Genome-wide association analyses for lung function and chronic obstructive pulmonary disease identify new loci and potential druggable targets

Louise V Wain, Nick Shrine, María Soler Artigas, A Mesut Erzurumluoglu, Boris Noyvert, Lara Bossini-Castillo, Ma'en Obeidat, Amanda P Henry, Michael A Portelli, Robert J Hall, Charlotte K Billington,Tracy L Rimington, Anthony G Fenech, Catherine John, Tineka Blake, Victoria E Jackson, Richard J Allen, Bram P Prins, Understanding Society Scientific Group, Archie Campbell, David J Porteous, Marjo-Riitta Jarvelin, Matthias Wielscher, Alan L James, Jennie Hui, Nicholas J Wareham, Jing Hua Zhao, James F Wilson, Peter K Joshi, Beate Stubbe, Rajesh Rawal, Holger Schulz, Medea Imboden, Nicole M Probst-Hensch, Stefan Karrasch, Christian Gieger, Ian J Deary, Sarah E Harris, Jonathan Marten, Igor Rudan, Stefan Enroth, Ulf Gyllensten, Shona M Kerr, Ozren Polasek, Mika Kähönen, Ida Surakka, Veronique Vitart, Caroline Hayward, Terho Lehtimäki, Olli T Raitakari, David M Evans, A John Henderson, Craig E Pennell, Carol A Wang, Peter D Sly, Emily S Wan, Robert Busch, Brian D Hobbs, Augusto A Litonjua, David W Sparrow, Amund Gulsvik, Per S Bakke, James D Crapo, Terri H Beaty, Nadia N Hansel, Rasika A Mathias, Ingo Ruczinski, Kathleen C Barnes, Yohan Bossé, Philippe Joubert, Maarten van den Berge, Corry-Anke Brandsma, Peter D Paré, Don D Sin, David C Nickle, Ke Hao, Omri Gottesman, Frederick E Dewey, Shannon E Bruse, David J Carey, H Lester Kirchner, Geisinger-Regeneron DiscovEHR Collaboration, Stefan Jonsson, Gudmar Thorleifsson, Ingileif Jonsdottir, Thorarinn Gislason, Kari Stefansson, Claudia Schurmann, Girish Nadkarni, Erwin P Bottinger, Ruth JF Loos, Robin G Walters, Zhengming Chen, Iona Y Millwood, Julien Vaucher, Om P Kurmi, Liming Li, Anna L Hansell, Chris Brightling, Eleftheria Zeggini, Michael H Cho, Edwin K Silverman, Ian Sayers, Gosia Trynka, Andrew P Morris, David P Strachan, Ian P Hall & Martin D Tobin

Supplementary Information

Contents

Supplementary Note

United Kingdom Household Longitudinal Study (UKHLS)

United Kingdom Household Longitudinal Study (UKHLS): The United Kingdom Household Longitudinal Study, also known as Understanding Society (https://www.understandingsociety.ac.uk) is a longitudinal panel survey of 40,000 UK households (England, Scotland, Wales and Northern Ireland) representative of the UK population. Participants are surveyed annually since 2009 and contribute information relating to their socioeconomic circumstances, attitudes, and behaviours via a computer assisted interview. The study includes phenotypical data for a representative sample of participants for a wide range of social and economic indicators as well as a biological sample collection encompassing biometric, physiological, biochemical, and haematological measurements and self-reported medical history and medication use. The United Kingdom Household Longitudinal Study has been approved by the University of Essex Ethics Committee and informed consent was obtained from every participant.

Lung function measurements were used from samples in England and Wales only where the electronic NDD Easy On-PC spirometer was used. For each participant the two highest FVC and FEV1 measurements are taken. Measurements were not taken from individuals who were pregnant, had abdominal or chest surgery or a heart attack in the last three months, had a detached retina or eye or ear surgery in the past 3 months, admitted to hospital with a heart complaint in the preceding month, had a resting pulse rate more than 120 beats/minute, or currently taking medications for the treatment of Tuberculosis.

10,484 UKHLS samples were genotyped using the Illumina Infinium HumanCoreExome (12v1-0) at the Wellcome Trust Sanger Institute, Hinxton, UK and genotypes were called using Illumina Genome Studio Gencall. Variants were mapped to NCBI build 37 (hg19) coordinates and strand was standardised (http://www.well.ox.ac.uk/~wrayner/strand/). Samples were excluded according to the following: call rate < 98%, autosomal heterozygosity outliers (> 3 SD), sex discrepancy, duplicates established using identity by descent (IBD) PI_HAT > 0.9, ethnic outliers after combining with 1000 Genomes Project data and carrying out IBD and multidimensional scaling. Variants were excluded with Hardy-Weinberg equilibrium (HWE) p-value < 1x10-4, call rate < 98% and poor genotype clustering values (< 0.4). Unrelated samples were determined by performing IBD and samples with PI_HAT > 0.2 were excluded resulting in 9,308 samples and 525,314 variants.

Prior to phasing additional variant QC was performed; duplicates, monomorphics and singletons were excluded. Will Rayners script was used for comparing alleles and frequencies with the 1000 Genomes Project haplotypes (http://www.well.ox.ac.uk/~wrayner/tools/). Samples were phased using SHAPEIT v2.r778. A combined reference panel was used consisting of 1000 Genomes Project[1] (27,449,245 variants and 1,092 samples), and UK10K[2] (25,109,897 variants and 3,781 samples). For 1000 Genomes Project the haplotypes used were 1000 Genomes Project (1000G) haplotypes Phase I integrated variant set release (ALL.integrated_phase1_SHAPEIT_16-06-14.nosing) downloaded from the IMPUTEv2 website (http://mathgen.stats.ox.ac.uk/impute/impute_v2.1.0.html). For UK10K the haplotypes were prepared and described previously[2,3]. IMPUTEv2[4,5] was used for imputation. Post imputation variant QC consisted of excluding variants with an IMPUTE info score < 0.4 and/or HWE p-value < $1x10^{-4}$.

Studies contributing to analyses of COPD susceptibility and risk of exacerbation

UK Biobank

In UK Biobank, COPD status was defined based on spirometry with individuals with % predicted FEV_1<80% and FEV_1/FVC<0.7 (indicative of moderate to severe COPD[6]) selected as COPD cases.

Individuals with FEV_1/FVC>0.7 and % predicted FEV_1>80% were selected as controls (in UK BiLEVE, controls were selected from the high % predicted FEV_1 group only and all had % predicted FEV_1>107%). Individuals were defined as exacerbation cases is they were COPD cases, as defined above, and had any of the following ICD-10 codes, according to the Hospital Episodes Statistics (HES) in UK Biobank: from J40 to J44 (excluding J43.0), J06.9, J13 to J16, J18 (excluding J18.2), J20.8, J20.9 or J22. Exacerbation controls were defined as COPD cases (as above) who were not exacerbation cases.
Analyses were carried out using the score test, implemented in SNPTEST v2.5b4 [7] assuming an additive genetic model of genotype dose. For never-smokers, sex, age, age^2, height and the first 10 ancestry principal components were included as covariates. For heavy-smokers, pack years were included as an additional covariate. The results for never and heavy-smokers were then combined, using inverse variance weighted meta-analysis. Due to minor differences in the array and imputation, analyses were carried out separately in the stage 1 UK BiLEVE subset and the stage 2 subset of UK Biobank and results were meta-analysed (inverse variance weighted).

deCODE COPD Study

deCODE genetics have collected spirometry data through their own phenotyping efforts and through epidemiological studies and clinical services carried out by collaborating physicians. The available measurements were performed between 1977 and 2010. Quality controlled spirometry data without prior administration of an inhaled bronchodilator medication was available for 4,872 individuals with genotype information. Based on the latest spirometry result available for each individual, a COPD diagnosis was made if the GOLD 2 criteria was fulfilled (FEV_1/FVC < 0.70 and FEV_1 % of predicted < 80). This resulted in a group of 1,964 spirometrically defined COPD patients with age at spirometry > 40 years. Of those, 1,248 were chip-typed and directly imputed; the remaining 716 were first or second degree relatives to chip typed individuals and had their genotypes inferred based on genealogy [8]. 1,236 were GOLD 2, 590 were GOLD 3 and 138 were GOLD 4 patients. Based on the available information on smoking status, subgroups of ever-smokers (1,015 chip typed, 535 relatives) and never-smokers (87, all chip typed) were defined.
Single variant association testing was performed using logistic regression, adjusting for sex, age and county, as previously described [9]. Genotypes were familially imputed into close relatives of chip typed individuals, achieving sample sizes of 1,964 for all COPD, 1,550 COPD smokers and 87 COPD non-smokers. Population controls (142,262) were used for analysis of the entire COPD cohort, but for the smoker and non-smoker subsets, selected control groups of 7,468 and 449 individuals, respectively, matched on sex, age, smoking status and genotyping status were used.
Familially imputed genotypes are not applicable to genetic risk score analysis by current in-house methodology, so only chip typed individuals were used for the risk scores, reducing case and control group sizes to 1,248/74,770 and 1,015/5,075 for the whole cohort and smoker subset, respectively.
To account for inflation in test statistics due to cryptic relatedness and stratification within the case and control sample sets, we applied an LD regression based genomic control correction factor[10] to the association analysis. The estimated correction factor was 1.14, 1.12 and 1.02 for the whole cohort, smoker subset and non-smoker subset, respectively.
Approval for these studies was provided by the National Bioethics Committee and the Icelandic Data Protection Authority.

Lung resection cohorts: Groningen, Laval and University of British Columbia (UBC)

The details and subjects' characteristics of the lung eQTL study population have been previously described[11,12]. All lung tissue samples were obtained in accordance with Institutional Review Board guidelines at the three sites: Laval University (Quebec, Canada), University of British-Columbia (Vancouver, Canada) and Groningen University (Groningen, The Netherlands). 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 ("Code of conduct; Dutch federation of biomedical scientific societies"; http://www.federa.org).

Briefly, Following standard microarray and genotyping quality controls, 1,111 patients were available including The University of British Columbia Centre for Heart and Lung Innovation (n=339, Vancouver, Canada), Laval University (n=409, Quebec City, Canada) and the University of Groningen (n=363, Groningen, The Netherlands). Gene expression profiling was performed using an Affymetrix custom array (GPL10379) testing 51,627 non-control probesets and normalization was performed using multi-array average (RMA)[13]. The expression data are available at NCBI Gene Expression Omnibus repository through accession numbers GSE23352, GSE23529 and GSE23545.
Genotyping was performed on DNA extracted from blood or lung tissue using the Illumina Human1M-Duo BeadChip array, and imputation was performed with MaCH/Minimac software[14] using the 1000G reference panel, March 2012 release. The eQTL analysis was adjusted for age, sex and smoking status in each study separately, and the results were meta-analysed using inverse variance weighting meta-analysis. The resulting eQTLs were categorized into cis-acting (less than 1Mb away from transcription start site) or trans eQTLs (further than 1Mb away or on a different chromosome). Genome-wide significant threshold was set using Benjamini-Hochberg 10% FDR.
COPD was defined dichotomously based on an $FEV_1/FVC < 0.7$ cutoff. Post-bronchodilator spirometry was used when available; otherwise, pre-bronchodilator values were used[15].

COPD case-control studies: COPDGene Study

Details of the COPDGene Study (NCT00608764, www.copdgene.org) have been previously described[16,17]. 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). Moderate to severe cases were defined using post-bronchodilator % predicted FEV_1 < 80% predicted and FEV_1/FVC < 0.7. Genotyping was performed by Illumina (San Diego, CA) on the HumanOmniExpress array. Subjects were excluded for missingness, heterozygosity, chromosomal aberrations, sex check, population outliers, and cryptic relatedness. Genotyping at the Z and S alleles was performed in all subjects. Subjects known or found to have severe 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 and minimac (version 2012-10-09). Reference panels for the non-Hispanic whites and African-Americans were the 1000 Genomes Phase I v3 European (EUR) and cosmopolitan reference panels, respectively. Variants with an r^2 value of ≤ 0.3 were removed from further analysis.
Exacerbation data were ascertained by questionnaire at enrolment; subjects were asked to recount up to 6 exacerbation episodes which occurred during the year prior to enrolment. Cases were defined as COPD subjects who reported an exacerbation requiring hospitalization or an emergency room (ER) visit. Controls were COPD subjects who did not report any exacerbations requiring hospitalization/ER visit.

COPD case-control studies: Evaluation of COPD Longitudinally to Identify Predictive Surrogate End-points (ECLIPSE)

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 [18,19]. 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 2 and above COPD, and controls had normal spirometry (% predicted FEV_1 > 85%). 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). Imputation was performed using MaCH and minimac (version 2012-10-09) and the 1000 Genomes Phase I v3 European (EUR) reference panel.
Exacerbation data were ascertained by questionnaire at enrolment; cases were defined as COPD subjects who reported ≥1 exacerbation requiring hospitalization during the year prior to enrolment. Control subjects did not report any exacerbations requiring hospitalization during the year prior to enrolment.

COPD case-control studies: National Emphysema Treatment Trial (NETT) and Normative Aging Study (NAS) (NETT/NAS)

Details of the National Emphysema Treatment Trial have been described previously [19,20]. NETT (www.nhlbi.nih.gov/health/prof/lung/nett/) was a multicentre clinical trial to evaluate lung volume reduction surgery. Enrolled subjects had severe airflow obstruction by post-bronchodilator spirometry (% predicted FEV_1 < 45%) 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)[19,21]. 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. Imputation was performed using MaCH and minimac (version 2012-10-09) and the 1000 Genomes Phase I v3 European (EUR) reference panel.

Exacerbations were ascertained using Medicare billing data during the year prior to enrolment. Subjects who were hospitalized for COPD exacerbations were considered cases; subjects who were not hospitalized for COPD exacerbations during the year before enrolment were considered controls.

COPD case-control studies: NORWAY-GenKOLS

Details on the Norwegian GenKOLS (Genetics of Chronic Obstructive Lung Disease, GSK code RES11080) study have been described previously[22]. Subjects with > 2.5 pack years of smoking history were recruited from Bergen, Norway; cases had post-bronchodilator GOLD 2 or greater disease, while controls had normal spirometry; subjects with severe alpha-1 antitrypsin deficiency and other lung diseases (aside from asthma) were excluded. 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 and minimac (version 2012-10-09) and the 1000 Genomes Phase I v3 European (EUR) reference panel.

Exacerbation data were ascertained by questionnaire at enrolment. Subjects who reported ≥1 hospitalization related to respiratory symptoms in the year prior to enrolment were considered cases. Subjects who did not report any hospitalizations for respiratory symptoms were considered controls.

eMR studies: Geisinger-Regeneron DiscovEHR Study (DiscovEHR)

The DiscovEHR[23] collaboration between the Regeneron Genetics Center and Geisinger Health System MyCode Community Health Initiative couples high throughput genetic data to a Healthcare Provider Organization utilizing longitudinal electronic health records (EHR). The study was approved by the institutional review board at the Geisinger Health System. A subset of individuals with available genome-wide genotyping data was included in the current study. Genotyping was performed using the Illumina OmniExpressExome BeadChip, with standard QC metrics applied. Imputation was performed with IMPUTE2 v2.3.2 using the 1000 Genomes cosmopolitan dataset (June 2014 version). COPD cases were defined using a combination of ICD-9 diagnosis codes and available lung function testing. ICD-9–based diagnoses required one or more of the following: a problem-list entry of the diagnosis code or an encounter diagnosis code entered for two separate outpatient visits on separate calendar days. To be considered a COPD case, individuals were required to have spirometry-confirmed airflow obstruction (FEV_1/FVC< 0.70) **and** any of the following ICD-9 diagnoses codes: 490, 491.0, 491.1, 491.8, 491.9, 492.8, 492.0, 491.22, 493.21, 491.21, 493.22, 491.20, 493.20 and 496. Controls were defined as individuals without an ICD-9 diagnosis code of either asthma or COPD. Asthmatics were excluded from the control group given that the shared features of these diseases complicate their diagnosis in a clinical setting. Both cases and controls were restricted to individuals of European genetic ancestry and with age > 40. For exacerbation analyses, cases were COPD patients (as described above) with one or more inpatient admissions attributed to COPD; controls were COPD patients with no inpatient admissions attributed to COPD.

eMR studies: Mount Sinai Bio*Me* Biobank (Bio*Me*)

The Bio*Me* Biobank is an ongoing, prospective, hospital- and outpatient- based population research program operated by The Charles Bronfman Institute for Personalized Medicine (IPM) at The Icahn School of Medicine at Mount Sinai and has enrolled over 33,000 participants since September 2007. Bio*Me* is an Electronic Medical Record (EMR)-linked biobank that integrates research data and clinical care information for consented patients at The Mount Sinai Medical Center, which serves diverse local communities of upper Manhattan with broad health disparities. Bio*Me* populations include 25% of African American ancestry (AA), 36% of Hispanic Latino ancestry (HL), 30% of white European ancestry (EA), and 9% of other ancestry. The Bio*Me* disease burden is reflective of health disparities in the local communities. Bio*Me* operations are fully integrated in clinical care processes, including direct recruitment from clinical sites waiting areas and phlebotomy stations by dedicated recruiters independent of clinical care providers, prior to or following a clinician standard of care visit. Recruitment currently occurs at a broad spectrum of over 30 clinical care sites.

Information on COPD cases status (ICD9 codes), height, age and sex was derived from participants' EMR. Case/control selection was restricted to individuals with age > 40 years, available genotyping data, as well as sex, height and smoking data. Case/control definition was carried out based on information retrieved from EMRs: COPD cases were defined as individuals with records of ICD-9 codes for COPD (491.xx-492.xx, 496.xx), whereas COPD controls were defined as individuals with none of the above listed ICD-9 codes for COPD.

Exacerbation cases and controls were defined as individuals with and without a primary COPD diagnosis (based on the ICD codes) at an inpatient visit, respectively.

Bio*Me* participants were genotyped with the Illumina HumanOmniExpressExome-8 v1.0 BeadChip array and imputed to the 1000 Genomes Project Phase 1 (March 2012) reference panel using IMPUTE2. SNPs of interest were extracted using gtool [http://www.well.ox.ac.uk/~cfreeman/software/gwas/gtool.html]. Out of the 95 COPD variants, 93 were available in the Bio*Me* data set either directly genotyped or imputed with good imputation quality (info>0.7), for two variants, proxies were used (rs12438269 for rs66650179 [r^2=0.618] and rs62070270 for rs59835752 [r^2=0.999]). Association analyses were carried out using generalized linear models in R stratified by self-reported ancestry (EA: 207 COPD cases and 1,817 controls).

Chinese ancestry: China Kadoorie Biobank prospective cohort (CKB)

The CKB study involved 512,891 participants, aged 30-79 years, recruited between 2004-8 from 10 diverse regions of China and who gave their informed written consent to proceed to an extensive collection of clinical and environmental data at baseline[24]. Subsets of ~25,000 survivors were actively followed up in 2008 (1st resurvey) and in 2013-14 (2nd resurvey) with additional collection of clinical and blood samples. Furthermore, all participants were followed up for cause-specific mortality and episodes of hospitalisation using:[24] (i) cross-checking with official death certificates collected by the regional Center for Disease Control (CDC) to code causes of death according to World Health Organisation ICD-10 codes; (ii) linkage with established disease registries to supplement information on non-fatal events for 4 major diseases (stroke, ischaemic heart disease (IHD), diabetes, and cancer); and (iii) electronic records from the the national Chinese health insurance (HI) system, to retrieve additional disease and hospitalisation events (e.g. COPD).

A genotyping study (hereafter¸ called SNP-Panel) of 384 single-nucleotide polymorphisms (SNPs) was conducted in 93,208 (after quality control [QC]) subjects in 2013-14. SNPs were selected based on previous association (mainly GWAS) with chronic diseases (e.g. stroke, IHD) and intermediate phenotypes (e.g. lung function, blood pressure, BMI), metabolic pathways (e.g. Vitamin D) and risk exposure (e.g. smoking). In addition, using a customised Affymetrix Axiom® CKB array (optimised for use with Han Chinese subjects) including ~700,000 markers before imputation (including all markers included on the SNP-panel), a genotyping study (GWAS) was conducted in 2014-15 in 32,201 (after QC) individuals, including ~14,000 with SNP-panel data. Subjects were selected for the GWAS who were part of a stroke nested case/control study (~20,000), had additional phenotypes of interest (ischaemic heart disease, ~2,000; COPD exacerbations, ~5,000), and ~5,000 participants who attended the 2nd resurvey. Participants with prior self-reported cardiovascular disease, cancer and/or statin use at baseline were excluded.

We excluded participants who were <40 years of age and those with prior cardiovascular diseases, cancer and/or statin use to be consistent with the exclusion criteria for the GWAS data (see above). Only pre-bronchodilator spirometry measurements were available for the analysis. GOLD 2-4 was defined based on (i) a FEV_1/FVC ratio<70; and (ii) % predicted FEV_1 values as derived from Quanjer *et al*.[25] For individuals with lung function measurements available at the baseline and in the 1st and/or 2nd follow-ups, we used the highest lung function measurement for the analysis. Exacerbation status was defined as any hospitalisation for COPD exacerbation, as recorded through the Chinese health insurance system.
The GWAS dataset (n=32,201) was combined with a non-overlapping dataset from the SNP-Panel study (n=78,884), which yielded a combined dataset of 111,085 individuals with genetic data. Based on the list of SNPs provided by UK BiLEVE, we were able to identify 71 lead or proxy SNPs in the CKB dataset.
We identified those COPD cases and controls for whom genetic data were available, which yielded a dataset of 87,966 individuals for the COPD analysis. The same approach was used to select exacerbation cases and controls (n=10,566).

In single variant analysis, logistic regression of each SNP on (i) COPD and (ii) COPD exacerbation status was performed adjusting for sex, age, height, geographical region (n=10) and disease status (to account for ascertainment of a subset of the cohort based on disease status; 5 categories: ischaemic stroke, intra-cerebral haemorrhage, subarachnoid haemorrhage; ischaemic heart disease; no cardiovascular disease ascertainment). Inflation estimates (λ) corresponding to COPD and COPD exacerbation status analyses were derived from the results of array-wide association using the GWAS dataset and were estimated according to the LDscore intercept method, with λ=1.0302 for COPD and λ=1.0056 for COPD exacerbation. Adjusted inflation estimates for SNPs also present on the SNP-Panel were derived based on the appropriate numbers of cases and controls. Standard errors of the logOR for these analyses were adjusted for the estimated inflation.
In the genetic risk score analysis, we restricted the analysis to the GWAS subsample with genotypes for all SNPspresent in the single variant analysis, except for one (rs153916) that was only available in the SNP-Panel dataset; the GRS analysis thus included 70 SNPs. Missing genotypes were imputed as the mean genotype (2 x MAF) for the region for that individual, based on MAFs derived from a pruned GWAS dataset with relatives (3rd cousin or closer) excluded. Logistic regression of the risk score on COPD and COPD exacerbation status adjusting for sex, age, age^2, height, regions (n=10) and disease status (n=5; see above) was conducted. Standard errors for the logOR were again adjusted for the estimated inflation.
Data management was conducted using Stata v.13.1 (Stata Corp, TX, USA) and Plink 1.90. Single variant and genetic risk score analyses were conducted using Plink 1.90 and Stata v.13.1, respectively.

Lung Health Study (LHS)

The LHS was a multicenter clinical study to evaluate the effect of bronchodilators and smoking cessation on lung function decline in current smokers with mild-moderate COPD[26,27].
The details of genotyping and quality control have been previously described[28]. Briefly, samples were genotyped using the Illumina Human660WQuad v.1_A BeadChip. Overall, 98.4 % of samples (n = 4,181) passed initial quality control standards and genotypes were released for 559,766 SNPs. Imputation was undertaken with the software IMPUTE2[5] using the all ancestries 1000G reference panel, March 2012 release[29].

Hospitalizations were defined in the following way. For all hospitalizations, copies of essential documents were obtained from hospital record rooms. Records that made significant mention of respiratory or cardiovascular disease (CVD) or cancer were forwarded to the study's mortality and morbidity review board for definitive coding. Thus, "respiratory" hospitalizations were all deemed by this board as being primarily driven by a respiratory condition (e.g. COPD exacerbation and pneumonia)[30]. Testing for association with exacerbations defined as respiratory hospitalizations was performed using data on the total number of respiratory hospitalizations reported on LHS study participants at year 5.

Studies contributing analyses of lung function in children

Avon Longitudinal Study of Parents and Children (ALSPAC)

The Avon Longitudinal Study of Parents and Children (ALSPAC) recruited 14,541 pregnant women resident in Avon, UK with expected dates of delivery 1st April 1991 to 31st December 1992. 14,541 is the initial number of pregnancies for which the mother enrolled in the ALSPAC study and had either returned at least one questionnaire or attended a "Children in Focus" clinic by 19/07/99. Of these initial pregnancies, there was a total of 14,676 foetuses, resulting in 14,062 live births and 13,988 children who were alive at 1 year of age.

When the oldest children were approximately 7 years of age, an attempt was made to bolster the initial sample with eligible cases who had failed to join the study originally. As a result, when considering variables collected from the age of seven onwards (and potentially abstracted from obstetric notes) there are data available for more than the 14,541 pregnancies mentioned above.

The number of new pregnancies not in the initial sample (known as Phase I enrolment) that are currently represented on the built files and reflecting enrolment status at the age of 18 is 706 (452 and 254 recruited during Phases II and III respectively), resulting in an additional 713 children being enrolled. The phases of enrolment are described in more detail in the cohort profile paper[31].

The total sample size for analyses using any data collected after the age of seven is therefore 15,247 pregnancies, resulting in 15,458 foetuses. Of this total sample of 15,458 fetuses, 14,775 were live births and 14,701 were alive at 1 year of age.

Spirometry was performed using the Vitalograph Spirotrac IV system (Vitalograph, Maids Moreton UK) and the hand-held Medikro Spirostar USB spirometer (Medikro, Kuopio, Finland) using methods described previously[32,33]. The machines were calibrated every day the medical examination took place. FVC and FEV_1 were measured in sitting position, while wearing a nose clip, by trained personnel, according to the ATS/ERS guidelines. For each child, at least three acceptable manoeuvres had to be obtained. The best results of three acceptable & repeatable (FVC +/- 150mL) flow-volume curves were accepted after post hoc quality control by a respiratory physician.

Genotyping details are described in Kemp *et al*. (2014)[34]. Briefly, a total of 9,912 subjects were genotyped using the Illumina HumanHap550 quad genome-wide SNP genotyping platform by the Wellcome Trust Sanger Institute, Cambridge, UK and the Laboratory Corporation of America (LabCorp Holdings., Burlington, NC, USA). PLINK software (v1.07) was used to carry out quality control measures[35]. Individuals were excluded from further analysis on the basis of having incorrect gender assignments, minimal or excessive heterozygosity (,0.320 and .0.345 for the Sanger data and ,0.310 and .0.330 for the LabCorp data), disproportionate levels of individual missingness (.3%), evidence of cryptic relatedness (.10% IBD) and being of non-European ancestry (as detected by a multidimensional scaling analysis seeded with HapMap 2 individuals). EIGENSTRAT analysis revealed no additional obvious population stratification and genome-wide analyses with other phenotypes indicate a low lambda)[36]. SNPs with a minor allele frequency of ,1% and call rate of ,95% were removed. Furthermore, only SNPs that passed an exact test of Hardy–Weinberg equilibrium (P. > 5x10^-7) were considered for analysis. After quality control, 8,365 unrelated individuals who were genotyped at 500,527 SNPs were available for analysis. Known autosomal variants were imputed with Markov Chain Haplotyping software (MACH 1.0.16)[37,38], using CEPH individuals from phase II of the HapMap project (hg18) as a reference set (release 22)[39].

Please note that the ALSPAC study website contains details of all the data that is available through a fully searchable data dictionary (http://www.bris.ac.uk/alspac/researchers/data-access/data-dictionary/).

Ethical approval for the study was obtained from the ALSPAC Ethics and Law Committee and the Local Research Ethics Committees.

	Males:Females	Age (mean (SD) [range])	FEV_1 (l) (mean (SD) [range])	FVC (l) (mean (SD) [range])	FEV_1/FVC (mean (SD) [range])
ALSPAC	2547:2515	8.64 (0.30) [7.42-10.33]	1.70 (0.26) [0.68-2.80]	1.93 (0.32) [0.77-3.13]	0.88 (0.06) [0.50-1]

Raine study

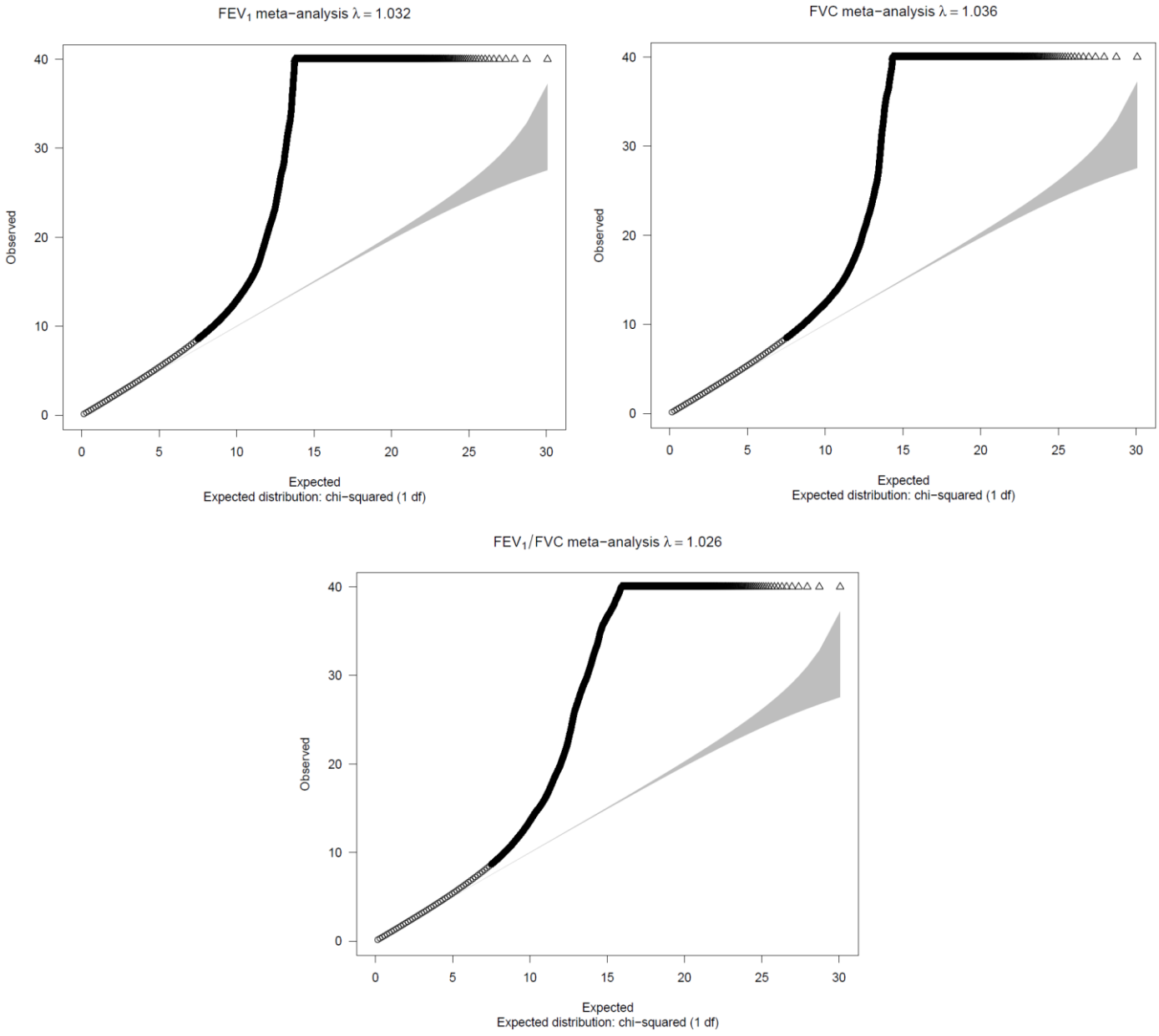
The Raine Study is a cohort of children formed in 1989-91 where approximately 2900 pregnant women volunteered to be part of the study at King Edward Memorial Hospital in Perth, Australia. Ethical approval was obtained from the University of Western Australia Human Research Ethics Committee.

Raine samples were genotyped using Illumina 660W Quad Array. Individuals genotyped were excluded if they had low genotyping success (>3% missing), excessive heterozygosity (which may indicate sample contamination), or had gender discrepancies between the core data and genotyped data. Individuals who were related with $\pi > 0.1875$ (in between second and third degree relatives – e.g. between half siblings and cousins) were investigated and the individual with a lower proportion of missing data was kept in the data set. Plate controls and replicates were removed from the data set. With replicates, the sample with a lower proportion of missing data was kept in the data set. A total number of 1494 individuals passed QC criteria and were used in genetics analyses. GWAS SNP QC was carried out in accordance to the Wellcome Trust Case Control Consortium thresholds (HWE $p < 5.7E-07$, call rate < 95%, MAF < 1%, A/T and G/C SNPs were also removed due to possible strand ambiguity). Imputation was then performed against the 1000G Phase 1 v3 reference using MACH/Minimac.

	Males:Females	Age (mean (SD) [range])	FEV_1 (l) (mean (SD) [range])	FVC (l) (mean (SD) [range])	FEV_1/FVC (mean (SD) [range])
Raine	590:630	8.1 (0.35) [7.13-9.98]	1.56 (0.25) [0.59-2.39]	1.65 (0.28) [0.59-2.92]	0.95 (0.05) [0.65-1.07]

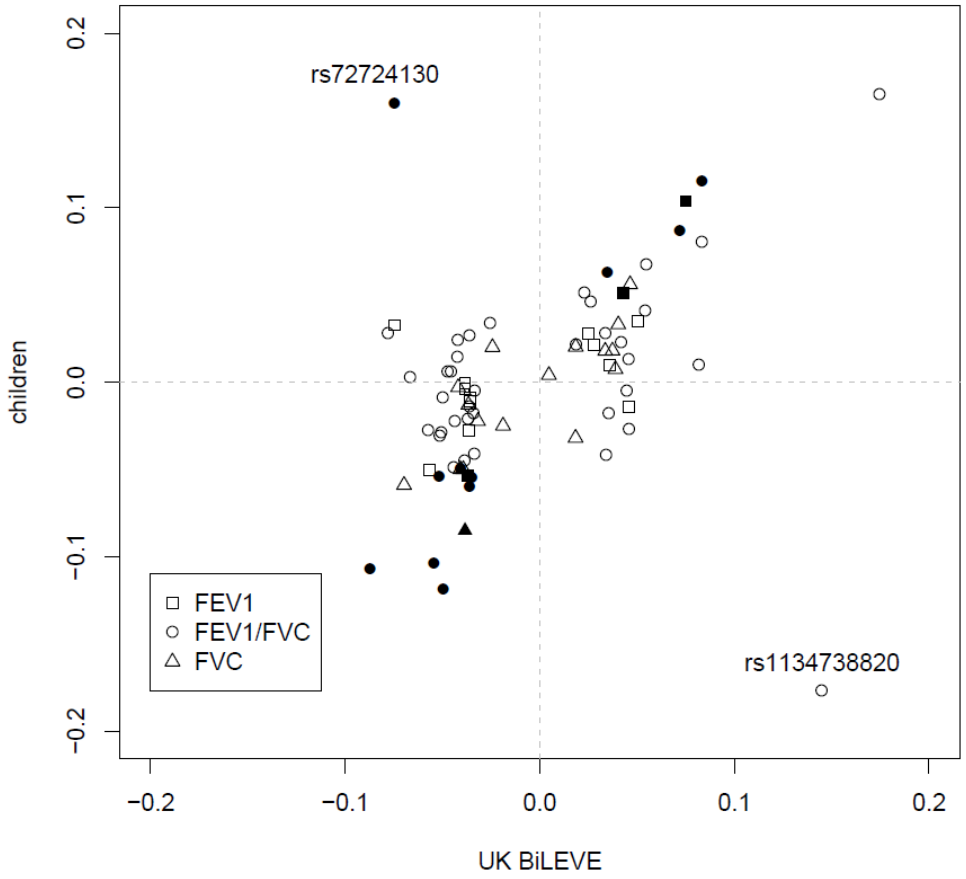
Supplementary Figures

Supplementary Figure 1: Quantile-Quantile (QQ)-plots and genomic inflation factor (λ) for discovery stage 1 (n= 48,943) association tests of FEV_1, FVC and FEV_1/FVC meta-analyses of heavy and never smokers.

