Data Supplement for:

Twenty-two genetic loci in COPD overlap with population-based lung function and pulmonary fibrosis

Brian D. Hobbs[1,2], Kim de Jong[3,4], Maxime Lamontagne[5], Yohan Bossé[5], Nick Shrine[6], María Soler Artigas[6], Louise V. Wain[6], Ian P. Hall[7], Victoria E. Jackson[6,8], Annah B. Wyss[9], Stephanie J. London[10], Kari E. North[11], Nora Franceschini[11], David P. Strachan[12], Terri H. Beaty[13], John E. Hokanson[14], James D. Crapo[15,57], Peter J. Castaldi[1,16], Robert P. Chase[1], Traci M. Bartz[17,18], Susan R. Heckbert[17,19,20], Bruce M. Psaty[17,21,20], Sina A. Gharib[22], Pieter Zanen[23], Jan W. Lammers[24], Matthijs Oudkerk[25], H. J. Groen[26], Nick Locantore[27], Ruth Tal-Singer[27], Stephen I. Rennard[28,29], Wim Timens[30], Peter D. Paré[31], Jeanne C. Latourelle[32], Josée Dupuis[33,34], George T. O'Connor[34,35], Jemma B. Wilk[34], Woo Jin Kim[36], Mi Kyeong Lee[36], Yeon-Mok Oh[37], Judith M. Vonk[3,4], Harry J. de Koning[38], Shuguang Leng[39], Steven A. Belinsky[39], Yohannes Tesfaigzi[39], Ani Manichaikul[40,41], Xin-Qun Wang[41], Stephen S. Rich[40,41], R Graham Barr[42], David Sparrow[43], Augusto L. Litonjua[1,2], Per Bakke[44], Amund Gulsvik[44], Lies Lahousse[45,46], Guy G. Brusselle[47], Bruno H. Stricker[45,48,49,50], André G. Uitterlinden[45,49,50], Elizabeth J. Ampleford[51], Eugene R. Bleecker[51], Prescott G. Woodruff[52], Deborah A. Meyers[51], Dandi Qiao[1], David A. Lomas[53], Jae-Joon Yim[54], Deog Kyeom Kim[55], Iwona Hawrylkiewicz[56], Pawel Sliwinski[56], Megan Hardin[1,2], Tasha E. Fingerlin[57,58], David A. Schwartz[57,59,60], Dirkje S. Postma[26,4], William MacNee[61], Martin D. Tobin[6,62], Edwin K. Silverman[1,2], H. Marike Boezen[3,4], *Michael H. Cho[1,2], COPDGene Investigators, ECLIPSE Investigators, LifeLines Investigators, SPIROMICS Research Group, International COPD Genetics Network Investigators, UK BiLEVE Investigators, International COPD Genetics Consortium

1 Channing Division of Network Medicine, Brigham and Women's Hospital, Boston, MA, USA
2 Division of Pulmonary and Critical Care Medicine, Brigham and Women's Hospital, Boston, MA, USA
3 University of Groningen, University Medical Center Groningen, Department of Epidemiology, Groningen, the Netherlands
4 University of Groningen, University Medical Center Groningen, Groningen Research Institute for Asthma and COPD (GRIAC), Groningen, the Netherlands
5 Institut universitaire de cardiologie et de pneumologie de Québec, Québec, Canada
6 Genetic Epidemiology Group, Department of Health Sciences, University of Leicester, Leicester, UK
7 Division of Respiratory Medicine, Queen's Medical Centre, University of Nottingham, Nottingham, UK
8 Department of Health Sciences, University of Leicester, Leicester, UK
9 Epidemiology Branch, National Institute of Environmental Health Sciences, Research Triangle Park, NC, USA
10 Epidemiology Branch, National Institute of Environmental Health Sciences, National Institutes of Health, Department of Health and Human Services, Research Triangle Park, NC, USA
11 Department of Epidemiology, University of North Carolina, Chapel Hill, NC, USA
12 St George's, University of London, Cranmer Terrace, London SW17 0RE, UK
13 Johns Hopkins University Bloomberg School of Public Health, Baltimore, MD, USA
14 Department of Epidemiology, University of Colorado Anschutz Medical Campus, Aurora, CO, USA
15 National Jewish Health, Denver, CO, USA
16 Division of General Internal Medicine, Brigham and Women's Hospital, Boston, MA, USA
17 Cardiovascular Health Research Unit, University of Washington, Seattle, WA, USA
18 Departments of Medicine and Biostatistics, University of Washington, Seattle, WA, USA
19 Department of Epidemiology, University of Washington, Seattle, WA, USA
20 Group Health Research Institute, Group Health Cooperative, Seattle, WA, USA

21 Departments of Epidemiology, Medicine and Health Services, University of Washington, Seattle, WA, USA
22 Computational Medicine Core, Center for Lung Biology, UW Medicine Sleep Center, Department of Medicine, University of Washington, Seattle, WA, USA
23 University Medical Center Utrecht, Department of Pulmonary Diseases, Utrecht, the Netherlands
24 Department of Pulmonology, University Medical Center Utrecht, University of Utrecht, Utrecht, the Netherlands
25 Department of Radiology, University Medical Center Groningen, University of Groningen, Groningen, the Netherlands
26 University of Groningen, University Medical Center Groningen, Department of Pulmonology, Groningen, the Netherlands
27 GSK R&D, King of Prussia, PA, USA
28 Pulmonary, Critical Care, Sleep and Allergy Division, Department of Internal Medicine, University of Nebraska Medical Center, Omaha, NE, USA
29 Clincal Discovery Unit, AstraZeneca, Cambridge, UK
30 Department of Pathology and Medical Biology, University of Groningen, University Medical Center Groningen, GRIAC Research Institute, Groningen, the Netherlands
31 University of British Columbia Center for Heart Lung Innovation and Institute for Heart and Lung Health, St Paul's Hospital, Vancouver, British Columbia, Canada
32 Department of Neurology, Boston University School of Medicine, Boston, MA, USA
33 Department of Biostatistics, Boston University School of Public Health, Boston, MA, USA
34 The National Heart, Lung, and Blood Institute's Framingham Heart Study, Framingham, MA, USA
35 Pulmonary Center, Department of Medicine, Boston University School of Medicine, Boston, MA, USA
36 Department of Internal Medicine and Environmental Health Center, School of Medicine, Kangwon National University, Chuncheon, South Korea
37 Department of Pulmonary and Critical Care Medicine, and Clinical Research Center for Chronic Obstructive Airway Diseases, Asan Medical Center, University of Ulsan College of Medicine, Seoul, South Korea
38 Department of Public Health, Erasmus Medical Center Rotterdam, Rotterdam, the Netherlands
39 Lovelace Respiratory Research Institute, Albuquerque, NM, USA
40 Center for Public Health Genomics, University of Virginia, Charlottesville, VA, USA
41 Department of Public Health Sciences, University of Virginia, Charlottesville, VA, USA
42 Department of Medicine, College of Physicians and Surgeons and Department of Epidemiology, Mailman School of Public Health, Columbia University, New York, NY, USA
43 VA Boston Healthcare System and Department of Medicine, Boston University School of Medicine, Boston, MA, USA
44 Department of Clinical Science, University of Bergen, Bergen, Norway
45 Department of Epidemiology, Erasmus Medical Center, Rotterdam, the Netherlands
46 Department of Respiratory Medicine, Ghent University Hospital, Ghent, Belgium
47 Department of Public Health, Erasmus Medical Center, Rotterdam, the Netherlands
48 Netherlands Health Care Inspectorate, The Hague, the Netherlands
49 Department of Internal Medicine, Erasmus Medical Center, Rotterdam, the Netherlands
50 Netherlands Genomics Initiative (NGI)-sponsored Netherlands Consortium for Healthy Aging (NCHA), Leiden, the Netherlands
51 Center for Genomics and Personalized Medicine Research, Wake Forest University School of Medicine, Winston Salem, NC, USA
52 Cardiovascular Research Institute and the Department of Medicine, Division of Pulmonary, Critical Care, Sleep, and Allergy, University of California at San Francisco, San Francisco, CA, USA
53 University College London, London, UK

54 Division of Pulmonary and Critical Care Medicine, Department of Internal Medicine, Seoul National University College of Medicine, South Korea
55 Seoul National University College of Medicine, SMG-SNU Boramae Medical Center, Seoul, South Korea
56 National Tuberculosis and Lung Disease Research Institute, Warsaw, Poland
57 Center for Genes, Environment and Health, National Jewish Health, Denver, CO, USA
58 Department of Biostatistics and Informatics, University of Colorado
Denver, Aurora, CO, USA
59 Department of Medicine, School of Medicine, University of Colorado Denver, Aurora, CO, USA
60 Department of Immunology, School of Medicine, University of Colorado Denver, Aurora, CO, USA
61 University of Edinburgh, Edinburgh, UK
62 National Institute for Health Research (NIHR) Leicester Respiratory Biomedical Research Unit, Glenfield Hospital, Leicester, UK

*Corresponding author:
Michael H. Cho (remhc@channing.harvard.edu)
tel: 617-525-0897
fax: 888-487-1078

Supplemental Figures:

Figures S1a-v: Forest plots illustrate the association results at the individual cohort level in each stage of the analysis for each of the 22 genome-wide significant loci. For the Stage 1 analysis portion of the forest plots, the studies are grouped into studies containing individuals of European ancestry and studies containing individuals of non-European ancestry. The "Stage 1" odds ratio (OR) and 95% confidence interval (CI) represents a meta-analysis across the 26 Stage 1 analysis studies. The "Stage 2: BiLEVE" OR represents the meta-analysis of UK BiLEVE never smokers and heavy smokers. The "Overall" OR represents the cumulative meta-analysis across the Stage 1 results and the UK BiLEVE Stage 2 results. If the marker was not available in a specific cohort, the cohort is missing from the forest plot. Study abbreviations: ARIC = Atherosclerosis Risk in Communities, B58 = British 1958 Birth Cohort, CHS = Cardiovascular Health Study, COPACETIC = COPD Pathology: Addressing Critical gaps, Early Treatment & Diagnosis and Innovative Concepts, ECLIPSE = Evaluation of COPD Longitudinally to Identify Predictive Surrogate End-points, eQTL = Lung Expression Quantitative Trait Loci Study, FHS = Framingham Heart Study, KARE = Korean Association Resource project, MESA = Multi-Ethnic Study of Atherosclerosis, NETT-NAS = National Emphysema Treatment Trial / Normative Aging Study, RS = Rotterdam Study, SPIROMICS = Subpopulations and intermediate outcome measures in COPD study , EOCOPD = Boston Early-Onset COPD Study, ICGN = International COPD Genetics Network, TCGS = Transcontinental COPD Genetics Study, BiLEVE = UK Biobank Lung Exome Variant Evaluation; NHW = Non-Hispanic white, AA = African American, EA = European American.

Figure S1a: Forest plot for rs13141641 (*HHIP*)

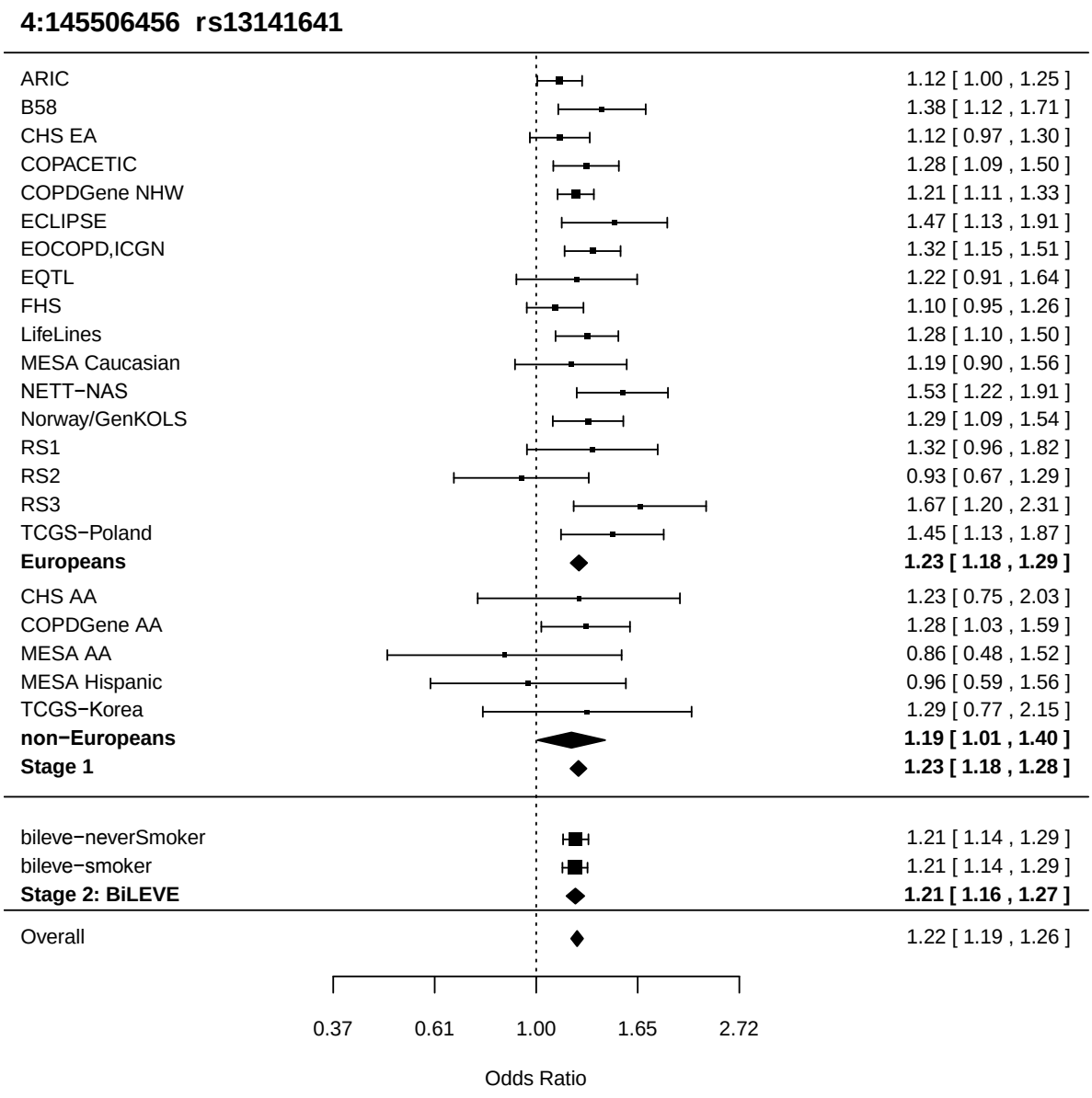


Figure S1b: Forest plot for rs17486278 (*CHRNA5*)

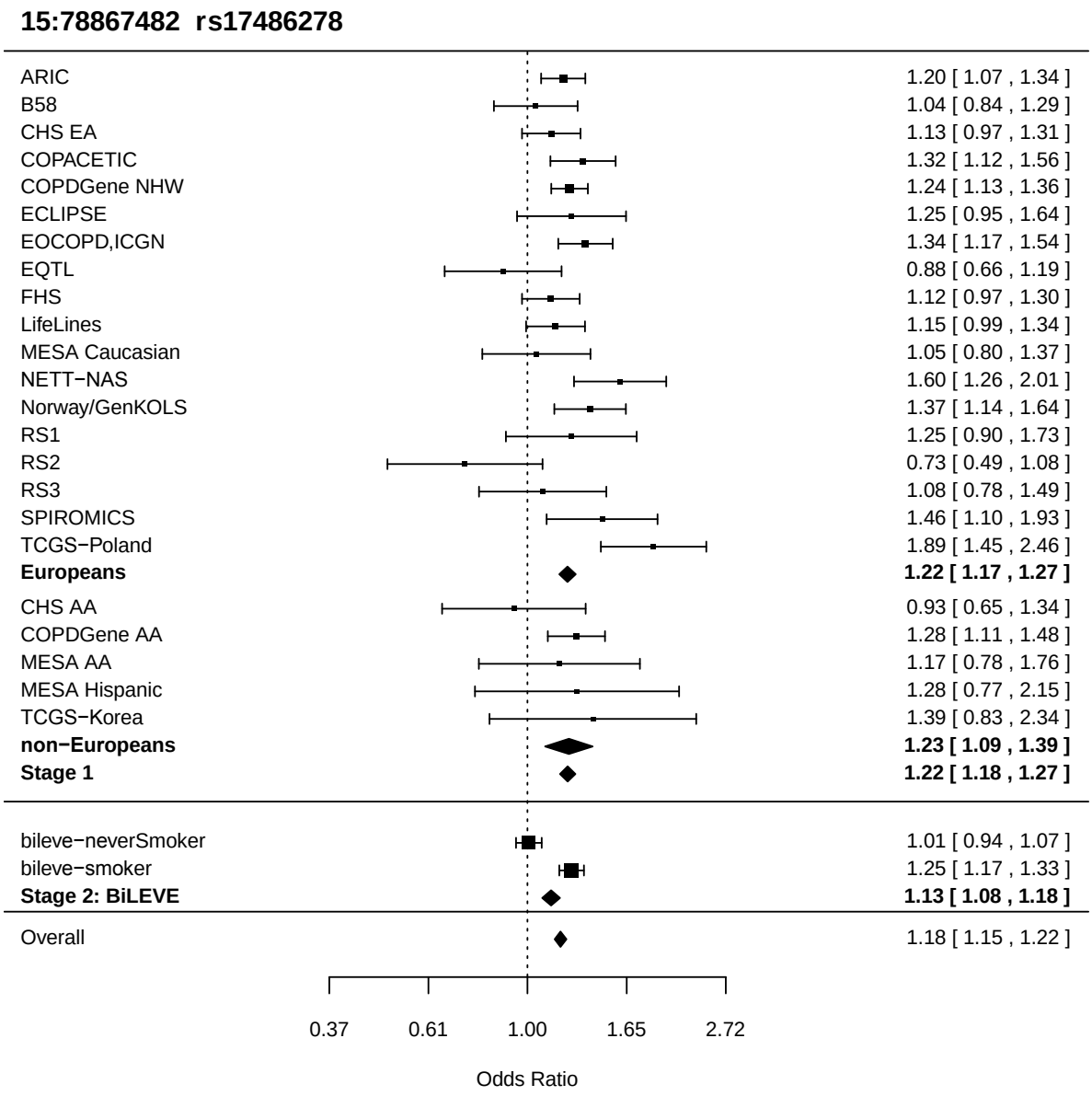


Figure S1c: Forest plot for rs7733088 (*HTR4*)

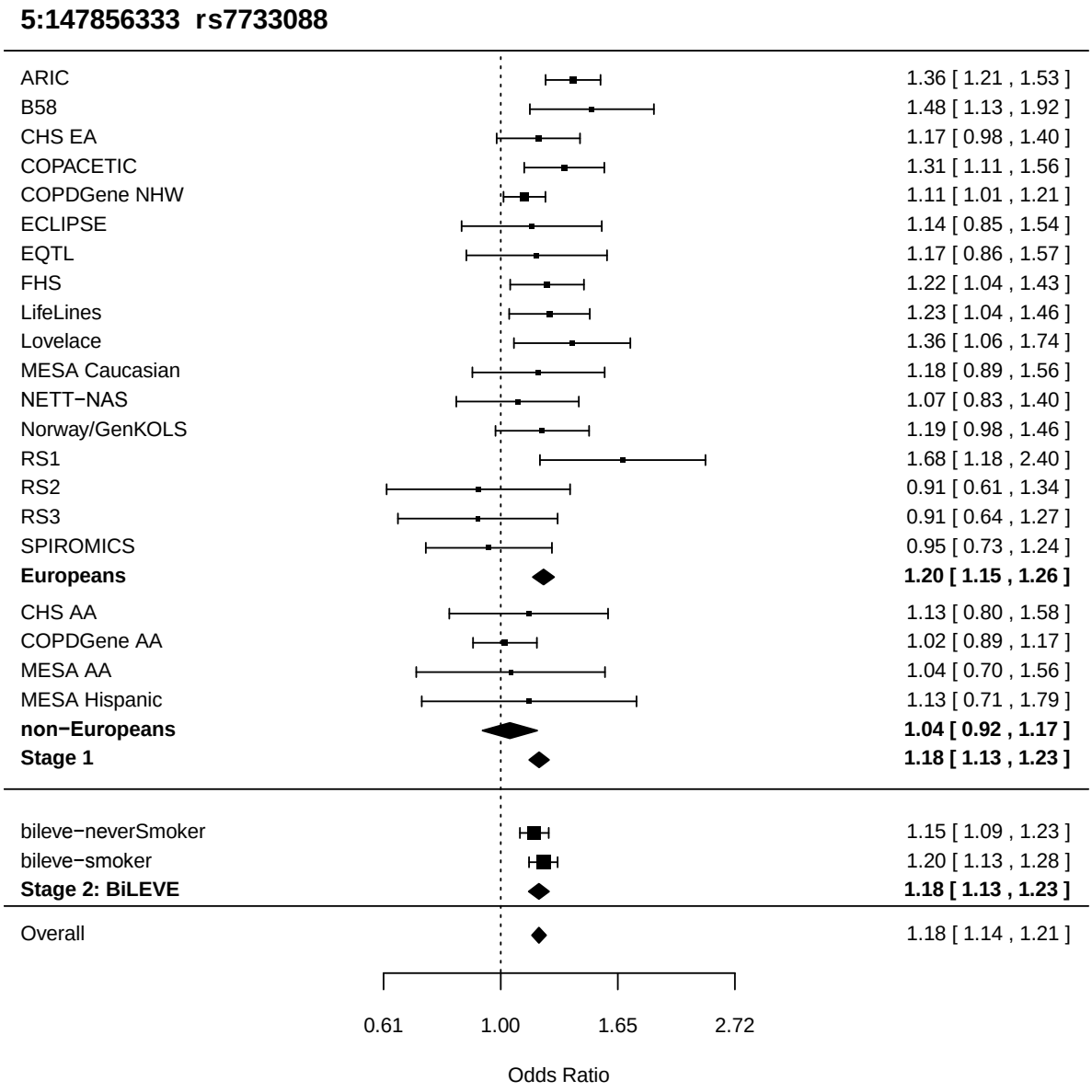


Figure S1d: Forest plot for rs9399401 (*ADGRG6*)

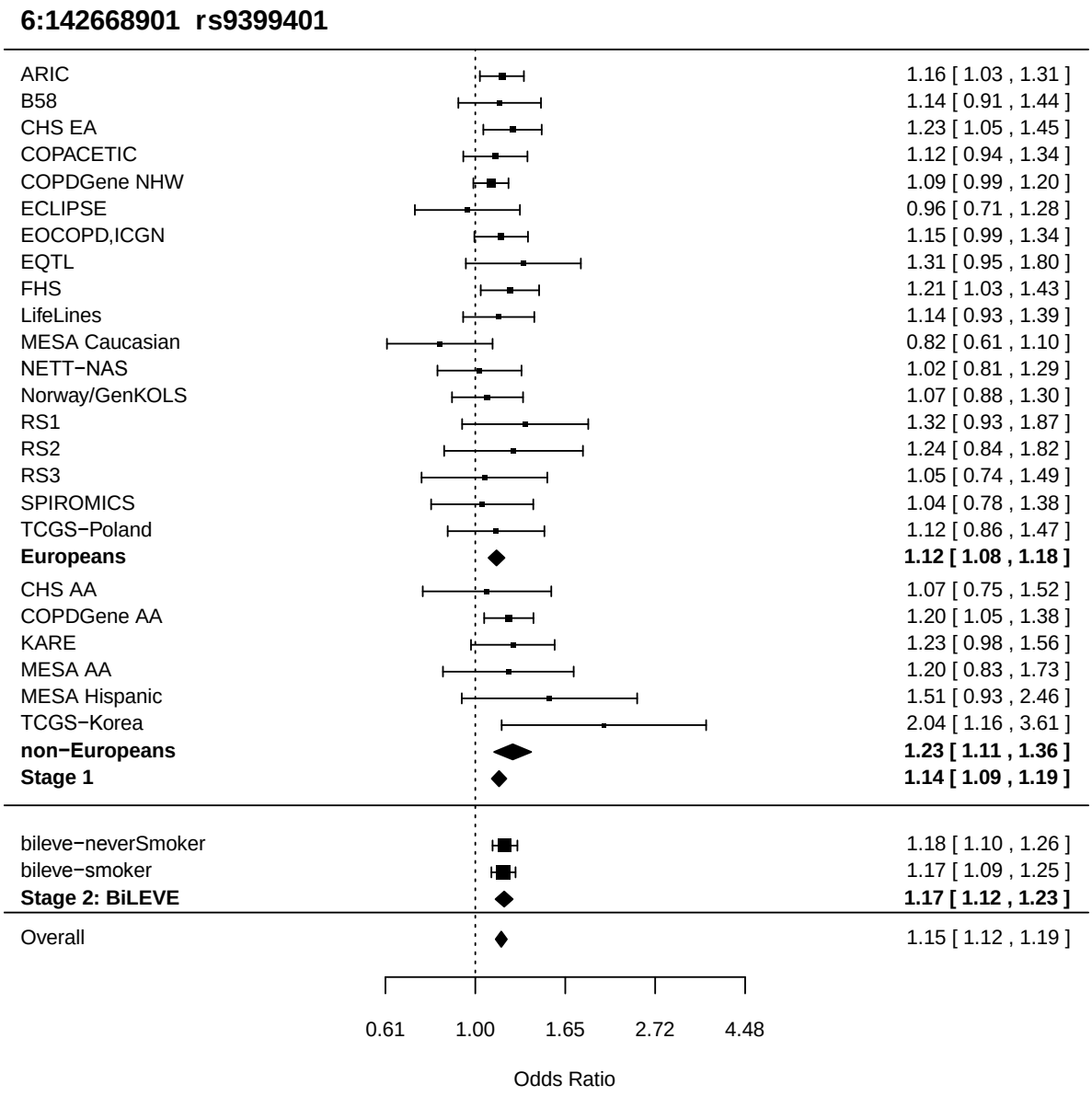


Figure S1e: Forest plot for rs1441358 (*THSD4*)

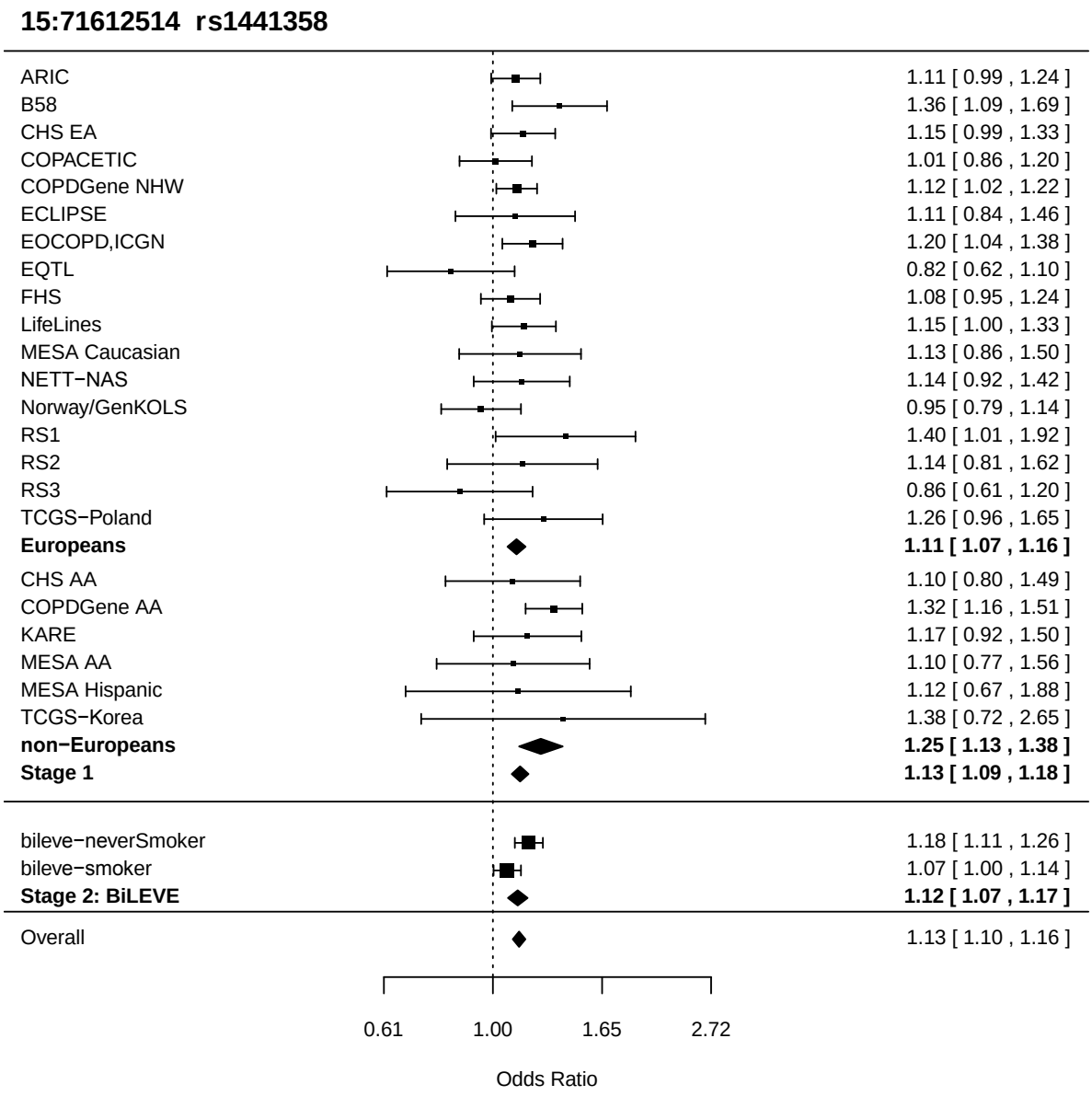


Figure S1f: Forest plot for rs6837671 (*FAM13A*)

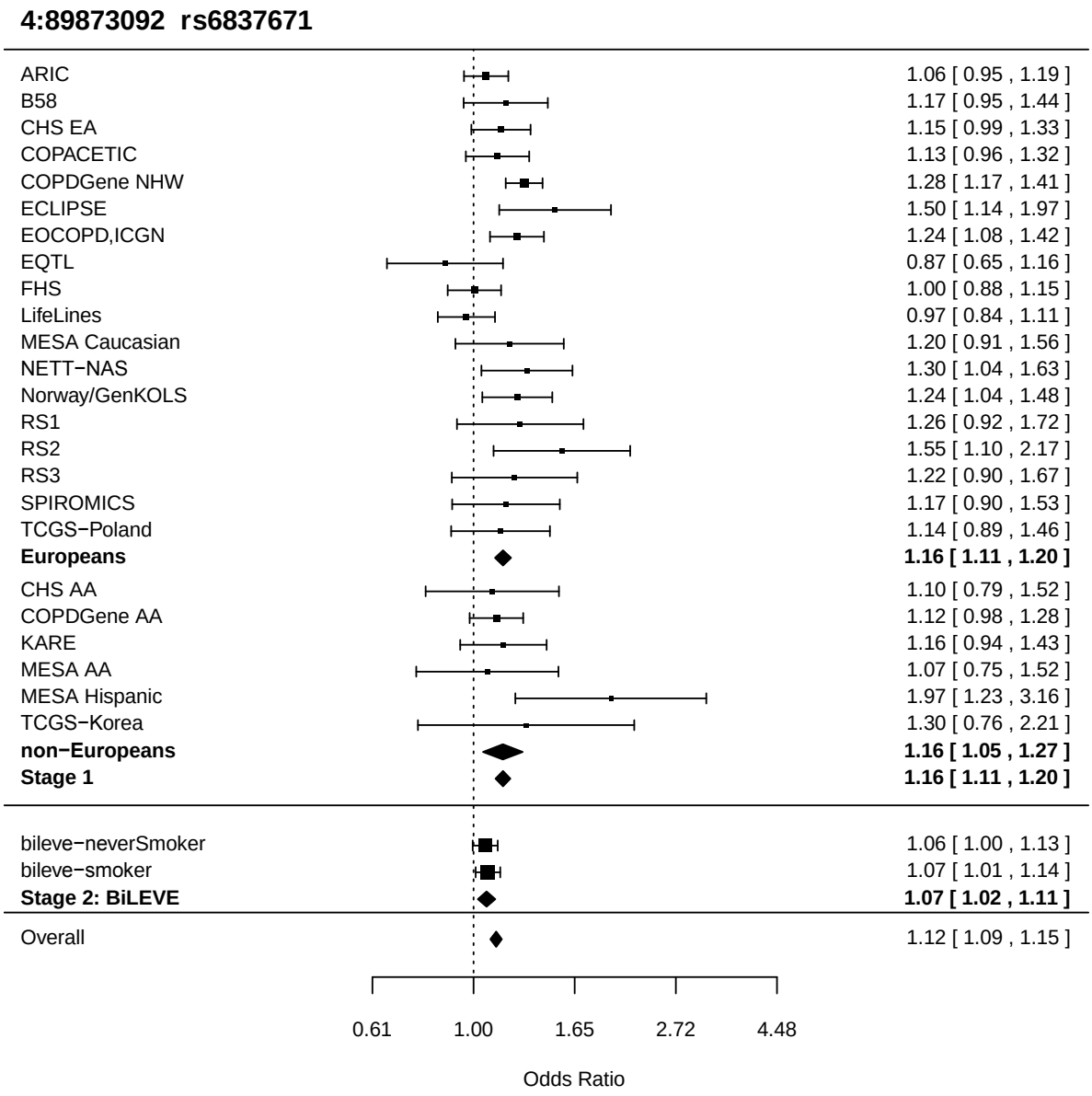


Figure S1g: Forest plot for rs11727735 (*GSTCD*)

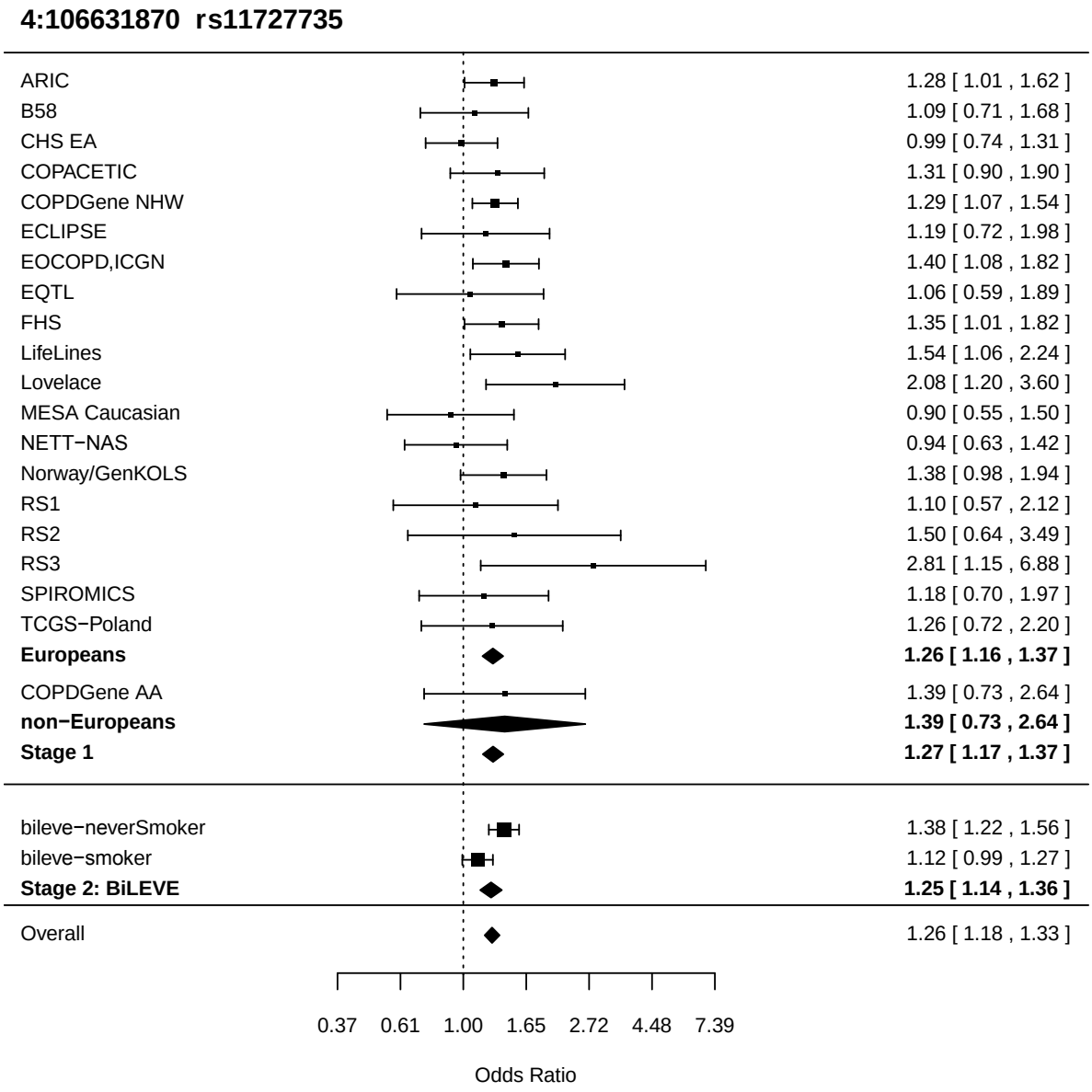


Figure S1h: Forest plot for rs754388 (*RIN3*)

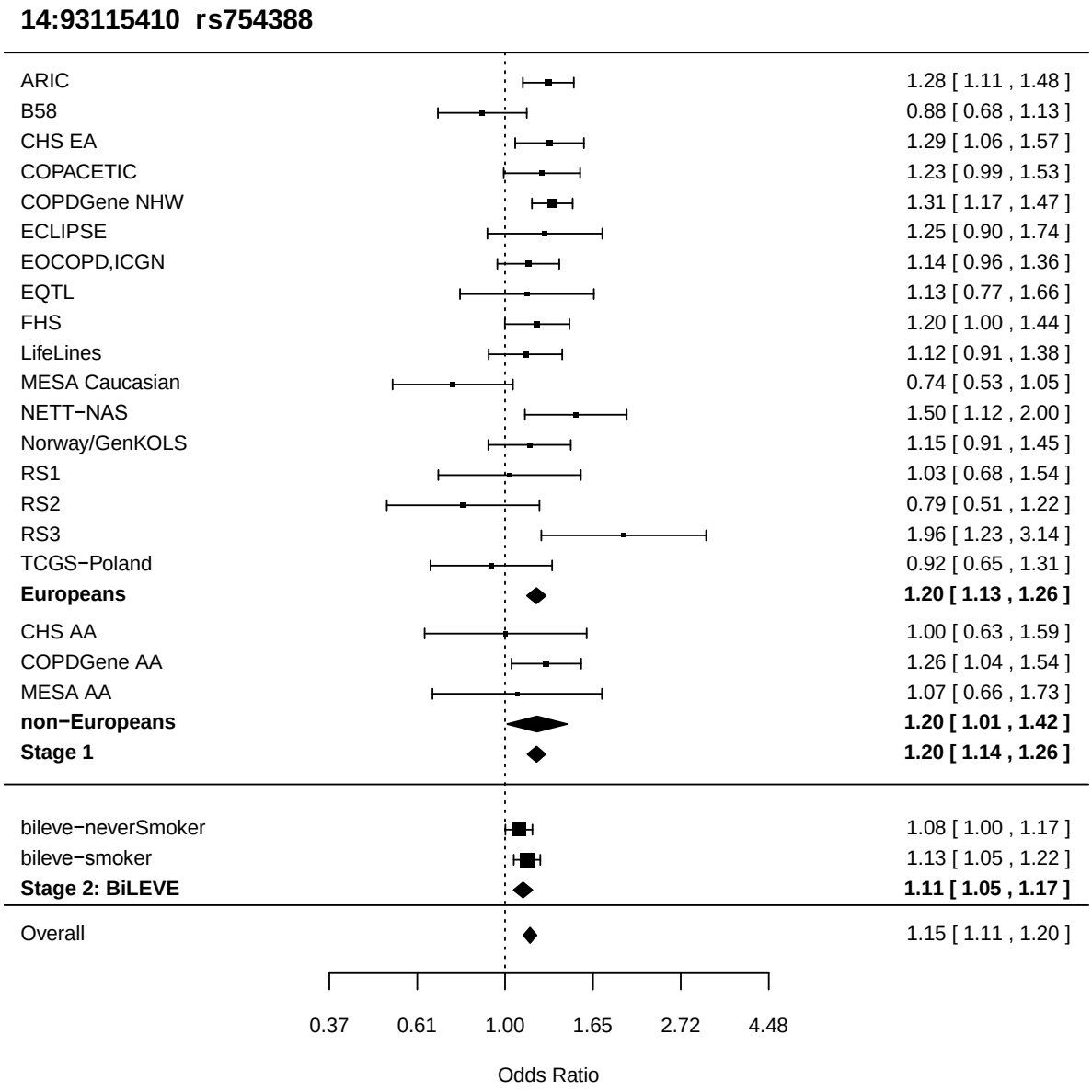


Figure S1i: Forest plot for rs113897301 (*ADAM19*)

