Genome-wide association analysis of blood biomarkers in COPD

Deog Kyeom Kim, Michael H. Cho, Craig P. Hersh, David A. Lomas, Bruce E. Miller, Xiangyang Kong, Per Bakke, Amund Gulsvik, Alvar Agusti, Emiel Wouters, Bartolome Celli, Harvey Coxson, Jørgen Vestbo, William MacNee, Julie C. Yates, Stephen Rennard, Augusto Litonjua, Weiliang Qiu, Terri H. Beaty, James D. Crapo, John H. Riley, Ruth Tal-Singer, and Edwin K. Silverman on behalf of the ECLIPSE, ICGN, and COPDGene Investigators

Online Data Supplement

Materials and Methods

Study populations

In the Norway-Bergen cohort, COPD cases were included with post-bronchodilator $FEV_1/FVC < 0.7$, post-bronchodilator $FEV_1 < 80\%$ predicted, and ≥ 2.5 pack years of smoking. The details of the NETT-NAS studies have been published elsewhere.(1-3) COPD subjects with FEV1≤45% predicted and bilateral emphysema on chest CT were included in the National Emphysema Treatment Trial (NETT) study, and control subjects who had normal spirometry and at least 10 pack-years of cigarette smoking history were included in the Normative Aging Study (NAS) cohort.

The International COPD Genetics Network (ICGN) recruited COPD cases as probands, and included their siblings and available parents. Probands aged 45-65 years had post-bronchodilator FEV1< 60% predicted, FEV1/VC ratio < 90% predicted, and smoking history ≥ 5 pack-years.

The COPDGene Study (www.COPDGene.org) analysis included COPD cases with GOLD stage II or higher ($FEV_1 < 80\%$ predicted and $FEV_1/FVC < 0.7$) and smoking controls who were between the ages of 45 and 80 years, with normal spirometry and at least 10-pack-years of smoking history.

Measurement of biomarkers

Briefly, whole blood was collected into vacutainer tubes at the beginning of the study. Serum was prepared by centrifugation at 1500 g for 10–15 min. The serum was collected and stored at -

80 degrees C until analyzed. Serum CC16 and SP-D were measured by operators who were blinded to an individual's lung disease using a colorimetric sandwich immunoassay method (BioVendor GmbH, Heidelberg, Germany) according to the manufacturer's instructions. Serum samples were diluted 5 to 20-fold with the dilution buffer supplied by the manufacturer. The concentration of CC16 was determined by comparison with a standard curve prepared with known concentrations of CC16. The concentration of SP-D in the diluted samples was interpolated from the standard curve of recombinant human SP-D (molecular mass 41 kDa) and then corrected for the dilution factor.

A high sensitivity, sandwich enzyme-linked immunoassay (SearchLight Protein Array Technology, Aushon Biosystems, Inc., Billerica, MA USA) was used to measure CRP. Serum samples were diluted 500- to 10,000-fold for analysis. The lower limit of quantification was 6 ng/ml. Serum concentrations of IL-6 and IL-8 were determined by validated multiplexed immunoassays (SearchLight Array Technology, Thermo Fisher Scientific, Rockford, IL, USA). The limits of quantification for IL-6 and IL-8 were 0.4 pg/ml, and 0.8 pg/ml respectively. TNF-α and fibrinogen were also measured using validated immunoassays (4).

Sputum induction, RNA Isolation, and Microarray analysis

Sputum induction was performed with standard methods as previously described (5-7). Briefly, a cell pellet was acquired after processing the induced sputum with 0.1% DTT on ice in a ratio of 4:1. The cell pellet was re-suspended in cold PBS so that a cell count could be performed and a cytospin slide prepared for differential count. Cytospin preparations were air dried, fixed with methanol and stained with Rapi-diff (Triangle, Skelmersdale, UK). Five hundred leukocytes

were counted by two independent readers at a central laboratory and the results expressed as a percentage of the total leukocyte count and a total cell number/ml. RNA was extracted from sputum pellets using TRIzol reagent (Invitrogen, Paisley, UK) and gene expression profiling was performed using the Affymetrix Human U133 Plus2 array (Affymetrix, Santa Clara, CA, USA) following standard procedures (7-8). After scanning arrays using a GeneChip Scanner 3000, fluorescence intensity was obtained by using GeneChip Operating Software (Affymetrix, Santa Clara, CA, USA). Standard MAS5.0 Affymetrix quality control criteria were examined to determine the quality of the GeneChip data (6-8)

Statistical analysis

To minimize the effects of confounders, we adjusted biomarker GWAS and subsequent association analyses for some covariates. Biomarker GWAS were adjusted for covariates including age, sex, amount of smoking in pack-years, current smoking status, and principal components for genetic ancestry produced by a modified EIGENSTRAT method (9). To select the most appropriate additional covariates crossing all biomarkers, multivariate analyses were done for each biomarker adjusting for well-known confounders in COPD (Table E1), and the above four variables were the most consistently significant variables associated with each biomarker.

These four covariates were adjusted in each linear model for all biomarkers in GWAS analyses.