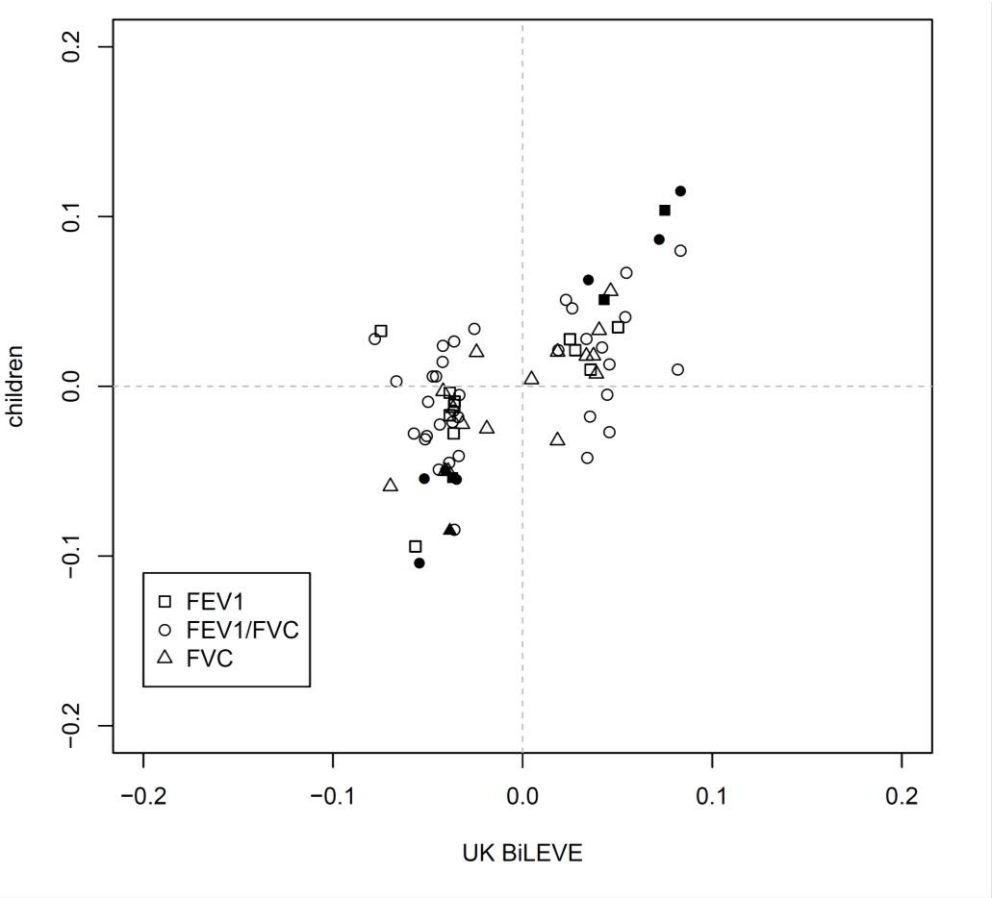


Supplementary Figure 2: **Comparison of effect sizes for lung function associated variants in adults and children**. a) Results available in children for 81 of the 97 variants with imputation quality >0.5 (79 variants in ALSPAC and 35 in Raine). Correlation coefficient r =0.417. Filled shapes indicate P<0.05 in children A genetic risk score of all 81 variants showed a per risk allele β (s.e.) on FEV_1, FVC and FEV_1/FVC of -0.0162 (0.003955) ($P=4.14x10^{-5}$), -0.0005 (0.003965) (P=0.894) and -0.0229 (0.003541) ($P=1.04x10^{-10}$). The two clear outliers were rs72724130 (novel signal in an intron of *MGA*, imputation quality=0.65, MAF=4.9% in ALSPAC) and rs113473882 (previously reported signal in an intron of *LTBP4*, imputation quality =0.76, MAF 1.34% in ALSPAC). Neither were available in Raine. Exclusion of these two SNPs gives a correlation coefficient r=0.71 for the remaining 79 variants. b) Seventy-three of the 81 variants had imputation quality >0.8 (71 variants in ALSPAC and 35 in Raine). Correlation coefficient r =0.651. Filled shapes indicate P<0.05 in children. A genetic risk score of all 73 variants showed a per risk allele β (s.e.) on FEV_1, FVC and FEV_1/FVC of -0.0177 (0.0040) ($P=1.03x10^{-5}$), -0.0037 (0.0041) (P=0.366) and -0.0213 (0.0037) ($P=1.27x10^{-8}$).

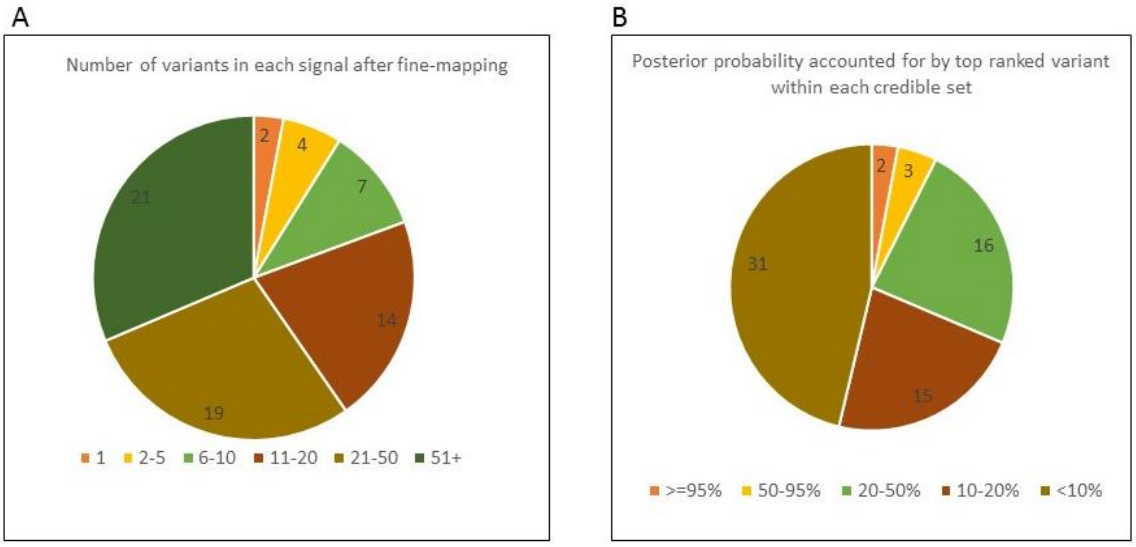
a


rs72724130
rs1134738820
FEV1
FEV1/FVC
FVC
children
UK BiLEVE
0.2
0.1
0.0
−0.1
−0.2
−0.2
−0.1
0.0
0.1
0.2

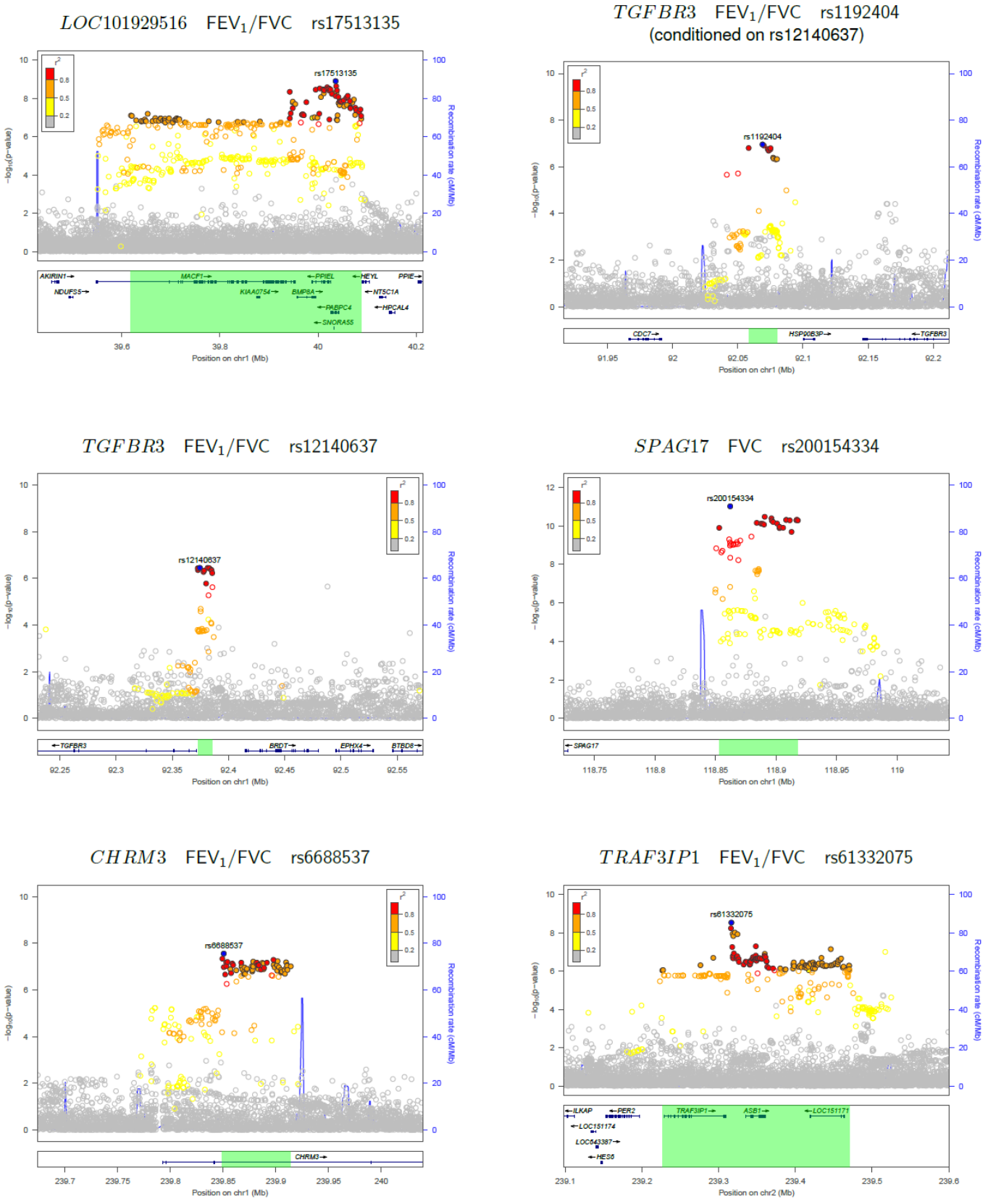
b


FEV1
FEV1/FVC
FVC
children
UK BiLEVE
0.2
0.1
0.0
−0.1
−0.2
−0.2
−0.1
0.0
0.1
0.2

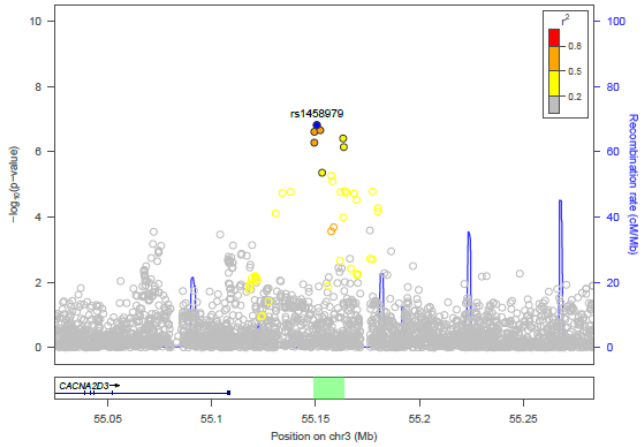
Supplementary Figure 3: Summary of Bayesian fine-mapping to 95% credible sets for lung function signals. The 95% credible set is the set of variants that are 95% likely to contain the underlying causal variant based on Bayesian refinement. Following exclusion of signals in the HLA region, one chromosome X signal and 23 previously-reported signals which did not reach $P<10^{-5}$ for association with lung function in stage 1 of this study, 67 signals underwent Bayesian fine-mapping to identify the 95% credible set. A: Numbers of signals fine-mapped to 1, 2-5, 6-10, etc variants. B: Numbers of signals for which a single variant accounts for >=95%, 50-95%, 20-50%, etc, of the posterior probability.



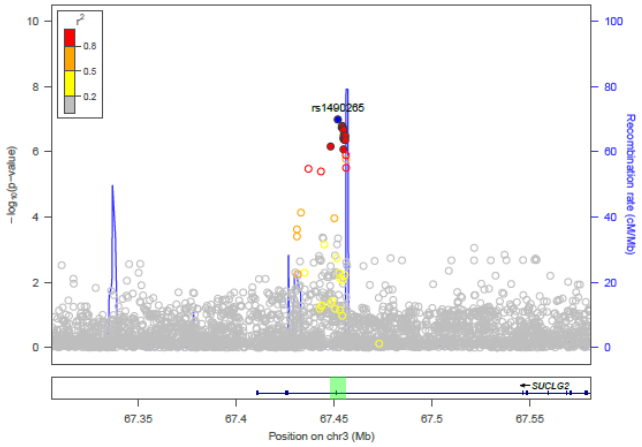
Supplementary Figure 4: Region plots with credible sets shown for 43 novel variants. Variants in the 95% credible set are shown as filled circles, those not in the credible set as open circles with the span of the credible set shaded in green on the gene track below. Credible sets were not calculated for 2 signals in the HLA region on chromosome 6 (labelled as *LST1* and *HLA-DQB1*). Where a "conditioned on" variant is given, the novel signal is a secondary or tertiary signal after conditioning and accordingly the region plot shows –log10 P values from stage 1 after conditioning on the corresponding variant.



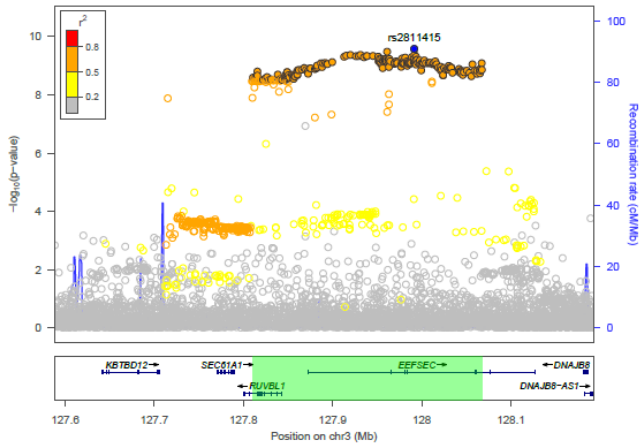
CACNA2D3 FEV_1/FVC rs1458979



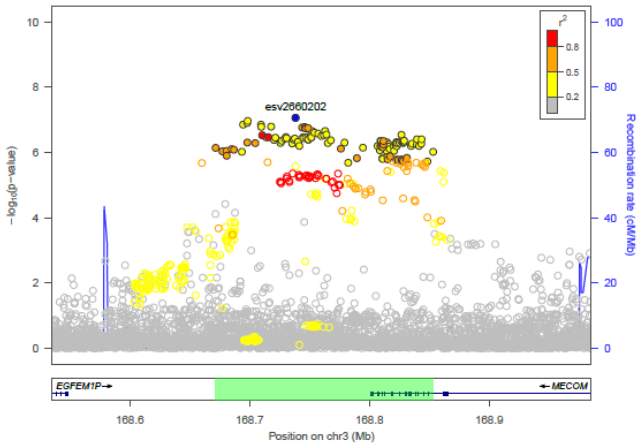
SUCLG2 FVC rs1490265



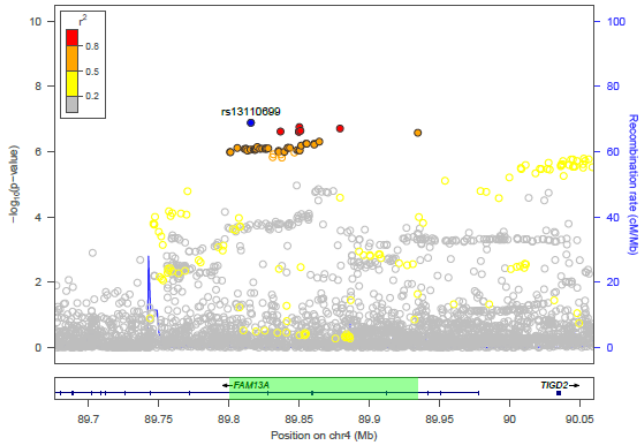
EEFSEC FEV_1/FVC rs2811415



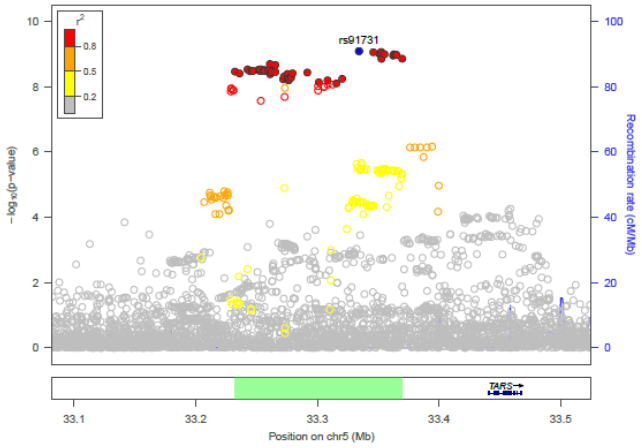
MECOM FEV_1/FVC esv2660202



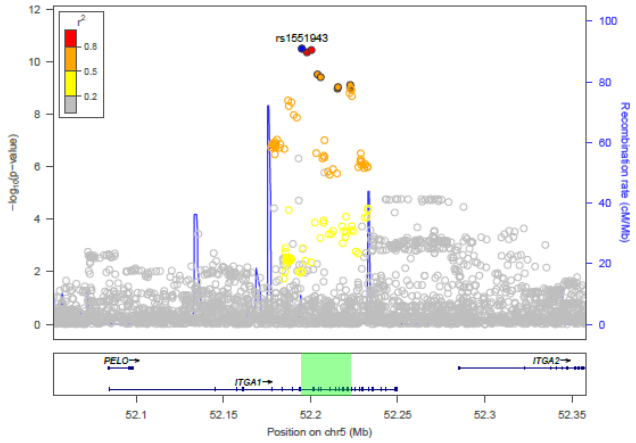
FAM13A FEV_1/FVC rs13110699
(conditioned on rs2045517)



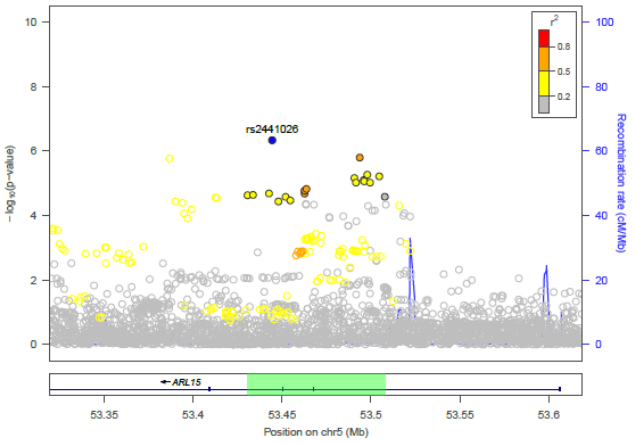
TARS FVC rs91731



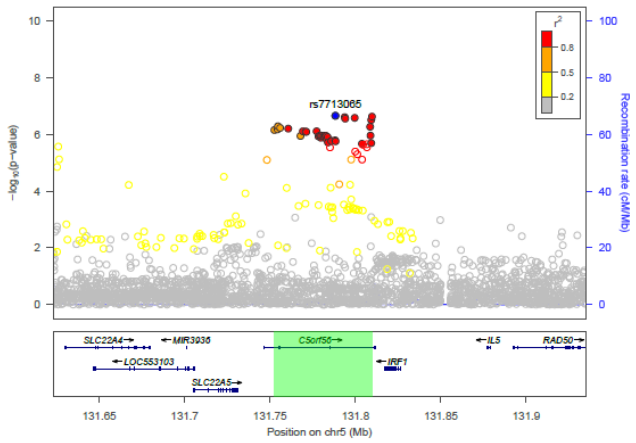
ITGA1 FEV_1/FVC rs1551943



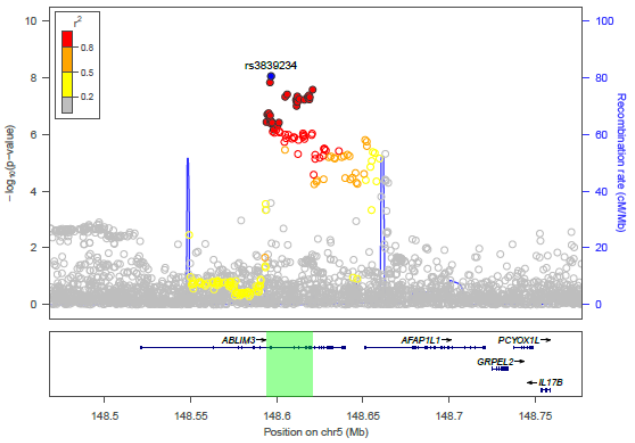
ARL15 FVC rs2441026



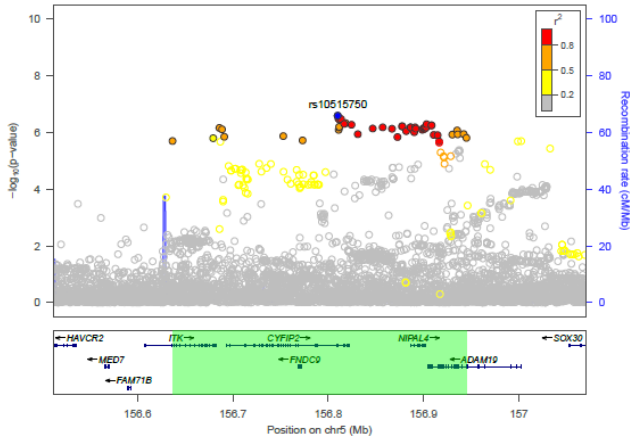
C5orf56 FEV_1/FVC rs7713065



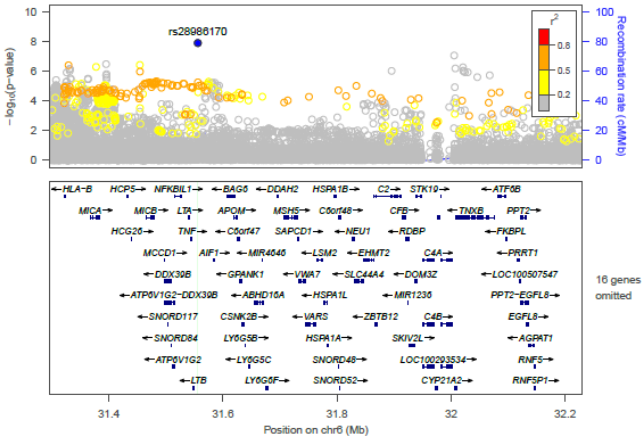
ABLIM3 FEV_1 rs3839234



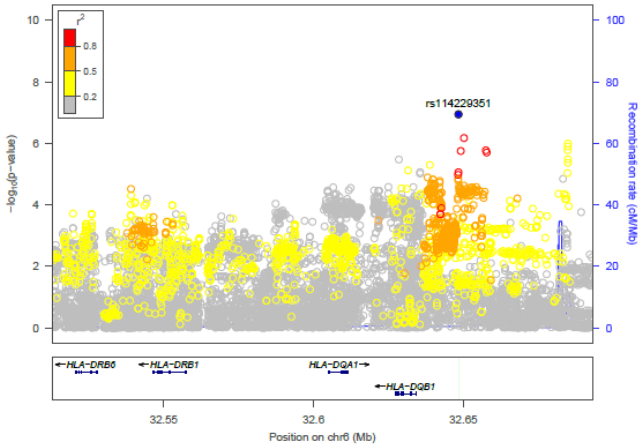
CYFIP2 FEV_1/FVC rs10515750
(conditioned on rs1990950)



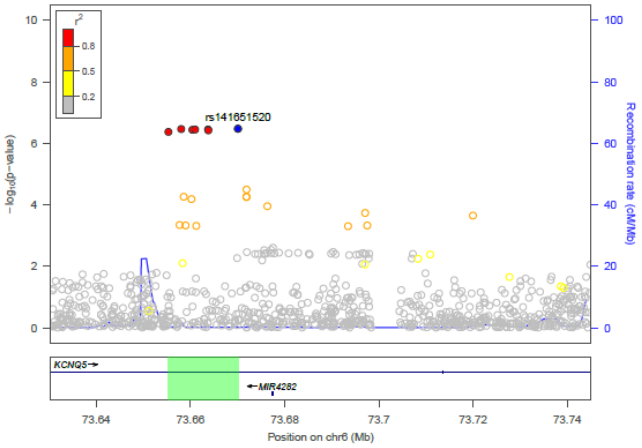
LST1 FEV_1/FVC rs28986170
(conditioned on rs2070600 & rs201002132)



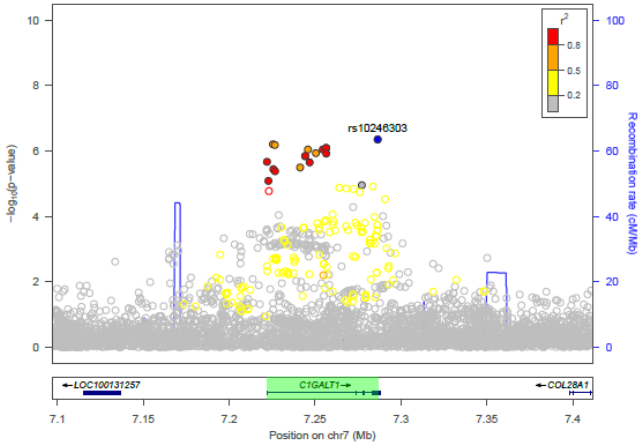
$HLA-DQB1$ FEV_1 rs114229351



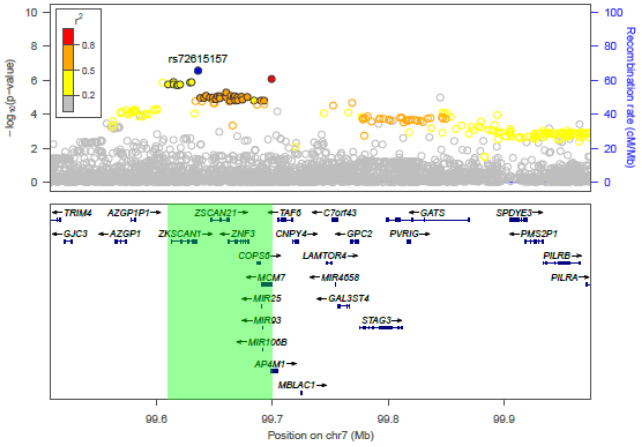
$KCNQ5$ FEV_1/FVC rs141651520



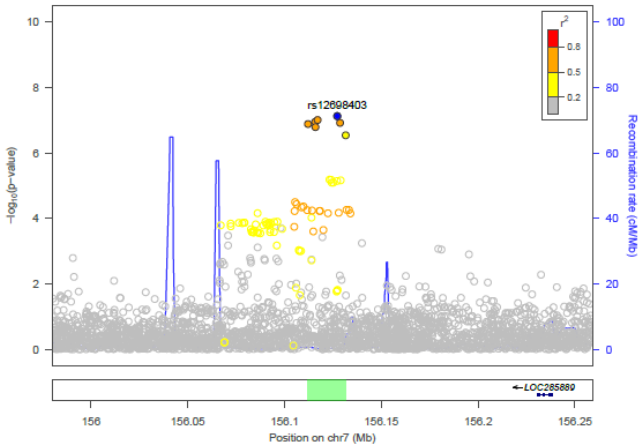
$C1GALT1$ FEV_1/FVC rs10246303



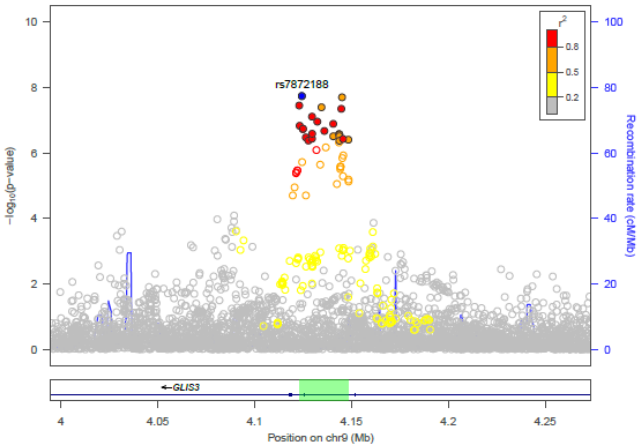
$ZKSCAN1$ FEV_1/FVC rs72615157



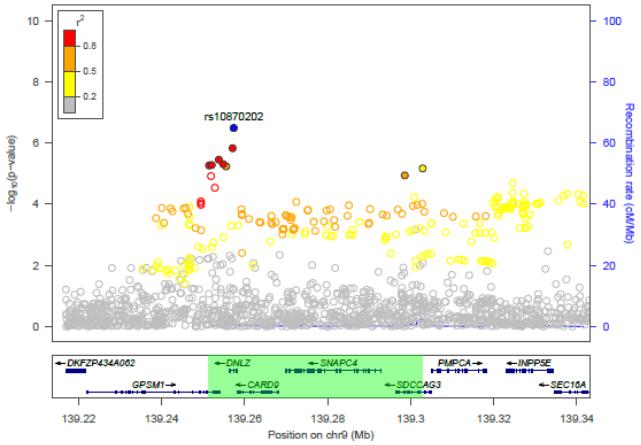
$LOC285889$ FEV_1 rs12698403



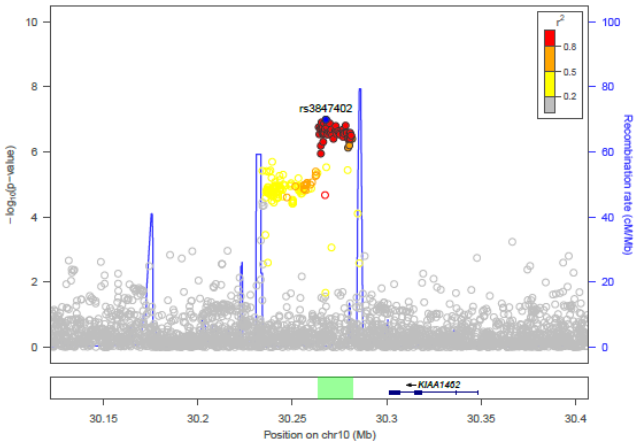
$GLIS3$ FEV_1 rs7872188



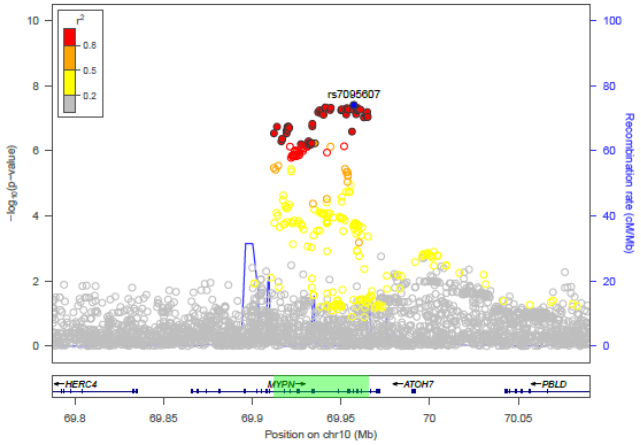
DNLZ FVC rs10870202
(conditioned on rs10858246)



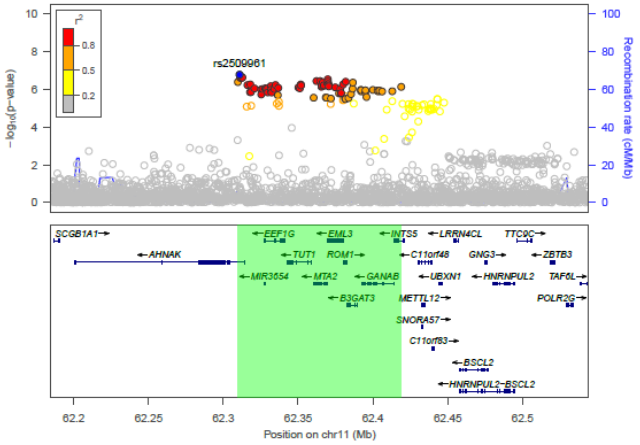
KIAA1462 FEV_1/FVC rs3847402



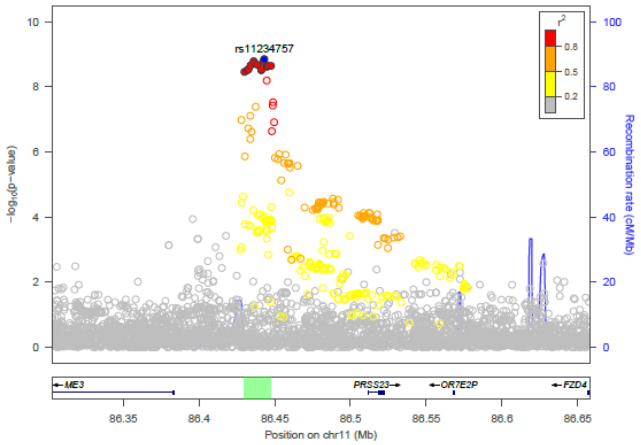
MYPN FVC rs7095607



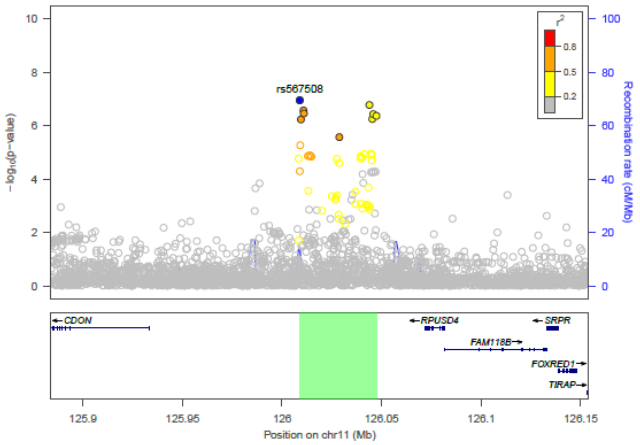
AHNAK FEV_1 rs2509961



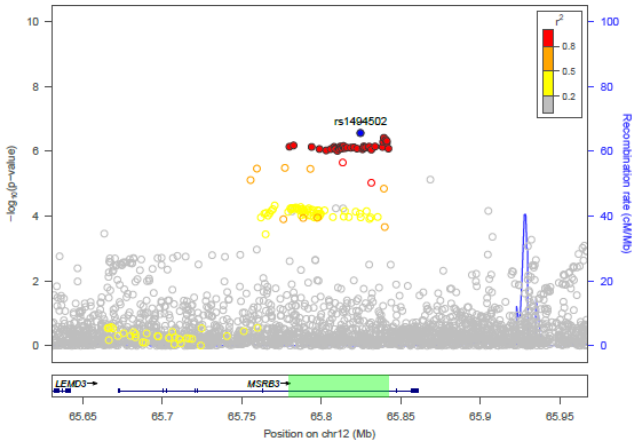
PRSS23 FEV_1 rs11234757



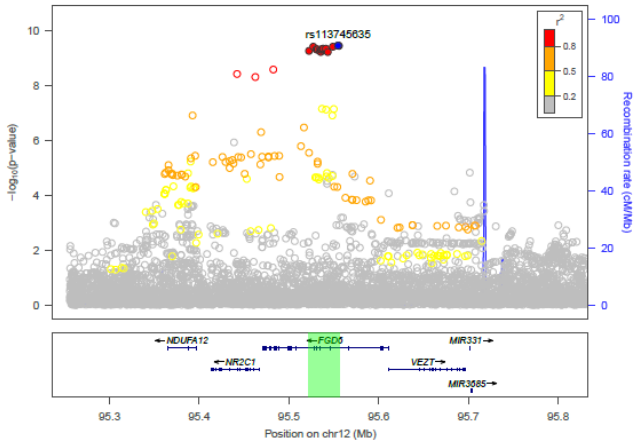
RPUSD4 FEV_1 rs567508



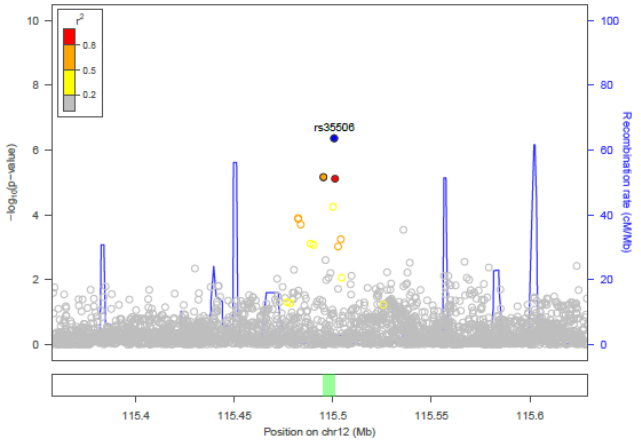
MSRB3 FEV$_1$ rs1494502



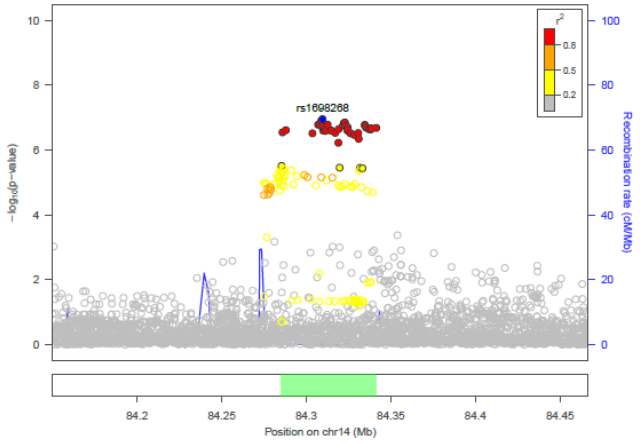
FGD6 FEV$_1$/FVC rs113745635



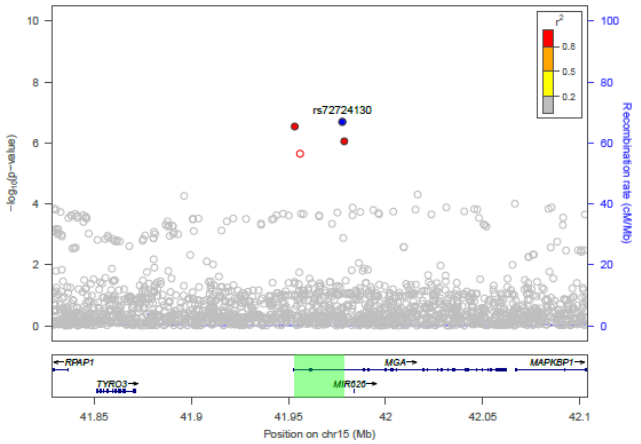
TBX3 FVC rs35506



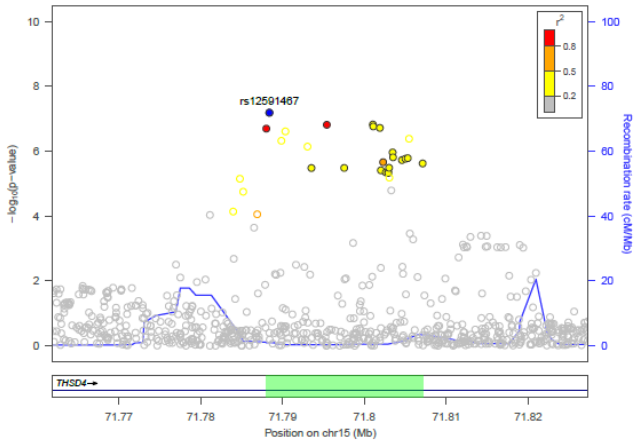
LINC00911 FEV$_1$/FVC rs1698268



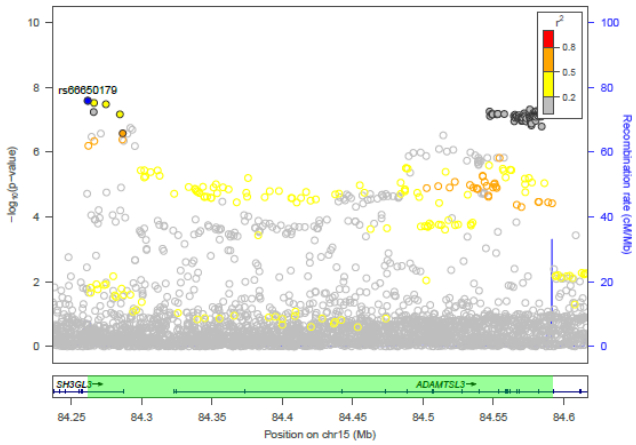
MGA FEV$_1$/FVC rs72724130



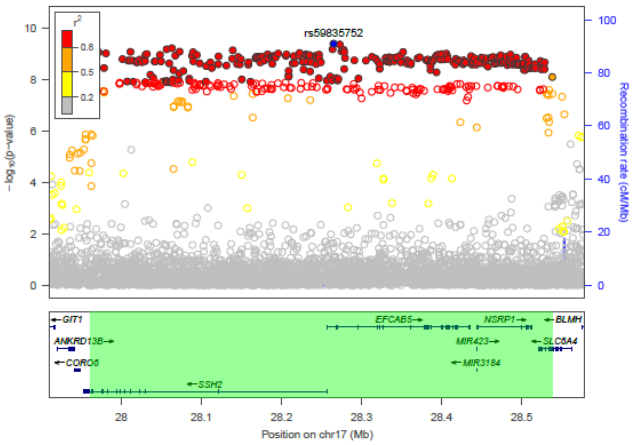
THSD4 FEV$_1$/FVC rs12591467
(conditioned on rs10851839)



