Supplementary Information

Contents

Supplementary Figures

Supplementary Tables

Supplementary Note

Selecting individuals from UK Biobank

Spirometry Quality Control

UK Biobank contains data for 502,682 individuals. Of these, 445,754 had at least two measures of FEV_1 (VariableID: 3063) and FVC (VariableID: 3062), complete information for spirometry method used (VariableID: 23), age (VariableID: 21022), sex (VariableID: 31) standing height (VariableID: 50), and for whom ever smoking status could be derived (derivation of ever smoking status described below). For quality control of spirometry, the pre-derived FEV_1, FVC and PEF measurements (VariableIDs: 3063, 3062 and 3064), the blow curve time series measurements (VariableID: 3066) and the Vitalograph spirometer blow quality metrics (VariableID: 20031) were used.

Acceptability of blows

To identify "acceptable" blows for inclusion in the analyses of FEV_1, FVC, FEV_1/FVC and PEF, the following quality control steps were undertaken;

- Blows were initially deemed to be acceptable if they contained the following values in the Vitalograph spirometer blow quality metrics; "blank", "ACCEPT", BELOW6SEC ACCEPT" and "BELOW6SEC". A total of 777,676 blows from 387,430 participants were deemed acceptable.
- Next, start of blow quality was examined. Blows were excluded if the back-extrapolated volume (as defined using the blow curve time series measurements [1]) was less than 5% of FVC or less than 150ml. Following this exclusion, a total of 776,927 blows from 387,277 participants remained.
- Finally, a comparison of the pre-derived FEV_1 and FVC measurements (VariableID: 3063 and VariableID: 3062) and FEV_1 and FVC newly derived from the blow curve time series measurements (VariableID: 3066) was undertaken. Blows where the pre-derived and newly-derived values differed by 5% were excluded. Following this exclusion, a total of 776,318 "acceptable" blows from 387,052 participants remained for further analysis of FEV_1, FVC, FEV_1/FVC and PEF. Whilst PEF was also pre-derived, we identified a subset of individuals had unusually low recorded values, which were inconsistent with the PEF values derived from the time series curves; the predefined PEF values were deemed to be erroneous, therefore no exclusions were undertaken based on comparisons of pre-derived and newly-derived PEF, and the newly-derived PEF values were used for association analyses.

Identification of best measures

The "best measure" per individual was defined as the highest measure from the "acceptable" blows for FEV_1, FVC. FEV_1/FVC was derived from the selected FEV_1 and FVC. For PEF, which is a measure of flow, the best measure was defined as the blow with the highest acceptable measure of the sum of FEV_1 and FVC. This definition meant that a participant's "best measures" did not necessarily have to be derived from the same blow.

Reproducibility of measures

To meet the criterion for reproducibility in our analysis, the "best measures" of FEV_1 and FVC had to be within 250ml of those measures from any other blow. The other blow did not need to be acceptable. Where an individual's best measures for FEV_1 and FVC were not both found to be reproducible, that individual was excluded. 348,936 individuals had acceptable and reproducible measures of both FEV_1 and FVC and were eligible for inclusion in analyses of FEV_1, FVC, FEV_1/FVC and PEF.

Differences in approach from previous analyses

The previous approach used for quality control of spirometry data was described in [2]. This previous approach utilised the Vitalograph spirometer blow quality metrics to define acceptability only. In the present analysis, following recommendations based on work conducted for the UK Biobank Outcomes Adjudication Working Group [Strachan, personal communication], we have additionally included quality control steps based on the volume-time curves recorded (at 10ms intervals) for each spirogram. Metrics derived from these curve datasets allowed a more comprehensive and systematic assessment of: start of blow quality; end of blow quality; length of blow; and derivation of flow rates. They also permitted a comparison between FEV_1, FVC and PEF derived from the curve datasets and those pre-derived by the spirometer.

The quality control of spirometry data used in our previous publication [1] applied the ATS/ERS criteria for assessing reproducibility. These criteria, which are widely used in clinical practice, recommend that the best measures of FEV_1 and FVC are within 150ml of any other blow. However, within UK Biobank a subset of 20,347 participants were re-examined after an interval of 2-7 years, of whom 14,238 (70%) performed two or more spirograms with good start-of-blow and end-of blow quality on both occasions. Analysis of the within-subject between-occasion correlation (reliability coefficient) of FEV_1 and FVC in relation to the reproducibility of these measures at the entry examination suggested that the ATS/ERS reproducibility threshold was unduly conservative. For epidemiological studies, where spirometric comparisons are being made between groups rather than for monitoring of individual patients, a more relaxed reproducibility threshold of 250ml could be applied, increasing the available sample size without jeopardising the reliability of FEV_1 or FVC.

For illustration, among the participants with good start-of-blow and end-of-blow quality, using a reproducibility threshold of 250mL, FVC reliability was 0.9199, 0.9033, 0.8886, 0.9086 and 0.9071, respectively, for subjects with intervals of 2, 3, 4, 5 and 6-7 years between the two examinations. The corresponding figures for FEV_1 reliability were 0.9152, 0.9014, 0.8753, 0.8981 and 0.8992.

Definition of smoking status for covariate adjustment of association analyses

Smoking initiation (123,890 ever smoked vs 151,706 never smoked) was inferred using answers from questionnaire. Never smokers are those individuals who do not smoke at present and never smoked in the past [code 1239=0 & 1249=4] or do not smoke at present, smoked occasionally or just tried once or twice in the past, but had less than 100 smokes in their lifetime [1239=0 &1249=2/3 & 2644=0]. Ever smokers include current smokers (who smoke at present, on most or all days or occasionally [1239=1/2]), previous smokers (who do not smoke at present and smoked on most or all days in the past [1239=0 &1249=1] or do not smoke at present, smoked occasionally or just tried once or twice in the past, and had more than 100 smokes in their lifetime [1239=0 &1249=2/3 & 2644=1]) and individuals who smoked on most/all days or occasionally in the past, and smoked more than 100 times in their life, but prefer not to answer about current smoking [1239=-3 & 1249=1 or 1239=-3 & 1249=2 & 2644=1].

Genotyping quality control

The genotyping procedure, genotype quality control and imputation of the UK Biobank individuals is described in detail elsewhere [ref]. 968 individuals with outlying heterozygosity or missingness were already excluded from the provided imputed genotypes. We further excluded 378 individuals for whom the submitted gender did not match the genetically inferred gender, 977 samples related to >200 other samples, 188 samples with >10 3^{rd} degree relatives and 471 samples with putative sex chromosome aneuploidy, giving 2,008 excluded samples in total leaving 486,369 samples from which to select our discovery set.

Identification of individuals of European ancestry for inclusion in the genome-wide association analysis of lung function

K-means clustering was used to identify the set of European- ancestry individuals to include in the genome-wide association analysis of lung function. The steps taken to define the sets of non-European ancestry individuals to include in the analysis of heterogeneity of signals is described below.

Principal components (PCs) were provided with the UK Biobank genetic data. K-means clustering using the first two PCs was undertaken for between 3 and 8 clusters after excluding 2,008 samples failing genotyping quality control. The 6 cluster k-means model was selected as most appropriately clustering the 486,369 samples remaining after genotype quality control (QC) into broad ethnic groups giving 453,958 samples of “European ancestry” (**Supplementary Figure**). This resulted in an additional 45,865 individuals being eligible for inclusion in addition to the 408,093 passing genotype QC and defined as “white British” by UK Biobank[3].

Selecting individuals passing spirometry and genotyping quality control for genome-wide association testing

There was an overlap of 341,102 individuals (321,057 European) between 348,936 passing spirometry quality control for FEV_1, FVC, FEV_1/FVC and PEF and 486,369 passing genotyping quality control.

Removal of outlying lung function measures in European samples for discovery GWAS

Adjustment for sex, age, age^2, height, and smoking status (ever/never) of each lung function measure was undertaken in each ancestry category. 10 European individuals were excluded that were obvious outliers in plots of the adjusted phenotype distributions and the adjustment was repeated. This left 321,047 European individuals for the discovery GWAS of FEV_1, FVC, FEV_1/FVC and PEF.

Power Calculations

Power calculations were performed with the GeneticsDesign R package (https://bioconductor.org/packages/GeneticsDesign/) (**Supplementary Figure 7**) to:
A) calculate the power to detect a signal passing Tier 1 or Tier 2 criteria i.e. $P<10^{-3}$ in the SpiroMeta cohort of 79,055 samples. At this threshold, there would be 75% power to detect an effect size of 0.0325 standard deviations for a variant with MAF 10% and 95% power to detect an effect size of 0.122 standard deviations for a variant with MAF 1% in SpiroMeta.
B) calculate the power to confirm a previously reported lung function quantitative trait association in UK Biobank at $P<10^{-5}$ (n=321,047).

Overlap of samples

Each trait FEV1, FVC, FEV_1/FVC and PEF were regressed against the LD score of each variant using LDSC[4]. The proportion of total inflation due to confounding is (Intercept-1)/(Mean $\chi 2$ -1), where $\chi 2$ is the mean statistic from the association testing and the intercept is the intercept of the LD score regression (estimate of inflation due to confounding but not polygenicity). The proportion of inflation due to confounding in the meta-analysis was low (<4%) (**Supplementary Table 25**), hence we did not conclude overlap of samples between UK Biobank and SpiroMeta.

Conditional analysis with GCTA

All SNPs ±1Mb were extracted around each sentinel variant. GCTA[5] was then used to perform stepwise conditional analysis in order to select independently associated SNPs within each 2Mb region using the single SNP association statistics combined with LD information from reference genotypes representative of the samples in the association testing. For UK Biobank the same genotype data as used for the initial discovery association testing was used as an LD reference; for SpiroMeta, genotypes from 48,943 unrelated participants[6] formed the LD reference set

Smoking behaviour association analyses in UKB.

Association analyses with smoking behaviour phenotypes were performed in the 335,641 UKB individuals out of the full 488,377 included in the final release of genetic data that were not in the 152,736 in the interim release (http://www.ukbiobank.ac.uk/scientists-3/genetic-data/), as part of an independent replication for the GSCAN study that included samples from the UK Biobank interim release.

Genotyping quality control was performed using the same criteria as for the lung function analysis (individuals excluded on the basis of sex mismatches, heterozygosity and missingness). Only individuals of European ancestry were included in the association analyses. These were identified by first calculating the minimum and maximum value of the first 4 PCs of the samples defined as white British in UK Biobank [ref to QC paper] and then we included any individual in this PC range regardless of their self-reported ancestry. Individuals who were related to UK Biobank individuals included in previous releases with a kinship coefficient > 0.075 were excluded from the analyses. Only variants imputed on the HRC panel and with MAC >= 3 were included in the analyses.

Smoking initiation (123,890 ever smoked vs 151,706 never smoked) was inferred using answers from questionnaire as for the smoking covariate adjustment above.

The average number of cigarettes smoked per day (CPD) for all individuals who smoke, or smoked, on most or all days was binned as follows: 1 = 1-5, 2 = 6-15, 3 = 16-25, 4 = 26-35, 5 = 36+. Cigarettes per day was available for 80,015 samples.

All phenotypes used age, age squared, sex, and genetic principal components 1-15 as covariates. Residuals were calculated for each phenotype by linear regression, with the phenotype as the dependent variable and the corresponding covariates as the independent variables. These residuals were then inverse normalized, and the corresponding z-scores were used as the input phenotype values for the association analysis.

BOLT-LMM version 2.3 was used to conduct association analysis on each chromosome. The variants included in the mixed model were extracted from the genotyped variants by applying the following filters: missingness < 5%, minor allele frequency > 1%, HWE p > 10^{-6}, pruning for LD r2 < 0.2. The hg19 reference map was used to interpolate genetic map coordinates. BOLT-LMM statistic was calibrated using the 1000 genomes LD scores reference table.

Smoking interaction testing

Association testing for lung function was calculated separately in ever and never smoker subgroups and meta-analysed across UK Biobank and SpiroMeta for up to 176,701 ever smokers and 197,999 never smokers. The Welch test was used to compare genetic effect between ever and never smokers:

$$t = \frac{\beta_1 - \beta_2}{\sqrt{{se_1}^2 + {se_2}^2}}$$

with degrees of freedom:

$$d.f. = \frac{({se_1}^2 + {se_2}^2)^2}{\frac{{se_1}^2}{n_1 - 1} + \frac{{se_2}^2}{n_2 - 1}}$$

A deviation from equality (P<$1.8\text{x}10^{-4}$, i.e. 0.05/279 tests) was considered significant evidence of interaction. For these analyses, phenotypes were inverse normalised after regressing on sex, age, age^2, and height. Genotyping array was included as a covariate.

Using the *European only* sample as input for relatedness exclusion here resulted in a marginally bigger sample size than that produced when including all ancestries (N = 303,619 cf. N = 303,570).

Population Attributable Risk Calculation

We calculated the population attributable risk fraction (PARF) as follows:

$$PARF = \frac{P(E)(OR - 1)}{1 + P(E)(OR - 1)}$$

where *P(E)* is set to 0.9, i.e. the probability of possessing more risk alleles than those in the lowest decile of the risk score (the 'probability of the exposure'). *OR* above refers to the odds of having COPD in individuals across deciles 2 to 10 of the risk score compared to the odds of having COPD for individuals in the lowest decile (decile 1) of the risk score.

Before calculating the PARF, we used the European meta-analysis OR of 1.546 (95CI: 1.476-1.620) per SD of the genetic risk score (GRS) to estimate the *OR* for COPD, comparing individuals in deciles 2-10 vs those in decile 1. We assume that the GRS is normally distributed so that log(1.546) is the additive effect on a standard normal variable.

The expected GRS, given that an individual is in decile j of the GRS, is

$$\frac{1}{0.1}\int_{\Phi^{-1}\left(\frac{j-1}{10}\right)}^{\Phi^{-1}\left(\frac{j}{10}\right)} x\phi(x)dx$$

The limits of the integral are the lower and upper values of the GRS for individuals in decile j, assuming the GRS is standard normal. The division by 0.1 ensures the expectation is conditional on the individual being in the decile, which is 1/10 by definition.

Then the expected log OR for decile j is

$$\frac{\log(1.546)}{0.1}\int_{\Phi^{-1}\left(\frac{j-1}{10}\right)}^{\Phi^{-1}\left(\frac{j}{10}\right)} x\phi(x)dx$$

and comparing with decile 1 gives

$$\frac{\log(1.546)}{0.1}\left[\int_{\Phi^{-1}\left(\frac{j-1}{10}\right)}^{\Phi^{-1}\left(\frac{j}{10}\right)} x\phi(x)dx - \int_{-\infty}^{\Phi^{-1}\left(\frac{1}{10}\right)} x\phi(x)dx\right]$$

We can now proceed to estimate the log OR for deciles 2-10 vs decile 1 as

$$\log(1.546)\left[\frac{1}{0.9}\int_{\Phi^{-1}\left(\frac{1}{10}\right)}^{\infty} x\phi(x)dx - \frac{1}{0.1}\int_{-\infty}^{\Phi^{-1}\left(\frac{1}{10}\right)} x\phi(x)dx\right] = \log(2.339)$$

The estimated bounds of the 95% confidence interval around this new estimate are then calculated using the same method, and entered into the PARF equation, above.

SpiroMeta consortium study details

This section provides study descriptions for the cohorts contributing to the SpiroMeta consortium. All participants provided written informed consent and studies were approved by local Research Ethics Committees and/or Institutional Review boards.

Details of the **British 1958 Birth Cohort** biomedical follow-up have been previously reported[7]. Spirometry at age 44–45 years was done in the standing position without nose clips, using a Vitalograph handheld spirometer as previously described[8]. In the analysis, all readings with a best-test variation greater than 10% were excluded.

The **Busselton Health Study** (BHS) is a longitudinal survey of the town of Busselton in the south-western region of Western Australia that began in 1966. In 1994/1995 a cross-sectional community follow-up study was undertaken where blood was taken for DNA extraction. A sample of 1,168 European-ancestry individuals were genotyped using the Illumina 610-Quad BeadChip (BHS1), and subsequent genotyping was carried out on an independent group of 3,428 European-ancestry individuals using Illumina 660W-Quad (BHS2). Spirometric measures of forced expired volume in one second (FEV_1) and forced vital capacity (FVC) were assessed.

The CROATIA study was initiated to investigate the use of isolated rather than urban populations for the identification of genes associated with medically-relevant quantitative traits. Three cohorts have been recruited as part of the CROATIA study: **CROATIA-Vis**[9], **CROATIA-Korcula**[10] and **CROATIA-Split**[11]. CROATIA-Vis was the first to be collected when 1,008 Croatians aged 18-93 recruited from the villages of Komiza and Vis on the Dalmatian island of Vis. Recruitment occurred from 2003 to 2004 with participants donating blood for DNA extraction and biochemical measurements as well as undergoing some anthropometric measurements and physiological tests to measure traits such as height, weight and blood pressure, and finally completing several questionnaires relating to general health, medical history, diet and lifestyle. CROATIA-Korcula was recruited from 2007 to 2008 from the town of Korcula and the villages of Lumbarda, Zrnovo and Racisce on the island of Korcula, Croatia with 969 adults aged 18-98 agreeing to participate. This study followed the same recruitment procedures as CROATIA-Vis and the same samples and tests were collected with a few additions to reflect the research interests and expertise in Edinburgh. Volunteers were recruited to be part of the CROATIA-Split cohort in 2009-2010 from the Dalmatian mainland city of Split. This is the main ferry port to the islands and is the second largest city in Croatia and the largest along the Dalmatian coast. 1,012 adults aged 18-85 were recruited using the same methodology and with the same samples collected as in CROATIA-Korcula. Ethical approval was obtained from appropriate regulatory bodies in both Scotland and Croatia and participants gave informed consent prior to joining the study.

European Prospective Investigation of Cancer (EPIC)-Norfolk is an ongoing UK-based prospective cohort and part of the Europe-wide multi-centre EPIC study. Details of the study design were described previously.[12] Briefly, 25,639 men and women aged 40û79 in eastern England were recruited through general practice registers and underwent baseline assessment between 1993 and 1997. Participants were further invited to the follow-up assessment (1998 to 2000), and were followed up by 2009 for incident outcomes and by 2013 for mortality.

The **Generation Scotland: Scottish Family Health Study** is a collaboration between the Scottish Universities and the NHS, funded by the Chief Scientist Office of the Scottish Government. GS:SFHS is a family-based genetic epidemiology cohort with DNA, other biological samples (serum, urine and cryopreserved whole blood) and socio-demographic and clinical data from ~24,000 volunteers, aged 18-98 years, in ~7,000 family groups. Participants were recruited across Scotland, with some family members from further afield, from 2006-2011. Most (87%) participants were born in Scotland and 96% in the UK or Ireland. The cohort profile has been published[13]. GS:SFHS operates under appropriate ethical approvals, and all participants gave written informed consent. Generation Scotland is a collaboration between the University Medical Schools and National Health Service in Aberdeen, Dundee, Edinburgh and Glasgow (UK).

The DNA archive established from the **Health 2000** Survey Cohort was used. Details of this study population and phenotyping procedures have been previously reported[14]. Genome-wide genotyping was available for 2124 individuals selected from the Health 2000 cohort as metabolic syndrome cases and their matched controls[15]. Spirometry was done in the standing position without nose clips, using a Vitalograph 2150 spirometer. In the analysis, the maximum permissible difference between the two highest FEV_1 and FVC values was 10%.