To assess the relationships between candidate SNPs and level of mRNA expression, linear regression models were used adjusting for covariates including age, sex, total amount of smoking in pack-years, and current smoking status.

For testing the association of COPD affection status and biomarker GWAS SNPs in the combined case-control cohort, a logistic regression model was used with adjustment for age, pack-years of smoking, and principal components for genetic ancestry. All of the COPD cases from NETT, ECLIPSE, and Norway/GenKOLS were combined into one group and compared to all of the smoking controls from NAS, ECLIPSE, and Norway/GenKOLS.

In the ICGN pedigrees, family-based association analysis for weighted COPD affection status was performed using Golden Helix PBAT (http://www.goldenhelix.com/SNP_Variation/PBAT/index.html) (10). Adjusting for age, sex, amount of smoking in pack-years, and current smoking status, the association of candidate SNPs with COPD affection status in ICGN was tested.

In the COPDGene data, as different genotyping chips from Illumina (San Diego, CA, USA) were used in ECLIPSE and COPDGene (11-13), genotype imputation was performed using MaCH 1.0.16 (14) using 100 rounds of iteration to estimate model parameters and CEU samples from HapMap2 (15) and the 1000 Genomes Project (12) (phased CEU data, March 2010) as reference populations. The details of this imputation process were described previously (16). Imputed genotypes were also analyzed using SNP dosage data in PLINK with adjustment for age, sex, body mass index, amount of smoking in pack-years, and current smoking status to test the association with COPD susceptibility.

References

1. Rationale and design of the national emphysema treatment trial: A prospective randomized trial of lung volume reduction surgery. The national emphysema treatment trial research group. *Chest* 1999;116:1750-1761.

2. Fishman A, Martinez F, Naunheim K, Piantadosi S, Wise R, Ries A, Weinmann G, Wood DE. A randomized trial comparing lung-volume-reduction surgery with medical therapy for severe emphysema. *N Engl J Med* 2003;348:2059-2073.

3. Pillai SG, Ge D, Zhu G, Kong X, Shianna KV, Need AC, Feng S, Hersh CP, Bakke P, Gulsvik A, Ruppert A, Lodrup Carlsen KC, Roses A, Anderson W, Rennard SI, Lomas DA, Silverman EK, Goldstein DB. A genome-wide association study in chronic obstructive pulmonary disease (copd): Identification of two major susceptibility loci. *PLoS Genet* 2009;5:e1000421.

4. Hanania NA, Mullerova H, Locantore NW, Vestbo J, Watkins ML, Wouters EF, Rennard SI, Sharafkhaneh A. Determinants of depression in the eclipse chronic obstructive pulmonary disease cohort. *Am J Respir Crit Care Med* 2011;183:604-611.

5. Pizzichini E, Pizzichini MM, Efthimiadis A, Evans S, Morris MM, Squillace D, Gleich GJ, Dolovich J, Hargreave FE. Indices of airway inflammation in induced sputum: Reproducibility and validity of cell and fluid-phase measurements. *Am J Respir Crit Care Med* 1996;154:308-317.

6. Singh D, Edwards L, Tal-Singer R, Rennard S. Sputum neutrophils as a biomarker in copd: Findings from the eclipse study. *Respir Res* 2010;11:77.

7. Singh D, Fox SM, Tal-Singer R, Plumb J, Bates S, Broad P, Riley JH, Celli B. Induced sputum genes associated with spirometric and radiological disease severity in copd ex-smokers. *Thorax* 2011;66:489-495.

8. Qiu W, Cho MH, Riley JH, Anderson WH, Singh D, Bakke P, Gulsvik A, Litonjua AA, Lomas DA, Crapo JD, Beaty TH, Celli BR, Rennard S, Tal-Singer R, Fox SM, Silverman EK, Hersh CP. Genetics of sputum gene expression in chronic obstructive pulmonary disease. *PLoS One* 2011;6:e24395.

9. 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.

10. Lange C, DeMeo D, Silverman EK, Weiss ST, Laird NM. Pbat: Tools for family-based association studies. *Am J Hum Genet* 2004;74:367-369.

11. The international hapmap project. *Nature* 2003;426:789-796.

12. A map of human genome variation from population-scale sequencing. *Nature* 2010;467:1061-1073.

13. Siedlinski M, Cho MH, Bakke P, Gulsvik A, Lomas DA, Anderson W, Kong X, Rennard SI, Beaty TH, Hokanson JE, Crapo JD, Silverman EK. Genome-wide association study of smoking behaviours in patients with copd. *Thorax* 2011.

14. 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;34:816-834.

15. Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, Gibbs RA, Belmont JW, Boudreau A, Hardenbol P, Leal SM, Pasternak S, Wheeler DA, Willis TD, Yu F, Yang H, Zeng C, Gao Y, Hu H, Hu W, Li C, Lin W, Liu S, Pan H, Tang X, Wang J, Wang W, Yu J, Zhang B,

Zhang Q, Zhao H, Zhou J, Gabriel SB, Barry R, Blumenstiel B, Camargo A, Defelice M, Faggart M, Goyette M, Gupta S, Moore J, Nguyen H, Onofrio RC, Parkin M, Roy J, Stahl E, Winchester E, Ziaugra L, Altshuler D, Shen Y, Yao Z, Huang W, Chu X, He Y, Jin L, Liu Y, Sun W, Wang H, Wang Y, Xiong X, Xu L, Waye MM, Tsui SK, Xue H, Wong JT, Galver LM, Fan JB, Gunderson K, Murray SS, Oliphant AR, Chee MS, Montpetit A, Chagnon F, Ferretti V, Leboeuf M, Olivier JF, Phillips MS, Roumy S, Sallee C, Verner A, Hudson TJ, Kwok PY, Cai D, Koboldt DC, Miller RD, Pawlikowska L, Taillon-Miller P, Xiao M, Tsui LC, Mak W, Song YQ, Tam PK, Nakamura Y, Kawaguchi T, Kitamoto T, Morizono T, Nagashima A, Ohnishi Y, Sekine A, Tanaka T, Tsunoda T, Deloukas P, Bird CP, Delgado M, Dermitzakis ET, Gwilliam R, Hunt S, Morrison J, Powell D, Stranger BE, Whittaker P, Bentley DR, Daly MJ, de Bakker PI, Barrett J, Chretien YR, Maller J, McCarroll S, Patterson N, Pe'er I, Price A, Purcell S, Richter DJ, Sabeti P, Saxena R, Schaffner SF, Sham PC, Varilly P, Stein LD, Krishnan L, Smith AV, Tello-Ruiz MK, Thorisson GA, Chakravarti A, Chen PE, Cutler DJ, Kashuk CS, Lin S, Abecasis GR, Guan W, Li Y, Munro HM, Qin ZS, Thomas DJ, McVean G, Auton A, Bottolo L, Cardin N, Eyheramendy S, Freeman C, Marchini J, Myers S, Spencer C, Stephens M, Donnelly P, Cardon LR, Clarke G, Evans DM, Morris AP, Weir BS, Mullikin JC, Sherry ST, Feolo M, Skol A, Zhang H, Matsuda I, Fukushima Y, Macer DR, Suda E, Rotimi CN, Adebamowo CA, Ajayi I, Aniagwu T, Marshall PA, Nkwodimmah C, Royal CD, Leppert MF, Dixon M, Peiffer A, Qiu R, Kent A, Kato K, Niikawa N, Adewole IF, Knoppers BM, Foster MW, Clayton EW, Watkin J, Muzny D, Nazareth L, Sodergren E, Weinstock GM, Yakub I, Birren BW, Wilson RK, Fulton LL, Rogers J, Burton J, Carter NP, Clee CM, Griffiths M, Jones MC, McLay K, Plumb RW, Ross MT, Sims SK, Willey DL, Chen Z, Han H, Kang L, Godbout M, Wallenburg JC, L'Archeveque P, Bellemare G, Saeki K, An D, Fu H, Li Q, Wang Z, Wang R, Holden AL,

Brooks LD, McEwen JE, Guyer MS, Wang VO, Peterson JL, Shi M, Spiegel J, Sung LM, Zacharia LF, Collins FS, Kennedy K, Jamieson R, Stewart J. A second generation human haplotype map of over 3.1 million snps. *Nature* 2007;449:851-861.

16. Cho MH, Castaldi PJ, Wan ES, Siedlinski M, Hersh CP, Demeo DL, Himes BE, Sylvia JS, Klanderman BJ, Ziniti JP, Lange C, Litonjua AA, Sparrow D, Regan EA, Make B, Hokanson JE, Murray T, Hetmanski JB, Pillai SG, Kong X, Anderson WH, Tal-Singer R, Lomas DA, Coxson HO, Edwards LD, MacNee W, Vestbo J, Yates JC, Agusti A, Calverley PM, Celli B, Crim C, Rennard S, Wouters E, Bakke P, Gulsvik A, Crapo JD, Beaty TH, Silverman EK. A genome-wide association study of copd identifies a susceptibility locus on chromosome 19q13. *Hum Mol Genet* 2011;21:947-957.

Table E1. Results of multivariate analysis to identify major confounders for the levels of biomarkers in COPD (p-values for each variable in the multivariate model constructed for each biomarker are shown)

Covariates	CC16	SP-D	Fibrinogen	IL6	IL8	TNF-α	CRP
Age, year	<.0001	<.0001	0.0001	<.0001	0.55	0.24	0.02
Sex, female	<.0001	0.28	0.009	0.53	0.004	0.65	0.92
Amount of smoking, pack- years	0.56	0.048	<.0001	<.0001	0.002	0.59	0.001
Smoking status, current	<.0001	<.0001	0.76	0.30	0.19	0.0006	0.67

CC16; Clara cell protein, SP-D; surfactant protein D, IL6; interleukin-6, IL8; interleukin-8, TNF- α; tumor necrosis factor-α, CRP; C-reactive protein

Table E2. Top 10 SNPs associated with circulating level of fibrinogen, IL-6, IL-8, TNF-α, and CRP

