphysema ≤5%	991	-3.00	6.01	0.62
Total emphysema > 5%	422	-14.00	5.36	***0.01***
Change of FEV_1 between Visit 1 and Visit 2 (percent predicted/year)				
Overall	1413	-0.16	0.13	0.22
Total emphysema ≤5%	991	-0.12	0.25	0.63
Total emphysema > 5%	422	-0.30	0.24	0.21

FEV_1: Forced expiratory volume in 1 second; MMRC: Modified Medical Research Council; SGRQ: St. George's Respiratory Questionnaire; Total emphysema: Percent of CT densitometry low attenuation area below -950 Hounsfield units at end-inspiration using Thirona software (% LAA-950); Percent gas trapping is measured at end-exhalation and defined as the percentage of lung voxels with density less than -856 Hounsfield units.

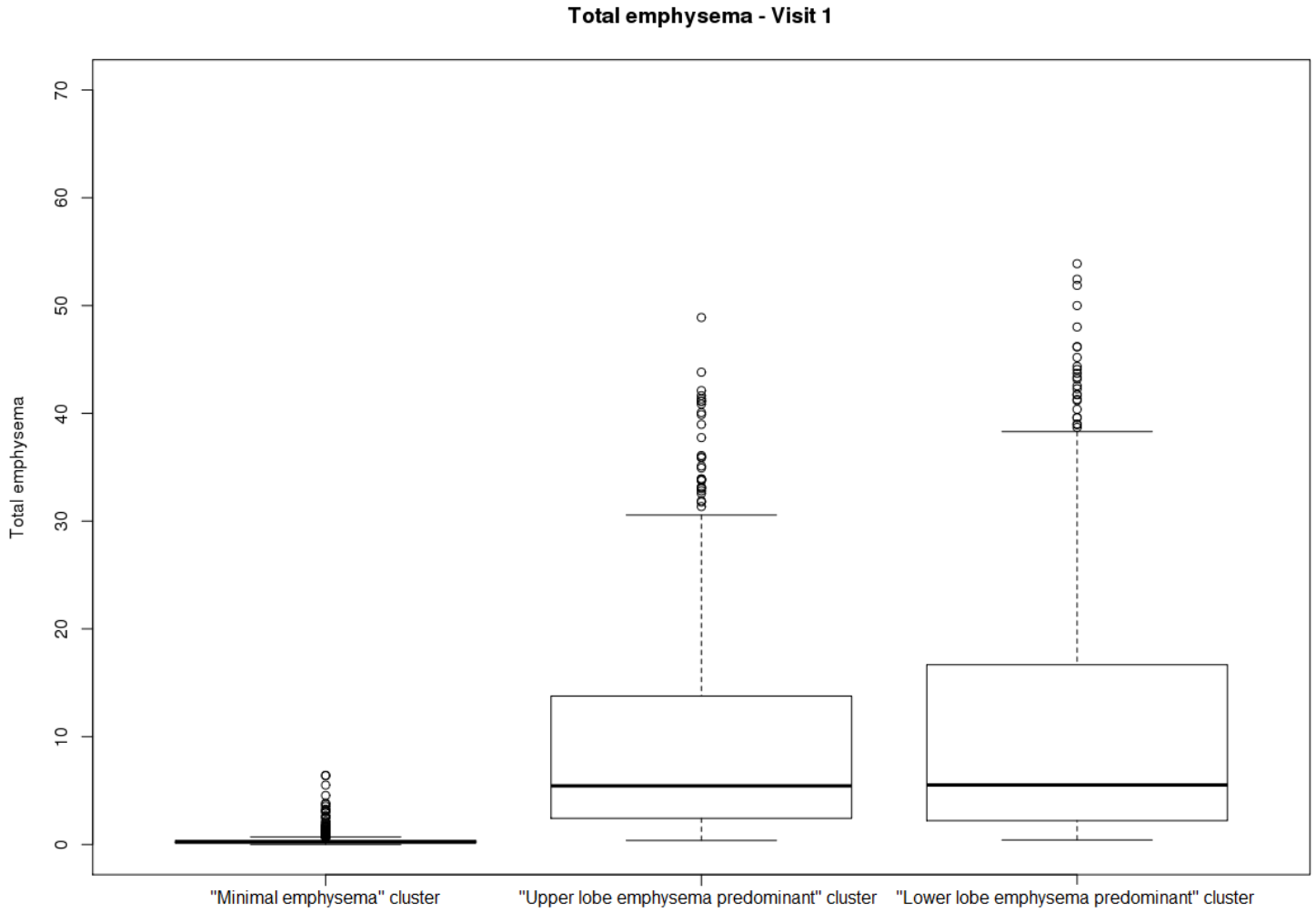
The longitudinal outcomes of change per year for each individual are defined as (Value at Visit 2 - Value at Visit 1) / (Time between Visit 1 and Visit 2). Negative values represent a lower value of the outcome measured at Visit 2. Median quantile regression models were used for all the variables except MMRC where ordinal logistic regression model was used. Lower-lobe predominant cluster is the reference cluster. P-values < 0.05 are bolded and italicized.