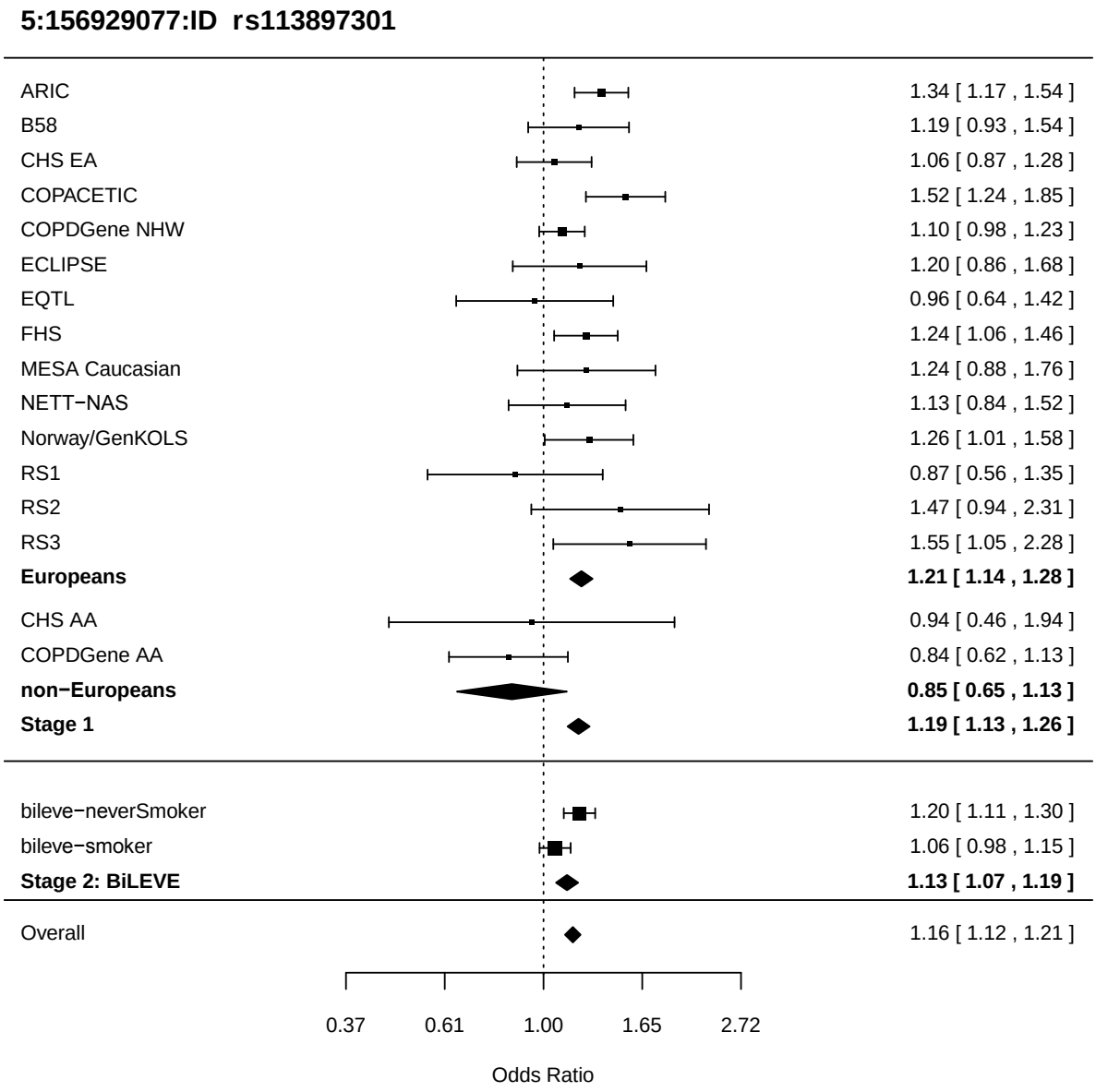


Figure S1j: Forest plot for rs2047409 (*TET2*)

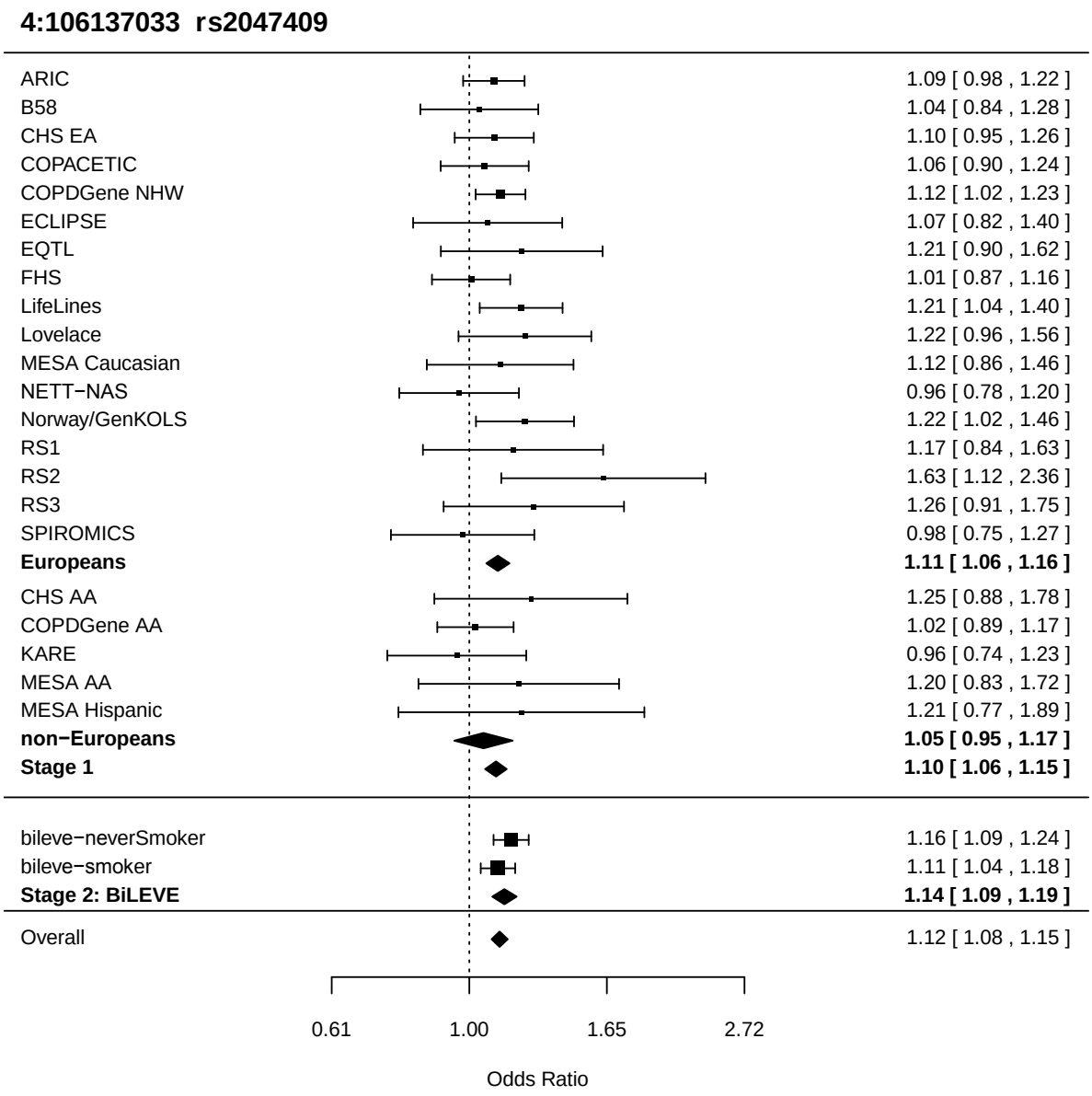


Figure S1k: Forest plot for rs2955083 (*EEFSEC*)

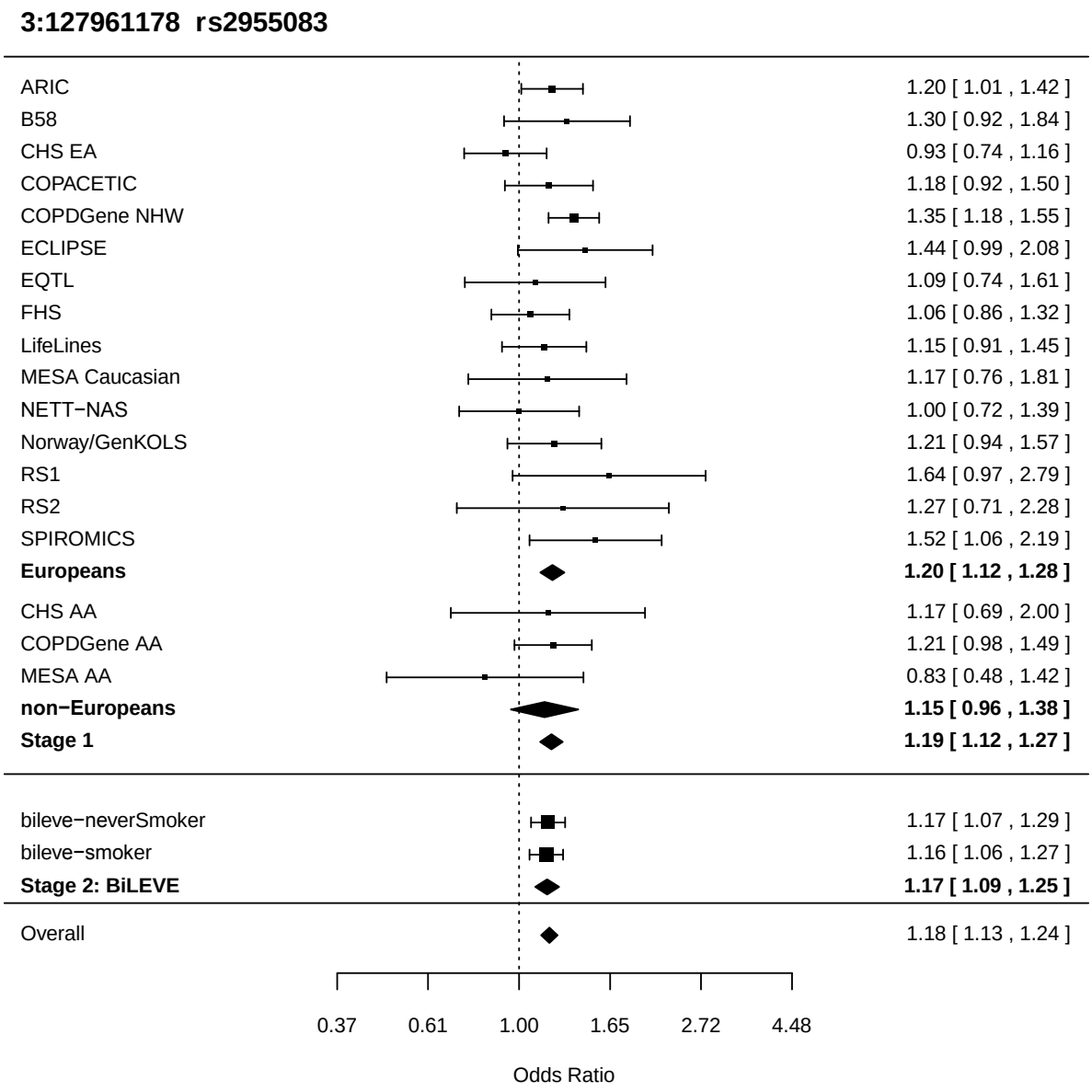


Figure S1l: Forest plot for rs7186831 (*CFDP1*)

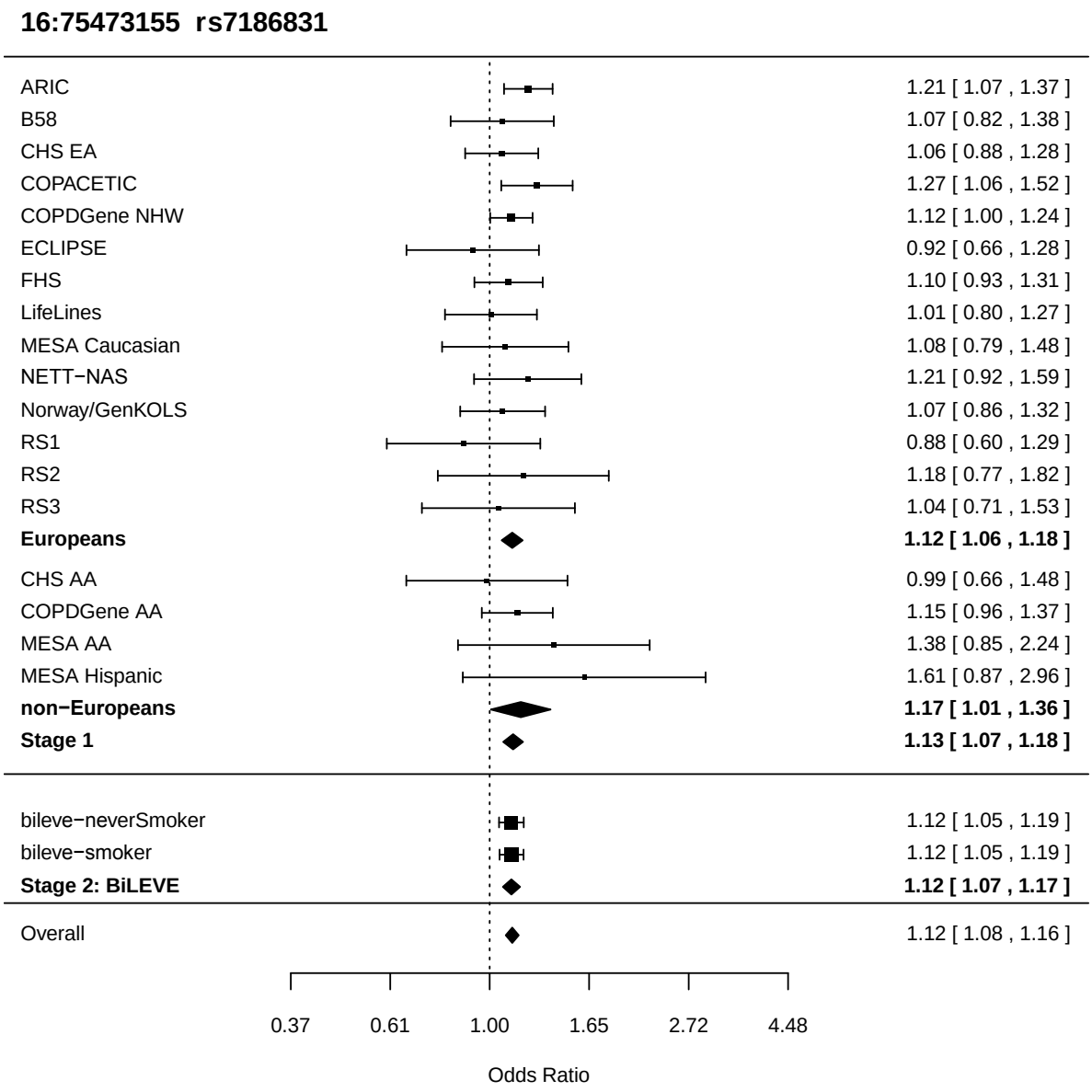


Figure S1m: Forest plot for rs10429950 (*TGFB2*)

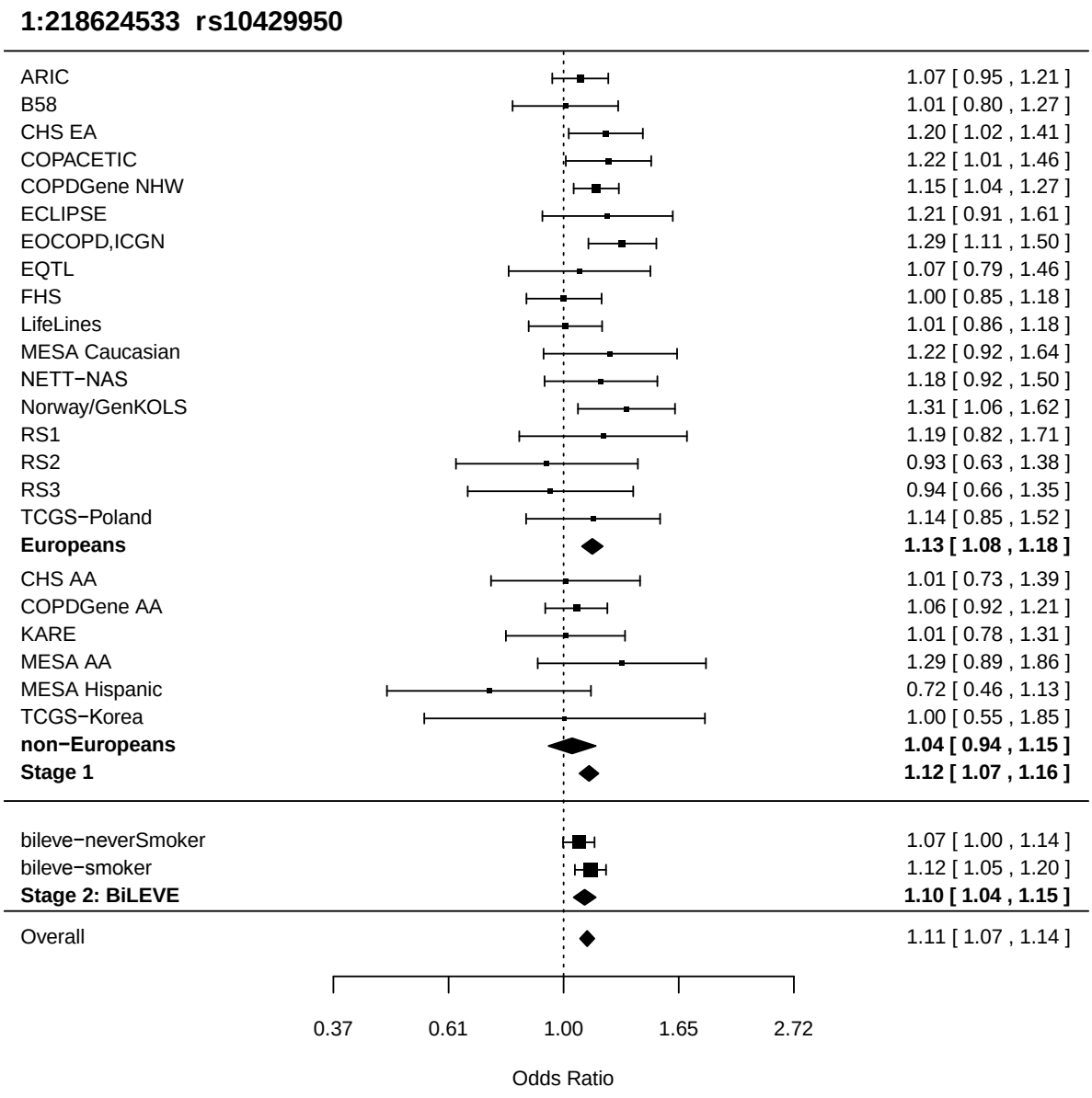


Figure S1n: Forest plot for rs2070600 (*AGER*)

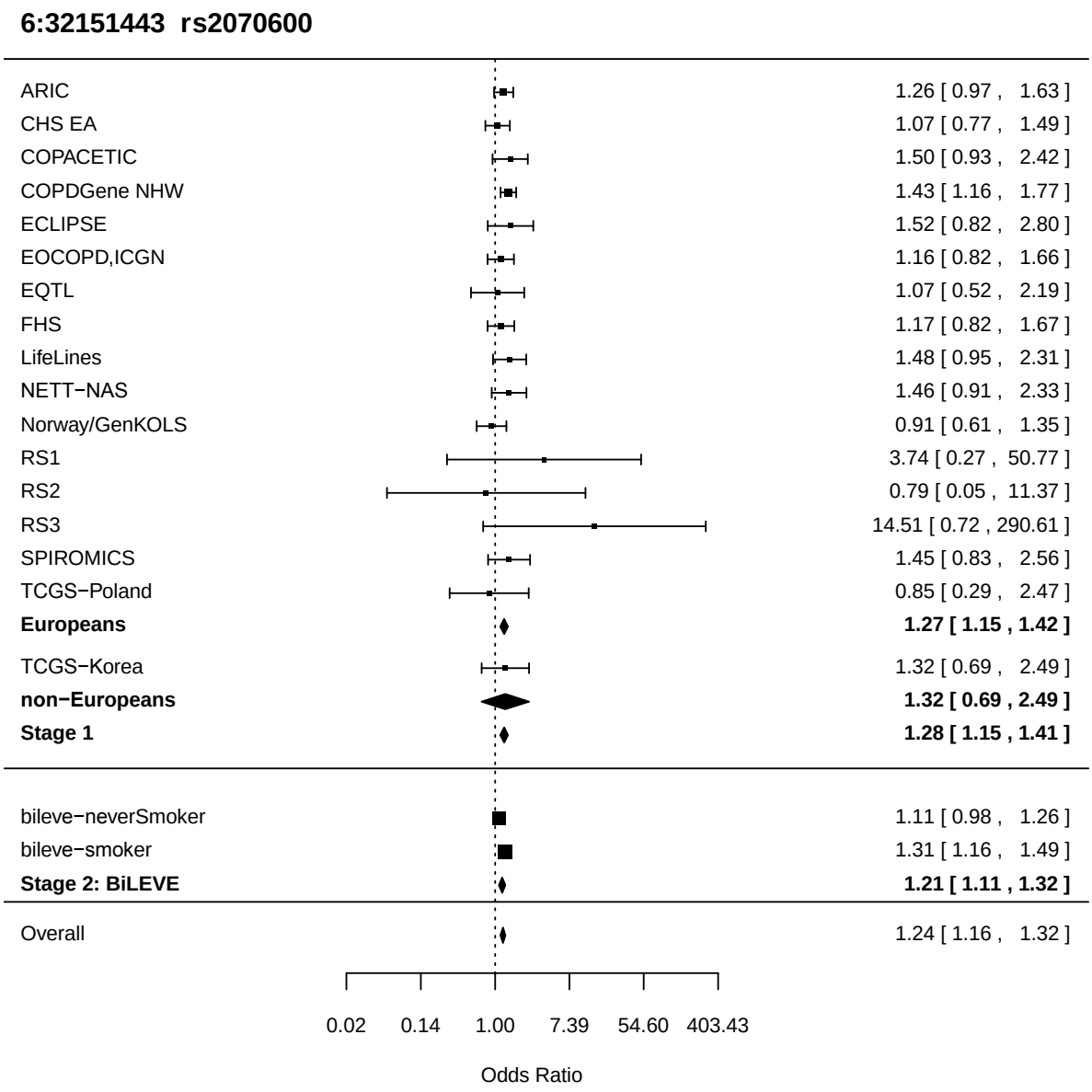


Figure S1o: Forest plot for rs17707300 (*CCDC101*)

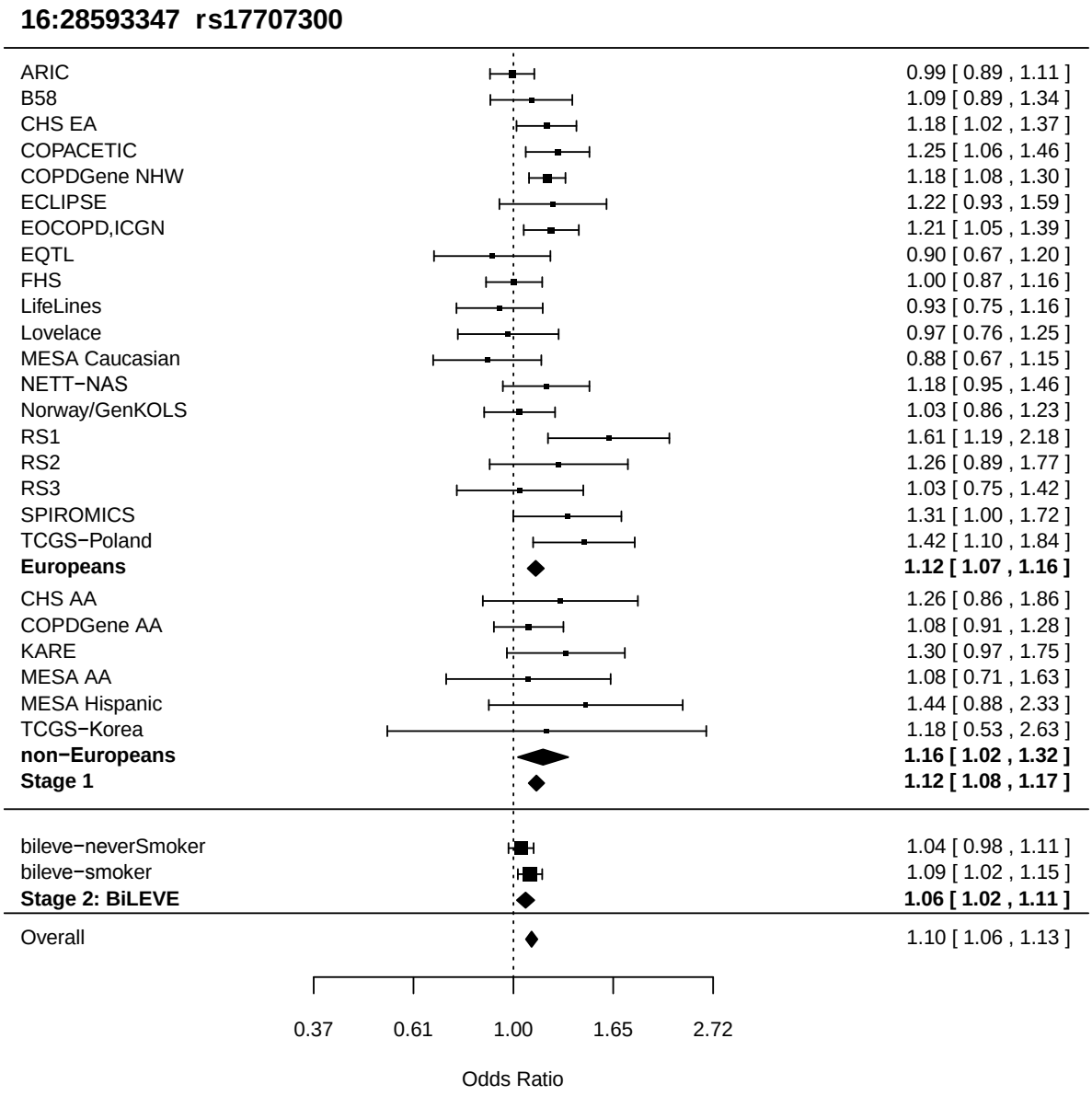


Figure S1p: Forest plot for rs2806356 (*ARMC2*)

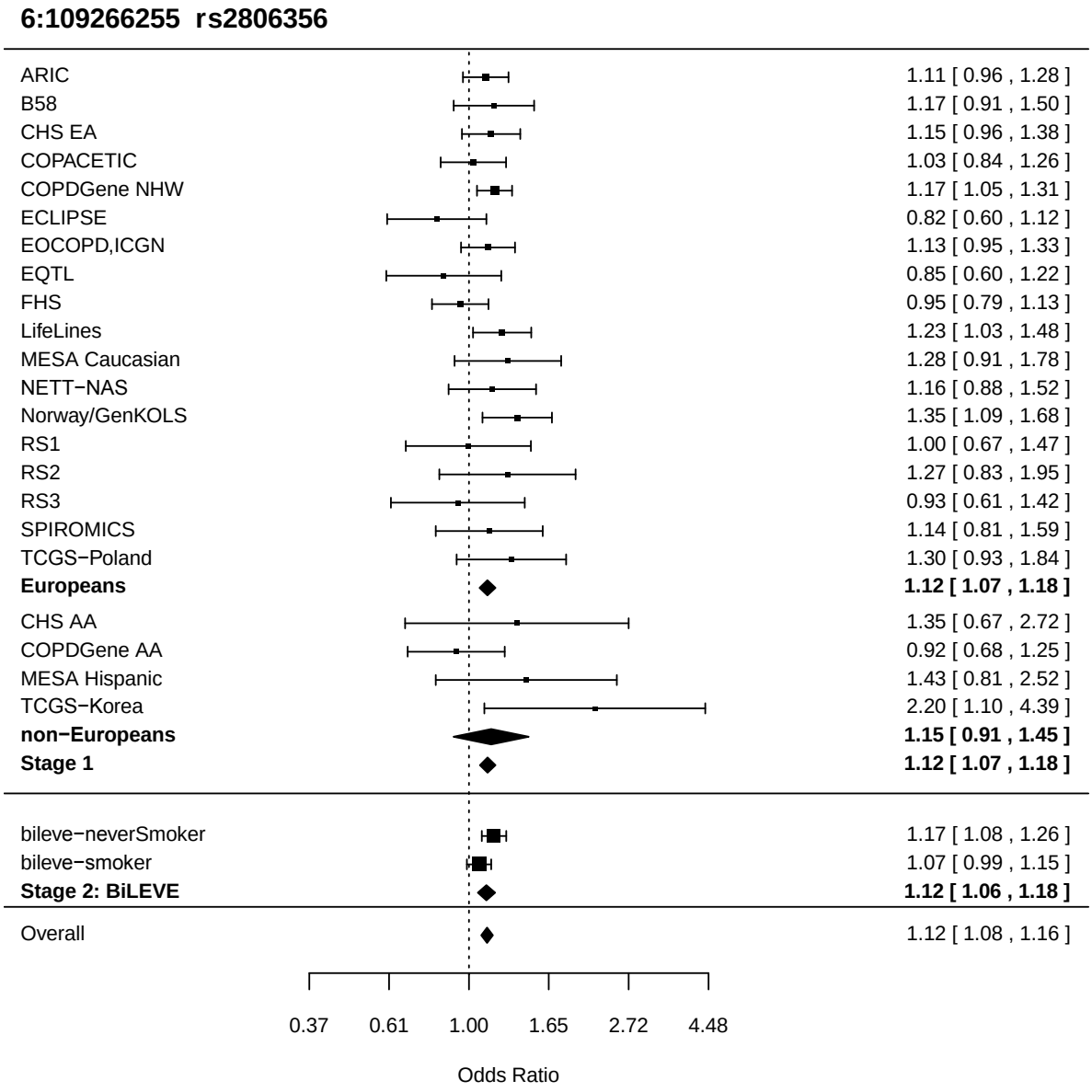


Figure S1q: Forest plot for rs16825267 (*PID1*)

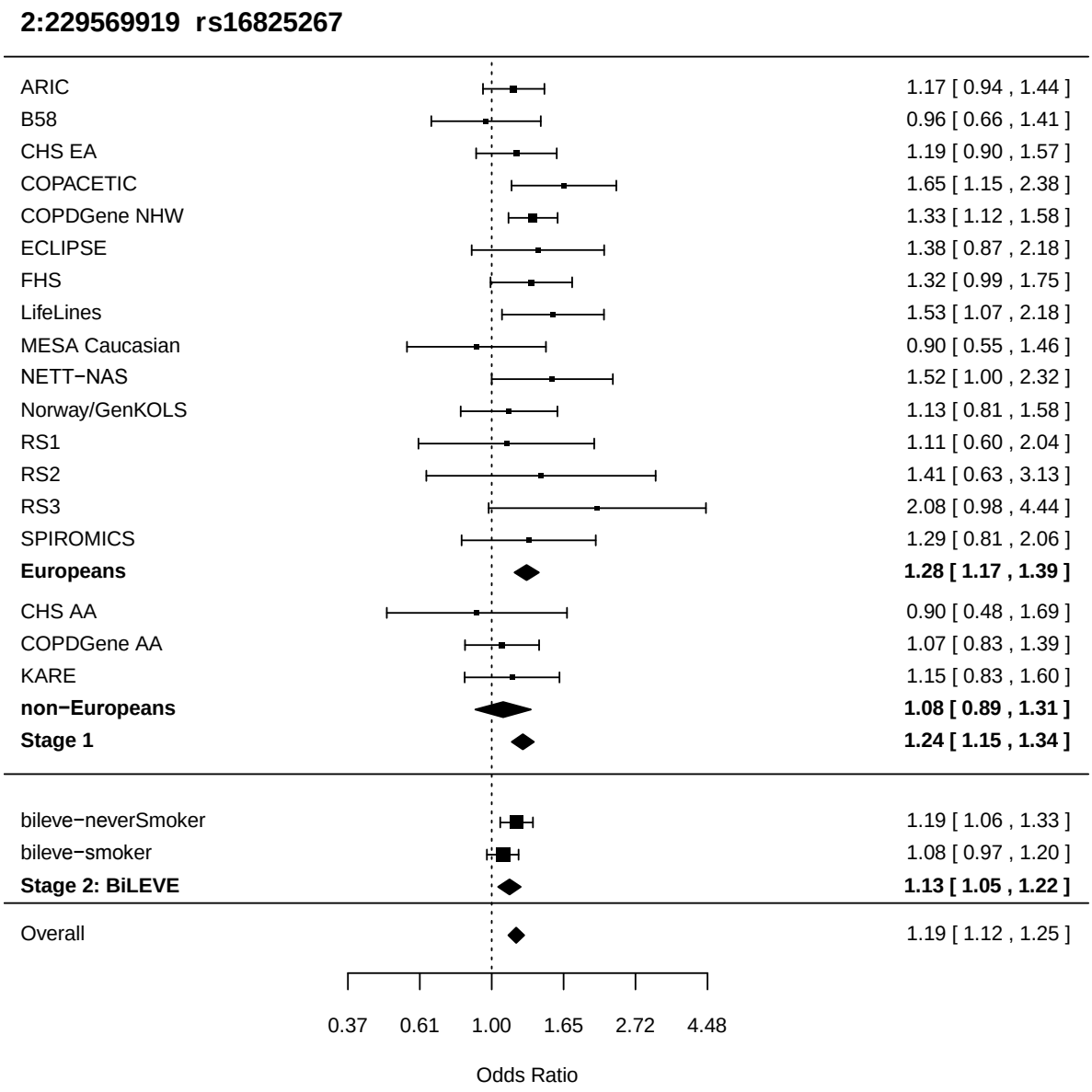


Figure S1r: Forest plot for rs2076295 (*DSP*)

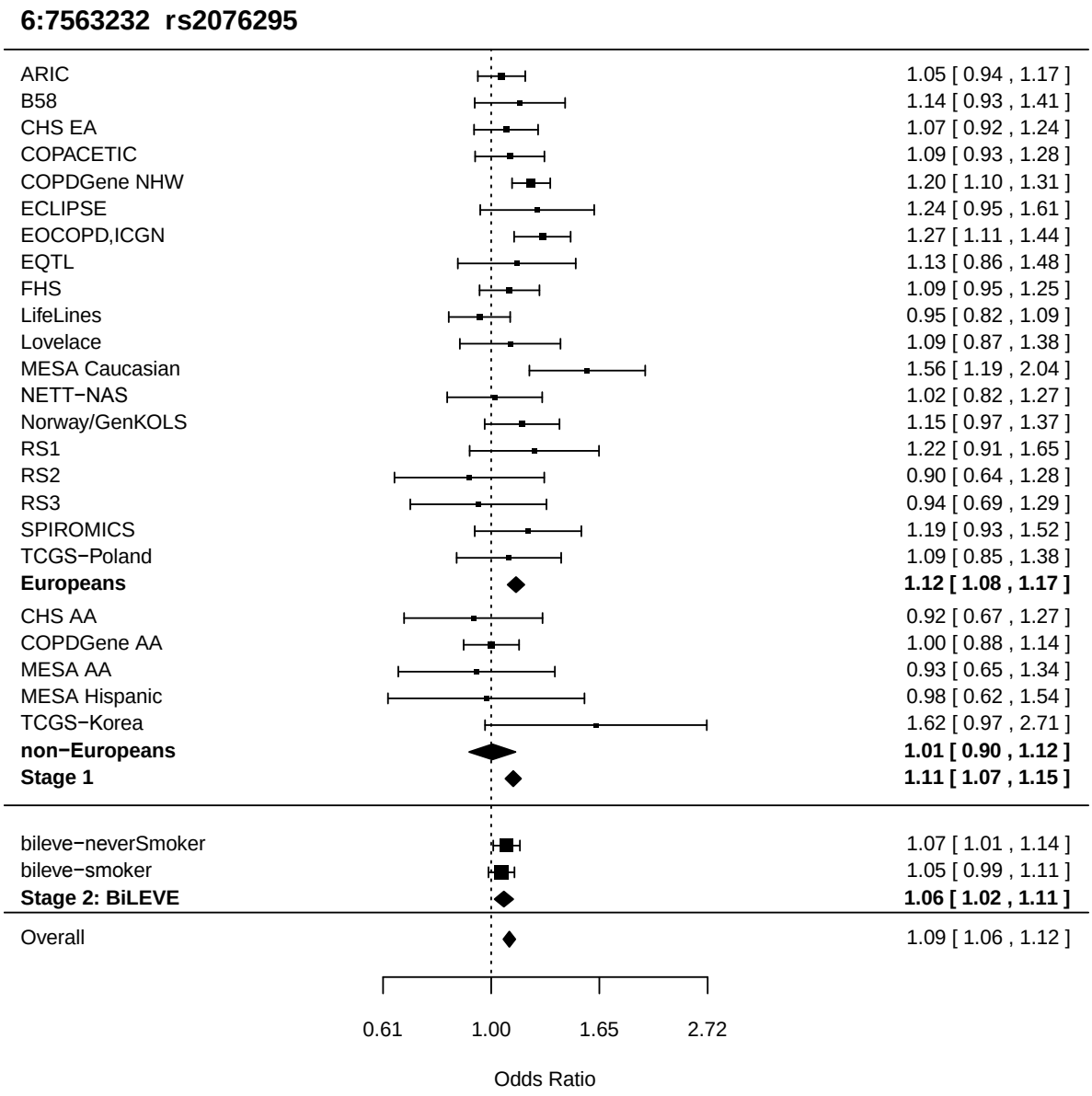


Figure S1s: Forest plot for rs647097 (*MTCL1*)