Biomarkers	SNP	Rank	P-value	Chr	Coordinate	Type	Closest gene	Distance to gene	A1	MAF	HWE
Fibrinogen	rs9951925	1	2.36E-06	18	71878517	INTERGENIC	AC090398.2	-16224	C	0.47	0.96
	rs4508864	2	5.57E-06	4	155481289	UPSTREAM	FGB	-2819	T	0.22	0.36
	rs13181561	3	6.47E-06	5	138850905	UPSTREAM	AC138517.1	-1206	G	0.27	0.73
	rs722989	4	1.22E-05	8	76441911	INTRONIC	HNF4G	0	A	0.14	0.64
	rs12377896	5	1.42E-05	9	83873695	INTERGENIC	RP11-232A1.2	-151499	A	0.02	0.13
	rs3729848	6	1.65E-05	8	11607930	INTRONIC	GATA4	0	T	0.14	0.18
	rs512625	7	1.95E-05	20	3648378	UPSTREAM	ADAM33	234	A	0.29	0.17
	rs7380062	8	2.24E-05	5	138847901	UPSTREAM	AC138517.1	-4210	T	0.14	0.78
	rs7912637	9	2.40E-05	10	45723417	DOWNSTREAM	RP11-432I13.1	4057	T	0.31	0.46
	rs12378661	10	3.18E-05	9	83881459	INTERGENIC	RP11-232A1.2	-143735	T	0.04	0.005
IL6	rs2823743	1	1.42E-06	21	17667720	WITHIN_NON_	C21orf34	0	C	0.14	0.09

						CODING_GENE					
						WITHIN_NON_					
	rs2823735	2	3.16E-06	21	17657785	CODING_GENE	C21orf34	0	G	0.14	0.08
	rs954820	3	3.71E-06	10	133957761	INTRONIC	JAKMIP3	0	C	0.44	0.65
	rs6667220	4	7.70E-06	1	15347640	INTRONIC	RP1-21O18.1	0	G	0.26	0.64
	rs1124480	5	9.46E-06	3	13857969	3PRIME_UTR	WNT7A	0	C	0.46	0.28
	rs346658	6	9.50E-06	5	135758057	INTERGENIC	AC112178.1	-6524	T	0.37	0.63
	rs854505	7	9.51E-06	1	201293985	INTRONIC	PKP1	0	G	0.26	0.04
	rs10873629	8	1.01E-05	15	26621229	INTERGENIC	AC009878.2	-18990	A	0.17	0.93
	rs3814258	9	1.30E-05	13	113915485	INTRONIC	CUL4A	0	A	0.17	0.07
	rs12294685	10	1.33E-05	11	76423840	INTERGENIC	AP001189.1	-8538	C	0.05	1
IL8	rs903614	1	3.08E-06	8	119830680	INTERGENIC	KB-1137H10.1;	-56242	C	0.10	0.45
						NON_SYNONY					
	rs3751143	2	4.20E-06	12	121622304	MOUS_CODING	P2RX7	0	C	0.18	0.25
	rs7006821	3	5.11E-06	8	72270082	INTRONIC	EYA1	0	C	0.07	0.21
	rs12149070	4	7.50E-06	16	70913984	INTRONIC	HYDIN	0	T	0.06	0.001

	rs7680050	5	1.20E-05	4	185181276	DOWNSTREAM	RP11-162O12.1	4793	T	0.48	0.32
	rs2000059	6	1.28E-05	1	207137338	INTRONIC	FCAMR	0	A	0.11	0.30
	rs2791400	7	1.41E-05	1	245966754	INTRONIC	SMYD3	0	C	0.32	0.61
	rs13127455	8	1.49E-05	4	185181119	DOWNSTREAM	RP11-162O12.1	4636	C	0.37	0.63
	rs637736	9	1.53E-05	9	84392044	DOWNSTREAM	RP11-154D17.1	229	A	0.42	0.12
	rs6930161	10	1.63E-05	6	122910337	INTRONIC	PKIB	0	A	0.42	0.41
TNF-α	rs10007052	1	1.17E-07	4	142005573	INTRONIC	RNF150	0	A	0.23	0.13
	rs7147624	2	4.61E-06	14	65865625	INTERGENIC	FUT8	-11685	T	0.15	0.29
	rs17832777	3	6.44E-06	14	56886686	INTERGENIC	PELI2	118442	C	0.17	0.58
	rs4468361	4	8.24E-06	11	132010117	INTRONIC	NTM;OPCML	0	T	0.32	0.21
	rs1468013	5	9.70E-06	11	91612894	INTERGENIC	RP11-447G14.1	-281379	T	0.33	0.006
	rs2898816	6	1.06E-05	14	66008160	INTRONIC	FUT8	0	T	0.14	0.31
	rs13003408	7	1.09E-05	2	197045688	UPSTREAM	STK17B	-4461	A	0.27	0.46
	rs12891725	8	1.11E-05	14	48392595	INTERGENIC	MDGA2	-248642	C	0.37	0.59
	rs8021889	9	1.17E-05	14	66214811	DOWNSTREAM	FUT8	3972	A	0.15	0.17

