Genome-Wide Association Analysis of Body Mass in Chronic Obstructive Pulmonary Disease

Emily S. Wan, Michael H. Cho, Nadia Boutaoui, Barbara J. Klanderman, Jody S. Sylvia, John P. Ziniti, Sungho Won, Christoph Lange, Sreekumar G. Pillai, Wayne H. Anderson, Xiangyang Kong, David A. Lomas, Per S. Bakke, Amund Gulsvik, Elizabeth A. Regan, James R. Murphy, Barry J. Make, James D. Crapo, Emiel F. Wouters, Bartolome R. Celli, Edwin K. Silverman, Dawn L. DeMeo, on behalf of the ECLIPSE, Norway-Bergen cohort, NETT, and COPD Gene investigators

Online Data Supplement

Materials and Methods

Cohorts and variables

Evaluation of COPD Longitudinally to Identify Predictive Surrogate End-points (ECLIPSE): COPD subjects were recruited based on spirometry criteria outlined in the main text. Reference values for spirometry were derived from the European Community for Coal and Steel standards(1, 2). Subjects with a history of other known respiratory diseases, including severe alpha-1 antitrypsin (A1AT) deficiency were excluded. Height, weight, and spirometry as measured by study staff on Visit 1 were used in the current analysis. Single frequency (50Hz) bioimpedance measurements were obtained using the Bodystat® 1500 as outlined below. Self-reported history of hypertension and diabetes was obtained from questionnaire data administered at enrollment. Subjects underwent low-dose volumetric chest computed tomography (CT) scan (120 kV peak, 40 mA and 1.00 or 1.25-mm slice thickness) at full inspiration. Quantitative evaluation of emphysema at -950HU was performed using Pulmonary Workstation 2.0 (VIDA Diagnostics, Iowa City, IA, USA). All CT scans were evaluated at the central imaging unit at the University of British Columbia in Vancouver.

Norway-Bergen cohort:
COPD subjects were recruited based on spirometry criteria outlined in the main text. Reference equations for spirometry were derived from Johannessen et al(3). Subjects with alpha-1 antitrypsin genotypes PiZZ, PiZ-Null, Pi Null-Null, and Pi SZ were excluded. Height, weight, bioimpedance, and post-bronchodilator spirometry at enrollment were utilized in this analysis. Computed tomography scans were obtained in a subset of subjects using a GE LightSpeed Ultra CT scanner (120 kVp, 200 mA) employing 1 mm slice thickness at 20mm intervals at full inspiration(20). Quantitative densitometric assessment of emphysema was performed using a density mask of -950 Hounsfield units in subjects with genotype data. Self-reported history of hypertension and diabetes was obtained from questionnaire data obtained at enrollment.

National Emphysema Treatment Trial (NETT) Genetics Ancillary Study cohort:
COPD subjects were recruited based on spirometry criteria outlined in the main text. Reference equations for spirometry were derived from Crapo et al(4). Height, weight, body mass index, and spirometry

measurements following completion of pulmonary rehabilitation were used in the current analysis. CT scan images were obtained at full inspiration using a minimum of 200mA with a slice thickness of 1-5 mm at 20 mm intervals. Quantitative assessment of emphysema at -950 HU was analyzed using Pulmonary Analysis Software Suite (PASS, Iowa City, IA, USA).
Co-morbidities such as hypertension and diabetes were based upon reported use of anti-hypertensive medications and insulin.

COPDGene cohort:
COPD subjects were recruited based on spirometry criteria outlined in the main text. Reference equations for spirometry were derived from Hankinson et al(5). Height, weight, and post-bronchodilator spirometry were obtained at initial enrollment. CT scans were obtained at full inspiration using Siemens, General Electric, or Philips chest CT scanners with at least 16 detectors at 120 kV, 200 mA, with sub-millimeter contiguous slices. Quantitative assessment of emphysema at -950 HU was obtained using Slicer version 2.0 (www.slicer.org) software. Self-reported history of hypertension or diabetes was assessed by questionnaire.

Bioimpedance Measurements and Fat Free Mass calculation
Single frequency (50 Hz) bioimpedance measurements were made in the ECLIPSE and Norway-Bergen cohorts according to the following protocol. Subjects were evaluated in the supine position after a minimum 3 hour fast and after voiding to empty the bladder. Two electrodes were placed on the right hand and right foot and raw bioimpedance measurements were made using the Bodystat® 1500. Fat free mass index in both the ECLIPSE and the Norway cohorts was derived using the equations from Steiner et al(6).