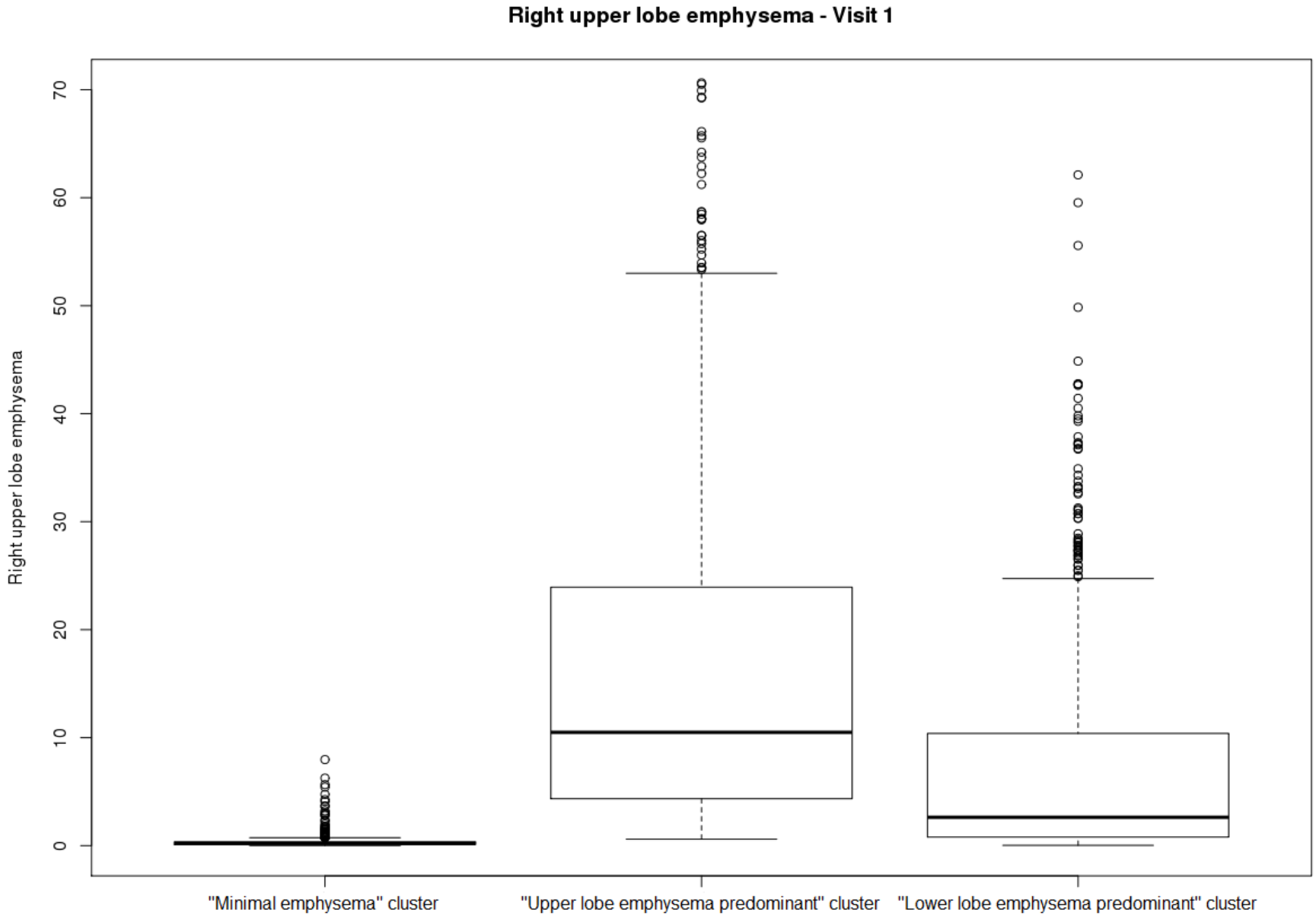
e-Table 10. Most significant variants *(P-values<5x10^{-6})* from the genome-wide association studies of COPDGene non-Hispanic white (COPDGeneNHW) and African American (COPDGeneAA) subjects comparing the upper and lower lobe emphysema predominant clusters.

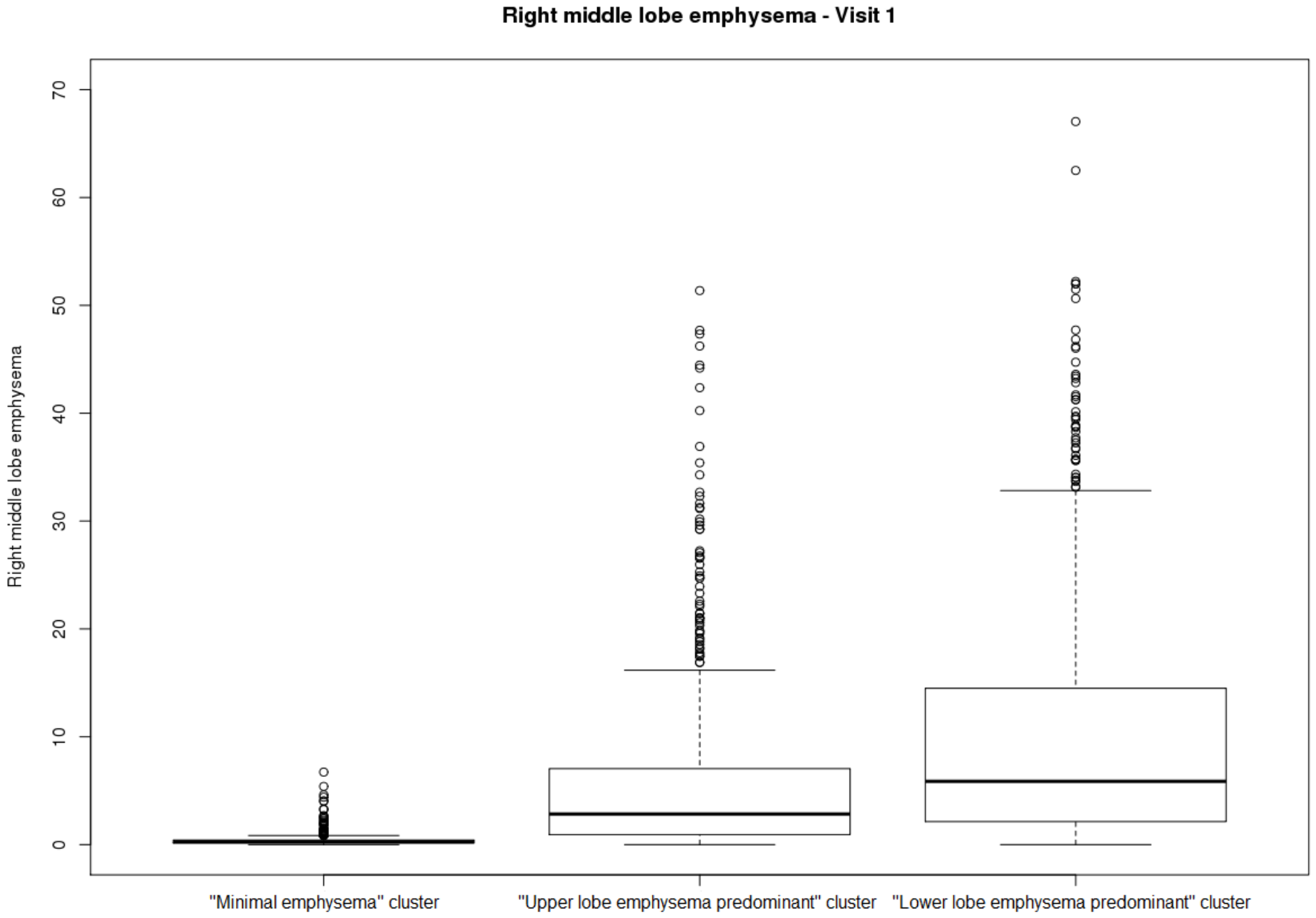
						Meta-analysis			COPDGeneNHW			COPDGeneAA		