The KORA studies (Cooperative Health Research in the Region of Augsburg) are a series of independent population based studies from the general population living in the region of Augsburg, Southern Germany[16,17]. **KORA F4** including 3,080 individuals was conducted from 2006-2008 as a follow-up study to KORA S4 (1999-2001). Lung function tests were performed in a random subsample of subjects born between 1946 and 1965 (age range 41–63 years). Spirometry was performed in line with the ATS/ERS recommendations[1] using a pneumotachograph-type spirometer (Masterscreen PC, CardinalHealth, Würzburg, Germany) before and after inhalation of 200µg salbutamol. The present study is based on maximum values of FEV_1 and FVC measured before bronchodilation. The spirometer was calibrated daily using a calibration pump (CardinalHealth, Würzburg, Germany), and additionally, an internal control was used to ensure constant instrumental conditions. For KORA F4 participants without spirometry measurements in 2006-2008, we used measurements from the KORA-Age time point conducted in 2008/09. KORA Age contains subjects from all KORA studies born until 1943 (aged 65-90 years)[18]. Spirometry was measured in 935 randomly selected participants. Conditions including the examiner were the same as in 2008/09 except that inhalation of salbutamol was not performed due to the high number of contraindications anticipated in this aged population.

The KORA studies (Cooperative Health Research in the Region of Augsburg) are a series of independent population based studies from the general population living in the region of Augsburg, Southern Germany[16,17]. The **KORA S3** study including 4,856 individuals was conducted in 1994/95. Spirometry was measured during a follow up in 1997/98 for all participants younger than 60 years who did not smoke or use inhalers one hour before the test. All spirometric tests were performed strictly adhering to the ECRHS protocol[19,20] using Biomedin Spirometers (Biomedin srl, Padova, Italy). Tests were accounted valid if at least two technically satisfactory manoeuvres could be obtained throughout a maximum of nine trials. FEV_1 and FVC were defined as the maximum value within all valid manoeuvers. For KORA S3 participants without spirometry measurements in 1997/98 we used measurements from the KORA-Age time point conducted in 2008/09. KORA Age contains subjects from all KORA studies born until 1943 (aged 65–90 years) [18]. Spirometry was measured in 935 randomly selected participants. Conditions including the examiner were the same as in KORA F4 (see below) except that inhalation of salbutamol was not performed due to the high number of contraindications anticipated in this aged population.

The **Lothian Birth Cohort 1936** consists of 1,091 relatively healthy individuals assessed on cognitive and medical traits at about 70 years of age. They were all born in 1936 and most took part in the Scottish Mental Survey of 1947. At baseline the sample of 548 men and 543 women had a mean age 69.6 years (s.d. = 0.8). They were all Caucasian, community-dwelling, and almost all lived in the Lothian region (Edinburgh city and surrounding area) of Scotland. A full description of participant recruitment and testing can be found elsewhere[21]. Genotyping was performed at the Wellcome Trust Clinical Research Facility, Edinburgh. Quality control measures were applied and 1,005 participants remained. Lung function assessing peak expiratory flow rate, forced expiratory volume in 1 second, and forced vital capacity (each the best of three), using a Micro Medical Spirometer was assessed, sitting down without nose clips, at age 70 years. The accuracy of the spirometer is ±3% (to ATS recommendations Standardisation of Spirometry 1994 update for flows and volumes).

The **Northern Finland Birth Cohort 1966 (NFBC1966)** is a prospective follow-up study of children from the two northernmost provinces of Finland born in 1966.[22]All individuals still living in northern Finland or the Helsinki area (n = 8,463) were contacted and invited for clinical examination. A total of 6007 participants attended the clinical examination at the participants' age of 31 years. DNA was extracted from blood samples given at the clinical examination (5,753 samples available).[23]The subset with DNA is representative of the original cohort in terms of major environmental and social factors. Informed consent was obtained from all subjects. After performing standard sample QC we included 5,402 NFBC1966 participants that were genotyped on an Illumina HumanCNV370DUO Analysis BeadChip. 329,401 variants were included in the imputation scaffold. Variants were imputed to the HRC reference r1.1 2016 on the Michigan Imputation Server. Prior to analysis we excluded variants monomorphic in this dataset. In NFBC1966, we used a Vitalograph P-model spirometer (Vitalograph Ltd., Buckingham, UK), with a volumetric accuracy of ±2% or ±50 mL whichever was greater. The spirometer was calibrated regularly using a 1-Litre precision syringe. The spirometric manoeuvre was performed three times but was repeated if the coefficient of variation between two maximal readings was >4%.

The **Northern Finland Birth Cohort 1986 (NFBC1986)** consists of 99% of all children, who were born in the provinces of Oulu and Lapland in Northern Finland between 1 July 1985 and 30 June 1986. 9,203 live-born individuals entered the study.[24] At the age of 16, the subjects living in the original target area or in the capital area (n=9,215) were invited to participate in a follow-up study including a clinical examination. 7,344 participants attend the study in year 2001/2002, of which 5,654 completed the postal questionnaire, the clinical examination and provided a blood sample.[25] DNA was extracted from all 5,654 blood samples. An informed consent for the use of the data including DNA was obtained from all subjects. After performing standard sample QC we included 3,743 NFBC1986 participants that were genotyped on an Illumina Human Omni Express Exome 8v1.2 BeadChip. 889,119 variants were included in the imputation scaffold. Variants were imputed to the HRC reference r1.1 2016 on the Michigan Imputation Server. For Spirometry measurements, we used a Vitalograph Gold Standard (Model 2150) (Vitalograph Ltd., Buckingham, UK). The machines were calibrated every day the medical examination took place. The spirometric manoeuvre was performed in an upright sitting position while wearing a nose clip. At least three acceptable manoeuvres were performed. Acceptable manoeuvers did not exceed a difference between two maximal FEV 1 and FVC values of 4 %. The results were recorded with a 0.05 litre accuracy.

The **Northern Sweden Population Health Study** (NSPHS) represents a cross-sectional study conducted in the communities of Karesuando (samples gathered in 2006) and Soppero (2009) in the subarctic region of the County of Norrbotten, Sweden. Spirometry was performed in sitting position without noseclips using a MicroMedicalSpida 5 spirometer (http://www. medisave.co.uk). Three consecutive 28 lung function measurements per participant were done and the maximum value per measured lung function parameter was used for further analysis. Relatedness was taken into account by applying the "polygenic" linear mixed effects model. Genome-wide association analysis was performed using a score test, a family-based association test[26] which uses the residuals and the variance-covariance matrix from the polygenic model and the SNP fixed effect coded under an additive model.

The **Orkney Complex Disease Study** (ORCADES) is an ongoing family-based, cross-sectional study in the isolated Scottish archipelago of Orkney. Spirometry was performed in the sitting position without nose clips, using a Spida handheld spirometer. Measurements were repeated once and the better reading was used for analysis.

The **Prospective Investigation of the Vasculature in Uppsala Seniors** (PIVUS)[27] is a population-based study of cardiovascular health in the elderly. Mailed invitations were sent to subjects who lived in Uppsala, Sweden, within 2 months after their 70th birthday. The subjects were randomly selected from the community register. A total of 1,016 men and women participated in the baseline investigation (participation rate, 50.1%). Spirometry was performed in 901 subjects at baseline in accordance with American Thoracic Society recommendations (α spirometer; Vitalograph Ltd; Buckingham, UK). The best value from three recordings was used. The Ethics Committee of the University of Uppsala approved the study, and the participants gave their informed consent. Genotyping of all samples was undertaken using the Illumina Omni Express and CardioMetabochip. Genotypes were called using GENCALL. A total of 738,879 SNPs passed quality control (thresholds: call rate < 0.95, and call rate < 0.99 for MAF<5%; HWE *P* < 10-6). SNPs with MAF<1% were removed from the imputation scaffold. Imputation was performed using IMPUTE up to haplotypes from the Haplotype Reference Consortium.

The **SAPALDIA** cohort is a population-based multi-center study in eight geographic areas representing the range of environmental, meteorological and socio-demographic conditions in Switzerland[28,29]. It was initiated in 1991 (SAPALDIA 1) with a follow-up assessment in 2002 (SAPALDIA 2) and 2010 (SAPALDIA3). This study has specifically been designed to investigate longitudinally lung function, respiratory and cardiovascular health; to study and identify the associations of these health indicators with individual long term exposure to air pollution, other toxic inhalants, life style and molecular factors.

The **Study of Health In Pomerania (SHIP)[30]** is a cross-sectional and prospective longitudinal population-based cohort study in Western Pomerania assessing the prevalence and incidence of common diseases and their risk factors. SHIP encompasses the two independent cohorts **SHIP** and **SHIP-TREND**. A total of 4,308 participants were recruited between 1997 and 2001 in the SHIP cohort. Between 2008 and 2012 a total of 4,420 participants were recruited in the SHIP-TREND cohort. Individuals were invited to the SHIP study centre for a computer-assisted personal interviews and extensive physical examinations.

The examinations for **SHIP** were conducted using a body plethysmograph equipped with a pneumotachograph (VIASYS Healthcare, JAEGER, Hoechberg, Germany) which meets the American Thoracic Society (ATS) criteria.[31] The volume signal of the equipment was calibrated with a 3.0 litre syringe connected to the pneumotachograph in accordance with the manufacturer´s recommendations and at least once on each day´s testing. Barometric pressure, temperature and relative humidity were registered every morning. Calibration of reference gas and volume was examined under ATS-conditions (Ambient Temperature Pressure) and the integrated volumes were BTPS (Body Temperature Pressure Saturated) corrected.[31,32] Lung function variables were measured continuously throughout the baseline breathing and the forced manoeuvres using a VIASYS HEALTHCARE system (MasterScreen Body/Diff.). Spirometry flow volume loops were conducted in accordance with ATS recommendations[32] in a sitting position and with wearing nose clips. The participants performed at least three forced expiratory lung function manoeuvres in order to obtain a minimum of two acceptable and reproducible values.[33] Immediate on-screen error codes indicating the major acceptability (including start, duration and end of test) and reproducibility criteria supported the attempt for standardised procedures. The procedure was continuously monitored by a physician. The best results for FVC, FEV1, peak expiratory flow (PEF) and expiratory flow at 75%, 50%, 25% of FVC (MEF 75, MEF 50, MEF 25) were taken. The ratio of FEV1 to FVC was calculated from the largest FEV1 and FVC.

In terms of the pulmonary items the computer-assisted interview in **SHIP-TREND** was nearly identical to that of the SHIP. Of the 4.420 subjects who have been investigated in the study, 2.678 (60.6 %) of the subjects have undergone spirometry, body plethysmography, and measurements of diffusing capacity (CO and NO), IOS and respiratory muscle strength. In SHIP-TREND, the following additional methods that are of particular interest in terms of lung health and comorbidities have been applied: polysomnography, analysis of volatile compounds in the exhaled breath, and whole-body MRI. The following devices have been used for the pulmonary investigations in SHIP-TREND: a MasterScreen for body plethysmography, diffusing capacity measurements (single breath) and measurements of respiratory muscle strength (Viasys Healthcare, Hoechberg, Germany), an ABL 500 and later an ABL 80 for blood gas analyses (Radiometer, Copenhagen, Denmark), a MasterScreen PFT Pro CO-NO-Diffusion (CareFusion, Hoechberg, Germany), a MasterSreen IOS for Impuls-Oscillometry (CareFusion, Hoechberg, Germany), and a MicroCO carbon monoxide monitor (CareFusion, Hoechberg, Germany).

The **United Kingdom Household Longitudinal Study (UKHLS)**, 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.
For a subset of individuals who took part in a nurse health assessment, blood samples were taken and genomic DNA extracted. Of these, 10,484 samples were genotyped at the Wellcome Trust Sanger Institute using the Illumina Infinium HumanCoreExome-12 v1.0BeadChip.
Lung function measures in samples from England and Wales were conducted with the NDD Easy On-PC spirometer (NDD Medical Technologies, Zurich, Switzerland). Participants were excluded in the following cases: pregnancy, having had abdominal or chest surgery (past 3 weeks), admitted to the hospital with a heart complaint (in the past 6 weeks), having had recent eye surgery (past 4 weeks), or in case of having a tracheostomy. Subjects were asked to perform up to 8 blows that ideally lasted at least 6 seconds, uninterrupted by coughing, glottis closure, laughing or leakage of air. Upon completion, the measurements were rated either acceptable or unacceptable by the NDD Easy On-PC software.

The Viking Health Study - Shetland (**VIKING**) is a family-based, cross-sectional study that seeks to identify genetic factors influencing cardiovascular and other disease risk in the population isolate of the Shetland Isles in northern Scotland. Genetic diversity in this population is decreased compared to Mainland Scotland, consistent with the high levels of endogamy historically. Participants were recruited between 2013 and 2015, each having at least three grandparents from Shetland. Fasting blood samples were collected and over 300 health-related phenotypes and environmental exposures were measured in each individual. All participants gave informed consent and the study was approved by the South East Scotland Research Ethics Committee.

The **Young Finns Study** (YFS) is a population-based follow up-study started in 1980[34]. The main aim of the YFS is to determine the contribution made by childhood lifestyle, biological and psychological measures to the risk of cardiovascular diseases in adulthood. In 1980, over 3,500 children and adolescents all around Finland participated in the baseline study. The follow-up studies have been conducted mainly with 3-year intervals. The latest 30-year follow-up study was conducted in 2010-2011 (ages 33-49 years) with 2,063 participants. The study was approved by the local ethics committees (University Hospitals of Helsinki, Turku, Tampere, Kuopio and Oulu) and was conducted following the guidelines of the Declaration of Helsinki. All participants gave their written informed consent.

Cohort contributors

UNDERSTANDING SOCIETY SCIENTIFIC GROUP

The UK Household Longitudinal Study: 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

Grant Support and Disclaimer

The project described was supported by Award Number U01 HL089897 and Award Number U01 HL089856 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.

COPD Foundation Funding

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.

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; 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; Lystra P. Hayden, MD, MMSc; 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; Edwin K. Silverman, MD, PhD; Emily S. Wan, MD; Sungho Won, Ph.D.; Phuwanat Sakornsakolpat, M.D.; Dmitry Prokopenko, Ph.D.

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

Mortality Adjudication Core: Surya Bhatt, MD; Jessica Bon, MD; MeiLan K. Han, MD, MS; Barry Make, MD; Carlos Martinez, MD, MS; Susan Murray, ScD; Elizabeth Regan, MD; Xavier Soler, MD; Carla G. Wilson, MS

Biomarker Core: Russell P. Bowler, MD, PhD; Katerina Kechris, PhD; Farnoush Banaei-Kashani, Ph.D

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; Karin F. Hoth, PhD; 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

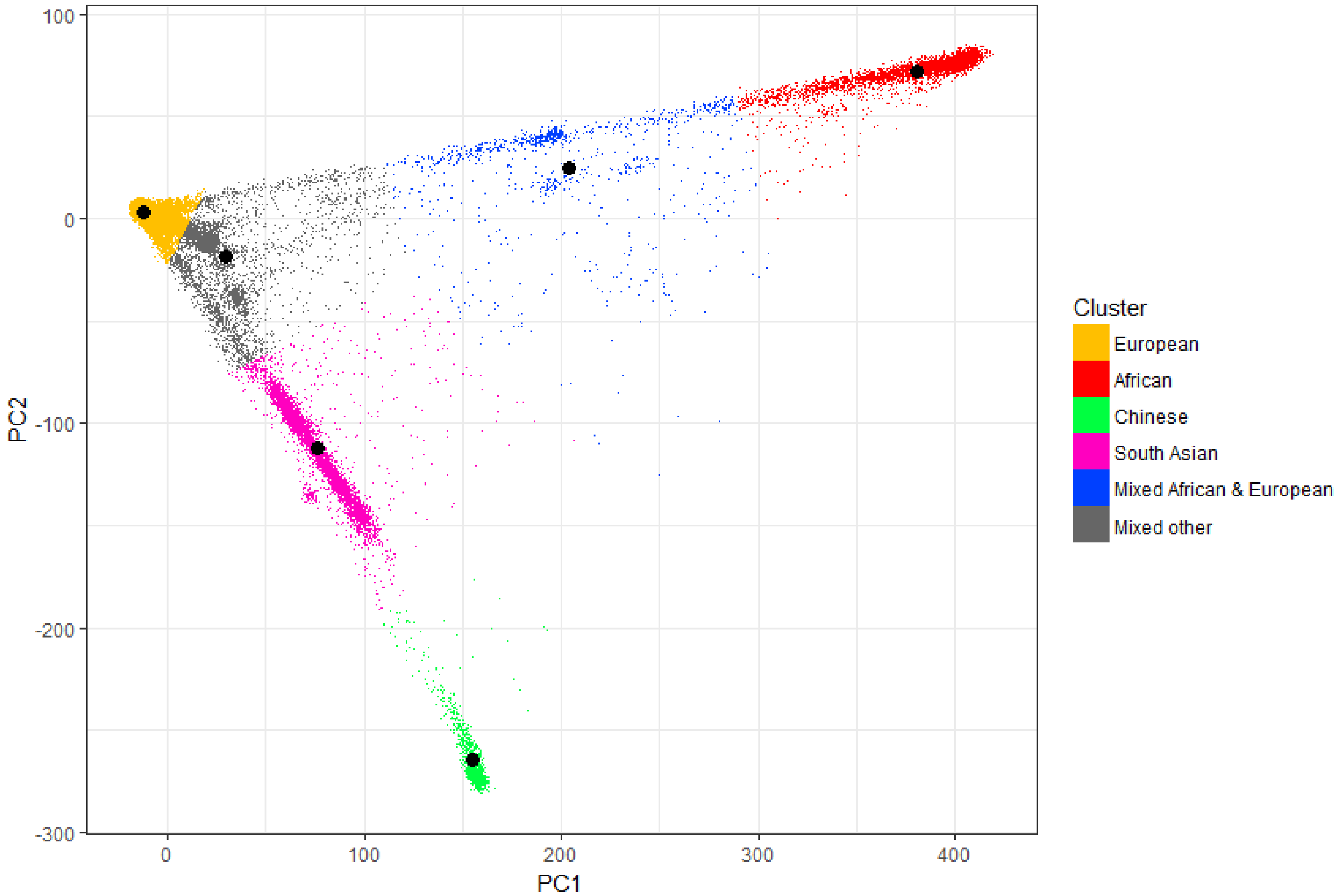
University of Pittsburgh, Pittsburgh, PA: Frank Sciurba, MD; Jessica Bon, MD; Divay Chandra, MD, MSc; Carl Fuhrman, MD; Joel Weissfeld, MD, MPH

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

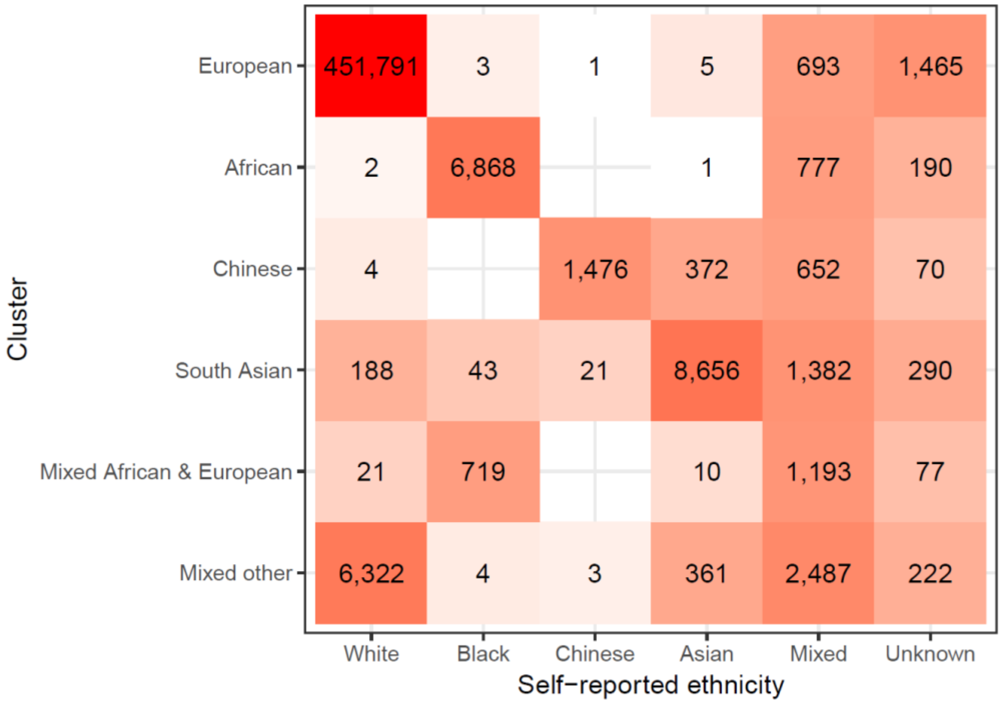
Supplementary Figures

Supplementary Figure 1: 6 ethnic grouping clusters chosen by K-means clustering

A) K-means clustering was performed on the first 2 principal components. 6 clusters were chosen to infer ancestry groupings. The black dots show the cluster centres.