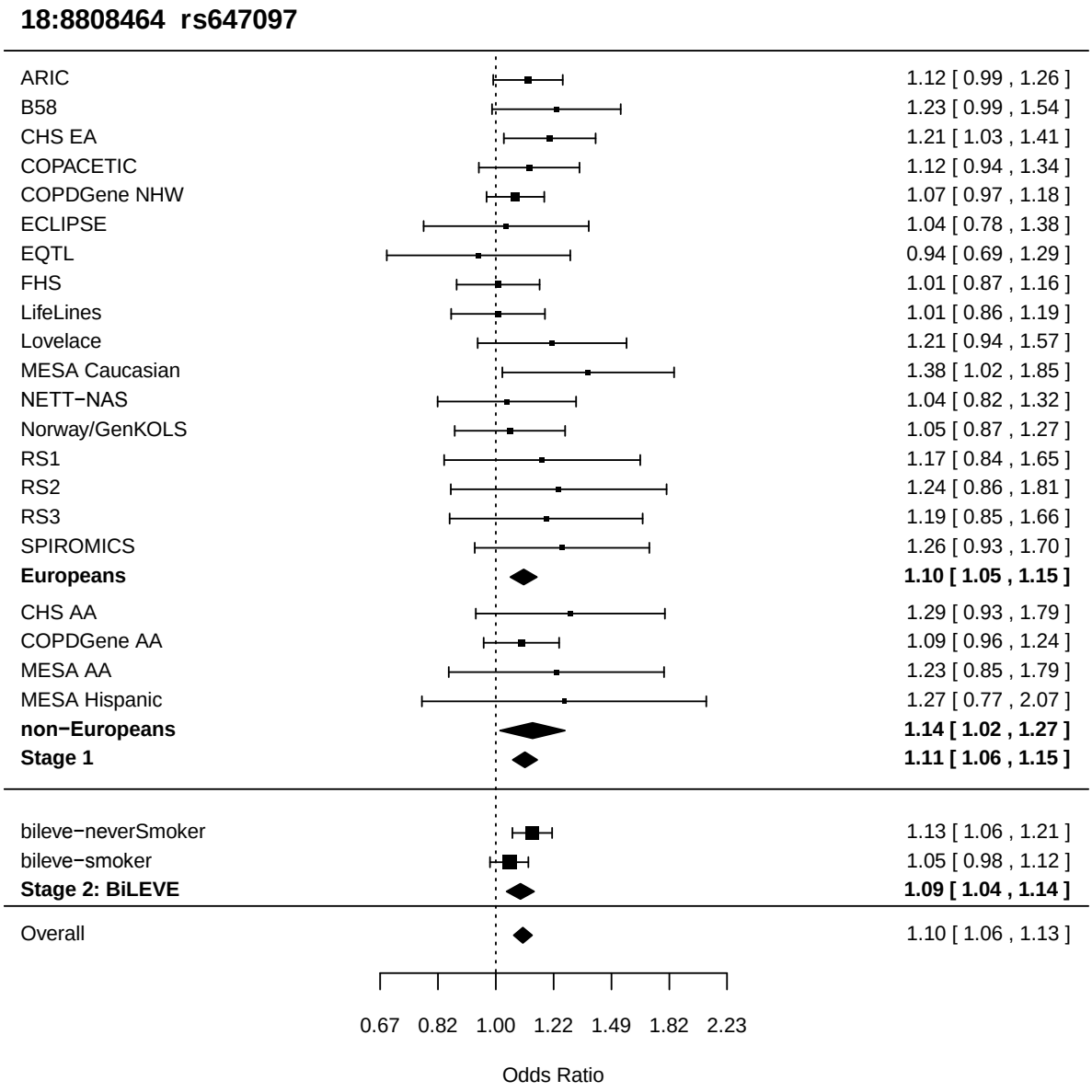


Figure S1t: Forest plot for rs1529672 (*RARB*)

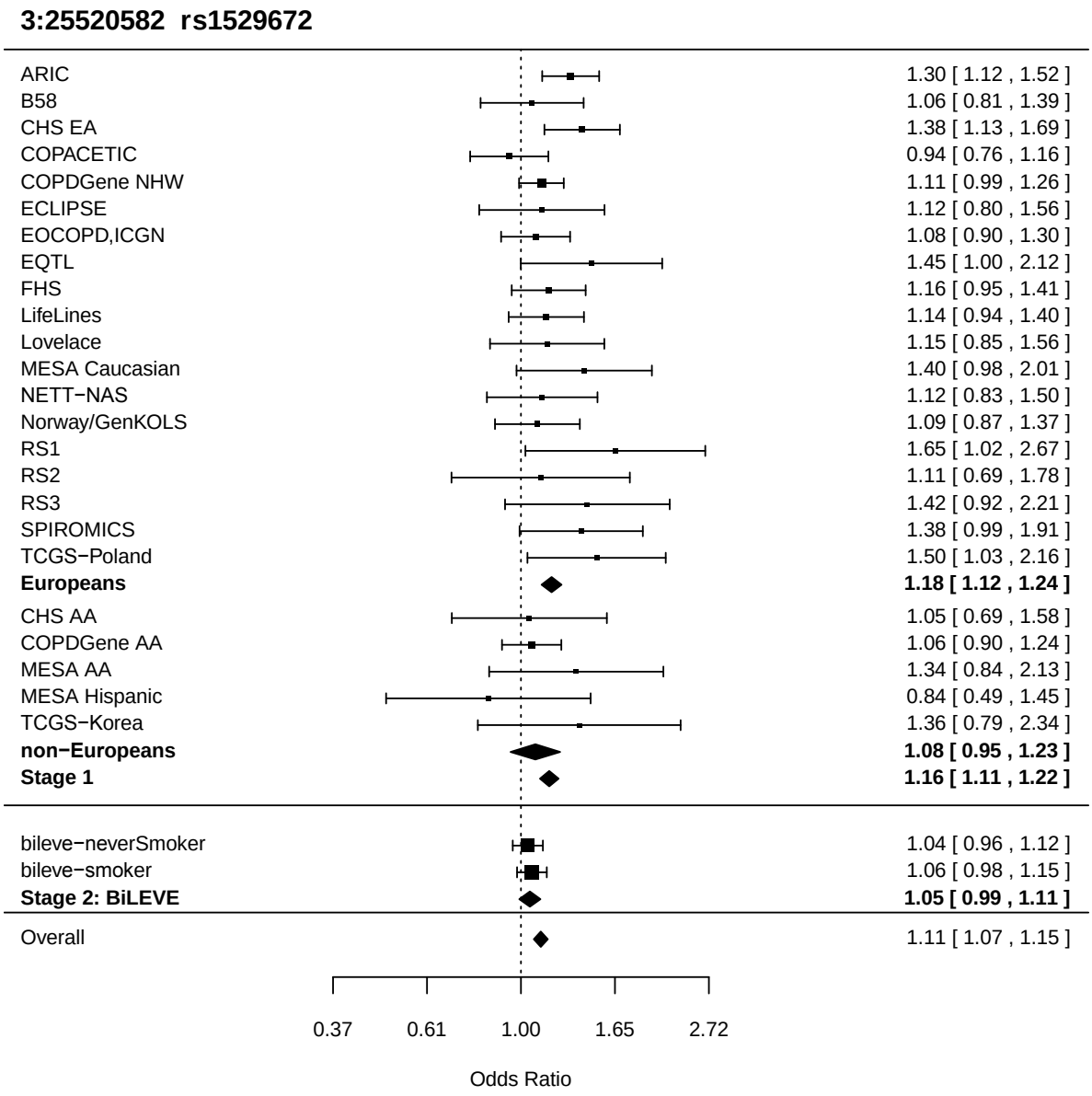


Figure S1u: Forest plot for rs721917 (*STFPD*)

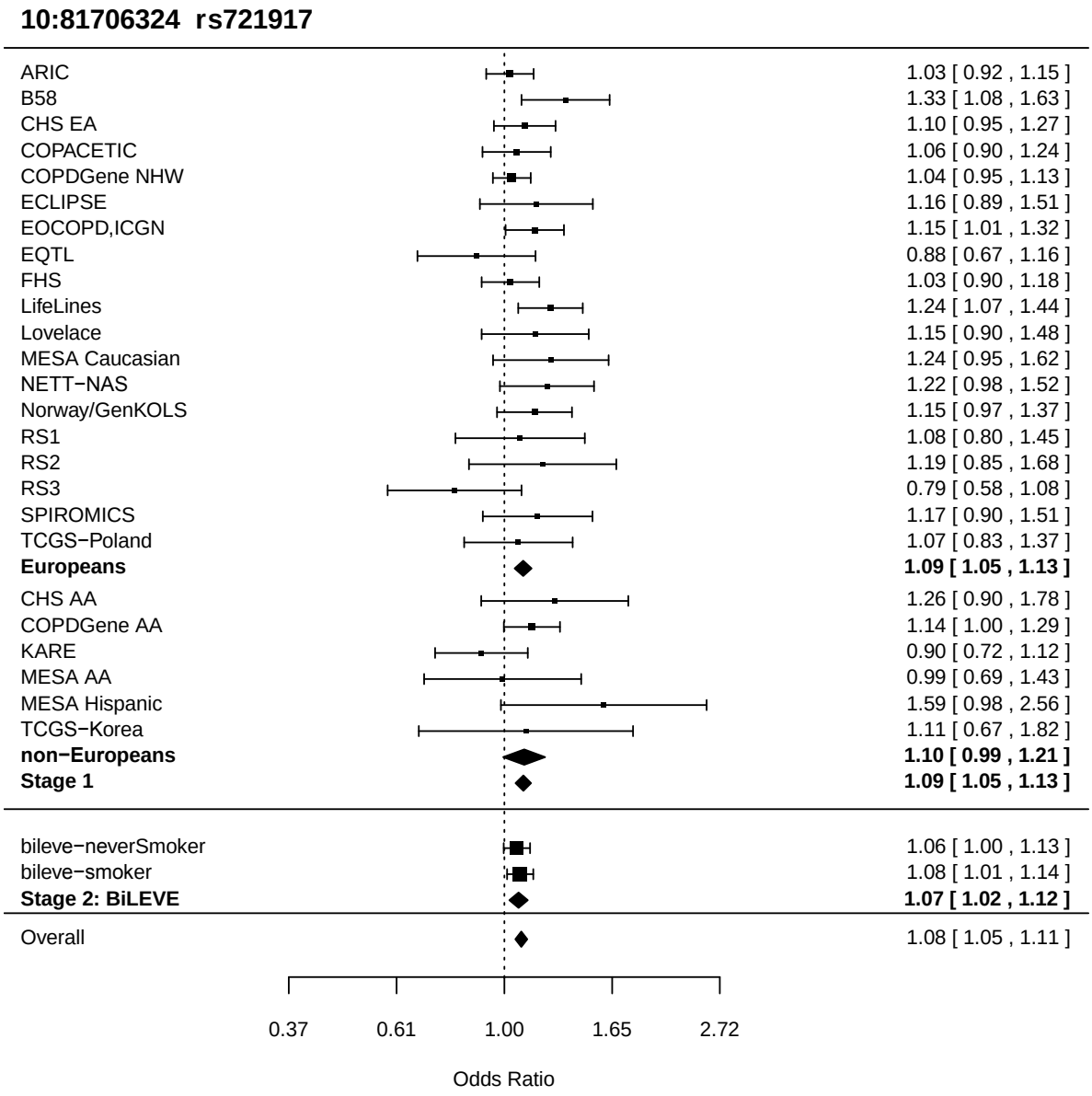


Figure S1v: Forest plot for rs12459249 (*CYP2A6*)

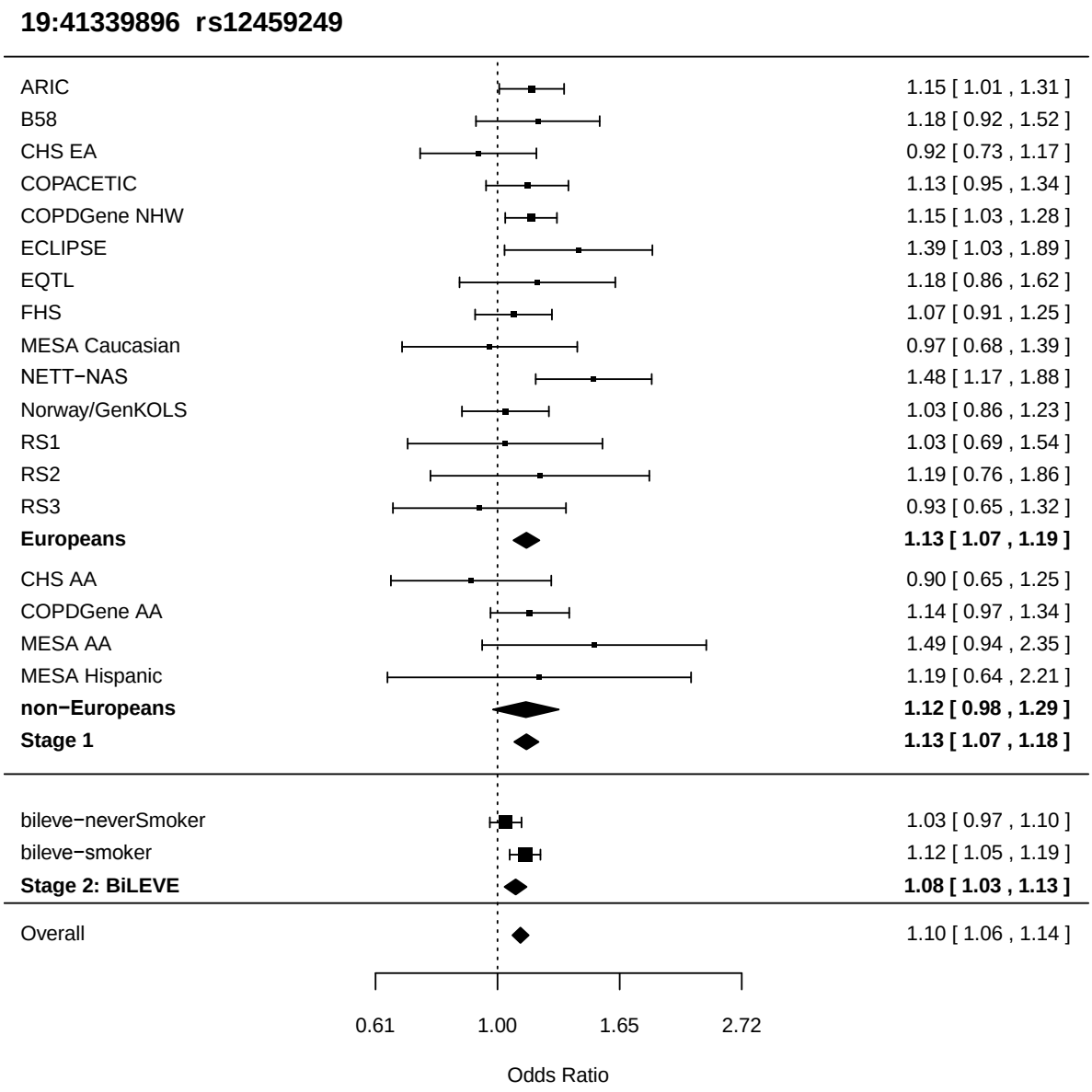


Figure S2a-v: LocusZoom[1] plots for each of the 22 genome-wide significant loci. The point size for each SNP is determined by the sample size for the SNP. The most significant variant in each plot as well as the largest points in each plot were tested in all available cohorts including UK BiLEVE. Smaller points indicate the variant was not tested in UK BiLEVE Stage 2 analysis and thus the -log10 P value is taken from the Stage 1 analysis.

Figure S2a: LocusZoom for rs13141641 (*HHIP*)

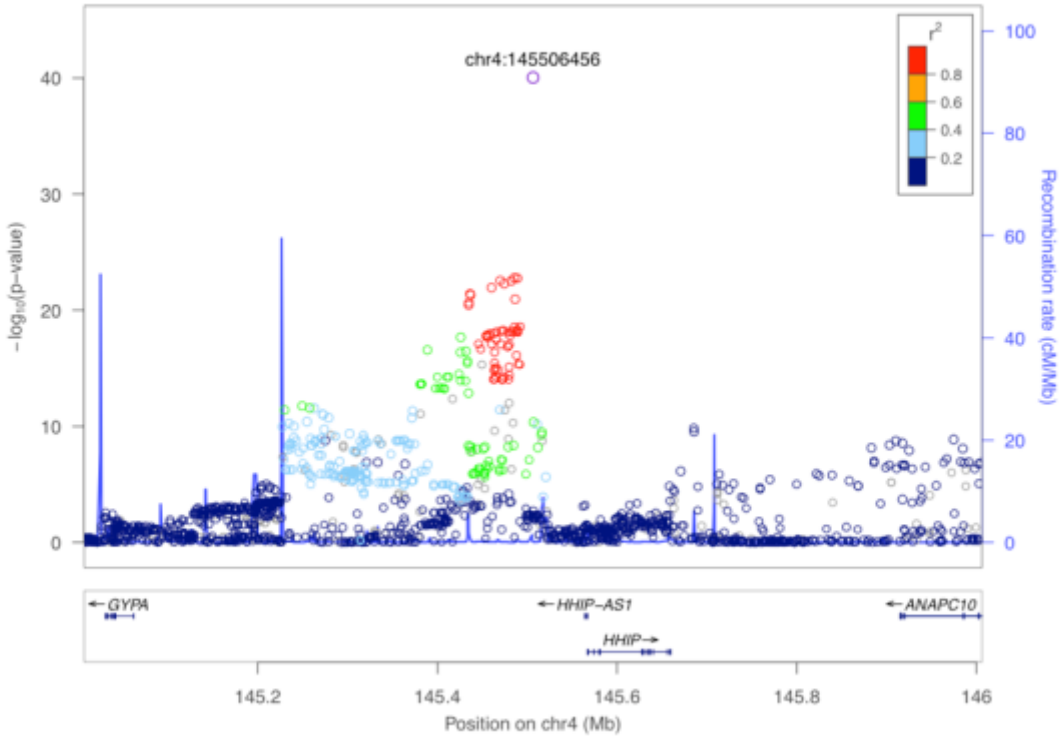


Figure S2b: LocusZoom for rs17486278 (*CHRNA5*)

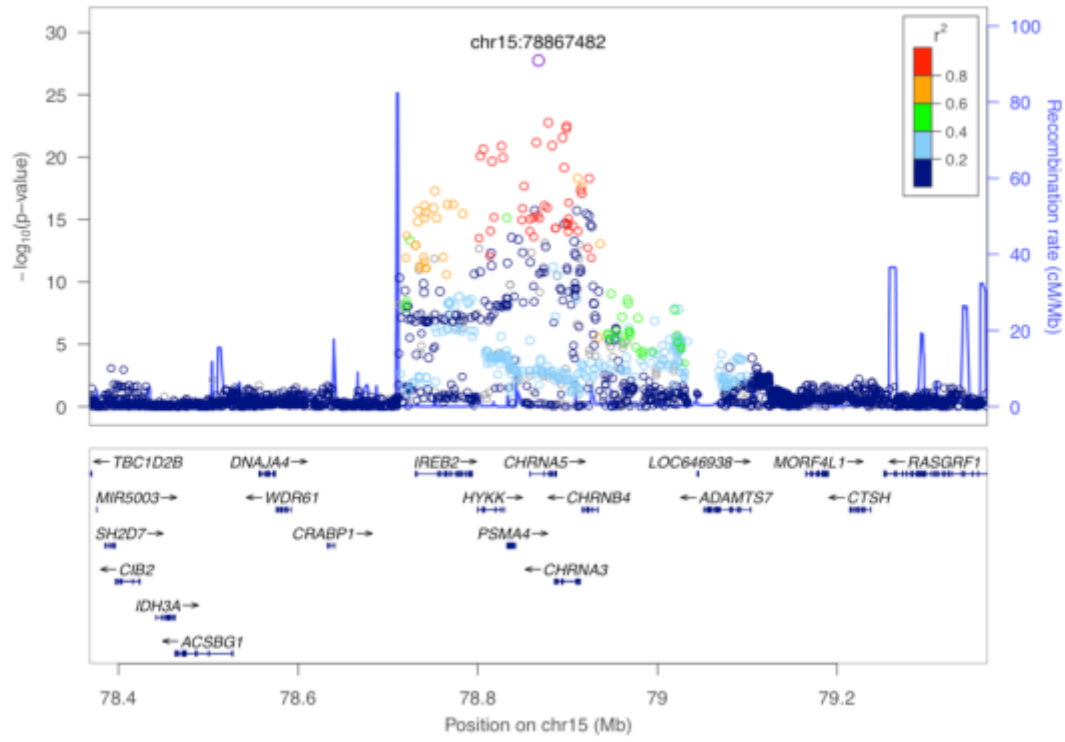


Figure S2c: LocusZoom for rs7733088 (*HTR4*)

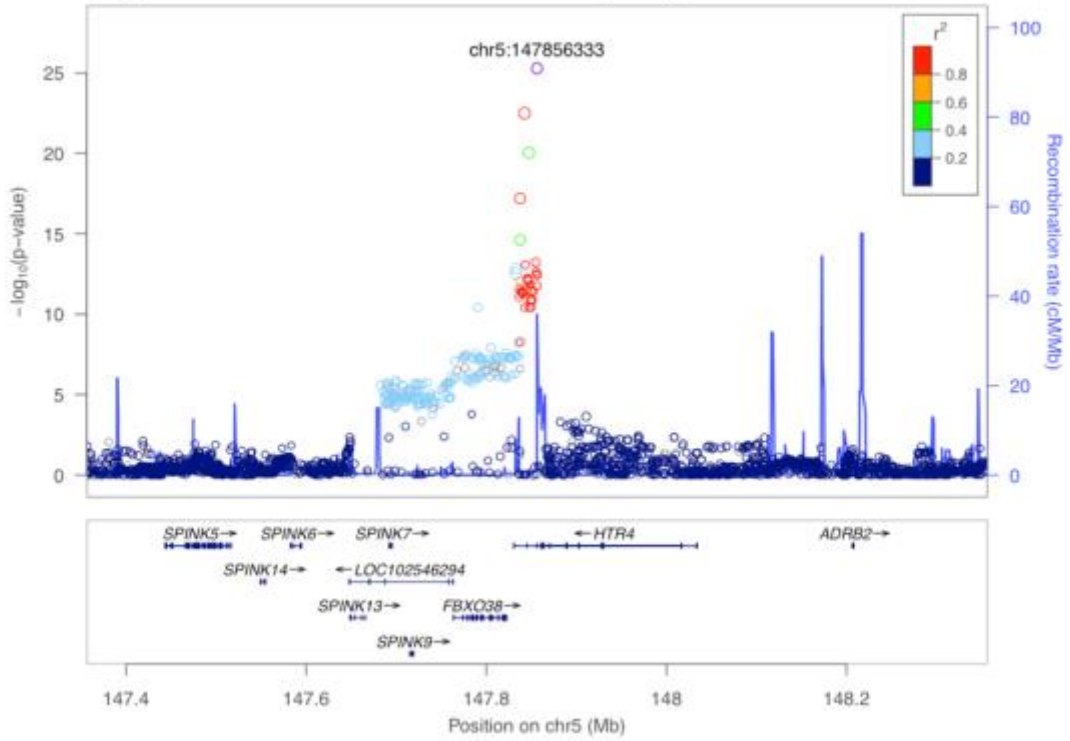


Figure S2d: LocusZoom for rs9399401 (*ADGRG6*)

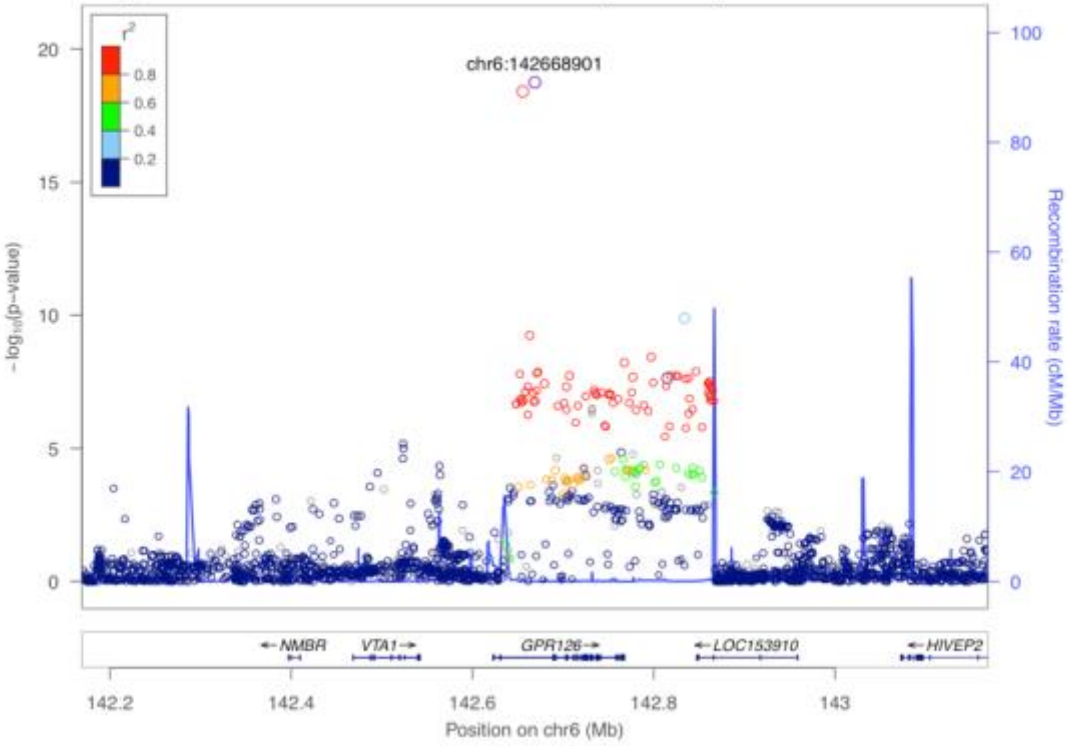


Figure S2e: LocusZoom for rs1441358 (*THSD4*)

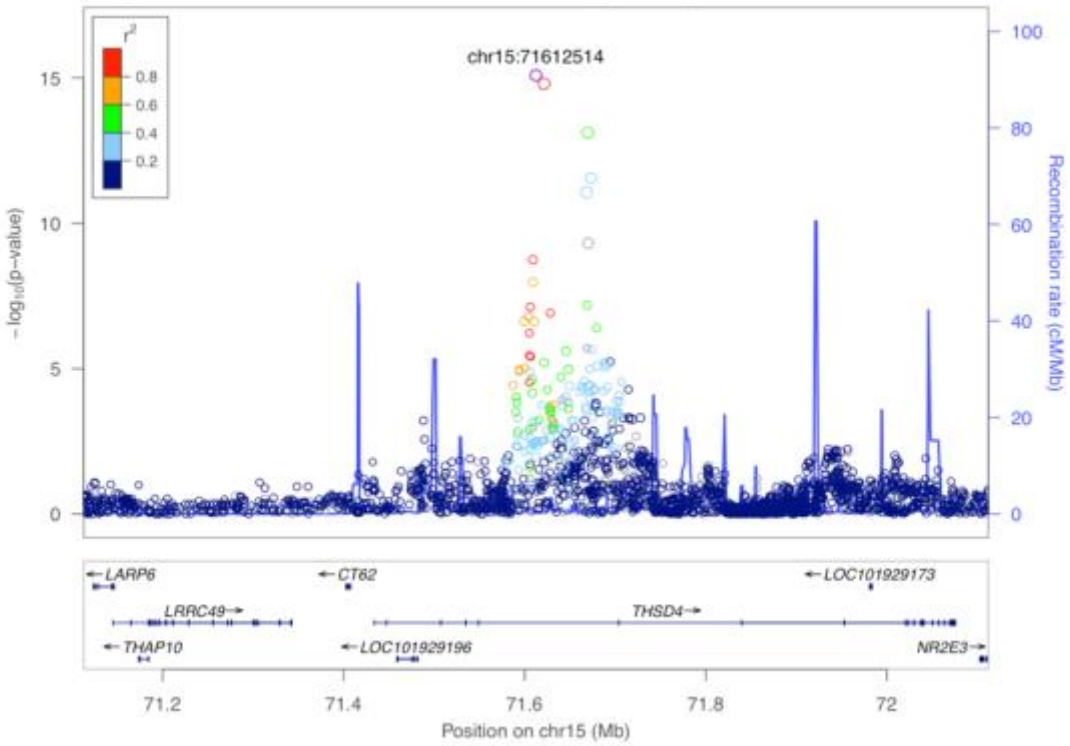


Figure S2f: LocusZoom for rs6837671 (*FAM13A*)

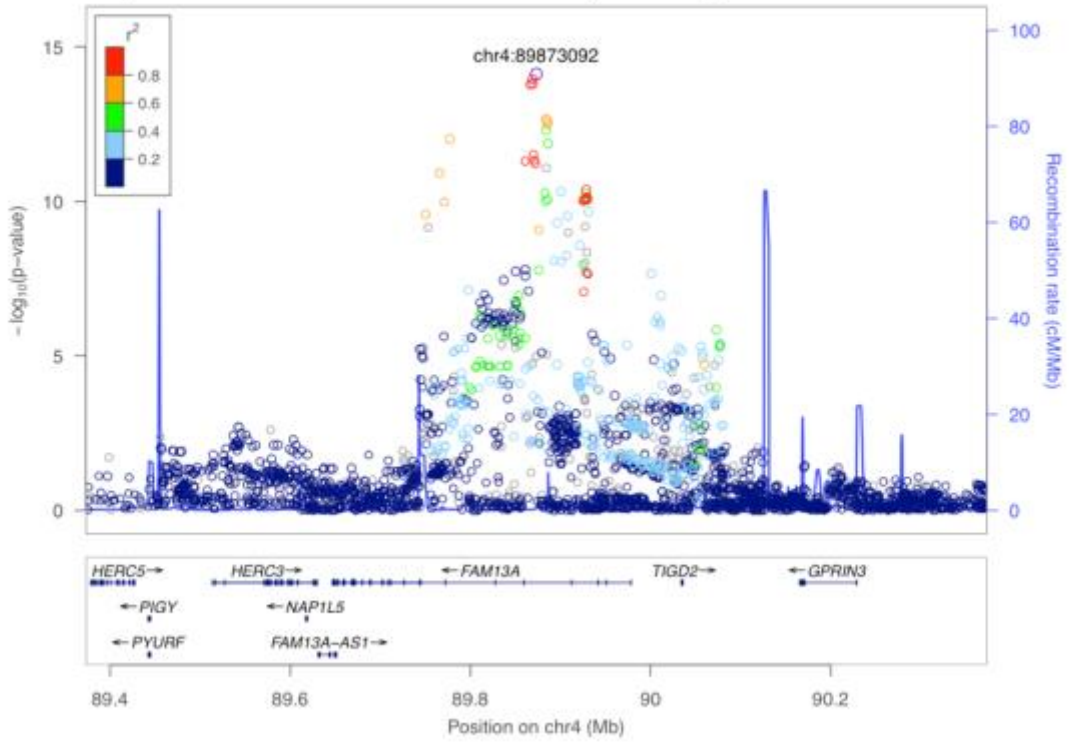


Figure S2g: LocusZoom for rs11727735 (*GSTCD*)

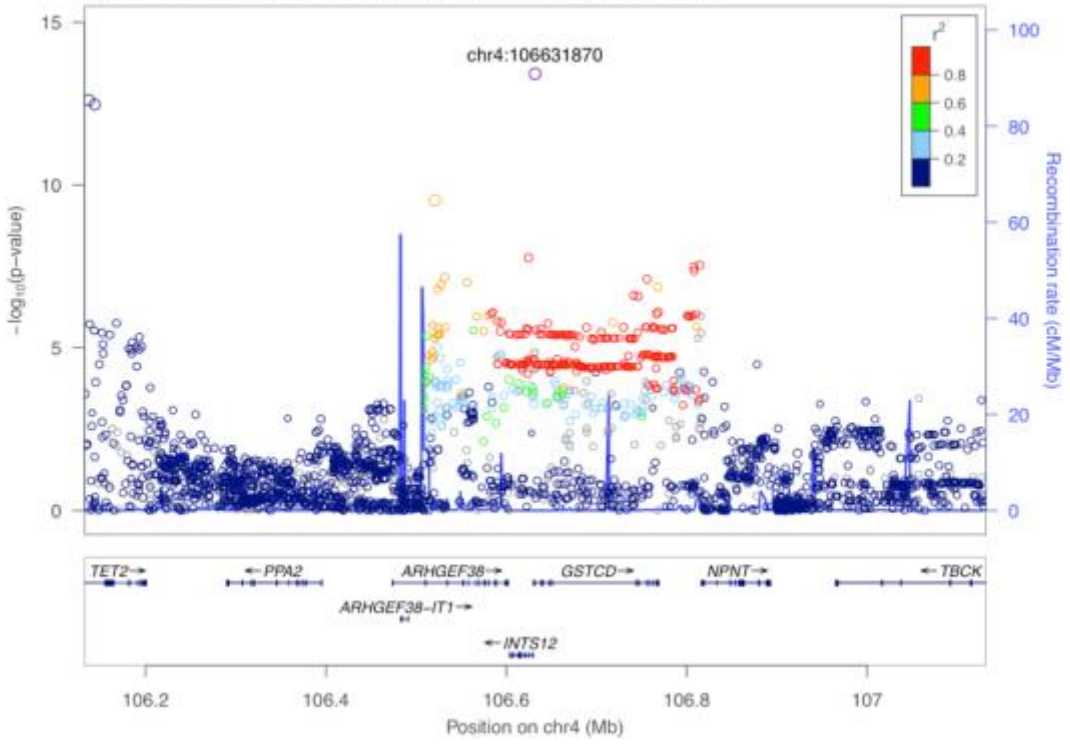


Figure S2h: LocusZoom for rs754388 (*RIN3*)

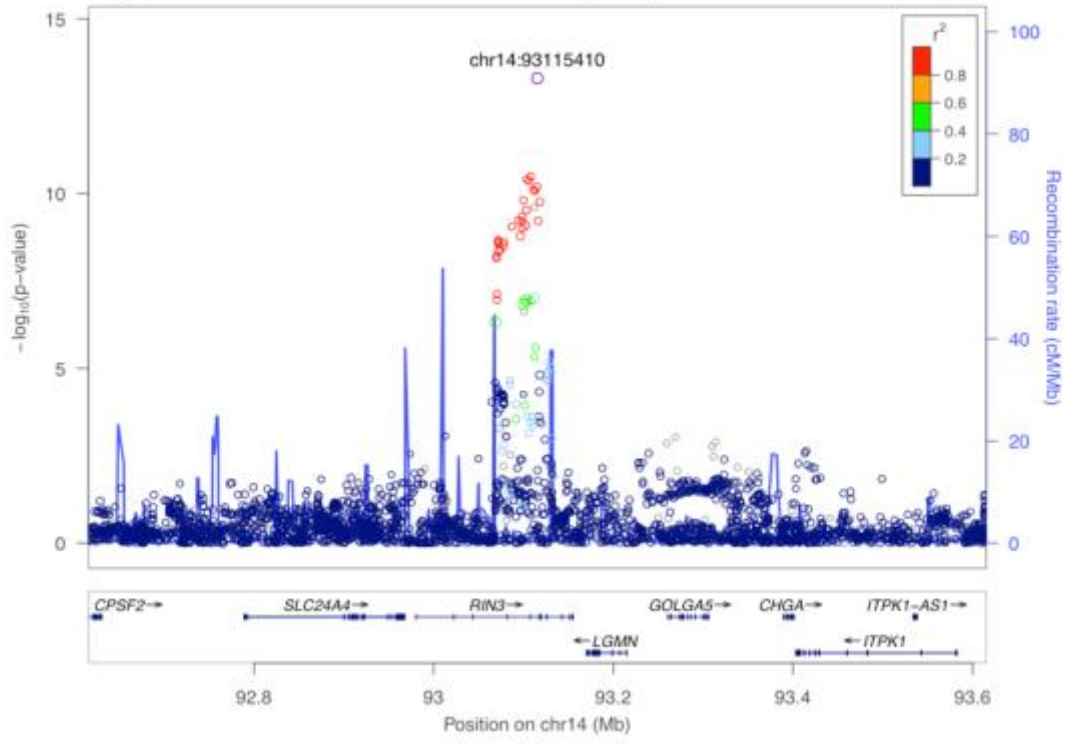


Figure S2i: LocusZoom for rs113897301 (*ADAM19*)

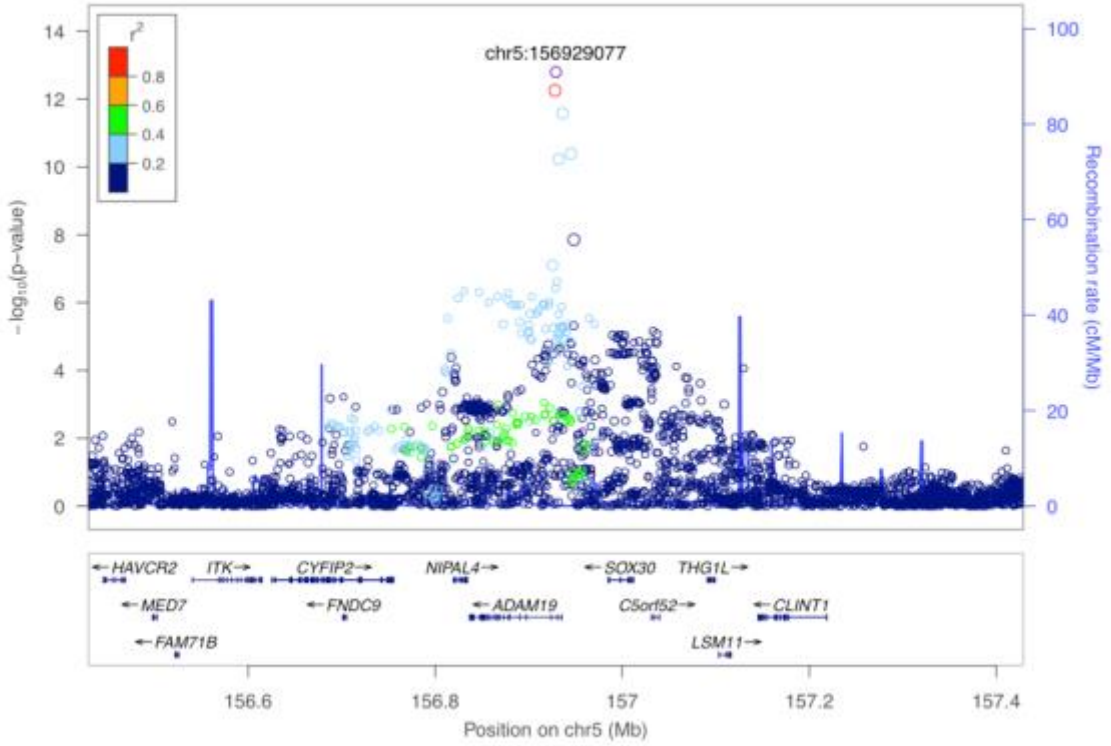


Figure S2j: LocusZoom for rs2047409 (*TET2*)

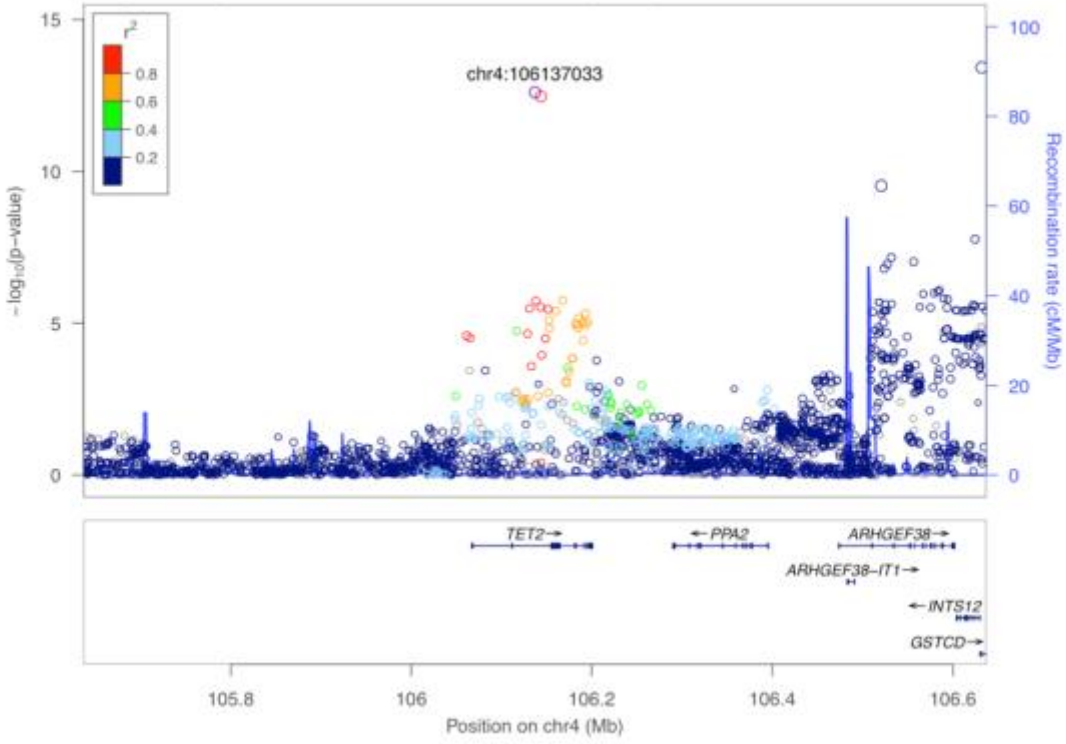


Figure S2k: LocusZoom for rs2955083 (*EEFSEC*)