SNP	**Locus**	**Nearest Gene**	**Effect Allele**	**Effect Allele Frequency NHW**	**Effect Allele Frequency AA**	**OR**	**SE**	**P-value**	**OR**	**SE**	**P-value**	**OR**	**SE**	**P-value**
rs10904076	10p15	*PFKP*	A	0.80	0.74	0.60	1.10	$2.8x10^{-7}$	0.65	1.12	0.0001	0.48	1.22	0.0003
rs9571261	13q21	*KLHL1*	C	0.92	0.97	2.12	1.17	$1.2x10^{-6}$	2.11	1.17	$2.9x10^{-6}$	2.20	1.81	0.18
rs2796423	1q32	*SOX13*	A	0.86	0.52	1.61	1.10	$1.8x10^{-6}$	1.37	1.13	0.009	2.24	1.19	$3.7x10^{-6}$
rs7757142	6q24	*PLAGL1*	T	0.20	0.11	1.61	1.11	$3.0x10^{-6}$	1.82	1.12	$9.8x10^{-8}$	0.91	1.27	0.69
rs1991095	9q21	*ALDH1A1*	A	0.05	0.04	2.28	1.19	$3.5x10^{-6}$	2.53	1.22	$2.5x10^{-6}$	1.46	1.51	0.36
rs12271406	11q14	*FAM181B*	T	0.98	0.87	2.71	1.24	$3.8x10^{-6}$	3.57	1.43	0.0003	2.31	1.31	0.002