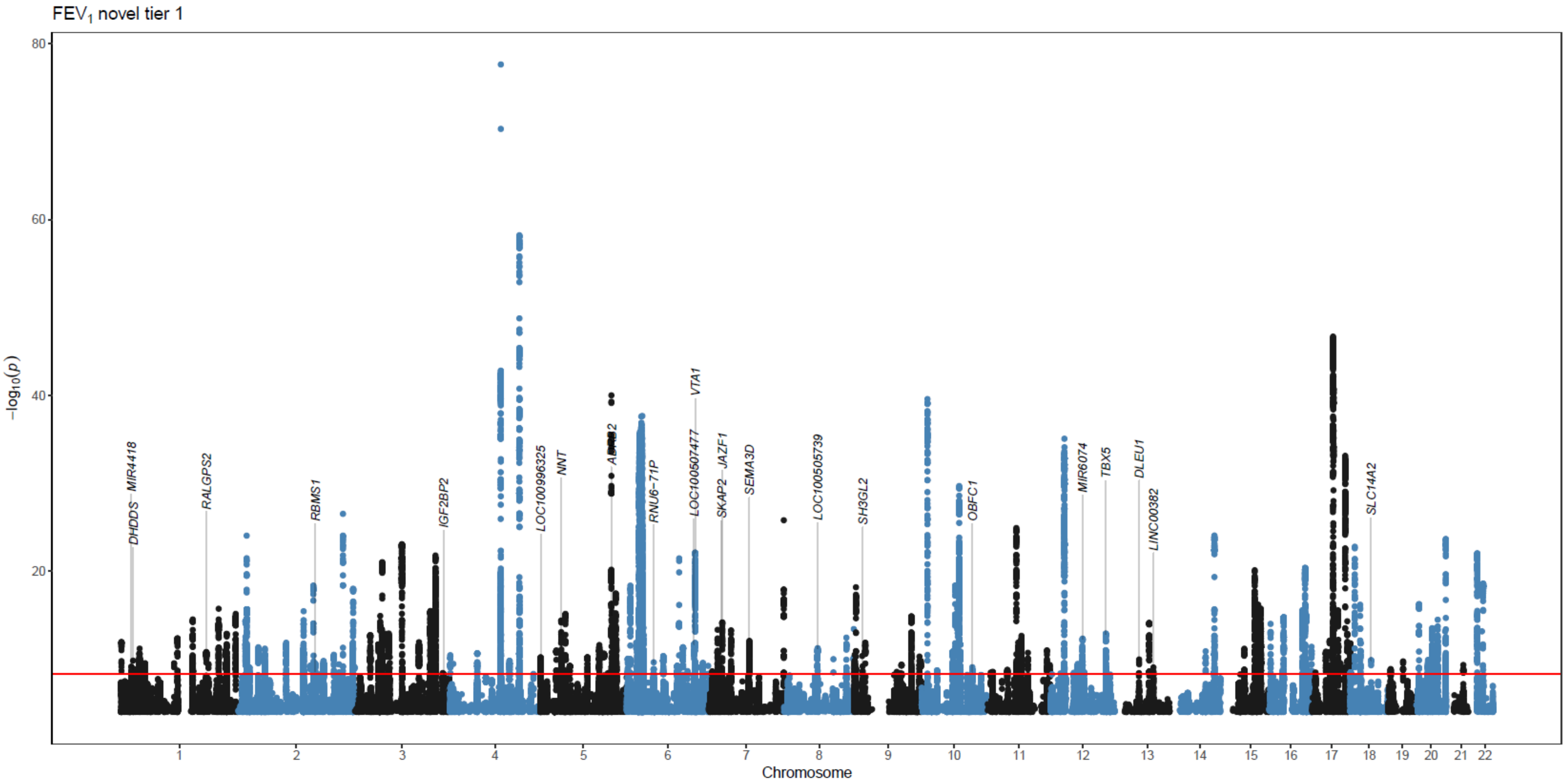


B) Correlation between ancestry groups derived from K-means clustering and self-reported ethnicity; the numbers of samples in each K-means cluster (y-axis) with each self-reported ethnicity (x-axis) are shown.

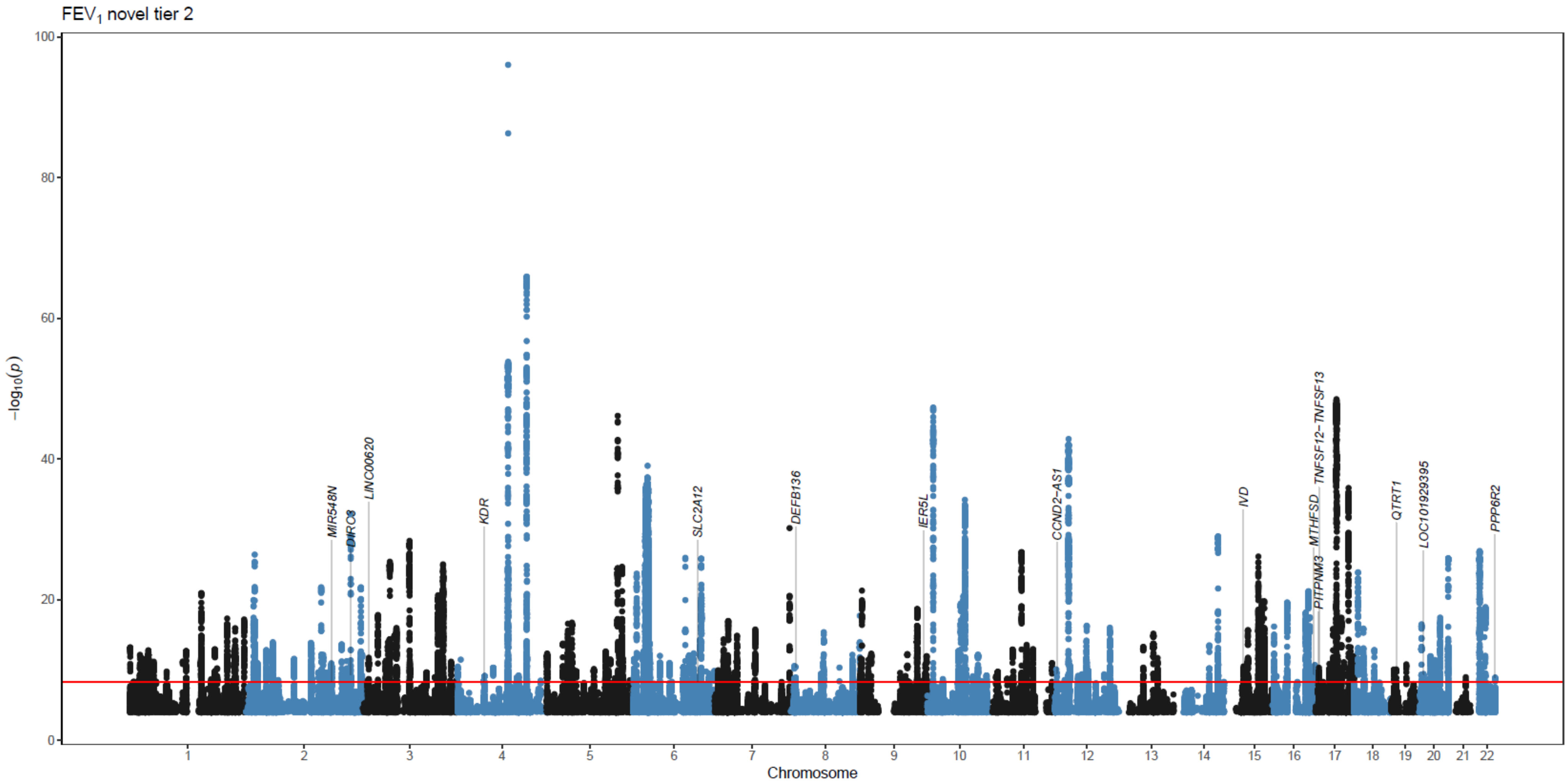


Supplementary Figure 2: Manhattan plots

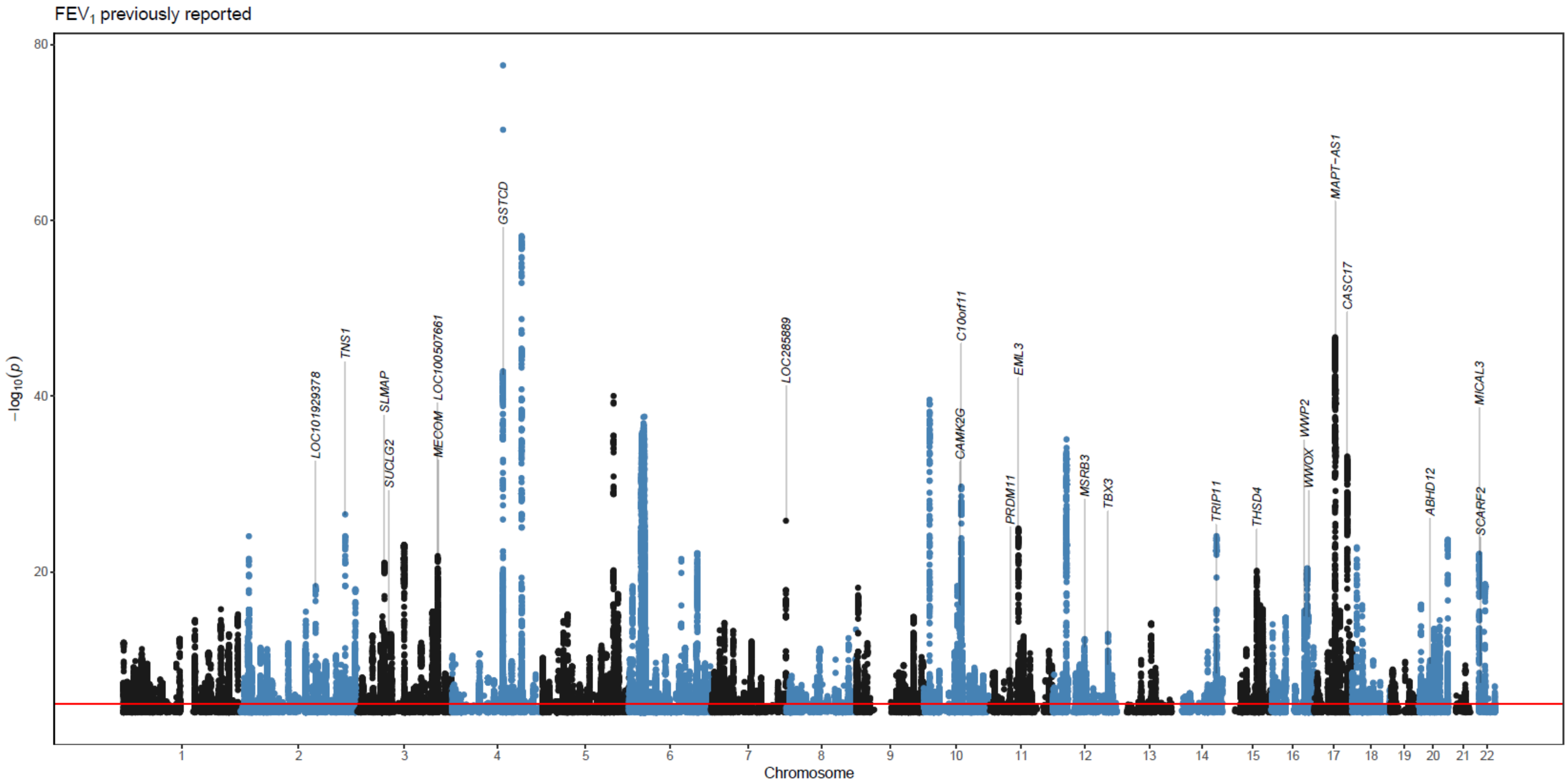
A) FEV_1 novel tier 1 signals. P values from UK Biobank with $P=5\times10^{-9}$ (Tier 1 UK Biobank threshold) shown in red.



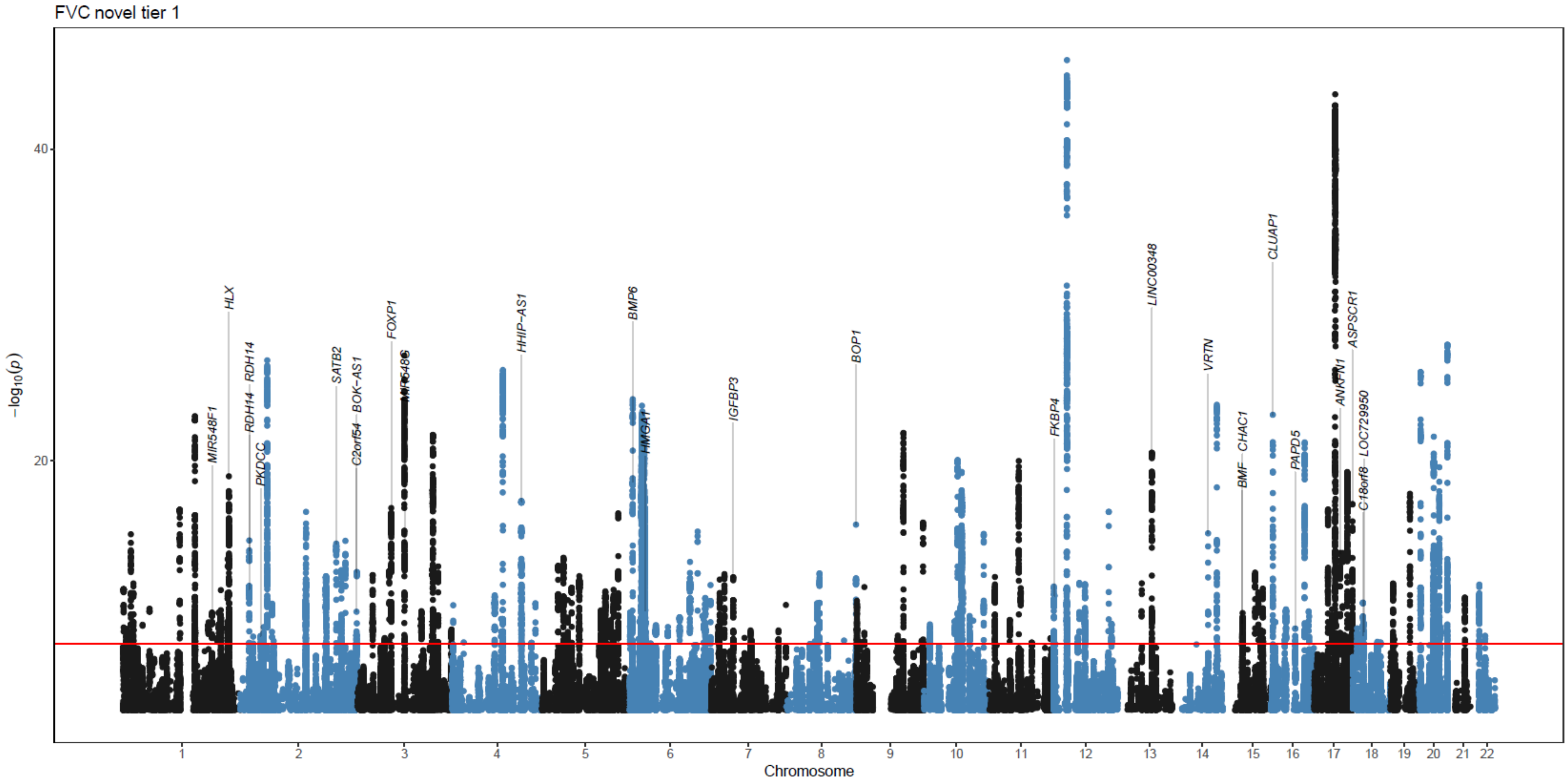
B) FEV_1 novel tier 2 signals. P values from meta-analysis of UK Biobank and SpiroMeta with $P=5\times10^{-9}$ (Tier 2 meta-analysis threshold) shown in red.