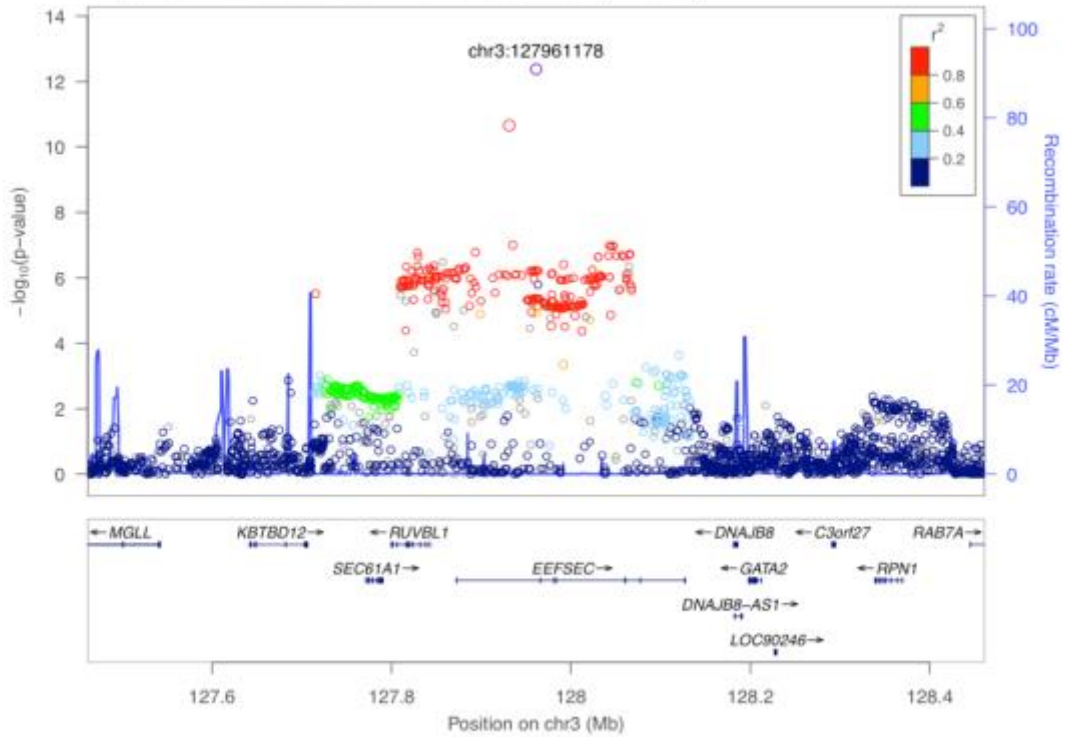


Figure S2l: LocusZoom for rs7186831 (*CFDP1*)

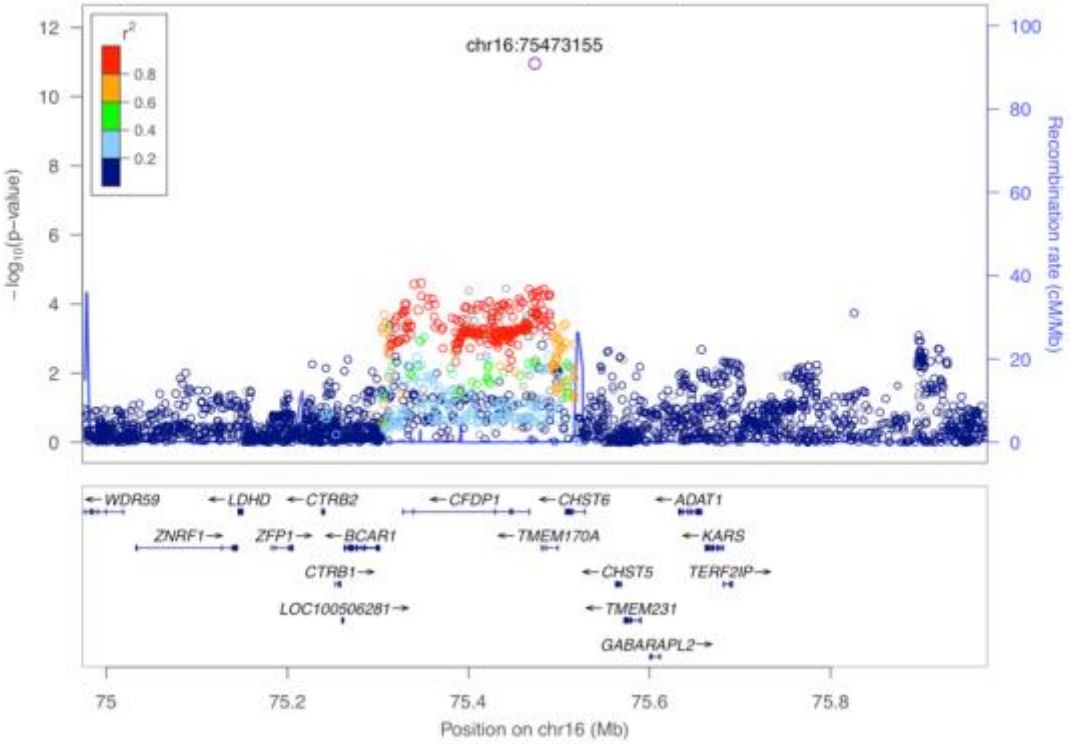


Figure S2m: LocusZoom for rs10429950 (*TGFB2*)

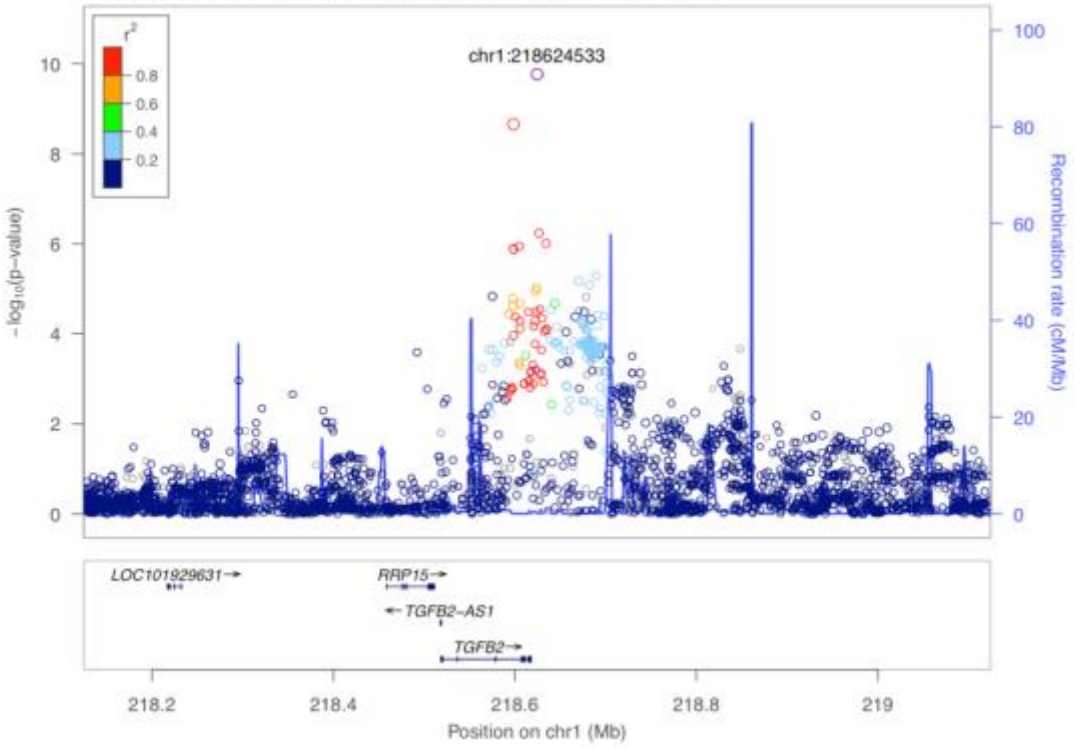


Figure S2n: LocusZoom for rs2070600 (*AGER*)

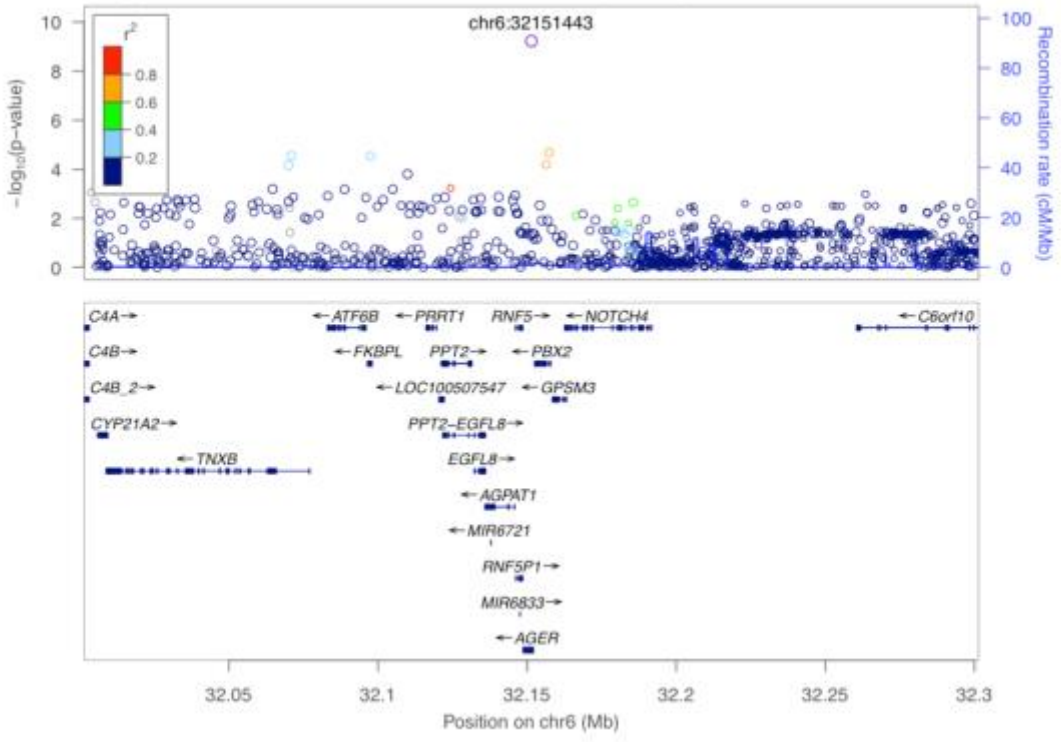


Figure S2o: LocusZoom for rs17707300 (*CCDC101*)

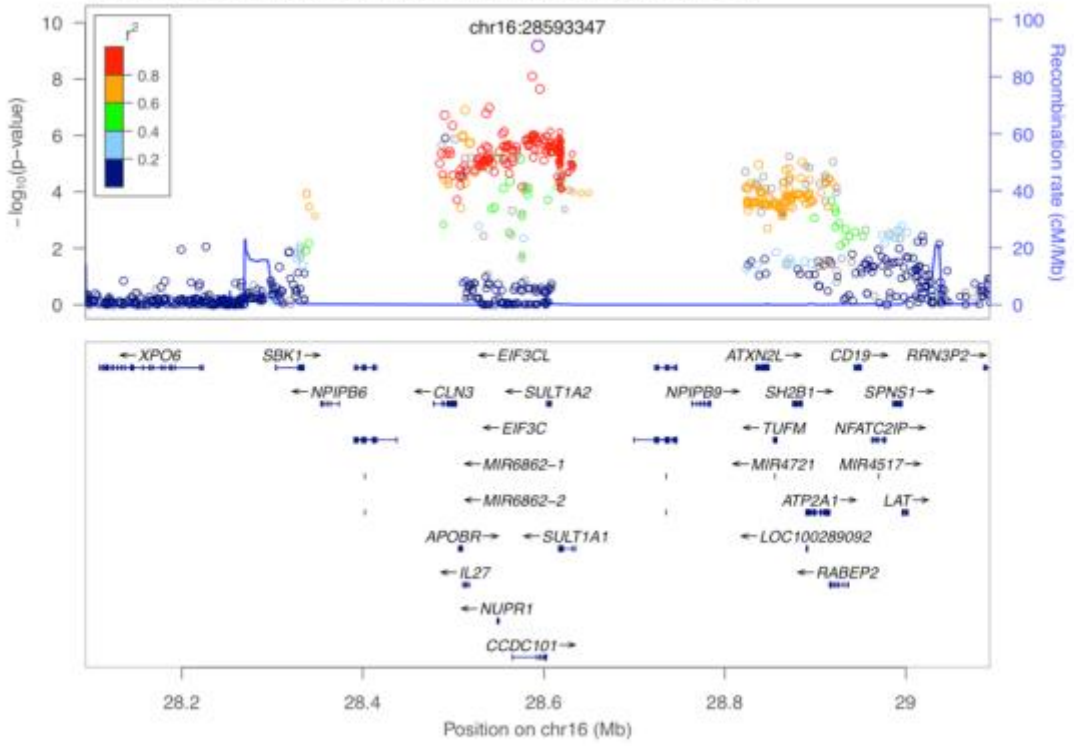


Figure S2p: LocusZoom for rs2806356 (*ARMC2*)

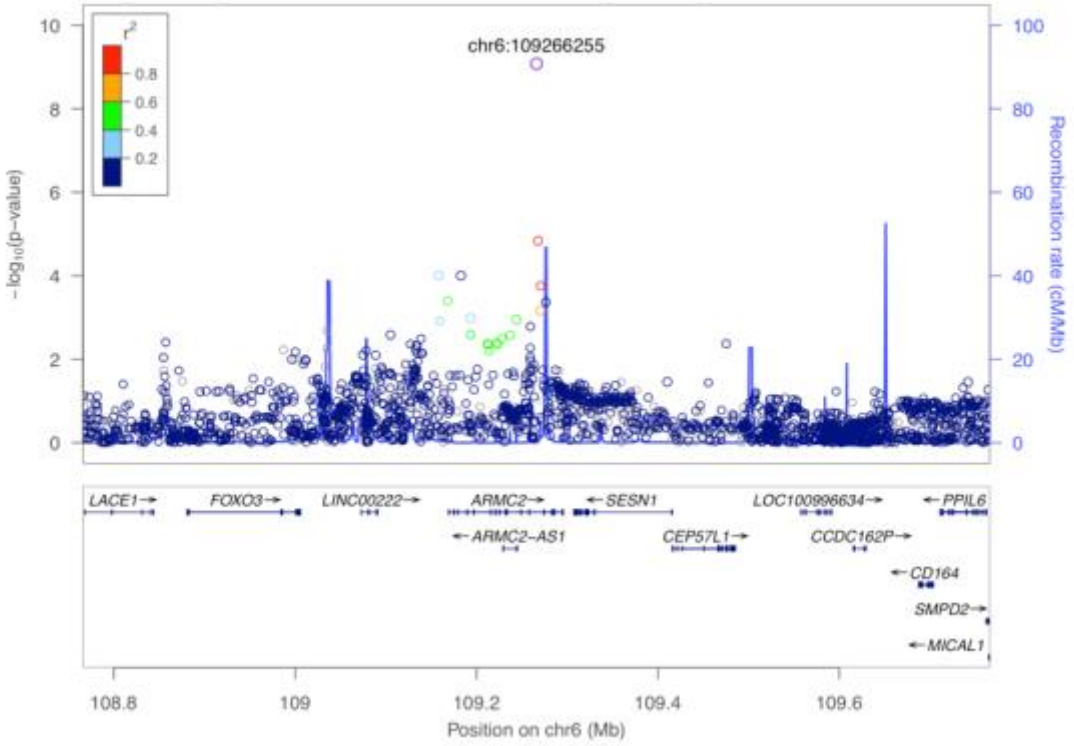


Figure S2q: LocusZoom for rs16825267 (*PID1*)

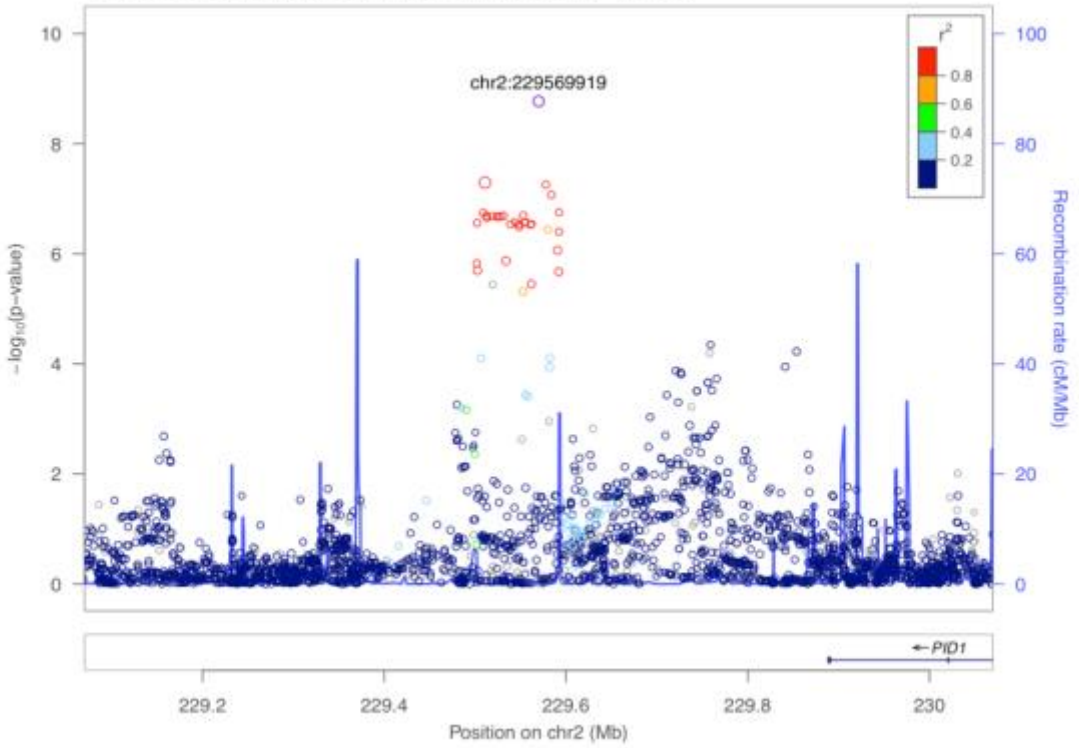


Figure S2r: LocusZoom for rs2076295 (*DSP*)

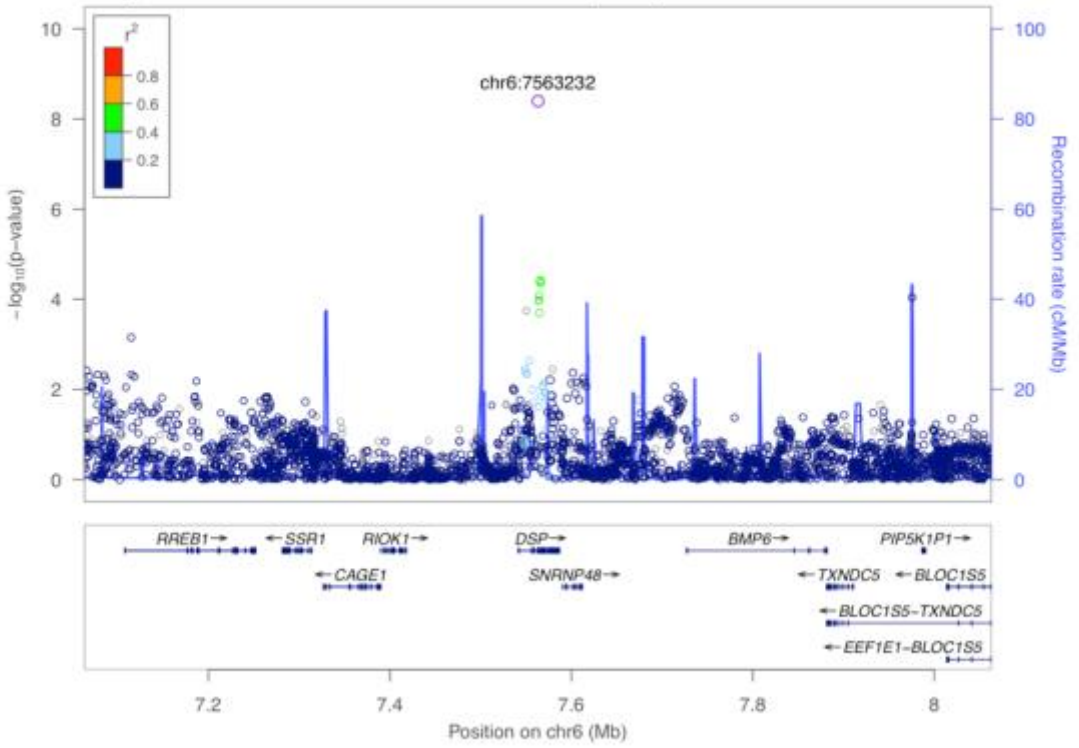


Figure S2s: LocusZoom for rs647097 (*MTCL1*)

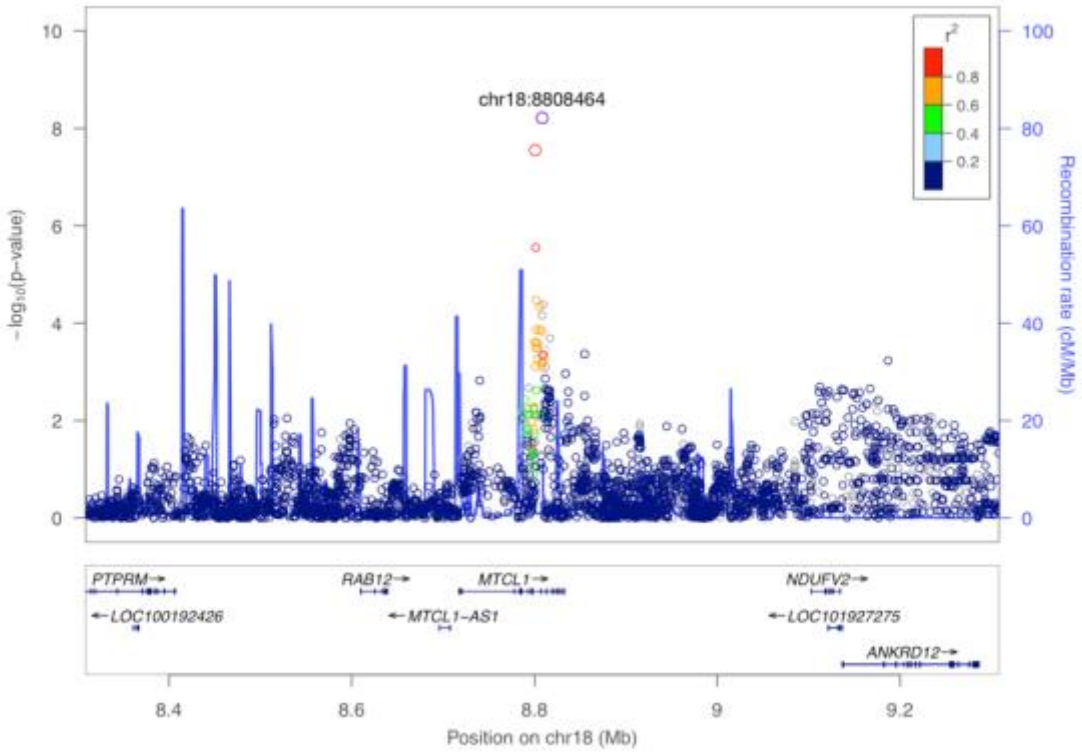


Figure S2t: LocusZoom for rs1529672 (*RARB*)

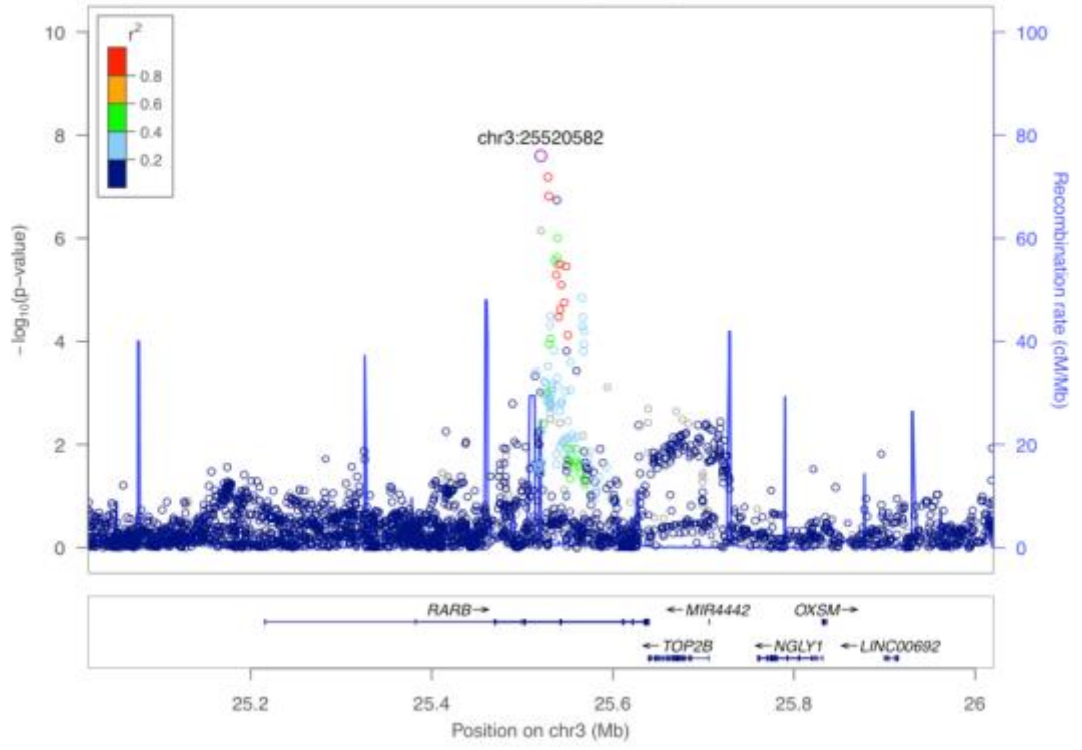


Figure S2u: LocusZoom for rs721917 (*STFPD*)

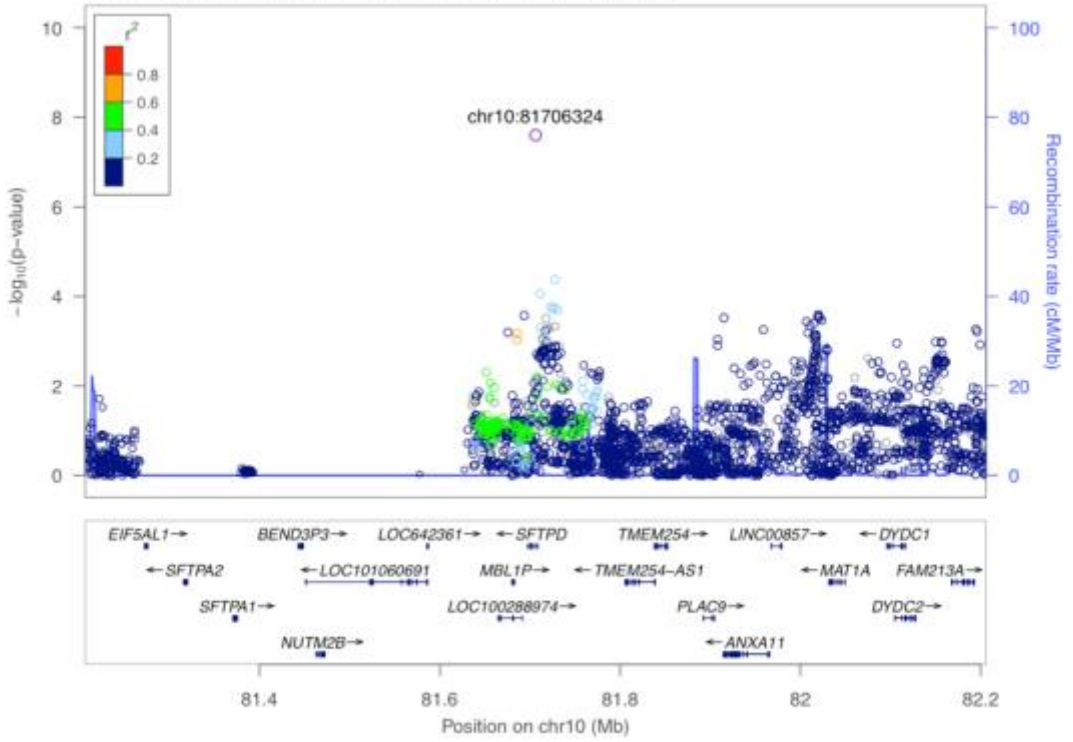


Figure S2v: LocusZoom for rs12459249 (*CYP2A6*)

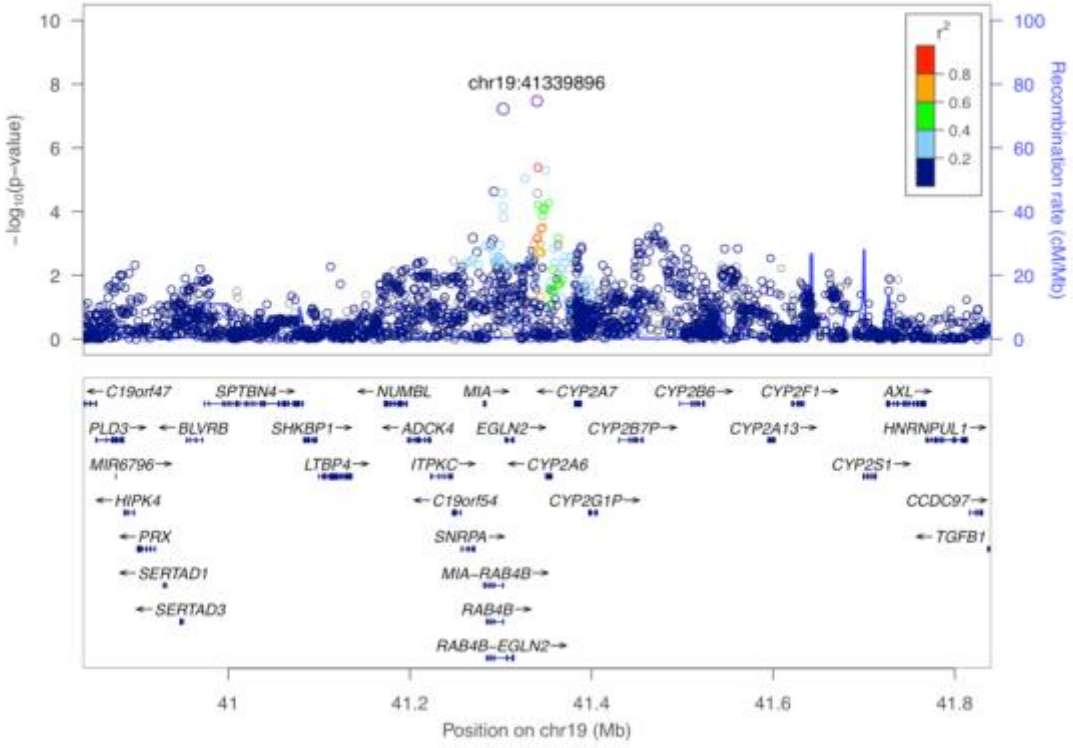
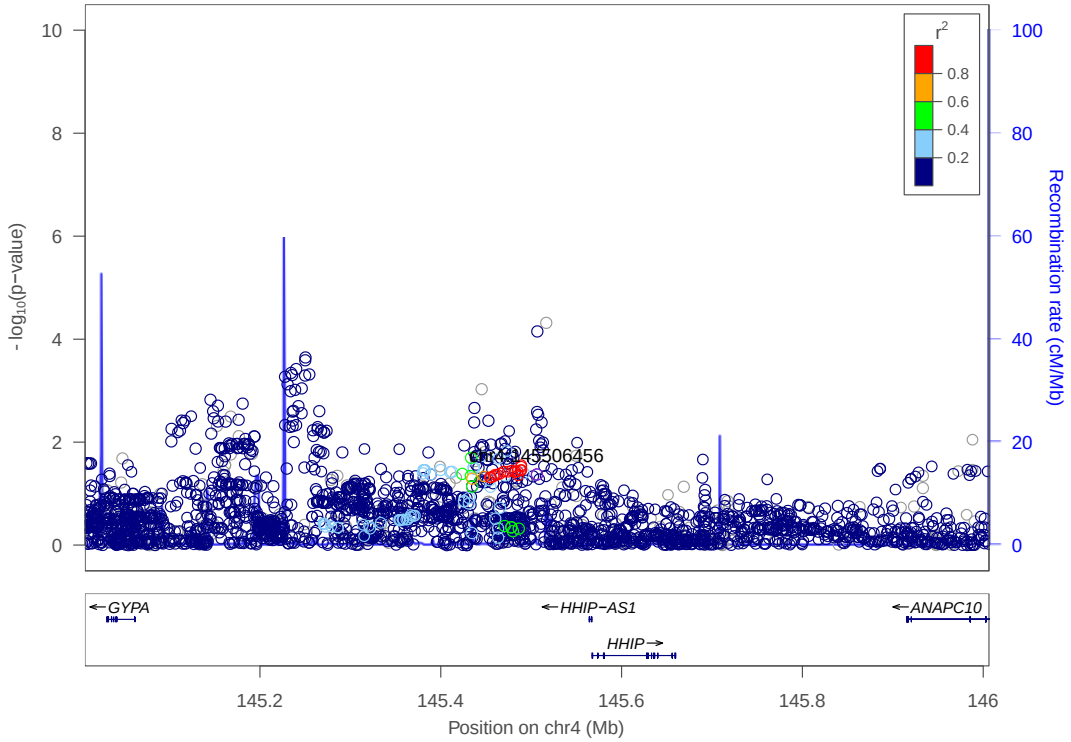
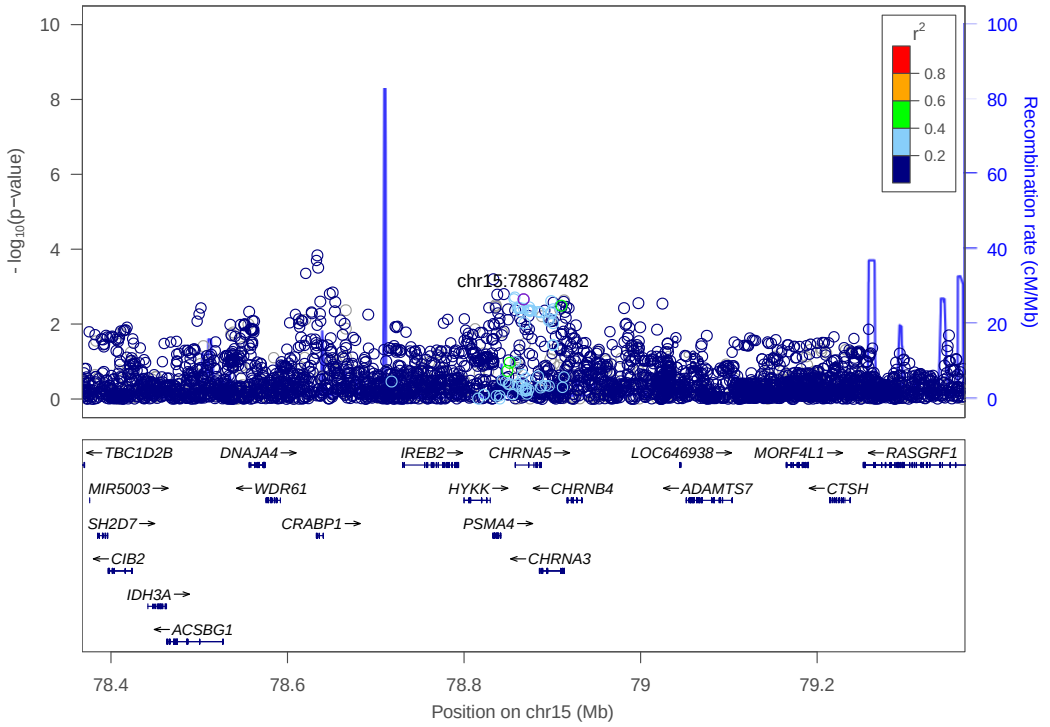
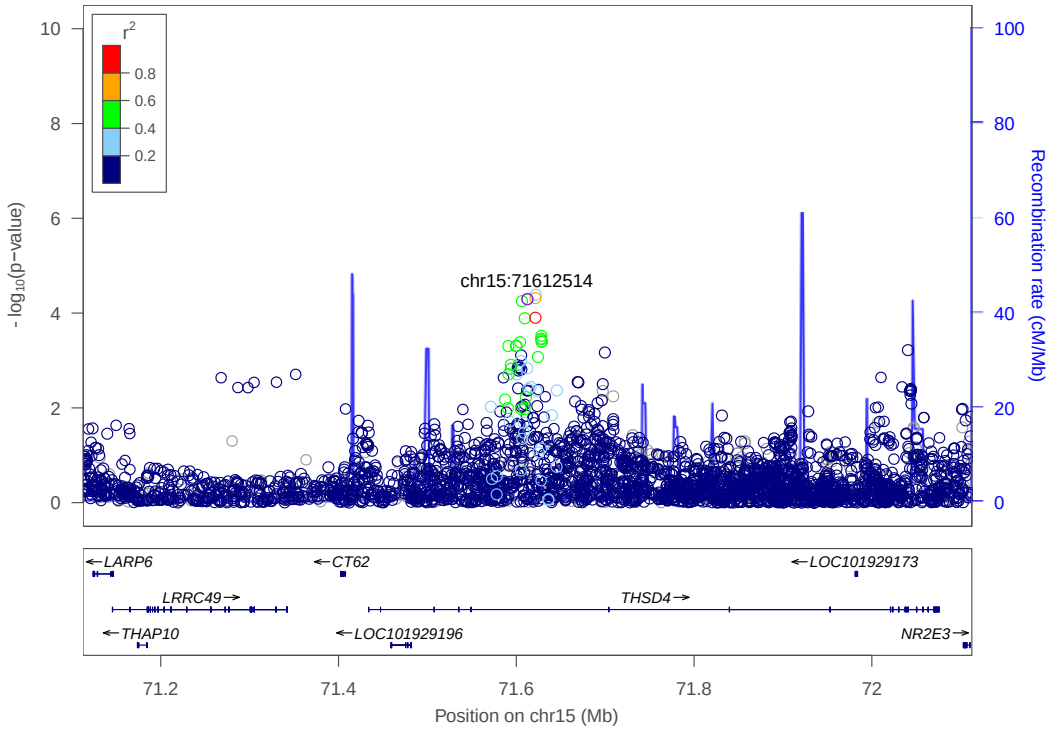
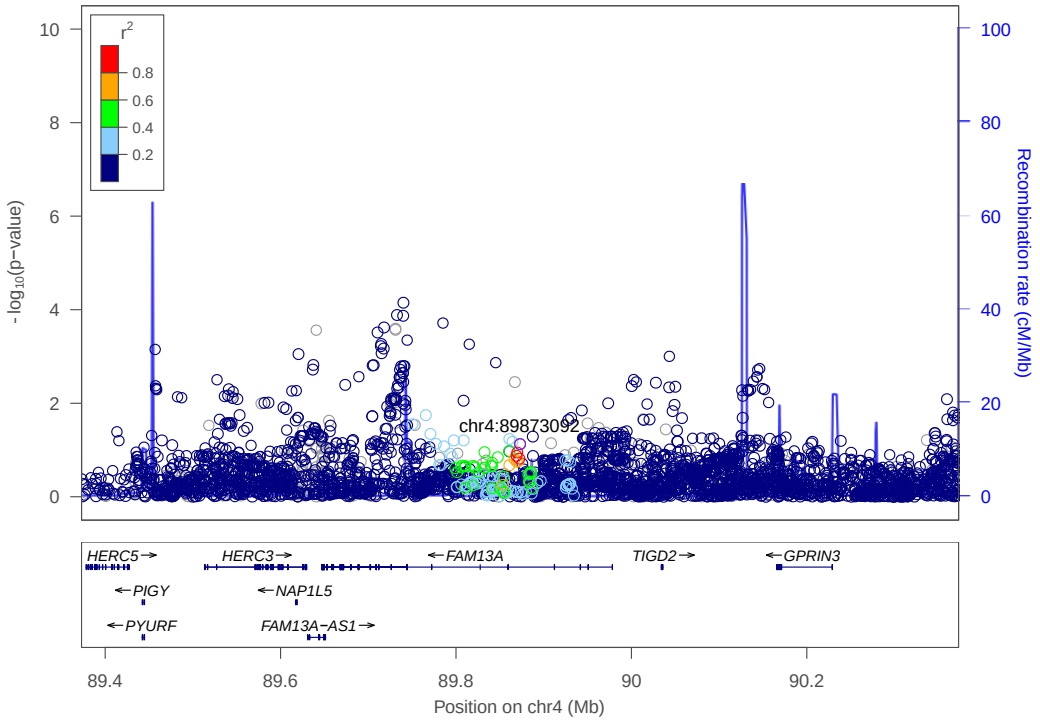


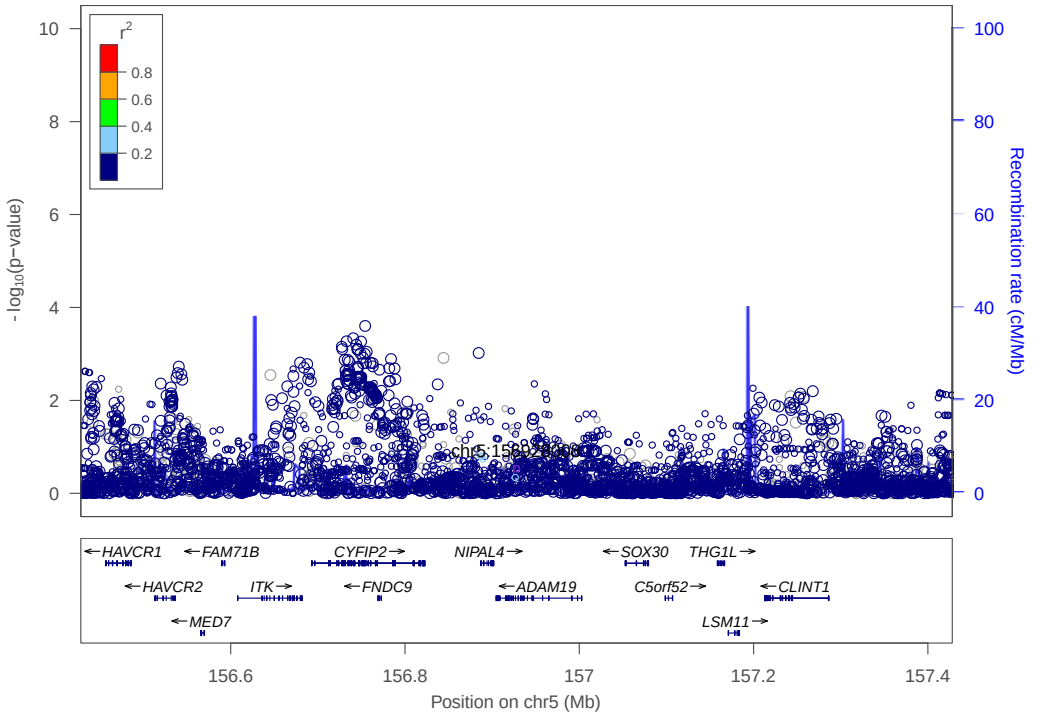
Figure S3a-xxxxx: LocusZoom[1] plots for the meta-analysis across the three Stage 1 cohorts of African Ancestry, including the Cardiovascular Health Study (CHS) African Americans, COPDGene study African American, and the Multi-Ethnic Study of Atherosclerosis (MESA) African Americans. The loci shown each had an association signal with P < 0.001 in the meta-analysis across Stage 1 cohorts of African ancestry. The labeled variant is the top variant at the locus from the overall meta-analysis and the r^2 represents the African ancestry LD pattern relative to the top variant from the overall meta-analysis. Point size indicates the sample size for each variant.