	rs12151959	10	1.22E-05	21	38215923	INTRONIC	HLCS	0	G	0.19	0.24
CRP	rs7953249	1	1.16E-06	12	121403724	INTERGENIC	HNF1A;TCF1	-12622	G	0.44	0.89
	rs652520	2	1.86E-06	6	93713805	INTERGENIC	RP1-23E21.2	-66300	C	0.48	0.12
	rs2650000	3	6.34E-06	12	121388962	INTERGENIC	HNF1A;TCF1	-27384	A	0.37	0.66
	rs10774579	4	7.35E-06	12	121405210	INTERGENIC	HNF1A;TCF1	-11136	C	0.46	0.68
	rs12420082	5	1.04E-05	11	23565554	INTERGENIC	RP11-713P14.1	22806	T	0.20	0.44
	rs12677017	6	1.20E-05	8	22676135	INTRONIC	PEBP4	0	T	0.45	0.09
	rs11027306	7	1.48E-05	11	23569160	INTERGENIC	RP11-713P14.1	26412	C	0.19	0.36
	rs2076904	8	1.83E-05	9	130310173	INTRONIC	FAM129B	0	G	0.27	0.11
	rs10514583	9	2.04E-05	16	83289317	INTRONIC	CDH13	0	G	0.15	0.42
	rs7310409	10	2.16E-05	12	121424861	INTRONIC	HNF1A;TCF1	0	A	0.41	0.55

Notes: A1 = Minor allele; MAF = Minor Allele Frequency; HWE = p-value for deviation from Hardy-Weinberg Equilibrium

Table E3. The contribution of relatively independent top SNPs to plasma levels of CC16 and SP-D.

Biomarkers	Variables included in models with relatively independent SNPs*	R-squared	Adjusted R-squared
CC16	Age, Sex, Pack-years, Smoking status	0.158	0.156
	rs7929679	0.166	0.164
	rs2463822	0.170	0.168
	rs3741240	0.201	0.199
	rs2077224	0.182	0.179
	rs17157266	0.172	0.169
	All five CC16 SNPs, Age, Sex, Pack-years, and Smoking status	0.220	0.216
SP-D	Age, Sex, Pack-years, Smoking status	0.044	0.042
	rs3130559	0.061	0.059
	rs1265093	0.062	0.059
	rs2074488	0.071	0.069
	rs9266629	0.064	0.061

	rs1923539	0.059	0.057
	rs7078012	0.062	0.059
	rs3923564	0.098	0.095
	rs12220777	0.067	0.064
	rs728616	0.068	0.066
	rs3851050	0.066	0.063
	rs6585424	0.062	0.059
	rs8048576	0.072	0.070
	All twelve SFTPD SNPs, Age, Sex, Pack-years, and Smoking status	0.217	0.209

* Independent SNPs were selected based on r^2 threshold of 0.5.

Table E4.The correlation of mRNA expression of biomarker genes in sputum and level of biomarkers in blood*

Gene (Probes)	Pearson correlation coefficient (P-value)
SCGB1A1	0.17(0.06)
SFTPD	0.22 (0.009)
Probe 1 for FGB	0.02(0.80)
Probe 2 for FGB	0.04(0.66)
Probe 1 for FGG	-0.04(0.70)
Probe 2 for FGG	0.01(0.91)
Probe 1 for FGA	0.22(0.02)
Probe 2 for FGA	-0.001(0.99)
IL6	-0.001(0.99)
Probe 1 for IL8	0.08(0.39)
Probe 2 for IL8	0.06(0.51)
TNF-α	0.12(0.17)
Probe 1 for CRP	0.12(0.25)
Probe 2 for CRP	0.09(0.41)
Probe 3 for CRP	-0.23(0.03)

FGA, FGB, and FGG =Fibrinogen alpha, beta, and gamma genes, respectively.

Table E5. The association of SNPs related to mRNA expression in sputum with circulating protein levels

			Linear association with circulating level of biomarkers*			
CHR	SNP	BP	Minor allele	Beta	95% CI	P
CC 16						
11	rs10466455	34737512	C	-0.11	-0.14, -0.07	2.6E-10
11	rs10836312	34767019	C	-0.11	-0.14, -0.08	6.7E-11
11	rs906902	34736854	A	-0.11	-0.14, -0.07	2.4E-10
11	rs3741240	61943118	A	-0.18	-0.21, -0.15	1.4E-26
11	rs2509956	61953299	C	-0.14	-0.17, -0.10	1.0E-13
SP-D						
10	rs1923539	81684930	A	0.11	0.08, 0.15	5.0E-9
10	rs1885551	81702333	G	-0.36	-0.42, -0.31	1.2E-39
10	rs2146192	81705718	C	-0.36	-0.42, -0.31	1.2E-39

* Circulating blood levels of CC16 and SP-D were transformed to a natural log scale to approximate a normal distribution and age, gender, pack-years, current smoking status, and principal components for genetic ancestry were adjusted as covariates.

CHR= Chromosome, BP=Physical position (base-pair), Beta= Regression coefficient, CI=Confidence interval (lower, upper)

Table E6. The association of top hits SNPs with risk of COPD in each collaborative COPD GWAS population and the COPDGene and ICGN cohorts.