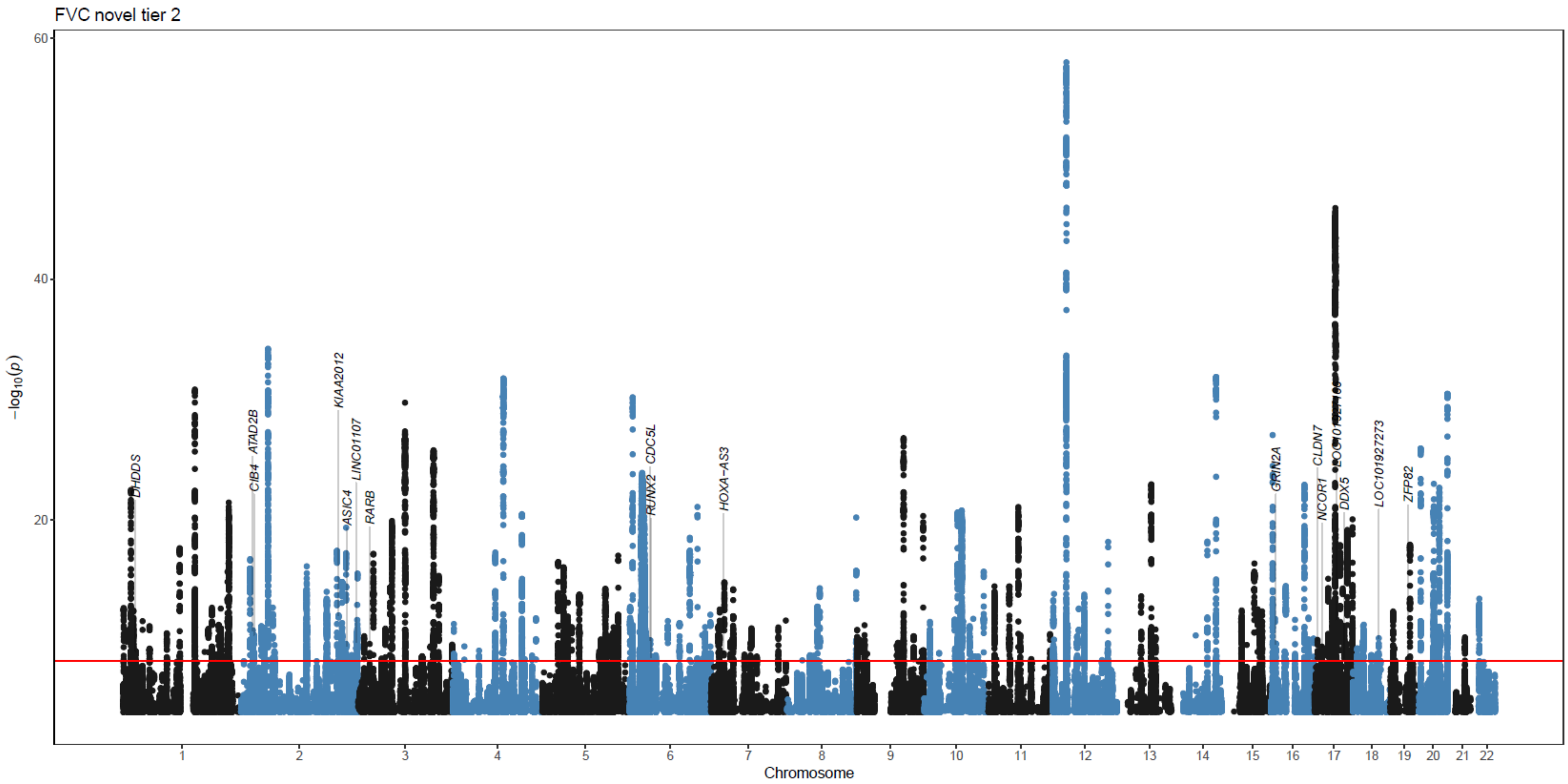
C) FEV_1 previously reported signals. P values from UK Biobank with $P=10^{-5}$ (threshold for inclusion of previously reported signals in downstream analyses) shown in red.



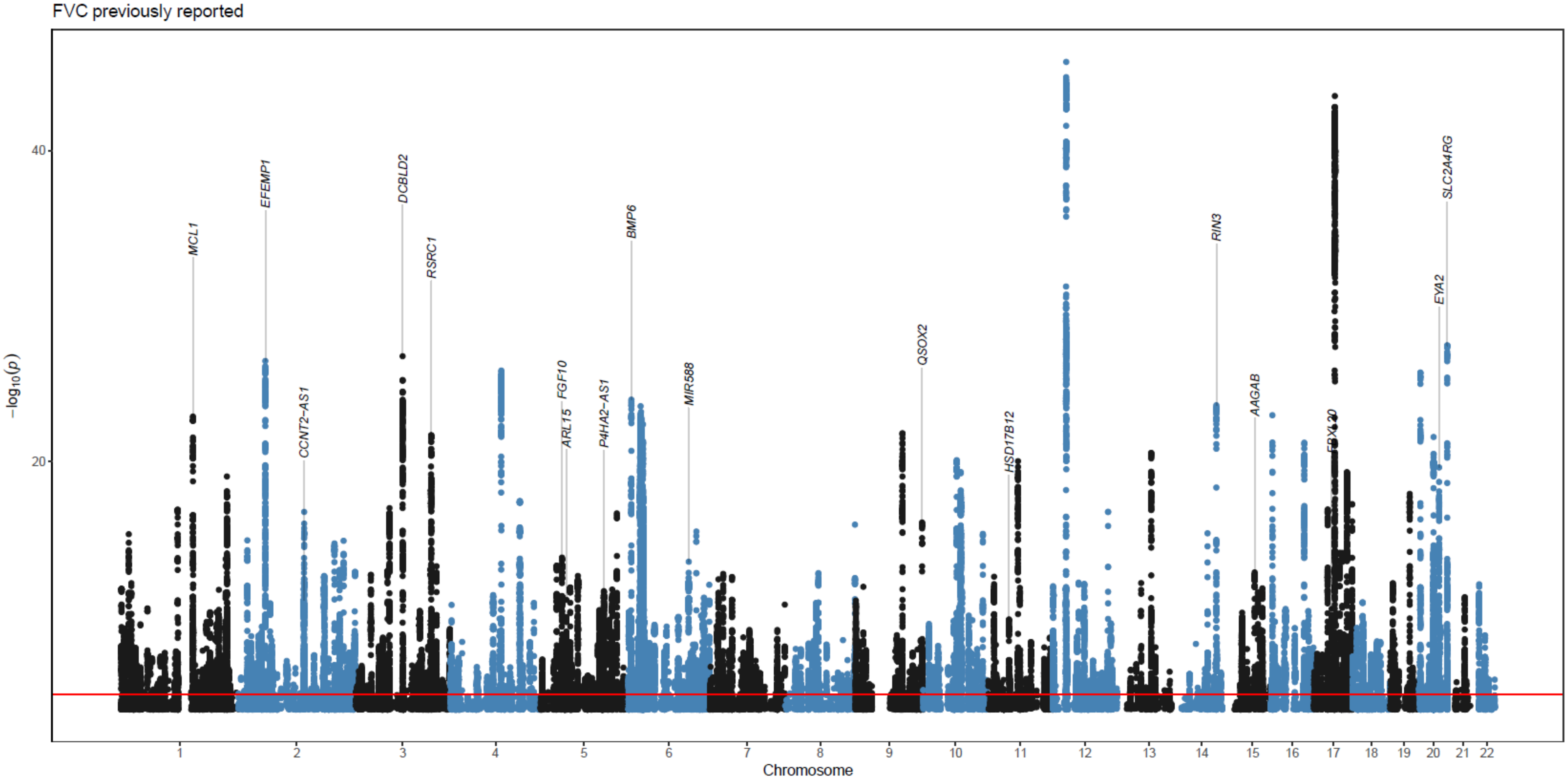
D) FVC novel tier 1 signals. P values from UK Biobank with $P=5\times10^{-9}$ (Tier 1 UK Biobank threshold) shown in red.



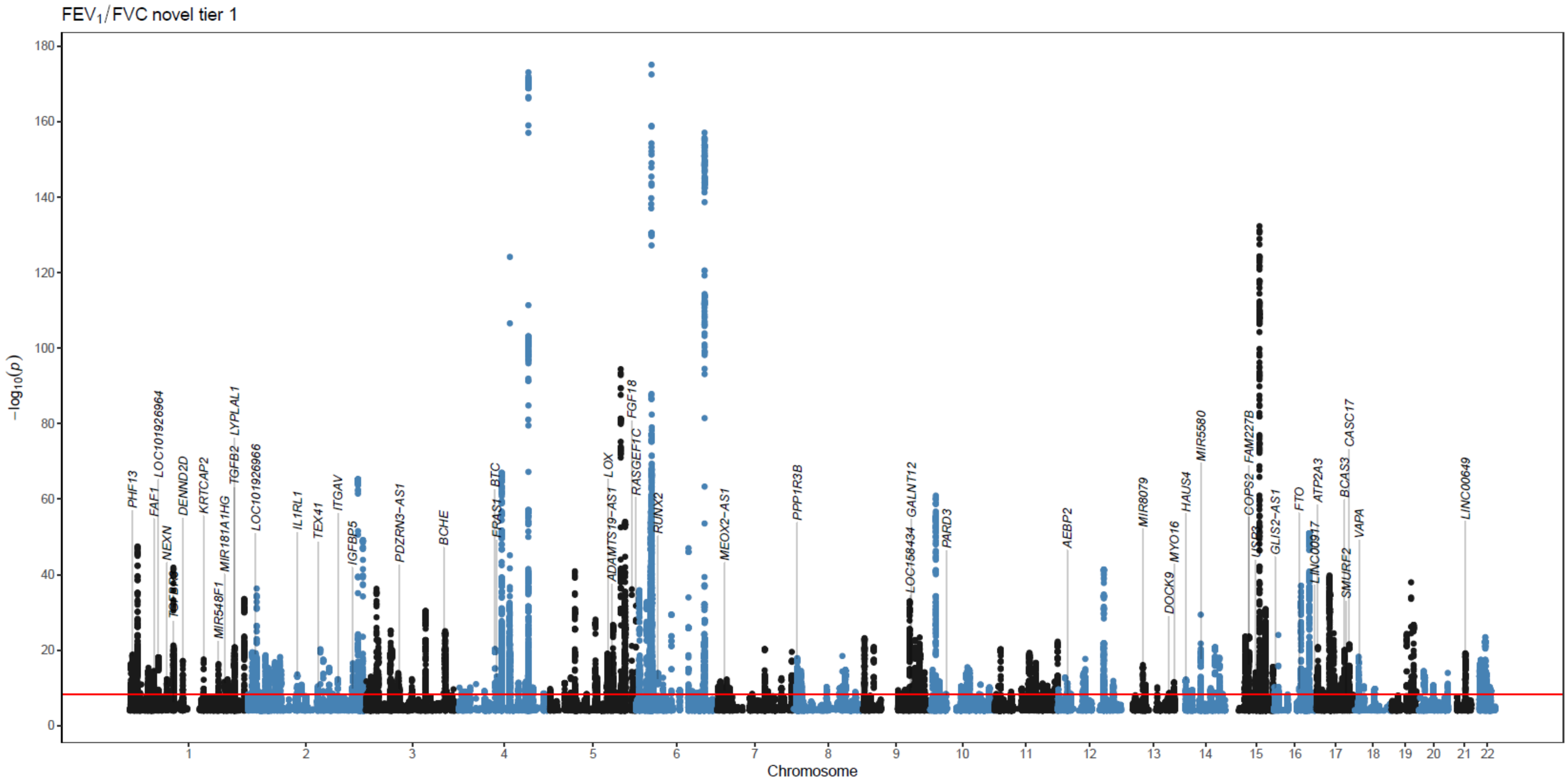
E) FVC novel tier 2 signals. P values from meta-analysis of UK Biobank and SpiroMeta with P=5×10^{-9} (Tier 2 meta-analysis threshold) shown in red.



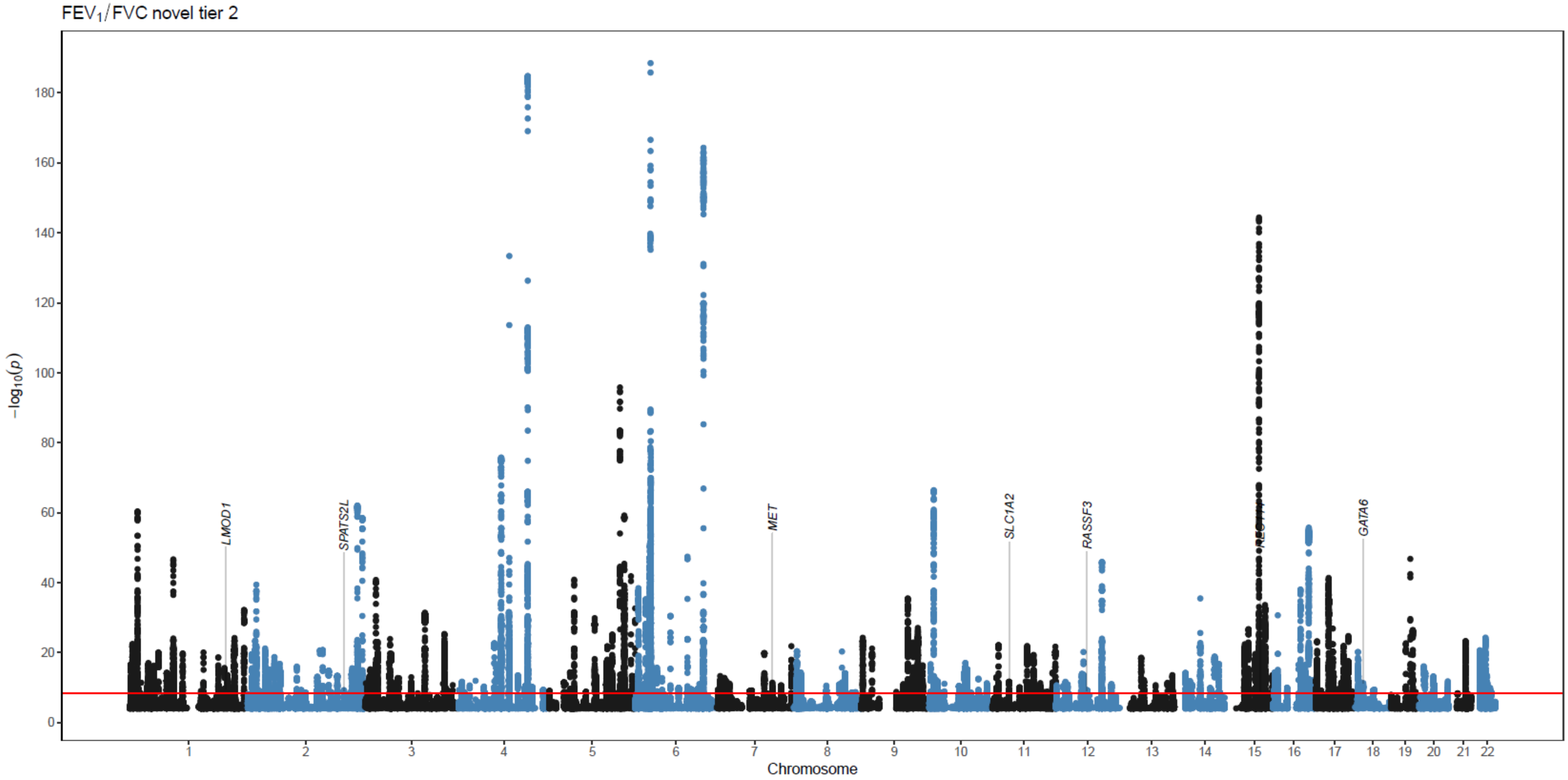
F) FVC previously reported signals. P values from UK Biobank with $P=10^{-5}$ (threshold for inclusion of previously reported signals in downstream analyses) shown in red.



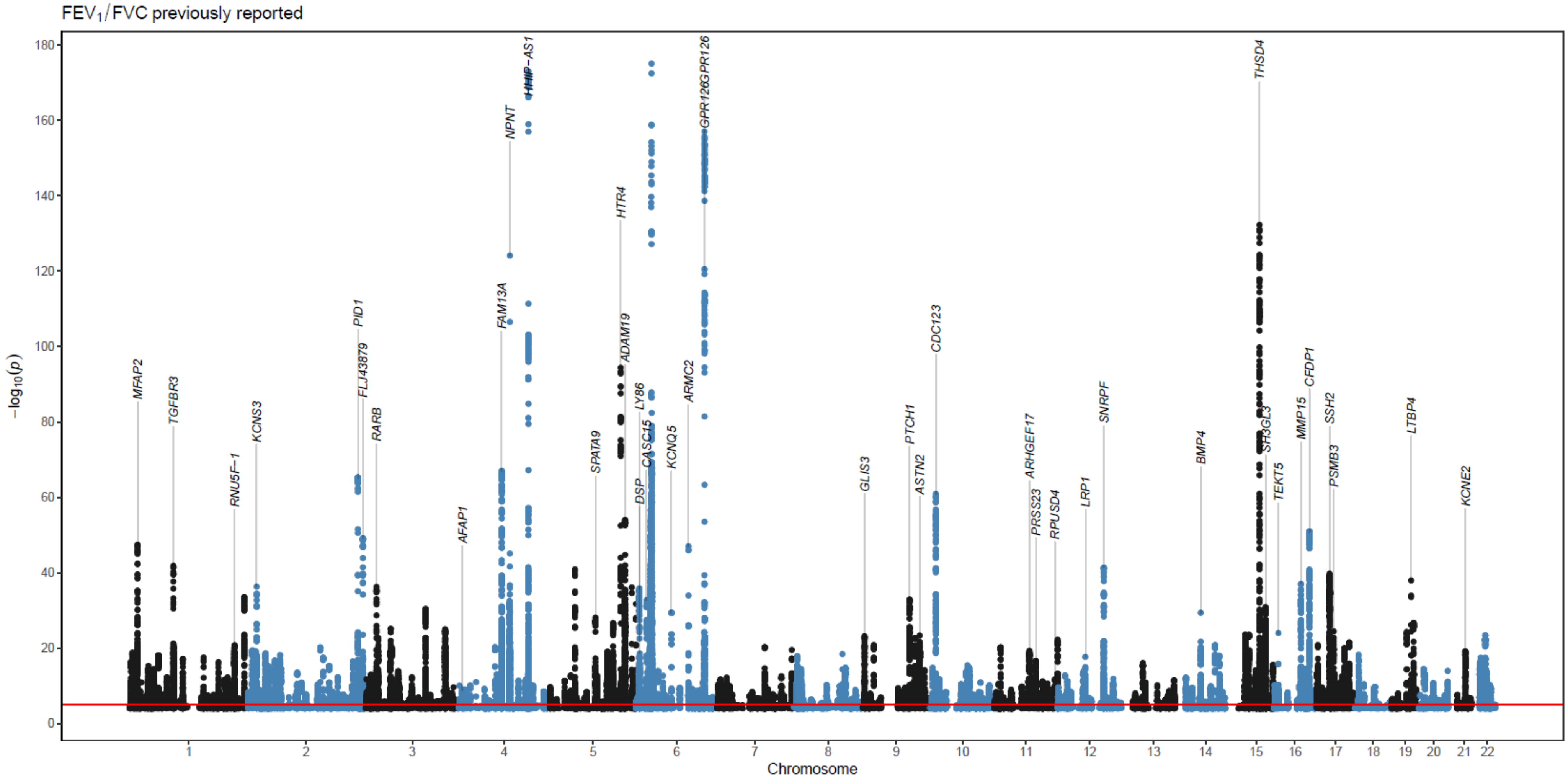
G) FEV$_1$/FVC novel tier 1 signals. P values from UK Biobank with P=5×10^{-9} (Tier 1 UK Biobank threshold) shown in red.