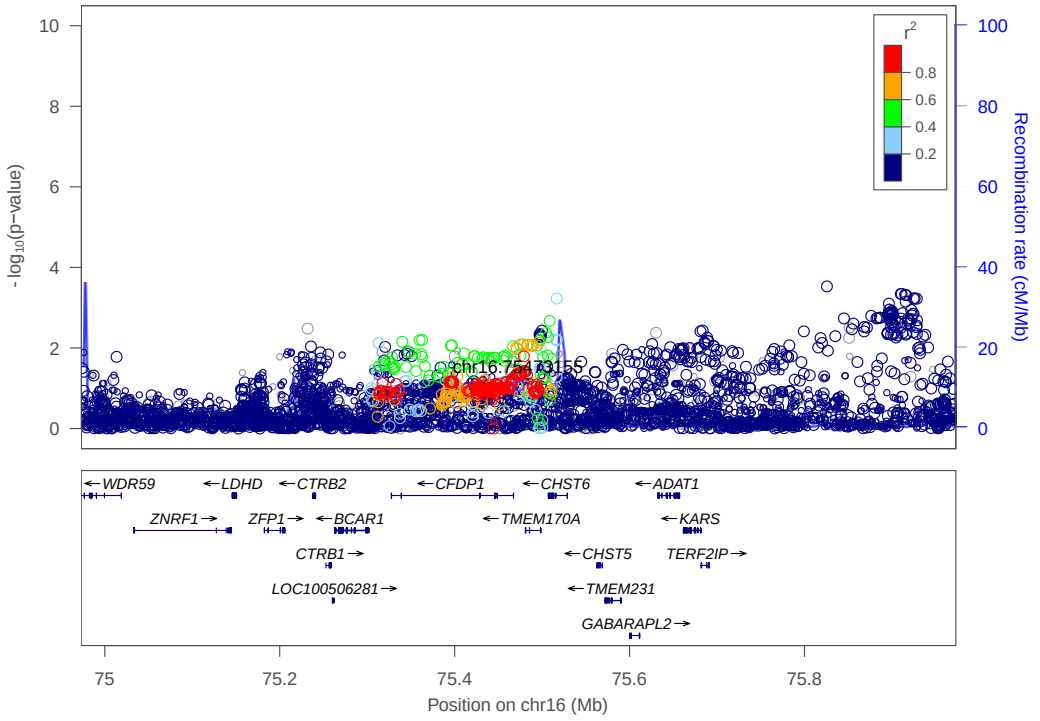


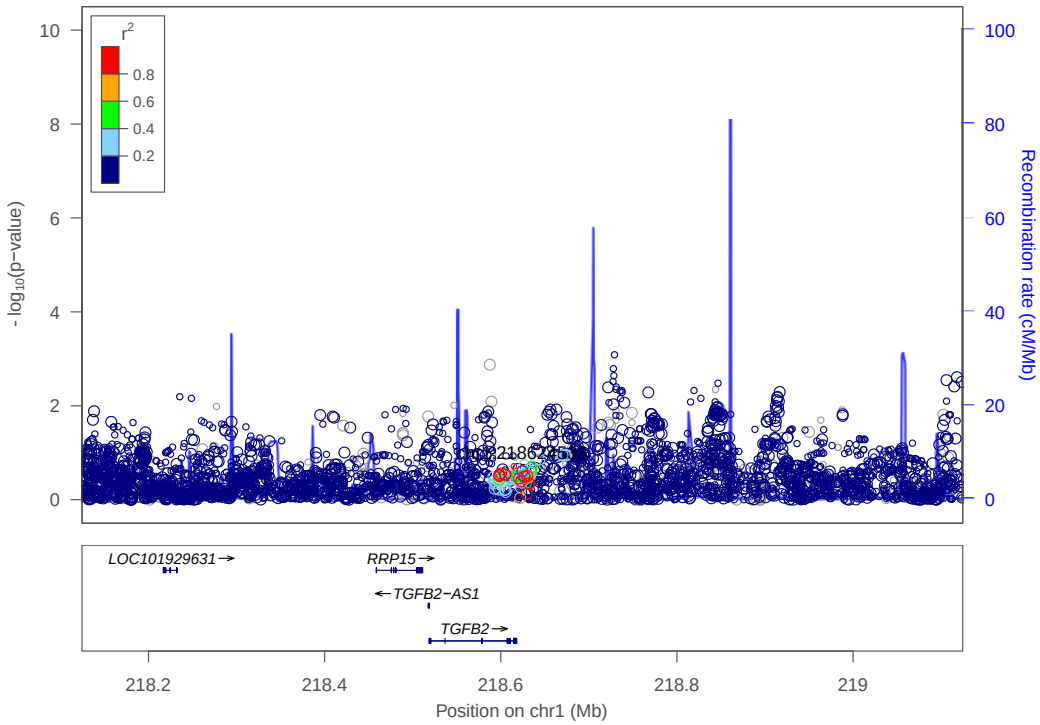

chr15:78867482
- log10(p-value)
Recombination rate (cM/Mb)
r2
0.8
0.6
0.4
0.2
TBC1D2B
DNAJA4
IREB2
CHRNA5
LOC646938
MORF4L1
RASGRF1
MIR5003
WDR61
HYKK
CHRNB4
ADAMTS7
CTSH
SH2D7
CRABP1
PSMA4
CIB2
CHRNA3
IDH3A
ACSBG1
Position on chr15 (Mb)

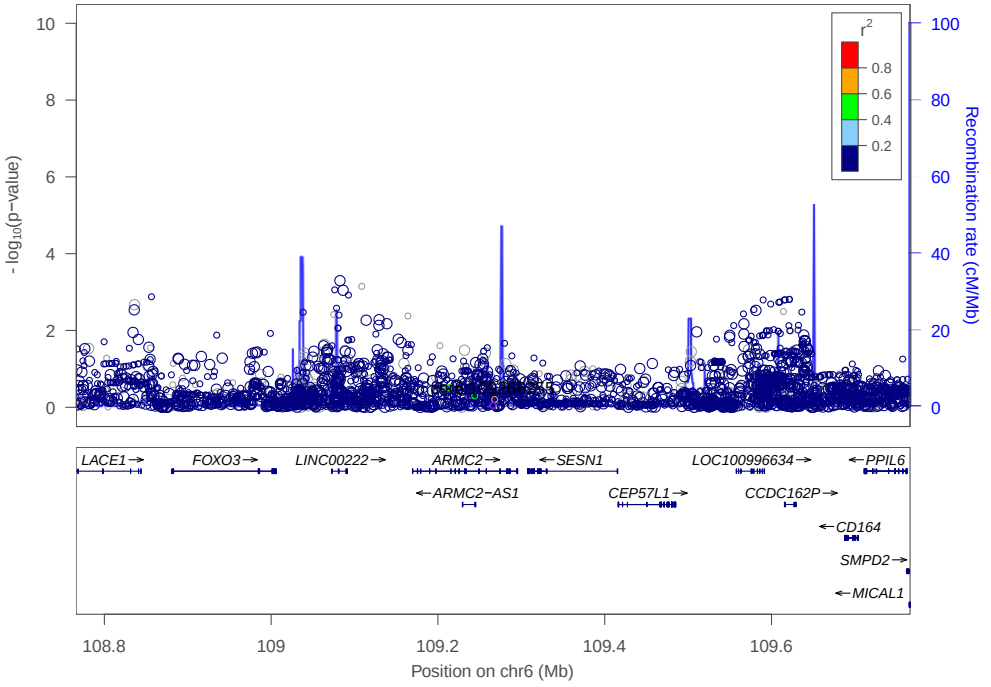

chr15:71612514
- log10(p-value)
Recombination rate (cM/Mb)
r2
0.8
0.6
0.4
0.2
LARP6
CT62
LOC101929173
LRRC49
THSD4
THAP10
LOC101929196
NR2E3
Position on chr15 (Mb)

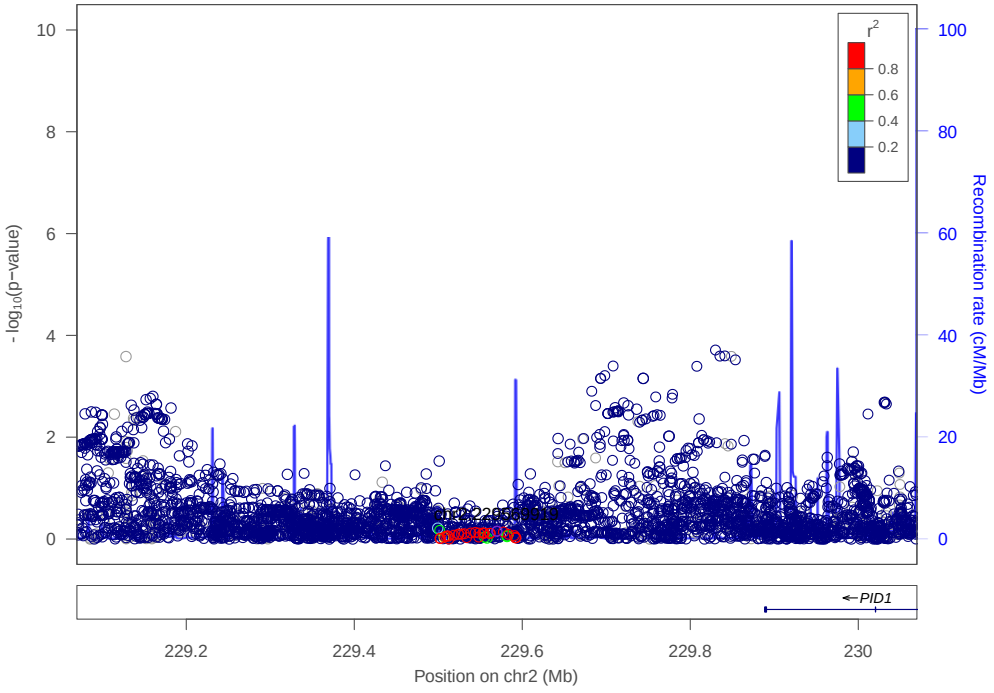

r²
0.8
0.6
0.4
0.2
- log10(p−value)
Recombination rate (cM/Mb)
chr4:89873092
HERC5
HERC3
FAM13A
TIGD2
GPRIN3
PIGY
NAP1L5
PYURF
FAM13A−AS1
89.4
89.6
89.8
90
90.2
Position on chr4 (Mb)

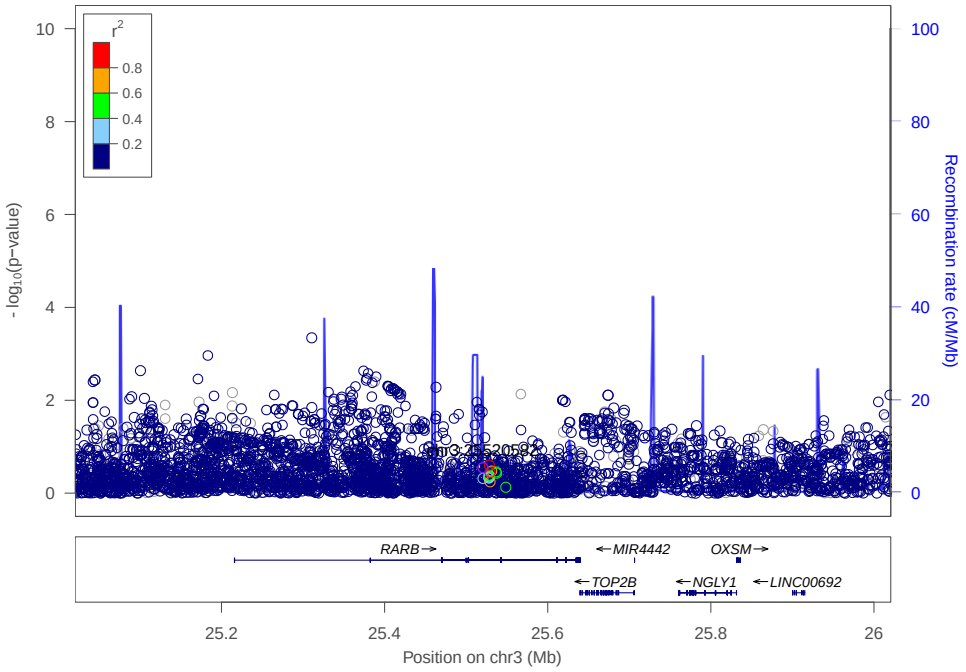

r²
0.8
0.6
0.4
0.2
- log10(p−value)
Recombination rate (cM/Mb)
HAVCR1
FAM71B
CYFIP2
NIPAL4
SOX30
THG1L
HAVCR2
ITK
FNDC9
ADAM19
C5orf52
CLINT1
MED7
LSM11
156.6
156.8
157
157.2
157.4
Position on chr5 (Mb)

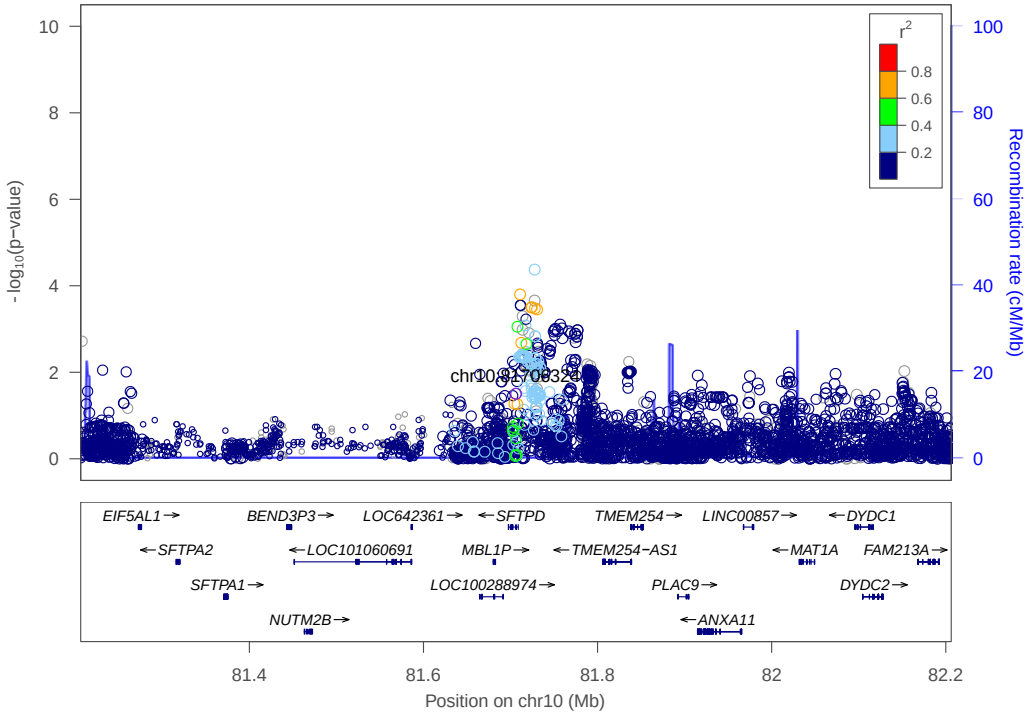

r²
0.8
0.6
0.4
0.2
- log10(p-value)
Recombination rate (cM/Mb)
WDR59
LDHD
CTRB2
CFDP1
CHST6
ADAT1
ZNRF1
ZFP1
BCAR1
TMEM170A
KARS
CTRB1
CHST5
TERF2IP
LOC100506281
TMEM231
GABARAPL2
Position on chr16 (Mb)


r²
0.8
0.6
0.4
0.2
- log10(p-value)
Recombination rate (cM/Mb)
LOC101929631
RRP15
TGFB2-AS1
TGFB2
Position on chr1 (Mb)


- log10(p−value)
Recombination rate (cM/Mb)
r2
0.8
0.6
0.4
0.2
LACE1
FOXO3
LINC00222
ARMC2
SESN1
LOC100996634
PPIL6
ARMC2−AS1
CEP57L1
CCDC162P
CD164
SMPD2
MICAL1
Position on chr6 (Mb)


- log10(p−value)
Recombination rate (cM/Mb)
r2
0.8
0.6
0.4
0.2
PID1
Position on chr2 (Mb)


r2
0.8
0.6
0.4
0.2
- log10(p-value)
Recombination rate (cM/Mb)
chr3:25520582
RARB
MIR4442
OXSM
TOP2B
NGLY1
LINC00692
Position on chr3 (Mb)


r2
0.8
0.6
0.4
0.2
- log10(p-value)
Recombination rate (cM/Mb)
chr10:81708324
EIF5AL1
BEND3P3
LOC642361
SFTPD
TMEM254
LINC00857
DYDC1
SFTPA2
LOC101060691
MBL1P
TMEM254-AS1
MAT1A
FAM213A
SFTPA1
LOC100288974
PLAC9
DYDC2
NUTM2B
ANXA11
Position on chr10 (Mb)

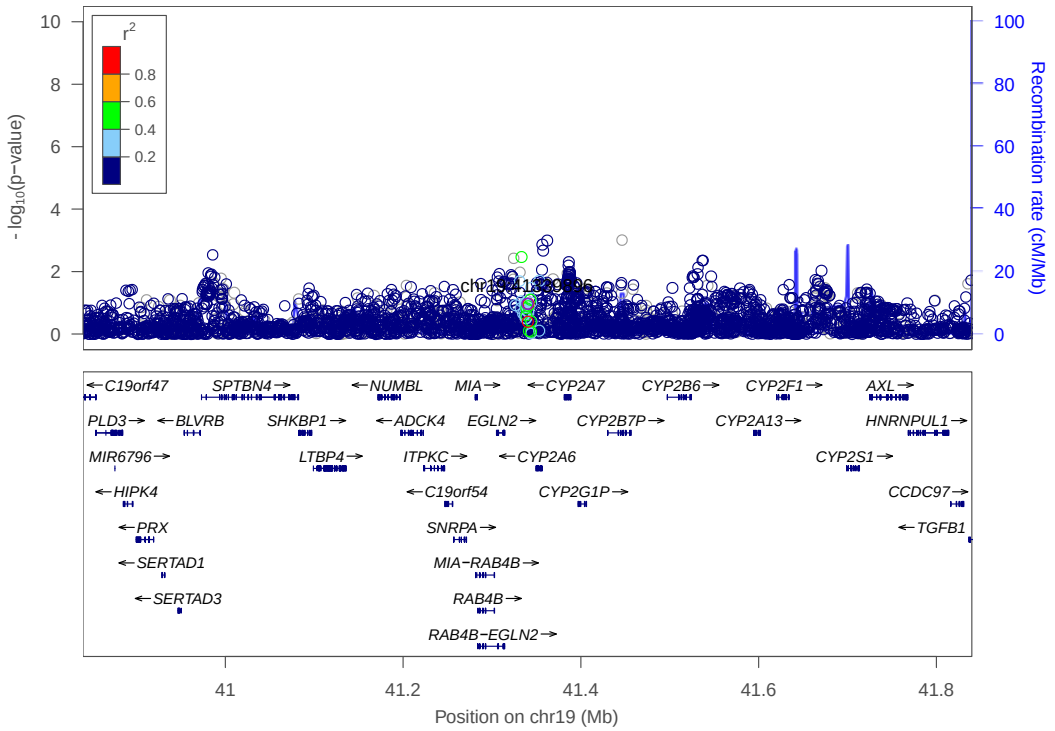

r2
0.8
0.6
0.4
0.2
- log10(p−value)
Recombination rate (cM/Mb)
chr19:41329896
C19orf47
SPTBN4
NUMBL
MIA
CYP2A7
CYP2B6
CYP2F1
AXL
PLD3
BLVRB
SHKBP1
ADCK4
EGLN2
CYP2B7P
CYP2A13
HNRNPUL1
MIR6796
LTBP4
ITPKC
CYP2A6
CYP2S1
HIPK4
C19orf54
CYP2G1P
CCDC97
PRX
SNRPA
TGFB1
SERTAD1
MIA−RAB4B
SERTAD3
RAB4B
RAB4B−EGLN2
Position on chr19 (Mb)

Figure SXXX: Comparison of effect size (beta) in the sensitivity analysis evaluating pre- and post-bronchodilator definitions of COPD for each of the 22 genome-wide significant loci.

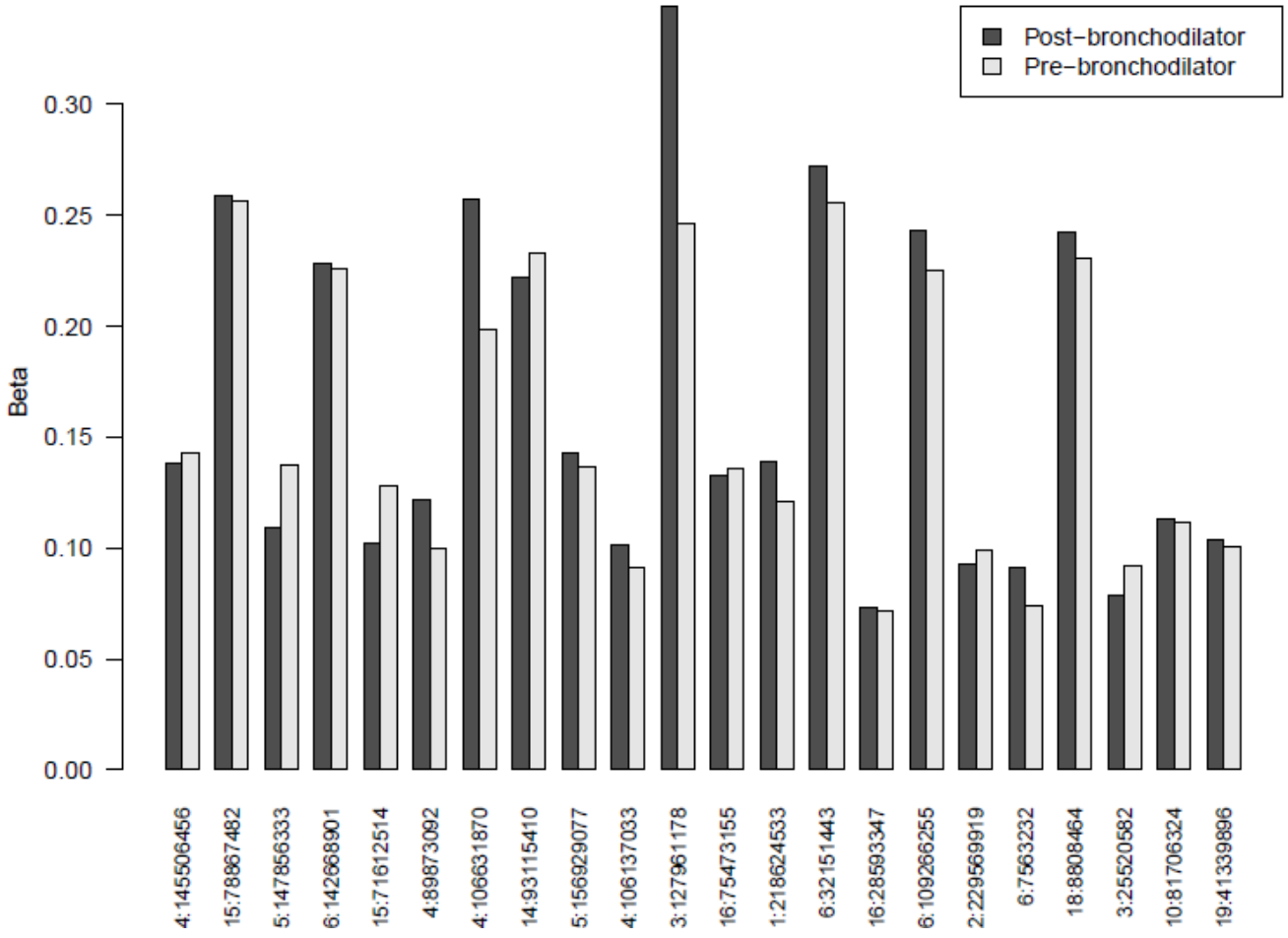
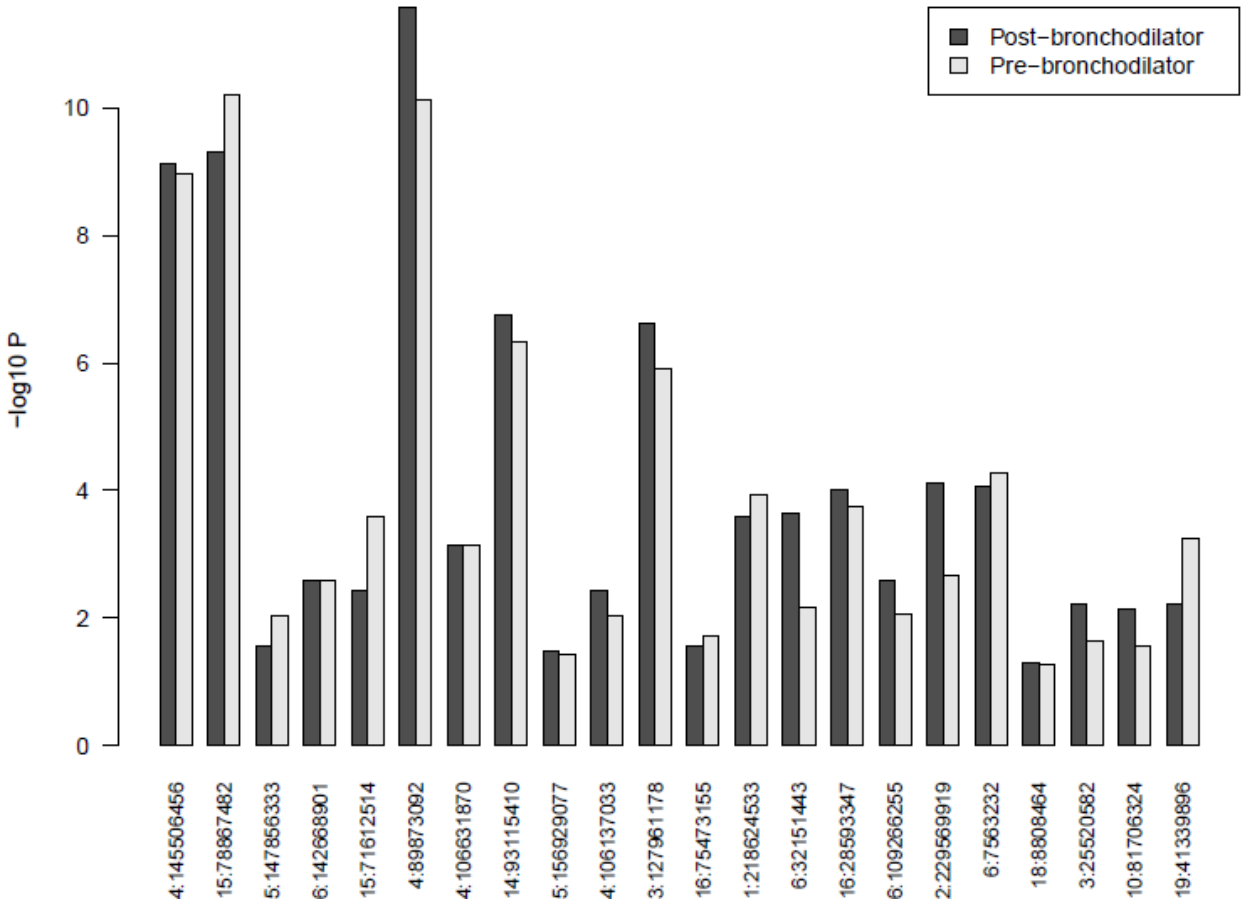


Figure SXXX: Comparison of -log10P values in the sensitivity analysis evaluating pre- and post-bronchodilator definitions of COPD for each of the 22 genome-wide significant loci.



Supplemental Tables:

Supplemental Table S1. Cohort baseline characteristics in COPD cases and controls.

Cohort	N		Age		% Female		% Ever smokers		Pack-years		FEV_1 % predicted	
	Cases	Controls	Cases	Controls	Cases	Controls	Cases	Controls	Cases	Controls	Cases	Controls
ARIC	1060	6164	56.8 (5.4)	54.2 (5.6)	43.5	55.9	87.6	51.9	37.7 (25.7)	11.9 (17.9)	64.7 (12.5)	99.9 (11)
B58	205	3665	45 (0.3)	45 (0.3)	53.7	51.2	75	60.4	16.2 (14.2)	8.6 (11.4)	68.1 (9.5)	99.1 (11)
CHS EA	736	1586	73.1 (5.6)	71.7 (5)	44.6	71.4	72.7	40.2	31.4 (31.2)	10 (18.8)	59.8 (16.1)	100 (13.6)
COPACETIC	397	1906	61.9 (6.1)	59.7 (5.4)	2	1.6	100	100	45.6 (19.6)	38.9 (17.1)	65.3 (11.8)	106.9 (12.8)
COPDGene NHW	3068	2110	64.4 (8.3)	59.2 (8.6)	45.9	49.8	100	100	54.9 (27.1)	37.3 (20.1)	48.4 (18.5)	95.7 (10.7)
ECLIPSE	1741	149	63.6 (7.1)	57.3 (9.5)	33.1	42.8	100	100	50.4 (27.4)	30.9 (25.8)	43.3 (14.8)	105.2 (12.1)
EOCOPD *	394	495	53.6 (12.5)	40.5 (17.8)	60.2	57.2	88.3	63.3	39.8 (25.3)	10.4 (18.3)	38.4 (22.1)	97.2 (10.8)
ICGN *	1852	557	59 (7.1)	54.4 (8.9)	42.3	49.5	100	100	50.7 (28.2)	28.8 (20)	42.6 (18.1)	101.8 (13.9)
EQTL	252	224	64.6 (8.8)	61.5 (10.3)	38.9	50	98.4	86.2	52.6 (29.5)	38.6 (23.4)	61.1 (15.3)	97.7 (11.5)
FHS	701	5110	62.2 (11.8)	51.9 (13.3)	47.8	55	80.3	48.4	31.5 (27.6)	7.4 (14)	65.6 (11.9)	102.1 (11.9)
LifeLines	466	9863	53.9 (11.8)	47 (10.7)	48.9	59.9	77.9	54.4	17.2 (16.6)	6.4 (9.5)	70 (9.1)	105.1 (12)
Lovelace	259	641	64.9 (8.1)	54.4 (9.1)	78.7	81.3	100	100	52.7 (24.3)	36.6 (18.4)	61.5 (14.5)	100.1 (11.8)
MESA Caucasian	167	754	68.4 (8.9)	64.5 (9.7)	48.5	53.3	79.6	53.1	33.9 (34.2)	12.3 (20.4)	66 (12.4)	99 (11.7)
NETT-NAS	376	435	67.5 (5.8)	69.8 (7.5)	35.9	0	100	100	66.4 (30.7)	40.7 (27.8)	24.9 (6.6)	100 (13.2)
Norway/GenKOLS	846	695	65.5 (10.1)	55.4 (9.7)	41.7	50.2	100	100	31.9 (18.6)	19.4 (13.6)	46.4 (17)	92.6 (8.9)
RS1	112	815	79.5 (5)	78.5 (4.5)	38	60.5	83.9	61.8	28.6 (24.3)	12.3 (18)	64.5 (11.7)	113.4 (18)
RS2	94	811	73 (6)	71.6 (4.9)	44	55.6	90.4	61.3	33.4 (26.2)	12.1 (18.9)	64.2 (11.8)	108.6 (15.7)
RS3	106	1596	63 (6)	61.6 (5.4)	49	58.2	90.6	64	37.4 (28.2)	11.1 (16.4)	64.6 (12.7)	108.8 (14.3)
SPIROMICS	571	175	66.6 (7.6)	64.1 (9.2)	44.1	54	100	100	55.6 (24.2)	43.8 (23.8)	48.4 (18)	95.8 (10.8)
TCGS-Poland *	307	311	62.6 (7.4)	59.1 (7.2)	31.5	32.6	100	100	44.6 (22.5)	34.1 (15.1)	29.3 (9.6)	102.9 (12.6)
CHS AA	138	292	72.8 (5.2)	73 (5.3)	53.6	69.9	70.3	48.6	20.6 (25.1)	9.7 (17.1)	60.6 (14.2)	102.5 (14.6)
COPDGene AA	910	1556	58.6 (8.1)	52.8 (6)	46.8	43.8	100	100	42.7 (23.5)	36.1 (19.1)	51.2 (18.1)	97.2 (11.8)
KARE	199	6741	60.8 (8.4)	54.8 (8.6)	18	56	72.4	34.4	34.7 (20.1)	25.2 (17.9)	68 (9.5)	113.5 (15.2)
MESA AA	94	532	67.6 (9)	64.4 (9.4)	33	58.3	78.7	49.4	27.7 (26.7)	9.8 (16.2)	64.1 (13.9)	100.2 (14.5)
MESA Hispanic	52	548	68.4 (9.3)	63.3 (9.4)	32.7	55.8	71.2	41.8	25.4 (31.8)	6.9 (14.4)	65.1 (13.6)	100.4 (12.4)
TCGS-Korea *	153	205	68.5 (6.8)	52.8 (8.4)	1.6	3.4	100	100	44.3 (24.5)	27.6 (15)	34.8 (10.2)	95.5 (9.2)
UK BiLEVE Never Smokers	3737	4871	58.3 (7.7)	56.9 (7.9)	47.6	49.9	0	0	0	0	65.4 (11.4)	130.3 (8.3)
UK BiLEVE Heavy Smokers	5761	4877	59.3 (6.9)	56.9 (7.9)	45.5	49.8	100	100	41.6 (20.9)	29.8 (13.9)	61.2 (11.8)	118 (8.1)

Shaded studies have participants of European ancestry. All values represented as N, %, or mean (sd). ARIC = Atherosclerosis Risk in Communities, B58 = British 1958 Birth Cohort, CHS = Cardiovascular Health Study, COPACETIC = COPD Pathology: Addressing Critical gaps, Early Treatment & Diagnosis and Innovative Concepts, ECLIPSE = Evaluation of COPD Longitudinally to Identify Predictive Surrogate End-points, eQTL = Lung Expression Quantitative Trait Loci Study, FHS = Framingham Heart Study, KARE = Korean Association Resource project, MESA = Multi-Ethnic Study of Atherosclerosis, NETT-NAS = National Emphysema Treatment Trial / Normative Aging Study, RS = Rotterdam Study, SPIROMICS = Subpopulations and intermediate outcome measures in COPD study , EOCOPD = Boston Early-Onset COPD Study, ICGN = International COPD Genetics Network, TCGS = Transcontinental COPD Genetics Study, UK BiLEVE = UK Biobank Lung Exome Variant Evaluation; NHW = Non-Hispanic white, AA = African American, EA = European American. * Indicates the four studies with custom genotyping content (i.e. they did not have genome-wide data)

Supplemental Table S2. Ancestry and genomic control factor for each cohort in the Stage 1 meta-analysis.

Cohort	Ancestry	Effective Sample Size	Genomic Control Factor
ARIC	European	3617.9	1.026
B58	European	776.6	1.009
CHS EA	European	2010.8	1.016
COPACETIC	European	1314.3	1.013
COPDGene NHW	European	5000.8	1.017
ECLIPSE	European	549	1
EOCOPD, ICGN *	European	2590.4	NA
EQTL	European	474.4	1.003
FHS	European	2465.7	1.035
LifeLines	European	1779.9	1.019
Lovelace	European	737.9	1.032
MESA Caucasian	European	546.9	1.03
NETT-NAS	European	806.7	1.038
Norway/GenKOLS	European	1526.2	1.009
RS1	European	393.9	1.025
RS2	European	336.9	1.036
RS3	European	397.6	1.012
SPIROMICS	European	535.8	1.017
TCGS-Poland *	European	618	NA
CHS AA	African American	374.8	1.044
COPDGene AA	African American	2296.8	1.009
KARE	Korean	773.2	1
MESA AA	African American	319.5	1.026
MESA Hispanic	Hispanic	190	1.006
TCGS-Korea *	Korean	350.4	NA

* These studies did not have genome-wide data available for analysis.

Supplemental Table S3. Full results table for all 79 variants submitted for testing in UK BiLEVE Stage 2 analysis. The top 22 variants with overall meta-analysis P value < $5x10^{-8}$ are shaded. Supplemental Table S3 is attached in Excel format "icgcSuppTableS3_fullStage2Testing.xlsx".