CC 16			ECLIPSE			NETT/NAS			NORWAY			COPDGene *		ICGN	
CHR	SNP	Nearest Gene	NMISS	OR	P	NMISS	OR	P	NMISS	OR	P	OR	P	#informative families	P
CC16															
11	rs17157266	AHNAK	1912	1.69	0.004	801	1.29	0.11	1657	1.15	0.19	1.05	0.72	319	0.55
Surfactant protein D															
16	rs8063863	ATP2C2	1912	0.69	0.02	801	0.72	0.047	1658	0.84	0.10	0.88	0.36	326	0.52
16	rs8048576	ATP2C2	1912	0.68	0.02	801	0.74	0.10	1658	0.89	0.30	1.10	0.51	319	0.53
10	rs7078012	SFTPD	1911	0.68	0.01	801	0.84	0.30	1651	0.85	0.17	1.16	0.27	284	0.30
10	rs1885553	SFTPD	1912	0.51	0.61	801	1.06	0.63	1658	1.15	0.10	NA	NA	NA	NA
10	rs1923539	RP11-479O17.4	1912	1.20	0.20	801	1.05	0.73	1658	1.12	0.22	1.02	0.86	425	0.20

* Untyped markers were imputed from HapMap reference panel (phase II)

NA: not available

NMISS: Number of subjects with non-missing data included in association analysis

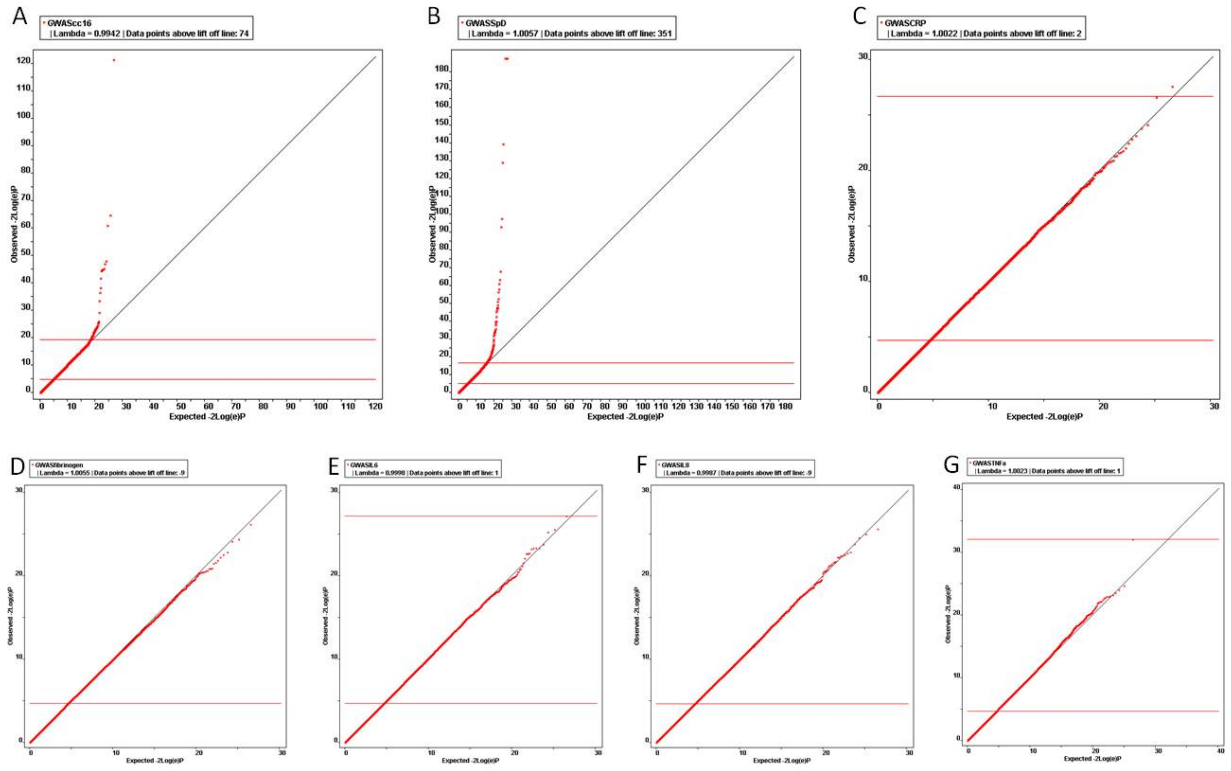

Figure E1. Q-q plots of GWAS for biomarkers. (A) CC16, (B) SP-D, (C) CRP, (D) fibrinogen, (E) IL-6, (F) IL-8, and (G) TNF-α. The lower red line denotes the 90th percentile, while the upper one (if present) indicates the point where the frequency of low P values exceeds expectations.