rs8043894	16p13	*DNAJA3*	T	0.95	0.84	0.43	1.20	$3.8x10^{-6}$	0.55	1.25	0.007	0.27	1.36	$2.9x10^{-5}$
rs61931524	12q14	*C12orf56*	T	0.77	0.57	1.53	1.10	$4.0x10^{-6}$	1.61	1.12	$1.5x10^{-5}$	1.35	1.18	0.07

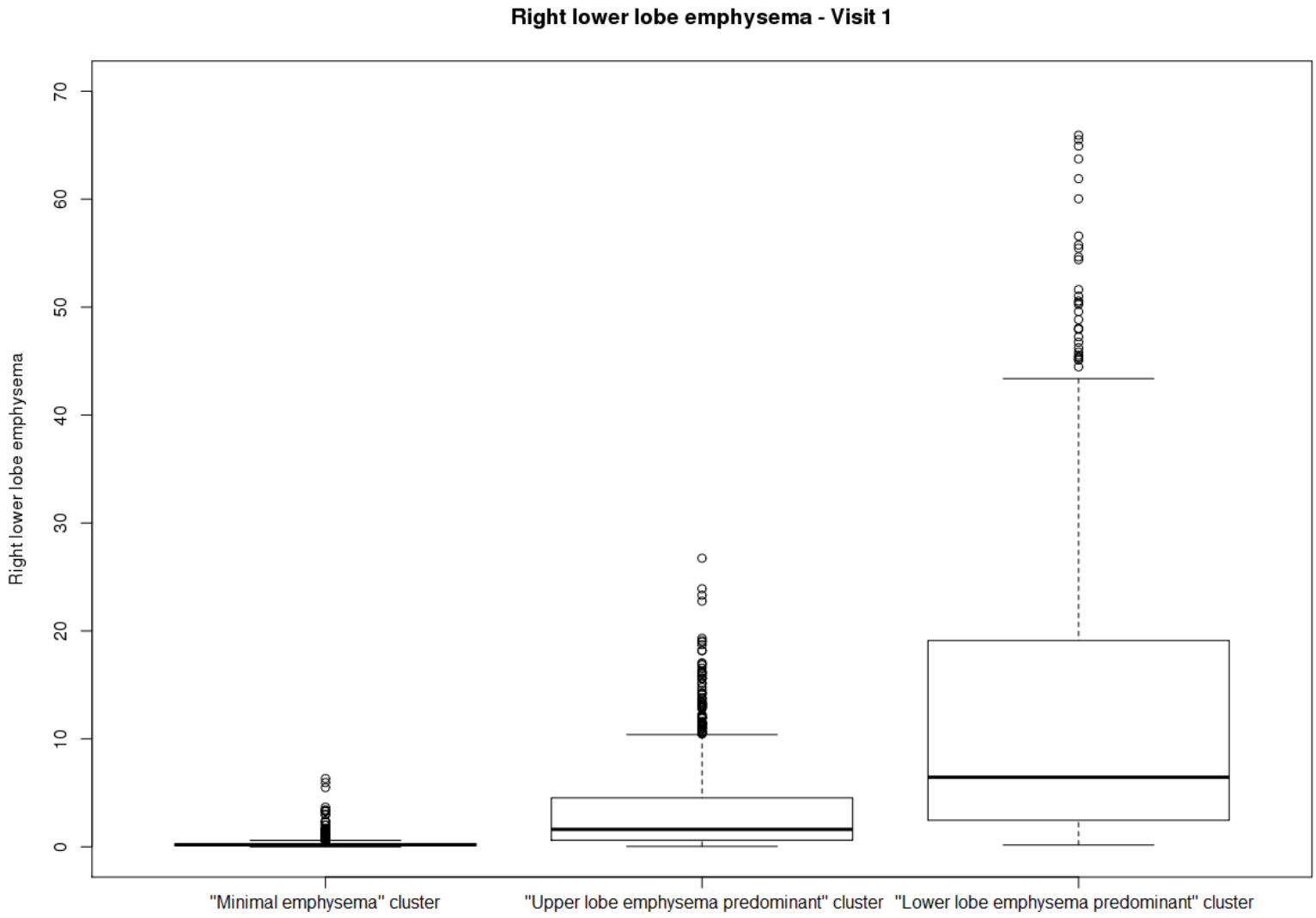
The analyses were adjusted for pack-years of smoking, current smoking, genetic ancestry, and sex. OR: Odds ratio with the upper lobe predominant group specified as the reference group; SE: Standard error.