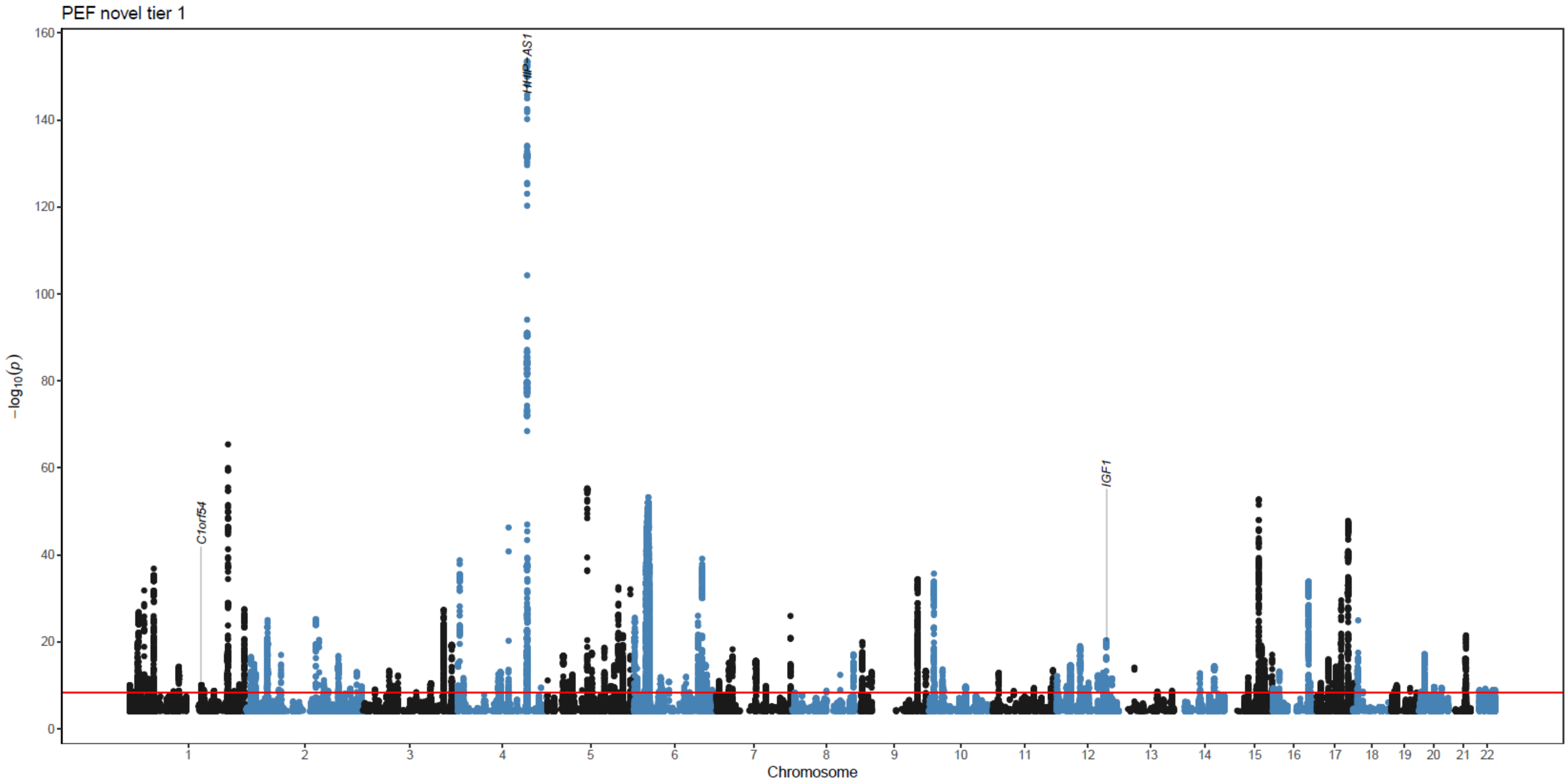
H) FEV_1/FVC novel tier 2 signals. P values from meta-analysis of UK Biobank and SpiroMeta with $P=5\times10^{-9}$ (Tier 2 meta-analysis threshold) shown in red.



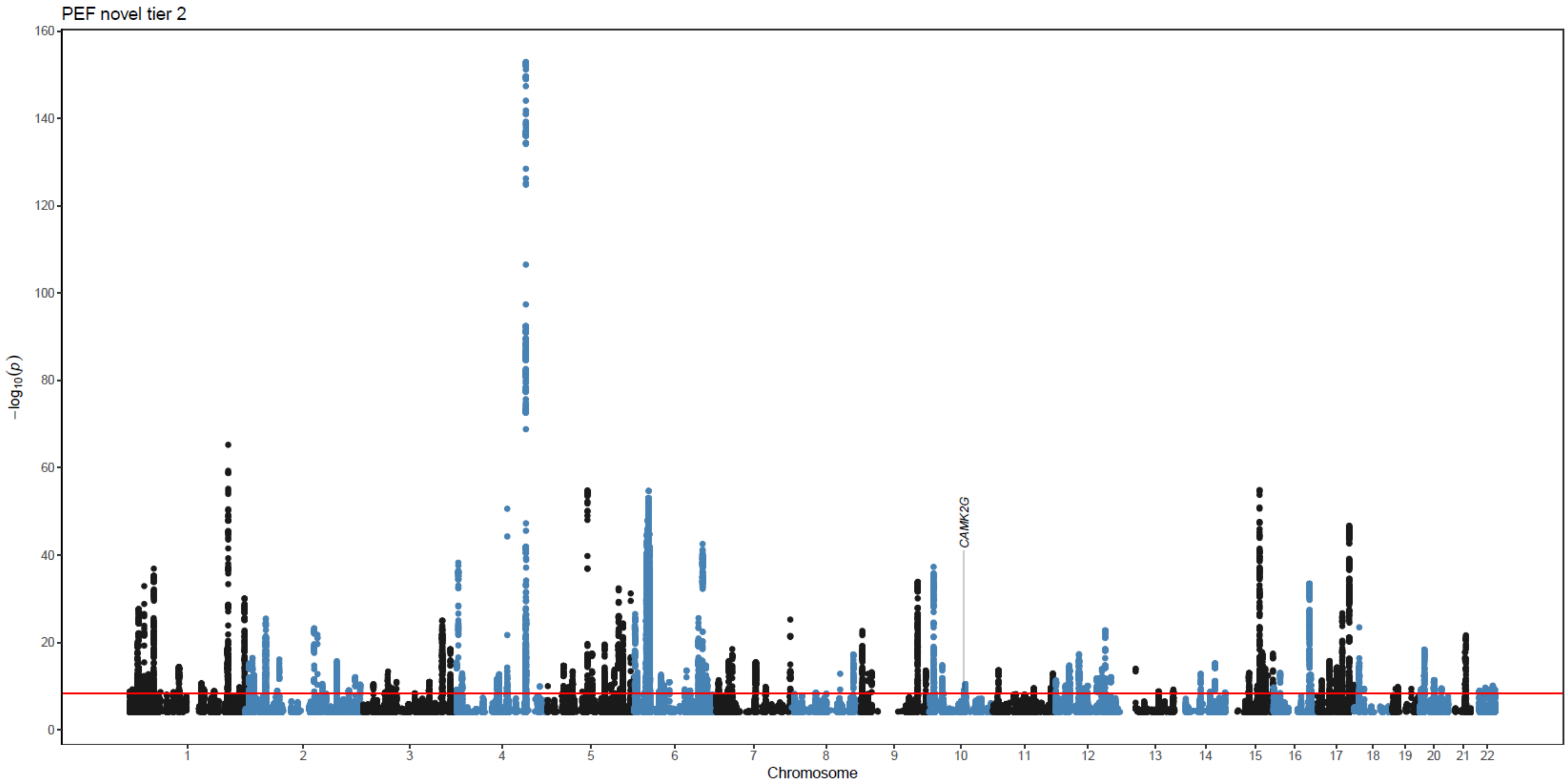
I) FEV_1/FVC previously reported signals. P values from UK Biobank with $P=10^{-5}$ (threshold for inclusion of previously reported signals in downstream analyses) in red.



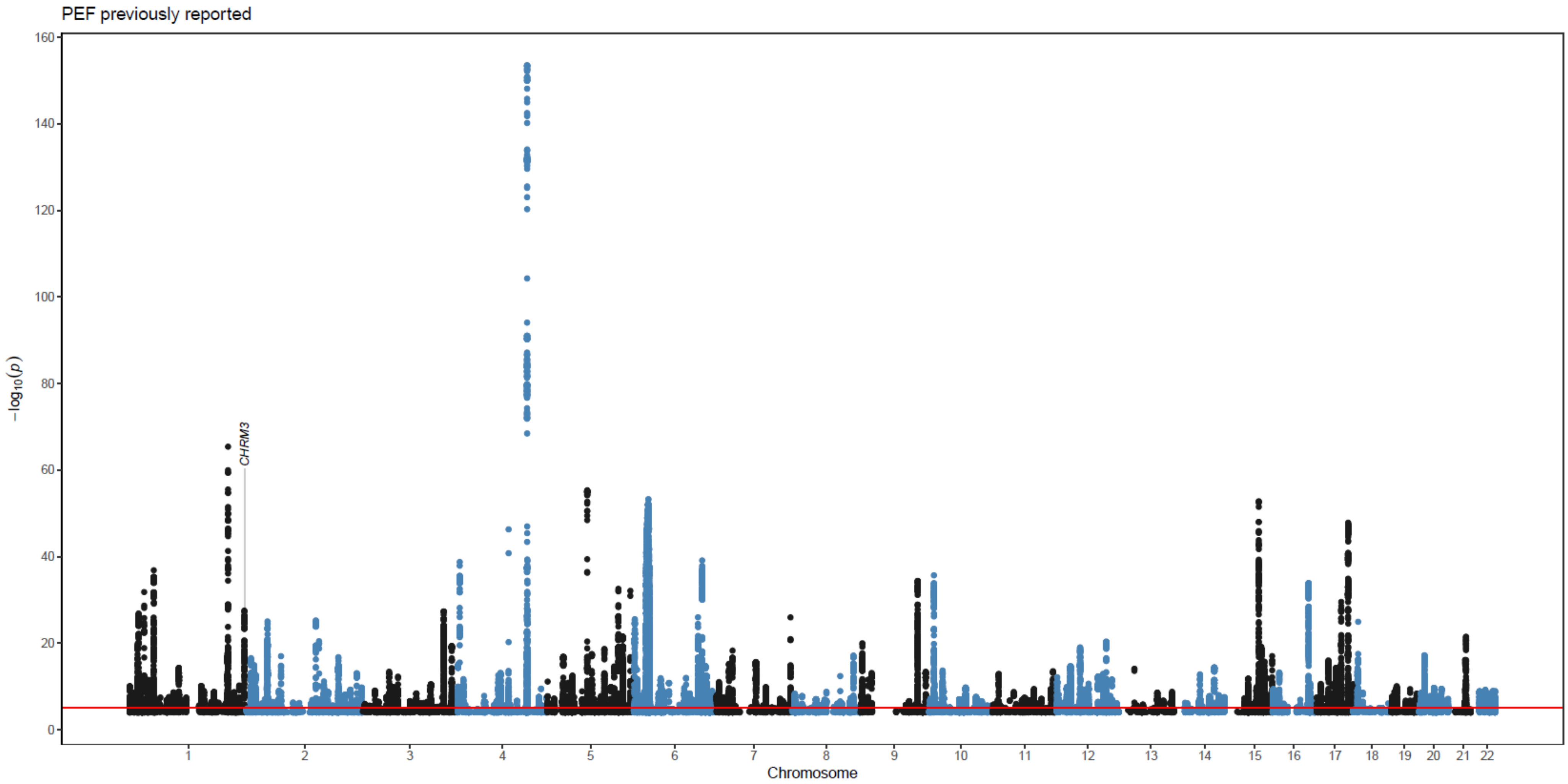
J) PEF novel tier 1 signals. P values from UK Biobank with $P=5\times10^{-9}$ (Tier 1 UK Biobank threshold) shown in red.



K) PEF novel tier 2 signals. P values from meta-analysis of UK Biobank and SpiroMeta with $P=5\times10^{-9}$ (Tier 2 meta-analysis threshold) shown in red.



L) PEF previously reported signals. P values from UK Biobank with $P=10^{-5}$ (threshold for inclusion of previously reported signals in downstream analyses) in red.



Supplementary Figure 3: Assessment of previously reported signals

Description of assessment of 184 signals for lung function or COPD previously reported in the literature. Signals were pruned for independence, leaving 157 signals. Corroboration of association was found for 142/157 signals. 2/142 signals were known to be associated with smoking behaviour, and in the current study, we replicated these findings, and also showed no association in non-smokers. After removing these signals, 140 remained for assessment. After combining with 139 novel signals, 279 signals entered downstream analyses.

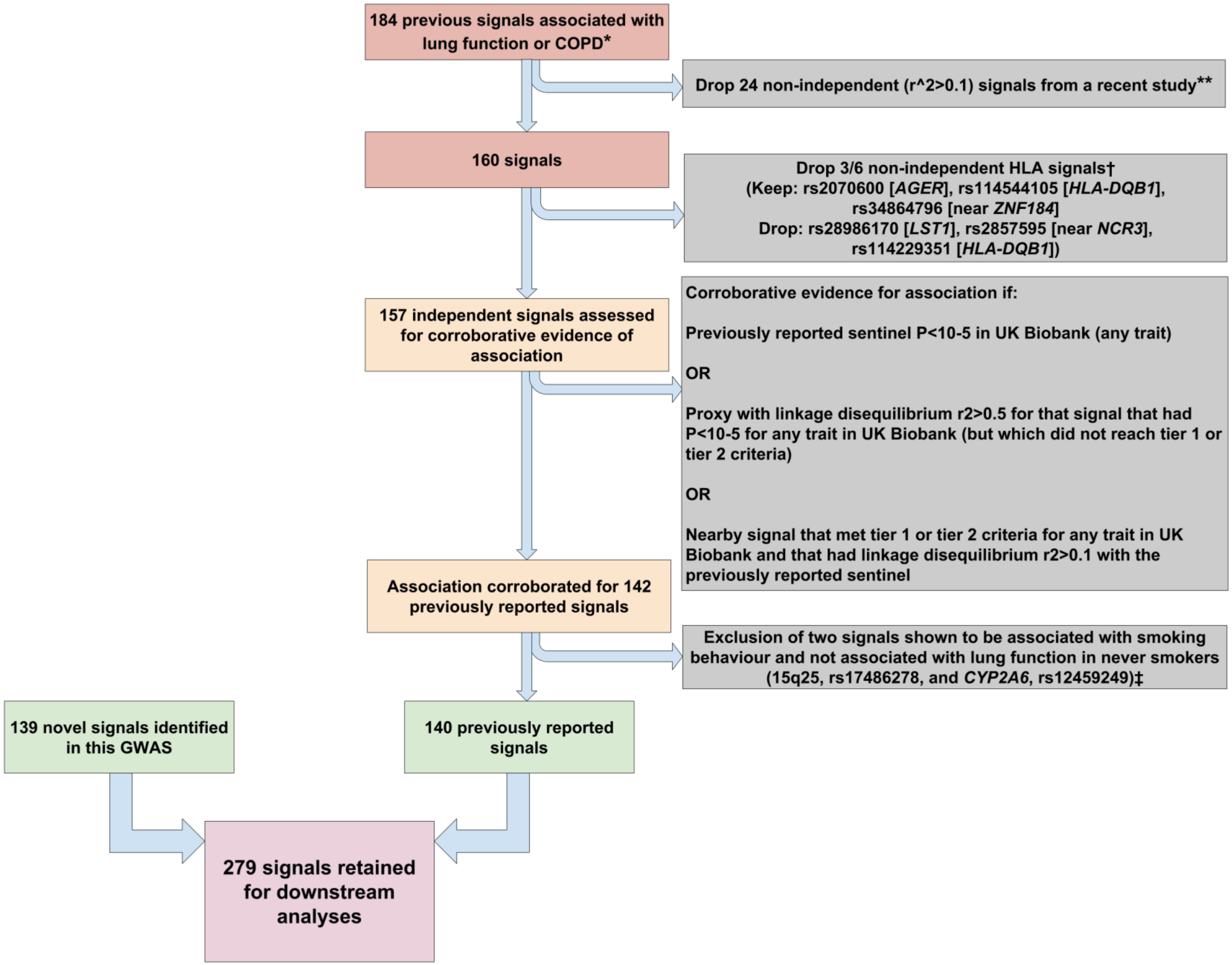
*=Wilk *et al.* 2009 [PMID: 19300500];[35] Hancock *et al.* 2010 [PMID: 20010835];[36] Repapi *et al.* 2010 [PMID: 20010834];[37] Soler Artigas *et al.* 2011 [PMID: 21946350];[38] Cho *et al.* 2012 [PMID: 22080838];[39] Loth *et al.* 2014 [PMID: 24929828];[40] Lutz *et al.* 2015 [PMID: 26634245];[41] Soler Artigas *et al.* 2015 [PMID: 21946350];[42] Wain *et al.* 2015 [PMID: 28166213];[2] Hobbs *et al.* 2016 [PMID: 26771213];[43] Hobbs *et al.* 2017 [PMID: 28166215];[44] Wain *et al.* 2017 [PMID: 26423011];[45] Wyss *et al.* 2017 [https://www.biorxiv.org/content/early/2017/10/05/196048][46]; Jackson *et al.* 2018 [https://wellcomeopenresearch.org/articles/3-4/v1];[47]

**=Wyss *et al.* 2017 [https://www.biorxiv.org/content/early/2017/10/05/196048][46]

†=See Wain *et al.* 2017 [PMID: 28166213] for details of HLA independence analysis.[45]

‡=Lutz *et al.* 2015 [PMID: 26634245];[41] Cho *et al.* 2012 [PMID: 22080838][39]

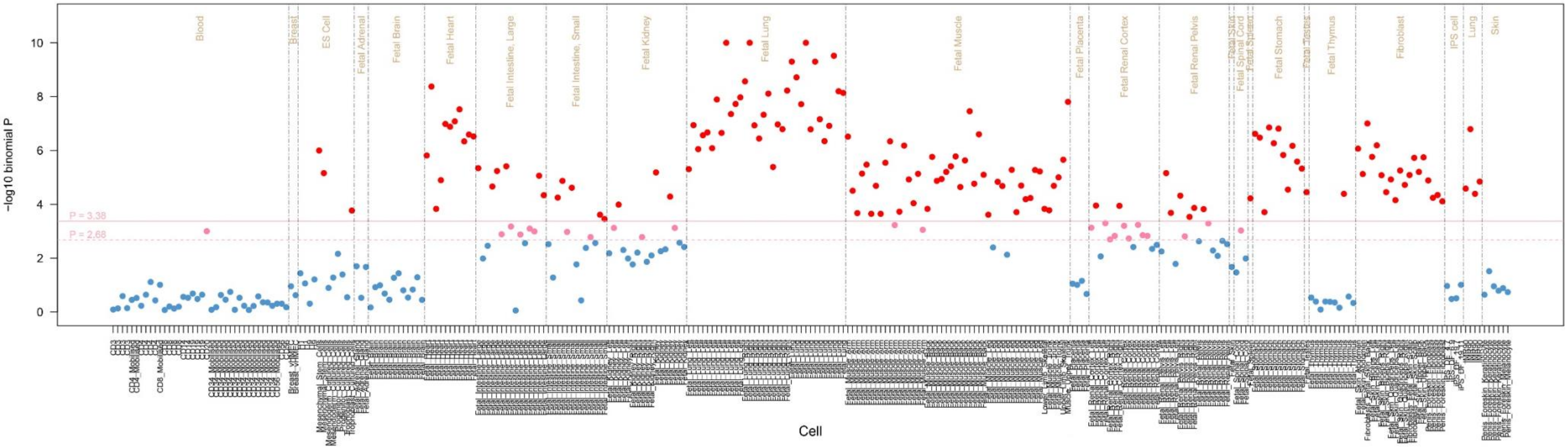
Figure on next page.