Genotyping and Quality Control
In the three cohorts utilized in the primary analysis (ECLIPSE, Norway-Bergen, NETT), genome-wide SNP genotype data was obtained using the Illumina platform (Illumina, Inc., San Diego, CA, USA). Genotyping of the NETT cohort was performed on the Illumina Human 610-Quad BeadChip at the Channing Laboratory. Genotyping of the Norway and ECLIPSE cohorts was performed on the Illumina HumanHap 550 (1, V3, and Duo) and HumanHap550 V3 respectively. Raw two channel intensities from the ECLIPSE and Norway cohorts were imported into Beadstudio (Illumina, Inc; San Diego, CA, USA) and reclustering of SNPs with a cluster separation <0.3 was performed by a core group of operators.

Subjects who failed the following quality control measures were removed: genotyping call rate of <95%, discordance rate >1% on replicates, relatedness and inbreeding, sex discordance, and lack of genotypic or phenotypic data. For the Norway cohort, where genotyping was conducted on multiple platforms, the genotyping call rate was determined based upon a shared set of markers. For the NETT cohort, discordance rates of >5% compared to previous (historical) genotyping calls were used to exclude subjects. Markers that failed the following quality control measures were removed: missingness >5%, monoallelic or singleton SNPs, and extreme deviation from Hardy Weinberg equilibrium (p-value $<10^{-8}$). The overall genotyping success rate using the Taqman assay (Applied Biosystems, Carlsbad, CA, USA) was 99.4 %.

Population Stratification and Association Analysis

Principal components to adjust for population stratification in each cohort were generated from a subset of autosomal SNP markers. SNPs with a minor allele frequency >0.01 and HWE p-value of $>10^{-3}$ were used to generate a pruned dataset using an r^2 of 0.1, window size of 1500, and step size of 150 in PLINK(E7). Principal components analysis and outlier removal were performed using EIGENSTRAT software(E8). A summary of the number of subjects removed at each step of data cleaning and as principal component outliers is summarized is Supplementary Table E1.

Genome wide association studies were performed in each of the three primary cohorts. SNPs with a minor allele frequency <0.01 in each cohort were excluded from the corresponding individual cohort analysis. Fat free mass index was not available in NETT. Four subjects in the Norway cohort and two subjects in the ECLIPSE cohort had non-physiological FFMI values which were set to missing.

References

E1. Standardized lung function testing. Report working party. *Bull Eur Physiopathol Respir* 1983;19 Suppl 5:1-95.

E2. Quanjer PH, Tammeling GJ, Cotes JE, Pedersen OF, Peslin R, Yernault JC. Lung volumes and forced ventilatory flows. Report working party standardization of lung function tests, european community for steel and coal. Official statement of the european respiratory society. *Eur Respir J Suppl* 1993;16:5-40.

E3. Johannessen A, Lehmann S, Omenaas ER, Eide GE, Bakke PS, Gulsvik A. Post-bronchodilator spirometry reference values in adults and implications for disease management. *Am J Respir Crit Care Med* 2006;173:1316-1325.

E4. Crapo RO, Morris AH, Gardner RM. Reference spirometric values using techniques and equipment that meet ats recommendations. *Am Rev Respir Dis* 1981;123:659-664.

E5. 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;159:179-187.

E6. Steiner MC, Barton RL, Singh SJ, Morgan MD. Bedside methods versus dual energy x-ray absorptiometry for body composition measurement in copd. *Eur Respir J* 2002;19:626-631.

E7. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC. Plink: A tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007;81:559-575.

E8. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* 2006;38:904-909.

Figure Legend

Figure E1 – Linkage disequilibrium (LD) structure (R-squared values) of a region of the first intron of the fat mass and obesity associated (FTO) gene in the HapMap CEU population. The most highly associated SNPs from our analysis are in strong LD with each other and with rs9939609.

Tables-

	ECLIPSE	**Norway**	**NETT**
Genotyped subjects	1887	933	385
Concordance rate in replicates (mean/median)	99.9%/>99.9%	>99.9%/>99.9%	99.8%/99.9%
Poor reproducibility	1	0	0
Missingness >5% or missing genotype information	11	17	5
Relatedness	51	50	0
Gender mismatch	12	2	3
Non-white race	48	0	0
Passing subjects	**1764**	**864**	**377**
Subjects removed as principal component outliers	30	13	12
Subjects included in final analysis	**1734**	**851**	**365**

Table E1 - Summary of subject cleaning in three primary cohorts.