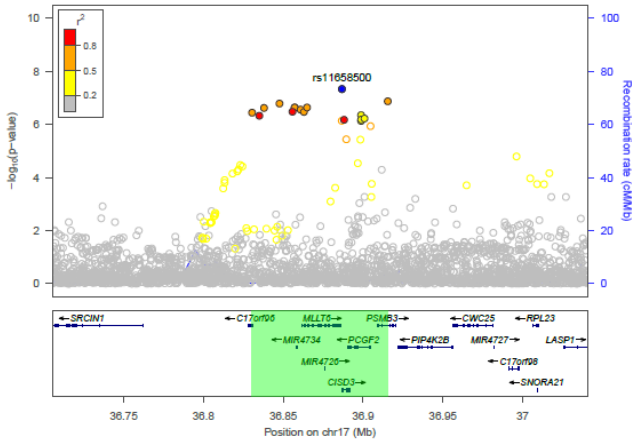
SH3GL3 FEV$_1$/FVC rs66650179



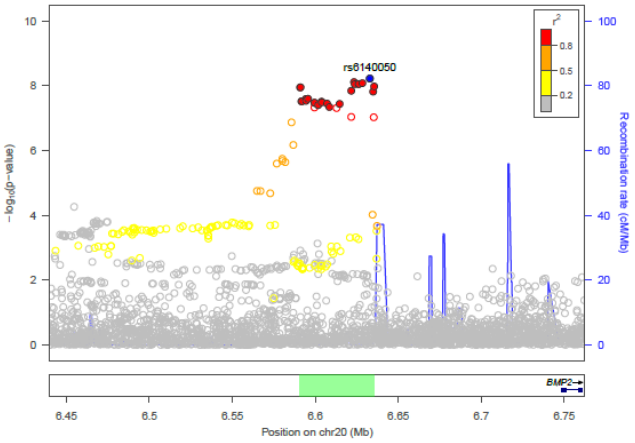
EFCAB5 FEV$_1$/FVC rs59835752



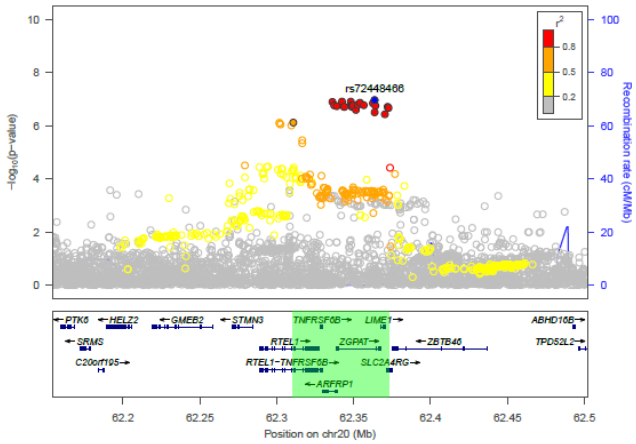
CISD3 FEV$_1$/FVC rs11658500



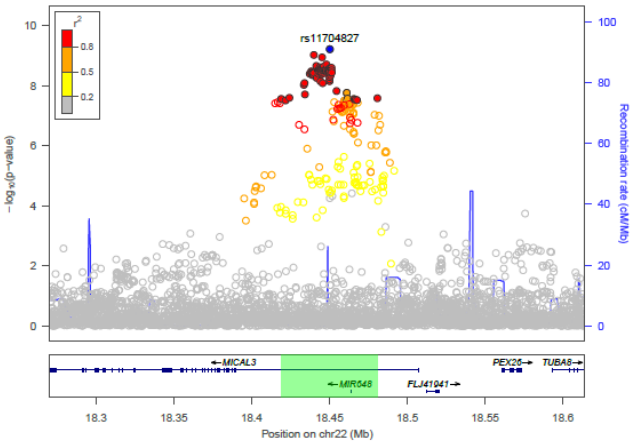
BMP2 FVC rs6140050



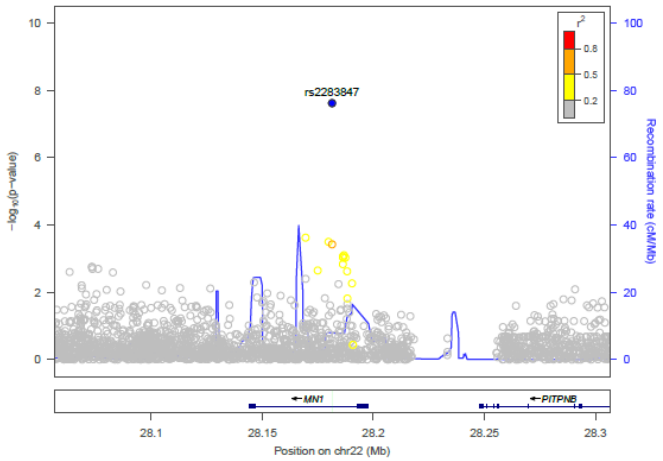
ZGPAT FEV$_1$ rs72448466



MICAL3 FEV$_1$ rs11704827

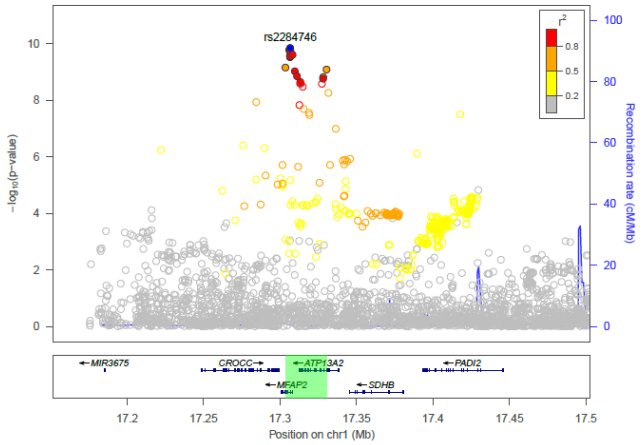


$MN1$ FEV_1 rs2283847

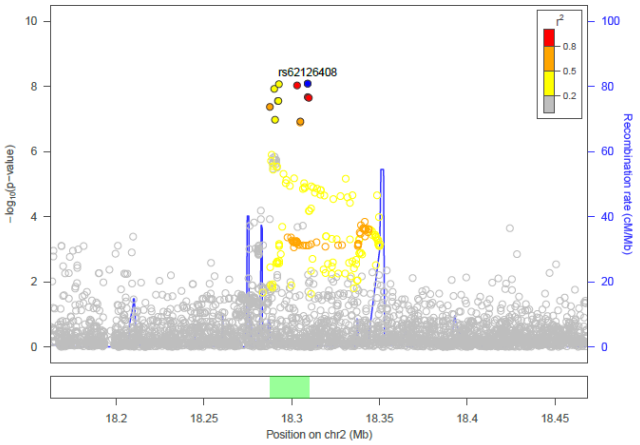


Supplementary Figure 5: Region plots with credible sets shown for 26 previously-reported signals that reached P <10^{-5} in stage 1 in this study and are not in the HLA region. Variants in the 95% credible set are shown as filled circles, those not in the credible set as open circles with the span of the credible set shaded in green on the gene track below. Where a "conditioned on" variant is given, the previously discovered signal is conditioned on a novel secondary signal.

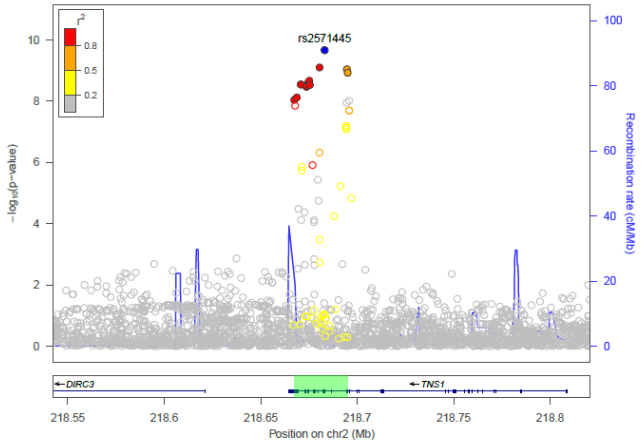
MFAP2 FEV_1/FVC rs2284746



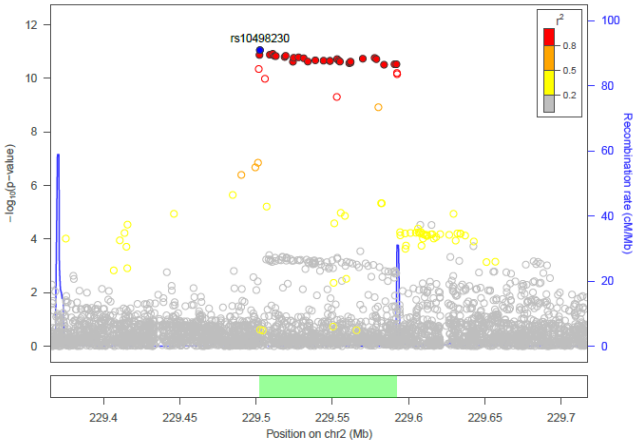
KCNS3 FEV_1/FVC rs62126408



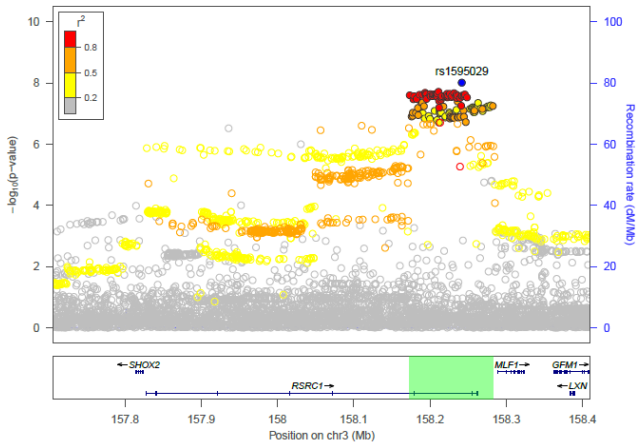
TNS1 FEV_1 rs2571445



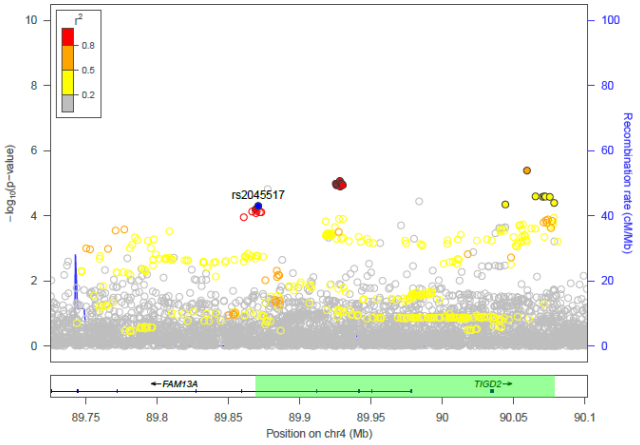
PID1 FEV_1/FVC rs10498230



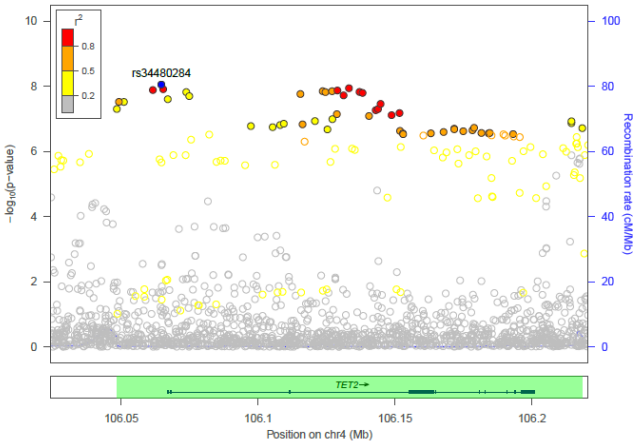
RP11 – 538P18.2 FVC rs1595029



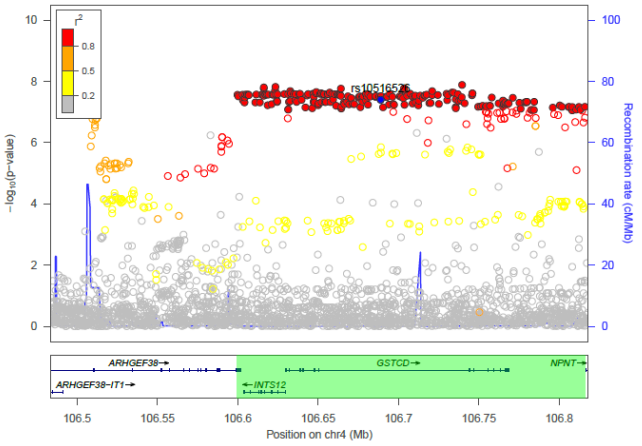
FAM13A FEV_1/FVC rs2045517
(conditioned on rs13110699)



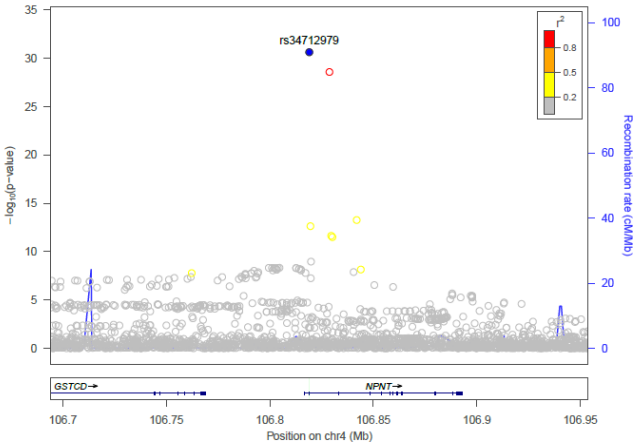
$TET2$ FEV_1 rs34480284



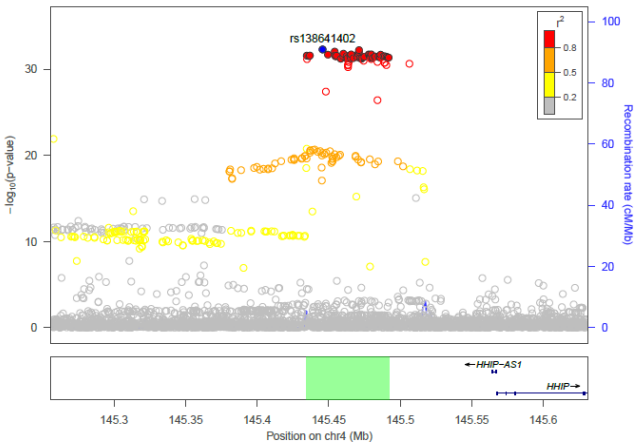
$GSTCD$ FEV_1 rs10516526



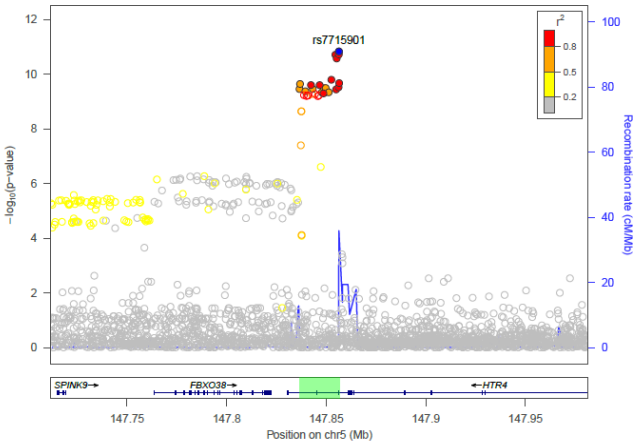
$NPNT$ FEV_1/FVC rs34712979



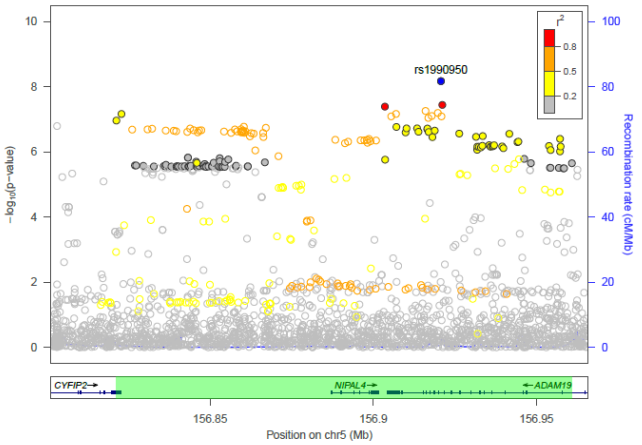
$HHIP$ FEV_1/FVC rs138641402



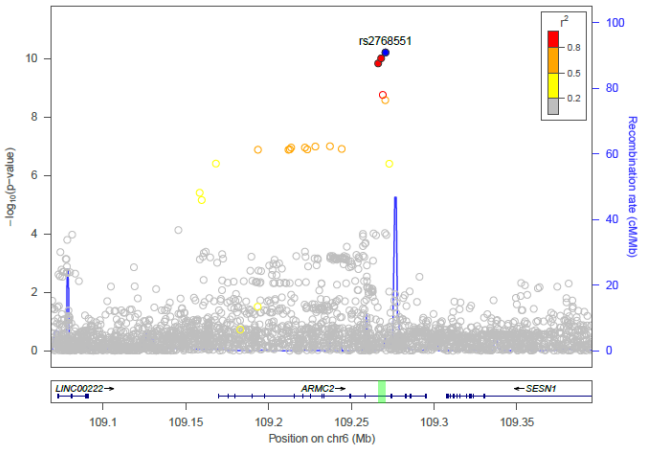
$HTR4$ FEV_1 rs7715901



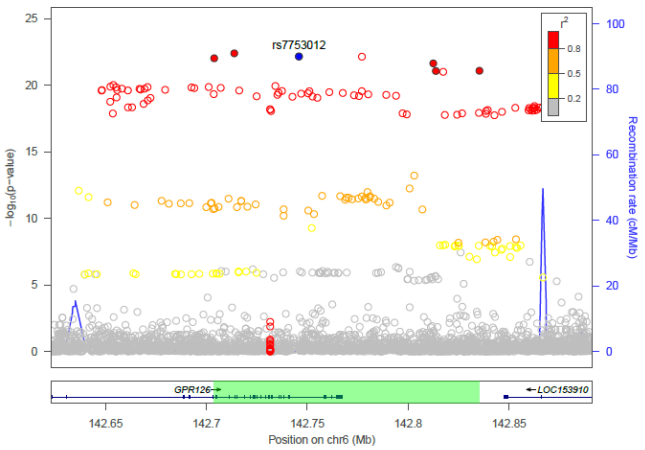
$ADAM19$ FEV_1/FVC rs1990950
(conditioned on rs10515750)



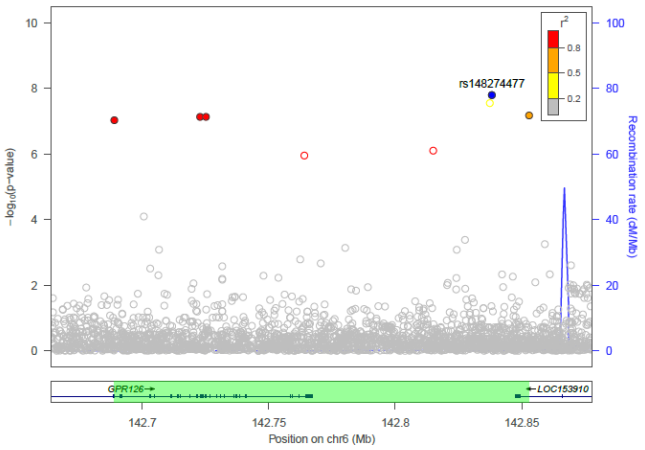
ARMC2 FEV_1/FVC rs2768551



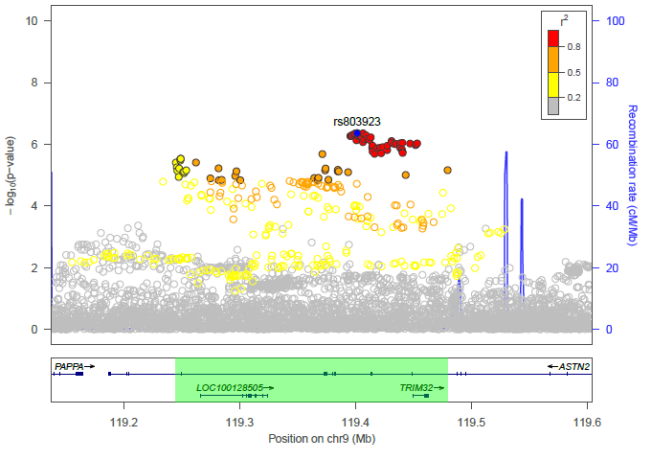
LOC153910 FEV_1/FVC rs7753012
(conditioned on rs148274477)



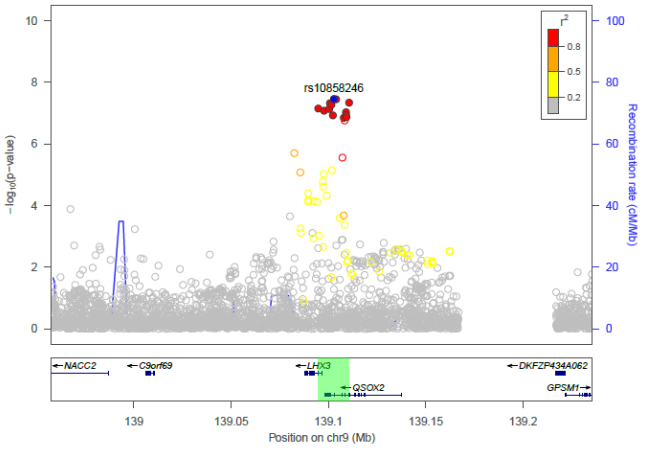
GPR126 FEV_1/FVC rs148274477
(conditioned on rs7753012)



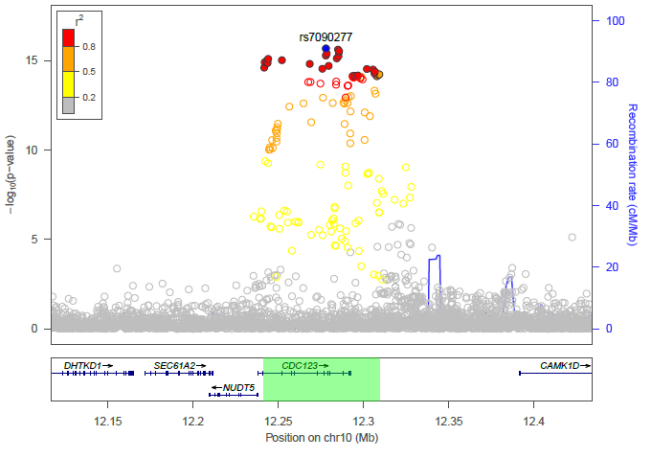
ASTN2 FEV_1/FVC rs803923



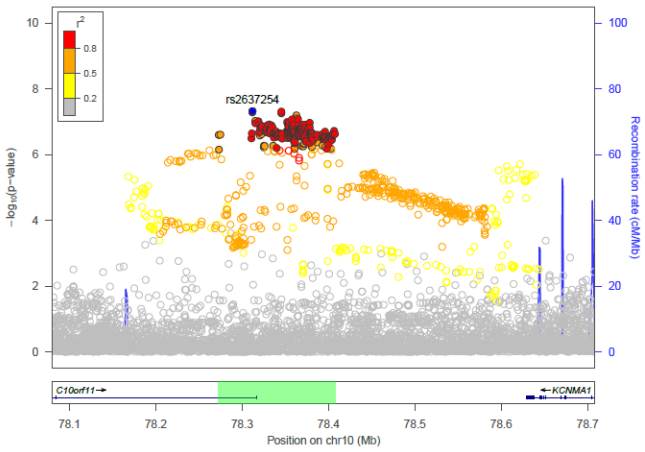
LHX3 FVC rs10858246



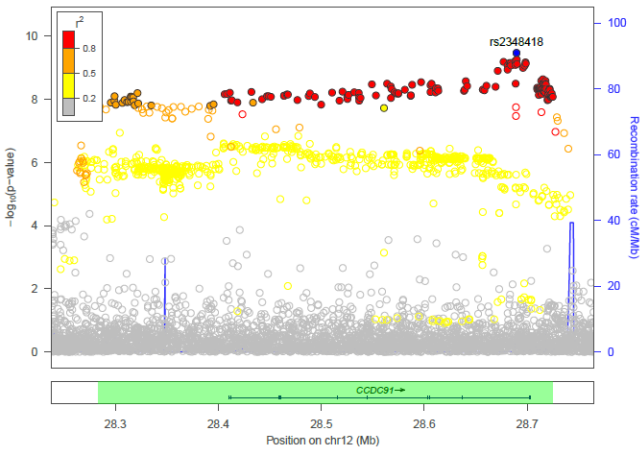
CDC123 FEV_1/FVC rs7090277



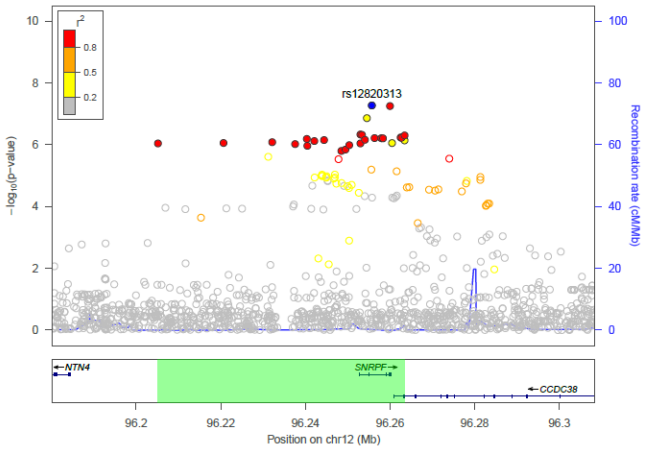
$C10orf11$ FEV_1 rs2637254



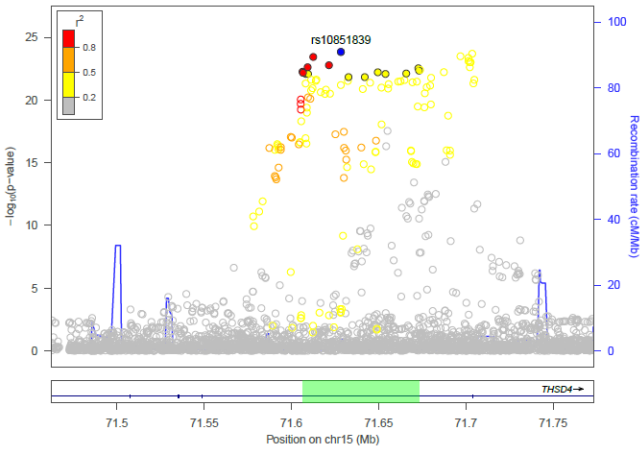
$CCDC91$ FVC rs2348418



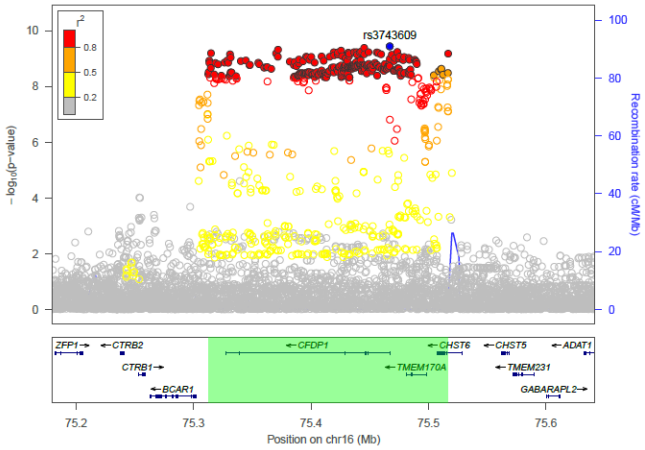
$CCDC38$ FEV_1/FVC rs12820313



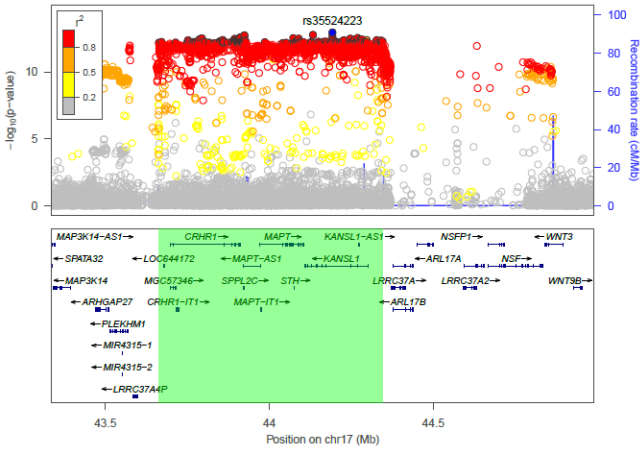
$THSD4$ FEV_1/FVC rs10851839



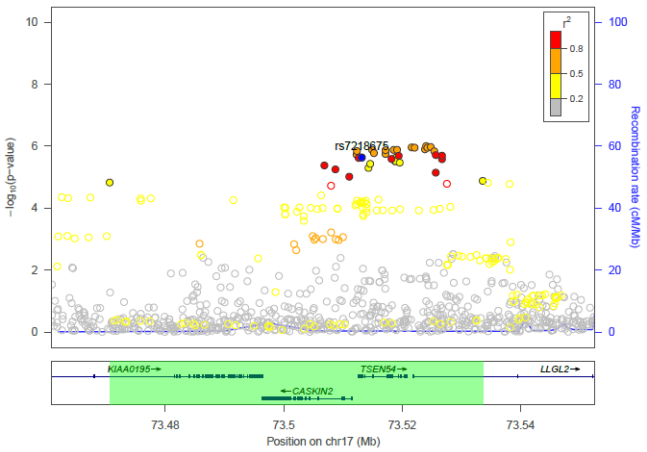
$CFDP1$ FEV_1/FVC rs3743609



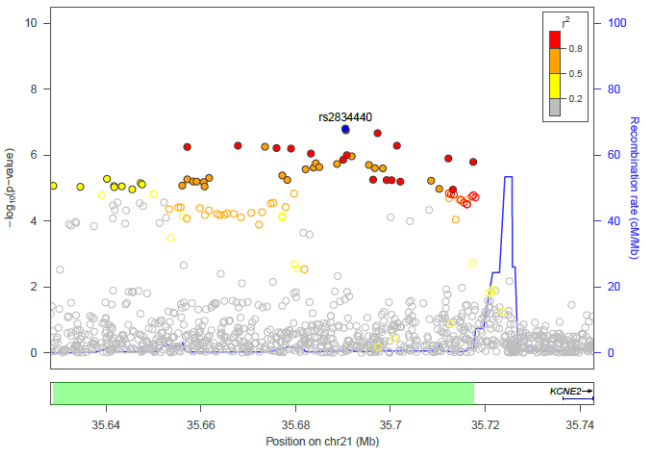
$KANSL1$ FEV_1 rs35524223



$TSEN54$ FEV_1 rs7218675

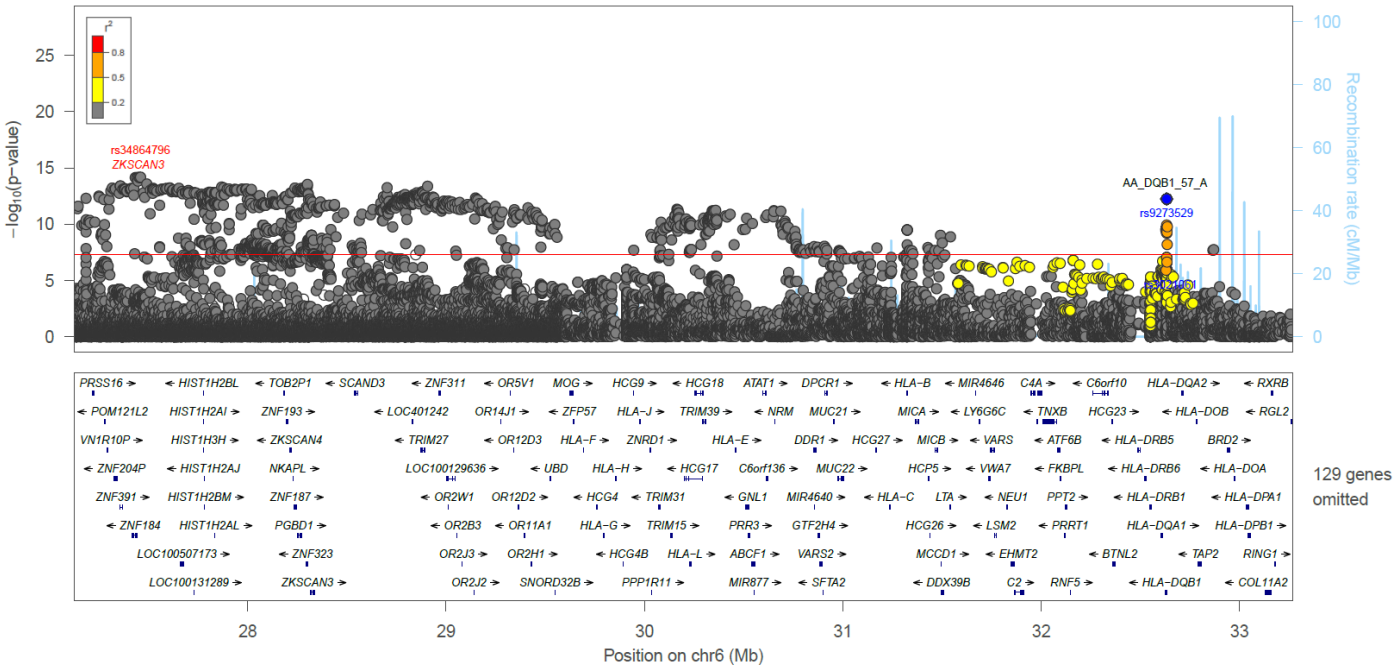


$KCNE2$ FEV_1/FVC rs2834440

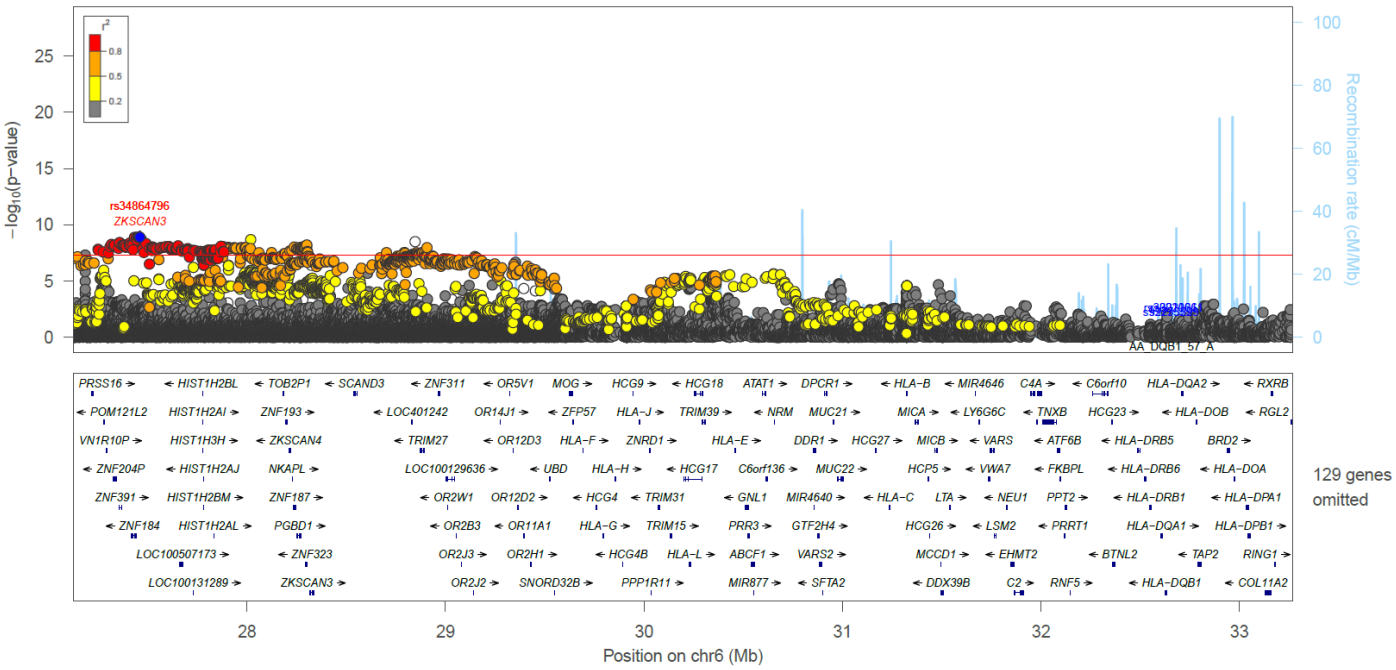


Supplementary Figure 6: Region plots for imputation of HLA haplotypes and amino acids. Results are shown for FEV_1 (a and b) and FEV_1/FVC (c and d) both before and after conditioning on HLA-DQβ1 amino acid position 57.