184 previous signals associated with lung function or COPD*
Drop 24 non-independent (r^2>0.1) signals from a recent study**
160 signals
Drop 3/6 non-independent HLA signals†
(Keep: rs2070600 [AGER], rs114544105 [HLA-DQB1], rs34864796 [near ZNF184]
Drop: rs28986170 [LST1], rs2857595 [near NCR3], rs114229351 [HLA-DQB1])
157 independent signals assessed for corroborative evidence of association
Corroborative evidence for association if:
Previously reported sentinel P<10-5 in UK Biobank (any trait)
OR
Proxy with linkage disequilibrium r2>0.5 for that signal that had P<10-5 for any trait in UK Biobank (but which did not reach tier 1 or tier 2 criteria)
OR
Nearby signal that met tier 1 or tier 2 criteria for any trait in UK Biobank and that had linkage disequilibrium r2>0.1 with the previously reported sentinel
Association corroborated for 142 previously reported signals
Exclusion of two signals shown to be associated with smoking behaviour and not associated with lung function in never smokers (15q25, rs17486278, and CYP2A6, rs12459249)‡
139 novel signals identified in this GWAS
140 previously reported signals
279 signals retained for downstream analyses

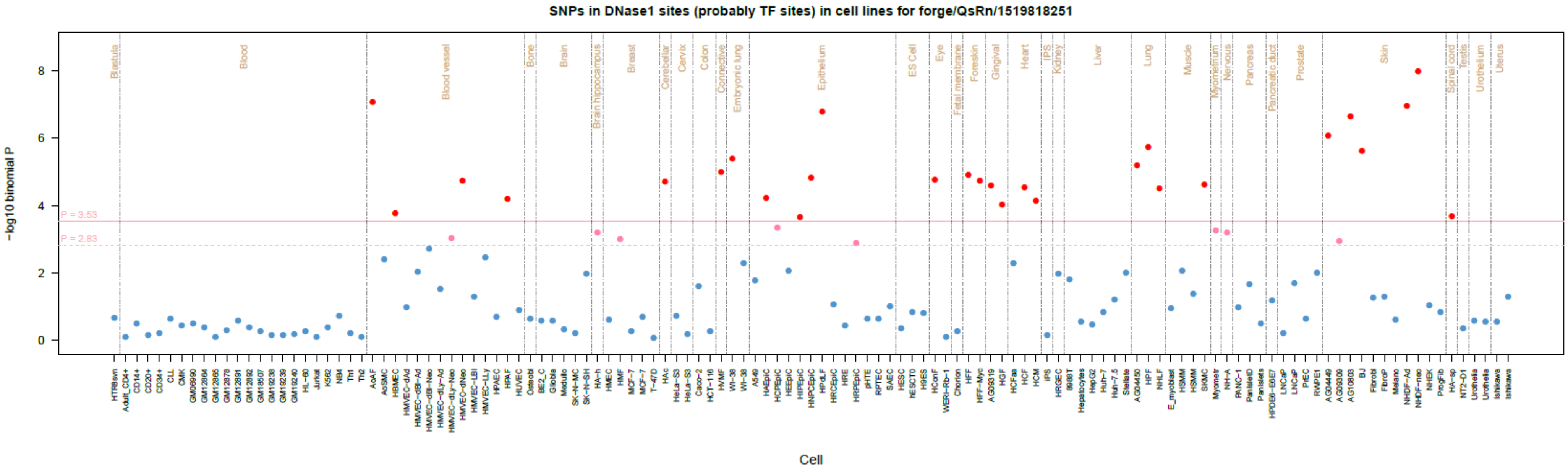
Supplementary Figure 4: Tissue-specific enrichment of overlap with DNase I hotspots

Tissue-specific enrichment of overlap for the 279 SNPs for each DNase I hotspot sample in the RoadMap Epigenome project (n=299) and ENCODE project (n=125). Each point represents the -log10 binomial *P*-value (y axis) of the enrichment of the 279 SNPs compared to matched background SNPs on a single DNase I hotspot sample, organised by tissue as indicated by the brown labels at the top of the figure (e.g. Blood, Breast), and alphabetically by cell sample (x axis; e.g. CD3). Points with Bonferroni adjusted $P \le 0.01$ and $P < 0.05$ are coloured in red and pink, respectively.

RoadMap Epigenome project

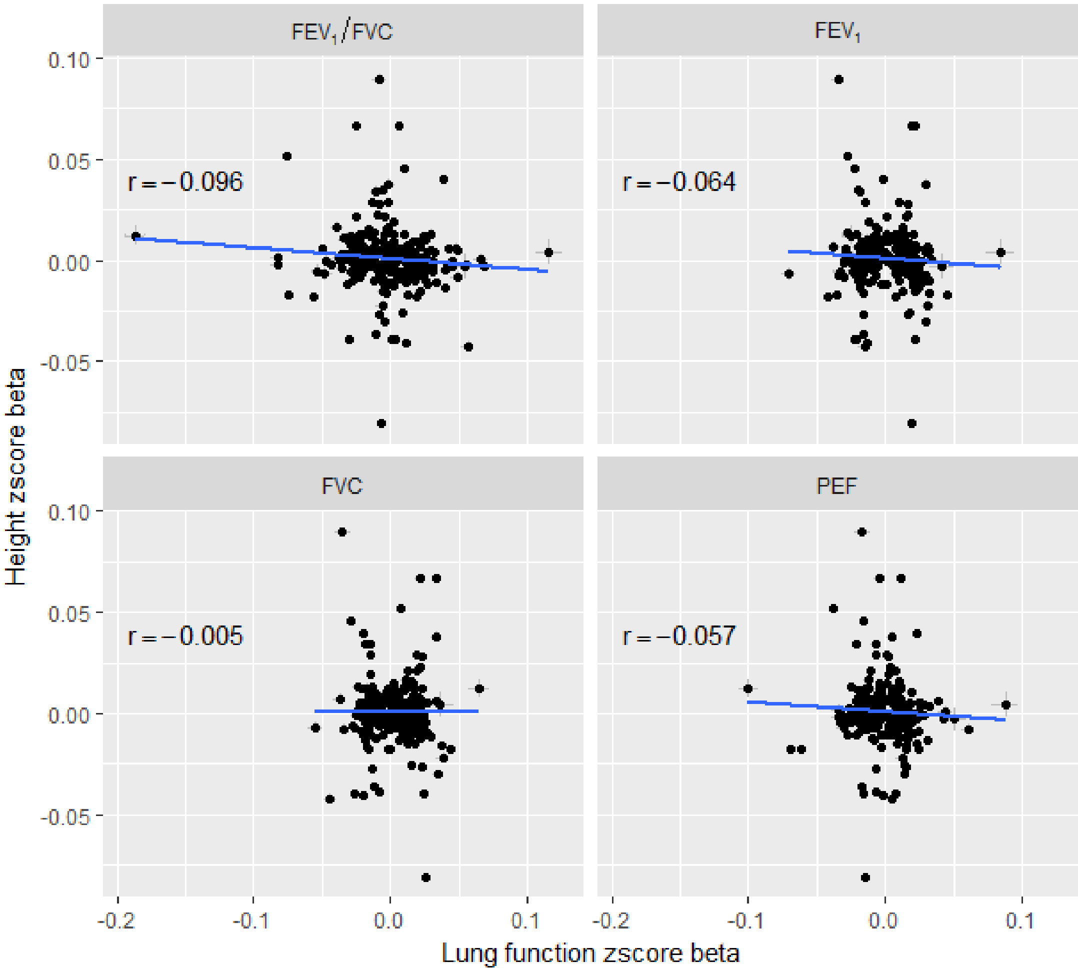


ENCODE project



Supplementary Figure 5: Comparison of genetic effects for height and lung function.

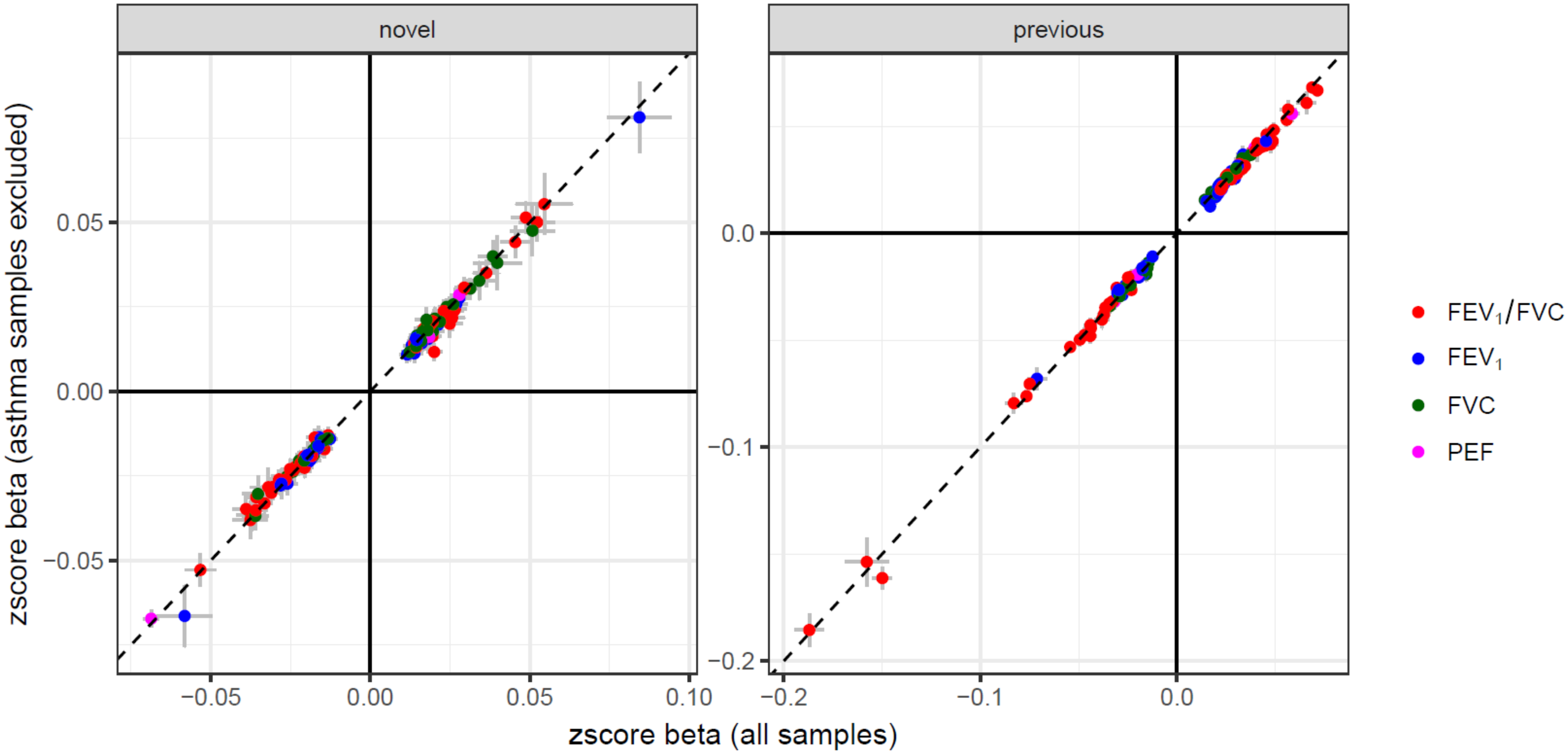
Height effect look up in GIANT[48] for 247 proxies (r^2>0.4) of our 276 credible set sentinel variants plotted against lung function effect in UK Biobank. All traits are rank inverse-normal transformed. There is no significant correlation (minimum P value 0.13 for FEV_1/FVC).



30 SNPs for which there was no proxy with r^2 > 0.4 in GIANT: rs72673461, rs141942982, rs55884799, rs72902177, rs12715478, rs34712979, rs2353940, rs7733410, rs10059996, rs79898473, rs12698403, rs7838717, rs7090277, rs10998018, rs56196860, rs2812208, rs35107139, rs56383987, rs3751837, rs78442819, rs76219171, rs35420030, rs2345443, rs62070648, rs35246838, rs77672322, rs59606152, rs34093919, rs2283847, rs113111175

Supplementary Figure 6: Comparison of effect sizes after excluding asthma samples

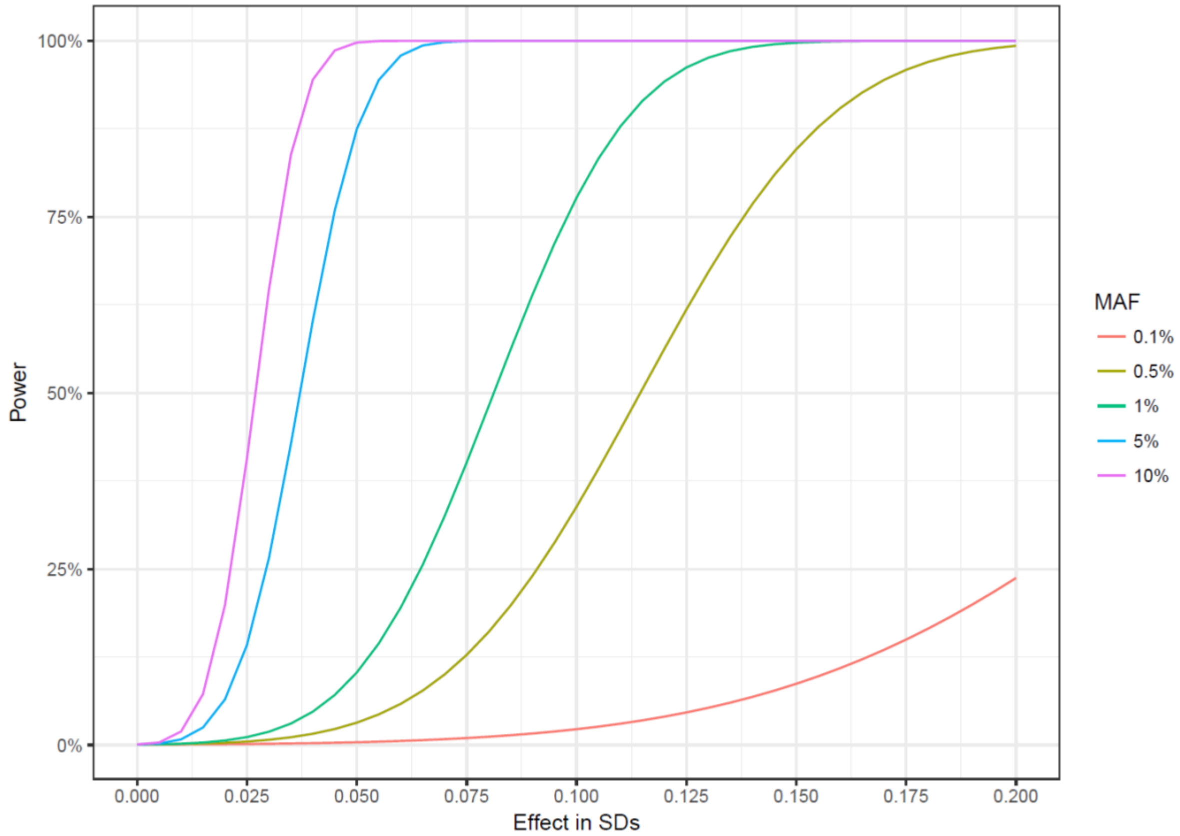
Comparison of effects (inverse-normal rank-transformed zscores) in UK Biobank for 139 novel and 140 previously reported signals in all UK Biobank samples with lung function data (x-axis, N=321,047) and after excluding 37,868 samples with doctor diagnosed asthma (y-axis, N=283,179). Doctor diagnosed asthma is self-reported touchscreen answer (UK Biobank field ID: 6152).



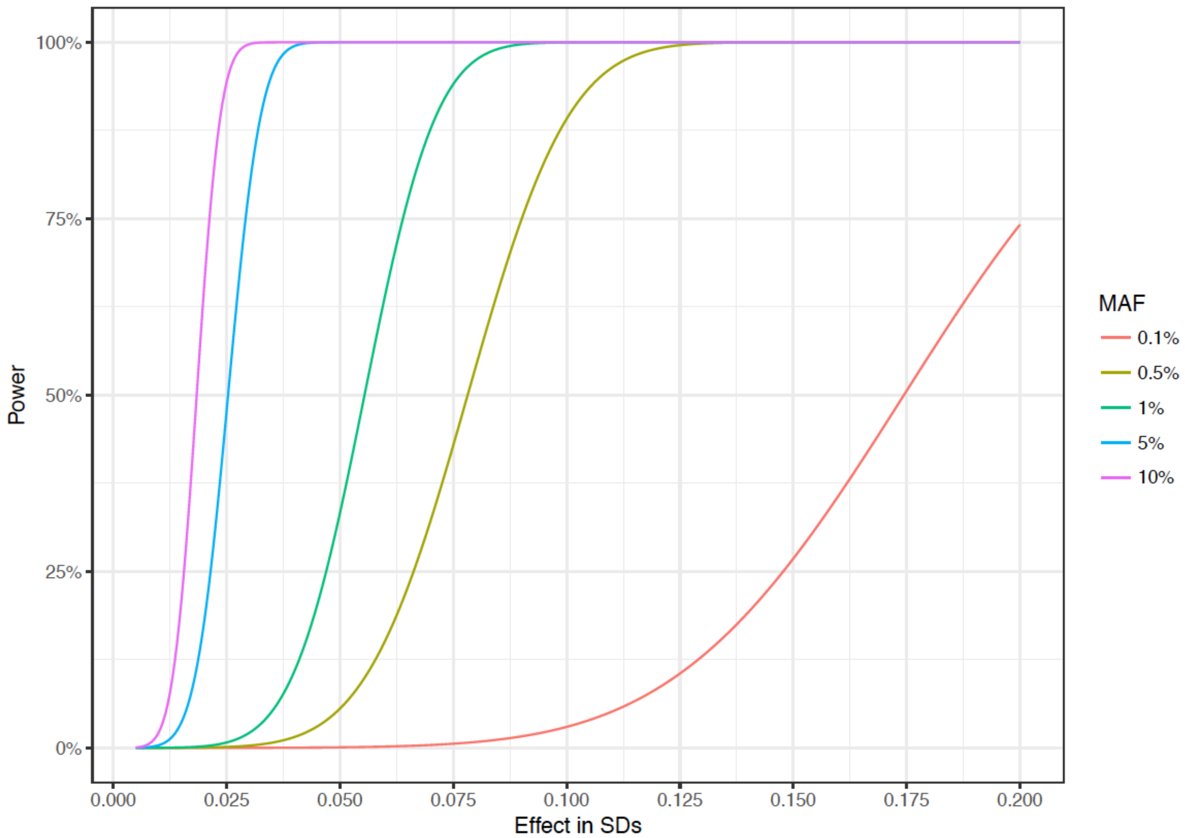
Supplementary Figure 7: Power calculations

Power to detect a range of single variant effect sizes, as standard deviations (SDs) of the continuous lung function phenotypes (FEV_1, FVC, FEV_1/FVC or PEF), over a range of minor allele frequencies (MAF).