Supplemental Table S4. Association of top 22 loci from meta-analysis of Stage 1 cohorts of European ancestry.

rsID	Closest Gene	Locus	Risk Allele	Alt Allele	Average Risk Allele Freq (Range)	Stage 1 OR (95% CI)	Stage 1 P Value
rs13141641	*HHIP*	4q31.21	T	C	0.594 (0.524-0.627)	1.229 (1.178-1.281)	4.05E-22
rs17486278	*CHRNA5*	15q25.1	C	A	0.358 (0.305-0.442)	1.231 (1.181-1.284)	3.04E-22
rs7733088	*HTR4*	5q32	G	A	0.602 (0.575-0.633)	1.193 (1.139-1.25)	8.30E-14
rs9399401	*ADGRG6*	6q24.1	T	C	0.724 (0.707-0.748)	1.124 (1.073-1.177)	6.13E-07
rs1441358	*THSD4*	15q23	G	T	0.333 (0.316-0.349)	1.106 (1.06-1.153)	3.30E-06
rs6837671	*FAM13A*	4q22.1	G	A	0.406 (0.364-0.454)	1.156 (1.11-1.204)	2.19E-12
rs11727735	*GSTCD*	4q24	A	G	0.936 (0.926-0.949)	1.271 (1.169-1.383)	2.14E-08
rs754388	*RIN3*	14q32.12	C	G	0.824 (0.812-0.851)	1.213 (1.148-1.281)	5.39E-12
rs113897301	*ADAM19*	5q33.3	AT	A	0.175 (0.15-0.187)	1.213 (1.154-1.271)	1.09E-10
rs2047409	*TET2*	4q24	A	G	0.617 (0.585-0.649)	1.112 (1.065-1.162)	1.45E-06
rs2955083	*EEFSEC*	3q21.3	A	T	0.881 (0.854-0.893)	1.197 (1.119-1.28)	1.58E-07
rs7186831	*CFDP1*	16q23.1	A	G	0.434 (0.417-0.452)	1.122 (1.063-1.185)	2.82E-05
rs10429950	*TGFB2*	1q41	T	C	0.733 (0.698-0.773)	1.138 (1.087-1.191)	4.14E-08
rs2070600	*AGER*	6p21.32	C	T	0.957 (0.945-0.987)	1.275 (1.149-1.415)	5.10E-06
rs17707300	*CCDC101*	16p11.2	C	T	0.374 (0.338-0.433)	1.119 (1.073-1.167)	1.19E-07
rs2806356	*ARMC2*	6q21	C	T	0.185 (0.158-0.198)	1.122 (1.065-1.181)	1.17E-05
rs16825267	*PID1*	2q36.3	C	G	0.929 (0.92-0.942)	1.294 (1.186-1.412)	7.19E-09
rs2076295	*DSP*	6p24.3	T	G	0.543 (0.442-0.581)	1.121 (1.077-1.166)	1.78E-08
rs647097	*MTCL1*	18p11.22	C	T	0.269 (0.259-0.297)	1.096 (1.046-1.148)	1.23E-04
rs1529672	*RARB*	3p24.2	C	A	0.834 (0.821-0.862)	1.181 (1.119-1.247)	1.80E-09
rs721917	*SFTPD*	10q22.3	G	A	0.425 (0.403-0.458)	1.083 (1.041-1.127)	8.85E-05
rs12459249	*CYP2A6*	19q13.2	C	T	0.664 (0.617-0.702)	1.124 (1.064-1.186)	2.81E-05

Supplemental Table S5. Association of top 22 loci from meta-analysis across the three Stage 1 cohorts of African ancestry.

rsID	Closest Gene	Locus	Risk Allele	Alt Allele	Average Risk Allele Freq	Stage 1 OR (95% CI)	Stage 1 P Value
rs13141641	*HHIP*	4q31.21	T	C	0.88	1.22 (1.01 - 1.47)	0.042
rs17486278	*CHRNA5*	15q25.1	C	A	0.29	1.22 (1.07 - 1.39)	2.2E-03
rs7733088	*HTR4*	5q32	G	A	0.67	1.03 (0.916 - 1.17)	0.59
rs9399401	*ADGRG6*	6q24.1	T	C	0.62	1.19 (1.05 - 1.34)	5.3E-03
rs1441358	*THSD4*	15q23	G	T	0.53	1.26 (1.13 - 1.42)	5.1E-05
rs6837671	*FAM13A*	4q22.1	G	A	0.58	1.11 (0.989 - 1.25)	0.075
rs11727735	*GSTCD*	4q24	A	G	---	---	---
rs754388	*RIN3*	14q32.12	C	G	0.86	1.2 (1.01 - 1.42)	0.036
rs113897301	*ADAM19*	5q33.3	AT	A	0.05	0.852 (0.645 - 1.13)	0.26
rs2047409	*TET2*	4q24	A	G	0.32	1.07 (0.944 - 1.2)	0.30
rs2955083	*EEFSEC*	3q21.3	A	T	0.88	1.15 (0.955 - 1.38)	0.14
rs7186831	*CFDP1*	16q23.1	A	G	0.23	1.15 (0.981 - 1.34)	0.086
rs10429950	*TGFB2*	1q41	T	C	0.64	1.07 (0.953 - 1.21)	0.25
rs2070600	*AGER*	6p21.32	C	T	---	---	---
rs17707300	*CCDC101*	16p11.2	C	T	0.16	1.1 (0.952 - 1.28)	0.19
rs2806356	*ARMC2*	6q21	C	T	0.053	0.981 (0.742 - 1.3)	0.89
rs16825267	*PID1*	2q36.3	C	G	0.93	1.05 (0.826 - 1.33)	0.70
rs2076295	*DSP*	6p24.3	T	G	0.49	0.984 (0.879 - 1.1)	0.78
rs647097	*MTCL1*	18p11.22	C	T	0.40	1.13 (1.01 - 1.27)	0.039
rs1529672	*RARB*	3p24.2	C	A	0.80	1.08 (0.938 - 1.24)	0.28
rs721917	*SFTPD*	10q22.3	G	A	0.41	1.13 (1.01 - 1.27)	0.031
rs12459249	*CYP2A6*	19q13.2	C	T	0.67	1.12 (0.976 - 1.29)	0.11

Supplemental Table S6. Summary of prior evidence for genome-wide association of our top 22 loci with either COPD or lung function. Novel loci have been highlighted with bold text.

ICGC Lead Variant	Closest Gene	Locus	Prior genome-wide COPD locus?	References for COPD association	Prior COPD SNP (LD with ICGC variant)	Consistent direction of effect?	Prior genome-wide lung function locus?	References for lung function association	Prior lung function SNP (LD with ICGC variant)	Consistent direction of effect?
rs13141641	*HHIP*	4q31.21	✓	Pillai et al. (2009), Cho et al. (2014)	rs1828591 (0.98), rs13141641 (same)	YES YES	✓	Wilk et al. (2009)	rs13147758 (0.90)	YES
rs17486278	*CHRNA5*	15q25.1	✓	Pillai et al. (2009), Cho et al. (2014)	rs8034191 (0.90), rs12914385 (0.77)	YES YES	☐			
rs7733088	*HTR4*	5q32	✓	Wilk et al. (2012)	rs7733088 (same)	YES	✓	Hancock et al. (2010), Repapi et al. (2010)	rs11168048 (0.90), rs3995090 (0.98)	YES YES
rs9399401	*ADGRG6*	6q24.1	☐				✓	Hancock et al. (2010)	rs3817928 (0.62)	YES
rs1441358	*THSD4*	15q23	☐				✓	Repapi et al. (2010)	rs12899618 (0.29)	YES
rs6837671	*FAM13A*	4q22.1	✓	Cho et al. (2010)	rs7671167 (0.64)	YES	✓	Hancock et al. (2010)	rs2869967 (1)	YES
rs11727735	*GSTCD*	4q24	✓	Wain et al. (2015)	rs10516528 (1)	YES	✓	Hancock et al. (2010), Repapi et al. (2010)	rs11097901 (1), rs10516526 (1)	YES YES
rs754388	*RIN3*	14q32.12	✓	Cho et al. (2014)	rs754388 (same)	YES	✓	Soler Artigas et al. (2015)	rs117068593 (0.96)	YES
rs113897301	*ADAM19*	5q33.3	☐				✓	Hancock et al. (2010)	rs2277027 (0.37)**	YES
rs2047409	*TET2*	4q24	☐				✓	Wain et al. (2015)	rs2047409 (same)	YES
rs2955083	*EEFSEC*	3q21.3	☐				☐			
rs7186831	*CFDP1*	16q23.1	☐				✓	Soler Artigas et al. (2011)	rs2865531 (0.96)	YES
rs10429950	*TGFB2*	1q41	✓	Cho et al. (2014)	rs4846480 (1)	YES	✓	Soler Artigas et al. (2011)	rs993925 (0.009)	NA (no LD)
rs2070600	*AGER*	6p21.32	☐				✓	Hancock et al. (2010), Repapi et al. (2010)	rs2070600 (same)	YES
rs17707300	*CCDC101*	16p11.2	✓*	Hobbs et al. (2016)	rs181206 (0.74)	YES	☐			
rs2806356	*ARMC2*	6q21	☐				✓	Soler Artigas et al. (2011)	rs2798641 (1)	YES
rs16825267	*PID1*	2q36.3	☐				✓	Hancock et al. (2010)	rs1435867 (1)	YES
rs2076295	*DSP*	6p24.3	☐				☐			
rs647097	*MTCL1*	18p11.22	☐				☐			
rs1529672	*RARB*	3p24.2	☐				✓	Soler Artigas et al. (2011)	rs1529672 (same)	YES
rs721917	*SFTPD*	10q22.3	☐				☐			
rs12459249	*CYP2A6*	19q13.2	✓	Cho et al. (2012)	rs7937 (0.24)	YES	☐			

All LD was calculated from 1000 genomes Phase 3 CEU unless otherwise noted. * locus was exome-wide significant.
** LD was calculated from COPDGene non-Hispanic whites..

Supplemental Table S7. Details of the four novel COPD association loci.

Lead Novel SNP Chr:position (hg19) Cytogenetic band	Lead SNP and SNPs with r^2 > 0.3 (1000 Genomes CEU) gene annotations [2]	Details of genes annotated to lead SNPs and SNPs with r^2 > 0.3 with lead SNP	Regulatory information via lead SNP look-up in HaploReg v4.1 [3,4]	NHGRI-EBI GWAS catalog query for GWAS loci (P < $5x10^{-8}$) within ± 500 kb of lead SNP and with r^2 > 0.3 with lead SNP
rs2955083 3:127961178 3q21.3	**rs2955083** (lead), ***EEFSEC*** intronic. rs11714256 (r^2=0.37), ***RUVBL1*** intronic. rs1042912 (r^2=0.37), ***SEC61A1***, 3' UTR. rs17203687 (r^2=0.31), ***KBTBD12*** intronic.	***EEFSEC*** codes for selenocysteine-specific elongation factor, is annotated to GTP and tRNA binding, and is a paralog to *TUFM* [5]. ***RUVBL1*** codes for RuvB-like 1 protein, which participates in ATPase and DNA helicase activity [5]. ***SEC61A1*** codes for protein transport protein Sec61 subunit alpha isoform 1, which is associated with membrane-bound ribosomes and is important to protein translocation [6]. ***KBTBD12*** codes for the Kelch repeat and BTB domain-containing protein 12, which is involved in protein ubiquitination [5].	**rs2955083** near *EEFSEC* and variants in high LD (r^2 > 0.8, 1000 genomes CEU) possess DNase hypersensitivity sites as well as promoter and enhancer site histone marks across nearly all indexed tissue types including lung, lung fibroblasts, fetal lung, and fetal lung fibroblasts. Additionally, 25 different regulatory motifs are altered by **rs2955083**.	rs2687729 (r^2=1.0), menarche age at onset [7] rs2999052 (r^2=0.38), hypospadias [8] rs10934853 (r^2=0.33), prostate cancer [9]
rs2076295 6:7563232 6p24.3	**rs2076295** (lead), ***DSP*** intronic. (No SNPs with r^2 > 0.3 in any additional genes)	***DSP*** codes for desmoplakin, a major protein of desmosomes [5], which is required for epidermal integrity [10]. ***DSP*** mutations have been linked to several Mendelian syndromes involving palmoplantar keratoderma [11], left ventricular cardiomyopathy [12], familial arrhythmogenic right ventricular dysplasia [13], and lethal ancantholytic epidermolysis bullosa [14]. Further, our lead SNP in ***DSP*** has been previously associated with fibrotic idiopathic interstitial pneumonias in GWAS [15]; however, the minor allele for **rs2076295** has opposite direction of effect for pulmonary fibrosis and COPD risk..	**rs2076295** reported to have DNase hypersensitivity sites and enhancer histone marks in blood and several solid organs; however, none in lung tissues	**rs2076295** (lead), interstitial lung disease [15,16]
rs647097 18:8808464 18p11.22	**rs647097** (lead), ***MTCL1*** intronic. (No SNPs with r^2 > 0.3 in any additional genes)	***MTCL1*** codes for the microtubule cross-linking factor 1, important in epithelial-cell-specific microtubule stabilization and accumulation around the Golgi apparatus [17,18].	**rs647097** and variants in high LD (r^2 > 0.8 in 1000 genomes phase 1 CEU) have enhancer histone marks in fetal lung fibroblasts. Several regulatory motifs are altered by **rs647097**	None.
rs721917 10:81706324 10q22.3	**rs721917** (lead), ***SFTPD*** non-synonymous (missense). (No SNPs with r^2 > 0.3 in any additional genes)	***SFTPD*** codes for the pulmonary surfactant-associated protein D, which functions in lung innate immunity [5]. *Sftpd* (-/-) mice have been observed to have chronic inflammation, emphysema, and fibrosis [19]. ***SFTPD*** variants have been examined in a COPD candidate gene study [20] and SFTPD levels have been research as a COPD biomarker [21].	**rs721917** has enhancer histone marks in fetal lung tissue	rs1923539 (r^2=0.38), circulating surfactant protein D levels in COPD [22]

Supplemental Table S8. Full results ($P < 0.05$ in meta-analysis) for lung expression quantitative trait locus (eQTL) analysis. All cis- and trans- eQTL with $P < 0.05$ in meta-analysis are shown. eQTL associations meeting Bonferroni correction of $P < 5.2 \times 10^{-8}$ are highlighted. * refers to proxy variant. Supplemental Table S8 is attached in Excel format "icgcSuppTableS8_fullEqtlResults.xlsx".

Supplemental Table 9a-b: Co-localization results. Probe = affymetrix probe ID, tile center given as chromosome position for hg19. ppH4 = posterior probability of co-localization of the GWAS and eQTL association. Results with posterior probability of co-localization > 0.6 are shown.

Supplemental Table 9a: Co-localization using the Lung eQTL data[23].

Probe	Gene	Tile center	ppH4
100149674_TGI_at	*RARB*	3:25520582	0.85
100139086_TGI_at	*NA*	4:145506456	0.95
100311674_TGI_at	*HHIP*	4:145506456	0.95
100133899_TGI_at	*NA*	4:145506456	0.94
100148028_TGI_at	*HHIP*	4:145506456	0.94
100312038_TGI_at	*FAM13A*	4:89873092	0.77
100313510_TGI_at	*AGER*	6:32151443	0.93
100131291_TGI_at	*NA*	6:7563232	0.99
100304914_TGI_at	*DSP*	6:7563232	0.99
100313405_TGI_at	*DSP*	6:7563232	0.99
100125626_TGI_at	*TGFB2*	1:218624533	0.62
100142979_TGI_at	*THSD4*	15:71612514	0.99
100308491_TGI_at	*THSD4*	15:71612514	0.99
100310339_TGI_at	*THSD4*	15:71612514	0.99
100130228_TGI_at	*NA*	15:71612514	0.68
100123649_TGI_at	*CHRNA3*	15:78867482	0.89
100129347_TGI_at	*SBK1*	16:28593347	0.79
100142003_TGI_at	*SPNS1*	16:28593347	0.67
100300284_TGI_at	*MTCL1*	18:8808464	0.76

Supplemental Table 9b: Co-localization using lung tissue from GTEx[24].

ensembleID	Gene	Tile center	ppH4
ENSG00000270720.1	*RP11-84C13.2*	4:89373092	0.75
ENSG00000135074.11	*ADAM19*	5:156429077	0.79
ENSG00000248544.2	*CTB-47B11.3*	5:156429077	0.67
ENSG00000176476.4	*SGF29*	16:28093347	0.89
ENSG00000178188.10	*SH2B1*	16:28093347	0.79
ENSG00000196502.7	*SULT1A1*	16:28093347	0.75
ENSG00000264455.1	*MIR4721*	16:28093347	0.7
ENSG00000168502.13	*MTCL1*	18:8308464	0.92

Supplemental Table S10. Sensitivity analysis results for COPD association of top 22 loci using Post- versus Pre-Bronchodilator Definitions of COPD in the COPDGene NHW and AA, ECLIPSE, NETT-NAS, and Norway/GenKOLS cohorts**.**

Location	rsID	Beta		P	
		Post-	Pre-	Post-	Pre-
4:145506456	rs13141641	0.23	0.23	7.80E-10	1.10E-09
15:78867482	rs17486278	0.22	0.23	5.10E-10	6.40E-11
5:147856333	rs7733088	0.08	0.09	0.027	0.0094
6:142668901	rs9399401	0.11	0.11	0.0026	0.0026
15:71612514	rs1441358	0.1	0.13	0.0037	0.00026
4:89873092	rs6837671	0.24	0.22	2.70E-12	7.40E-11
4:106631870	rs11727735	0.26	0.26	7.00E-04	0.00074
14:93115410	rs754388	0.24	0.23	1.80E-07	4.60E-07
5:156929077	rs113897301	0.1	0.1	0.032	0.037
4:106137033	rs2047409	0.1	0.09	0.0037	0.0091
3:127961178	rs2955083	0.27	0.26	2.40E-07	1.20E-06
16:75473155	rs7186831	0.09	0.1	0.027	0.019
1:218624533	rs10429950	0.14	0.14	0.00025	0.00012
6:32151443	rs2070600	0.34	0.25	0.00023	0.0068
16:28593347	rs17707300	0.14	0.14	1.00E-04	0.00018
6:109266255	rs2806356	0.14	0.12	0.0026	0.0087
2:229569919	rs16825267	0.26	0.2	7.50E-05	0.0022
6:7563232	rs2076295	0.13	0.14	8.40E-05	5.40E-05
18:8808464	rs647097	0.07	0.07	0.049	0.052
3:25520582	rs1529672	0.12	0.1	0.0059	0.023
10:81706324	rs721917	0.09	0.07	0.0072	0.028
19:41339896	rs12459249	0.11	0.14	0.0059	0.00055

Supplemental Table S11. Look-up of NHGRI-EBI GWAS Catalog asthma-associated trait genome-wide significant GWAS loci in our COPD association Stage 1 meta-analysis results. Supplemental Table S1 is attached in Excel format "icgcSuppTableS11_asthmaGwasCopdAssoc.xlsx". Only asthma trait GWAS loci that were present in our Stage 1 analysis are reported.

Supplemental Table S12. Association of top 22 COPD association loci from overall meta-analysis in the GABRIEL Consortium asthma GWAS[25]

Chr	SNP*	Asthma OR (Random Effects)	Asthma P Value (Random Effects)	Proxy SNP Used?	COPD Index SNP	LD (r^2) with COPD Index SNP
1	rs10429950	1.01	0.61	NO	--	--
3	rs1529672	1.03	0.27	NO	--	--
3	rs2811518	0.93	0.026	YES	rs2955083	1
4	rs2869967	1.02	0.44	YES	rs6837671	0.97
4	rs1391441	0.99	0.69	YES	rs2047409	0.90
4	rs11727735	0.96	0.26	NO	--	--
4	rs6817273	1.02	0.36	YES	rs13141641	0.93
5	rs3995090	0.98	0.29	YES	rs7733088	0.97
5	rs3734031	1.03	0.48	YES	rs56168343	0.56
6	rs2076295	1.00	0.86	NO	--	--
6	rs2070600	1.01	0.92	NO	--	--
6	rs2798641	1.03	0.18	YES	rs2806356	1
6	rs6570507	1.03	0.16	YES	rs9399401	0.96
10	rs721917	0.98	0.38	NO	--	--
14	rs11627032	1.03	0.26	YES	rs754388	0.63
15	rs6494904	1.00	0.84	YES	rs1441358	0.75
15	rs1051730	0.99	0.64	YES	rs17486278	0.93
16	rs4788073	1.00	0.92	YES	rs17707300	1
18	rs647097	1.04	0.13	NO	--	--
19	rs7251418	0.99	0.86	YES	rs12459249	0.88

No proxy SNPs with r2 > 0.5 could be found for top COPD SNPs rs7186831 (*CFDP1*) and rs16825267 (*PID1*)

Supplemental Table S13. Look-up of NHGRI-EBI GWAS Catalog "Nicotine Dependence" and "Smoking Behaviors" genome-wide significant GWAS loci in our COPD association Stage 1 meta-analysis results**.**

Chr	Position	SNP	Region	NHGRI-EBI Catalog Mapped Genes	Effect Allele	Average Effect Allele Freq	OR (COPD)	P value (COPD)
1	99445471	rs61784651	1p21.1	*LPPR5*	T	0.15	0.98	0.55
2	146316319	rs10193706	2q22.3	*TEX41 - PABPC1P2*	A	0.50	0.96	0.032
6	38901867	rs10807199	6p21.2	*DNAH8*	T	0.45	1.00	0.95
8	42546711	rs1451240	8p11.21	*SMIM19 - LOC105379396*	A	0.29	0.96	0.086
8	42550498	rs6474412	8p11.21	*SMIM19 - LOC105379396*	T	0.72	1.05	0.045
9	136478355	rs3025343	9q34.2	*FAM163B - DBH*	A	0.11	1.03	0.43
10	93348120	rs1329650	10q23.32	*HECTD2-AS1*	T	0.27	1.02	0.33
11	27679916	rs6265	11p14.1	*BDNF - BDNF-AS*	T	0.20	1.03	0.24
11	112861434	rs4466874	11q23.2	*NCAM1*	T	0.59	0.98	0.40
15	78851615	rs2036527	15q25.1	*PSMA4 - CHRNA5*	A	0.34	1.20	2.2E-18
15	78894339	rs1051730	15q25.1	*CHRNA3*	A	0.34	1.21	2.8E-22
19	41310571	rs3733829	19q13.2	*EGLN2 - RAB4B-EGLN2*	A	0.65	0.96	0.030
19	41363765	rs8102683	19q13.2	*CYP2A6 - CYP2A7*	T	0.25	0.95	0.017

Supplemental Table S14. Association of top 22 COPD loci from overall meta-analysis in the cigarettes per day (cig/day) GWAS by the Tobacco and Genetics Consortium[26]

Chr	SNP	Effect Allele	Effect Allele Freq	Effect Size (cig/day)	Standard Error	P Value (Cig/day)	Proxy SNP Used?	COPD Index SNP	LD (r^2) with COPD Index SNP
1	rs10429950	T	0.72	-0.012	0.091	0.90	NO	--	--
2	rs16825267	C	0.93	-0.037	0.17	0.83	NO	--	--
3	rs1529672	A	0.17	0.0038	0.11	0.97	NO	--	--
3	rs2999073	C	0.88	-0.078	0.13	0.56	YES	rs2955083	1
4	rs2045517	T	0.40	-0.069	0.083	0.40	YES	rs6837671	0.97
4	rs2047409	A	0.62	-0.067	0.084	0.42	NO	--	--
4	rs11727735	A	0.93	0.12	0.16	0.48	NO	--	--
4	rs13141641	T	0.58	0.21	0.083	0.012	NO	--	--
5	rs7733088	A	0.38	0.034	0.094	0.72	NO	--	--
5	rs3734031	A	0.90	-0.098	0.13	0.46	YES	rs56168343	0.56
6	rs2076295	T	0.55	0.022	0.084	0.79	NO	--	--
6	rs2070600	T	0.04	0.069	0.21	0.74	NO	--	--
6	rs2806356	T	0.81	0.031	0.10	0.77	NO	--	--
6	rs9399401	T	0.70	0.077	0.089	0.39	NO	--	--
10	rs721917	A	0.58	-0.12	0.083	0.15	NO	--	--
14	rs754388	C	0.82	-0.14	0.12	0.23	NO	--	--
15	rs11853359	A	0.33	0.059	0.087	0.50	YES	rs1441358	0.98
15	rs17486278	A	0.65	-1.04	0.087	5.91E-33	NO	--	--
16	rs17707300	T	0.63	-0.030	0.085	0.72	NO	--	--
18	rs647097	T	0.74	-0.12	0.092	0.18	NO	--	--
19	rs7251418	A	0.28	-0.21	0.16	0.19	YES	rs12459249	0.88

Supplemental Table S15. Association of top 22 COPD loci from overall meta-analysis in the ever-smoking GWAS by the Tobacco and Genetics Consortium[26]

Chr	SNP	Effect Allele	Effect Allele Freq	OR (Ever-Smoking)	P Value (Ever-Smoking)	Proxy SNP Used?	COPD Index SNP	LD (r^2) with COPD Index SNP
1	rs10429950	T	0.71	1.00	0.84	NO	--	--
2	rs16825267	C	0.89	0.97	0.24	NO	--	--
3	rs1529672	A	0.20	0.99	0.66	NO	--	--
3	rs2999073	C	0.84	1.03	0.18	YES	rs2955083	1
4	rs2045517	T	0.41	0.98	0.18	YES	rs6837671	0.97
4	rs2047409	A	0.61	1.02	0.058	NO	--	--
4	rs11727735	A	0.91	0.98	0.38	NO	--	--
4	rs13141641	T	0.57	1.00	0.89	NO	--	--
5	rs7733088	A	0.39	0.99	0.61	NO	--	--
5	rs3734031	A	0.87	0.96	0.024	YES	rs56168343	0.56
6	rs2076295	T	0.54	0.98	0.038	NO	--	--
6	rs2070600	T	0.07	1.02	0.42	NO	--	--
6	rs2806356	T	0.79	0.99	0.41	NO	--	--
6	rs9399401	T	0.69	1.00	0.72	NO	--	--
10	rs721917	A	0.57	1.00	0.68	NO	--	--
14	rs754388	C	0.79	1.01	0.74	NO	--	--
15	rs11853359	A	0.34	0.98	0.17	YES	rs1441358	0.98
15	rs17486278	A	0.65	1.00	0.94	NO	--	--
16	rs17707300	T	0.62	1.00	0.95	NO	--	--
18	rs647097	T	0.72	1.00	0.89	NO	--	--
19	rs7251418	A	0.28	1.02	0.38	YES	rs12459249	0.88

Supplemental Table S16. Look-up of NHGRI-EBI GWAS Catalog coronary artery disease and osteoporosis-related genome-wide significant GWAS loci in our COPD association Stage 1 meta-analysis results. Only comorbidity loci with COPD P value < 0.05 are reported.

SNP	Region	Trait	GWAS Reference PMID	Mapped Gene	Comorbid Trait Risk Allele	Freq of Comorbid Trait Risk Allele in COPD	COPD OR (95% CI) for Comorbid Trait Risk Allele	COPD Association P value
rs3825807	15q25.1	Coronary heart disease	21378990	*ADAMTS7*	A	0.576	0.927 (0.891 - 0.964)	1.5E-04
rs1994016	15q25.1	Coronary artery disease	21239051	*ADAMTS7*	C	0.608	0.93 (0.892 - 0.97)	6.9E-04
rs4380028	15q25.1	Coronary heart disease	21378988	*ADAMTS7 - TRK-CTT1-2*	C	0.605	0.94 (0.906 - 0.975)	0.0010
rs974819	11q22.3	Coronary heart disease	21378988	*LOC105369463 - LOC102723862*	T	0.332	0.938 (0.902 - 0.976)	0.0014
rs2128739	11q22.3	Coronary artery disease	26343387	*LOC105369463 - LOC102723862*	A	0.311	0.94 (0.901 - 0.98)	0.0037
rs13336428	16p13.3	Bone mineral density	22504420	*CLCN7 - PTX4*	A	0.447	0.942 (0.904 - 0.982)	0.0045
rs228769	17q21.31	Bone mineral density (hip)	19801982	*LOC105371789 - HDAC5*	G	0.261	0.938 (0.896 - 0.982)	0.0062
rs7932354	11p11.2	Bone mineral density (hip)	19801982	*ARHGAP1 - ZNF408*	T	0.345	1.06 (1.01 - 1.11)	0.0097
rs2519093	9q34.2	Coronary artery disease	26343387	*ABO*	T	0.192	1.06 (1.01 - 1.11)	0.018
rs56062135	15q22.33	Coronary artery disease	26343387	*SMAD3*	C	0.774	0.946 (0.901 - 0.992)	0.024
rs579459	9q34.2	Coronary heart disease	21378990	*ABO - SURF6*	C	0.215	1.05 (1.01 - 1.1)	0.028
rs10048146	16q24.1	Bone mineral density (spine)	19801982	*LOC105371390 - LOC105371391*	G	0.187	1.05 (1.01 - 1.11)	0.029
rs16921914	11p13	Bone mineral density (spine)	19801982	*DCDC5 - DCDC1*	A	0.279	1.05 (1 - 1.09)	0.030
rs964184	11q23.3	Coronary heart disease	21378990	*LOC105369514 - ZPR1*	G	0.152	0.945 (0.898 - 0.995)	0.032
rs6689306	1q21.3	Coronary artery disease	26343387	*IL6R*	A	0.412	0.958 (0.92 - 0.997)	0.034
rs736825	12q13.13	Bone mineral density	22504420	*HOXC5 - HOXC4 - HOXC6*	C	0.643	0.956 (0.917 - 0.997)	0.035
rs4468572	15q25.1	Coronary artery disease	26343387	*ADAMTS7 - TRK-CTT1-2*	C	0.566	0.958 (0.921 - 0.997)	0.036
rs344081	3q25.31	Bone mineral density	22504420	*LEKR1*	T	0.787	1.06 (1 - 1.12)	0.043
rs163879	11p14.1	Bone mineral density	22504420	*DCDC5*	T	0.651	0.96 (0.921 - 1)	0.048

Supplemental Table S17. Quantitative imaging GWAS results[27] for each of the 22 genome-wide significant loci from our COPD association overall meta-analysis.

				Emphysema				Airways			
rsID	Closest Gene	Locus	COPD Risk Allele	%LAA-950 Beta (95% CI)	%LAA-950 P Value	Perc15 Beta (95% CI)	Perc15 P Value	Pi10 Beta (95% CI)	Pi10 P Value	Wall Area Percent Beta (95% CI)	Wall Area Percent P Value
rs13141641	*HHIP*	4q31.21	T	0.12 (0.16, 0.091)	1.7E-12	-2.2 (-1.5, -2.9)	8.4E-10	-4e-04 (0.0033, -0.0041)	0.41	0.1 (0.19, 0.016)	0.027
rs17486278	*CHRNA5*	15q25.1	C	0.094 (0.13, 0.06)	4.9E-08	-1.4 (-0.75, -2.1)	4.4E-05	0.0031 (0.0066, -0.00043)	0.099	0.0036 (0.089, -0.082)	0.41
rs7733088	*HTR4*	5q32	G	0.022 (0.056, -0.011)	0.24	-0.5 (0.2, -1.2)	0.20	0.002 (0.0055, -0.0015)	9.5E-03	0.16 (0.24, 0.076)	2.7E-04
rs9399401	*ADGRG6*	6q24.1	T	0.065 (0.1, 0.031)	3.1E-04	-1.2 (-0.47, -1.9)	1.6E-03	-0.0013 (0.0024, -0.005)	0.27	0.021 (0.11, -0.067)	0.59
rs1441358	*THSD4*	15q23	G	0.05 (0.082, 0.017)	4.1E-03	-0.91 (-0.22, -1.6)	0.012	-1e-04 (0.0034, -0.0036)	0.96	0.043 (0.13, -0.041)	0.38
rs6837671	*FAM13A*	4q22.1	G	0.066 (0.098, 0.034)	8.4E-05	-1.2 (-0.55, -1.9)	1.6E-04	0.0036 (0.0071, 7.2e-05)	0.0525	0.13 (0.21, 0.05)	2.1E-03
rs11727735	*GSTCD*	4q24	A	0.036 (0.11, -0.036)	0.30	-0.67 (0.8, -2.1)	0.27	0.012 (0.019, 0.0038)	5.7E-03	0.39 (0.57, 0.21)	4.47E-05
rs754388	*RIN3*	14q32.12	C	0.093 (0.14, 0.05)	2.7E-05	-1.8 (-0.96, -2.7)	5.5E-05	0.0021 (0.0068, -0.0026)	0.45	0.11 (0.22, 0.0018)	0.040
rs2047409	*TET2*	5q33.3	A	0.034 (0.067, 0.0018)	0.012	-0.78 (-0.1, -1.5)	0.016	-3e-04 (0.0032, -0.0038)	0.92	0.058 (0.14, -0.026)	0.21
rs2955083	*EEFSEC*	4q24	A	0.058 (0.11, 0.0085)	0.029	-0.96 (0.075, -2)	0.087	0.0095 (0.015, 0.0042)	6.7E-04	0.093 (0.22, -0.033)	0.19
rs56168343*	*ADAM19*	3q21.3	T	0.052 (0.098, 0.007)	0.031	-1.3 (-0.38, -2.2)	7.7E-03	0.0022 (0.0073, -0.0029)	0.45	0.037 (0.15, -0.079)	0.60
rs7186831	*CFDP1*	16q23.1	A	0.0095 (0.049, -0.03)	0.45	-0.023 (0.8, -0.85)	0.98	0.0053 (0.0096, 0.00099)	0.020	0.14 (0.24, 0.036)	0.011
rs10429950	*TGFB2*	1q41	T	0.08 (0.11, 0.045)	1.1E-05	-1.3 (-0.6, -2.1)	4.9E-04	0 (0.0037, -0.0037)	0.99	0.084 (0.17, -0.0057)	0.087
rs2070600	*AGER*	6p21.32	C	0.24 (0.32, 0.15)	4.6E-09	-3.8 (-2, -5.5)	2.2E-06	-0.0036 (0.0058, -0.013)	0.53	0.073 (0.29, -0.15)	0.58
rs17707300	*CCDC101*	16p11.2	C	0.0051 (0.039, -0.029)	0.83	0.28 (0.98, -0.42)	0.49	0.0035 (0.0072, -0.00022)	0.085	0.055 (0.14, -0.032)	0.26
rs2806356	*ARMC2*	6q21	C	0.02 (0.064, -0.023)	0.42	-0.36 (0.54, -1.3)	0.49	0.0039 (0.0086, -8e-04)	0.13	0.19 (0.31, 0.082)	9.5E-04
rs16825267	*PID1*	2q36.3	C	0.082 (0.14, 0.019)	0.014	-1.7 (-0.38, -3)	0.015	-4e-04 (0.0063, -0.0071)	0.94	0.023 (0.18, -0.14)	0.83
rs2076295	*DSP*	6p24.3	T	0.031 (0.063, -0.00085)	0.014	-0.64 (0.014, -1.3)	0.066	-0.0012 (0.0021, -0.0045)	0.57	0.035 (0.12, -0.046)	0.46
rs647097	*MTCL1*	18p11.22	C	-0.001 (0.034, -0.036)	0.093	0.03 (0.75, -0.69)	0.32	0.0012 (0.0049, -0.0025)	0.59	0.077 (0.16, -0.011)	0.11
rs1529672	*RARB*	3p24.2	C	0.07 (0.11, 0.029)	8.4E-04	-1.7 (-0.81, -2.5)	2.0E-04	0.0012 (0.0057, -0.0033)	0.47	0.092 (0.2, -0.014)	0.11
rs721917	*SFTPD*	10q22.3	G	0.0099 (0.042, -0.022)	0.62	-0.085 (0.58, -0.75)	0.86	3e-04 (0.0036, -0.003)	0.71	0.07 (0.15, -0.012)	0.12
rs12459249	*CYP2A6*	19q13.2	C	0.073 (0.11, 0.035)	2.2E-04	-1.2 (-0.46, -2)	2.4E-03	8e-04 (0.0049, -0.0033)	0.78	0.08 (0.18, -0.016)	0.14

* The indel rs113897301 was not available in the quantitative imaging results, so rs56168343, the second most significant variant at the 3q21.3 locus, was used as a proxy.

Supplemental Table S18. For each of the 22 genome-wide significant loci, the highest LD (r^2) proxy SNP that was present in HapMap phase III was chosen as the input variant for the PrixFixe method of gene prioritization.

Chr:position index variant	Chr:position HapMap III proxy variant	rsID HapMap III proxy variant	P value of proxy variant*	LD (r^2) between index and proxy
1:218624533	1:218624533	rs10429950	1.83E-07	1.00
2:229569919	2:229592304	rs3732192	1.78E-07	0.99
3:127961178	3:127991938	rs2811416	3.91E-07	1.00
3:25520582	3:25520582	rs1529672	2.37E-09	1.00
4:106137033	4:106131210	rs2007403	3.32E-06	1.00
4:106631870	4:106631870	rs11727735	1.55E-08	1.00
4:145506456	4:145474473	rs1489759	5.48E-15	0.92
4:89873092	4:89869332	rs2869967	1.55E-14	0.99
5:147856333	5:147856333	rs7733088	4.41E-14	1.00
5:156929077:ID	5:156932376	rs2277027	8.31E-06	0.37
6:109266255	6:109268050	rs2798641	1.46E-05	1.00
6:142668901	6:142670955	rs7756305	1.52E-08	1.00
6:32151443	6:32062687	rs2071293	4.12E-03	0.01
6:7563232	6:7563232	rs2076295	4.95E-08	1.00
10:81706324	10:81706324	rs721917	2.11E-06	1.00
14:93115410	14:93111120	rs11624512	7.76E-11	0.99
15:71612514	15:71612514	rs1441358	2.06E-10	1.00
15:78867482	15:78878541	rs951266	1.72E-23	0.99
16:28593347	16:28593347	rs17707300	6.25E-09	1.00
16:75473155	16:75479153	rs2161684	2.85E-04	0.62
18:8808464	18:8808464	rs647097	3.03E-06	1.00
19:41339896	19:41341589	rs7251418	1.93E-03	0.85

* The COPD association Stage 1 meta-analysis P value for the HapMap III compatible proxy variant.

Supplemental Table S19. Gene prioritization scores using PrixFixe method of identifying co-functional genes across the genome-wide significant loci in our study.

Gene	Region	PrixFixe Score
TGFB2	chr1:q41	0.480
RARB	chr3:p24.2	0.188
RUVBL1	chr3:q21.3	0.160
EEFSEC	chr3:q21.3	0.021
FAM13A	chr4:q22.1	0.164
PPA2	chr4:q24	0.179
TET2	chr4:q24	0.056
HHIP	chr4:q31.21	0.184
HTR4	chr5:q32	0.026
ADAM19	chr5:q33.3	0.129
CFB	chr6:p21.32--p21.33	0.199
TNXB	chr6:p21.32--p21.33	0.187
EGFL8	chr6:p21.32--p21.33	0.111
C2	chr6:p21.32--p21.33	0.083
SLC44A4	chr6:p21.32--p21.33	0.052
EHMT2	chr6:p21.32--p21.33	0.051
RNF5	chr6:p21.32--p21.33	0.022
CYP21A2	chr6:p21.32--p21.33	0.011
CYP21A1P	chr6:p21.32--p21.33	0.009
LSM2	chr6:p21.32--p21.33	0.008
RDBP	chr6:p21.32--p21.33	0.006
DSP	chr6:p24.3	0.259
SNRNP48	chr6:p24.3	0.007
GPR126	chr6:q24.1--q24.2	0.053
SFTPD	chr10:q22.3	0.321
C10orf57	chr10:q22.3	0.073
RIN3	chr14:q32.12	0.102
THSD4	chr15:q23	0.164
CHRNB4	chr15:q25.1	0.262
PSMA4	chr15:q25.1	0.261
CHRNA3	chr15:q25.1	0.124
CHRNA5	chr15:q25.1	0.104
IREB2	chr15:q25.1	0.057
CLN3	chr16:p11.2	0.333
NUPR1	chr16:p11.2	0.157
CCDC101	chr16:p11.2	0.115
IL27	chr16:p11.2	0.111
APOBR	chr16:p11.2	0.078
SULT1A2	chr16:p11.2	0.071
SULT1A1	chr16:p11.2	0.070
BCAR1	chr16:q23.1	0.135
GABARAPL2	chr16:q23.1	0.107
CFDP1	chr16:q23.1	0.076
CHST6	chr16:q23.1	0.057
TMEM170A	chr16:q23.1	0.053
CTRB2	chr16:q23.1	0.047
CCDC165	chr18:p11.22	0.031
EGLN2	chr19:q13.2	0.135
NUMBL	chr19:q13.2	0.080
CYP2A6	chr19:q13.2	0.067
SNRPA	chr19:q13.2	0.055
ADCK4	chr19:q13.2	0.025
MIA	chr19:q13.2	0.018
ITPKC	chr19:q13.2	0.011
CYP2G1P	chr19:q13.2	0.002
C19orf54	chr19:q13.2	0.002

The 2q36.2 locus (rs16825267; closest gene *PID1*) is not present in PrixFixe's cofunction networks, and is consequently not represented in the table.
*Genes with PrixFixe prioritization score of "0" are not shown. Thus, the 6q21 locus (rs2806356; closest genes *ARMC2*) is not represented in the table.