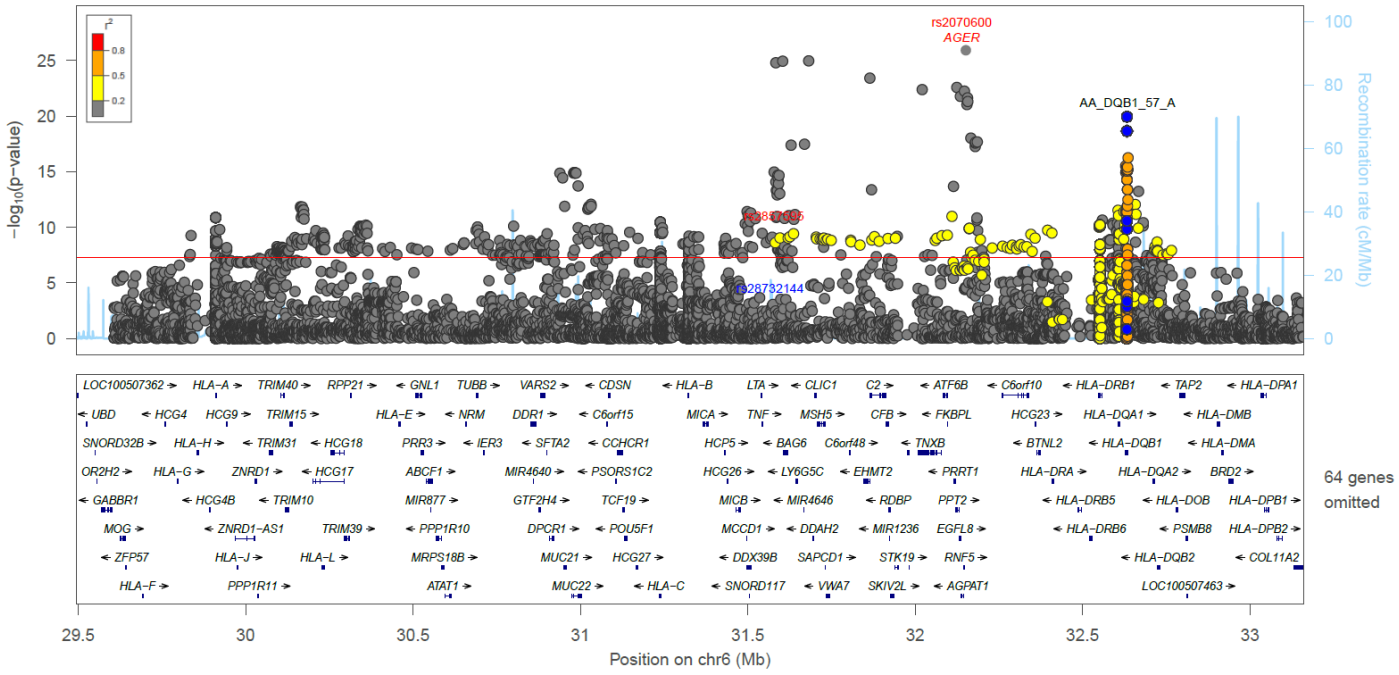
a) FEV_1 (no conditioning)



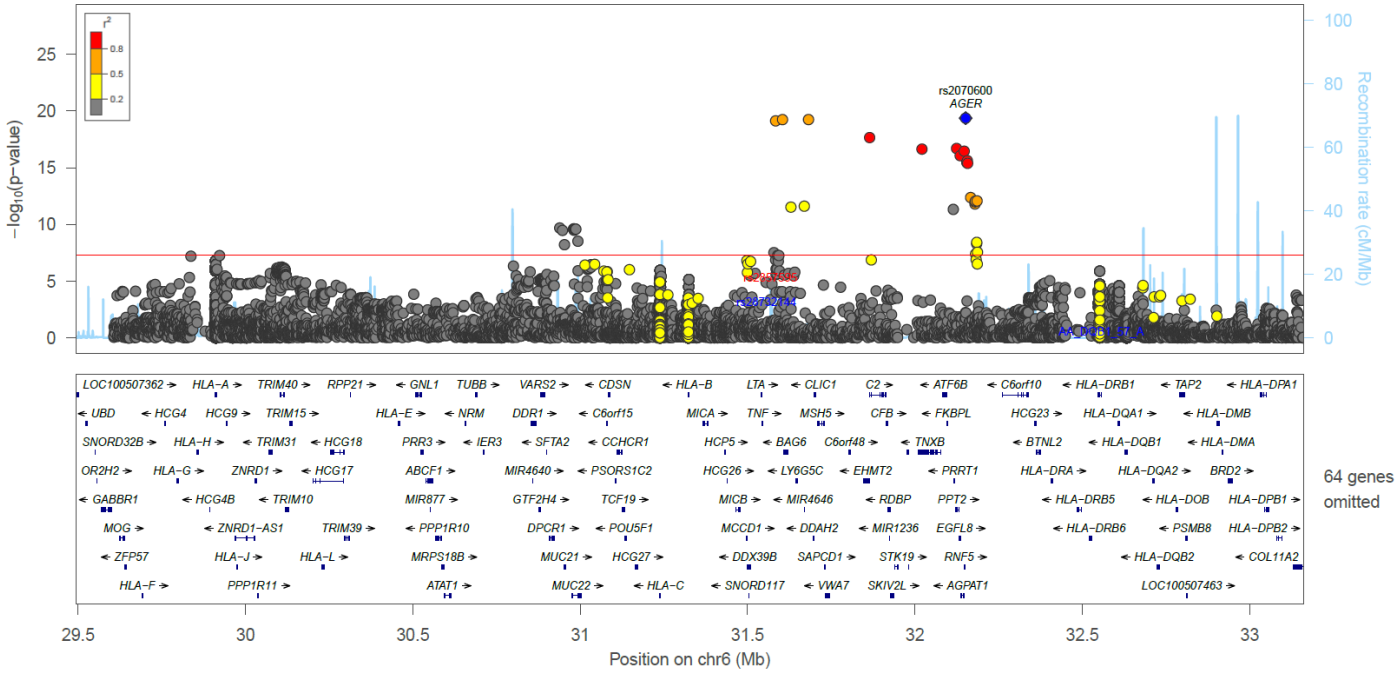
b) FEV_1 conditioned on HLA-DQβ1 amino acid position 57



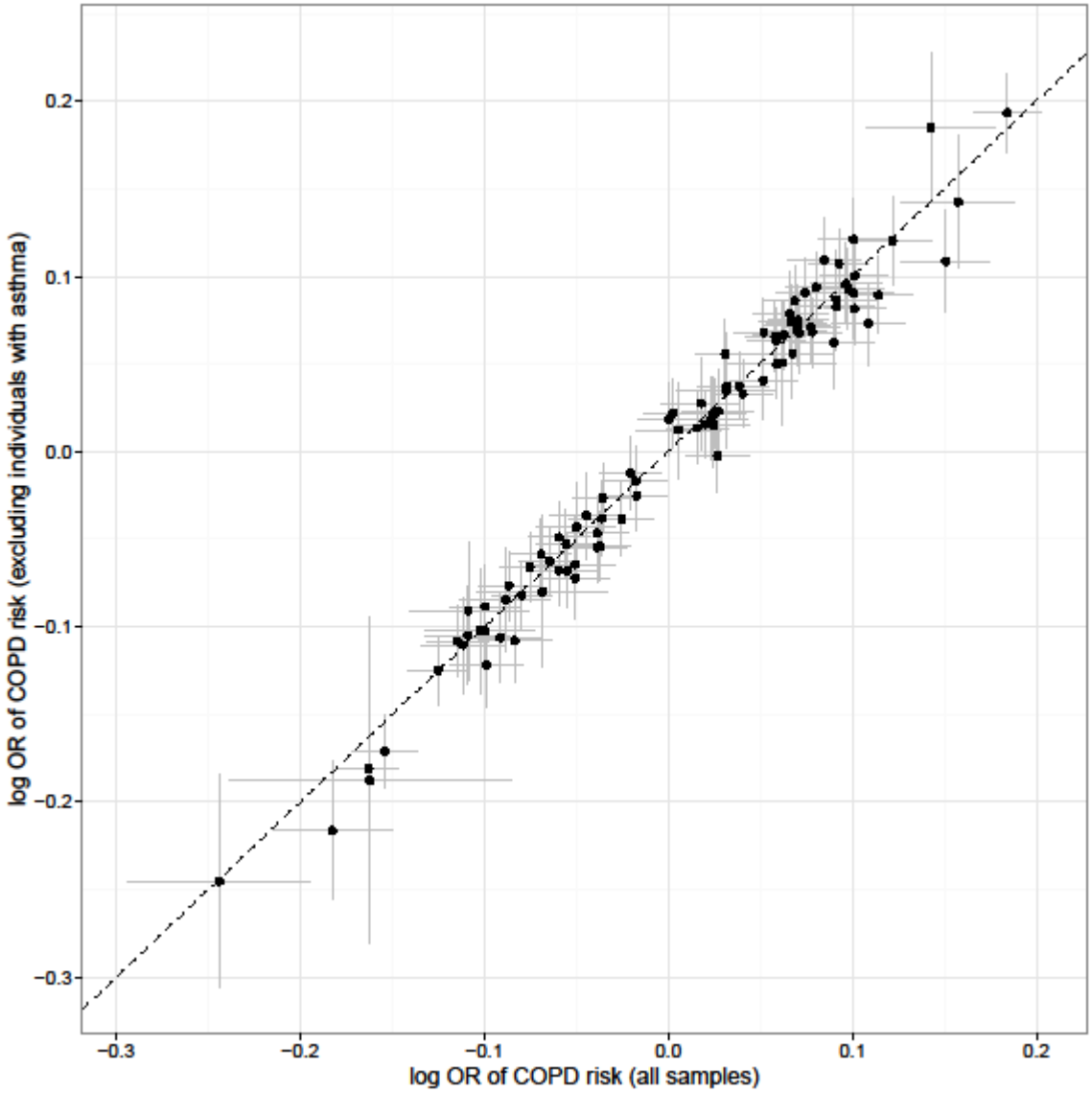
c) FEV_1/FVC (no conditioning)



d) FEV_1/FVC conditioned on HLA-DQβ1 amino acid position 57

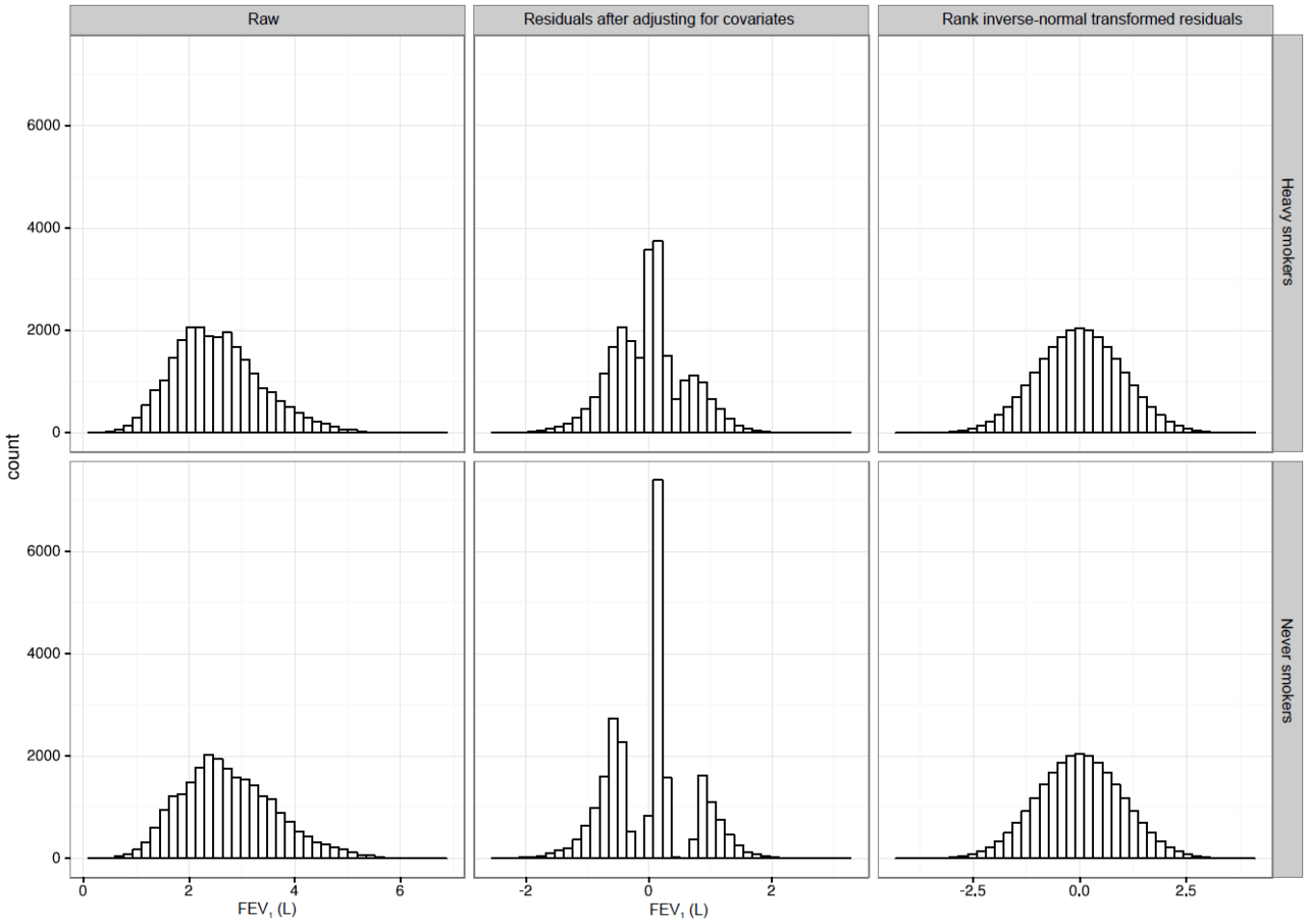


Supplementary Figure 7: Log odds ratio of COPD risk in UK Biobank samples excluding individuals with a doctor diagnosis of asthma (n=56,195) vs. log odds ratio of COPD risk in all available UK Biobank samples (n=64,484) for 97 lung function signals. Error bars are the standard errors of the effect estimates.

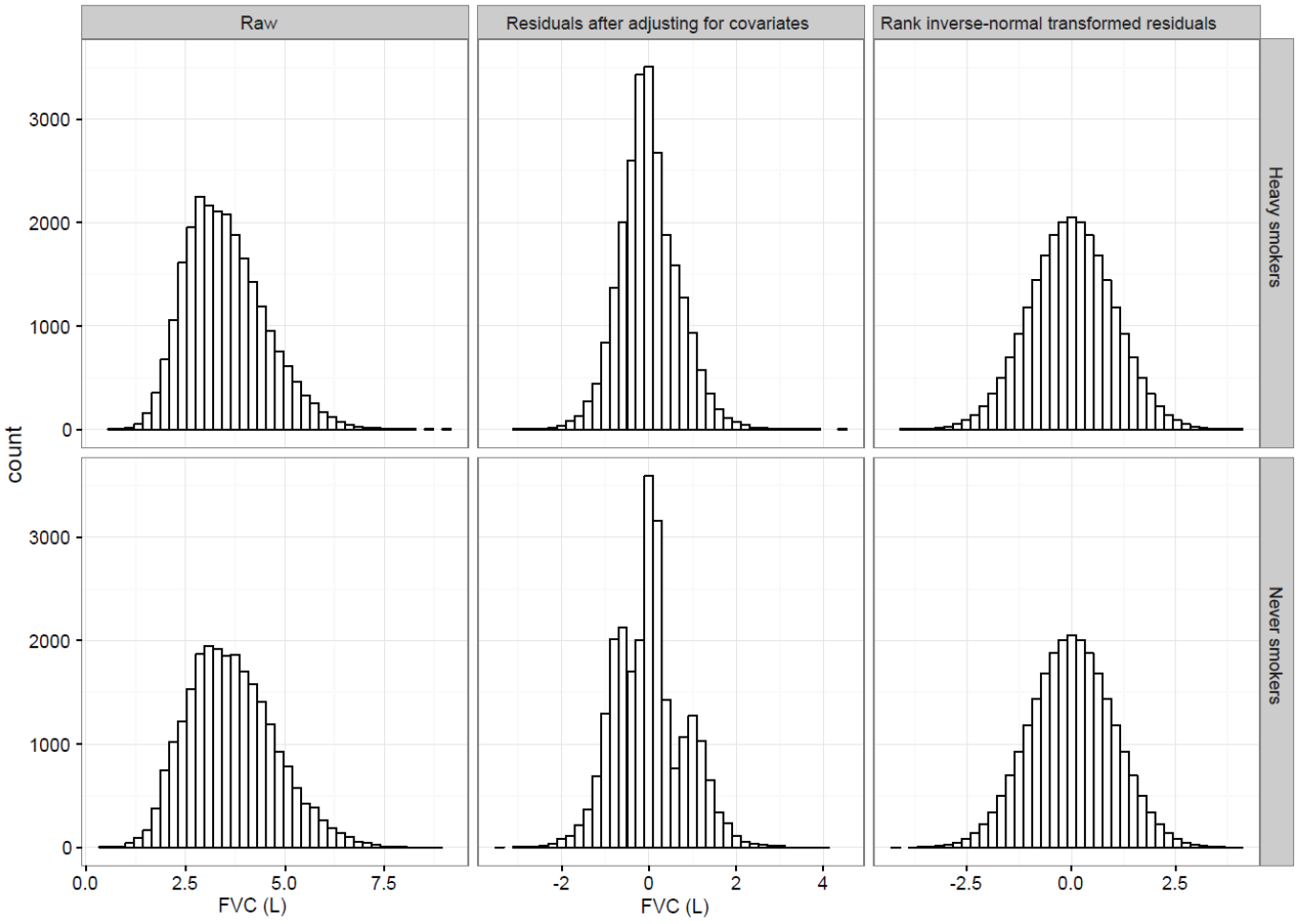


Supplementary Figure 8: Distribution of a) FEV_1, b) FVC and c) FEV_1/FVC in stage 1 (UK BiLEVE) for 48,493 stage 1 samples. Plots show distributions before adjustment (Raw), residuals after adjusting for covariates (age, age^2, sex, height and first 10 ancestry principal components) and residuals after rank inverse-normal transformation. Data are presented separately for heavy (top row) and never smokers (bottom row).

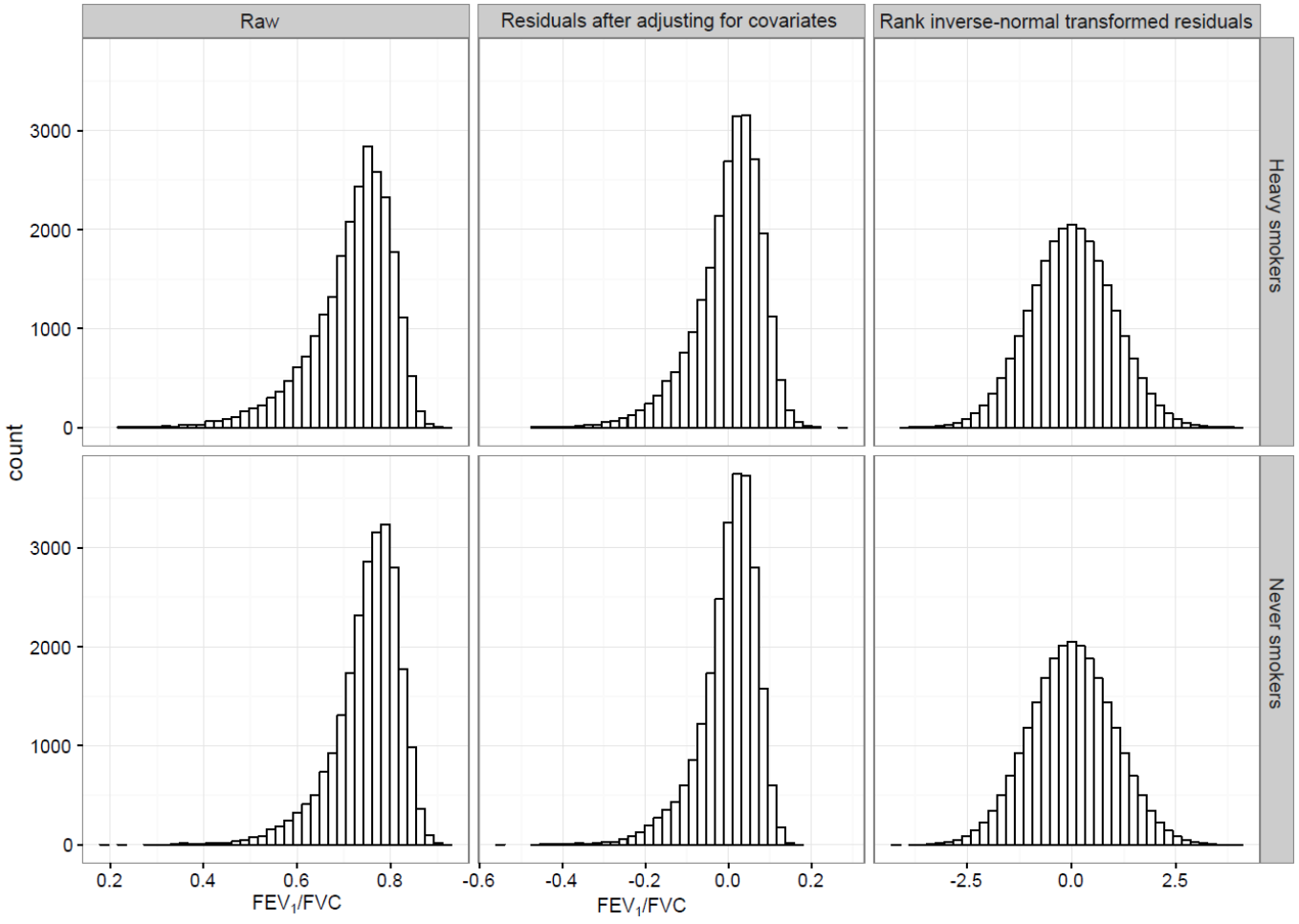
a) FEV_1



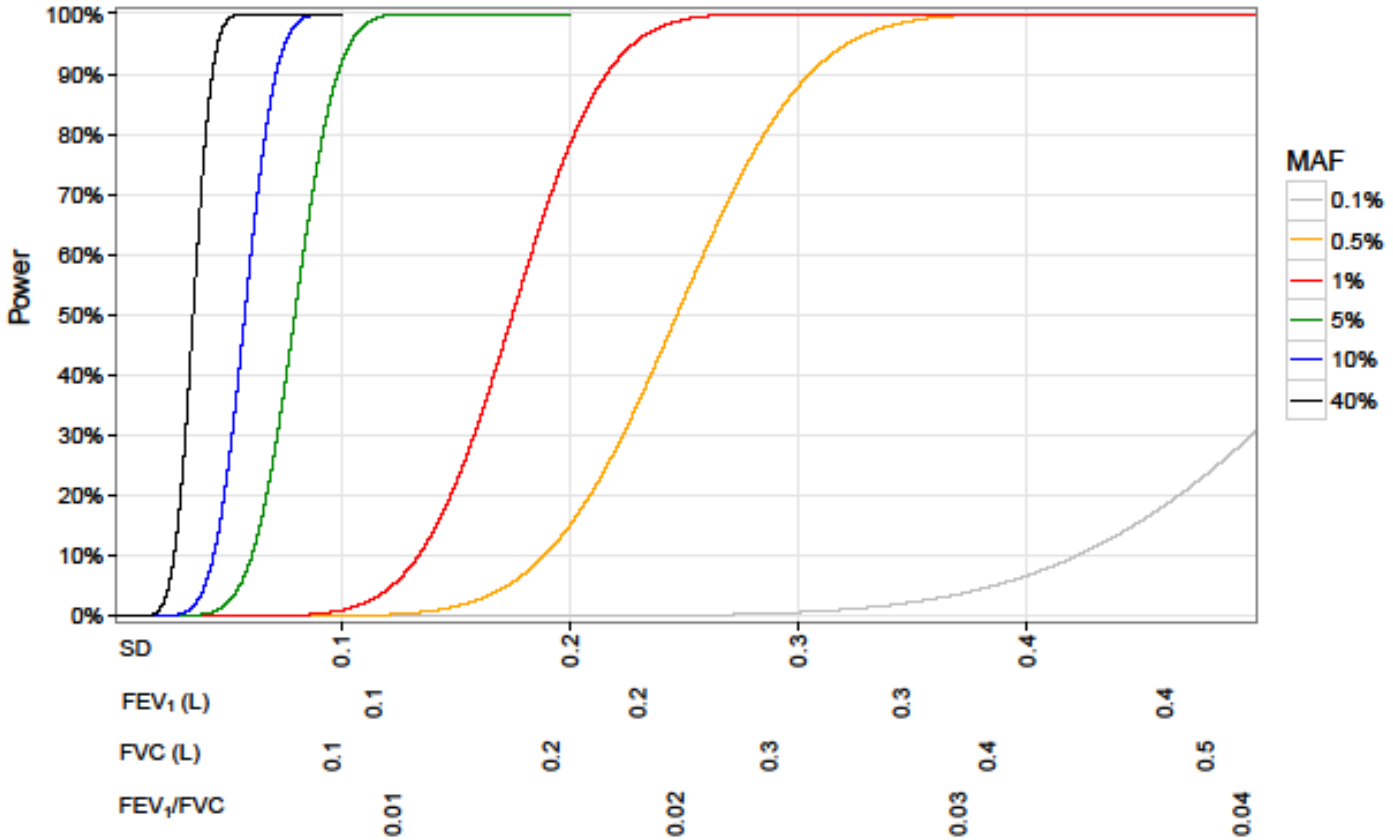
b) FVC



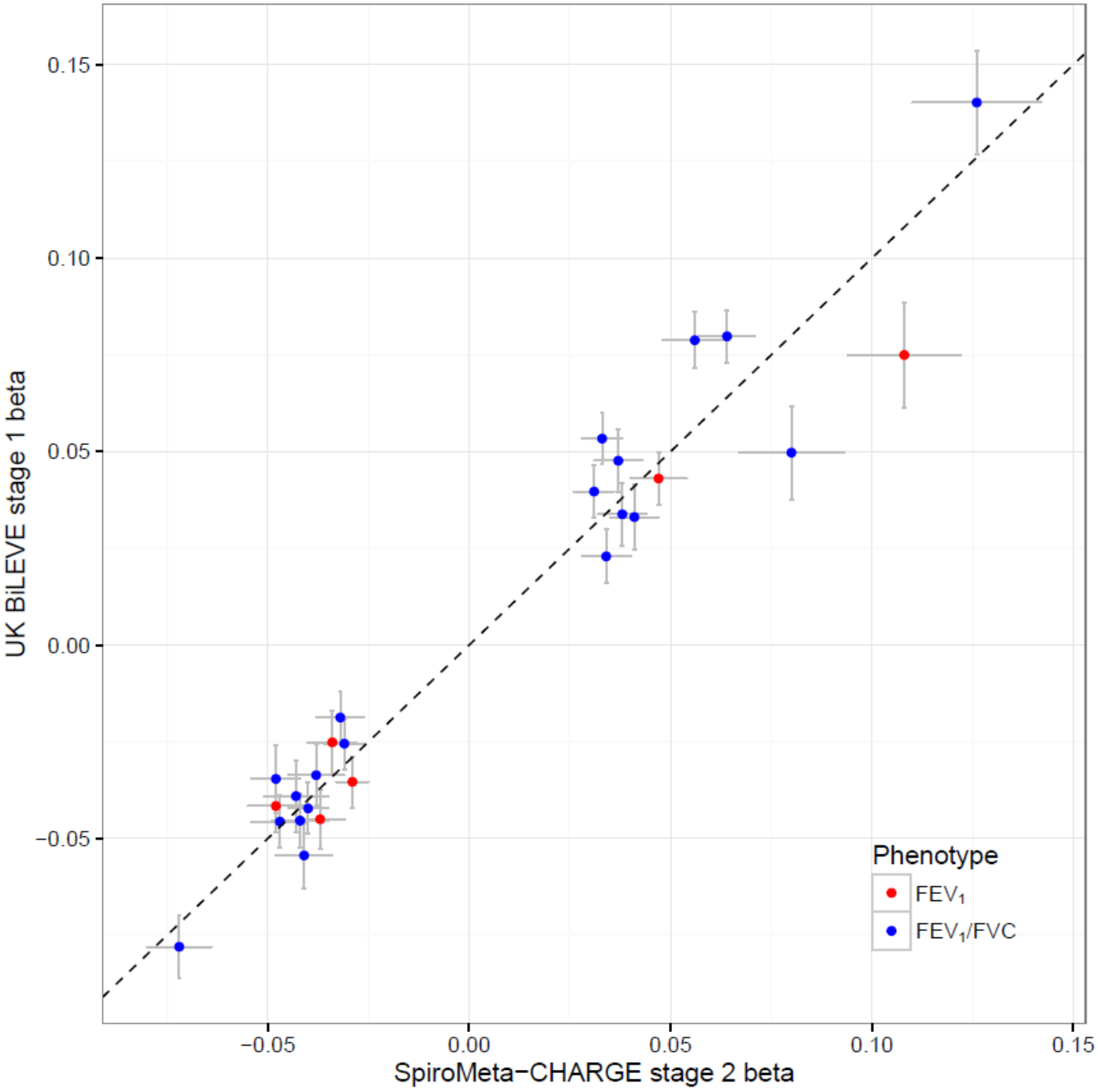
c) FEV_1/FVC



Supplementary Figure 9: Power calculations. Statistical power (y-axis) for detecting genome-wide significant association under an additive genetic model in a population of size 48,493 for varying minor allele frequency (MAF, coloured lines) and effect sizes (x-axis). Simplifying assumptions have been utilised to produce conservative estimates. A single stage design in a population drawn from a general population at random and a P-value threshold $5x10^{-8}$ is assumed. Power would be expected to be greater with enrichment for extremes values of a quantitative outcome variable, and with a higher p-value threshold and follow-up in an independent population. A study with such conservative assumptions applied would be powered to detect variants of and MAF≥5% and modest effect size (e.g. power >90% at MAF 5% and effect size 0.1 SD) and powered to detect lower frequency variants that have a larger effect size (e.g. power >75% for MAF 1% and effect size 0.2 SD).



Supplementary Figure 10: Comparison of effect estimates between SpiroMeta-CHARGE stage 2[40] and UK BiLEVE stage 1 for 26 variants reported for lung function before UK BiLEVE. Error bars are the standard errors of the effect estimates. Betas are quantiles of normal distribution (phenotypes rank inverse-normal transformed).



Supplementary Tables

Supplementary Table 1: Summaries of stage 1 (UK BiLVE) and stage 2 (UK Biobank, SpiroMeta and UKHLS) studies. *Details of all 17 studies that contributed to SpiroMeta can be found in Soler Artigas *et al* 2015[41]

Study Name	Smoking group	Lung function group	n	n (%) Male	Smokers, n (%)	Age, mean (SD)	FEV_1, litres. mean (SD)	FVC, litres. mean (SD)	FEV_1/FVC, mean (SD)
Stage 1									
UK BiLEVE	All		48,943	24,489 (50.0%)	24,460 (50.0%)	56.9 (7.89)	2.65 (0.87)	3.59 (1.05)	0.733 (0.081)
	Heavy smokers	High	4,907	2,459 (50.1%)	4,907 (100%)	56.9 (7.90)	3.49 (0.72)	4.49 (0.96)	0.778 (0.044)
		Average	9,803	4,908 (50.1%)	9,803 (100%)	56.9 (7.89)	2.68 (0.56)	3.62 (0.78)	0.743 (0.054)
		Low	9,750	4,886 (50.1%)	9,750 (100%)	56.9 (7.88)	1.93 (0.55)	2.92 (0.75)	0.663 (0.096)
	Never smokers	High	4,902	2,457 (50.1%)	0	56.9 (7.90)	3.83 (0.73)	4.85 (0.95)	0.791 (0.041)
		Average	9,831	4,905 (49.9%)	0	56.9 (7.89)	2.92 (0.57)	3.81 (0.79)	0.769 (0.047)
		Low	9,750	4,874 (50.0%)	0	56.9 (7.88)	2.05 (0.54)	2.92 (0.79)	0.707 (0.084)
Stage 2									
UK Biobank			49,727	20,682 (41.6%)	31,952 (64.3%)	56.4 (7.95)	2.85 (0.71)	3.75 (0.91)	0.762 (0.055)
SpiroMeta*			38,199	*	*	*	*	*	*
UKHLS			7,449	3,293 (44.2%)	4,509 (60.5%)	53.10 (15.94)	2.89 (0.90)	3.83 (1.08)	0.753 (0.090)

Supplementary Table 2: LD score regression analysis to estimate extent of overlap between SpiroMeta (stage 2) and the two UK Biobank subsets; UK BiLEVE (stage 1) and UK Biobank (stage 2). Results for the regression of each trait FEV_1, FVC and FEV_1/FVC against the LD score of each variant are shown**.** Total Observed scale h2: Estimate of heritability, Lambda GC: Usual lambda used for genomic control: inflation due to both confounding and polygenicity, Mean χ^2: Mean χ^2 statistic from the association testing, Intercept: Intercept of the LD score regression (estimate of inflation due to confounding but not polygenicity; suggested as a more appropriate genomic-control factor), Ratio: Proportion of total inflation due to confounding (Intercept-1)/(Mean χ^2-1). 95% confidence intervals are shown in brackets. A) Meta-analysis of UK BiLEVE (stage 1) and UK Biobank (stage 2) shown for comparison as overlapping samples were excluded. B) Meta-analysis of UK BiLEVE and SpiroMeta, C) Meta-analysis of UK Biobank and SpiroMeta, D) Genetic covariance intercept (95% C.I.) for bivariate LD score regression

A) Meta-analysis of UK BiLEVE (stage 1) and UK Biobank (stage 2):

N = 98,670	FEV_1	FVC	FEV_1/FVC
Total Observed scale h2	0.212 (0.187, 0.236)	0.209 (0.186, 0.233)	0.230 (0.198, 0.263)
Lambda GC	1.344	1.372	1.331
Mean χ^2	1.498	1.496	1.548
Intercept	1.040 (1.018, 1.062)	1.049 (1.025, 1.072)	1.055 (1.030, 1.079)
Ratio	0.080 (0.036, 0.124)	0.098 (0.050, 0.146)	0.100.055, 0.144)

B) Meta-analysis of UK BiLEVE and SpiroMeta

N = 87,142	FEV_1	FVC	FEV_1/FVC
Total Observed scale h2	0.208 (0.184, 0.233)	0.210 (0.186, 0.234)	0.185 (0.157, 0.213)
Lambda GC	1.297	1.313	1.24
Mean χ^2	1.427	1.419	1.371
Intercept	1.036 (1.016, 1.055)	1.026 (1.006, 1.046)	1.025 (1.002, 1.048)
Ratio	0.084 (0.039, 0.128)	0.062 (0.015, 0.110)	0.67 .006, 0.129)

C) Meta-analysis of UK Biobank and SpiroMeta

N = 87,926	FEV_1	FVC	FEV_1/FVC
Total Observed scale h2	0.158 (0.136, 0.179)	0.157 (0.136, 0.178)	0.169 (0.142, 0.196)
Lambda GC	1.25	1.25	1.236
Mean χ^2	1.325	1.326	1.356
Intercept	1.029 (1.008, 1.050)	1.031 (1.010, 1.052)	1.038 (1.018, 1.059)
Ratio	0.088 (0.024, 0.152)	0.096 (0.032, 0.160)	0.108.050, 0.166)

D) Genetic covariance intercept (95% C.I.) for bivariate LD score regression

	FEV_1	FVC	FEV_1/FVC
UK BiLEVE & UK Biobank	0.008 (-0.008, 0.023)	0.021 (0.005, 0.036)	0.007 (-0.011, 0.026)
UK BiLEVE & SpiroMeta	0.012 (-0.002, 0.026)	0.006 (-0.008, 0.021)	0.001 (-0.014, 0.015)
UK Biobank & SpiroMeta	0.009 (-0.005, 0.024)	0.013 (-0.000, 0.026)	0.007 (-0.007, 0.022)

Supplementary Table 3: Full results for all 81 variants followed up in stage 2. The 81 variants showing suggestive association ($P < 5x10^{-7}$) with a lung function quantitative trait in discovery, their lung function association results in stage 1 and stage 2 studies separately, the results of the meta-analysis of the stage 2 studies and the meta-analysis of the stage 1 and stage 2 studies are shown. The 43 variants with $P < 5x10^{-8}$ following meta-analysis of Stage 1 and Stage 2 are presented first (sorted by chromosome and position), followed by the remaining 38 signals with $P > 5x10^{-8}$ following meta-analysis of Stage 1 and Stage 2. Values are missing from stage 2 studies where there was quality control failure due to poor imputation (info < 0.5) or low minor allele count (MAC < 3). Where the discovery variant was not available in replication cohorts but a proxy with $r^2 > 0.7$ was available, the proxy was used for replication in all cohorts (proxies are marked with * in the list of discovery variants). For discovery the standard errors and P values are genomic control (GC) corrected except for conditional analyses ("Conditioned on" column non-empty) where unadjusted standard errors and P values are given. GC corrected results were used for SpiroMeta 1000 genomes. Unadjusted results are used for UK Biobank and UKHLS where genome-wide inflation factors were not available. In the meta-analysis of the Stage 2 replication cohorts the 39 variants showing independent replication (Bonferroni correction for 81 tests: $P < 6.17x10^{-4}$) have P value in bold. In the meta-analysis of the discovery and replication stages (Stage 1 + 2) the variants showing genome-wide significant association ($P < 5x10^{-8}$) have P value in bold.