A) Power to meet tier 1 and tier 2 criterion $P<10^{-3}$ in SpiroMeta (n=79,055)



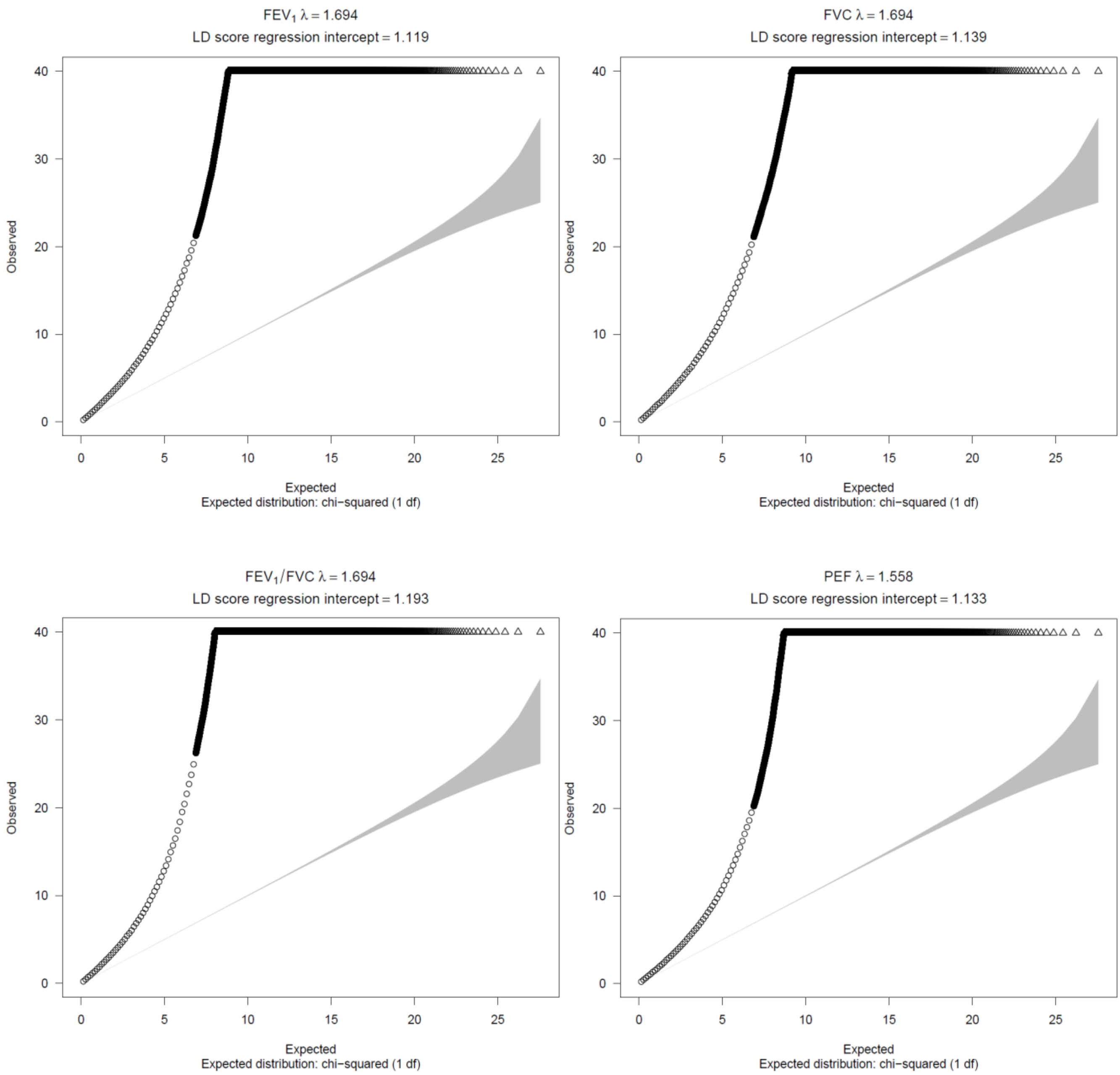
B) Power for association of previous signals $P<10^{-5}$ in UK Biobank (n=321,047)



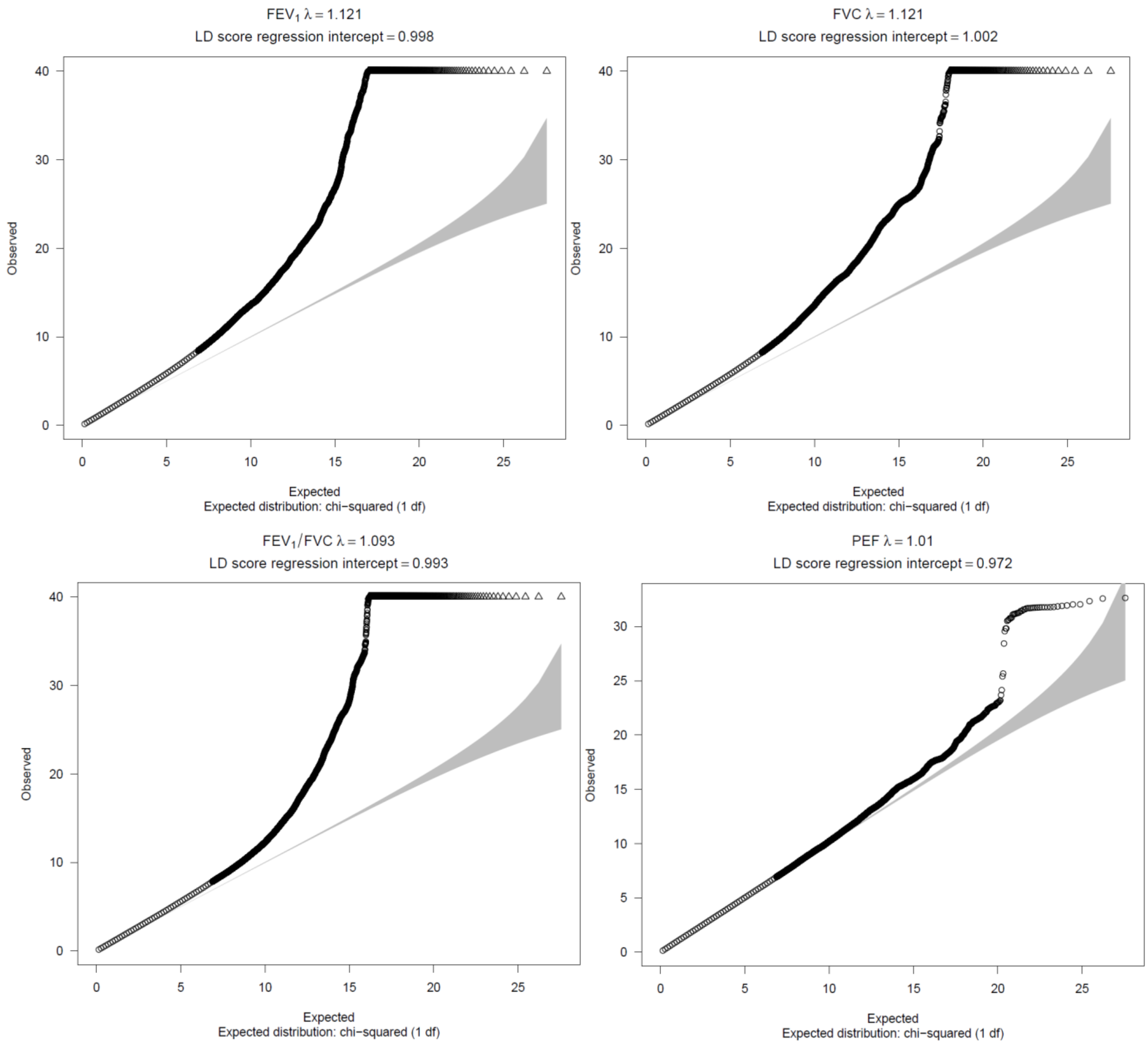
Supplementary Figure 8: QQ plots

LD score regression implemented in LDSC[49] was used to estimate inflation of test statistics due to confounding. The unadjusted genomic inflation factors λ are shown as well as the LD score regression intercept which is the inflation factor adjusted for polygenicity. Genomic control was applied if the LD score regression intercept was larger than 1.05 suggesting residual inflation. Accordingly, genomic control was applied to UK Biobank but not SpiroMeta.

UK Biobank



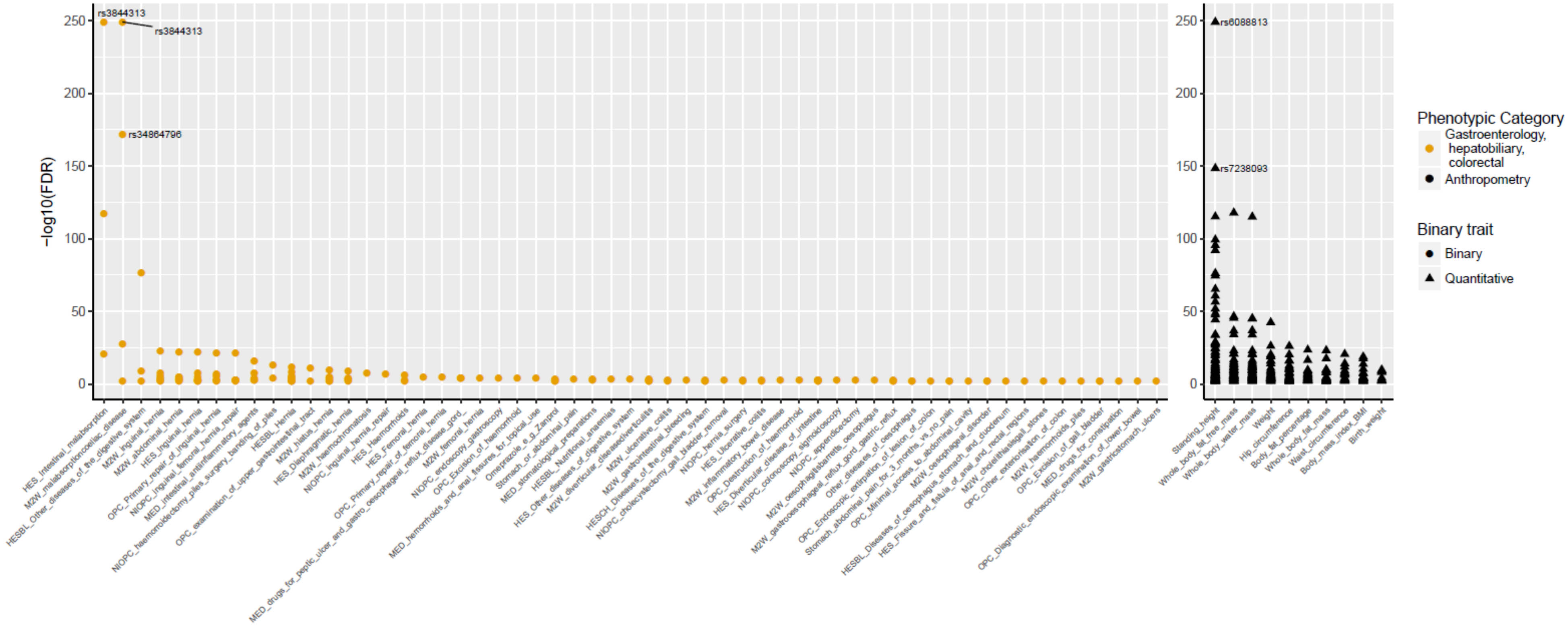
SpiroMeta



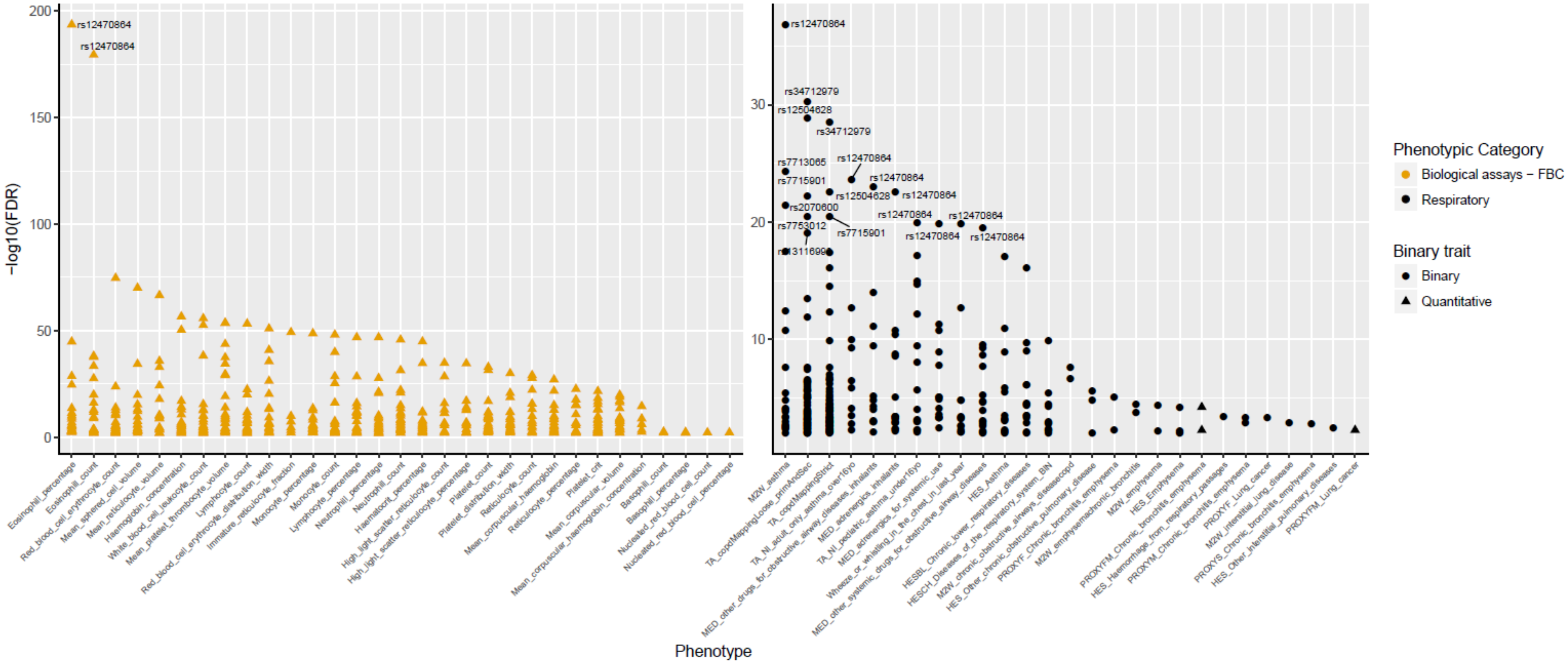
FEV$_1$ λ = 1.121
LD score regression intercept = 0.998
FVC λ = 1.121
LD score regression intercept = 1.002
FEV$_1$/FVC λ = 1.093
LD score regression intercept = 0.993
PEF λ = 1.01
LD score regression intercept = 0.972
Observed
Expected
Expected distribution: chi−squared (1 df)

Supplementary Figure 9: Individual PheWAS results, separately by trait category

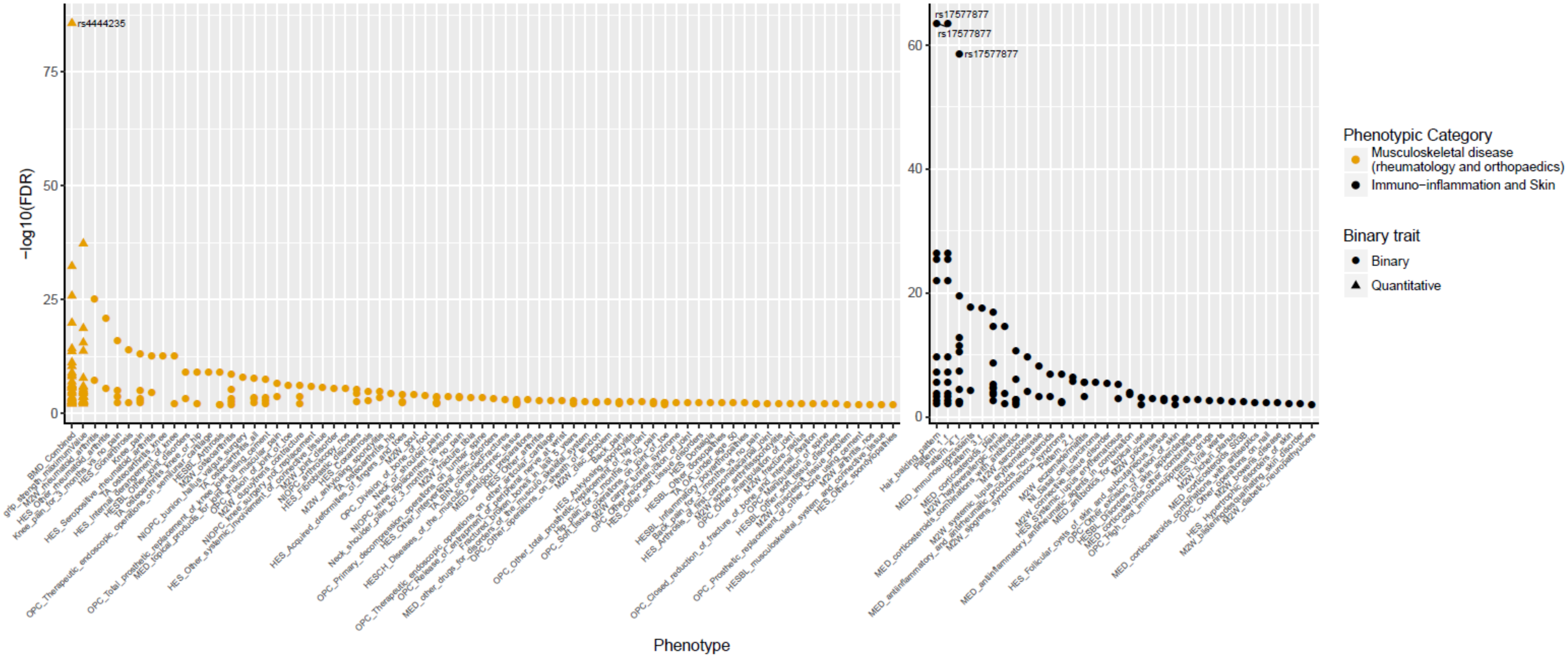
In these extended plots, individual associations passing FDR 1% between the 279 lung function signals and 2411 traits are shown (y-axis: -log10(FDR)). Each category has its own separate subplot. Categories are presented in order of their most significant (according to FDR) association, and within each subplot, results for each trait are presented in decreasing order of FDR. Triangular points indicate quantitative traits, and circular points indicate binary traits. For each category, associations that are >50% of the highest –log10(FDR) value are labelled with their rsID. Individual SNP results are also available in a separate file for download. Due to the size of these data, individual results of the 279*2411 SNP-trait associations, along with details of their categorisation, and the plain English labels used in **Figure 4** of the main manuscript are available from the authors on request.