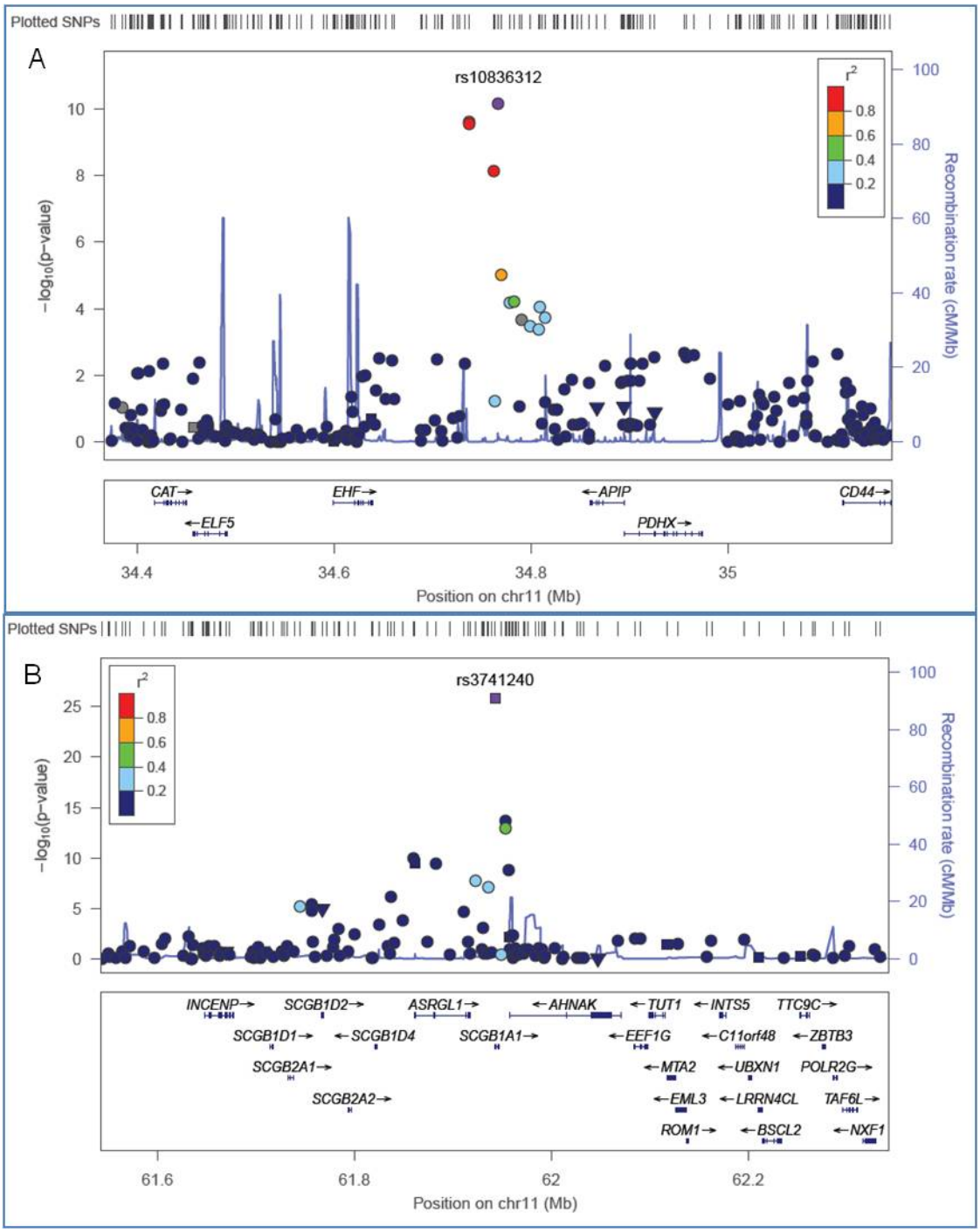

Figure E2. Regional association plots of genotyped SNPs on chromosome 11 associated with circulating levels of CC16. (A) A plot of genetic loci approximately 25 Mb away from *SCGB1A1* across the centromere. (B) A plot of genetic loci located near *SCGB1A1*. rs10836312

(A) and rs3741240 (B), the most highly associated index SNPs, are indicated by purple color while the colors of the remaining SNPs indicate the linkage disequilibrium with the index SNP. Displayed LD estimates were obtained from HapMap phase II (CEU). Symbols reflecting genomic annotation represent as following; ▼= nonsynonymous, ■=synonymous or UTR, ●=intergenic or intronic.