See accompanying Excel file.

Supplementary Table 4: Stage 1 results for 97 variants associated with lung function (all traits). The 97 variants showing association with lung function comprising (a) 43 novel variants and (b) 54 previously-reported variants (the most significant variant in this study for the previously reported signal is given). Association results are from the discovery stage (48,943 UK BiLEVE samples). In (a), the trait for which the variant showed the most significant association is given in the "trait" column and the effect and P value for the reported trait is in bold. In (b), the trait for which the variant was previously reported as showing the most significant association is given in the "trait" column and the effect and P value for the reported trait is in bold. The effect estimate beta is on the inverse-normal rank scale, standard errors and P values are Genomic Control (GC) corrected for unconditional association results. In (a), the variant upon which the association was conditioned is given in the "Conditioned on" column (conditional results are not GC corrected). The nearest genes, or location of variant within the gene, is indicated. In (b), the published study that first reported the signal is given. *The listed gene is the gene name used to describe that signal in the previous study publication. References for previous studies are as follows: Wilk *et al* (2009)[42], Repapi *et al* (2010)[43], Hancock *et al* (2010)[44], Soler Artigas *et al* (2011)[40], Loth *et al* (2014)[45], Wain *et al* (2015)[46], Soler Artigas *et al* (2015)[41].

See accompanying Excel file.

Supplementary Table 5: Bayesian estimation of 95% credible sets. A summary of the number of variants in the 95% credible sets for the novel association signals and the previous signals having association $P < 10^{-5}$. The table includes the number of variants in the credible set, the top ranked variant and its posterior probability. The posterior probabilities and the credible sets were calculated as described in Wakefield[47]. Six HLA signals, 1 chromosome X signal and 23 previously-reported signals with $P > 10^{-5}$ could not be refined using this method resulting in sets being defined for 41 novel signals and 26 previously-reported signals. Conditional results were used for rs1192404 (conditioned on rs12140637), rs13110699 (rs2045517), rs2045517 (rs13110699), rs10515750 (rs1990950), rs1990950 (rs10515750), rs7753012 (rs148274477) and rs148274477 (rs7753012). The posterior probabilities of rs2045517 (rank: 20), rs10516526 (114), rs7753012 (2) and rs7218675 (20) are 0.01316, 0.00404, 0.1959 and 0.0214 respectively.

Sentinel variant ID and Genomic position	Locus	Number of variants in credible set	Trait	Nearest genes to Sentinel variant	Top ranked variant (Posterior probability)
Novel signals					
rs17513135 chr1: 40035686	Chr 1: 39527963-40113043	104	FEV_1/FVC	*LOC101929516* (intron)	Sentinel (0.09118)
rs1192404 chr1: 92068967	Chr 1: 92016515-92112240	12	FEV_1/FVC	*CDC7/TGFBR3*	Sentinel (0.149)
rs12140637 chr1: 92374517	Chr 1:92330156-92472668	12	FEV_1/FVC	*TGFBR3/BRDT*	Sentinel (0.1021)
rs200154334 chr1: 118862070	Chr 1:118824762-118942956	21	FVC	*SPAG17/TBX15*	Sentinel (0.2355)
rs6688537 chr1: 239850588	Chr 1:239773921-239939160	60	FEV_1/FVC	*CHRM3* (intron)	Sentinel (0.0523)
rs61332075 chr2: 239316560	Chr 2:239198478-239500420	115	FEV_1/FVC	*TRAF3IP1/ASB1*	Sentinel (0.2538)
rs1458979 chr3: 55150677	Chr 3:55124454-55183751	7	FEV_1/FVC	*CACNA2D3/WNT5A*	Sentinel (0.2813)
rs1490265 chr3: 67452043	Chr 3:67406108-67481222	16	FVC	*SUCLG2* (intron)	Sentinel (0.1378)
rs2811415 chr3: 127991527	Chr 3:127688264-128092441	197	FEV_1/FVC	*EEFSEC* (intron)	Sentinel (0.01469)
esv2660202 chr3: 168738454	Chr 3:168635231-168885010	119	FEV_1/FVC	*LOC100507661/MECOM*	Sentinel (0.03174)
rs13110699 chr4: 89815695	Chr 4:89775892-89959645	43	FEV_1/FVC	*FAM13A* (intron)	Sentinel (0.0874)
rs91731 chr5: 33334312	Chr 5:33182002-33424894	52	FVC	*LOC340113/TARS*	Sentinel (0.04772)
rs1551943 chr5: 52195033	Chr 5:52152346-52257838	9	FEV_1/FVC	*ITGA1* (intron)	Sentinel (0.3193)
rs2441026 chr5: 53444498	Chr 5:53419498-53518744	20	FVC	*ARL15* (intron)	Sentinel (0.4559)
rs7713065 chr5: 131788334	Chr 5:131723241-131834757	36	FEV_1/FVC	*C5orf56* (intron)	Sentinel (0.07636)
rs3839234 chr5: 148596693	Chr 5:148568202-148677363	33	FEV_1	*ABLIM3* (intron)	Sentinel (0.1756)
rs10515750 chr5: 156810072	Chr 5:156611712-156970148	47	FEV_1/FVC	*CYFIP2* (intron)	Sentinel (0.05234)
rs28986170 chr6: 31556155	Chr 6:31296753-32229882	HLA	FEV_1/FVC	*LST1* (intron)	HLA
rs114229351 chr6: 32648418	Chr 6:32512879-32693100	HLA	FEV_1	*HLA-DQB1/HLA-DQA2*	HLA
rs141651520 chr6: 73670095	Chr 6:73630333-73744982	7	FEV_1/FVC	*KCNQ5* (intron)	Sentinel (0.1527)
rs10246303 chr7: 7286445	Chr 7:7196968-7311445	18	FEV_1/FVC	*C1GALT1* (3′ UTR)	Sentinel (0.136)

Sentinel variant ID and Genomic position	Locus	Number of variants in credible set	Trait	Nearest genes to Sentinel variant	Top ranked variant (Posterior probability)
rs72615157 chr7: 99635967	Chr 7:99608739-99874854	36	FEV_1/FVC	*ZKSCAN1* (3' UTR)	Sentinel (0.306)
rs12698403 chr7: 156127246	Chr 7:156080037-156159055	7	FEV_1	*LOC389602/LOC285889*	Sentinel (0.2177)
rs7872188 chr9: 4124377	Chr 9:4094707-4173531	24	FEV_1	*GLIS3* (intron)	Sentinel (0.1887)
rs10870202 chr9: 139257411	Chr 9:139213707-139343071	9	FVC	*DNLZ* (intron)	Sentinel (0.4887)
rs3847402 chr10: 30267810	Chr 10:30222165-30306732	58	FEV_1/FVC	*SVIL/KIAA1462*	Sentinel (0.03702)
rs7095607 chr10: 69957350	Chr 10:69887278-69990177	61	FVC	*MYPN* (intron)	Sentinel (0.03546)
rs2509961 chr11: 62310909	Chr 11:62284787-62443921	78	FEV_1	*AHNAK* (intron)	Sentinel (0.04564)
rs11234757 chr11: 86443072	Chr 11:86403024-86557868	14	FEV_1	*ME3/PRSS23*	Sentinel (0.1066)
rs567508 chr11: 126008910	Chr 11:125983910-126053787	9	FEV_1	*CDON/RPUSD4*	Sentinel (0.3015)
rs1494502 chr12: 65824670	Chr 12:65730543-65867258	39	FEV_1	*MSRB3* (intron)	Sentinel (0.05955)
rs113745635 chr12: 95554771	Chr 12:95336610-95733206	18	FEV_1/FVC	*FGD6* (intron)	Sentinel (0.07072)
rs35506 chr12: 115500691	Chr 12:115457443-115529071	4	FVC	*TBX3/MED13L*	Sentinel (0.819)
rs1698268 chr14: 84309664	Chr 14:84250124-84366454	40	FEV_1/FVC	*LINC00911*	Sentinel (0.04836)
rs72724130 chr15: 41977690	Chr 15:41928211-42003725	3	FEV_1/FVC	*MGA* (intron)	Sentinel (0.4877)
rs12591467 chr15: 71788387	Chr 15:71761905-71827290	20	FEV_1/FVC	*THSD4* (intron)	Sentinel (0.3553)
rs66650179 chr15: 84261689	Chr 15:84236689-84616675	105	FEV_1/FVC	*SH3GL3* (intron)	Sentinel (0.0299)
rs59835752 chr17: 28265330	Chr 17:27910546-28578639	273	FEV_1/FVC	*EFCAB5* (intron)	Sentinel (0.01471)
rs11658500 chr17: 36886828	Chr 17:36805562-36940540	17	FEV_1/FVC	*CISD3* (intron)	Sentinel (0.2799)
rs6140050 chr20: 6632901	Chr 20:6539919-6662234	24	FVC	*CASC20/BMP2*	Sentinel (0.09918)
rs72448466 chr20: 62363640	Chr 20:62254332-62401939	24	FEV_1	*ZGPAT* (intron)	Sentinel (0.06342)
rs11704827 chr22: 18450287	Chr 22:18370241-18513883	84	FEV_1	*MICAL3* (intron)	Sentinel (0.06432)
rs2283847 chr22: 28181399	Chr 22:28156399-28206436	1	FEV_1	*LINC01422/MN1*	Sentinel (1)
Previously-reported lung function signals					
rs2284746 chr1: 17306675	Chr1: 17251627-17402956	15	FEV_1/FVC	*MFAP2* (intron)	Sentinel (0.1464)
rs62126408 chr2: 18309132	Chr 2:18262623-18368845	11	FEV_1/FVC	*KCNS3/RDH14*	Sentinel (0.1967)
rs2571445 chr2: 218683154	Chr 2:218642372-218720848	14	FEV_1	*TNS1* (exon)	Sentinel (0.3905)
rs10498230 chr2: 229502503	Chr 2:229465307-229617415	29	FEV_1/FVC	*SPHKAP/PID1*	Sentinel (0.06795)
rs1595029 chr3: 158241767	Chr 3:157805916-158310280	121	FVC	*RSRC1* (intron)	Sentinel (0.03169)

Sentinel variant ID and Genomic position	**Locus**	**Number of variants in credible set**	**Trait**	**Nearest genes to Sentinel variant**	**Top ranked variant (Posterior probability)**
rs2045517 chr4: 89870964	Chr 4:89725361-90102090	21	FEV_1/FVC	*FAM13A* (intron)	rs6828137 (0.1448)
rs34480284 chr4: 106064626	Chr 4:106024147-106220572	51	FEV_1	*LOC101929468/TET2*	Sentinel (0.07098)
rs10516526 chr4: 106688904	Chr 4:106483526-106818063	209	FEV_1	*GSTCD* (intron)	rs10516528 (0.006794)
rs34712979 chr4: 106819053	Chr 4:106794053-106853795	1	FEV_1/FVC	*NPNT* (intron)	Sentinel (0.9913)
rs138641402 chr4: 145445779	Chr 4:145355633-145531456	48	FEV_1/FVC	*GYPA/HHIP-AS1*	Sentinel (0.09656)
rs7715901 chr5: 147856392	Chr 5:147811609-147881522	22	FEV_1	*HTR4* (intron)	Sentinel (0.1958)
rs1990950 chr5: 156920756	Chr 5:156801152-156965873	103	FEV_1/FVC	*ADAM19* (intron)	Sentinel (0.3326)
rs34864796 chr6: 27459923	Chr 6:26437104-28478618	HLA	FEV_1	*ZNF184/LINC01012*	HLA
rs2857595 chr6: 31568469	Chr 6:31263877-31943860	HLA	FEV_1/FVC	*NCR3/AIF1*	HLA
rs2070600 chr6: 32151443	Chr 6:31558841-32210605	HLA	FEV_1/FVC	*AGER* (exon)	HLA
rs114544105 chr6: 32635629	Chr 6:32084979-32671184	HLA	FEV_1	*HLA-DQB1/HLA-DQA2*	HLA
rs2768551 chr6: 109270656	Chr 6:109168639-109295656	3	FEV_1/FVC	*ARMC2* (intron)	Sentinel (0.4661)
rs7753012 chr6: 142745883	Chr 6:142623056-142891387	7	FEV_1/FVC	*GPR126* (intron)	rs6570508 (0.2339)
rs148274477 chr6: 142838173	Chr 6:142663969-142877897	5	FEV_1/FVC	*GPR126/LOC153910*	Sentinel (0.5099)
rs803923 chr9: 119401650	Chr 9:119237495-119504774	78	FEV_1/FVC	*ASTN2* (intron)	Sentinel (0.03569)
rs10858246 chr9: 139102831	Chr 9:139057491-139135654	13	FVC	*QSOX2* (intron)	Sentinel (0.1345)
rs7090277 chr10: 12278021	Chr 10:12216815-12334390	31	FEV_1/FVC	*CDC123* (intron)	Sentinel (0.1363)
rs2637254 chr10: 78312002	Chr 10:78180071-78608611	224	FEV_1	*C10orf11* (intron)	Sentinel (0.01745)
rs2348418 chr12: 28689514	Chr 12:28237880-28764845	152	FVC	*CCDC91* (intron)	Sentinel (0.05737)
rs12820313 chr12: 96255704	Chr 12:96180161-96308432	26	FEV_1/FVC	*SNRPF* (intron)	Sentinel (0.2313)
rs10851839 chr15: 71628370	Chr 15:71562373-71673497	15	FEV_1/FVC	*THSD4* (intron)	Sentinel (0.5145)
rs3743609 chr16: 75467021	Chr 16:75279623-75541739	270	FEV_1/FVC	*CFDP1* (intron)	Sentinel (0.01521)
rs35524223 chr17: 44192590	Chr 17:43435181-44890603	279	FEV_1	*KANSL1* (intron)	Sentinel (0.01611)
rs7218675 chr17: 73513185	Chr 17:73460781-73552560	34	FEV_1	*TSEN54* (intron)	rs146301005 (0.05408)
rs2834440 chr21: 35690499	Chr 21:35628304-35742962	48	FEV_1/FVC	*LINC00310/KCNE2*	Sentinel (0.1445)

Supplementary Table 6: Association results for the 6 previously reported MHC region GWAS signals before and after conditioning on HLA-DQβ1 amino acid position 57. Unconditional P values and standard errors are Genomic Control corrected. P values in bold meet genome-wide significance ($P<5x10^{-8}$).

a) FEV_1

MHC signal	Chr:pos	FEV_1 beta	FEV_1 se	FEV_1 P	FEV_1 (conditioned on HLA-DQβ1 amino acid position 57) beta	se	P
rs34864796 (*ZKSCAN3*)	6:27459923	-0.074	0.010	**6.14E-14**	-0.058	0.010	**1.26E-09**
rs28986170* (*LST1*)	6:31556155	0.056	0.013	3.07E-05	0.042	0.013	1.74E-03
rs2857595 (*NCR3*)	6:31568469	-0.039	0.008	2.05E-06	-0.023	0.008	3.52E-03
rs2070600 (*AGER*)	6:32151443	0.039	0.014	4.15E-03	0.023	0.013	7.32E-02
rs114544105 (*HLA-DQB1*)	6:32635629	-0.049	0.008	**8.84E-11**	-0.006	0.007	4.04E-01
rs114229351† (*HLA-DQB1*)	6:32648418	-0.046	0.009	**1.15E-07**	-0.015	0.009	7.75E-02

b) FEV_1/FVC

MHC signal	Chr:pos	FEV_1/FVC beta	FEV_1/FVC se	FEV_1/FVC P	FEV_1/FVC (conditioned on HLA-DQβ1 amino acid position 57) beta	se	P
rs34864796 (*ZKSCAN3*)	6:27459923	-0.062	0.010	**3.52E-10**	-0.041	0.010	2.07E-05
rs28986170* (*LST1*)	6:31556155	0.077	0.013	**1.23E-08**	0.065	0.013	1.11E-06
rs2857595 (*NCR3*)	6:31568469	-0.048	0.008	**3.50E-09**	-0.028	0.008	4.27E-04
rs2070600 (*AGER*)	6:32151443	0.140	0.014	**3.11E-25**	0.120	0.013	**4.23E-20**
rs114544105 (*HLA-DQB1*)	6:32635629	-0.063	0.008	**5.20E-17**	-0.008	0.007	2.96E-01
rs114229351† (*HLA-DQB1*)	6:32648418	-0.050	0.009	**6.79E-09**	-0.006	0.009	5.20E-01

*****Already conditioned on rs2070600 & rs201002132.
†Already conditioned on rs34864796.

Supplementary Table 7: GRASP and/or GWAS Catalog-reported genome-wide associations for the 97 lung function signals. *Where signals for which a credible set was not defined, variants within 2Mb and LD $r^2 \geq 0.8$ were used to query the databases. The previously reported signals of association with COPD and lung function are not shown. For signals associated with height, the consistency of direction of effect on lung function with height is indicated for all 3 traits (FEV_1, FVC, FEV_1/FVC), where "+" indicates that the allele associated with increased height is also associated with an increase in the lung function trait and "-" indicates that the allele associated with increased height is associated with decreased lung function.

Trait	Sentinel lung function association SNP	Locus name	GWAS catalog/GRASP reported trait(s)
Novel signals			
$\underline{FEV_1}$ FVC	rs17513135 chr1:40035686	*LOC101929516*	HDL cholesterol, C-reactive protein levels, Mean corpuscular hemoglobin, Triglycerides
$\underline{FEV_1}$ FVC	rs1192404 chr1: 92068967	*CDC7-TGFBR3*	Optic disc area, Vertical cup disc ratio, PC2 (Disc area), FAC2 (Disc area, cup shape measure, and oppositely directed rim to disc area ratio and linear cup to disc ratio)
FVC	rs200154334 chr1:118862070	*SPAG17-TBX15*	Height (---), Infant length, Height tails (upper and lower 5th percentiles)
$\underline{FEV_1}$ FVC	rs61332075 chr2:239316560	*TRAF3IP1-ASB1*	Iris furrow contractions
$\underline{FEV_1}$ FVC	rs13110699 chr4: 89815695	*FAM13A*	Fibrotic idiopathic interstitial pneumonias (pulmonary fibrosis)
$\underline{FEV_1}$ FVC	rs7713065 chr5: 131788334	*C5orf56*	Juvenile idiopathic arthritis (including oligoarticular and rheumatoid factor negative polyarticular JIA), Crohn's disease
$\underline{FEV_1}$ FVC	rs10515750 chr5: 156810072	*CYFIP2*	Bipolar disorder and schizophrenia, Bipolar disorder (body mass index interaction), Several serum metabolites
FVC	rs10870202 chr9: 139257411	*DNLZ*	Inflammatory bowel disease (Crohn's disease & Ulcerative colitis), IgA nephropathy
FVC	rs7095607 chr10: 69957350	*MYPN*	Height (---)
FEV_1	rs1494502 chr12: 65824670	*MSRB3*	Temperament
$\underline{FEV_1}$ FVC	rs66650179 chr15: 84261689	*SH3GL3*	Height (+++)
$\underline{FEV_1}$ FVC	rs59835752 chr17: 28265330	*EFCAB5*	Coffee consumption (cups per day), Psoriasis (HLA-C risk allele negative)
FVC	rs6140050 chr20: 6632901	*CASC20-BMP2*	Height (--+), Waist to hip ratio adjusted for body mass index, Sitting height ratio
FEV_1	rs72448466 chr20: 62363640	*ZGPAT*	Inflammatory bowel disease (Crohn's disease & Ulcerative colitis), Prostate cancer, Atopic dermatitis
FEV_1	rs11704827 chr22: 18450287	*MICAL3*	Liver enzyme levels (gamma glutamyl transferase), Presence of antiphospholipid antibodies
Previously-reported lung function signals			
$\underline{FEV_1}$ FVC	rs2284746 chr1:17306675	*MFAP2*	Height (adults, males and females) (-+-), Height tails (upper and lower 5th percentiles)
$\underline{FEV_1}$	rs993925*	*MIR548F3*	Acne (severe)

Trait	Sentinel lung function association SNP	Locus name	GWAS catalog/GRASP reported trait(s)
FVC	chr1: 218860068		
FVC	rs1595029 chr3: 158241767	*RSRC1*	Height (+++), Height tails (upper and lower 5th percentiles)
$\underline{FEV_1}$ FVC	rs2045517 chr4: 89870964	*FAM13A*	Fibrotic idiopathic interstitial pneumonias (pulmonary fibrosis)
FEV_1	rs34480284 chr4: 106064626	*TET2*	Prostate cancer
FEV_1	rs34864796* chr6: 27459923	*ZNF184-LINC01012*	Schizophrenia, Bipolar disorder
$\underline{FEV_1}$ FVC	rs2857595* chr6: 31568469	*NCR3-AIF1*	Type 1 Diabetes, Laryngeal squamous cell carcinoma
$\underline{FEV_1}$ FVC	rs7753012 chr6: 142745883	*GPR126*	Height (---), Scoliosis
$\underline{FEV_1}$ FVC	rs803923 chr9: 119401650	*ASTN2*	Hippocampal volume
$\underline{FEV_1}$ FVC	rs11172113* chr12: 57527283	*LRP1*	Cervical artery dissection, Migraine
FEV_1	rs7155279* chr14: 92485881	*TRIP11*	Height (---)
FEV_1	rs117068593* chr14: 93118229	*RIN3*	Bone mineral density (lower limb and total body less head), Paget's disease
FEV_1	rs35524223 chr17: 44192590	*KANSL1*	Parkinson's disease, Intracranial volume, Male pattern baldness, Subcortical brain region volumes, Ovarian cancer in BRCA1 mutation carriers, Epithelial ovarian cancer, Progressive supranuclear palsy, Hematocrit (Hct), Hemoglobin (Hb), Primary biliary cirrhosis, Fibrotic idiopathic interstitial pneumonias (pulmonary fibrosis)
$\underline{FEV_1}$ FVC	rs2834440 chr21: 35690499	*KCNE2*	Height (+-+), BMI

Supplementary Table 8: Look up for association with smoking behaviour for the 97 lung function variants. Smoking association results from a previously-reported study which compared 24,457 heavy-smokers vs. 24,474 never-smokers in UK BiLEVE[46]. One variant shows evidence of association with smoking behaviour using a 5% Bonferroni-corrected threshold for 97 tests ($P < 5.15x10^{-4}$, shown in bold). P values for smoking association are genomic-control corrected (λ=1.101) except where the association is conditioned on another variant. For the 5 novel variants with P<0.05 (*), a further look-up was undertaken in results from the TAG consortium study of smoking behaviour (PMID:20418890). Four traits were analysed: cigarettes per day, likelihood of smoking initiation, likelihood of quitting smoking and (log) age of onset. Associations (P<0.05) with smoking-related traits were observed for; rs72448466 (P=0.01, likelihood of quitting) and rs113745635 (P=0.02, age of onset of smoking). Both associations had a consistent direction of effect to that shown in the table below.

trait	rsid	Position b37	Gene	Coded Allele	Conditioned on	Smoking OR (95% C.I.)	Smoking P
43 novel variants							
$\frac{FEV_1}{FVC}$	rs17513135	1:40035686	*LOC101929516*	T		0.99 (0.96,1.03)	0.708
$\frac{FEV_1}{FVC}$	rs1192404	1:92068967	*TGFBR3*	G	rs12140637	1.03 (1.00,1.07)	0.053
$\frac{FEV_1}{FVC}$	rs12140637	1:92374517	*TGFBR3*	T		1.00 (0.97,1.03)	0.897
FVC	rs200154334	1:118862070	*SPAG17*	C		1.00 (0.97,1.03)	0.913
$\frac{FEV_1}{FVC}$	rs6688537	1:239850588	*CHRM3*	A		0.99 (0.96,1.02)	0.417
$\frac{FEV_1}{FVC}$	rs61332075	2:239316560	*TRAF3IP1*	C		1.01 (0.97,1.05)	0.627
$\frac{FEV_1}{FVC}$	rs1458979	3:55150677	*CACNA2D3*	G		0.98 (0.96,1.01)	0.243
FVC	rs1490265	3:67452043	*SUCLG2*	A		0.98 (0.95,1.01)	0.204
$\frac{FEV_1}{FVC}$	rs2811415	3:127991527	*EEFSEC*	G		1.01 (0.97,1.05)	0.609
$\frac{FEV_1}{FVC}$	esv2660202	3:168738454	*MECOM*	C		0.97 (0.94,1.00)	0.021*
$\frac{FEV_1}{FVC}$	rs13110699	4:89815695	*FAM13A*	G	rs2045517	1.00 (0.97,1.04)	0.813
FVC	rs91731	5:33334312	*TARS*	A		0.99 (0.95,1.04)	0.791
$\frac{FEV_1}{FVC}$	rs1551943	5:52195033	*ITGA1*	A		1.01 (0.97,1.04)	0.746
FVC	rs2441026	5:53444498	*ARL15*	T		1.01 (0.99,1.04)	0.297
$\frac{FEV_1}{FVC}$	rs7713065	5:131788334	*C5orf56*	C		1.03 (1.00,1.07)	0.029*
FEV_1	rs3839234	5:148596693	*ABLIM3*	T		1.00 (0.98,1.03)	0.781
$\frac{FEV_1}{FVC}$	rs10515750	5:156810072	*CYFIP2*	T	rs1990950	0.98 (0.93,1.03)	0.450
$\frac{FEV_1}{FVC}$	rs28986170	6:31556155	*LST1*	AA	rs2070600 rs201002132	1.00 (0.94,1.05)	0.889
FEV_1	rs114229351	6:32648418	*HLA-DQB1*	C	rs34864796	0.97 (0.94,1.01)	0.112
$\frac{FEV_1}{FVC}$	rs141651520	6:73670095	*KCNQ5*	A		1.00 (0.97,1.04)	0.852
$\frac{FEV_1}{FVC}$	rs10246303	7:7286445	*C1GALT1*	T		1.01 (0.98,1.04)	0.580
$\frac{FEV_1}{FVC}$	rs72615157	7:99635967	*ZKSCAN1*	A		1.02 (0.98,1.05)	0.371
FEV_1	rs12698403	7:156127246	*LOC285889*	A		0.98 (0.96,1.01)	0.224
FEV_1	rs7872188	9:4124377	*GLIS3*	T		0.99 (0.96,1.02)	0.463

trait	rsid	Position b37	Gene	Coded Allele	Conditioned on	Smoking OR (95% C.I.)	Smoking P
FVC	rs10870202	9:139257411	*DNLZ*	C	rs10858246	0.99 (0.97,1.02)	0.453
$\frac{FEV_1}{FVC}$	rs3847402	10:30267810	*KIAA1462*	A		1.02 (0.99,1.05)	0.124
FVC	rs7095607	10:69957350	*MYPN*	A		1.00 (0.98,1.03)	0.881
FEV_1	rs2509961	11:62310909	*AHNAK*	C		1.00 (0.98,1.03)	0.770
FEV_1	rs11234757	11:86443072	*PRSS23*	A		1.00 (0.96,1.04)	0.972
FEV_1	rs567508	11:126008910	*RPUSD4*	A		1.01 (0.97,1.05)	0.645
FEV_1	rs1494502	12:65824670	*MSRB3*	G		1.01 (0.98,1.04)	0.566
$\frac{FEV_1}{FVC}$	rs113745635	12:95554771	*FGD6*	T		0.97 (0.94,1.00)	0.041*
FVC	rs35506	12:115500691	*TBX3*	A		0.99 (0.96,1.02)	0.577
$\frac{FEV_1}{FVC}$	rs1698268	14:84309664	*LINC00911*	T		1.00 (0.97,1.03)	0.894
$\frac{FEV_1}{FVC}$	rs72724130	15:41977690	*MGA*	T		1.04 (0.98,1.10)	0.224
$\frac{FEV_1}{FVC}$	rs12591467	15:71788387	*THSD4*	T	rs10851839	1.00 (0.97,1.02)	0.860
$\frac{FEV_1}{FVC}$	rs66650179	15:84261689	*SH3GL3*	C		0.99 (0.96,1.03)	0.637
$\frac{FEV_1}{FVC}$	rs59835752	17:28265330	*EFCAB5*	T		1.00 (0.97,1.02)	0.777
$\frac{FEV_1}{FVC}$	rs11658500	17:36886828	*CISD3*	A		1.00 (0.96,1.03)	0.861
FVC	rs6140050	20:6632901	*BMP2*	A		1.00 (0.97,1.03)	0.951
FEV_1	rs72448466	20:62363640	*ZGPAT*	C		1.03 (1.00,1.06)	0.047*
FEV_1	rs11704827	22:18450287	*MICAL3*	T		0.99 (0.96,1.03)	0.751
FEV_1	rs2283847	22:28181399	*MN1*	T		0.97 (0.95,1.00)	0.048*
54 previously-reported variants							
$\frac{FEV_1}{FVC}$	rs2284746	1:17306675	*MFAP2*	G		1.00 (0.97,1.02)	0.885
FEV_1	rs6681426	1:150586971	*ENSA*	A		1.00 (0.97,1.02)	0.816
$\frac{FEV_1}{FVC}$	rs993925	1:218860068	*TGFB2*	T		1.02 (1.00,1.05)	0.082
$\frac{FEV_1}{FVC}$	rs4328080	1:219963088	*RNU5F-1*	A		1.04 (1.02,1.07)	0.002
$\frac{FEV_1}{FVC}$	rs62126408	2:18309132	*KCNS3*	C		0.98 (0.95,1.02)	0.340
FVC	rs1430193	2:56120853	*EFEMP1*	T		1.00 (0.97,1.03)	0.910
FEV_1	rs2571445	2:218683154	*TNS1*	G		1.00 (0.97,1.02)	0.747
$\frac{FEV_1}{FVC}$	rs10498230	2:229502503	*PID1*	T		1.05 (1.00,1.11)	0.040
$\frac{FEV_1}{FVC}$	rs12477314	2:239877148	*HDAC4*	T		1.01 (0.98,1.05)	0.511
$\frac{FEV_1}{FVC}$	rs1529672	3:25520582	*RARB*	A		0.98 (0.95,1.01)	0.244
FVC	rs1595029	3:158241767	*RP11-538P18.2*	C		0.98 (0.96,1.01)	0.158
FEV_1	rs1344555	3:169300219	*MECOM*	T		1.02 (0.98,1.05)	0.321
$\frac{FEV_1}{FVC}$	rs2045517	4:89870964	*FAM13A*	T		1.03 (1.01,1.06)	0.018
FEV_1	rs34480284	4:106064626	*TET2*	TA		1.02 (1.00,1.05)	0.091
FEV_1	rs10516526	4:106688904	*GSTCD*	G		1.00 (0.95,1.05)	0.954
$\frac{FEV_1}{FVC}$	rs34712979	4:106819053	*NPNT*	A		0.98 (0.95,1.01)	0.239

trait	rsid	Position b37	Gene	Coded Allele	Conditioned on	Smoking OR (95% C.I.)	Smoking P
$\frac{FEV_1}{FVC}$	rs138641402	4:145445779	*HHIP*	T		1.01 (0.98,1.04)	0.420
$\frac{FEV_1}{FVC}$	rs153916	5:95036700	*SPATA9*	T		0.99 (0.96,1.02)	0.470
FEV_1	rs7715901	5:147856392	*HTR4*	G		1.00 (0.98,1.03)	0.843
$\frac{FEV_1}{FVC}$	rs1990950	5:156920756	*ADAM19*	T		1.01 (0.99,1.04)	0.340
FVC	rs6924424	6:7801611	*BMP6*	G		0.99 (0.96,1.03)	0.657
FEV_1	rs34864796	6:27459923	*ZKSCAN3*	A		0.96 (0.92,1.00)	0.034
$\frac{FEV_1}{FVC}$	rs2857595	6:31568469	*NCR3*	A		1.00 (0.97,1.04)	0.833
$\frac{FEV_1}{FVC}$	rs2070600	6:32151443	*AGER*	T		0.97 (0.92,1.03)	0.297
FEV_1	rs114544105	6:32635629	*HLA-DQB1*	A		0.99 (0.96,1.02)	0.484
$\frac{FEV_1}{FVC}$	rs2768551	6:109270656	*ARMC2*	A		0.96 (0.93,1.00)	0.032
$\frac{FEV_1}{FVC}$	rs7753012	6:142745883	*LOC153910*	G		1.00 (0.97,1.03)	0.973
$\frac{FEV_1}{FVC}$	rs148274477	6:142838173	*GPR126*	T		0.93 (0.86,1.02)	0.111
$\frac{FEV_1}{FVC}$	rs16909859	9:98204792	*PTCH1*	A		1.02 (0.97,1.07)	0.467
$\frac{FEV_1}{FVC}$	rs803923	9:119401650	*ASTN2*	A		1.02 (0.99,1.05)	0.143
FVC	rs10858246	9:139102831	*LHX3*	C		0.99 (0.96,1.02)	0.378
$\frac{FEV_1}{FVC}$	rs7090277	10:12278021	*CDC123*	A		1.00 (0.98,1.03)	0.717
FEV_1	rs2637254	10:78312002	*C10orf11*	A		1.00 (0.98,1.03)	0.712
FVC	rs4237643	11:43648368	*HSD17B12*	G		0.99 (0.97,1.02)	0.641
FVC	rs2863171	11:45250732	*PRDM11*	C		1.04 (1.00,1.08)	0.036
FVC	rs2348418	12:28689514	*CCDC91*	C		1.02 (0.99,1.04)	0.235
$\frac{FEV_1}{FVC}$	rs11172113	12:57527283	*LRP1*	C		1.01 (0.98,1.03)	0.695
$\frac{FEV_1}{FVC}$	rs12820313	12:96255704	*CCDC38*	C		1.02 (0.99,1.06)	0.142
FEV_1	rs569058293	12:114743533	*RBM19*	C		1.73 (1.17,2.55)	0.006
FEV_1	rs10850377	12:115201436	*TBX3*	A		0.98 (0.95,1.01)	0.172
FEV_1	rs7155279	14:92485881	*TRIP11*	T		1.02 (0.99,1.04)	0.286
FEV_1	rs117068593	14:93118229	*RIN3*	T		1.00 (0.96,1.03)	0.857
$\frac{FEV_1}{FVC}$	rs10851839	15:71628370	*THSD4*	A		1.01 (0.99,1.04)	0.350
$\frac{FEV_1}{FVC}$	rs12149828	16:10706328	*TEKT5*	A		0.98 (0.95,1.02)	0.376
$\frac{FEV_1}{FVC}$	rs12447804	16:58075282	*MMP15*	T		0.97 (0.94,1.01)	0.112
$\frac{FEV_1}{FVC}$	rs3743609	16:75467021	*CFDP1*	C		1.00 (0.98,1.03)	0.819
FVC	rs1079572	16:78187138	*WWOX*	A		1.00 (0.98,1.03)	0.843
FEV_1	rs35524223	17:44192590	*KANSL1*	A		0.94 (0.91,0.97)	**4.79E-04**
FVC	rs6501431	17:68976415	*KCNJ2*	T		1.00 (0.97,1.03)	0.930
FEV_1	rs7218675	17:73513185	*TSEN54*	A		1.00 (0.97,1.03)	0.839
$\frac{FEV_1}{FVC}$	rs113473882	19:41124155	*LTBP4*	C		0.86 (0.75,0.99)	0.033
$\frac{FEV_1}{}$	rs2834440	21:35690499	*KCNE2*	A		0.98 (0.95,1.00)	0.091

trait	rsid	Position b37	Gene	Coded Allele	Conditioned on	Smoking OR (95% C.I.)	Smoking P
FVC							
FEV_1	rs134041	22:28056338	*MN1*	C		0.99 (0.97,1.02)	0.598
$\frac{FEV_1}{FVC}$	rs7050036	X:15964845	*AP1S2*	A		1.00 (0.98,1.02)	0.971

Supplementary Table 9: Summary of the number of variants analysed and the standard deviation of the COPD risk score in each of the studies included in risk score and single variant analyses of COPD susceptibility and risk of COPD exacerbations.