Supplemental Table S20: Results of fine-mapping analysis (see text).

Chr:Pos Index variant	European GWAS		African-American GWAS Top variant P < 0.001		
	Markers in 95% credible set	Posterior probability, top-ranked marker	r2 (with top overall variant)	D' (with top overall variant)	Present in European credible set
1:218624533	121	0.13	4.30E-02	0.42	N
2:229569919	29	0.11	7.30E-04	0.45	N
3:127961178	192	0.043			
3:25520582	10	0.61	1.10E-03	0.057	N
4:106137033	29	0.17			
4:106631870	267	0.033			
4:145506456	29	0.66	7.90E-02	0.71	N
4:89873092	9	0.85	1.40E-03	0.085	N
5:147856333	19	0.36			
5:156929077	1	0.99	4.30E-05	0.014	N
6:109266255	384	0.28	7.30E-03	1	N
6:142668901	71	0.057			
6:32151443	30	0.37			
6:7563232	8	0.52			
10:81706324	1193	0.15	4.20E-01	0.93	Y
14:93115410	16	0.24			
15:71612514	37	0.44	3.40E-01	0.93	N
15:78867482	18	0.37	1.90E-03	0.082	N
16:28593347	97	0.11			
16:75473155	542	0.22	2.20E-01	0.72	Y
18:8808464	172	0.29			
19:41339896	22	0.52	2.10E-02	0.39	N

Supplemental Methods:

Cohort descriptions and cohort-specific methods

ARIC: Atherosclerosis Risk in Communities (ARIC)[28] (NCT00005131), is a population based study of risk factors for atherosclerosis and its sequelae in adults from four U.S. field centers aged 45-64 at recruitment in 1987-1989. Institutional Review Board (IRB) approval was obtained at all associated study centers and informed consent was obtained for all participants. ARIC spirometry measurements were made with a Collins Survey II water-seal spirometer (Collins Medical, Inc.) and Pulmo-Screen II software (PDS Healthcare Products, Inc.). Genotyping was done using the AffymetrixGeneChip SNP Array 6.0 and imputation was performed using IMPUTE2. Quality control steps for genotyping data included exclusions for call rate <95%, minor allele frequency <1%, HWE P<10-5, no chromosomal location, suspected first-degree relative of an included individual based on genotype data, or more than 8 standard deviations for any of the first ten principal components. The current analysis includes 7,224 Caucasian subjects with genotyping data, pulmonary function measures and complete covariate information. Imputation was performed using the 1000 Genomes[29] cosmopolitan reference panel, Phase I v3, using MaCH[30,31]. Logistic regression was performed using ProbABEL[32] v0.1-3.

B58C: The British 1958 birth cohort is a long-term follow-up of persons born in England, Scotland and Wales during one week in 1958. At age 44-45 years, cohort members were invited to participate in a biomedical examination in their home at which spirometry was performed in the standing position, without noseclips, using a Vitalograph Micro spirometer. Informed consent was obtained for all participants and ethical approval for the medical examination of the British 1958 Birth Cohort was obtained from the South-East Multicentre Research Ethics Committee (SE-MREC). Details of this fieldwork and the spirometric protocol have been published elsewhere[33,34]. At the same examination, blood samples were collected with consent for DNA extraction and production

of immortalized cell lines. The resulting DNA collection has been widely used as a nationally representative reference set, including the Wellcome Trust Case-Control Consortium[35], the Type 1 Diabetes Genetic Consortium[36] and the GABRIEL asthma consortium[25]. Genotyping platform was performed using Illumina 550K/610K. Imputation was performed using the 1000 Genomes[29] March 2012, cosmopolitan reference panel and MaCH/minimac[30,31,37]. ProbABEL[32] 0.1-9 was used for analysis.

Cardiovascular Health Study (CHS): The Cardiovascular Health Study (CHS) is a population-based cohort study of risk factors for coronary heart disease and stroke in adults ≥65 years conducted across four field centers[38] (NCT00005133 and NCT00149435). Local IRB approval was obtained at participating centers and written informed consent was obtained for all participants. The original predominantly European ancestry cohort of 5,201 persons was recruited in 1989-1990 from random samples of the Medicare eligibility lists; subsequently, an additional predominantly African-American cohort of 687 persons were enrolled for a total sample of 5,888. Blood samples were drawn from all participants at their baseline examination and DNA was subsequently extracted from available samples. European ancestry participants were excluded from the GWAS study sample due to the presence at study baseline of coronary heart disease, congestive heart failure, peripheral vascular disease, valvular heart disease, stroke or transient ischemic attack. Genotyping was performed at the General Clinical Research Center's Phenotyping/Genotyping Laboratory at Cedars-Sinai among CHS participants who consented to genetic testing and had DNA available using the Illumina 370CNV BeadChip system (for European ancestry participants, in 2007) or the Illumina HumanOmni1-Quad_v1 BeadChip system (for African-American participants, in 2010). Additional genotypes were provided from the ITMAT-Broad-CARe (IBC) Illumina iSELECT chip. Imputation was performed using 1000 Genomes[29] Phase 1 v3 haplotypes and minimac[37] (2012-11-16). Logistic regression was performed in R, adjusting for CHS clinic (4 sites) and PCs 1-5.

COPD Pathology: Addressing Critical gaps, Early Treatment & Diagnosis and Innovative Concepts (COPACETIC): The COPACETIC[39] cohort includes Dutch (all Caucasian-ancestry) participants from the NELSON[40] lung cancer screening trial recruited by the University Medical Centers of Groningen and Utrecht. The trials were approved by the Dutch Minister of Health and the ethics board at each participating centre. All participants gave written informed consent. All participants were substantial smokers (16.5 or more pack-years) aged between 50–75 years. Pre-bronchodilator pulmonary function tests were performed with standardized equipment according to the American Thoracic Society (ATS) and European Respiratory Society (ERS) guidelines. All participants provided written informed consent. Blood samples were genotyped using the Illumina 610 Quad BeadChip. Genotype calling was done with the standard algorithm provided by Illumina and implemented in GenomeStudio software. Quality control was performed using PLINK[41] software. Samples were excluded if more than 5% of genotype data was missing (n = 19), if samples were duplicated (n = 29), detected as an ethnic outlier (based on genetic distance derived from principal components c1 and c2; n = 49), derived from a relative of another participant (based on IBS estimation, Phat > 0.5, n= 13), or if the participant had a diagnosis of lung cancer (n =16). Imputation was performed using the Impute2 pipeline developed by the GoNL - Impute team. In summary: The study data was lifted over from human genome build 36 to build 37 using PLINK and UCSC liftOver, followed by alignment to reference data and filtering on MAF larger than 1%, Hardy-Weinberg Equilibrium p-value of 1e-4 and a call rate higher than 0.95. Afterwards the study data was pre-phased per chromosome using SHAPEIT2 v.2.644[42]. Finally the imputation over genome chunks of 5Mb was performed using IMPUTE2 2.3.0[43]. 1000 Genomes[29] phase1 integrated version 3 was used as reference panel. Associations between genomic dosages with moderate/severe COPD were assessed with logistic regression models adjusted for age, pack years smoked, and sex; current smoking status was not

available at the time of analysis. All analysis were performed in software package PLINK version 1.07[41,44].

COPDGene: Details of the COPDGene Study (NCT00608764, www.copdgene.org) have been described previously[45,46]. Local IRB approval was obtained at all study centers and all study participants provided written informed consent. Eligible subjects were of non-Hispanic white or African-American ancestry, aged 45-80 years old, with a minimum of 10 pack-years of smoking and no lung disease (other than COPD or asthma). Genotyping was performed by Illumina (San Diego, CA) on the HumanOmniExpress array. Subjects were excluded for missingness, heterozygosity, chromosomal aberrations, gender check, population outliers, and cryptic relatedness. Genotyping at the Z and S alleles was performed in all subjects. Subjects known or found to have alpha-1 antitrypsin deficiency were excluded. Markers were excluded based on missingness, Hardy-Weinberg P-values, and low minor allele frequency. Imputation on the COPDGene cohorts was performed using MaCH[30,31] and minimac[37] (version 2012-10-09). Reference panels for the non-Hispanic whites and African-Americans were the 1000 Genomes[29] Phase I v3 European (EUR) and cosmopolitan reference panels, respectively. Variants with an r^2 value of ≤ 0.3 were removed from further analysis. Logistic regression was performed on pre-bronchodilator spirometry defined cases and controls, adjusting for age, sex, pack-years, current smoking, and principal components of ancestry, separately in non-Hispanic whites and African-Americans, using PLINK2[47,48].

Evaluation of COPD Longitudinally to Identify Predictive Surrogate End-points (ECLIPSE; SCO104960, NCT00292552, www.eclipse-copd.com**):** Details of the ECLIPSE study and genome-wide association analysis have been described previously[49]. The ECLIPSE study was conducted in accordance with the Declaration of Helsinki and good clinical practice guidelines, and was approved by the relevant ethics and review boards at the participating centers. All participants provided

written informed consent. ECLIPSE was an observational 3-year study of COPD. Both cases and controls were aged 40-75 with at least a 10 pack-year smoking history without other respiratory diseases; cases were post-bronchodilator GOLD[50] Grade 2 and above COPD, and controls had normal spirometry (FEV_1 > 85% predicted). Genotyping was performed using the Illumina HumanHap 550 V3 (Illumina, San Diego, CA). Subjects and markers with a call rate of < 95% were excluded. Population stratification exclusion and adjustment on self-reported white subjects was performed using EIGENSTRAT (EIGENSOFT Version 2.0)[51]. Imputation was performed using MaCH[30,31] and minimac[37] (version 2012-10-09) and the 1000 Genomes[29] Phase I v3 European (EUR) reference panel. Logistic regression was performed on pre-bronchodilator spirometry defined cases and controls, adjusting for age, sex, pack-years, current smoking, and principal components of ancestry using PLINK2[47,48].

Expression Quantitative Trait Loci – Lung (eQTL): The lung eQTL study has been described previously[23]. Briefly, patients who underwent lung surgery were recruited at three academic sites: Laval University, University of British Columbia (UBC), and University of Groningen, henceforth referred to as Laval, UBC, and Groningen, respectively. Patients from Laval were those undergoing lung cancer surgery, the majority of the UBC patients had lung resection for small peripheral lung lesions with some samples derived from autopsy or at the time of lung transplantation. At Groningen, patients were recruited from those having surgery for various lung diseases, including patients that underwent therapeutic resection for lung tumors and lung transplantation. All patients provided written informed consent and the study was approved by the ethics committees of the Institut universitaire de cardiologie et de pneumologie de Québec and the UBC-Providence Health Care Research Institute Ethics Board for Laval and UBC, respectively. The study protocol was consistent with the Research Code of the University Medical Center Groningen and Dutch national ethical and professional guidelines. COPD diagnosis and severity were determined according to the

GOLD[50] recommendations. Patients whose lung function could have been influenced by lung diseases other than COPD and lung cancer were excluded. This includes patients with alpha-1 antitrypsin deficiency (n=11), amyloidosis (n=1), bronchiectasis (n=3), bronchiolitis obliterans (n=2), bronchopulmonary dysplasia (n=2), cystic fibrosis (n=14), idiopathic pulmonary fibrosis (n=13), langerhans cell histiocytosis (n=1), lymphangioleiomyomatosis (n=1), primary pulmonary hypertension (n=4), sarcoidosis (n=3) and vascular malformation (n=1).

Genotyping was carried out using the Illumina Human1M-Duo BeadChip. Standard genotyping quality controls were performed independently in the Laval, UBC and Groningen cohorts. Each genotyping set was filtered for low-quality loci with 10th percentile of Illumina GenCall score ≤0.1, call rate <97%, Hardy-Weinberg equilibrium $P < 1x10^{-7}$, and minor allele frequency (MAF) <1%. Samples were excluded after consideration for the 10th percentile of Illumina GenCall score ≤0.2, genotype completion rate <90%, genotypic and phenotypic gender mismatch, unexpected duplicates and genetic relatedness, and genetic background outliers detected by STRUCTURE (k=4) with HapMap[52] subjects as internal controls. PC1 to PC10 were calculated with EIGENSOFT 4.2. Imputation was performed with SHAPEIT[53] (Shapeit.v1.ESHG.linux.x64) and IMPUTE2[43] (impute_v2.2.2_x86_64_dynamic), and the reference set from the 1000 genomes[29] project (ALL_1000G_phaselintegrated_v3). Only SNPs that passed the cutoff $r^2 \geq 0.95$ were retained. Single-marker association tests were performed with PLINK v1.90[47,48] adjusting for age, sex, smoking status, pack-years, ever smoking status, center and PC1 to PC10.

Framingham Heart Study (FHS; NCT00005121**):** Details on pulmonary function in the FHS have been previously published[54,55]. The study was IRB-approved at the relevant institutions and all participants provided written informed consent. Data from the most recent exam for each of the three generations of families participating in the FHS were analyzed. Genotypes were from the

Affymetrix 500K array supplemented by the Affymetrix MIPS 50K. From a total number of 549,781 genotyped SNPs, 412,053 were used with the MaCH[30,31] program for phasing. A total of 137,728 genotyped SNPs were removed based on the following filtering criteria: 22,018 SNPs for Hardy-Weinberg Equilibrium p-value of less than $1x10^{-6}$, 48,285 SNPs for a call rate of less than 96.9%, 66,063 SNPs for a minor allele frequency of less than 0.01, 82 SNPs due to not mapping correctly from Build 36 to Build 37 locations, 428 SNPs missing a physical location, 25 SNPs for number of Mendelian errors greater than 1000, 786 SNPs due to not being on chromosomes 1-22 or X and 41 SNPs because they were duplicates. MaCH/minimac[30,31,37] were used in this imputation to impute the FHS sample using the November 2010 release of the 1000 Genomes[29] multi-ethnic panel comprised of 1,092 samples. We used GEE implemented in the R package geepack with independent correlation matrix and clustering based on family, adjusted for sex, age, smoking status, pack years and principal component 1 (to adjust for population stratification).

KARE: Details on the Korean Association Resource project (KARE) have been previously published[56,57]. KARE was initiated in 2007 to undertake genome-wide analyses analysis among 10,038 participants in the rural-based Ansung and city-based Ansan South Korean cohorts. The study was approved at appropriate IRBs from participating institutions and participants provided informed consent. KARE3 data were obtained from the third phenotype collection in 2008; lung function was collecting using a portable spirometer (Vmax-2130, Sensor Medics, Yorba Linda, CA, USA) according to standardized protocols of the American Thoracic Society (ATS). Genotyping was performed using the Affymetrix Genome-Wide Human array 5.0 (Affymetrix, Inc., Santa Clara, CA, USA). We performed imputation using IMPUTE2 and the 1000 Genomes[29] Phase 3 cosmopolitan panel. Markers were converted to genotype from dosage with call rate ≥ 95%, minor allele frequency ≥ 1%, p for HWE ≥ $1.0x10^{-5}$, imputation quality score ≥ 0.9. Logistic regression was performed using PLINK[44].

LifeLines: The LifeLines Cohort Study is a large population-based cohort study and biobank that was established as a resource for research on complex interactions between environmental, phenotypic and genomic factors in the development of chronic diseases and healthy aging[58-60]. Between 2006 and 2013, inhabitants of the northern part of The Netherlands and their families were invited to participate, thereby contributing to a three-generation design. Participants visited one of the LifeLines research sites for a physical examination, including pre-bronchodilator spirometry following ATS guidelines. All participants signed an informed consent form before they received an invitation for the physical examination. The LifeLines Cohort Study is conducted according to the principles of the Declaration of Helsinki and in accordance with research code University Medical Center Groningen (UMCG), The Netherlands. The LifeLines study is approved by the medical ethical committee of the UMCG.

Blood samples for a subset of individuals were genotyped using the Illumina CytoSNP-12v2 array. Independent Caucasian-ancestry samples (n = 13,436) have been imputed using the 1000 Genomes[29] phase1 v3 reference panels. Quality control of the data is based on SNP filtering on MAF above 0.001, HWE p-value > $1x10^{-4}$, call rate of 0.95 using PLINK[41], and PCA to check for population outliers resulting in 268,407 SNPs and 13,436 samples kept for genome-wide association analysis. Before imputation, the genotypes were pre-phased using SHAPEIT2[42] and aligned to the reference panels using Genotype Harmonizer (www.molgenis.org/systemsgenetics) in order to resolve strand issues. The samples were imputed using minimac[37] (version 2012-10-09), yielding 28,681,763 SNPs. Associations between genomic dosages with moderate/severe COPD were assessed with logistic regression models adjusted for age, smoking status (never/ever), current smoking (no/yes), pack years smoked and sex. All analysis were performed in software package PLINK version 1.07[41,44].

Lovelace: The LSC has been actively enrolling smokers from the Albuquerque, NM metropolitan area since 2001 and has been previously decribed[61]. All participants provided written informed consent and the study was approved by the relevant IRB. Longitudinal studies to predict lung cancer and other chronic pulmonary diseases are conducted through assessing biomarkers present in sputum, blood, and urine samples. Enrollment was restricted to current and former smokers age 40 to 74 y with a minimum of 10 pack-years of smoking and no personal history of lung cancer. A detailed questionnaire written in English was used to collect information on demographics, medical, cigarette smoking, and exposure history, socioeconomic status, diet, and quality of life. Sputum samples were collected by induction and stored in Saccomanno's fixative. Pulmonary function testing was performed at each visit. All participants signed a consent form, and the Western Institutional Review Board approved this project. The GWAS discovery set was comprised of 1200 Caucasian (self-reported) smokers with methylation status of 12 tumor suppressor genes measured in sputum DNA samples.

The HumanOmni2.5-4v1-H BeadChip (Illumina, San Diego, CA) was used to genotype 2,450,000 SNPs in 1200 Caucasian smokers from the LSC. We removed 37 subjects due to low call rate (< 95%, n = 7), low heterozygosity (n = 1), low Caucasian ancestry (< 85%, n = 2) and high relatedness with other samples (n = 27). Furthermore, SNPs were excluded if they had a call rate of < 90%, a minor allele frequency (MAF) < 0.008, or $P < 10^{-8}$ for Hardy-Weinberg equilibrium test, or were on Y or pseudo-autosomal region of X. The MAF cutoff is a technical one to identify at least 20 heterozygotes for accurate genotype clustering required by GenomeStudio. After quality assessment, 1163 subjects with 1,599,980 SNPs remained in the genetic association analysis. Logistic regression was performed using PLINK.

Multi-Ethnic Study of Atherosclerosis (MESA): MESA is a longitudinal study of subclinical cardiovascular disease and risk factors that predict progression to clinically overt cardiovascular

disease or progression of the subclinical disease[62]. Between 2000 and 2002, MESA recruited 6,814 men and women 45 to 84 years of age from Forsyth County, North Carolina; New York City; Baltimore; St. Paul, Minnesota; Chicago; and Los Angeles. Exclusion criteria were clinical cardiovascular disease, weight exceeding 136 kg (300 lb.), pregnancy, and impediment to long-term participation. The MESA Family Study recruited 1,595 African American and Hispanic participants, generally siblings of MESA participants, using the same inclusion and exclusion criteria as MESA except that clinical cardiovascular disease was permitted. The MESA Air Pollution Study recruited an additional 257 participants from Los Angeles and Riverside County, CA, and Rockland County, NY, using the same criteria as MESA, except that participants were ages 50 to 89 who lived in the area more than 50% of the year and had no plans to move in the next five years[63]. The MESA Lung Study performed spirometry following the 2005 ATS/ERS guidelines in a subset of the MESA and MESA Family Studies and all of the new recruits in the MESA Air Pollution Study, as previously described[64]. All participants provided informed consent and the protocols of MESA were approved by the IRBs of collaborating institutions and the National Heart, Lung and Blood Institute.

Participants in the original MESA cohort, the MESA Family Study and the MESA Air Pollution Study who consented to genetic analyses were genotyped in 2009 using the Affymetrix Human SNP array 6.0. Genotype quality control for these data included filter on SNP level call rate < 95%, individual level call rate < 95%, heterozygosity > 53%, described previously[65]. The cleaned genotypic data was deposited with MESA phenotypic data into dbGaP as the MESA SHARe project (study accession phs000209); 8,224 consenting individuals (2,685 White, 2,588 non-Hispanic African-American, 2,174 Hispanic, 777 Chinese) were included, with 897,981 SNPs passing study specific quality control (QC). For GWAS, IMPUTE version 2.2.2 was used to perform imputation for the MESA SHARe participants using the cosmopolitan 1,000 Genomes[29] Phase 1 v3 March 2012 reference set. Logistic regression was performed using SNPTEST v2.4.0[66].

National Emphysema Treatment Trial (NETT; NCT00000606**,** www.nhlbi.nih.gov/health/prof/lung/nett/**) and Normative Aging Study (NAS)**: Details of the National Emphysema Treatment Trial have been described previously[67]. NETT was a multicenter clinical trial to evaluate lung volume reduction surgery. All participants provided written informed consent and the study was approved by the IRB at all participating institutions. Enrolled subjects had severe airflow obstruction by post-bronchodilator spirometry (FEV_1 < 45% predicted) and evidence of emphysema on computed tomography (CT) chest imaging; exclusion criteria included significant sputum production or bronchiectasis. A subset of 382 self-reported white subjects without severe alpha-1 antitrypsin deficiency were enrolled in the NETT Genetics Ancillary Study.

The Normative Aging Study is a longitudinal study of healthy men established in 1963 and conducted by the Veterans Administration (VA)[68]. Participants provided written informed consent at each visit, and the VA Boston Healthcare System Institutional Review Board approved the study. Men aged 21 to 80 years from the greater Boston area, free of known chronic medical conditions, were enrolled. Smoking controls were of self-reported white ancestry and at least 10 pack-years of cigarette smoking with no evidence of airflow obstruction on spirometry on their most recent visit. Genotyping for NETT-NAS was performed using the Illumina Quad 610 array (Illumina, San Diego, CA), with quality control, population stratification adjustment, as described previously[69,70]. Imputation was performed using MaCH[30,31] and minimac[37] (version 2012-10-09) and the 1000 Genomes Phase I v3 European (EUR) reference panel. Logistic regression was performed on pre-bronchodilator spirometry defined cases and controls, adjusting for age, sex, pack-years, current smoking, and principal components of ancestry using PLINK2[47,48].

GenKOLS (Norway): Details on the Norwegian GenKOLS (Genetics of Chronic Obstructive Lung Disease, GSK code RES11080) study have been described previously[71]. Subjects with > 2.5 pack

years of smoking history were recruited from Bergen, Norway; cases had post-bronchodilator GOLD 2 or greater disease[50], while controls had normal spirometry; subjects with severe alpha-1 antitrypsin deficiency and other lung diseases (aside from asthma) were excluded. The study was performed in accordance with the ethical standards laid down in the Helsinki Declaration. The Regional Committee for Medical Research Ethics (REK Vest), the Norwegian Data Inspectorate and the Norwegian Department of Health approved the case–control study. A written informed consent was obtained from all participants. Genotyping was performed using Illumina HumanHap 550 arrays (Illumina, San Diego, CA), with quality control, population stratification adjustment as previously described. Imputation was performed using MaCH[30,31] and minimac[37] (version 2012-10-09) and the 1000 Genomes[29] Phase I v3 European (EUR) reference panel. Logistic regression was performed on pre-bronchodilator spirometry defined cases and controls, adjusting for age, sex, pack-years, current smoking, and principal components of ancestry using PLINK2[47,48].

The Rotterdam Study: The Rotterdam Study is a prospective population-based cohort study founded in 1990 in a suburb of Rotterdam, the Netherlands[72,73]. The first cohort (RS-I) consists of 7,983 participants, aged 55 years and over. The second cohort (RS-II) was recruited in 2000 with the same inclusion criteria. The third cohort (RS-III) consists of 3,932 participants, aged 45 years and over and was recruited in 2006. The Rotterdam Study was approved by the institutional review board (Medical Ethics Committee) of the Erasmus Medical Center and by the review board of The Netherlands Ministry of Health, Welfare and Sports. All participants provided written informed consent. Spirometry was performed by trained paramedical personnel using the Master Screen® PFT Pro (CareFusion, San Diego, CA) according to the American Thoracic Society(ATS)/European Respiratory Society (ERS) guidelines.

A total of 6,318 subjects were genotyped in RS I, 2,516 in RS II and 3,540 subjects in RS III. Exclusions included a call rate < 98%, Hardy-Weinberg $P < 1x10^{-6}$ and MAF < 0.01%. A total of

6,291 for RS I, 2,157 for RS II and 3,048 for RS III passed genotyping quality control. Regression coefficients and their standard errors were determined using the ProbABEL [32] program according to an additive model.

Subpopulations and intermediate outcome measures in COPD study (SPIROMICS; NCT01969344**):** Participants of the NHLBI SPIROMICS study were 40-80 years of age at baseline with a smoking history ≥20 pack-years. Recruitment included non-smokers, smokers without COPD, mild-moderate COPD, and severe COPD[74]. All participants provided written informed consent and the Institutional Review Boards/Ethics Committees of all the cooperating institutions approved the study protocols. All SPIROMICS study participants have been comprehensively characterized with standardized comprehensive questionnaires (including clinical outcomes such as exacerbations requiring healthcare utilization, medication use, St. George's Respiratory Questionnaire [SGRQ], and BODE index), pre- and post-bronchodilator lung function, and CT scan imaging. Genome wide genotyping was performed using the Illumina OmniExpress HumanExome BeadChip using standard techniques in the first 571 subjects with COPD and 175 controls. Imputation was performed against 1000 Genomes reference panels using Impute-v2.30 using a quality cutoff of 0.9, and association analysis performed using PLINK.

Studies with Custom Genotyping:

Boston Early-Onset COPD Study (BEOCOPD; ClinicalTrials.gov: NCT01177618): Details of the BECOPD study have been described previously[75,76]. IRB approval was obtained for the study and all participants provided written informed consent. BEOCOPD is an an extended pedigree study constructed based on probands under 53 years of age with severe COPD (defined as forced expiratory volume in one second (FEV_1) < 40% predicted) and without severe alpha-1 antitrypsin deficiency. Given the small number of non-white subjects in this cohort, only white subjects were

included in this analysis. Extended pedigrees from BEOCOPD, ICGN, TCGS-Poland, and TCGS-Korea were genotyped together on the Illumina HumanExome v1.2 array with custom content (see below). Subject quality control included kinship via KING[77] and kinship2[78], genotyping success rate, sex check, homozygosity, Mendelian errors. Markers were assessed for missingness, concordance between duplicate samples, and Hardy-Weinberg equilibrium. Assessment for population outliers and control for population stratification was performed using TRACE v1.0[79,80] and HapMap population [52] as the reference. Single-variant association analysis for pre-bronchodilator moderate-to-severe COPD, adjusting for age, pack-years, sex, and current and ever smoking, was performed in EOCOPD and ICGN together using a covariate additionally indicating study, via a logistic mixed model as implemented in GMMAT version 0.5 in R (version 3.2.0, http://www.R-project.org/)[81].

International COPD Genetics Network (ICGN): The ICGN has been described previously[82,83]. ICGN recruited subjects (FEV_1 < 60% predicted and FEV_1/FVC < 90% predicted between ages 45-65) as probands and then enrolled available siblings and parents of the proband. The study was IRB approved at all relevant institutions and participants provided written informed consent. Given the small number of non-white subjects in this cohort, only white subjects were included in this analysis. Genotyping and analysis was performed as detailed above (in the **BEOCOPD** section).

Transcontinental COPD Genetics Study (TCGS) – Korea and Poland: TCGS has been described previously[84] and was comprised of two case-control studies, based in Poland and in Korea. The study was approved by the appropriate IRBs and all participants provided written informed consent. Both studies recruited individuals between 40 and 80 years of age, with at least 10 pack-years of cigarette smoking; where cases had severe COPD (FEV1 < 50% predicted) and controls had normal spirometry. Subjects with other lung disease were excluded. TCGS-Poland enrolled white individuals, and TCGS-Korea enrolled Korean individuals. Details of genotyping are described

above (in the **BEOCOPD** section). We used PLINK2[47,48] to perform logistic regression, adjusting for age, pack-years, sex, and current smoking status.

UK BiLEVE Cohort Information: UK BiLEVE is a subset of UK Biobank and is described in detail elsewhere[85]. In brief, UK Biobank comprised 502,682 individuals of which 472,858 were of self-reported European ancestry and 275,939 had at least two measures of FEV_1 and FVC and passed ATS/ERS quality criteria[86]. Spirometry data was obtained using a Vitalograph Pneumotrac 6800 (Buckingham, UK); at least two measures were obtained. From these 275,939 individuals, 50,000 individuals were selected based on % predicted FEV_1 such that 10,000 individuals with low FEV_1, 10,000 individuals with near-average FEV_1, and 5,000 with high FEV_1 were selected from amongst never smokers (total n=105,272) and the same numbers from amongst the heavy smokers (total n=46,758). Equal numbers of males and females were selected and the number of individuals selected within each age-sex band was proportional to the number of individuals available for sampling for each band. Genotyping was undertaken using the Affymetrix Axiom UK BiLEVE array[85] and imputed to the 1000 Genomes Project[29] Phase 1 and UK10K[87] combined panel. Of the 50,008 samples selected, 48,943 unrelated individuals passed all genotype and sample quality control steps and were used as the sampling frame for selection of samples included in the analysis described here. The UK Biobank received ethics approval from the National Health Service National Research Ethics Service (Ref 11/NW/0382). All participants provided written informed consent.

In UK BiLEVE, COPD status was defined based on spirometry with individuals with % predicted FEV_1<80% and FEV_1/FVC<0.7 (indicative of COPD GOLD stage 2 or worse) selected as COPD cases. COPD controls were individuals that did not meet the case criteria and were selected from the high % predicted FEV_1 group only (all had % predicted FEV_1>80%). Analyses were carried out using the score test, implemented in SNPTEST v2.5b4 and assuming an additive genetic model of genotype dose. For never smokers, sex, age and the first 10 ancestry principal components were

included as covariates. For heavy smokers, current smoking status and pack years were included as additional covariates.

Additional Supplemental Methods

Fine Mapping Analysis Methods

We also attempted to determine, at each locus, whether we could identify a potentially causal variant. We performed these analyses using European ancestry subjects with genome-wide data, and excluded variants that were not present in at least 80% of the full sample. We used a +/- 250kb region around the top variant in each locus, and used two methods. We calculated approximate Bayes factors using the method of Wakefield[88], and determined the 95% credible set. While trans-ethnic mapping can significantly assist in fine-mapping, the number of non-European samples in our study was too low to leverage either PAINTOR2[89] or MANTRA[90].

Functional Enrichment Analysis Methods

To identify enriched cell types for our COPD associations, we applied LD score regression to GenoSkyline[91] lung tissue annotations (the default LD score regression annotations collapse lung into the cardiovascular tissue type), as well as cell-type specific annotations from LD score regression[92]. We also performed analysis using only the 22 genome-wide significant loci and tested for enrichment of imputed chromatin marks from ROADMAP[3]. Further, we applied a more sophisticated analysis adjusting for local linkage disequilibrium patterns, GoShifter[93], Finally, we examined overlap with gene expression datasets using SNPsea[94].

Quantitative Imaging Overlap Analysis Methods

To explore the relationship between our top COPD-associated variants and imaging features of emphysema and airway thickness, we queried data from a GWAS of COPD quantitative imaging features[27]. For each genome-wide significant COPD susceptibility locus in our overall meta-analysis, we assessed the corresponding quantitative imaging GWAS effect size, effect direction, and P value for association with the following quantitative imaging traits: %LAA-950 (percentage low attenuation area, using a threshold of -950 Hounsfield units); Perc15 (value of Hounsfield units at the 15th percentile of the density histogram); Pi10 (airway wall area: the value for a hypothetical

10-mm airway obtained by plotting a regression line of the square root of the airway wall area versus the airway internal perimeter); and WAP (percentage of the wall area compared to the total bronchial area).

Gene Set Enrichment Analysis Methods

As an attempt to minimize false positives in our gene set enrichment analysis, we divided the Stage 1 GWAS cohorts with full genome-wide data into two sets of roughly equal size. We then used i-GSEA4GWAS (http://gsea4gwas.psych.ac.cn/)[95] for each of the two GWAS data sets to assess enrichment of COPD GWAS loci in BioCarta (http://cgap.nci.nih.gov/Pathways/BioCarta_Pathways) and KEGG[96] pathways as well as gene ontology (GO) terms[97,98]. We first evaluated GO terms and pathways with a false-discovery rate (FDR) less than 5% in both analysis sets and then used a more stringent threshold of FDR < 1% to evaluate overlap of GO term and pathway enrichment in our two analysis sets.

Causal Gene Analysis Methods

For the genome-wide significant loci from the overall meta-analysis, we explored potential causative genes at each association locus using the PrixFixe method[99], assuming co-function of all significant loci. As required by the PrixFixe method, we assured our genome-wide significant loci were present in dbSNP v137[100] and were represented HapMap[52] phase III data; for loci not meeting these requirements, proxy SNPs from HapMap phase III were selected based on strongest LD (r^2) with index SNP (see Supplemental Table S15 for details of the proxy variant used at each genome-wide significant locus).

Supplemental Results:

Fine Mapping Analysis Results

We examined our top 22 GWAS loci from the overall meta-analysis to examine the size of the 95% credible set at each locus. We found that the association signals for these 22 loci were minimal in the African-Americans (our largest non-European ethnicity), likely due to inadequate power. At 12 loci with some evidence of association ($P < 0.001$) in individuals of African ancestry, we determined the linkage disequilibrium between top variants in African Americans (AA) and the top overall variants at these loci. Results of these fine mapping analyses are presented in Table S20. In the European only fine-mapping analysis, for 3 of the loci, the 95% credible set had < 10 variants, with an additional 2 with < 20. For two loci (*SFPTD* and *CFDP1*), we found variants in the African-Americans that were in modest LD (arbitrarily set at $r^2 > 0.3$ or $D' > 0.4$) that were in the 95% credible set. Overall, these results suggest some loci for which a reasonable number of variants can be tested, and others for which functional annotation, sample size, and studies of other ethnicities could be most useful.

Lung eQTL Analysis Results

Genotypes and gene expression was available for 409 patients from Laval, 287 patients from the UBC, and 342 patients from Groningen. One hundred ten eQTL at 20 loci were nominally ($P < 0.05$) significant in the meta-analysis; 14 eQTL (P value < 5.2×10^{-8}) involving 6 loci reached significance after Bonferroni correction. Supplemental Table S6 shows the most significant gene-SNP pairs at each locus. rs17486278 on chromosome 15 was also significantly associated with CHRNA5 and PSMA4 as previously reported[101].

Functional Enrichment Analysis Results

The LD score regression functional enrichment analysis identified lung as the primary tissue (coefficient Z-score, 2.54), and fetal lung H3K4me1 as the primary cell type (Z-score, 4.4). No inflammatory cell types were nominally significant (Z all < 1.75). In the ROADMAP[3] chromatin

marks analysis, no individual cell was significant after correction for multiple testing, though we identified nominal significance (P value, using GWAS background < 0.05) for A549 lung cancer, NHLF lung fibroblast, and primary helper and effector T cells from peripheral blood ($0.01 < P < 0.05$); top ranked associations were fetal intestine, thymus, and rectal mucosa. Using GoShifter[93] we identified nominal significance for fetal lung and IMR90 lung fibroblasts ($P = 0.006$ and 0.049 respectively); there was no evidence for immune cell enrichment.

With the SNPsea[94] analysis, lung related cells ranked in the top annotations in both the Affymetrix (Lung, Trachea, Bronchial epithelial cells with unadjusted $0.02 < P < 0.03$) and FANTOM5 (lung, mesothelioma, and alveolar cell carcinoma $0.001 < P < 0.03$). In the Immgen dataset, our top enrichment was for lung dendritic cells ($P = 0.002$). No enrichment was seen for neutrophils, T-cells, and lung macrophages (P all > 0.09).

In summary, multiple analyses identified enrichment of lung cell types; in contrast, we failed to find convincing evidence of enrichment for other cell types, including inflammatory cells. While this lack of enrichment could reflect a limited influence of genetically driven inflammatory factors contributing to COPD, this could also represent limitations in the power and assessment appropriate cell types in these datasets.

Quantitative Imaging Overlap Analysis Results

We evaluated our 22 COPD susceptibility loci in a GWAS of COPD-related quantitative imaging phenotypes[27] to assess differential effects of our COPD susceptibility loci on emphysema and airway imaging phenotypes (Supplemental Table S14).

Gene Set Enrichment Analysis Results

To assess enrichment of our Stage 1 association analysis results in biologic pathways, we divided the Stage 1 analysis cohorts into discovery and validation sets of roughly equal size and analyzed each set using i-GSEA4GWAS (http://gsea4gwas.psych.ac.cn/)[95]. We identified 32 GO terms and canonical pathways with significant enrichment (false discovery rate (FDR) < 5%) in

both sets. These enriched pathways related to cell homeostasis, cell differentiation, lymphoid organ and immune system development, biopeptide signaling and transcription factor complex activity. Using a more stringent FDR threshold of <1%, five pathways and GO terms overlapped between the discovery and validation analysis sets, including the T cell receptor pathway, sugar metabolism, ion and substrate channel activities, and response to organic substance stimulus.

Causal Gene Analysis Results

We used the PrixFixe method[99] to rank potential causal genes based on mutual function across all COPD genome-wide loci (Supplemental Table S16). In regions such as chr15q25.1, chr 16p11.2, chr16q23.1, and chr19q13.2 with multiple genes in an LD window with our top COPD variant, the PrixFixe method did not rank the closest gene as the most likely causative gene based on mutual function of genes at all top COPD loci.

Funding:

This work was supported by NHLBI R01 HL084323, R01 HL113264, R01 HL089856, and P01 HL105339 (E.K.S.); K08 HL097029 and R01 HL113264 (M.H.C.), R01 HL089897 (J.D.C.) and P01 HL114501 (Augustine Choi); the Alpha-1 Foundation (M.H.C.) and a VA Research Career Scientist award (D.S.).

The content of this study is solely the responsibility of the authors and does not necessarily represent the official views of the National Heart, Lung, And Blood Institute or the National Institutes of Health.

The **Atherosclerosis Risk in Communities Study** is carried out as a collaborative study supported by National Heart, Lung, and Blood Institute contracts (HHSN268201100005C, HHSN268201100006C, HHSN268201100007C, HHSN268201100008C, HHSN268201100009C, HHSN268201100010C, HHSN268201100011C, and HHSN268201100012C), R01HL087641, R01HL59367 and R01HL086694; National Human Genome Research Institute contract U01HG004402; and National Institutes of Health contract HHSN268200625226C. Infrastructure was partly supported by Grant Number UL1RR025005, a component of the National Institutes of Health and NIH Roadmap for Medical Research. Nora Franceschini is supported by R21HL123677-01. This work was also supported in part by the Intramural Research Program of the NIH, National Institute of Environmental Health Sciences.

We acknowledge use of phenotype and genotype data from the **British 1958 Birth Cohort** DNA collection, funded by the Medical Research Council grant G0000934 and the Wellcome Trust grant 068545/Z/02. Genotyping for the B58C-WTCCC subset was funded by the Wellcome Trust grant 076113/B/04/Z. The B58C-T1DGC genotyping utilized resources provided by the Type 1 Diabetes Genetics Consortium, a collaborative clinical study sponsored by the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), National Institute of Allergy and Infectious Diseases (NIAID), National Human Genome Research Institute (NHGRI), National Institute of Child Health and Human Development (NICHD), and Juvenile Diabetes Research Foundation International (JDRF) and supported by U01 DK062418. B58C-T1DGC GWAS data were deposited by the Diabetes and Inflammation Laboratory, Cambridge Institute for Medical Research (CIMR), University of Cambridge, which is funded by Juvenile Diabetes Research Foundation International, the Wellcome Trust and the National Institute for Health Research Cambridge Biomedical Research Centre; the CIMR is in receipt of a Wellcome Trust Strategic Award (079895). The B58C-GABRIEL genotyping was supported by a contract from the European Commission Framework Programme 6 (018996) and grants from the French Ministry of Research.