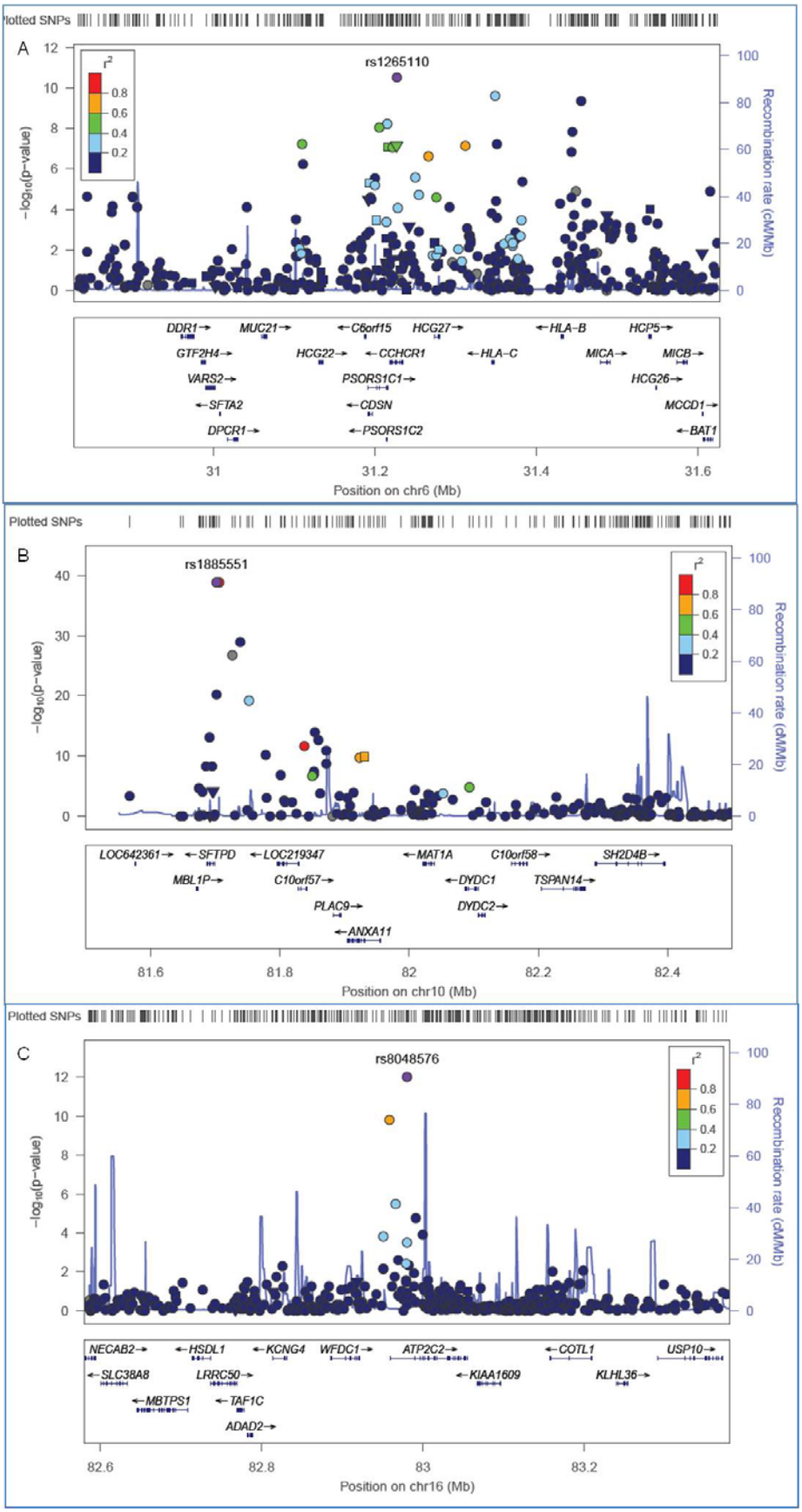
Plotted SNPs
A
rs1265110
r²
0.8
0.6
0.4
0.2
−log10(p-value)
Recombination rate (cM/Mb)
DDR1→ MUC21→ ←C6orf15 HCG27→ ←HLA-B HCP5→
GTF2H4→ HCG22→ ←CCHCR1 ←HLA-C MICA→ MICB→
VARS2→ PSORS1C1→ HCG26→
←SFTA2 ←CDSN MCCD1→
DPCR1→ ←PSORS1C2 ←BAT1
31 31.2 31.4 31.6
Position on chr6 (Mb)
Plotted SNPs
B
rs1885551
r²
0.8
0.6
0.4
0.2
−log10(p-value)
Recombination rate (cM/Mb)
LOC642361→ ←SFTPD ←LOC219347 ←MAT1A C10orf58→ SH2D4B→
MBL1P→ C10orf57→ ←DYDC1 TSPAN14→
PLAC9→ DYDC2→
←ANXA11
81.6 81.8 82 82.2 82.4
Position on chr10 (Mb)
Plotted SNPs
C
rs8048576
r²
0.8
0.6
0.4
0.2
−log10(p-value)
Recombination rate (cM/Mb)
NECAB2→ ←HSDL1 ←KCNG4 WFDC1→ ATP2C2→ ←COTL1 USP10→
←SLC38A8 LRRC50→ ←KIAA1609 KLHL36→
←MBTPS1 ←TAF1C
ADAD2→
82.6 82.8 83 83.2
Position on chr16 (Mb)

Figure E3. Regional association plots of genotyped SNPs associated with circulating level of SpD. Plots of genetic loci on chromosomes 6 (A), 10 (B), and 16 (C). The most highly associated index SNP in each chromosomal region is indicated by purple color while the colors of the remaining SNPs indicate the linkage disequilibrium with the index SNP. Displayed LD estimates were obtained from HapMap phase II (CEU). Symbols reflecting genomic annotation represent the following; ▼= nonsynonymous, ■=synonymous or UTR, ●=intergenic or intronic.