Study	Number of variants total	Number of proxies	Number of variants in risk score	Standard deviation of COPD risk score
European ancestry				
Bio*Me*	94	1	93	6.12
DiscovEHR	93	7	86	5.80
COPDGene	92	3	90	5.84
ECLIPSE	91	2	90	5.83
NETT/NAS	91	2	90	5.79
GenKOLS	91	2	90	5.84
Groningen	93	3	93	5.70
Laval	93	2	93	5.75
UBC	93	3	93	5.66
LHS	89	0	89	
deCODE COPD	95	3	95	5.85
UK Biobank	95	3	95	6.09
Chinese ancestry				
CKB	71	49	70	4.63

Supplementary Table 10: Single variant results for association with COPD risk. Results for COPD risk associations are provided for variants representing 95 lung-function-associated signals that could be followed up in case-control studies. The 47 variants for which UK BiLEVE data did not contribute to discovery are presented in (a), and the results for the 48 variants for which UK BiLEVE data did contribute to discovery are presented in (b). When the sentinel variant (Sentinel rsid) was not available in a study, a proxy (Proxy rsid) was analysed instead. For signals where different variants were analysed across studies we present results for the variants analysed in the largest number of COPD cases. Studies were clustered into 3 groups according to their study design and phenotype classification criteria: electronic health medical record (eMR), which included Bio*Me* and DiscovEHR; COPD case-control studies, which included COPDGene Study, ECLIPSE, NETT/NAS and the Norway GenKOLS study; and lung resection studies, which included Groningen, Laval and UBC. Overall sample sizes are given as N effective sample sizes (the sum of the products of the total sample size and imputation quality within each study). Results in the China Kadoorie Biobank prospective cohort (CKB) are presented in table (c). The coded allele presented in the tables is always the risk allele (defined as the allele associated with decreased lung function in UK BiLEVE). Odds ratios are bold in table (a) if directions of effect are consistent for lung function and COPD i.e. the same allele is associated both with decreased lung function and a higher risk of COPD. P values after meta-analysing all studies of European descent which reached a Bonferroni corrected threshold for 95 tests (5.26×10^{-4}) are presented in bold in table (a). In table (c), P values which reached a Bonferroni corrected threshold for 71 tests (7.04×10^{-4}) in CKB are indicated in bold. In table (c): *Consistency of direction of effect unavailable ("-") if OR=1 in either European Ancestry results or in CKB.

See accompanying Excel file.

Supplementary Table 11: Association of COPD risk with lung function risk score. Studies are grouped according to their study design and phenotyping: "eMR", electronic medical records, which used ICD codes to define COPD (DiscovEHR also used spirometry to refine the COPD definition); "case-control", COPD case-control, which used post-bronchodilator spirometry to define COPD; "lung resection cohort", which used a combination of pre and post-bronchodilator spirometry to define COPD; the Icelandic Biobank, deCODE, where cases were selected from a population based study and a study of COPD patients and defined using a spirometric definition, controls were selected as individuals within the cohort that were not known cases (no spirometric definition was used for controls); and UK Biobank, which used spirometry to define both COPD cases and controls. UK Biobank is separated into UK BiLEVE, which was the discovery population for 48 of the variants included in the risk score (43 discovered in this analysis and 5 in [46]) and the remaining of UK Biobank labelled "UK Biobank". Meta-analysed results within each of these groups and across all studies are presented, both per allele and as per standard deviation of the risk score (~6 alleles).

Study/ Study group	per allele		per sd		N cases	N controls
	OR (95% CI)	P	OR (95% CI)	P		
European ancestry						
eMR	1.01 (1,1.02)	5.56E-03	1.08 (1.02,1.14)	5.55E-03	1471	14849
COPD case control	1.05 (1.05,1.06)	5.52E-36	1.36 (1.3,1.43)	5.65E-36	5778	3950
lung resection	1.05 (1.02,1.08)	6.56E-04	1.33 (1.13,1.57)	6.74E-04	310	332
deCODE COPD	1.03 (1.02,1.04)	7.67E-09	1.18 (1.12,1.25)	7.67E-09	1248	74770
UK BiLEVE	1.06 (1.06,1.07)	5.03E-193	1.46 (1.42,1.50)	5.03E-193	9563	27387
UK Biobank	1.04 (1.03,1.05)	1.96E-12	1.27 (1.19,1.36)	1.96E-12	984	26561
UK BILEVE + UK Biobank	1.06 (1.06,1.06)	3.94E-205	1.42 (1.39,1.45)	3.94E-205	10547	53948
All	1.05 (1.05,1.05)	1.59E-223	1.35 (1.32,1.37)	1.59E-223	19354	147849
All excluding UK BiLEVE	1.04 (1.03,1.04)	5.05E-49	1.24 (1.20,1.27)	5.05E-49	9791	120462
Chinese ancestry						
CKB	1.02 (1.01,1.02)	4.22E-06	1.077 (1.044,1.112)	4.22E-06	7116	20919

Supplementary Table 12: Single variant results for association with COPD exacerbations. Results for COPD exacerbations associations are provided for 95 lung-function-associated signals that could be followed up in case-control studies. When the sentinel variant (Sentinel rsid) was not available in a study, a proxy (Proxy rsid) was analysed instead. For signals where different variants were analysed across studies we present results for the variants analysed in the largest number of COPD cases. Studies were clustered into 2 groups according to their study design and phenotype classification criteria: electronic health medical record (eMR), which included Bio*Me* and DiscovEHR; and COPD case-control studies, which included COPDGene Study, ECLIPSE, NETT/NAS and the Norway GenKOLS study. Meta-analysed results within each of these groups, as well as for LHS and UK Biobank, and across all studies are presented in table (a). Results in the China Kadoorie Biobank prospective cohort (CKB) are presented in table (b). The coded allele presented in the tables is always the risk allele (defined as the allele associated with decreased lung function in UK BiLEVE).

See accompanying Excel file.

Supplementary Table 13: Association of COPD exacerbations with lung function risk score. Results for COPD exacerbation risk score associations are provided. Studies that took part in these analyses were grouped according to their study design and phenotyping into: electronic health medical record (eMR), which included Bio*Me* and DiscovEHR and COPD case-control studies, which included COPDGene Study, ECLIPSE, NETT/NAS and the Norway GenKOLS study. Meta-analysed results within each of these groups and across all studies are presented per allele.

Study/ Study group	per allele		N cases	N controls
	OR (95% CI)	P		
European ancestry				
eMR	0.99 (0.97,1.01)	4.74E-01	773	664
COPD case control	1.01 (0.99,1.02)	3.41E-01	1042	4724
LHS	0.97 (0.94,1.01)	1.31E-01	100	4002
UK Biobank	1 (0.99,1.02)	5.61E-01	647	9900
All	1 (0.99,1.01)	7.25E-01	2562	19290
Chinese ancestry				
CKB	1 (0.99,1.02)	7.35E-01	5292	1824

Supplementary Table 14: Deleterious variants that explain the lung function association signal. Each of the 97 sentinel variants were conditioned on nearby coding functional variants as identified by Variant Effect Predictor. The unconditional association effect sizes and P values are shown for the sentinel variant with the conditional effect sizes and P values for the sentinel after conditioning on the functional variant shown in the consecutive rows. The LD of each functional variant with the sentinel is shown (r^2 with sentinel), the Combined Annotation Dependent Depletion (CADD), PHRED-scaled score and the gene implicated by the functional variant. Only sentinels and functional conditional variants are shown where P>0.01 after conditioning.
*Sentinel rs28986170 is a tertiary signal after conditioning on rs2070600 and rs201002132 and hence was conditioned on these in addition to any functional variants.

trait	Sentinel/ condition on	rsid	position	r^2 with sentinel	CADD PHRED	Beta (se) sentinel unconditional conditional	P sentinel unconditional conditional	Gene
Novel variants								
FEV_1 FVC	**sentinel**	**rs28986170***	**6:31556155**			**0.077 (0.013)**	**1.23E-08**	
	condition	rs41558312	6:31378864	0.688	12.3	0.033 (0.013)	0.013	*MICA*
	condition	rs41293883	6:31474820	0.757	12.5	0.030 (0.013)	0.025	*MICB*
FVC	**sentinel**	**rs7095607**	**10:69957350**			**-0.037 (0.007)**	**3.92E-08**	
	condition	rs7079481	10:69957350	0.993	27.0	0.000 (0.006)	0.947	*MYPN*
FEV_1	**sentinel**	**rs2509961**	**11:62310909**			**0.036 (0.007)**	**1.69E-07**	
	condition	rs13941	11:62310909	0.454	10.0	0.016 (0.007)	0.017	*C11orf83*
FEV_1 FVC	**sentinel**	**rs11658500**	**17:36886828**			**-0.051 (0.009)**	**4.69E-08**	
	condition	rs2879097	17:36886828	0.501	19.2	-0.021 (0.009)	0.024	*CISD3*
Previously-reported variants								
FEV_1	**sentinel**	**rs2571445**	**2:218683154**			**0.043 (0.007)**	**2.19E-10**	
	condition	rs1063281	2:218668732	0.925	17.99	0.005 (0.007)	0.410	*TNS1*
FEV_1	**sentinel**	**rs34864796**	**6:27459923**			**-0.075 (0.010)**	**6.14E-14**	
	condition	rs34788973	6:27459923	0.797	6.853	-0.010 (0.010)	0.277	*OR2B2*
FEV_1 FVC	**sentinel**	**rs2857595**	**6:31568469**			**-0.048 (0.008)**	**3.50E-09**	
	condition	rs3134900	6:31473957	0.580	8.773	-0.013 (0.008)	0.100	*MICB*
FEV_1	**sentinel**	**rs114544105**	**6:32635629**			**-0.049 (0.008)**	**8.84E-11**	
	condition	rs3891176	6:32634318	0.971	13.75	-0.005 (0.007)	0.516	*HLA-DQB1*
FEV_1	**sentinel**	**rs35524223**	**17:44192590**			**-0.061 (0.008)**	**1.13E-13**	
	condition	rs34579536	17:44108906	0.968	3.452	-0.005 (0.008)	0.508	*KANSL1*
	condition	rs17651549	17:44061278	0.981	18.18	-0.004 (0.008)	0.647	*MAPT*
	condition	rs12373123	17:43924073	0.977	17.99	-0.005 (0.008)	0.552	*SPPL2C*
FEV_1	**sentinel**	**rs7218675**	**17:73513185**			**-0.035 (0.007)**	**2.34E-06**	
	condition	rs991150	17:73513185	0.991	13.19	0.000 (0.007)	0.961	*TSEN54*
FEV_1 FVC	**sentinel**	**rs113473882**	**19:41124155**			**0.145 (0.035)**	**3.03E-05**	
	condition	rs34093919	19:41117300	0.878	18.35	-0.011 (0.034)	0.742	*LTBP4*

Supplementary Table 15: Plausible genes per locus. Summary of general and functional information with regards to each novel and previously-reported sentinel variant (where applicable). All plausible genes (for definition, see 'Implication of causal genes' section, Online Methods) with regards to each loci are presented. Non-high-priority genes at the HLA regions are excluded. *High-priority genes. #Variant did not reach $P<5.15x10^{-4}$ in this study for any trait.

Genome-wide significant trait (additional traits with $P<5.15x10^{-4}$)	Variant ID (position b37)	Nearest gene(s)	All plausible genes
Novel signals			
FEV_1/FVC (FVC)	rs17513135 (chr1:40,035,686)	*LOC101929516* (intron)	*PABPC4*, OXCT2, MACF1, HPCAL4, NDUFS5, BMP8A*
FEV_1/FVC (-)	rs1192404 (chr1:92,068,967)	*CDC7/TGFBR3*	*CDC7*
FEV_1/FVC (FEV_1)	rs6688537 (chr1:239,850,588)	*CHRM3* (intron)	*CHRM3**
FEV_1/FVC (-)	rs61332075 (chr2:239,316,560)	*TRAF3IP1/ASB1*	*ASB1, TRAF3IP1*
FVC (FEV_1)	rs1490265 (chr3:67,452,043)	*SUCLG2* (intron)	*SUCLG2*
FEV_1/FVC (FEV_1)	rs2811415 (chr3:127,991,527)	*EEFSEC* (intron)	*RUVBL1*, SEC61A1, EEFSEC*
FEV_1/FVC (-)	rs13110699 (chr4:89,815,695)	*FAM13A* (intron)	*FAM13A**
FEV_1/FVC (-)	rs1551943 (chr5:52,195,033)	*ITGA1* (intron)	*ITGA1*
FEV_1/FVC (-)	rs7713065 (chr5:131,788,334)	*C5orf56* (intron)	*SLC22A4, SLC22A5, RAD50, IRF1, PDLIM4, P4HA2*
FEV1 (FVC, FEV_1/FVC)	rs3839234 (chr5:148,596,693)	*ABLIM3* (intron)	*GRPEL2*, ABLIM3*, AFAP1L1*
FEV_1/FVC (FEV_1)	rs10515750 (chr5:156,810,072)	*CYFIP2* (intron)	*ADAM19*, ITK, FNDC9, NIPAL4, CYFIP2*
FEV_1/FVC (FEV_1)	rs200003338 (chr6:31,556,155)	*LST1* (intron)	*MICB*, MICA**
FEV_1/FVC (FEV_1)	rs10246303 (chr7:7,286,445)	*C1GALT1* (3' UTR)	*C1GALT1**
FEV_1/FVC (-)	rs72615157 (chr7:99,635,967)	*ZKSCAN1* (3' UTR)	*PILRB, TRIM4, AP4M1, PVRIG, COPS6, MCM7, STAG3, CNPY4, ZNF3, LAMTOR4, ZSCAN21, MEPCE, ZCWPW1, TAF6, TSC22D4, MBLAC1, NYAP1, GAL3ST4, ZKSCAN1, PILRA*
FVC (FEV_1)	rs10870202 (chr9:139,257,411)	*DNLZ* (intron)	*INPP5E*, CARD9*, SNAPC4, DNLZ, SDCCAG3, GPSM1, PMPCA, SEC16A*
FVC (FEV_1)	rs7095607 (chr10:69,957,350)	*MYPN* (intron)	*MYPN*, ATOH7*
FEV_1 (FVC)	rs2509961 (chr11:62,310,909)	*AHNAK* (intron)	*ROM1*, EML3*, MTA2*, GANAB*, C11orf83*, INTS5, BSCL2, ZBTB3, AHNAK, B3GAT3, TTC9C, HNRNPUL2, UBXN1*
FEV_1 (FVC, FEV_1/FVC)	rs567508 (chr11:126,008,910)	*RPUSD4/CDON*	*FOXRED1, RPUSD4, CDON*
FEV_1 (FVC)	rs1494502 (chr12:65,824,670)	*MSRB3* (intron)	*LEMD3*
FEV_1/FVC (FEV_1)	rs113745635 (chr12: 95,554,771)	*FGD6* (intron)	*FGD6, VEZT, NDUFA12,*

			NR2C1, SNRPF
FEV_1/FVC (-)	rs72724130 (chr15:41,977,690)	*MGA* (intron)	*SPTBN5, MAPKBP1*
FEV_1/FVC (FEV_1)	rs66650179 (chr15:84,261,689)	*SH3GL3* (intron)	*ADAMTSL3*
FEV_1/FVC (-)	rs59835752 (chr17: 28,265,330)	*EFCAB5* (intron)	*EFCAB5*, CRYBA1*, SSH2*, SLC6A4*, CPD, GOSR1, NSRP1, CORO6, ANKRD13B, GIT1, BLMH, TP53I13*
FEV_1/FVC (FEV_1)	rs11658500 (chr17:36,886,828)	*CISD3* (intron)	*CISD3*, PCGF2*
FEV_1 (FVC)	rs72448466 (chr20:62,363,640)	*ZGPAT* (intron)	*LIME1*, ZGPAT, RTEL1, EEF1A2, SLC2A4RG, STMN3*
FEV_1 (FVC)	rs11704827 (chr22:18,450,287)	*MICAL3* (intron)	*MICAL3*
Previously-reported lung function signals			
FEV_1 (FVC)	rs2284746 (chr1:17,306,675)	*MFAP2* (intron)	*MFAP2, PADI2, ATP13A2, CROCC, NBPF1, MACF1, SDHB*
FEV_1 (FVC)	rs6681426 (chr1:150,586,971)	*MCL1/ENSA*	*GOLPH3L* , FAM63A, ADAMTSL4, MRPS21, LASS2, HORMAD1, ARNT, CTSK, CTSS, CDC42SE1, BNIPL, C1orf138, MCL1, SETDB1, SCNM1, ANXA9*
FEV_1/FVC (-)	rs993925 (chr1:218,860,068)	*MIR548F3*	*TGFB2*
FEV_1/FVC (-)	rs4328080 (chr1:219,963,088)	*LYPLAL1/RNU5F-1*	*SLC30A10**
FEV_1/FVC (FEV_1, FVC)	rs62126408 (chr2:18,309,132)	*KCNS3/RDH14*	*KCNS3*
FVC# (-)	rs1430193 (chr2: 56,120,853)	*EFEMP1* (intron)	*EFEMP1*
FEV_1 (FVC, FEV_1/FVC)	rs2571445 (chr2:218,683,154)	*TNS1* (exon)	*TNS1**
FEV_1/FVC (-)	rs10498230 (chr2:229,502,503)	*SPHKAP/PID1*	*SPHKAP**
FVC (FEV_1)	rs1595029 (chr3: 158,241,767)	*RSRC1* (intron)	*RSRC1*, GFM1, MLF1, FLJ40475, MFSD1, LXN*
FEV_1# (-)	rs1344555 (chr3:169,300,219)	*MECOM* (intron)	*MECOM*
FEV_1/FVC (-)	rs2045517 (chr4: 89,870,964)	*FAM13A* (intron)	*FAM13A*
FEV_1 (FVC, FEV_1/FVC)	rs10516526 (chr4:106,688,904)	*GSTCD* (intron)	*INTS12*, GSTCD*, NPNT**
FEV_1/FVC (FEV_1, FVC)	rs34712979 (chr4:106,819,053)	*NPNT* (intron)	*NPNT**
FEV_1	rs34480284 (chr4: 106,064,626)	*LOC101929468/TET2*	*PPA2*
FEV_1/FVC (FEV_1)	rs138641402 (chr4:145,445,779)	*GYPA/HHIP-AS1*	*HHIP**
FEV_1/FVC (-)	rs153916 (chr5 95,036,700)	*SPATA9/RHOBTB3*	*RHOBTB3*, ARSK, SPATA9*
FEV_1 (FVC, FEV_1/FVC)	rs7715901 (chr5:147,856,392)	*HTR4* (intron)	*FBXO38, SPINK7*
FEV_1/FVC (FEV_1)	rs1990950 (chr5: 156,920,756)	*ADAM19* (intron)	*ADAM19*, NIPAL4, CYFIP2, THG1L*
FEV_1 (FVC, FEV_1/FVC)	rs34864796 (chr6:27,459,923)	*ZNF184/LINC01012*	*OR2B2**
FEV_1/FVC (FEV_1)	rs2857595 (chr6:31,568,469)	*NCR3/AIF1*	*MICB**
FEV_1/FVC (-)	rs2070600 (chr6:32,151,443)	*AGER* (exon)	*AGER**
FEV_1 (FVC, FEV_1/FVC)	rs114544105 (chr6:32,635,629)	*HLA-DQB1/HLA-DQA2*	*HLA-DQB1*, APOM*, RNF5**
FEV_1/FVC (-)	rs2768551 (chr6: 109,270,656)	*ARMC2* (intron)	*SESN1, ARMC2*
FEV_1/FVC (FEV_1)	rs113096699 (chr6:142,745,883)	*GPR126* (intron)	*GPR126**
FEV_1/FVC (-)	rs148274477 (chr6:142,838,173)	*GPR126/LOC153910*	*GPR126**

FEV_1/FVC (-)	rs16909859 (chr9: 98,204,792)	*PTCH1*	*PTCH1, NEFH*
FEV_1/FVC (-)	rs803923 (chr9:119,401,650)	*ASTN2* (intron)	*ASTN2*
FVC (FEV_1)	rs10858246 (chr9:139,102,831)	*QSOX2* (intron)	*QSOX2*, DNLZ, CARD9*
FEV_1/FVC (FEV_1)	rs7090277 (chr10:12,278,021)	*CDC123* (intron)	*CDC123, CAMK1D, NUDT5*
FEV_1 (FVC, FEV_1/FVC)	rs2637254 (chr10:78,312,002)	*C10orf11* (intron)	*C10orf11*
FVC# (-)	rs4237643 (chr11:43,648,368)	*MIR129-2/HSD17B12*	*HSD17B12*
FVC# (-)	rs2863171 (chr11:45,250,732)	*PRDM11* (3' UTR)	*SYT13*
FVC (FEV_1)	rs2348418 (chr12:28,689,514)	*CCDC91* (intron)	*FLJ35252*, CCDC91, PTHLH*
FEV_1/FVC (-)	rs11172113 (chr12:57,527,283)	*LRP1* (intron)	*LRP1*, STAT6, TMEM194A, ING2*
FEV_1/FVC (-)	rs12820313 (chr12:96,255,704)	*SNRPF* (intron)	*SNRPF, NTN4*
FEV_1 (-)	rs7155279 (chr14:92,485,881)	*TRIP11* (intron)	*ATXN3*, TRIP11, CPSF2, FBLN5, NDUFB1*
FEV_1# (-)	rs117068593 (chr14:93,118,229)	*RIN3* (exon)	*RIN3**
FEV_1/FVC (FEV_1)	rs10851839 (chr15:71,628,370)	*THSD4* (intron)	*THSD4*, SENP8*
FEV_1/FVC (-)	rs12149828 (chr16:10,706,328)	*EMP2/TEKT5*	*CLEC16A*
FEV_1/FVC (-)	rs12447804 (chr16:58,075,282)	*MMP15* (intron)	*MMP15*, ZNF319, C16orf57, C16orf80, CSNK2A2, TEPP*
FEV_1/FVC (FEV_1)	rs3743609 (chr16:75,467,021)	*CFDP1* (intron)	*TMEM170A*, BCAR1*, CFDP1*, ADAT1*
FVC (-)	rs1079572 (chr16:78,187,138)	*WWOX* (intron)	*WWOX*
FEV_1 (FVC, FEV_1/FVC)	rs35524223 (chr17:44,192,590)	*KANSL1* (intron)	*KANSL1*, MAPT*, ARL17B*, ARL17A*, LRRC37A4*, NUDT1*, LRRC37A*, CRHR1*, LRRC37A2*, ARHGAP27*, FMNL1*, PLEKHM1*, WNT3*, NSF*, SPPL2C*, TBC1D24, GOSR2, EPB41L5, CCDC43, DCAKD, SPPL2C*
FEV_1 (FVC)	rs7218675 (chr17:73,513,185)	*TSEN54* (intron)	*CASKIN2*, TSEN54*, TSEN54, MRPS7, KIAA0195, GRB2, LLGL2, NUP85, KIAA0195, MIF4GD*
FEV_1/FVC (-)	rs113473882 (chr19:41,124,155)	*LTBP4* (intron)	*LTBP4**
FEV_1/FVC (-)	rs2834440 (chr21:35,690,499)	*LINC00310/KCNE2*	*KCNE2, LINC00310, MRPS6*

Supplementary Table 16: Gene-based pathway analyses. Summary of gene sets overrepresented in known biological pathways and gene ontology (GO) terms. Pathway analysis results for (i) all high-priority genes (n=68) and (ii) analysis including all implicated causal genes (excluding non-high-priority genes at the HLA regions, n=234) are presented separately. GO term categories (m= molecular function, b= biological process, c= cellular component) and levels (1 to 5 with high level GO terms assigned to level 1) are indicated. The effective size is the number of genes present in that respective pathway or GO term. Pathways or gene sets represented by only 2 genes from the same association signal have been excluded. Pathways or gene sets which include 2 or more genes implicated via the same association signal have been noted. FDR: False discovery rate.

All high-priority genes (n=68)

Overrepresented biological pathways

None at FDR<0.05

Overrepresented gene ontology terms

P value	FDR	Name of GO term (GO term category/level)	Genes associated with GO term	Total size of GO geneset	Notes
5.42E-05	0.001	SH3 domain binding (m/4)	*MYPN, ADAM19, BCAR1, ARHGAP27, MAPT*	117	*ARHGAP27* and *MAPT* implicated by the same signal (rs35524223); and *MYPN* is a novel gene at a novel signal. *ADAM19* is implicated at both a novel and a previously-reported signal.
2.43E-04	0.037	fibroblast migration (b/5)	*TNS1, AGER, MTA2*	35	*MTA2* is a novel gene at a novel signal
7.70E-04	0.059	cellular response to misfolded protein (b/5)	*RNF5, ATXN3*	12	
1.06E-03	0.019	protein domain specific binding (m/3)	*MYPN, WNT3, NSF, CARD9, ARHGAP27, MAPT, ADAM19, BCAR1*	597	*WNT3, NSF, ARHGAP27* and *MAPT* are all implicated by rs35524223; and *CARD9* and *MYPN* are novel genes at different novel signals. *ADAM19* is implicated at both a novel and a previously-reported signal.
1.39E-03	0.019	apolipoprotein binding (m/3)	*LRP1, MAPT*	16	

1.48E-03	0.012	small GTPase binding (m/5)	*RHOBTB3, FMNL1, RIN3, NSF, SLC6A4*	240	*NSF* and *FMNL1* implicated by rs35524223; and *SLC6A4* is a novel gene at a novel signal
1.57E-03	0.012	syntaxin-1 binding (m/5)	*NSF, SLC6A4*	17	*SLC6A4* is a novel gene at a novel signal
2.14E-03	0.015	GTPase binding (m/4)	*RHOBTB3, FMNL1, RIN3, NSF, SLC6A4*	261	*NSF* and *FMNL1* implicated by rs35524223; and *SLC6A4* is a novel gene at a novel signal
2.40E-03	0.015	actin binding (m/4)	*SLC6A4, FMNL1, SSH2, ABLIM3, MYPN, TNS1*	392	*SSH2* and S*LC6A4* implicated by rs59835752; and *ABLIM3*, *MYPN*, and *SSH2* and *SLC6A4* are novel genes at three different novel signals
3.87E-03	0.035	protein complex binding (m/3)	*LRP1, SLC6A4, FMNL1, NSF, NPNT, LTBP4, MAPT, MTA2, CRHR1*	902	*MAPT*, *FMNL1*, *CRHR1* and *NSF* are implicated by rs35524223; and *MTA* and *SLC6A4* are novel genes at different novel signals
All plausible genes (excluding non-high-priority genes in HLA region, n=234)					
Overrepresented biological pathways					
P value	**FDR**	**Name of pathway**	**Genes in pathway**	**Total size of pathway geneset**	**Notes**
7.71E-06	0.003	Signaling events mediated by the Hedgehog family	*CDON, PTCH1, PTHLH, TGFB2, HHIP*	23	*CDON* is a novel gene at a novel signal; and *PTHLH* is a novel gene at a previously-reported signal
3.05E-05	0.006	Molecules associated with elastic fibres	*EFEMP1, TGFB2, LTBP4, MFAP2, FBLN5*	30	
6.60E-05	0.008	Elastic fibre formation	*EFEMP1, TGFB2, LTBP4, MFAP2, FBLN5*	35	
1.00E-04	0.010	Ligand-receptor interactions	*CDON, PTCH1, HHIP*	8	*CDON* is a novel gene at a novel signal

Overrepresented gene ontology terms

P value	FDR	Name of GO term (GO term category/level)	Genes associated with GO term	Total size of GO geneset	Flags
6.99E-05	0.029	extracellular matrix organization (b/4)	*HSD17B12, MMP15, TGFB2, CTSK, ADAMTSL4, EFEMP1, ITGA1, THSD4, NTN4, NPNT, LTBP4, MFAP2, CTSS, LEMD3, FBLN5*	388	*ADAMTSL4*, *CTSS* and *CTSK* implicated by the same signal (rs6681426)
7.20E-05	0.019	extracellular structure organization (b/3)	*HSD17B12, MMP15, TGFB2, CTSK, ADAMTSL4, EFEMP1, ITGA1, THSD4, NTN4, NPNT, LTBP4, MFAP2, CTSS, LEMD3, FBLN5*	389	*ADAMTSL4*, *CTSS* and *CTSK* implicated by the same signal (rs6681426); *LEMD3* and *ITGA1* are novel genes at different novel signals
3.24E-04	0.014	fibronectin binding (m/3)	*HSD17B12, CTSS, CTSK, MFAP2*	28	*CTSS* and *CTSK* implicated by the same signal (rs6681426)
4.23E-04	0.014	hedgehog family protein binding (m/3)	*PTCH1, HHIP*	3	
8.62E-04	0.020	protein domain specific binding (m/3)	*MLF1, MYPN, LLGL2, HPCAL4, STMN3, WNT3, EPB41L5, NSF, SLC22A4, SLC22A5, CARD9, GRB2, ARHGAP27, MCL1, MAPT, ADAM19, BCAR1*	597	*EPB41L5, WNT3, NSF, ARHGAP27* and *MAPT* are all implicated by rs35524223; also *SLC22A4* and *SLC22A5* are implicated by the same signal (rs7713065). *GRB2* and *LLGL2* are also implicated by the same signal (rs7218675). *CARD9, HPCAL4, STMN3* and *MYPN* are novel genes at different novel signals
1.22E-03	0.021	protein complex binding (m/3)	*HSD17B12, SLC6A4, ITGA1, MACF1, CTSK, MFAP2, CORO6, FMNL1, NEFH, NSF, FBLN5, TRAF3IP1, MTA2, LTBP4, CTSS, ING2, LRP1, NPNT, GIT1, MAPT, PTCH1, CRHR1*	902	*FMNL1, NSF, CRHR1 and MAPT* implicated by rs35524223; and *CTSS* and *CTSK* are implicated by the same signal (rs6681426). *NEFH* and *PTCH1, SLC6A4* and *GIT1*, and *ING2* and *LRP1* are also implicated by the same signals (rs16909859, rs59835752 and rs11172113 respectively). *MACF1, ITGA1, GIT1, CORO6, SLC6A4, MTA2* and *TRAF3IP1* are novel genes at different novel signals

2.82E-03	0.067	SH3 domain binding (m/4)	*ARHGAP27, GRB2, MYPN, MAPT, ADAM19, BCAR1*	117	*ARHGAP27* and *MAPT* implicated by the same signal (rs35524223); *MYPN* is a novel gene at a novel signal. *ADAM19* is implicated at both a novel and a previously-reported signal.
3.18E-03	0.036	organellar small ribosomal subunit (c/5)	*MRPS7, MRPS6, MRPS21*	25	
3.76E-03	0.036	Golgi stack (c/5)	*INPP5E, AP4M1, GOLPH3L, NSF, GOSR1, GAL3ST4*	124	*GAL3ST4* and *AP4M1* implicated by the same signal (rs72615157). *GOSR1, GAL3ST4* and *AP4M1* are also novel genes at novel signals. *INPP5E* is a high priority gene at a novel signal.
3.98E-03	0.036	MLL1/2 complex (c/5)	*TAF6, KANSL1, RUVBL1*	27	*TAF6* and *RUVBL1* are novel genes at different novel signals
5.39E-03	0.036	Golgi cisterna (c/5)	*AP4M1, GOSR1, GOLPH3L, GAL3ST4, INPP5E*	94	*GAL3ST4* and *AP4M1* implicated by the same signal. *GOSR1, INPP5E, GAL3ST4* and *AP4M1* are novel genes at novel signals.

Supplementary Table 17: Results of MAGENTA pathway analysis. Results (P value and FDR) presented for analyses run with the HLA region included and with the HLA region excluded. Green shading indicates FDR<5% for either analysis. PMF: PANTHER Molecular Functions, PBP: PANTHER Biological Processes, PP: PANTHER Pathways, GO: Gene Ontology term, KEGG: Kyoto Encyclopedia of Genes and Genomes.

Database Gene set	HLA included P value	HLA included FDR	HLA excluded P value	HLA excluded FDR
FEV_1				
KEGG SYSTEMIC LUPUS ERYTHEMATOSUS	1.60E-04	0.0080	3.97E-03	0.2489
KEGG ALLOGRAFT REJECTION	8.20E-05	0.0092	7.82E-02	0.5623
KEGG GRAFT VERSUS HOST DISEASE	2.18E-04	0.0100	0.146	0.4988
KEGG ARRHYTHMOGENIC RIGHT VENTRICULAR CARDIOMYOPATHY ARVC	9.00E-04	0.0319	1.90E-03	0.2317
KEGG ASTHMA	2.10E-03	0.0389	8.14E-02	0.5696
FEV_1/FVC				
PMF Major histocompatibility complex antigen	6.00E-06	0.0005	5.60E-02	0.8659
GO nucleosome	4.00E-06	0.0012	4.50E-05	0.0487
KEGG SYSTEMIC LUPUS ERYTHEMATOSUS	1.70E-05	0.0019	1.34E-03	0.1877
GO antigen processing and presentation of peptide antigen via MHC class I	9.00E-06	0.0019	1.34E-02	0.4215
PMF Histone	4.30E-05	0.0027	1.85E-04	0.0237
PBP MHCI-mediated immunity	2.50E-05	0.0030	1.13E-02	0.1534
KEGG CELL ADHESION MOLECULES CAMS	1.31E-04	0.0118	2.63E-02	0.4948
KEGG TYPE I DIABETES MELLITUS	4.66E-04	0.0134	0.670	0.9957
Ingenuity PXR.RXR.Activation	8.00E-04	0.0258	2.00E-03	0.1722
KEGG GRAFT VERSUS HOST DISEASE	1.22E-03	0.0272	0.644	0.9495
Ingenuity Interferon.Signaling	2.30E-03	0.0389	5.40E-03	0.0976
PBP Phagocytosis	1.20E-03	0.0392	3.60E-03	0.1309
KEGG ALLOGRAFT REJECTION	2.63E-03	0.0407	0.799	1.0000
PBP Cell communication	5.00E-04	0.0474	1.60E-03	0.1238
KEGG VIRAL MYOCARDITIS	2.50E-03	0.0475	0.363	0.9522
KEGG ANTIGEN PROCESSING AND PRESENTATION	2.46E-03	0.0487	0.901	1.0000
FVC				
PP FAS signaling pathway	3.00E-06	0.0001	8.00E-06	<0.00001
KEGG SYSTEMIC LUPUS ERYTHEMATOSUS	2.15E-04	0.0278	1.10E-03	0.2039
Ingenuity Hepatic.Cholestasis	1.10E-03	0.0348	3.50E-03	0.0657
GO positive regulation of apoptosis	2.80E-05	0.0399	2.40E-05	0.0369

Supplementary Table 18: Chromatin Mark enrichment. Results of analysis of enrichment for overlap of lung function signals with H3K4me1 and H3K4me3 histone marks in 127 tissues/cell types from the Roadmap/ENCODE projects. Tables A and B: overlap of H3K4me1 using hypergeometric test and GoShifter, respectively. Tables C and D: overlap of H3K4me3 using hypergeometric test and GoShifter, respectively. Tissue/cell types that were significant using both the hypergeometric test and GoShifter are in bold.

A) H3K4me1 overlap using hypergeometic test

	Tissue/cell type	P value	FDR
E083	**Fetal Heart**	**<0.001**	**0.016**
E076	Colon Smooth Muscle	<0.001	0.016
E078	Duodenum Smooth Muscle	0.001	0.024
E055	Foreskin Fibroblast Primary Cells skin01	0.001	0.031
E111	**Stomach Smooth Muscle**	**0.003**	**0.047**
E065	Aorta	0.003	0.047
E088	**Fetal Lung**	**0.004**	**0.053**
E126	NHDF-Ad Adult Dermal Fibroblast Primary Cells	0.005	0.053
E090	Fetal Muscle Leg	0.007	0.070
E056	Foreskin Fibroblast Primary Cells skin02	0.009	0.087
E075	Colonic Mucosa	0.010	0.087

B) H3K4me1 overlap using GoShifter method

	Tissue/cell type	P value
E072	Brain Inferior Temporal Lobe	0.008
E088	**Fetal Lung**	**0.017**
E128	NHLF Lung Fibroblast Primary Cells	0.018
E058	Foreskin Keratinocyte Primary Cells skin03	0.024
E061	Foreskin Melanocyte Primary Cells skin03	0.030
E083	**Fetal Heart**	**0.039**
E111	**Stomach Smooth Muscle**	**0.042**
E023	Mesenchymal Stem Cell Derived Adipocyte Cultured Cells	0.046
E089	Fetal Muscle Trunk	0.046

C) H3K4me3 overlap using hypergeometic test

	Tissue/cell type	P value	FDR
E065	Aorta	9.30E-05	0.006
E106	Sigmoid Colon	1.05E-03	0.026
E126	NHDF-Ad Adult Dermal Fibroblast Primary Cells	1.29E-03	0.026
E092	Fetal Stomach	1.32E-03	0.026
E013	hESC Derived CD56+ Mesoderm Cultured Cells	4.68E-03	0.060
E035	Primary hematopoietic stem cells	4.78E-03	0.060
E109	Small Intestine	6.58E-03	0.060
E090	Fetal Muscle Leg	6.61E-03	0.060
E005	H1 BMP4 Derived Trophoblast Cultured Cells	7.64E-03	0.060
E062	Primary mononuclear cells from peripheral blood	8.46E-03	0.060
E086	Fetal Kidney	8.89E-03	0.060
E026	Bone Marrow Derived Cultured Mesenchymal Stem Cells	9.75E-03	0.060
E084	Fetal Intestine Large	0.010	0.060
E029	Primary monocytes from peripheral blood	0.010	0.060

	Tissue/cell type	P value	FDR
E089	Fetal Muscle Trunk	0.010	0.060
E031	Primary B cells from cord blood	0.013	0.069
E085	Fetal Intestine Small	0.015	0.071
E104	Right Atrium	0.017	0.071
E046	Primary Natural Killer cells from peripheral blood	0.018	0.071
E095	Left Ventricle	0.019	0.071
E116	GM12878 Lymphoblastoid Cell Line	0.019	0.071
E088	Fetal Lung	0.020	0.071
E093	Fetal Thymus	0.021	0.071
E083	Fetal Heart	0.022	0.071
E037	Primary T helper memory cells from peripheral blood 2	0.022	0.071
E097	Ovary	0.022	0.071
E004	H1 BMP4 Derived Mesendoderm Cultured Cells	0.023	0.073
E078	Duodenum Smooth Muscle	0.024	0.073
E053	Cortex derived primary cultured neurospheres	0.025	0.076
E091	Placenta	0.026	0.078
E122	**HUVEC Umbilical Vein Endothelial Cells Cell Line**	**0.027**	**0.078**
E075	Colonic Mucosa	0.028	0.078
E098	Pancreas	0.035	0.088
E055	Foreskin Fibroblast Primary Cells skin01	0.035	0.088
E076	Colon Smooth Muscle	0.036	0.088
E001	ES-I3 Cell Line	0.037	0.089
E082	Fetal Brain Female	0.038	0.089
E028	Breast variant Human Mammary Epithelial Cells (vHMEC)	0.040	0.091
E044	Primary T regulatory cells from peripheral blood	0.044	0.095
E111	**Stomach Smooth Muscle**	**0.045**	**0.095**
E121	HSMM cell derived Skeletal Muscle Myotubes Cell Line	0.045	0.095
E128	NHLF Lung Fibroblast Primary Cells	0.049	0.100

D) H3K4me3 overlap using GoShifter method

	Tissue/cell type	P value
E122	**HUVEC Umbilical Vein Endothelial Cells Cell Line**	**0.010**
E111	**Stomach Smooth Muscle**	**0.025**
E063	Adipose Nuclei	0.035
E124	Monocytes-CD14+ RO01746 Cell Line	0.041

Supplementary Table 19: Druggability analysis. Genes encoding targets for which there are approved drugs and/or clinical candidates in ChEMBL. Indications were ordered by 'Max phase' (i.e. the maximum phase a clinical trial has reached). *High-priority genes. Phase 1: Testing of drug on healthy volunteers for dose-ranging; Phase 2: Testing of drug on patients to assess efficacy and safety; Phase 3: Testing of drug on patients to assess efficacy, effectiveness and safety; and Phase 4: Approval of drug and post-marketing surveillance. EFO: Experimental Factor Ontology; MeSH: Medical Subject Headings.