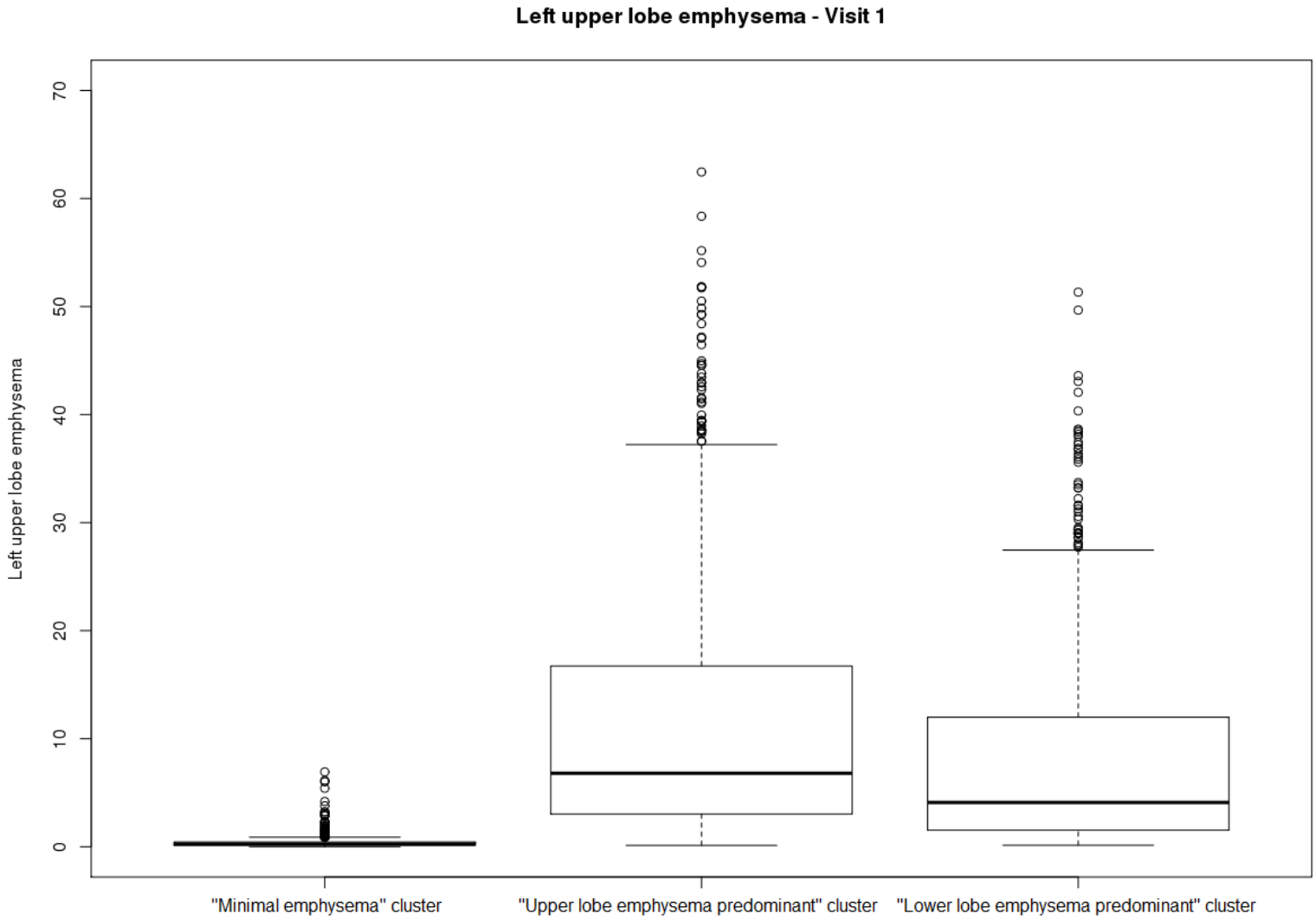
e-Figure 1. Box-and-whisker plots of total and lobar emphysema (% LAA-950) at Visit 1 by cluster ("minimal emphysema", "upper lobe emphysema predominant" and "lower lobe emphysema predominant" clusters).