	ECLIPSE		Norway		NETT		COPDGene	
Gender	Male	Female	Male	Female	Male	Female	Male	Female
BMI	26.9 (15.4-53.6)	26.4 (14.5-57.7)	25.5 (11.7-44.3)	25.1 (13.4-58.9)	25.3 (17.9-31.1)	24.4 (17.2-32.3)	28.7 (15.6-48.2)	27.5 (16.3-48.5)
FFMI	18.1 (11-30.7)	15.8 (10.7-28.2)	17.6 (10.9-27)	15.6 (10.2-28.1)	N/A	N/A	N/A	N/A
Pack-years	53.9 (10-220)	43.2 (6-161)	35.6 (3.2-130)	26.7 (3-125)	71.4 (12-260)	56.4 (16-117)	59.6 (11-237.6)	49.9 (10-161.7)

Table E2 - **Demographic differences by gender**. All data are presented as mean (range). All analyses were restricted to within-group comparisons, results summarized in table were statistically significant at p-value <0.05. The Student's t-test and Wilcoxon rank sum test were used to examine normal and non-normal data respectively.

Cohort	Rank	SNP	p-value	Type	Closest gene
ECLIPSE	1	rs9864349	1.13×10^{-6}	Intergenic	AC107623.1
	2	rs7875754	4.91×10^{-6}	Intronic	*EPB41L4B*
	3	rs952462	6.23×10^{-6}	Intronic	*SCN9A*
	4	rs2904266	6.87×10^{-6}	Upstream	AL034429.2
	5	rs4793325	8.07×10^{-6}	Intergenic	AC005181.1
	6	rs6020502	1.01×10^{-5}	Intergenic	AL034429.2
	7	rs3817193	1.21×10^{-5}	Intronic	*ZSCAN2*
	8	rs10485478	1.52×10^{-5}	Upstream	RP4-640H8.1
	9	rs1528481	1.53×10^{-5}	Intronic	*SCN9A*
	10	rs8050136	1.96×10^{-5}	Intronic	*FTO*
Norway	1	rs2086731	2.26×10^{-6}	Intergenic	*TECRL*
	2	rs1519980	4.18×10^{-6}	Intergenic	AC078940.1
	3	rs10126555	9.30×10^{-6}	Intronic	*GPR64*
	4	rs6534369	1.08×10^{-5}	Synonymous	*SPATA5*
	5	rs5955518	1.19×10^{-5}	Intronic	*GPR64*
	6	rs2208606	1.91×10^{-5}	Intergenic	AL590290.1
	7	rs17024477	1.93×10^{-5}	Intergenic	AC068538
	8	rs1648395	1.98×10^{-5}	Intergenic	*KIAA1239*
	9	rs17115765	2.21×10^{-5}	Intronic	*DAB1*
	10	rs7223919	2.76×10^{-5}	Intronic	*MS12*
NETT	1	rs10273260	3.01×10^{-6}	Downstream	AC005537.1
	2	rs10280281	3.60×10^{-6}	Intronic	AC004741.1
	3	rs880976	3.62×10^{-6}	In non-coding gene	AC005537.2
	4	rs6923737	6.63×10^{-6}	Intronic	ALO7934.3
	5	rs1931324	8.51×10^{-6}	Upstream	AL354983.1
	6	rs1362382	1.36×10^{-5}	Intergenic	AC007333.1
	7	rs4885124	1.52×10^{-5}	Intronic	*KLF12*
	8	rs423477	1.58×10^{-5}	Intergenic	*SLC44A1*
	9	rs9573321	1.60×10^{-5}	Intronic	*KLF12*
	10	rs2286234	1.67×10^{-5}	Intergenic	*GLI3*

Supplementary Table E3 - **Top association analysis results within individual cohorts for BMI.** Annotation performed on WGA Viewer (v 1.26G, 2009). Models used in each cohort were adjusted for sex and significant principal components for genetic ancestry. The ranks of SNP rs8050136 in the individual cohort analyses for Norway and NETT were 968 and 360288 respectively.

Gene	Analysis	Top SNP	p-value
TNF	BMI	rs6929796	3.1×10^{-2}
	FFMI	rs2734583	8.77×10^{-2}
IL1B	BMI	rs4849124	1.94×10^{-2}
	FFMI	rs12621220	1.03×10^{-2}
IL6	BMI	rs12055945	6×10^{-4}
	FFMI	rs12055945	2.1×10^{-3}

Supplementary Table E4 - p-values from the BMI and FFMI combined analysis for previously reported candidate genes (+/- 50 kb) in COPD-related cachexia

rs8054237
rs8050136
rs4783819
rs8051591
rs4783820
rs9935401
rs3751812
rs3751813
rs9936385
rs11075989
rs11075990
rs9939609
rs12597786
rs7201850
rs7185735
rs9931164
rs9941349
Block 1 (9 kb)
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
11
96