A) All genes

Lung function Sentinel SNP (trait), position, gene, ChEMBL Target ID, name	Approved drugs and clinical candidates [ChEMBL ID]	Approved drugs and Clinical candidates [Name]	Indications [MeSH/EFO term] (Max phase for indication)
rs1192404 (FEV1/FVC) chr1: 92,068,967 *CDC7* CHEMBL5443 Cell division cycle 7-related protein kinase	CHEMBL3544943	BMS-863233	Hematologic Cancer (2)
	CHEMBL3545090	RXDX-103	Cancer (N/A)
	CHEMBL3545321	NMS-1116354	Advanced Solid Tumors (1)
rs6688537 (FEV1/FVC) chr1: 239,850,588 **CHRM3* CHEMBL245 Muscarinic acetylcholine receptor M3	CHEMBL14	CARBACHOL	GLAUCOMA (4)
	CHEMBL550	PILOCARPINE	GLAUCOMA (4), URINARY INCONTINENCE (1)
	CHEMBL1133	OXYBUTYNIN CHLORIDE	HYPERHIDROSIS (4), POLYURIA (4), URINARY INCONTINENCE (4), URINARY BLADDER NEUROGENIC (3)
	CHEMBL1184	ACETYLCHOLINE CHLORIDE	GLAUCOMA (4)
	CHEMBL1231	OXYBUTYNIN	HYPERHIDROSIS (4), POLYURIA (4), URINARY INCONTINENCE (4), URINARY BLADDER NEUROGENIC (3)
	CHEMBL1240	PROPANTHELINE BROMIDE	DIGESTIVE SYSTEM DISEASES (4)
	CHEMBL517712	ATROPINE	DIGESTIVE SYSTEM DISEASES (4), PARKINSON'S DISEASE (4), PEPTIC ULCER (4), SEASONAL ALLERGIC RHINITIS (4), AMBLYOPIA (3), PAIN (3), GLUCOSE INTOLERANCE (1)
	CHEMBL1578	ANISOTROPINE METHYLBROMIDE	Peptic Ulcer (N/A)
	CHEMBL523299	UMECLIDINIUM BROMIDE	**CHRONIC OBSTRUCTIVE PULMONARY DISEASE** (4), **ASTHMA** (2), HYPERHIDROSIS (1)
	CHEMBL1724	MEPENZOLATE BROMIDE	DIGESTIVE SYSTEM DISEASES (4)
	CHEMBL551466	ACLIDINIUM BROMIDE	**CHRONIC OBSTRUCTIVE PULMONARY DISEASE** (4)
	CHEMBL1768	BETHANECHOL CHLORIDE	EOSINOPHILIC ESOPHAGITIS (2), TYPE 2 DIABETES MELLITUS (1)
	CHEMBL1200330	PILOCARPINE HYDROCHLORIDE	GLAUCOMA (4), URINARY INCONTINENCE (1)
	CHEMBL1200347	ISOPROPAMIDE IODIDE	DIGESTIVE SYSTEM DISEASES (4)
	CHEMBL1200473	CYCLOPENTOLATE HYDROCHLORIDE	Retinopathy of Prematurity (N/A)
	CHEMBL1200479	DICYCLOMINE HYDROCHLORIDE	DIGESTIVE SYSTEM DISEASES (4)
	CHEMBL1200604	TROPICAMIDE	SIALORRHEA (2)
	CHEMBL1200764	METHACHOLINE CHLORIDE	**ASTHMA** (4)
	CHEMBL1200771	TRIDIHEXETHYL CHLORIDE	DIGESTIVE SYSTEM DISEASES (4)
	CHEMBL1200803	SOLIFENACIN SUCCINATE	POLYURIA (4), URINARY INCONTINENCE (4)
	CHEMBL1200880	DIPHEMANIL METHYLSULFATE	DIGESTIVE SYSTEM DISEASES (4)
	CHEMBL1200891	OXYPHENCYCLIMINE HYDROCHLORIDE	DIGESTIVE SYSTEM DISEASES (4)
	CHEMBL1200906	OXYPHENONIUM BROMIDE	DIGESTIVE SYSTEM DISEASES (4)

Lung function Sentinel SNP (trait), position, gene, ChEMBL Target ID, name	Approved drugs and clinical candidates [ChEMBL ID]	Approved drugs and Clinical candidates [Name]	Indications [MeSH/EFO term] (Max phase for indication)
	CHEMBL1200935	DARIFENACIN HYDROBROMIDE	POLYURIA (4), URINARY INCONTINENCE (4)
	CHEMBL1200950	CLIDINIUM BROMIDE	DIGESTIVE SYSTEM DISEASES (4)
	CHEMBL1201024	METHSCOPOLAMINE BROMIDE	DIGESTIVE SYSTEM DISEASES (4)
	CHEMBL1201027	GLYCOPYRROLATE BROMIDE	**OBSTRUCTIVE LUNG DISEASE** (4), **CHRONIC OBSTRUCTIVE PULMONARY DISEASE** (3), DIGESTIVE SYSTEM DISEASES (4), **ASTHMA** (2)
	CHEMBL1201765	FESOTERODINE FUMARATE	POLYURIA (4), URINARY INCONTINENCE (4), NOCTURIA (2)
	CHEMBL1626570	HEXOCYCLIUM METHYLSULFATE	DIGESTIVE SYSTEM DISEASES (4)
	CHEMBL1722209	TOLTERODINE TARTRATE	POLYURIA (4), URINARY INCONTINENCE (4), KIDNEY CALCULI (2)
	CHEMBL2134724	IPRATROPIUM BROMIDE HYDRATE	**OBSTRUCTIVE LUNG DISEASE** (4), **CHRONIC OBSTRUCTIVE PULMONARY DISEASE** (4), NASAL OBSTRUCTION (4)
	CHEMBL2146146	ATROPINE SULFATE	DIGESTIVE SYSTEM DISEASES (4), PARKINSON'S DISEASE (4), PEPTIC ULCER (4), SEASONAL ALLERGIC RHINITIS (4), AMBLYOPIA (3), PAIN (3), GLUCOSE INTOLERANCE (1)
	CHEMBL2218917	CEVIMELINE HYDROCHLORIDE	Xerostomia (4)
	CHEMBL3084748	TROSPIUM CHLORIDE	POLYURIA (4), URINARY INCONTINENCE (4), **CHRONIC OBSTRUCTIVE PULMONARY DISEASE** (1)
	CHEMBL3545181	TIOTROPIUM BROMIDE	**ASTHMA** (4), **CHRONIC OBSTRUCTIVE PULMONARY DISEASE** (4), CYSTIC FIBROSIS (3)
	CHEMBL1779046	Tarafenacin	Overactive Bladder (2)
	CHEMBL3545222	AZD8683	**CHRONIC OBSTRUCTIVE PULMONARY DISEASE** (2)
rs62126408 (FEV_1/FVC - previous) chr2: 18,309,132 *KCNS3* CHEMBL2362996 Voltage-gated potassium channel	CHEMBL284348	DALFAMPRIDINE	MULTIPLE SCLEROSIS (4), STROKE (3), RENAL INSUFFICIENCY (1)
	CHEMBL1200728	GUANIDINE HYDROCHLORIDE	HEART FAILURE (3)
rs10515750 (FEV_1/FVC) chr5: 156,810,072 *ITK* CHEMBL2959 Tyrosine-protein kinase ITK/TSK	CHEMBL1201733	PAZOPANIB HYDROCHLORIDE	NEOPLASMS (4), RENAL CELL CARCINOMA (3), OVARIAN CARCINOMA (3), SARCOMA (3), NON-SMALL CELL LUNG CARCINOMA (2), HEAD AND NECK SQUAMOUS CELL CARCINOMA (2), GASTROINTESTINAL STROMAL TUMOR (2), LEIOMYOSARCOMA (2), ACUTE MYELOID LEUKEMIA (2), LIPOSARCOMA (2), LYMPHEDEMA (2), AGE-RELATED MACULAR DEGENERATION (2), PROSTATE ADENOCARCINOMA (2), GASTRIC CARCINOMA (2), HEREDITARY HEMORRHAGIC TELANGIECTASIA (2), THYROID CARCINOMA (2), VON HIPPEL-LINDAU DISEASE (2), CORNEAL NEOVASCULARIZATION (1)
rs113745635 (FEV_1/FVC) chr12: 95,554,771 *NDUFA12* CHEMBL2363065 Mitochondrial complex I (NADH dehydrogenase)	CHEMBL1703	METFORMIN HYDROCHLORIDE	TYPE I DIABETES MELLITUS (4), TYPE II DIABETES MELLITUS (4), FATTY LIVER (4), GESTATIONAL DIABETES (4), GLUCOSE INTOLERANCE (4), OBESITY (4), POLYCYSTIC OVARY SYNDROME (4), BRAIN NEOPLASMS (3), BREAST CARCINOMA (3), PROSTATIC NEOPLASMS (3), ADENOCARCINOMA (2), NON-SMALL CELL LUNG CARCINOMA (2), COLORECTAL NEOPLASMS (2), ENDOMETRIAL NEOPLASM (2), LUNG NEOPLASMS (2), PULMONARY HYPERTENSION (2), MELANOMA (2), MILD COGNITIVE IMPAIRMENT (2), PERIODONTITIS (2), RENAL INSUFFICIENCY (2), LI-FRAUMENI SYNDROME (1), NON-ALCOHOLIC FATTY LIVER DISEASE (1), PANCREATIC NEOPLASMS (1)
	CHEMBL3545320	ME-344	Solid Tumors (1)
rs59835752 (FEV_1/FVC) chr17: 28,265,330 **SLC6A4* CHEMBL228	CHEMBL1113	AMOXAPINE	DEPRESSIVE DISORDER (4)
	CHEMBL1118	DESVENLAFAXINE	DEPRESSIVE DISORDER (4), FIBROMYALGIA (2)
	CHEMBL1409	FLUVOXAMINE MALEATE	DEPRESSIVE DISORDER (4), OBSESSIVE-COMPULSIVE DISORDER (4), AUTISTIC DISORDER (3)

Lung function Sentinel SNP (trait), position, gene, ChEMBL Target ID, name	Approved drugs and clinical candidates [ChEMBL ID]	Approved drugs and Clinical candidates [Name]	Indications [MeSH/EFO term] (Max phase for indication)
Serotonin transporter	CHEMBL1692	IMIPRAMINE HYDROCHLORIDE	DEPRESSIVE DISORDER (4), GASTROESOPHAGEAL REFLUX (3), PAIN (3)
	CHEMBL1708	PAROXETINE HYDROCHLORIDE	ANXIETY (4), DEPRESSIVE DISORDER (4), POST-TRAUMATIC STRESS DISORDER (4), PREMATURE EJACULATION (3), HIV INFECTION (1)
	CHEMBL1709	SERTRALINE HYDROCHLORIDE	ANXIETY (4), DEPRESSIVE DISORDER (4), POST-TRAUMATIC STRESS DISORDER (4), PANIC DISORDER (4), AUTISM (3), INJURY (2)
	CHEMBL1200322	ESCITALOPRAM OXALATE	ANXIETY (4), DEPRESSIVE DISORDER (4), OBSESSIVE-COMPULSIVE DISORDER (4), POST-TRAUMATIC STRESS DISORDER (4), BIPOLAR DISORDER (3), CARCINOMA (3), PULMONARY HYPERTENSION (3), CANCER (3), BORDERLINE PERSONALITY DISORDER (2), COCAINE DEPENDENCE (2), HEPATITIS C (2)
	CHEMBL1200328	DULOXETINE HYDROCHLORIDE	ANXIETY (4), DEPRESSIVE DISORDER (4), DIABETIC NEPHROPATHY (4), FIBROMYALGIA (4), OSTEOARTHRITIS (4), PAIN (4), NEUROPATHY (4), MULTIPLE SCLEROSIS (3), OSTEOARTHRITIS OF THE KNEE (3), ALCOHOLISM (2), ATTENTION DEFICIT HYPERACTIVITY DISORDER (2), CHRONIC FATIGUE SYNDROME (2), NEURALGIA (2)
	CHEMBL1200332	PROTRIPTYLINE HYDROCHLORIDE	DEPRESSIVE DISORDER (4)
	CHEMBL1200492	NEFAZODONE HYDROCHLORIDE	DEPRESSIVE DISORDER (4)
	CHEMBL1200595	CHLORPHENTERMINE HYDROCHLORIDE	Anorexia (N/A)
	CHEMBL1200609	PAROXETINE MESYLATE	ANXIETY (4), DEPRESSIVE DISORDER (4), POST-TRAUMATIC STRESS DISORDER (4), PREMATURE EJACULATION (3), HIV INFECTION (1)
	CHEMBL1200631	IMIPRAMINE PAMOATE	DEPRESSIVE DISORDER (4), GASTROESOPHAGEAL REFLUX (3), PAIN (3)
	CHEMBL1200710	CLOMIPRAMINE HYDROCHLORIDE	ADEPRESSIVE DISORDER (4), PREMATURE EJACULATION (3)
	CHEMBL1200781	CITALOPRAM HYDROBROMIDE	DEPRESSIVE DISORDER (4), AUTISTIC DISORDER (2), COCAINE DEPENDENCE (2), STROKE (2), ALCOHOLISM (1), AUTISM SPECTRUM DISORDER (1)
	CHEMBL1200798	TRAZODONE HYDROCHLORIDE	DEPRESSIVE DISORDER (4), INSOMNIA (3), ALCOHOLISM (2)
	CHEMBL1200964	AMITRIPTYLINE HYDROCHLORIDE	DEPRESSIVE DISORDER (4), PAIN (4), MIGRAINE DISORDER (3), INSOMNIA (3), MOVEMENT DISORDER (2)
	CHEMBL1201066	VENLAFAXINE HYDROCHLORIDE	ANXIETY (4), DEPRESSIVE DISORDER (4), PROSTATE CARCINOMA (3), COCAINE DEPENDENCE (2), PAIN (2)
	CHEMBL1201082	FLUOXETINE HYDROCHLORIDE	DEPRESSIVE DISORDER (4), AUTISTIC DISORDER (3), GASTROESOPHAGEAL REFLUX (2), OBSESSIVE-COMPULSIVE DISORDER (2), STROKE (2)
	CHEMBL1201156	NORTRIPTYLINE HYDROCHLORIDE	DEPRESSIVE DISORDER (4), GASTROESOPHAGEAL REFLUX (3), GASTROPARESIS (3), IRRITABLE BOWEL SYNDROME (2), PSORIASIS (2), ATOPIC ECZEMA (1)
	CHEMBL1201728	DESVENLAFAXINE SUCCINATE	DEPRESSIVE DISORDER (4), FIBROMYALGIA (2)
	CHEMBL1615374	VILAZODONE HYDROCHLORIDE	ANXIETY (4), DEPRESSIVE DISORDER (4), MARIJUANA DEPENDENCE (2), MEMORY IMPAIRMENT (2)
	CHEMBL2096626	MILNACIPRAN HYDROCHLORIDE	DEPRESSIVE DISORDER (4), FIBROMYALGIA (4), PAIN (4), IRRITABLE BOWEL SYNDROME (2), NEURALGIA (2)
	CHEMBL2105732	LEVOMILNACIPRAN HYDROCHLORIDE	DEPRESSIVE DISORDER (4)
	CHEMBL2107387	VORTIOXETINE HYDROBROMIDE	DEPRESSIVE DISORDER (4), ANXIETY (3), LIVER DISEASE (1)
	CHEMBL3039565	DESVENLAFAXINE FUMARATE	DEPRESSIVE DISORDER (4), FIBROMYALGIA (2)
	CHEMBL2104986	TEDATIOXETINE	DEPRESSIVE DISORDER (2)
rs35524223 (FEV_1 - previous) chr17: 44,192,590	CHEMBL482950	PEXACERFONT	Generalized Anxiety Disorder (2), Irritable Bowel Syndrome (2), Major Depressive Disorder (1)

Lung function Sentinel SNP (trait), position, gene, ChEMBL Target ID, name	Approved drugs and clinical candidates [ChEMBL ID]	Approved drugs and Clinical candidates [Name]	Indications [MeSH/EFO term] (Max phase for indication)
**CRHR1* CHEMBL1800 Corticotropin releasing factor receptor 1	CHEMBL291657	SSR125543	Major Depression (2)
	CHEMBL514270	EMICERFONT	Irritable Bowel Syndrome (2)
	CHEMBL1287935	VERUCERFONT	Post-Traumatic Stress Disorder (2), Alcohol Dependence (2)

B) Genes encoding targets predicted to interact with high-priority gene products

Lung function Sentinel SNP (trait), position, high-priority gene	Genes encoding targets predicted to interact with high-priority gene products (ChEMBL ID), name	Approved drugs and clinical candidates [ChEMBL ID]	Approved drugs and Clinical candidates [Name]	Indications [MeSH/EFO term] (Max phase for indication)
rs10870202 (FVC) chr9: 139,257,411 *INPP5E	PIK3CD (CHEMBL3130), PIK3CA (CHEMBL4005), PI3-kinase p110-delta subunit	CHEMBL2216870	IDELALISIB	CHRONIC LYMPHOCYTIC LEUKEMIA (3), HODGKINS LYMPHOMA (2), NON-HODGKINS LYMPHOMA (2), ALLERGIC RHINITIS (1)
		CHEMBL3545397	Acalisib	Lymphoid Malignancies (1)
		CHEMBL3545048	AMG-319	Head and Neck cancer squamous cell carcinoma (2), Tumors (1)
		CHEMBL3545052	CUDC-907	Multiple Myeloma (1)
		CHEMBL3545112	ME-401	N/A
		CHEMBL3545141	RP-6530	T-Cell Lymphoma (1)
		CHEMBL3545205	INCB-040093	Refractory Hodgkin Lymphoma (2)
		CHEMBL3545247	CAL-263	Allergic Rhinitis (1)
		CHEMBL3545250	GSK-2269557	**Chronic Obstructive Pulmonary Disease** (2), **Asthma** (1)
		CHEMBL3545267	TGR-1202	Chronic Lymphocytic Leukemia (1)
rs2509961 (FEV1) Chr11: 62,310,909 *MTA2	HDAC3 (CHEMBL1829), Histone deacetylase 3	CHEMBL98	VORINOSTAT	CUTANEOUS T-CELL LYMPHOMA (3), BRAIN DISEASE (2), HIV-1 INFECTION (2), ACUTE MYELOID LEUKEMIA (2), LYMPHOMA (2), NEOPLASM (2), SARCOMA (2), BRAIN NEOPLASM (1), BREAST CARCINOMA (1), PANCREATIC CARCINOMA (1), OVARIAN CARCINOMA (1)
rs35524223 (FEV1 - previous) chr17:44,192,590 *KANSL1	MGA (CHEMBL2074), Maltase-glucoamylase	CHEMBL1561	MIGLITOL	TYPE II DIABETES MELLITUS
		CHEMBL1566	ACARBOSE	TYPE II DIABETES MELLITUS (4), METABOLIC SYNDROME X (3), NON-ALCOHOLIC FATTY LIVER DISEASE (2)
rs6688537 (FEV_1/FVC) chr1: 239,850,588 *CHRM3	HCRTR1 (CHEMBL5113), Orexin receptor 1	CHEMBL1272307	SB-649868	Insomnia (2)
		CHEMBL3545367	LEMBOREXANT	Driving performance (1)
rs3743609 (FEV_1/FVC - previous) chr16:75,467,021 *BCAR1	JAK2 (CHEMBL2971), Tyrosine-protein kinase JAK2	CHEMBL1795071	RUXOLITINIB PHOSPHATE	POLYCYTHEMIA VERA (3), PRIMARY MYELOFIBROSIS (3), ALOPECIA AREATA (2), BETA-THALASSEMIA (2), BREAST CARCINOMA (2), CACHEXIA (2), HODGKINS LYMPHOMA (2), MYELOPROLIFERATIVE DISORDER (2), METASTATIC PROSTATE CANCER (2), PSORIASIS (2), CHRONIC LYMPHOCYTIC LEUKEMIA (1)
		CHEMBL603469	LESTAURTINIB	Leukemia (2), Psoriasis (2)
		CHEMBL2035187	PACRITINIB	Hodgkin Lymphoma (2)
		CHEMBL1231124	AZD-1480	Primary Myelofibrosis (1)
		CHEMBL2107823	GANDOTINIB	N/A
		CHEMBL3545215	BMS-911543	Cancer (2)
		CHEMBL3545217	NS-018	Primary Myelofibrosis (2)
		CHEMBL3544997	LS-104	N/A
		CHEMBL3545241	AC-430	Rheumatoid Arthritis (1)

Lung function Sentinel SNP (trait), position, high-priority gene	Genes encoding targets predicted to interact with high-priority gene products (ChEMBL ID), name	Approved drugs and clinical candidates [ChEMBL ID]	Approved drugs and Clinical candidates [Name]	Indications [MeSH/EFO term] (Max phase for indication)
		CHEMBL3545328	XL-019	Polycythemia Vera (1)
rs11172113 (FEV_1/FVC - previous) chr12: 7,527,283 *LRP1	PLAT (CHEMBL1873), Tissue-type plasminogen activator	CHEMBL1046	AMINOCAPROIC ACID	HEMORRHAGE (4), CRANIOSYNOSTOSIS (2)
rs11172113 (FEV_1/FVC - previous) chr12:57,527,283 *LRP1	PDGFRB (CHEMBL1913), Platelet-derived growth factor receptor beta	CHEMBL1421	DASATINIB	CHRONIC MYELOGENOUS LEUKEMIA (4), BREAST CARCINOMA (2), NON-SMALL CELL LUNG CARCINOMA (2), POLYCYTHEMIA VERA (2), GLIOBLASTOMA (2), CENTRAL NERVOUS SYSTEM CANCER (2), SYSTEMIC SCLERODERMA (1)
		CHEMBL1642	IMATINIB MESYLATE	GASTROINTESTINAL STROMAL TUMOR (4), CHRONIC MYELOGENOUS LEUKEMIA (4), PULMONARY HYPERTENSION (3), SARCOMA (3), **ASTHMA** (2), OVARIAN CARCINOMA (2), POLYCYTHEMIA VERA (2), CENTRAL NERVOUS SYSTEM CANCER (1)
		CHEMBL1200485	SORAFENIB TOSYLATE	HEPATOCELLULAR CARCINOMA (4), RENAL CELL CARCINOMA (3), KIDNEY NEOPLASM (3), BREAST CARCINOMA (2), PORTAL HYPERTENSION (2), KELOID (2), MELANOMA (2), OVARIAN CARCINOMA (2), PULMONARY HYPERTENSION (1)
		CHEMBL124660	TANDUTINIB	Prostate Cancer (2), Glioblastoma (2), Acute Myelogenous Leukemia (1)
rs3743609 (FEV_1/FVC - previous) chr16:75,467,021 *BCAR1	SRC (CHEMBL267), Tyrosine-protein kinase SRC	CHEMBL24828	VANDETANIB	THYROID CARCINOMA (4), Various Cancers (3-1)
		CHEMBL288441	BOSUTINIB	CHRONIC MYELOGENOUS LEUKEMIA (4), GLIOBLASTOMA (2)
		CHEMBL571546	KX2-391	Prostate Cancer (2)
rs12447804 (FEV_1/FVC - previous) chr16:58,075,282 *MMP15	MMP1 (CHEMBL332), MMP8 (CHEMBL4588), MMP7 (CHEMBL4073), Matrix metalloproteinase-1,8,7	CHEMBL1200567	DOXYCYCLINE HYCLATE	ACNE (4), BLEPHARITIS (4), INFECTION (4), PERIODONTITIS (4), **CHRONIC OBSTRUCTIVE PULMONARY DISEASE** (4), ALZHEIMERS DISEASE (3), HEMORRHAGE (3), URETHRITIS (3), PRIMARY SYSTEMIC AMYLOIDOSIS (2), ABDOMINAL AORTIC ANEURYSM (2), COLORECTAL ADENOCARCINOMA (2), DIABETIC RETINOPATHY (2), INFLAMMATION (2), NEOPLASM OF MATURE B-CELLS (2), AGE-RELATED MACULAR DEGENERATION (2), MARFAN SYNDROME (2), PAIN (2), PLEURAL EFFUSION (2), RHEUMATOID ARTHRITIS (1)
		CHEMBL1200699	DOXYCYCLINE HYDRATE	ACNE (4), BLEPHARITIS (4), INFECTION (4), PERIODONTITIS (4), **CHRONIC OBSTRUCTIVE PULMONARY DISEASE** (4), ALZHEIMERS DISEASE (3), HEMORRHAGE (3), URETHRITIS (3), PRIMARY SYSTEMIC AMYLOIDOSIS (2), ABDOMINAL AORTIC ANEURYSM (2), COLORECTAL ADENOCARCINOMA (2), DIABETIC RETINOPATHY (2), INFLAMMATION (2), NEOPLASM OF MATURE B-CELLS (2), AGE-RELATED MACULAR DEGENERATION (2), MARFAN SYNDROME (2), PAIN (2), PLEURAL EFFUSION (2), RHEUMATOID ARTHRITIS (1)
		CHEMBL2364574	DOXYCYCLINE CALCIUM	ACNE (4), BLEPHARITIS (4), INFECTION (4), PERIODONTITIS (4), **CHRONIC OBSTRUCTIVE PULMONARY DISEASE** (4), ALZHEIMERS DISEASE (3), HEMORRHAGE (3), URETHRITIS (3), PRIMARY SYSTEMIC AMYLOIDOSIS (2), ABDOMINAL AORTIC ANEURYSM (2), COLORECTAL ADENOCARCINOMA (2), DIABETIC RETINOPATHY (2), INFLAMMATION (2), NEOPLASM OF MATURE B-CELLS (2),

Lung function Sentinel SNP (trait), position, high-priority gene	Genes encoding targets predicted to interact with high-priority gene products (ChEMBL ID), name	Approved drugs and clinical candidates [ChEMBL ID]	Approved drugs and Clinical candidates [Name]	Indications [MeSH/EFO term] (Max phase for indication)
				AGE-RELATED MACULAR DEGENERATION (2), MARFAN SYNDROME (2), PAIN (2), PLEURAL EFFUSION (2), RHEUMATOID ARTHRITIS (1)

Supplementary Table 20: Characteristics of studies contributing to analyses of COPD susceptibility and risk of exacerbation. Summaries are given separately for each analysis subgroup (i.e. cases and controls). SD: Standard Deviation. l: litres.

Study Name	Case/control status	n total	n (%) female	Age range	Age, mean (SD)	Height range (cm)	Height, mean (SD) (cm)	FEV_1, mean (SD) (l)	FEV_1/FVC, mean (SD)	FVC, mean (SD) (l)	% ever smokers	Pack-years range	Pack-years, mean (SD)
European ancestry													
Bio*Me*-EUR	COPD case	207	44.9	56-98	74.1 (9.7)	147.3-195.6	169.1 (10)	-	-	-	45.4	-	-
	COPD control	1,817	48.3	48-101	70.2 (9.2)	141.6-210.8	169.6 (10.3)	-	-	-	17.3	-	-
	Exacerbation case	8	62.5	62-87	77.5 (9.1)	149.9-182.9	166 (12.9)	-	-	-	37.5	-	-
	Exacerbation control	199	44.2	56-98	74 (9.8)	147.3-195.6	169.2 (9.9)	-	-	-	45.7	-	-
DiscovEHR *	COPD case	1,280	36.4	40-92	70.1 (10.8)	99.1-208.3	168.9 (10.1)	1.5(0.62)	0.55 (0.12)	2.7 (0.88)	92.8	-	-
	COPD control	13,321	54.6	40-92	64.5 (12.7)	119.4-203.2	168 (10.2)	2.7(0.72)	0.8 (0.05)	3.38 (0.92)	48.8	-	-
	Exacerbation case	774	33.9	40-92	71 (10.2)	99.1-208.3	169.2 (10.2)	1.44(0.59)	0.54 (0.12)	2.63 (0.85)	96.3	-	-
	Exacerbation control	472	39.6	40-92	68.4 (11.5)	137.2-198.1	168.6 (10.2)	1.6(0.64)	0.57 (0.12)	2.81 (0.93)	90.0	-	-
COPDGene	COPD case	2,812	44.3	45-81	64.7 (8.2)	138.9-195.6	169.7 (9.4)	1.46(0.64)	0.49 (0.13)	2.95 (0.91)	100.0	10-331.7	56.3 (28)
	COPD control	2,534	50.7	45-81	59.5 (8.7)	140-200.3	169.7 (9.4)	2.96(0.69)	0.78 (0.05)	3.81 (0.9)	100.0	10-172.5	37.8 (20.3)
	Exacerbation case	557	44.5	45-81	63.2 (8.5)	147.9-195.6	168.8 (9.1)	1.25(0.59)	0.45 (0.13)	2.74 (0.87)	100.0	10-237.6	58 (28)
	Exacerbation control	2,255	44.3	45-81	65 (8.1)	138.9-195	169.9 (9.5)	1.51(0.64)	0.5 (0.13)	3 (0.92)	100.0	10-331.7	55.8 (28)
ECLIPSE	COPD case	1,736	33.1	40-75	63.7 (7.1)	142-201	169.5 (9)	1.33(0.52)	0.45 (0.12)	3.01 (0.9)	100.0	6-220	50.4 (27.4)
	COPD control	176	42.6	40-75	57.5 (9.5)	151-196	171.7 (9.7)	3.27(0.82)	0.79 (0.06)	4.16 (1.04)	100.0	10-230	32.2 (25)
	Exacerbation case	278	31.3	40-75	63.8 (7.3)	144-189	168.4 (8.5)	1.14(0.44)	0.42 (0.11)	2.74 (0.84)	100.0	10-220	51.4 (29.6)

Study Name	Case/control status	n total	n (%) female	Age range	Age, mean (SD)	Height range (cm)	Height, mean (SD) (cm)	FEV_1, mean (SD) (l)	FEV_1/FVC, mean (SD)	FVC, mean (SD) (l)	% ever smokers	Pack-years range	Pack-years, mean (SD)
	Exacerbation control	1,458	33.4	40-75	63.7 (7)	142-201	169.7 (9.1)	1.37(0.52)	0.45 (0.12)	3.06 (0.91)	100.0	6-205	50.2 (27)
NETT/NAS	COPD case	376	35.9	40-85	67.5 (5.8)	142.7-190.5	168.8 (9.6)	0.82(0.26)	0.32 (0.06)	2.62 (0.83)	100.0	12-260	66.4 (30.7)
	COPD control	435	0.0	48-89	69.8 (7.5)	156.7-192	174.4 (6.8)	3.03(0.51)	0.79 (0.05)	3.83 (0.63)	100.0	10-185.5	40.7 (27.8)
	Exacerbation case	87	36.8	40-77	66.7 (5.7)	144.8-185.4	167.9 (8.6)	0.77(0.24)	0.31 (0.06)	2.52 (0.78)	100.0	22-193.5	71.8 (36.2)
	Exacerbation control	277	34.7	49-85	67.7 (5.8)	142.7-190.5	169.3 (9.6)	0.83(0.26)	0.32 (0.06)	2.66 (0.85)	100.0	12-260	64.3 (28.8)
GenKOLS	COPD case	854	39.8	40-90	65.5 (10.1)	146-197	169.9 (9)	1.57(0.71)	0.51 (0.13)	2.99 (0.96)	100.0	3-130	31.9 (18.5)
	COPD control	805	49.8	40-88	55.6 (9.7)	151-200	171.8 (8.8)	3.24(0.73)	0.79 (0.04)	4.11 (0.94)	100.0	2.5-90	19.7 (13.6)
	Exacerbation case	120	45.0	43-89	68.9 (9.5)	148-185	167.5 (8.5)	1.11(0.48)	0.44 (0.13)	2.48 (0.75)	100.0	3.9-130	34 (22.7)
	Exacerbation control	734	39.0	40.4-90	64.9 (10)	146-197	170.3 (9)	1.65(0.71)	0.53 (0.13)	3.07 (0.96)	100.0	3-125	31.6 (17.7)
Groningen	COPD case	98	50.0	35-81	58.4 (9.4)	154-194	170.7 (9.4)	0.78(0.47)	0.34 (0.12)	2.29 (1.01)	94.8	0-90	31.7 (17.4)
	COPD control	42	47.6	46-76	60.6 (8.5)	156-196	172.5 (8.3)	1.33(1.11)	0.81 (0.08)	1.61 (1.26)	90.5	0-70	32.3 (18.8)
Laval	COPD case	134	43.3	33-81	64.3 (8.4)	142-183	164.7 (8.4)	1.79(0.48)	0.59 (0.08)	3.07 (0.8)	98.5	0-157.5	53.1 (29.3)
	COPD control	164	49.4	34-80	60.5 (10.1)	145-188	164.4 (9.4)	2.12(0.53)	0.76 (0.04)	2.8 (0.68)	87.2	0-136	35.4 (26.5)
UBC	COPD case	78	38.5	41-84	63 (8.7)	147-195	170.6 (10.2)	1.87(0.67)	0.57 (0.12)	3.23 (1.02)	98.6	0-180	53.6 (33.8)
	COPD control	126	54.0	25-80	63.3 (10.2)	152-188	167.1 (8.4)	2.65(0.79)	0.77 (0.05)	3.45 (1.03)	91.1	0-125	36.6 (26.5)
LHS	Exacerbation case	100	41.0	36-60	49.5 (6.5)	148-198	170 (9.4)	2.57(0.62)	0.64 (0.06)	4.04 (0.95)	100.0	10-156	45.3 (22.1)
	Exacerbation control	4,002	36.9	35-62	48.5 (6.7)	142-216	172.1 (8.9)	2.78(0.63)	0.65 (0.06)	4.29 (0.95)	100.0	0-190	40.5 (18.6)
deCODE COPD **	COPD case	1,964	58.1	40-100	67.2 (10.7)	145-198	167.9 (8.9)	1.46(0.56)	0.59 (0.09)	2.46 (0.82)	78.9	1.7-124.8	45.9 (28)

Study Name	Case/control status	n total	n (%) female	Age range	Age, mean (SD)	Height range (cm)	Height, mean (SD) (cm)	FEV_1, mean (SD) (l)	FEV_1/FVC, mean (SD)	FVC, mean (SD) (l)	% ever smokers	Pack-years range	Pack-years, mean (SD)
	COPD control	142,262	49.6	40-100	61.2 (12.6)	146-198	169.1 (9.2)	2.53(0.8)	0.78 (0.06)	3.29 (1.03)	21.4	1-200.6	30.6 (24)
UK Biobank	COPD case	984	50.1	41-70	61.9 (6.2)	145-191	168.3 (8.6)	1.97 (0.47)	0.64 (0.06)	3.1 (0.72)	88	0-152.75	23 (20.4)
	COPD control	26561	61	39-70	55.9 (7.9)	139-200	167.6 (8.9)	2.91 (0.66)	0.78 (0.04)	3.74 (0.85)	39.5	0-210	16.5 (13.9)
UK BiLEVE	COPD case	9563	46.4	40-70	58.9 (7.2)	136-203	168.8 (9.2)	1.84 (0.54)	0.61 (0.07)	3.01 (0.82)	60.7	10.5-301	41.6 (20.9)
	COPD control	27387	50.8	40-70	56.4 (8)	122-201	168.8 (9)	3.1 (0.76)	0.78 (0.04)	3.99 (0.96)	47.8	10.125-180	31.2 (15.1)
UK Biobank + UK BiLEVE	Exacerbation case	647	47.0	40-70	61 (6.7)	136-193.5	167.6 (9.2)	1.57(0.53)	0.57 (0.1)	2.76 (0.8)	82.1	0-190	45 (23.6)
	Exacerbation control	9,900	47.0	40-70	59.1 (7.2)	138-203	168.9 (9.2)	1.87(0.53)	0.62 (0.07)	3.03 (0.81)	62.0	0-301	38.7 (21.6)
Chinese ancestry													
CKB	COPD case	7,116	48.1	40-79	62 (8.7)	101.9-186.4	156.3 (8.6)	1.45(0.66)	0.72 (0.14)	1.98 (0.75)	49.7	0-235	34.4 (24.2)
	COPD control	20,919	52.1	40-79	56.7 (9.5)	113.3-187.3	158.3 (8.3)	2.23(0.64)	0.83 (0.08)	2.71 (0.82)	38.8	0-199	27.8 (20.9)
	Exacerbation case	5,292	47.2	40-79	61.9 (8.7)	101.9-186.4	156.3 (8.6)	1.46(0.68)	0.74 (0.13)	1.93 (0.74)	51.5	0-196	35.1 (24.2)
	Exacerbation control	1,824	50.6	40-77	62.4 (8.8)	131.2-182.3	156 (8.5)	1.43(0.6)	0.66 (0.13)	2.14 (0.74)	44.2	0-235	31.9 (23.8)

*Spirometry results for COPD controls presented in the table for DiscovEHR are based only on 1120 individuals with spirometry data available.
** Spirometry results for COPD controls presented in the table for deCODE COPD are based only on 2502 individuals with spirometry data available.