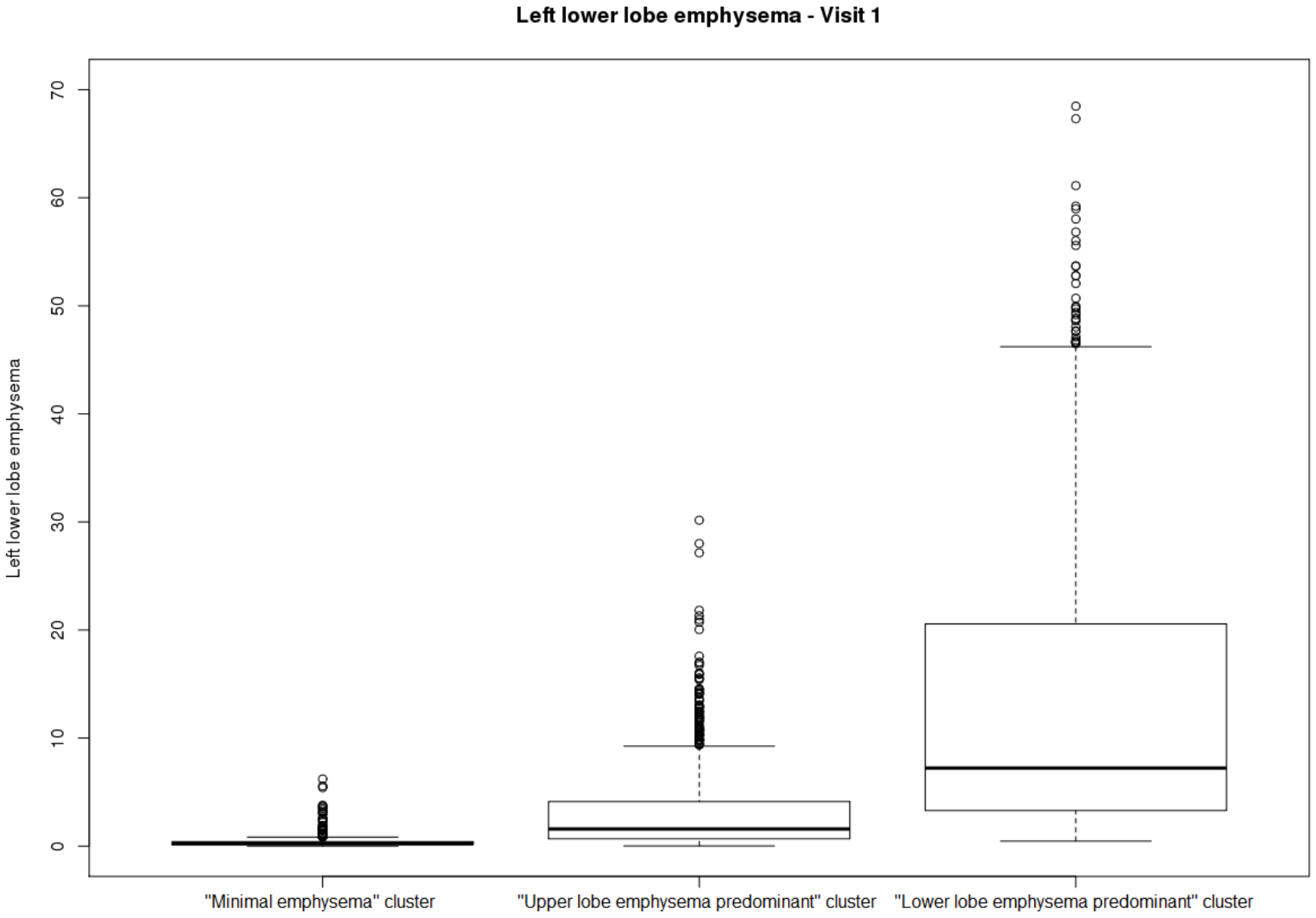




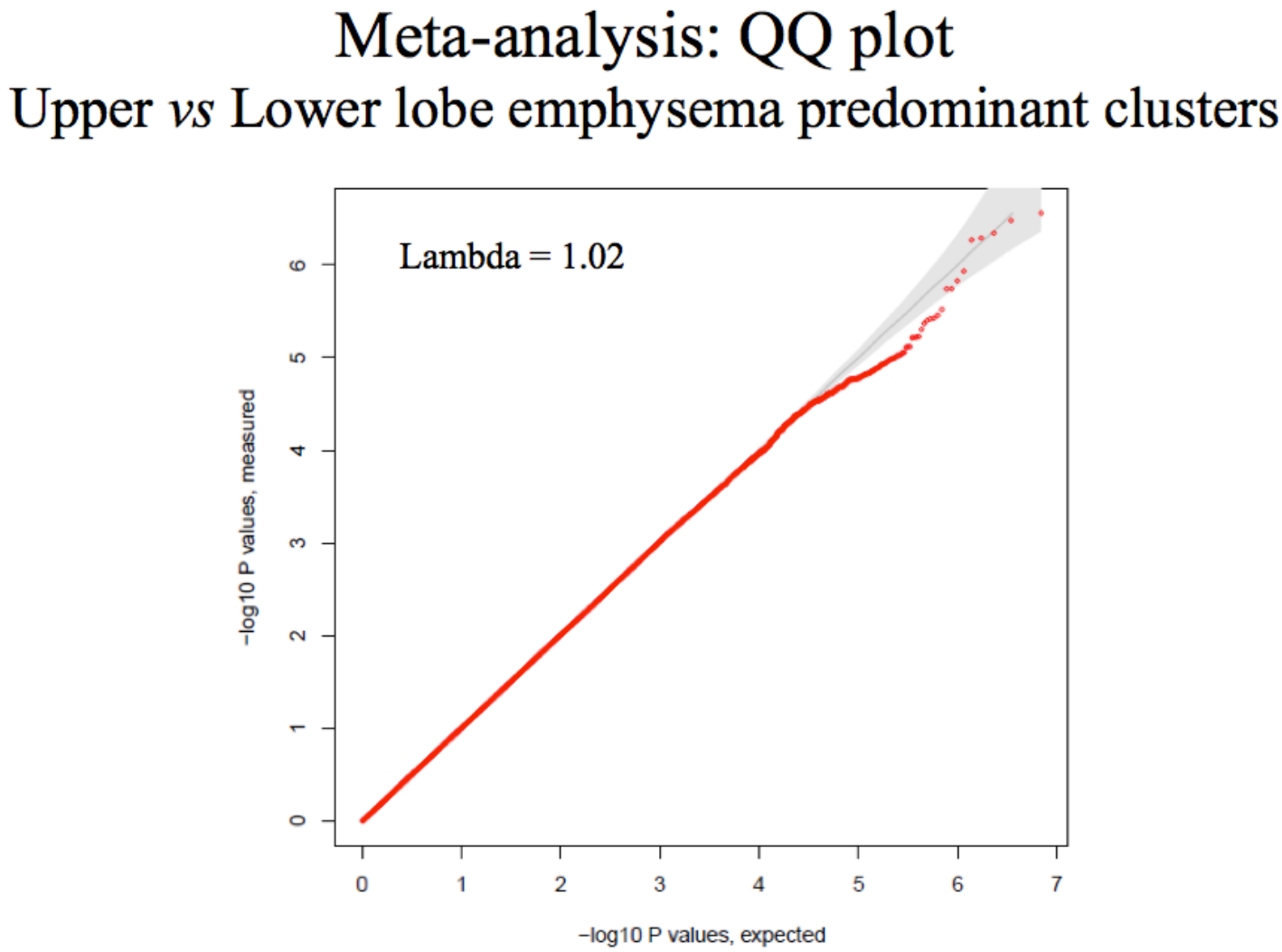








e-Figure 2. Quantile-Quantite (QQ) plot for GWAS meta-analysis performed in COPDGene Non-Hispanic White (NHW) and COPDGene African-American (AA) to test for SNP associations with upper versus lower lobe emphysema distribution clusters.



170793