This **CHS** research was supported by NHLBI contracts HHSN268201200036C, HHSN268200800007C, HHSN268200960009C, N01HC55222, N01HC85079, N01HC85080, N01HC85081, N01HC85082, N01HC85083, N01HC85086; and NHLBI grants U01HL080295, R01HL087652, R01HL105756, R01HL103612, R01HL085251, and R01HL120393 with additional contribution from the National Institute of Neurological Disorders and Stroke (NINDS). Additional support was provided through R01AG023629 from the National Institute on Aging (NIA). A full list of principal CHS investigators and institutions can be found at CHS-NHLBI.org.
The provision of genotyping data was supported in part by the National Center for Advancing Translational Sciences, CTSI grant UL1TR000124, and the National Institute of Diabetes and Digestive and Kidney Disease Diabetes Research Center (DRC) grant DK063491 to the Southern California Diabetes Endocrinology Research Center.

The **COPACETIC** study was supported by a European Union FP7 grant (201379, COPACETIC). NELSON was funded by 'Zorg Onderzoek Nederland-Medische Wetenschappen, KWF Kankerbestrijding, Stichting Centraal Fonds Reserves van Voormalig Vrijwillige Ziekenfondsverzekeringen, Siemens Germany, G. Ph. Verhagen Stichting, Rotterdam Oncologic Thoracic Steering Committee, Erasmus Trust Fund, Stichting tegen Kanker. Kim de Jong is supported by grant number 4.113.007 the Lung Foundation Netherlands.

The **COPDGene** project (NCT00608764) was supported by Award Number R01HL089897 and Award Number R01HL089856 from the National Heart, Lung, And Blood Institute. The COPDGene project is also supported by the COPD Foundation through contributions made to an Industry Advisory Board comprised of AstraZeneca, Boehringer Ingelheim, Novartis, Pfizer, Siemens, Sunovion, and GlaxoSmithKline.

The **ECLIPSE** study (NCT00292552; GSK code SCO104960) was funded by GSK.

This work was partially supported by the National Heart, Lung and Blood Institute's **Framingham Heart Study** (contract number N01-HC-25195) and its contract with Affymetrix, Inc for genotyping services (contract number N02-HL-6-4278). Also supported by NIH P01 AI050516.

KARE was funded by the Consortium for Large Scale Genome Wide Association Study III (2011E7300400), which was supported by the genotyping data (the Korean Genome Analysis Project, 4845-301) and the phenotype data (the Korean Genome Epidemiology Study, 4851-302) and a grant from the Centers for Disease Control and Prevention of Korea (budget 2012-E71011-00, 2011-E71011-00 and 2010-E71010-00) and also supported by the National Project for Personalized Genomic Medicine (A111218-11-GM02 and A111218-12-GM10).

The **Lung eQTL** study at Laval University was supported by the Chaire de pneumologie de la Fondation JD Bégin de l'Université Laval, the Fondation de l'Institut universitaire de cardiologie et de pneumologie de Québec, the Respiratory Health Network of the FRQS, the Canadian Institutes of Health Research (MOP - 123369), and the Cancer Research Society and Read for the Cure. Y.B. holds a Canada Research Chair in Genomics of Heart and Lung Diseases.

The **Norway GenKOLS** study (Genetics of Chronic Obstructive Lung Disease, GSK code RES11080) was funded by GSK.

The **ICGN** study was funded by GSK.

The **LifeLines** cohort study was supported by the Dutch Ministry of Health, Welfare and Sport, the Ministry of Economic Affairs, Agriculture and Innovation, the province of Groningen, the European Union (regional development fund), the Northern Netherlands Provinces (SNN), the Netherlands Organisation for Scientific Research (NWO), University Medical Center Groningen (UMCG), University of Groningen, de Nierstichting (the Dutch Kidney Foundation), and the Diabetes Fonds (the Diabetic Foundation).

The **Lovelace** cohort and analysis was primarily supported by National Cancer Institute grant R01 CA097356 (SAB). The State of New Mexico as a direct appropriation from the Tobacco Settlement Fund to SAB. through collaboration with University of New Mexico provided initial support to establish the LSC. Additional support was provided by NIH/NCI P30 CA118100 (SAB), HL68111 (Y.T.), and HL107873-01 (YT and SB).

MESA and the MESA SHARe project are conducted and supported by the National Heart, Lung, and Blood Institute (NHLBI) in collaboration with MESA investigators. Support for MESA is provided by contracts N01-HC-95159, N01-HC-95160, N01-HC-95161, N01-HC-95162, N01-HC-95163, N01-HC-95164, N01-HC-95165, N01-HC-95166, N01-HC-95167, N01-HC-95168, N01-HC-95169, UL1-TR-001079, UL1-TR-000040, and DK063491. MESA Family is conducted and supported by the National Heart, Lung, and Blood Institute (NHLBI) in collaboration with MESA investigators. Support is provided by grants and contracts R01HL071051, R01HL071205, R01HL071250, R01HL071251, R01HL071258, and R01HL071259 by the National Center for Research Resources, Grant UL1RR033176, and the National Center for Advancing Translational Sciences, Grant UL1TR000124. The MESA Lung study was supported by grants R01 HL077612, R01 HL093081 and RC1 HL100543 from the NHLBI. This publication was developed under a STAR research assistance agreement, No. RD831697 (MESA Air), awarded by the U.S Environmental protection Agency. It has not been formally reviewed by the EPA. The views expressed in this document are solely those of the authors and the EPA does not endorse any products or commercial services mentioned in this publication. Funding for SHARe genotyping was provided by NHLBI Contract N02-HL-64278. Genotyping was performed at Affymetrix (Santa Clara, California, USA) and the Broad Institute of Harvard and MIT (Boston, Massachusetts, USA) using the Affymetrix Genome-Wide Human SNP Array 6.0.

The **National Emphysema Treatment Trial** was supported by the NHLBI N01HR76101, N01HR76102, N01HR76103, N01HR76104, N01HR76105, N01HR76106, N01HR76107, N01HR76108, N01HR76109, N01HR76110, N01HR76111, N01HR76112, N01HR76113, N01HR76114, N01HR76115, N01HR76116, N01HR76118 and N01HR76119, the Centers for Medicare and Medicaid Services and the Agency for Healthcare Research and Quality. The **Normative Aging Study** is supported by the Cooperative Studies Program/ERIC of the US Department of Veterans Affairs and is a component of the Massachusetts Veterans Epidemiology Research and Information Center (MAVERIC). D.S. is supported by a VA Research Career Scientist award.

The **Rotterdam Study** is funded by Erasmus Medical Center and Erasmus University, Rotterdam, Netherlands Organization for the Health Research and Development (ZonMw), the Research Institute for Diseases in the Elderly (RIDE), the Ministry of Education, Culture and Science, the Ministry for Health, Welfare and Sports, the European Commission (DG XII), and the Municipality of Rotterdam.

The generation and management of GWAS genotype data for the **Rotterdam Study** (RS I, RS II, RS III) was executed by the Human Genotyping Facility of the Genetic Laboratory of the Department of Internal Medicine, Erasmus MC, Rotterdam, The Netherlands. The GWAS datasets are supported by the Netherlands Organisation of Scientific Research NWO Investments (nr. 175.010.2005.011, 911-03-012), the Genetic Laboratory of the Department of Internal Medicine, Erasmus MC, the Research Institute for Diseases in the Elderly (014-93-015; RIDE2), the Netherlands Genomics Initiative (NGI)/Netherlands Organisation for Scientific Research (NWO) Netherlands Consortium for Healthy Aging (NCHA), project nr. 050-060-810. The generation and management of spirometric data was supported by FWO project G035014N. Lies Lahousse is a Postdoctoral Fellow of the Fund for Scientific Research Foundation - Flanders (FWO).

The **UK BiLEVE** study was funded by a Medical Research Council (MRC) strategic award to M.D.T., I.P.H., D.P.S. and L.V.W. (MC_PC_12010). The research undertaken by M.D.T., M.S.A., L.V.W. and N.S. was partly funded by the National Institute for Health Research (NIHR). The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health.

M.D.T. holds a Medical Research Council Senior Clinical Fellowship (G0902313). This research used the ALICE High Performance Computing Facility at the University of Leicester. The Universities of Leicester and Nottingham acknowledge receipt of a Collaborative Research and Development grant from the Healthcare and Bioscience iNet, a project funded by the East Midlands Development Agency, part-financed by the European Regional Development Fund and delivered by Medilink East Midlands. I.P.H. holds a Medical Research Council programme grant (G1000861).

This research has been conducted using the UK Biobank Resource. The research undertaken by M.D.T., M.S.A., L.V.W. and N.S. was partly funded by the National Institute for Health Research (NIHR). The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health.

Acknowledgments:

The authors thank the staff and participants of the **ARIC** study for their important contributions. The **ARIC** study authors thank Dr. Frank Shuangshuang Dai of NIEHS and Dr. Jianping Jin of Westat, Inc (Research Triangle Park, NC) for expert computing assistance.

The authors would like to thank the staff at the Respiratory Health Network Tissue Bank of the FRQS for their valuable assistance with the **Lung eQTL** dataset at Laval University and the biobank staff and donors who contributed to the UBC lung tissue biobank.

The **MESA** authors thank the participants of the MESA study, the Coordinating Center, MESA investigators, and study staff for their valuable contributions. A full list of participating MESA investigators and institutions can be found at http://www.mesa-nhlbi.org.

The **Rotterdam Study** authors are grateful to the study participants, the staff from the Rotterdam Study and the participating general practitioners and pharmacists. We thank Pascal Arp, Mila Jhamai, Marijn Verkerk, Lizbeth Herrera and Marjolein Peters, MSc, and Carolina Medina-Gomez, MSc, for their help in creating the GWAS database, and Karol Estrada, PhD, Yurii Aulchenko, PhD, and Carolina Medina-Gomez, MSc, for the creation and analysis of imputed data.

COPDGene Investigators – Core Units

Administrative Core: James Crapo, MD (PI), Edwin Silverman, MD, PhD (PI), Barry Make, MD, Elizabeth Regan, MD, PhD

Genetic Analysis Core: Terri Beaty, PhD, Nan Laird, PhD, Christoph Lange, PhD, Michael Cho, MD, Stephanie Santorico, PhD, John Hokanson, MPH, PhD, Dawn DeMeo, MD, MPH, Nadia Hansel, MD, MPH, Craig Hersh, MD, MPH, Peter Castaldi, MD, MSc, Merry-Lynn McDonald, PhD, Emily Wan, MD, Megan Hardin, MD, Jacqueline Hetmanski, MS, Margaret Parker, PhD, Marilyn Foreman, MD, Brian Hobbs, MD, MMSc, Robert Busch, MD, MMSc, Adel El-Boueiz, MD, MMSc, Dandi Qiao, PhD, Elizabeth Regan, MD, Eitan Halper-Stromberg, Ferdouse Begum, Sungho Won, Sharon Lutz, PhD

Imaging Core: David A Lynch, MB, Harvey O Coxson, PhD, MeiLan K Han, MD, MS, MD, Eric A Hoffman, PhD, Stephen Humphries MS, Francine L Jacobson, MD, Philip F Judy, PhD, Ella A Kazerooni, MD, John D Newell, Jr., MD, Elizabeth Regan, MD, James C Ross, PhD, Raul San Jose Estepar, PhD, Berend C Stoel, PhD, Juerg Tschirren, PhD, Eva van Rikxoort, PhD, Bram van Ginneken, PhD, George Washko, MD, Carla G Wilson, MS, Mustafa Al Qaisi, MD, Teresa Gray, Alex Kluiber, Tanya Mann, Jered Sieren, Douglas Stinson, Joyce Schroeder, MD, Edwin Van Beek, MD, PhD

PFT QA Core, Salt Lake City, UT: Robert Jensen, PhD

Data Coordinating Center and Biostatistics, *National Jewish Health, Denver, CO*: Douglas Everett, PhD, Anna Faino, MS, Matt Strand, PhD, Carla Wilson, MS

Epidemiology Core, *University of Colorado Anschutz Medical Campus, Aurora, CO*: John E. Hokanson, MPH, PhD, Gregory Kinney, MPH, PhD, Sharon Lutz, PhD, Kendra Young PhD, Katherine Pratte, MSPH, Lindsey Duca, MS

COPDGene Investigators – Clinical Centers

Ann Arbor VA: Jeffrey L. Curtis, MD, Carlos H. Martinez, MD, MPH, Perry G. Pernicano, MD

Baylor College of Medicine, Houston, TX: Nicola Hanania, MD, MS, Philip Alapat, MD, Venkata Bandi, MD, Mustafa Atik, MD, Aladin Boriek, PhD, Kalpatha Guntupalli, MD, Elizabeth Guy, MD, Amit Parulekar, MD, Arun Nachiappan, MD

Brigham and Women's Hospital, Boston, MA: Dawn DeMeo, MD, MPH, Craig Hersh, MD, MPH, George Washko, MD, Francine Jacobson, MD, MPH

Columbia University, New York, NY: R. Graham Barr, MD, DrPH, Byron Thomashow, MD, John Austin, MD, Belinda D'Souza, MD, Gregory D.N. Pearson, MD, Anna Rozenshtein, MD, MPH, FACR

Duke University Medical Center, Durham, NC: Neil MacIntyre, Jr., MD, Lacey Washington, MD, H. Page McAdams, MD

Health Partners Research Foundation, Minneapolis, MN: Charlene McEvoy, MD, MPH, Joseph Tashjian, MD

Johns Hopkins University, Baltimore, MD: Robert Wise, MD, Nadia Hansel, MD, MPH, Robert Brown, MD, Karen Horton, MD, Nirupama Putcha, MD, MHS,

Los Angeles Biomedical Research Institute at Harbor UCLA Medical Center, Torrance, CA: Richard Casaburi, PhD, MD, Alessandra Adami, PhD, Janos Porszasz, MD, PhD, Hans Fischer, MD, PhD, Matthew Budoff, MD, Harry Rossiter, PhD

Michael E. DeBakey VAMC, Houston, TX: Amir Sharafkhaneh, MD, PhD, Charlie Lan, DO

Minneapolis VA: Christine Wendt, MD, Brian Bell, MD

Morehouse School of Medicine, Atlanta, GA: Marilyn Foreman, MD, MS, Gloria Westney, MD, MS, Eugene Berkowitz, MD, PhD

National Jewish Health, Denver, CO: Russell Bowler, MD, PhD, David Lynch, MD

Reliant Medical Group, Worcester, MA: Richard Rosiello, MD, David Pace, MD

Temple University, Philadelphia, PA: Gerard Criner, MD, David Ciccolella, MD, Francis Cordova, MD, Chandra Dass, MD, Gilbert D'Alonzo, DO, Parag Desai, MD, Michael Jacobs, PharmD, Steven Kelsen, MD, PhD, Victor Kim, MD, A. James Mamary, MD, Nathaniel Marchetti, DO, Aditi Satti, MD, Kartik Shenoy, MD, Robert M. Steiner, MD, Alex Swift, MD, Irene Swift, MD, Maria Elena Vega-Sanchez, MD

University of Alabama, Birmingham, AL: Mark Dransfield, MD, William Bailey, MD, J. Michael Wells, MD, Surya Bhatt, MD, Hrudaya Nath, MD

University of California, San Diego, CA: Joe Ramsdell, MD, Paul Friedman, MD, Xavier Soler, MD, PhD, Andrew Yen, MD

University of Iowa, Iowa City, IA: Alejandro Comellas, MD, John Newell, Jr., MD, Brad Thompson, MD

University of Michigan, Ann Arbor, MI: MeiLan Han, MD, Ella Kazerooni, MD, Carlos Martinez, MD

University of Minnesota, Minneapolis, MN: Joanne Billings, MD, Tadashi Allen, MD

University of Pittsburgh, Pittsburgh, PA: Frank Sciurba, MD, Divay Chandra, MD, MSc, Joel Weissfeld, MD, MPH, Carl Fuhrman, MD, Jessica Bon, MD

University of Texas Health Science Center at San Antonio, San Antonio, TX: Antonio Anzueto, MD, Sandra Adams, MD, Diego Maselli-Caceres, MD, Mario E. Ruiz, MD

ECLIPSE Investigators — *Bulgaria:* Y. Ivanov, Pleven; K. Kostov, Sofia. *Canada:* J. Bourbeau, Montreal; M. Fitzgerald, Vancouver, BC; P. Hernandez, Halifax, NS; K. Killian, Hamilton, ON; R. Levy, Vancouver, BC; F. Maltais, Montreal; D. O'Donnell, Kingston, ON. *Czech Republic:* J. Krepelka, Prague. *Denmark:* J. Vestbo, Hvidovre. *The Netherlands:* E. Wouters, Horn-Maastricht. *New Zealand:* D. Quinn, Wellington. *Norway:* P. Bakke, Bergen. *Slovenia:* M. Kosnik, Golnik. *Spain:* A. Agusti, J. Sauleda, P. de Mallorca. *Ukraine:* Y. Feschenko, V. Gavrisyuk, L. Yashina, Kiev; N. Monogarova, Donetsk. *United Kingdom:* P. Calverley, Liverpool; D. Lomas, Cambridge; W. MacNee, Edinburgh; D. Singh, Manchester; J. Wedzicha, London. *United States:* A. Anzueto, San Antonio, TX; S. Braman, Providence, RI; R. Casaburi, Torrance CA; B. Celli, Boston; G. Giessel, Richmond, VA; M. Gotfried, Phoenix, AZ; G. Greenwald, Rancho Mirage, CA; N. Hanania, Houston; D. Mahler, Lebanon, NH; B. Make, Denver; S. Rennard, Omaha, NE; C. Rochester, New Haven, CT; P. Scanlon, Rochester, MN; D. Schuller, Omaha, NE; F. Sciurba, Pittsburgh; A. Sharafkhaneh, Houston; T. Siler, St. Charles, MO; E. Silverman, Boston; A. Wanner, Miami; R. Wise, Baltimore; R. ZuWallack, Hartford, CT.
ECLIPSE Steering Committee: H. Coxson (Canada), C. Crim (GlaxoSmithKline, USA), L. Edwards (GlaxoSmithKline, USA), D. Lomas (UK), W. MacNee (UK), E. Silverman (USA), R. Tal Singer (Co-chair, GlaxoSmithKline, USA), J. Vestbo (Co-chair, Denmark), J. Yates (GlaxoSmithKline, USA).
ECLIPSE Scientific Committee: A. Agusti (Spain), P. Calverley (UK), B. Celli (USA), C. Crim (GlaxoSmithKline, USA), B. Miller (GlaxoSmithKline, USA), W. MacNee (Chair, UK), S. Rennard (USA), R. Tal-Singer (GlaxoSmithKline, USA), E. Wouters (The Netherlands), J. Yates (GlaxoSmithKline, USA).

The authors thank the **SPIROMICS** participants and participating physicians, investigators and staff for making this research possible. More information about the study and how to access SPIROMICS data is at www.spiromics.org. We would like to acknowledge the following current and former investigators of the SPIROMICS sites and reading centers: Neil E Alexis, PhD; Wayne H Anderson, PhD; R Graham Barr, MD, DrPH; Eugene R Bleecker, MD; Richard C Boucher, MD; Russell P Bowler, MD, PhD; Elizabeth E Carretta, MPH; Stephanie A Christenson, MD; Alejandro P Comellas, MD; Christopher B Cooper, MD, PhD; David J Couper, PhD; Gerard J Criner, MD; Ronald G Crystal, MD; Jeffrey L Curtis, MD; Claire M Doerschuk, MD; Mark T Dransfield, MD; Christine M Freeman, PhD; MeiLan K Han, MD, MS; Nadia N Hansel, MD, MPH; Annette T Hastie, PhD; Eric A Hoffman, PhD; Robert J Kaner, MD; Richard E Kanner, MD; Eric C Kleerup, MD; Jerry A Krishnan, MD, PhD; Lisa M LaVange, PhD; Stephen C Lazarus, MD; Fernando J Martinez, MD, MS; Deborah A Meyers, PhD; John D Newell Jr, MD; Elizabeth C Oelsner, MD, MPH; Wanda K O'Neal, PhD; Robert Paine, III, MD; Nirupama Putcha, MD, MHS; Stephen I. Rennard, MD; Donald P Tashkin, MD; Mary Beth Scholand, MD; J Michael Wells, MD; Robert A Wise, MD; and Prescott G Woodruff, MD, MPH. The project officers from the Lung Division of the National Heart, Lung, and Blood Institute were Lisa Postow, PhD, and Thomas Croxton, PhD, MD. SPIROMICS was supported by contracts from the NIH/NHLBI

(HHSN268200900013C, HHSN268200900014C, HHSN268200900015C, HHSN268200900016C, HHSN268200900017C, HHSN268200900018C HHSN268200900019C, HHSN268200900020C), which were supplemented by contributions made through the Foundation for the NIH from AstraZeneca; Bellerophon Therapeutics; Boehringer-Ingelheim Pharmaceuticals, Inc; Chiesi Farmaceutici SpA; Forest Research Institute, Inc; GSK; Grifols Therapeutics, Inc; Ikaria, Inc; Nycomed GmbH; Takeda Pharmaceutical Company; Novartis Pharmaceuticals Corporation; Regeneron Pharmaceuticals, Inc; and Sanofi.

The authors thank the SPIROMICS participants and participating physicians, investigators and staff for making this research possible. More information about the study and how to access SPIROMICS data is at www.spiromics.org. We would like to acknowledge the following current and former investigators of the SPIROMICS sites and reading centers: Neil E Alexis, PhD; Wayne H Anderson, PhD; R Graham Barr, MD, DrPH; Eugene R Bleecker, MD; Richard C Boucher, MD; Russell P Bowler, MD, PhD; Elizabeth E Carretta, MPH; Stephanie A Christenson, MD; Alejandro P Comellas, MD; Christopher B Cooper, MD, PhD; David J Couper, PhD; Gerard J Criner, MD; Ronald G Crystal, MD; Jeffrey L Curtis, MD; Claire M Doerschuk, MD; Mark T Dransfield, MD; Christine M Freeman, PhD; MeiLan K Han, MD, MS; Nadia N Hansel, MD, MPH; Annette T Hastie, PhD; Eric A Hoffman, PhD; Robert J Kaner, MD; Richard E Kanner, MD; Eric C Kleerup, MD; Jerry A Krishnan, MD, PhD; Lisa M LaVange, PhD; Stephen C Lazarus, MD; Fernando J Martinez, MD, MS; Deborah A Meyers, PhD; John D Newell Jr, MD; Elizabeth C Oelsner, MD, MPH; Wanda K O'Neal, PhD; Robert Paine, III, MD; Nirupama Putcha, MD, MHS; Stephen I. Rennard, MD; Donald P Tashkin, MD; Mary Beth Scholand, MD; J Michael Wells, MD; Robert A Wise, MD; and Prescott G Woodruff, MD, MPH. The project officers from the Lung Division of the National Heart, Lung, and Blood Institute were Lisa Postow, PhD, and Thomas Croxton, PhD, MD. SPIROMICS was supported by contracts from the NIH/NHLBI (HHSN268200900013C, HHSN268200900014C, HHSN268200900015C, HHSN268200900016C, HHSN268200900017C, HHSN268200900018C HHSN268200900019C, HHSN268200900020C), which were supplemented by contributions made through the Foundation for the NIH from AstraZeneca; Bellerophon Pharmaceuticals; Boehringer-Ingelheim Pharmaceuticals, Inc; Chiesi Farmaceutici SpA; Forest Research Institute, Inc; GSK; Grifols Therapeutics, Inc; Ikaria, Inc; Nycomed GmbH; Takeda Pharmaceutical Company; Novartis Pharmaceuticals Corporation; Regeneron Pharmaceuticals, Inc; and Sanofi.
The authors additionally thank Weiming Zhang for helping provide summary statistics for the pulmonary fibrosis analysis, and Ryan Welch at the University of Michigan for allowing us the use of pre-release, stand-alone LocusZoom software.

Supplement References:

1. Pruim, R.J. *et al.* LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics* **26**, 2336-7 (2010).
2. Johnson, A.D. *et al.* SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap. *Bioinformatics* **24**, 2938-9 (2008).
3. Ward, L.D. & Kellis, M. HaploReg v4: systematic mining of putative causal variants, cell types, regulators and target genes for human complex traits and disease. *Nucleic Acids Res* (2015).
4. Ward, L.D. & Kellis, M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res* **40**, D930-4 (2012).
5. UniProt, C. UniProt: a hub for protein information. *Nucleic Acids Res* **43**, D204-12 (2015).
6. Wiertz, E.J. *et al.* Sec61-mediated transfer of a membrane protein from the endoplasmic reticulum to the proteasome for destruction. *Nature* **384**, 432-8 (1996).
7. Perry, J.R. *et al.* Parent-of-origin-specific allelic associations among 106 genomic loci for age at menarche. *Nature* **514**, 92-7 (2014).
8. Geller, F. *et al.* Genome-wide association analyses identify variants in developmental genes associated with hypospadias. *Nat Genet* **46**, 957-63 (2014).
9. Gudmundsson, J. *et al.* Genome-wide association and replication studies identify four variants associated with prostate cancer susceptibility. *Nat Genet* **41**, 1122-6 (2009).
10. Vasioukhin, V., Bowers, E., Bauer, C., Degenstein, L. & Fuchs, E. Desmoplakin is essential in epidermal sheet formation. *Nat Cell Biol* **3**, 1076-85 (2001).
11. Armstrong, D.K. *et al.* Haploinsufficiency of desmoplakin causes a striate subtype of palmoplantar keratoderma. *Hum Mol Genet* **8**, 143-8 (1999).
12. Norgett, E.E. *et al.* Recessive mutation in desmoplakin disrupts desmoplakin-intermediate filament interactions and causes dilated cardiomyopathy, woolly hair and keratoderma. *Hum Mol Genet* **9**, 2761-6 (2000).
13. den Haan, A.D. *et al.* Comprehensive desmosome mutation analysis in north americans with arrhythmogenic right ventricular dysplasia/cardiomyopathy. *Circ Cardiovasc Genet* **2**, 428-35 (2009).
14. Jonkman, M.F. *et al.* Loss of desmoplakin tail causes lethal acantholytic epidermolysis bullosa. *Am J Hum Genet* **77**, 653-60 (2005).
15. Fingerlin, T.E. *et al.* Genome-wide association study identifies multiple susceptibility loci for pulmonary fibrosis. *Nat Genet* **45**, 613-20 (2013).
16. Mathai, S.K. *et al.* Desmoplakin (DSP) Variants are Associated with Idiopathic Pulmonary Fibrosis. *American Journal of Respiratory and Critical Care Medicine* (2015).
17. Sato, Y. *et al.* The novel PAR-1-binding protein MTCL1 has crucial roles in organizing microtubules in polarizing epithelial cells. *J Cell Sci* **126**, 4671-83 (2013).

18. Sato, Y. *et al.* MTCL1 crosslinks and stabilizes non-centrosomal microtubules on the Golgi membrane. *Nat Commun* **5**, 5266 (2014).
19. Wert, S.E. *et al.* Increased metalloproteinase activity, oxidant production, and emphysema in surfactant protein D gene-inactivated mice. *Proc Natl Acad Sci U S A* **97**, 5972-7 (2000).
20. Foreman, M.G. *et al.* Polymorphisms in surfactant protein-D are associated with chronic obstructive pulmonary disease. *Am J Respir Cell Mol Biol* **44**, 316-22 (2011).
21. Lomas, D.A. *et al.* Serum surfactant protein D is steroid sensitive and associated with exacerbations of COPD. *Eur Respir J* **34**, 95-102 (2009).
22. Kim, D.K. *et al.* Genome-wide association analysis of blood biomarkers in chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* **186**, 1238-47 (2012).
23. Hao, K. *et al.* Lung eQTLs to help reveal the molecular underpinnings of asthma. *PLoS Genet* **8**, e1003029 (2012).
24. Consortium, G.T. The Genotype-Tissue Expression (GTEx) project. *Nat Genet* **45**, 580-5 (2013).
25. Moffatt, M.F. *et al.* A large-scale, consortium-based genomewide association study of asthma. *N Engl J Med* **363**, 1211-21 (2010).
26. Consortium, T.a.G. Genome-wide meta-analyses identify multiple loci associated with smoking behavior. *Nat Genet* **42**, 441-7 (2010).
27. Cho, M.H. *et al.* A Genome-Wide Association Study of Emphysema and Airway Quantitative Imaging Phenotypes. *Am J Respir Crit Care Med* **192**, 559-69 (2015).
28. The Atherosclerosis Risk in Communities (ARIC) Study: design and objectives. The ARIC investigators. *Am J Epidemiol* **129**, 687-702 (1989).
29. Genomes Project, C. *et al.* An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56-65 (2012).
30. Li, Y., Willer, C., Sanna, S. & Abecasis, G. Genotype imputation. *Annu Rev Genomics Hum Genet* **10**, 387-406 (2009).
31. Li, Y., Willer, C.J., Ding, J., Scheet, P. & Abecasis, G.R. MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genet Epidemiol* **34**, 816-34 (2010).
32. Aulchenko, Y.S., Struchalin, M.V. & van Duijn, C.M. ProbABEL package for genome-wide association analysis of imputed data. *BMC Bioinformatics* **11**, 134 (2010).
33. Strachan, D.P. *et al.* Lifecourse influences on health among British adults: effects of region of residence in childhood and adulthood. *Int J Epidemiol* **36**, 522-31 (2007).
34. Marossy, A.E., Strachan, D.P., Rudnicka, A.R. & Anderson, H.R. Childhood chest illness and the rate of decline of adult lung function between ages 35 and 45 years. *Am J Respir Crit Care Med* **175**, 355-9 (2007).
35. Wellcome Trust Case Control, C. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* **447**, 661-78 (2007).
36. Barrett, J.C. *et al.* Genome-wide association study and meta-analysis find that over 40 loci affect risk of type 1 diabetes. *Nat Genet* **41**, 703-7 (2009).
37. Howie, B., Fuchsberger, C., Stephens, M., Marchini, J. & Abecasis, G.R. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nat Genet* **44**, 955-9 (2012).
38. Fried, L.P. *et al.* The Cardiovascular Health Study: design and rationale. *Ann Epidemiol* **1**, 263-76 (1991).

39. Mohamed Hoesein, F.A. *et al.* Variants in the 15q24/25 locus associate with lung function decline in active smokers. *PLoS One* **8**, e53219 (2013).
40. van den Bergh, K.A. *et al.* Short-term health-related quality of life consequences in a lung cancer CT screening trial (NELSON). *Br J Cancer* **102**, 27-34 (2010).
41. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* **81**, 559-75 (2007).
42. Delaneau, O., Zagury, J.F. & Marchini, J. Improved whole-chromosome phasing for disease and population genetic studies. *Nat Methods* **10**, 5-6 (2013).
43. Howie, B.N., Donnelly, P. & Marchini, J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet* **5**, e1000529 (2009).
44. Purcell, S. PLINK [version 1.07]. URL: http://pngu.mgh.harvard.edu/purcell/plink/. Vol. http://pngu.mgh.harvard.edu/purcell/plink/.
45. Regan, E.A. *et al.* Genetic epidemiology of COPD (COPDGene) study design. *COPD* **7**, 32-43 (2010).
46. Cho, M.H. *et al.* Risk loci for chronic obstructive pulmonary disease: a genome-wide association study and meta-analysis. *Lancet Respir Med* **2**, 214-25 (2014).
47. Purcell, S. & Chang, C. PLINK [version 1.9]. URL: https://www.cog-genomics.org/plink2. (2015).
48. Chang, C.C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015).
49. Vestbo, J. *et al.* Evaluation of COPD Longitudinally to Identify Predictive Surrogate End-points (ECLIPSE). *Eur Respir J* **31**, 869-73 (2008).
50. Vestbo, J. *et al.* Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: GOLD executive summary. *Am J Respir Crit Care Med* **187**, 347-65 (2013).
51. Price, A.L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* **38**, 904-9 (2006).
52. International HapMap, C. The International HapMap Project. *Nature* **426**, 789-96 (2003).
53. Delaneau, O., Marchini, J. & Zagury, J.F. A linear complexity phasing method for thousands of genomes. *Nat Methods* **9**, 179-81 (2012).
54. Hancock, D.B. *et al.* Meta-analyses of genome-wide association studies identify multiple loci associated with pulmonary function. *Nat Genet* **42**, 45-52 (2010).
55. Wilk, J.B. *et al.* Genome-wide association studies identify CHRNA5/3 and HTR4 in the development of airflow obstruction. *Am J Respir Crit Care Med* **186**, 622-32 (2012).
56. Cho, Y.S. *et al.* A large-scale genome-wide association study of Asian populations uncovers genetic factors influencing eight quantitative traits. *Nat Genet* **41**, 527-34 (2009).
57. Kim, W.J. *et al.* Genome-wide association studies identify locus on 6p21 influencing lung function in the Korean population. *Respirology* **19**, 360-8 (2014).
58. de Jong, K. *et al.* NOS1: a susceptibility gene for reduced level of FEV1 in the setting of pesticide exposure. *Am J Respir Crit Care Med* **190**, 1188-90 (2014).
59. de Jong, K. *et al.* Genome-wide interaction study of gene-by-occupational exposure and effects on FEV1 levels. *J Allergy Clin Immunol* **136**, 1664-1672 e14 (2015).

60. Scholtens, S. *et al.* Cohort Profile: LifeLines, a three-generation cohort study and biobank. *Int J Epidemiol* **44**, 1172-80 (2015).
61. Sood, A. *et al.* Difference in airflow obstruction between Hispanic and non-Hispanic White female smokers. *COPD* **5**, 274-81 (2008).
62. Bild, D.E. *et al.* Multi-Ethnic Study of Atherosclerosis: objectives and design. *Am J Epidemiol* **156**, 871-81 (2002).
63. Kaufman, J.D. *et al.* Prospective study of particulate air pollution exposures, subclinical atherosclerosis, and clinical cardiovascular disease: The Multi-Ethnic Study of Atherosclerosis and Air Pollution (MESA Air). *Am J Epidemiol* **176**, 825-37 (2012).
64. Hankinson, J.L. *et al.* Performance of American Thoracic Society-recommended spirometry reference values in a multiethnic sample of adults: the multi-ethnic study of atherosclerosis (MESA) lung study. *Chest* **137**, 138-45 (2010).
65. Manichaikul, A. *et al.* Association of SCARB1 variants with subclinical atherosclerosis and incident cardiovascular disease: the multi-ethnic study of atherosclerosis. *Arterioscler Thromb Vasc Biol* **32**, 1991-9 (2012).
66. Marchini, J. & Howie, B. Genotype imputation for genome-wide association studies. *Nat Rev Genet* **11**, 499-511 (2010).
67. Fishman, A. *et al.* A randomized trial comparing lung-volume-reduction surgery with medical therapy for severe emphysema. *N Engl J Med* **348**, 2059-73 (2003).
68. Bell, B., Rose, C. & Damon, D. The Normative Aging Study: an interdisciplinary and longitudinal study of health and aging. *Aging Hum Dev* **3**, 5-17 (1972).
69. Cho, M.H. *et al.* Variants in FAM13A are associated with chronic obstructive pulmonary disease. *Nat Genet* **42**, 200-2 (2010).
70. Cho, M.H. *et al.* A genome-wide association study of COPD identifies a susceptibility locus on chromosome 19q13. *Hum Mol Genet* **21**, 947-57 (2012).
71. Sorheim, I.C. *et al.* Case-control studies on risk factors for chronic obstructive pulmonary disease: how does the sampling of the cases and controls affect the results? *Clin Respir J* **4**, 89-96 (2010).
72. Hofman, A. *et al.* The Rotterdam Study: 2016 objectives and design update. *Eur J Epidemiol* **30**, 661-708 (2015).
73. Kreiner-Moller, E., Medina-Gomez, C., Uitterlinden, A.G., Rivadeneira, F. & Estrada, K. Improving accuracy of rare variant imputation with a two-step imputation approach. *Eur J Hum Genet* **23**, 395-400 (2015).
74. Couper, D. *et al.* Design of the Subpopulations and Intermediate Outcomes in COPD Study (SPIROMICS). *Thorax* **69**, 491-4 (2014).
75. Hobbs, B.D. *et al.* Exome Array Analysis Identifies a Common Variant in IL27 Associated with Chronic Obstructive Pulmonary Disease. *Am J Respir Crit Care Med* **194**, 48-57 (2016).
76. Silverman, E.K. *et al.* Gender-related differences in severe, early-onset chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* **162**, 2152-8 (2000).
77. Manichaikul, A. *et al.* Robust relationship inference in genome-wide association studies. *Bioinformatics* **26**, 2867-73 (2010).
78. Therneau, T., Atkinson, E., Sinnwell, J., Schaid, D. & McDonnell, S. kinship2: Pedigree functions. R package version 1.6.0. URL: http://CRAN.R-project.org/package=kinship2. (2014).

79. Wang, C. TRACE: fasT and Robust Ancestry Coordinate Estimation version 1.0. URL: http://www.sph.umich.edu/csg/chaolong/LASER/. (2014).
80. Wang, C., Zhan, X., Liang, L., Abecasis, G.R. & Lin, X. Improved ancestry estimation for both genotyping and sequencing data using projection procrustes analysis and genotype imputation. *Am J Hum Genet* **96**, 926-37 (2015).
81. R Core Team (2016). R: A language and environment for statistical computing. (R Foundation for Statistical Computing. URL: http://www.R-project.org/, Vienna, Austria).
82. Zhu, G. *et al.* The SERPINE2 gene is associated with chronic obstructive pulmonary disease in two large populations. *Am J Respir Crit Care Med* **176**, 167-73 (2007).
83. Patel, B.D. *et al.* Airway wall thickening and emphysema show independent familial aggregation in chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* **178**, 500-5 (2008).
84. Zhou, X. *et al.* Identification of a chronic obstructive pulmonary disease genetic determinant that regulates HHIP. *Hum Mol Genet* **21**, 1325-35 (2012).
85. Wain, L.V. *et al.* Novel insights into the genetics of smoking behaviour, lung function, and chronic obstructive pulmonary disease (UK BiLEVE): a genetic association study in UK Biobank. *Lancet Respir Med* **3**, 769-81 (2015).
86. Miller, M.R. *et al.* Standardisation of spirometry. *Eur Respir J* **26**, 319-38 (2005).
87. Consortium, U.K. *et al.* The UK10K project identifies rare variants in health and disease. *Nature* **526**, 82-90 (2015).
88. Wakefield, J. Bayes factors for genome-wide association studies: comparison with P-values. *Genet Epidemiol* **33**, 79-86 (2009).
89. Kichaev, G. & Pasaniuc, B. Leveraging Functional-Annotation Data in Trans-ethnic Fine-Mapping Studies. *Am J Hum Genet* **97**, 260-71 (2015).
90. Morris, A.P. Transethnic meta-analysis of genomewide association studies. *Genet Epidemiol* **35**, 809-22 (2011).
91. Lu, Q., Powles, R.L., Wang, Q., He, B.J. & Zhao, H. Integrative Tissue-Specific Functional Annotations in the Human Genome Provide Novel Insights on Many Complex Traits and Improve Signal Prioritization in Genome Wide Association Studies. *PLoS Genet* **12**, e1005947 (2016).
92. Finucane, H.K. *et al.* Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat Genet* **47**, 1228-35 (2015).
93. Trynka, G. *et al.* Disentangling the Effects of Colocalizing Genomic Annotations to Functionally Prioritize Non-coding Variants within Complex-Trait Loci. *Am J Hum Genet* **97**, 139-52 (2015).
94. Slowikowski, K., Hu, X. & Raychaudhuri, S. SNPsea: an algorithm to identify cell types, tissues and pathways affected by risk loci. *Bioinformatics* **30**, 2496-7 (2014).
95. Zhang, K., Cui, S., Chang, S., Zhang, L. & Wang, J. i-GSEA4GWAS: a web server for identification of pathways/gene sets associated with traits by applying an improved gene set enrichment analysis to genome-wide association study. *Nucleic Acids Res* **38**, W90-5 (2010).
96. Kanehisa, M. & Goto, S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* **28**, 27-30 (2000).
97. Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* **25**, 25-9 (2000).

98. Gene Ontology, C. Gene Ontology Consortium: going forward. *Nucleic Acids Res* **43**, D1049-56 (2015).
99. Tasan, M. *et al.* Selecting causal genes from genome-wide association studies via functionally coherent subnetworks. *Nat Methods* **12**, 154-9 (2015).
100. Database of Single Nucleotide Polymorphisms (dbSNP). Bethesda, MD: National Center for Biotechnology Information, National Library of Medicine. (dbSNP Build ID: 137). URL: http://www.ncbi.nlm.nih.gov/SNP/.
101. Nguyen, T., Khosravi, A., Creighton, D. & Nahavandi, S. Spike sorting using locality preserving projection with gap statistics and landmark-based spectral clustering. *J Neurosci Methods* **238**, 43-53 (2014).