Supplementary Table 21: Weights for risk score in UK Biobank. Weights for each of the 95 variants were selected from studies free of winner's curse bias as follows: weights from UK Biobank were used for 47 variants not discovered in UK Biobank, weights from a meta-analysis of COPD case-control studies (COPDGene, ECLIPSE, NETT/NAS, GenKOLS) were used for a further 41 variants with data available in those studies, weights from a meta-analysis of lung resection cohort studies and deCODE (lungeQTL+deCODE) were used for a further 4 variants and weights from deCODE were used for variants that did not have data in either COPD case-control or lung resection cohort studies but had data available in deCODE (3 variants). Given the limited sample sizes available to estimate some of these weights, 9 variants had opposite direction of effect on COPD risk to what would be expected given their effect on lung function. We assigned a small weight (the smallest positive logOR across variants = 4.97×10^{-5}) to all these variants.

Markername	Chromosome	Position	Risk allele	Non-risk allele	Study used for weight	Beta	weight
rs2284746	1	17,306,675	G	C	UK Biobank	0.0587	0.985
rs17513135	1	40,035,686	T	C	COPD case-control studies	0.0673	1.130
rs1192404	1	92,068,967	G	A	COPD case-control studies	0.0555	0.933
rs12140637	1	92,374,517	T	C	COPD case-control studies	0.0152	0.255
rs200154334	1	118,862,070	CAT	C	COPD case-control studies	0.0215	0.362
rs6681426	1	150,586,971	A	G	UK Biobank	0.0156	0.262
rs993925	1	218,860,068	C	T	UK Biobank	0.0171	0.286
rs4328080	1	219,963,088	G	A	UK Biobank	0.0555	0.932
rs6688537	1	239,850,588	A	C	COPD case-control studies	0.0277	0.465
rs62126408	2	18,309,132	T	C	UK Biobank	0.1087	1.826
rs1430193	2	56,120,853	T	A	UK Biobank	4.97E-05	0.001
rs2571445	2	218,683,154	A	G	UK Biobank	0.0865	1.453
rs10498230	2	229,502,503	C	T	UK Biobank	0.1024	1.719
rs61332075	2	239,316,560	G	C	COPD case-control studies	0.0814	1.367
rs12477314	2	239,877,148	C	T	UK Biobank	0.0833	1.400
rs1529672	3	25,520,582	C	A	UK Biobank	0.0500	0.840
rs1458979	3	55,150,677	G	A	COPD case-control studies	0.0261	0.439
rs1490265	3	67,452,043	C	A	COPD case-control studies	0.0064	0.107
rs2811415	3	127,991,527	G	A	COPD case-control studies	0.2078	3.490
rs1595029	3	158,241,767	C	A	UK Biobank	0.0317	0.533
rs56341938*	3	168,715,808	G	A	COPD case-control studies	4.97E-05	0.001
rs1344555	3	169,300,219	T	C	UK Biobank	0.0247	0.416
rs13110699	4	89,815,695	G	T	COPD case-control studies	0.1933	3.246
rs2045517	4	89,870,964	T	C	UK Biobank	0.0782	1.314
rs2047409*	4	106,137,033	G	A	lungeQTL+deCODE	4.97E-05	0.001
rs10516526	4	106,688,904	A	G	UK Biobank	0.1086	1.824
rs34712979	4	106,819,053	A	G	COPD case-control studies	0.1792	3.009
rs138641402	4	145,445,779	A	T	UK Biobank	0.1628	2.733
rs91731	5	33,334,312	A	C	COPD case-control studies	0.0222	0.372
rs1551943	5	52,195,033	A	G	COPD case-control studies	0.1291	2.169
rs2441026	5	53,444,498	C	T	COPD case-control studies	0.0211	0.354
rs153916	5	95,036,700	T	C	UK Biobank	0.0405	0.680
rs7713065	5	131,788,334	A	C	COPD case-control studies	0.0032	0.054
rs7715901	5	147,856,392	A	G	UK Biobank	0.1252	2.102
rs3839234	5	148,596,693	T	TG	COPD case-control studies	0.0172	0.289
rs10515750	5	156,810,072	T	C	COPD case-control studies	0.1836	3.084

rs1990950	5	156,920,756	G	T	UK Biobank	0.0752	1.263
rs6924424	6	7,801,611	G	T	UK Biobank	0.0056	0.093
rs34864796	6	27,459,923	A	G	UK Biobank	0.1507	2.530
rs28986170	6	31,556,155	G	GAA	COPD case-control studies	4.97E-05	0.001
rs2857595	6	31,568,469	A	G	UK Biobank	0.1087	1.825
rs2070600	6	32,151,443	C	T	UK Biobank	0.1825	3.064
rs114544105	6	32,635,629	A	G	lungeQTL+deCODE	0.0575	0.965
rs114229351	6	32,648,418	C	T	lungeQTL+deCODE	0.0231	0.389
rs141651520	6	73,670,095	ATTCTAT	A	COPD case-control studies	0.0251	0.422
rs2768551	6	109,270,656	A	G	UK Biobank	0.0662	1.112
rs7753012	6	142,745,883	T	G	UK Biobank	0.1540	2.586
rs148274477	6	142,838,173	C	T	UK Biobank	0.2439	4.095
rs10246303	7	7,286,445	T	A	COPD case-control studies	0.0444	0.745
rs72615157	7	99,635,967	G	A	COPD case-control studies	0.0100	0.168
rs12698403	7	156,127,246	A	G	COPD case-control studies	0.0947	1.590
rs7872188	9	4,124,377	T	C	COPD case-control studies	0.0254	0.427
rs16909859	9	98,204,792	A	G	UK Biobank	0.0618	1.038
rs803923	9	119,401,650	A	G	UK Biobank	0.0519	0.871
rs10858246	9	139,102,831	C	G	UK Biobank	0.0245	0.411
rs10870202	9	139,257,411	C	T	COPD case-control studies	4.97E-05	0.001
rs7090277	10	12,278,021	T	A	UK Biobank	0.0995	1.671
rs3847402	10	30,267,810	A	G	COPD case-control studies	0.0564	0.947
rs7095607	10	69,957,350	A	G	COPD case-control studies	0.0355	0.596
rs2637254	10	78,312,002	A	G	UK Biobank	0.0773	1.298
rs4237643	11	43,648,368	T	G	UK Biobank	0.0253	0.424
rs2863171	11	45,250,732	A	C	UK Biobank	0.0507	0.851
rs2509961	11	62,310,909	T	C	COPD case-control studies	0.0168	0.283
rs145729347*	11	86,442,733	G	C	deCODE	0.0377	0.633
rs567508	11	126,008,910	G	A	COPD case-control studies	0.0081	0.136
rs2348418	12	28,689,514	C	T	UK Biobank	0.0201	0.338
rs11172113	12	57,527,283	T	C	UK Biobank	0.0386	0.649
rs1494502	12	65,824,670	A	G	COPD case-control studies	0.0721	1.211
rs113745635	12	95,554,771	T	C	COPD case-control studies	0.0728	1.223
rs12820313	12	96,255,704	C	T	UK Biobank	0.0846	1.420
rs10850377	12	115,201,436	G	A	UK Biobank	0.0205	0.345
rs35506	12	115,500,691	T	A	COPD case-control studies	4.97E-05	0.001
rs1698268	14	84,309,664	T	A	COPD case-control studies	0.0139	0.233
rs7155279	14	92,485,881	G	T	UK Biobank	0.0594	0.998
rs117068593	14	93,118,229	C	T	UK Biobank	0.0443	0.743
rs72724130	15	41,977,690	T	A	COPD case-control studies	0.1461	2.454
rs10851839	15	71,628,370	T	A	UK Biobank	0.1144	1.921
rs12591467	15	71,788,387	C	T	COPD case-control studies	0.0638	1.072
rs66650179	15	84,261,689	C	CA	deCODE	0.0387	0.651
rs12149828	16	10,706,328	A	G	UK Biobank	0.0675	1.134
rs12447804	16	58,075,282	T	C	UK Biobank	0.0274	0.460
rs3743609	16	75,467,021	C	G	UK Biobank	0.0704	1.182
rs1079572	16	78,187,138	A	G	UK Biobank	0.0026	0.044
rs59835752	17	28,265,330	TA	T	deCODE	4.97E-05	0.001

rs11658500	17	36,886,828	A	G	COPD case-control studies	0.0721	1.210
rs35524223	17	44,192,590	A	T	lungeQTL+deCODE	0.0080	0.134
rs6501431	17	68,976,415	C	T	UK Biobank	4.97E-05	0.001
rs7218675	17	73,513,185	A	C	COPD case-control studies	4.97E-05	0.001
rs113473882	19	41,124,155	T	C	UK Biobank	0.1620	2.721
rs6140050	20	6,632,901	C	A	COPD case-control studies	0.0154	0.258
rs72448466	20	62,363,640	C	CGT	COPD case-control studies	0.0371	0.622
rs2834440	21	35,690,499	G	A	UK Biobank	0.0691	1.160
rs11704827	22	18,450,287	A	T	COPD case-control studies	0.0184	0.310
rs134041	22	28,056,338	T	C	UK Biobank	0.0645	1.084
rs2283847	22	28,181,399	T	C	COPD case-control studies	0.0329	0.553

Acknowledgements and Funding

M.D. Tobin is supported by MRC fellowships (G0501942 and G0902313). M.D. Tobin and L.V. Wain are supported by the MRC (MR/N011317/1). M.D. Tobin and C. Brightling are both supported by AirPROM. I.P. Hall and I. Sayers are supported by the MRC (G1000861). L. Bossini-Castillo is supported by the Medical Research Council (MR/N014995/1). M. Obeidat is a Postdoctoral Fellow of the Michael Smith Foundation for Health Research (MSFHR) and the Canadian Institute for Health Research (CIHR) Integrated and Mentored Pulmonary and Cardiovascular Training program (IMPACT). He is also a recipient of British Columbia Lung Association Research Grant. E. Zeggini and B.P. Prins are supported the Economic & Social Research Council (ES/H029745/1) and the Wellcome Trust (WT098051). Generation Scotland was funded by the Scottish Executive Health Department, Chief Scientist Office (CZD/16/6) and the Scottish Funding Council (HR03006). Genotyping was funded by the MRC and the Wellcome Trust. We acknowledge use of phenotype and genotype data from the British 1958 Birth Cohort DNA collection, funded by the MRC (G0000934) and the Wellcome Trust (068545/Z/02). Genotyping for the B58C-WTCCC subset was funded by the Wellcome Trust (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. NFBC1966 received financial support from the Academy of Finland (project grants 104781, 120315, 129269, 1114194, 24300796, Center of Excellence in Complex Disease Genetics and SALVE), University Hospital Oulu, Biocenter, University of Oulu, Finland (75617), NHLBI grant 5R01HL087679-02 through the STAMPEED program (1RL1MH083268-01), NIH/NIMH (5R01MH63706:02), ENGAGE project and grant agreement HEALTH-F4-2007-201413, EU FP7 EurHEALTHAgeing -277849, the Medical Research Council, UK (G0500539, G0600705, G1002319, PrevMetSyn/SALVE) and the MRC, Centenary Early Career Award. The program is currently being funded by the H2020-633595 DynaHEALTH action and academy of Finland EGEA-project (285547) and EU H2020 ALEC project (Grant Agreement 633212). The EPIC Norfolk Study is funded by Cancer Research UK and the MRC. ORCADES was supported by the Chief Scientist Office of the Scottish Government (CZB/4/276, CZB/4/710), the Royal Society, the MRC Human Genetics Unit, Arthritis Research UK and the European Union framework program 6 EUROSPAN project (contract no. LSHG-CT-2006-018947). DNA extractions were performed at the Wellcome Trust Clinical Research Facility in Edinburgh. SHIP is part of the Community Medicine Research net (CMR) of the University of Greifswald, Germany, which is funded by the Federal Ministry of Education and Research (ZZ9603, 01ZZ0103, 01ZZ0403), Competence Network Asthma/ COPD (FKZ 01GI0881-0888), the Ministry of Cultural Affairs as well as the Social Ministry of the Federal State of Mecklenburg-West Pomerania. The CMR encompasses several research projects which are sharing data of the population-based Study of Health in Pomerania (SHIP; http://ship.community-medicine.de). The Cooperative Health Research in the region of Augsburg (KORA) research platform was initiated and financed by the Helmholtz Zentrum München – German Research Center for Environmental Health, which is funded by the German Federal Ministry of Education and Research and by the State of Bavaria. This work was supported by the Competence Network Asthma and COPD (ASCONET), network COSYCONET (subproject 2, BMBF FKZ 01GI0882), and the KORA Age project (FKZ 01ET0713 and 01ET1003A) funded by the German Federal Ministry of Education and Research (BMBF). SAPALDIA is funded by the Swiss National Science Foundation (33CS30-148470/1&2, 33CSCO-134276/1, 33CSCO-108796, 324730_135673, 3247BO-104283, 3247BO-104288, 3247BO-104284, 3247-065896, 3100-059302, 3200-052720, 3200-042532, 4026-028099, PMPDP3_129021/1, PMPDP3_141671/1), the Federal Office for the Environment, the Federal Office of Public Health, the Federal Office of Roads and Transport, the canton's government of Aargau, Basel-Stadt, Basel-Land, Geneva, Luzern, Ticino, Valais, and Zürich, the Swiss Lung League, the canton's Lung League of Basel Stadt/ Basel Landschaft, Geneva, Ticino,

Valais, Graubünden and Zurich, Stiftung ehemals Bündner Heilstätten, SUVA, Freiwillige Akademische Gesellschaft, UBS Wealth Foundation, Talecris Biotherapeutics GmbH, Abbott Diagnostics, European Commission 018996 (GABRIEL) and the Wellcome Trust (WT 084703MA). Phenotype collection in the Lothian Birth Cohort 1936 was supported by Age UK (The Disconnected Mind project). Genotyping was funded by the Biotechnology and Biological Sciences Research Council (BBSRC). The work was undertaken by The University of Edinburgh Centre for Cognitive Ageing and Cognitive Epidemiology, part of the cross council Lifelong Health and Wellbeing Initiative (MR/K026992/1). Funding from the BBSRC and MRC is gratefully acknowledged. I. Rudan, C. Hayward, S.M. Kerr, O. Polasek, V. Vitart, and J. Marten are funded by the MRC, the Ministry of Science, Education and Sport in the Republic of Croatia (216-1080315-0302) and the Croatian Science Foundation (grant 8875). The Northern Swedish Population Health Study (NSPHS) was funded by the Swedish Medical Research Council (K2007-66X-20270-01-3, 2011-5252, 2012-2884 and 2011-2354), the Foundation for Strategic Research (SSF). NSPHS as part of European Special Populations Research Network (EUROSPAN) was also supported by the European Commission FP6 STRP (01947, LSHG-CT-2006-01947). Health 2000 was financially supported by the Medical Research Fund of the Tampere University Hospital. The UK Medical Research Council and the Wellcome Trust (Grant ref: 102215/2/13/2) and the University of Bristol provide core support for ALSPAC. ALSPAC GWAS data was generated by Sample Logistics and Genotyping Facilities at the Wellcome Trust Sanger Institute and LabCorp (Laboratory Corporation of America) using support from 23andMe. Lung function data collection was funded by MRC (G0401540). The COPDGene project (NCT00608764) was supported by Award Number R01HL089897 and Award Number R01HL089856 from the National Heart, Lung, And Blood Institute. The content 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 COPDGene project is also supported by the COPD Foundation through contributions made to an Industry Advisory Board comprised of AstraZeneca, Boehringer Ingelheim, GlaxoSmithKline, Novartis, Pfizer, Siemens and Sunovion. The ECLIPSE study (NCT00292552; GSK code SCO104960) was funded by GSK. The Norway GenKOLS study (Genetics of Chronic Obstructive Lung Disease, GSK code RES11080) was funded by GSK. 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). M.H. Cho is supported by NHLBI R01HL113264. The China Kadoorie Biobank prospective cohort (CKB) has received the following funding: Baseline survey: Kadoorie Charitable Foundation, Hong Kong. Long-term continuation: UK Wellcome Trust (088158/Z/09/Z, 104085/Z/14/Z), Chinese National Natural Science Foundation (81390541). DNA extraction and genotyping: GlaxoSmithKline, Merck & Co. Inc., UK Medical Research Council (MC_PC_13049). The British Heart Foundation, UK Medical Research Council and Cancer Research UK provide core funding to CTSU. J. Vaucher is supported by the Swiss National Science Foundation (P2LAP3_155086) for a postdoctoral research fellowship at the University of Oxford, UK. G.Trynka is supported by the Wellcome Trust (WT098051). A.P. Morris is a Wellcome Trust Senior Fellow in Basic Biomedical Science (WT098017). The Raine study was supported by the National Health and Medical Research Council of Australia [grant numbers 403981, 003209 and 572613] and the Canadian Institutes of Health Research [grant number MOP-82893]. The Lung Health Study I was supported by contract NIH/N01-HR-46002 and genotyping by GENEVA (U01HG004738).
The UK Household Longitudinal Study is led by the Institute for Social and Economic Research at the University of Essex and funded by the Economic and Social Research Council. The survey was conducted by NatCen and the genome-wide scan data were analysed and deposited by the Wellcome Trust Sanger Institute. Information on how to access the data can be found on the Understanding Society website https://www.understandingsociety.ac.uk/. The Busselton Health Study (BHS) acknowledges the generous support for the 1994/5 follow-up study from Healthway, Western Australia and the numerous Busselton community volunteers who assisted with data collection and the study participants from the Shire of Busselton. The Busselton Health Study is supported by The Great Wine Estates of the Margaret River region of Western Australia. The SAPALDIA study could not have been done without the help of the study

participants, technical and administrative support and the medical teams and field workers at the local study sites. Local fieldworkers : Aarau: S Brun, G Giger, M Sperisen, M Stahel, Basel: C Bürli, C Dahler, N Oertli, I Harreh, F Karrer, G Novicic, N Wyttenbacher, Davos: A Saner, P Senn, R Winzeler, Geneva: F Bonfils, B Blicharz, C Landolt, J Rochat, Lugano: S Boccia, E Gehrig, MT Mandia, G Solari, B Viscardi, Montana: AP Bieri, C Darioly, M Maire, Payerne: F Ding, P Danieli A Vonnez, Wald: D Bodmer, E Hochstrasser, R Kunz, C Meier, J Rakic, U Schafroth, A Walder. China Kadoorie Biobank acknowledges the participants, the project staff, and the China National Centre for Disease Control and Prevention (CDC) and its regional offices for access to death and disease registries. The Chinese National Health Insurance scheme provides electronic linkage to all hospital treatment. We are extremely grateful to all the families who took part in the ALSPAC study, the midwives for their help in recruiting them, and the whole ALSPAC team, which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists and nurses. The authors are grateful to the Raine Study participants and their families, and to the Raine Study research staff for cohort coordination and data collection. The authors gratefully acknowledge the NH&MRC for their long term contribution to funding the study over the last 20 years and also the following Institutions for providing funding for Core Management of the Raine Study: The University of Western Australia (UWA), Curtin University, Raine Medical Research Foundation, UWA Faculty of Medicine, Dentistry and Health Sciences, the Telethon Kids Institute, the Women and Infants Research Foundation and Edith Cowan University. This work was supported by resources provided by the Pawsey Supercomputing Centre with funding from the Australian Government and the Government of Western Australia. 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. The principal investigators and senior staff of the clinical and coordinating centers, the NHLBI, and members of the Safety and Data Monitoring Board of the Lung Health Study are as follows: Case Western Reserve University, Cleveland, OH: M.D. Altose, M.D. (Principal Investigator), C.D. Deitz, Ph.D. (Project Coordinator); Henry Ford Hospital, Detroit, MI: M.S. Eichenhorn, M.D. (Principal Investigator), K.J. Braden, A.A.S. (Project Coordinator), R.L. Jentons, M.A.L.L.P. (Project Coordinator); Johns Hopkins University School of Medicine, Baltimore, MD: R.A. Wise, M.D. (Principal Investigator), C.S. Rand, Ph.D. (Co-Principal Investigator), K.A. Schiller (Project Coordinator); Mayo Clinic, Rochester, MN: P.D. Scanlon, M.D. (Principal Investigator), G.M. Caron (Project Coordinator), K.S. Mieras, L.C. Walters; Oregon Health Sciences University, Portland: A.S. Buist, M.D. (Principal Investigator), L.R. Johnson, Ph.D. (LHS Pulmonary Function Coordinator), V.J. Bortz (Project Coordinator); University of Alabama at Birmingham: W.C. Bailey, M.D. (Principal Investigator), L.B. Gerald, Ph.D., M.S.P.H. (Project Coordinator); University of California, Los Angeles: D.P. Tashkin, M.D. (Principal Investigator), I.P. Zuniga (Project Coordinator); University of Manitoba, Winnipeg: N.R. Anthonisen, M.D. (Principal Investigator, Steering Committee Chair), J. Manfreda, M.D. (Co-Principal Investigator), R.P. Murray, Ph.D. (Co-Principal Investigator), S.C. Rempel-Rossum (Project Coordinator); University of Minnesota Coordinating Center, Minneapolis: J.E. Connett, Ph.D. (Principal Investigator), P.L. Enright, M.D., P.G. Lindgren, M.S., P. O'Hara, Ph.D., (LHS Intervention Coordinator), M.A. Skeans, M.S., H.T. Voelker; University of Pittsburgh, Pittsburgh, PA: R.M. Rogers, M.D. (Principal Investigator), M.E. Pusateri (Project Coordinator); University of Utah, Salt Lake City: R.E. Kanner, M.D. (Principal Investigator), G.M. Villegas (Project Coordinator); Safety and Data Monitoring Board: M. Becklake, M.D., B. Burrows, M.D. (deceased), P. Cleary, Ph.D., P. Kimbel, M.D. (Chairperson; deceased), L. Nett, R.N., R.R.T. (former member), J.K. Ockene, Ph.D., R.M. Senior, M.D. (Chairperson), G.L. Snider, M.D., W. Spitzer, M.D. (former member), O.D. Williams, Ph.D.; Morbidity and Mortality Review Board: T.E. Cuddy, M.D., R.S. Fontana, M.D., R.E. Hyatt, M.D., C.T. Lambrew, M.D., B.A. Mason, M.D., D.M. Mintzer, M.D., R.B. Wray, M.D.; National Heart, Lung, and Blood Institute staff, Bethesda, MD: S.S. Hurd, Ph.D. (Former Director, Division of Lung Diseases), J.P. Kiley, Ph.D. (Former Project Officer and Director, Division of Lung Diseases), G. Weinmann, M.D. (Former Project Officer and Director, Airway Biology and Disease Program, DLD), M.C. Wu, Ph.D. (Division of Epidemiology and Clinical Applications).

Cohort contributors

Understanding Society Scientific Group

Michaela Benzeval[1], Jonathan Burton[1], Nicholas Buck[1], Annette Jäckle[1], Meena Kumari[1], Heather Laurie[1], Peter Lynn[1], Stephen Pudney[1], Birgitta Rabe[1], Dieter Wolke[2]
[1]Institute for Social and Economic Research, [2]University of Warwick

COPDGene

COPDGene Investigators – Core Units

Administrative Center: James D. Crapo, MD (PI); Edwin K. Silverman, MD, PhD (PI); Barry J. Make, MD; Elizabeth A. Regan, MD, PhD
Genetic Analysis Center: Terri Beaty, PhD; Ferdouse Begum, PhD; Robert Busch, MD; Peter J. Castaldi, MD, MSc; Michael Cho, MD; Dawn L. DeMeo, MD, MPH; Adel R. Boueiz, MD; Marilyn G. Foreman, MD, MS; Eitan Halper-Stromberg; Nadia N. Hansel, MD, MPH; Megan E. Hardin, MD; Craig P. Hersh, MD, MPH; Jacqueline Hetmanski, MS, MPH; Brian D. Hobbs, MD; John E. Hokanson, MPH, PhD; Nan Laird, PhD; Christoph Lange, PhD; Sharon M. Lutz, PhD; Merry-Lynn McDonald, PhD; Margaret M. Parker, PhD; Dandi Qiao, PhD; Elizabeth A. Regan, MD, PhD; Stephanie Santorico, PhD; Edwin K. Silverman, MD, PhD; Emily S. Wan, MD; Sungho Won
Imaging Center: Mustafa Al Qaisi, MD; Harvey O. Coxson, PhD; Teresa Gray; MeiLan K. Han, MD, MS; Eric A. Hoffman, PhD; Stephen Humphries, PhD; Francine L. Jacobson, MD, MPH; Philip F. Judy, PhD; Ella A. Kazerooni, MD; Alex Kluiber; David A. Lynch, MB; John D. Newell, Jr., MD; Elizabeth A. Regan, MD, PhD; James C. Ross, PhD; Raul San Jose Estepar, PhD; Joyce Schroeder, MD; Jered Sieren; Douglas Stinson; Berend C. Stoel, PhD; Juerg Tschirren, PhD; Edwin Van Beek, MD, PhD; Bram van Ginneken, PhD; Eva van Rikxoort, PhD; George Washko, MD; Carla G. Wilson, MS;
PFT QA Center, Salt Lake City, UT: Robert Jensen, PhD
Data Coordinating Center and Biostatistics, National Jewish Health, Denver, CO: Douglas Everett, PhD; Jim Crooks, PhD; Camille Moore, PhD; Matt Strand, PhD; Carla G. Wilson, MS
Epidemiology Core, University of Colorado Anschutz Medical Campus, Aurora, CO: John E. Hokanson, MPH, PhD; John Hughes, PhD; Gregory Kinney, MPH, PhD; Sharon M. Lutz, PhD; Katherine Pratte, MSPH; Kendra A. Young, PhD

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; Mustafa Atik, MD; Venkata Bandi, MD; Aladin Boriek, PhD; Kalpatha Guntupalli, MD; Elizabeth Guy, MD; Arun Nachiappan, MD; Amit Parulekar, MD;
Brigham and Women's Hospital, Boston, MA: Dawn L. DeMeo, MD, MPH; Craig Hersh, MD, MPH; Francine L. Jacobson, MD, MPH; George Washko, MD
Columbia University, New York, NY: R. Graham Barr, MD, DrPH; John Austin, MD; Belinda D'Souza, MD; Gregory D.N. Pearson, MD; Anna Rozenshtein, MD, MPH, FACR; Byron Thomashow, MD
Duke University Medical Center, Durham, NC: Neil MacIntyre, Jr., MD; H. Page McAdams, MD; Lacey Washington, MD
HealthPartners Research Institute, Minneapolis, MN: Charlene McEvoy, MD, MPH; Joseph Tashjian, MD
Johns Hopkins University, Baltimore, MD: Robert Wise, MD; Robert Brown, MD; Nadia N. Hansel, MD, MPH; Karen Horton, MD; Allison Lambert, MD, MHS; Nirupama Putcha, MD, MHS
Los Angeles Biomedical Research Institute at Harbor UCLA Medical Center, Torrance, CA: Richard Casaburi, PhD, MD; Alessandra Adami, PhD; Matthew Budoff, MD; Hans Fischer, MD; Janos Porszasz, MD, PhD; Harry Rossiter, PhD; William Stringer, MD
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 G. Foreman, MD, MS; Eugene Berkowitz, MD, PhD; Gloria Westney, MD, MS
National Jewish Health, Denver, CO: Russell Bowler, MD, PhD; David A. Lynch, MB

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; Surya Bhatt, MD; Anand Iyer, MD; Hrudaya Nath, MD; J. Michael Wells, 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 P. Comellas, MD; John Newell, Jr., MD; Brad Thompson, MD
University of Michigan, Ann Arbor, MI: MeiLan K. Han, MD, MS; Ella Kazerooni, MD; Carlos H. Martinez, MD, MPH
University of Minnesota, Minneapolis, MN: Joanne Billings, MD; Abbie Begnaud, MD; Tadashi Allen, MD
University of Pittsburgh, Pittsburgh, PA: Frank Sciurba, MD; Jessica Bon, MD; Divay Chandra, MD, MSc; Carl Fuhrman, MD; Joel Weissfeld, MD, MPH
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

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).

Lung Health Study (LHS)

The principal investigators and senior staff of the clinical and coordinating centers, the NHLBI, and members of the Safety and Data Monitoring Board of the Lung Health Study are as follows:
Case Western Reserve University, Cleveland, OH: M.D. Altose, M.D. (Principal Investigator), C.D. Deitz, Ph.D. (Project Coordinator); Henry Ford Hospital, Detroit, MI: M.S. Eichenhorn, M.D. (Principal Investigator), K.J. Braden, A.A.S. (Project Coordinator), R.L. Jentons, M.A.L.L.P. (Project Coordinator); Johns Hopkins University School of Medicine, Baltimore, MD: R.A. Wise, M.D. (Principal Investigator), C.S. Rand, Ph.D. (Co-Principal Investigator), K.A. Schiller (Project Coordinator); Mayo Clinic, Rochester, MN: P.D. Scanlon, M.D. (Principal Investigator), G.M. Caron (Project Coordinator), K.S. Mieras, L.C. Walters; Oregon Health Sciences University, Portland: A.S. Buist, M.D. (Principal Investigator), L.R. Johnson, Ph.D. (LHS Pulmonary Function Coordinator), V.J. Bortz (Project Coordinator); University of Alabama at Birmingham: W.C. Bailey, M.D. (Principal Investigator), L.B. Gerald, Ph.D., M.S.P.H. (Project Coordinator); University of California, Los Angeles: D.P. Tashkin, M.D. (Principal Investigator), I.P. Zuniga (Project Coordinator); University of Manitoba, Winnipeg: N.R. Anthonisen, M.D. (Principal Investigator, Steering Committee Chair), J. Manfreda, M.D. (Co-Principal Investigator), R.P. Murray, Ph.D.

(Co-Principal Investigator), S.C. Rempel-Rossum (Project Coordinator); University of Minnesota Coordinating Center, Minneapolis: J.E. Connett, Ph.D. (Principal Investigator), P.L. Enright, M.D., P.G. Lindgren, M.S., P. O'Hara, Ph.D., (LHS Intervention Coordinator), M.A. Skeans, M.S., H.T. Voelker; University of Pittsburgh, Pittsburgh, PA: R.M. Rogers, M.D. (Principal Investigator), M.E. Pusateri (Project Coordinator); University of Utah, Salt Lake City: R.E. Kanner, M.D. (Principal Investigator), G.M. Villegas (Project Coordinator); Safety and Data Monitoring Board: M. Becklake, M.D., B. Burrows, M.D. (deceased), P. Cleary, Ph.D., P. Kimbel, M.D. (Chairperson; deceased), L. Nett, R.N., R.R.T. (former member), J.K. Ockene, Ph.D., R.M. Senior, M.D. (Chairperson), G.L. Snider, M.D., W. Spitzer, M.D. (former member), O.D. Williams, Ph.D.; Morbidity and Mortality Review Board: T.E. Cuddy, M.D., R.S. Fontana, M.D., R.E. Hyatt, M.D., C.T. Lambrew, M.D., B.A. Mason, M.D., D.M. Mintzer, M.D., R.B. Wray, M.D.; National Heart, Lung, and Blood Institute staff, Bethesda, MD: S.S. Hurd, Ph.D. (Former Director, Division of Lung Diseases), J.P. Kiley, Ph.D. (Former Project Officer and Director, Division of Lung Diseases), G. Weinmann, M.D. (Former Project Officer and Director, Airway Biology and Disease Program, DLD), M.C. Wu, Ph.D. (Division of Epidemiology and Clinical Applications

Geisinger-Regeneron DiscovEHR Collaboration

URL: http://www.discovehrshare.com

References

1. Abecasis, G.R. *et al.* An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56-65 (2012).
2. Walter, K. *et al.* The UK10K project identifies rare variants in health and disease. *Nature* **526**, 82-89 (2015).
3. Huang, J. *et al.* Improved imputation of low-frequency and rare variants using the UK10K haplotype reference panel. *Nature Communications* **6**(2015).
4. Delaneau, O., Zagury, J.F. & Marchini, J. Improved whole-chromosome phasing for disease and population genetic studies. *Nat Methods* **10**, 5-6 (2013).
5. 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).
6. Global Initiative for Chronic Obstructive Lung Disease. Global Strategy for the Diagnosis Management and Prevention of COPD. *http://goldcopd.org/* (2015).
7. Marchini, J. & Band, G. SNPTEST, https://mathgen.stats.ox.ac.uk/genetics_software/snptest/snptest.html. (2016).
8. Styrkarsdottir, U. *et al.* Nonsense mutation in the LGR4 gene is associated with several human diseases and other traits. *Nature* **497**, 517-20 (2013).
9. Gudbjartsson, D.F. *et al.* Large-scale whole-genome sequencing of the Icelandic population. *Nat Genet* **47**, 435-44 (2015).
10. Bulik-Sullivan, B.K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet* **47**, 291-295 (2015).
11. Hao, K. *et al.* Lung eQTLs to help reveal the molecular underpinnings of asthma. *PLoS Genet* **8**, e1003029 (2012).
12. Obeidat, M. *et al.* GSTCD and INTS12 regulation and expression in the human lung. *PLoS One* **8**, e74630 (2013).
13. Irizarry, R.A. *et al.* Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* **4**, 249-64 (2003).
14. 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).
15. Lamontagne, M. *et al.* Refining susceptibility loci of chronic obstructive pulmonary disease with lung eqtls. *PLoS One* **8**, e70220 (2013).
16. Regan, E.A. *et al.* Genetic epidemiology of COPD (COPDGene) study design. *COPD* **7**, 32-43 (2010).
17. 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).
18. Vestbo, J. *et al.* Evaluation of COPD Longitudinally to Identify Predictive Surrogate End-points (ECLIPSE). *Eur Respir J* **31**, 869-73 (2008).
19. Cho, M.H. *et al.* Variants in FAM13A are associated with chronic obstructive pulmonary disease. *Nat Genet* **42**, 200-2 (2010).
20. 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).
21. Bell, B., Rose, C. L. & Damon, H. The Normative Aging Study: an interdisciplinary and longitudinal study of health and aging. *Aging Hum Dev* **3**, 5–17 (1972).
22. Pillai, S.G. *et al.* A genome-wide association study in chronic obstructive pulmonary disease (COPD): identification of two major susceptibility loci. *PLoS Genet* **5**, e1000421 (2009).
23. Dewey, F.E. *et al.* Inactivating Variants in ANGPTL4 and Risk of Coronary Artery Disease. *N Engl J Med* **374**, 1123-33 (2016).
24. Chen, Z. *et al.* China Kadoorie Biobank of 0.5 million people: survey methods, baseline characteristics and long-term follow-up. *Int J Epidemiol* **40**, 1652-66 (2011).
25. Quanjer, P.H. *et al.* Multi-ethnic reference values for spirometry for the 3-95-yr age range: the global lung function 2012 equations. *Eur Respir J* **40**, 1324-43 (2012).
26. Anthonisen, N.R. *et al.* Effects of smoking intervention and the use of an inhaled anticholinergic bronchodilator on the rate of decline of FEV1. The Lung Health Study. *JAMA* **272**, 1497-505 (1994).
27. Kanner, R.E., Connett, J.E., Williams, D.E. & Buist, A.S. Effects of randomized assignment to a smoking cessation intervention and changes in smoking habits on respiratory symptoms in smokers with early chronic obstructive pulmonary disease: the Lung Health Study. *Am J Med* **106**, 410-6 (1999).

28. Hansel, N.N. *et al.* Genome-wide study identifies two loci associated with lung function decline in mild to moderate COPD. *Hum Genet* **132**, 79-90 (2013).
29. The 1000 Genomes Project consortium *et al.* An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56-65 (2012).
30. Anthonisen, N.R., Connett, J.E., Enright, P.L. & Manfreda, J. Hospitalizations and mortality in the Lung Health Study. *Am J Respir Crit Care Med* **166**, 333-9 (2002).
31. Boyd, A. *et al.* Cohort Profile: the 'children of the 90s'--the index offspring of the Avon Longitudinal Study of Parents and Children. *Int J Epidemiol* **42**, 111-27 (2013).
32. Cremers, E., Thijs, C., Penders, J., Jansen, E. & Mommers, M. Maternal and child's vitamin D supplement use and vitamin D level in relation to childhood lung function: the KOALA Birth Cohort Study. *Thorax* (2011).
33. Kotecha, S.J. *et al.* Spirometric lung function in school-age children: effect of intrauterine growth retardation and catch-up growth. *American journal of respiratory and critical care medicine* **181**, 969-974 (2010).
34. Kemp, J.P. *et al.* Phenotypic dissection of bone mineral density reveals skeletal site specificity and facilitates the identification of novel loci in the genetic regulation of bone mass attainment. *PLoS Genet* **10**, e1004423 (2014).
35. 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).
36. Price, A.L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* **38**, 904-909 (2006).
37. Li, Y., Willer, C., Sanna, S. & Abecasis, G. Genotype Imputation. *Annu. Rev. Genom. Human Genet.* **10**, 387-406 (2011).
38. Li, Y., Willer, C.J., Ding, J., Scheet, P. & Abecasis, G.R. MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genetic Epidemiology* **34**, 816-834 (2010).
39. International HapMap Consortium *et al.* A second generation human haplotype map of over 3.1 million SNPs. *Nature* **449**, 851-61 (2007).
40. Soler Artigas, M. *et al.* Genome-wide association and large-scale follow up identifies 16 new loci influencing lung function. *Nat Genet* **43**, 1082-90 (2011).
41. Soler Artigas, M. *et al.* Sixteen new lung function signals identified through 1000 Genomes Project reference panel imputation. *Nat Commun* **6**, 8658 (2015).
42. Wilk, J.B. *et al.* A genome-wide association study of pulmonary function measures in the Framingham Heart Study. *PLoS Genet* **5**, e1000429 (2009).
43. Repapi, E. *et al.* Genome-wide association study identifies five loci associated with lung function. *Nat Genet* **42**, 36-44 (2010).
44. 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).
45. Loth, D.W. *et al.* Genome-wide association analysis identifies six new loci associated with forced vital capacity. *Nat Genet* **46**, 669-77 (2014).
46. 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).
47. Wakefield, J. A Bayesian Measure of the Probability of False Discovery in Genetic Epidemiology Studies. *The American Journal of Human Genetics* **81**, 208-227 (2007).