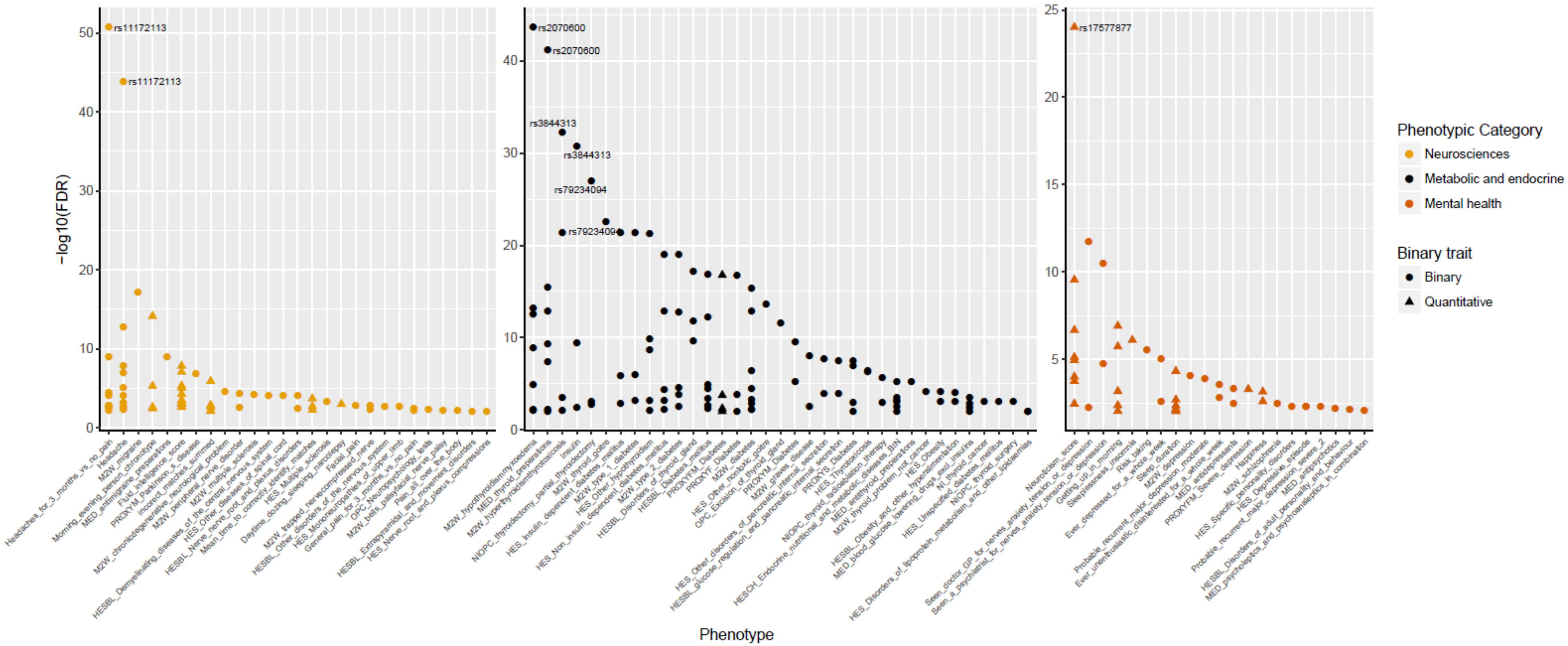


rs3844313
rs3844313
rs34864796
rs6088813
rs7238093
−log10(FDR)
250
200
150
100
50
0
Phenotype
Standing_height
Whole_body_fat_free_mass
Whole_body_water_mass
Weight
Hip_circumference
Body_fat_percentage
Whole_body_fat_mass
Waist_circumference
Body_mass_index_BMI
Birth_weight
Phenotypic Category
Gastroenterology, hepatobiliary, colorectal
Anthropometry
Binary trait
Binary
Quantitative

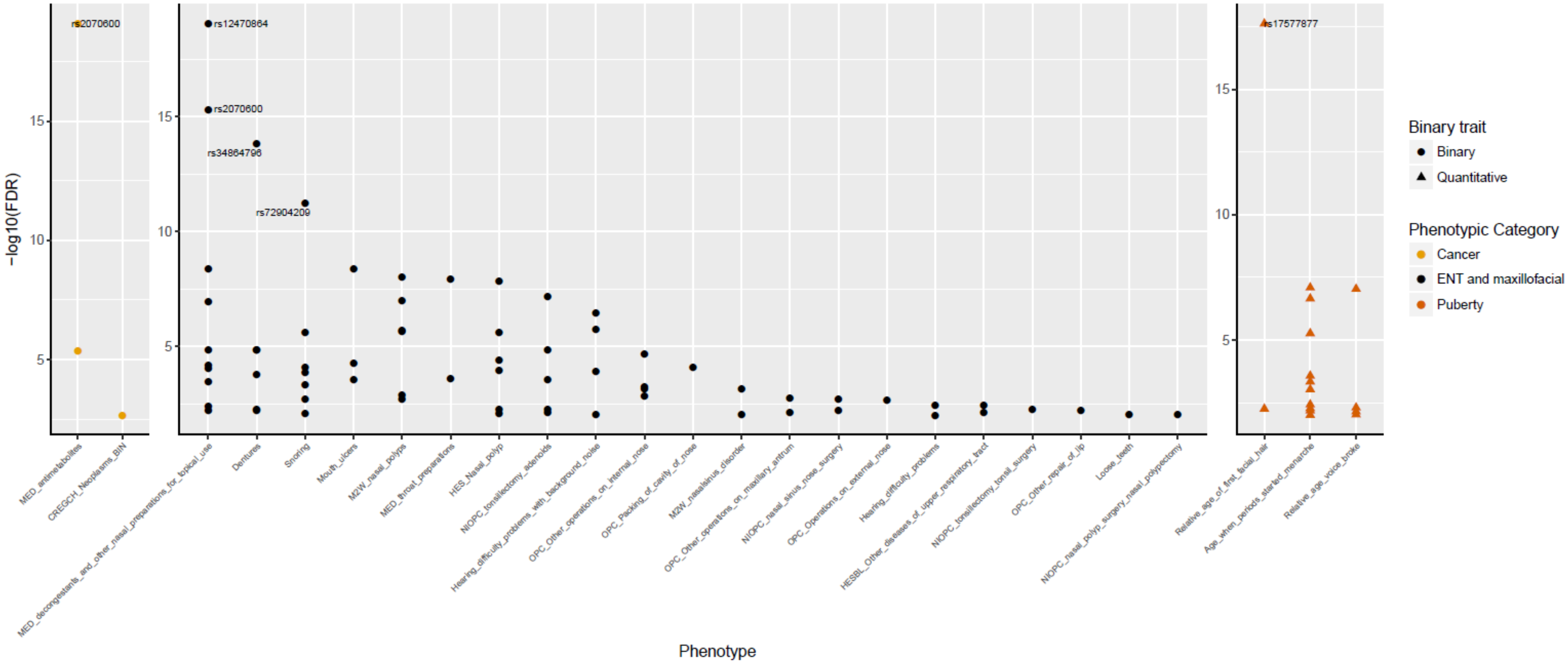


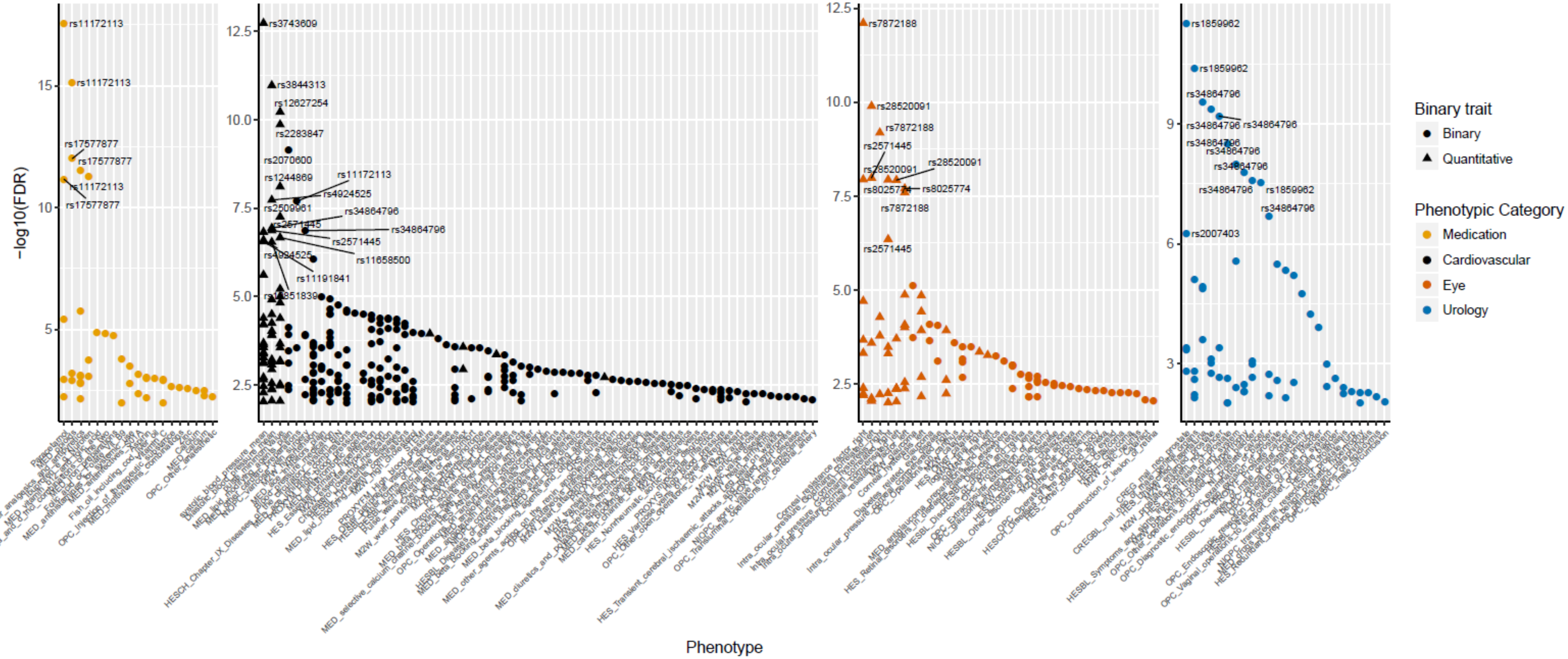
-log10(FDR)
Phenotype
Phenotypic Category
Biological assays − FBC
Respiratory
Binary trait
Binary
Quantitative
rs12470864
rs34712979
rs12504628
rs7713065
rs7715901
rs2070600
rs7753012

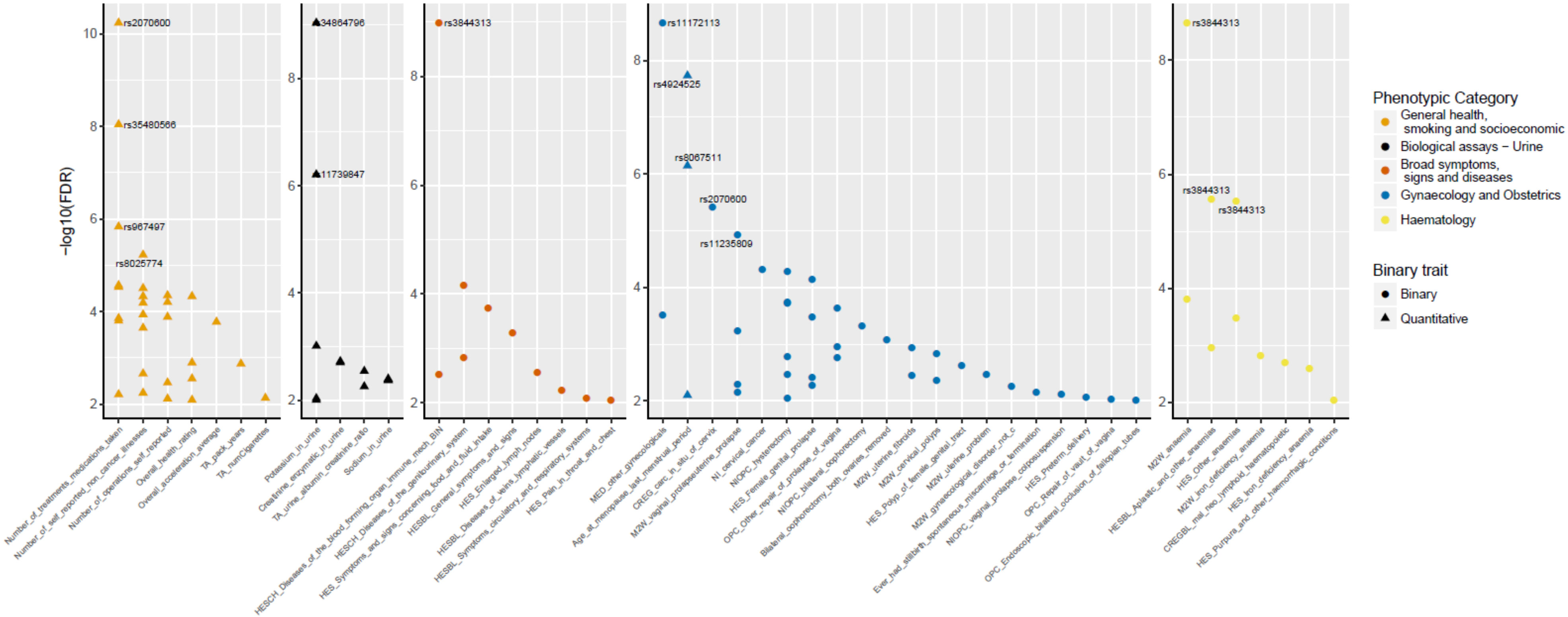

rs4444235
rs17577877
rs17577877
rs17577877
-log10(FDR)
Phenotype
Phenotypic Category
Musculoskeletal disease (rheumatology and orthopaedics)
Immuno−inflammation and Skin
Binary trait
Binary
Quantitative

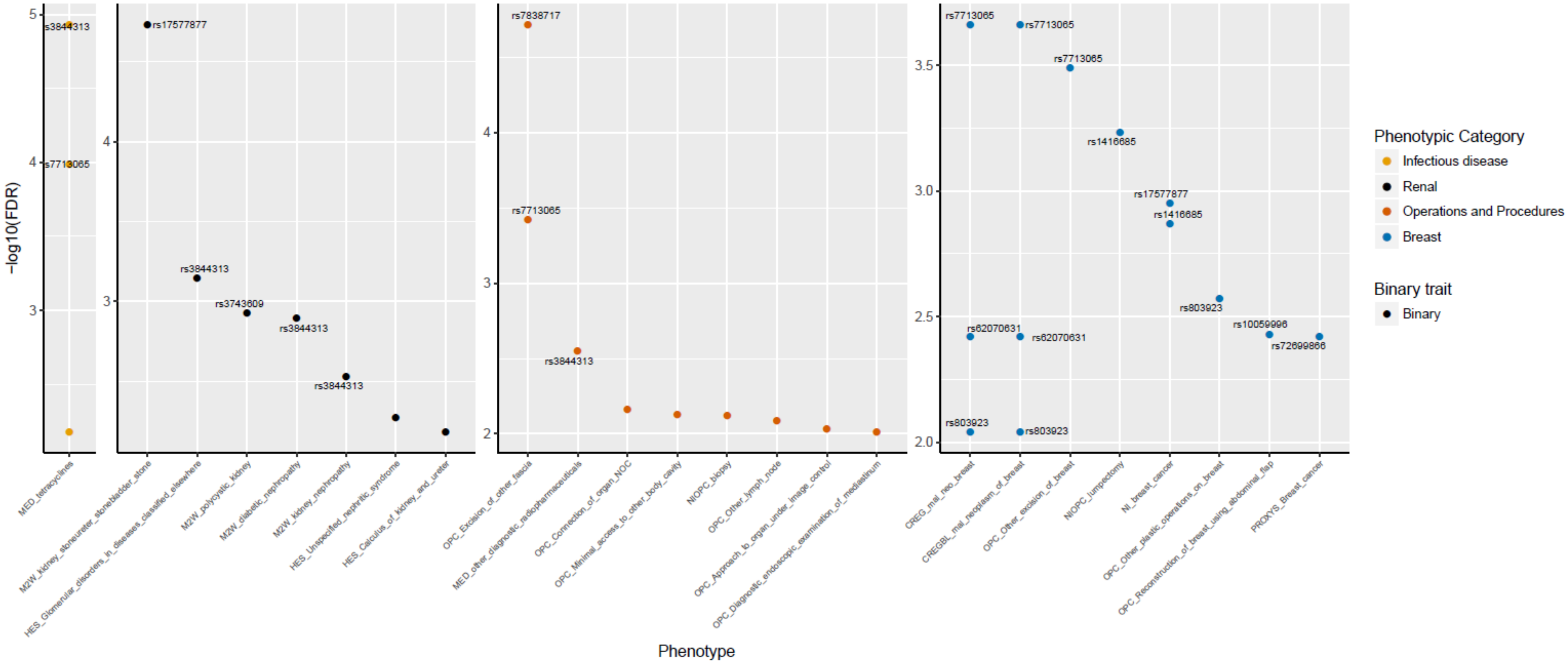


rs11172113
rs11172113
rs2070600
rs2070600
rs3844313
rs3844313
rs79234094
rs79234094
rs17577877
−log10(FDR)
Phenotype
Phenotypic Category
Neurosciences
Metabolic and endocrine
Mental health
Binary trait
Binary
Quantitative


-log10(FDR)
Phenotype
rs2070600
rs12470864
rs2070600
rs34864796
rs72904209
rs17577877
MED_antimetabolites
CREGCH_Neoplasms_BIN
MED_decongestants_and_other_nasal_preparations_for_topical_use
Dentures
Snoring
Mouth_ulcers
M2W_nasal_polyps
MED_throat_preparations
HES_Nasal_polyp
NIOPC_tonsillectomy_adenoids
Hearing_difficulty_problems_with_background_noise
OPC_Other_operations_on_internal_nose
OPC_Packing_of_cavity_of_nose
M2W_nasalsinus_disorder
OPC_Other_operations_on_maxillary_antrum
NIOPC_nasal_sinus_nose_surgery
OPC_Operations_on_external_nose
Hearing_difficulty_problems
HESBL_Other_diseases_of_upper_respiratory_tract
NIOPC_tonsillectomy_tonsil_surgery
OPC_Other_repair_of_lip
Loose_teeth
NIOPC_nasal_polyp_surgery_nasal_polypectomy
Relative_age_of_first_facial_hair
Age_when_periods_started_menarche
Relative_age_voice_broke
Binary trait
Binary
Quantitative
Phenotypic Category
Cancer
ENT and maxillofacial
Puberty


-log10(FDR)
Phenotype
Binary trait
Binary
Quantitative
Phenotypic Category
Medication
Cardiovascular
Eye
Urology
rs11172113
rs17577877
rs3743609
rs3844313
rs12627254
rs2283847
rs2070600
rs1244869
rs4924525
rs2509961
rs34864796
rs2571445
rs11658500
rs11191841
rs7872188
rs28520091
rs8025774
rs1859962
rs2007403


-log10(FDR)
Phenotype
Phenotypic Category
General health, smoking and socioeconomic
Biological assays − Urine
Broad symptoms, signs and diseases
Gynaecology and Obstetrics
Haematology
Binary trait
Binary
Quantitative
rs2070600
rs35480566
rs967497
rs8025774
rs34864796
rs11739847
rs3844313
rs11172113
rs4924525
rs8067511
rs2070600
rs11235809
rs3844313
rs3844313
rs3844313



-log10(FDR)
Phenotype
Phenotypic Category
Infectious disease
Renal
Operations and Procedures
Breast
Binary trait
Binary
MED_tetracyclines
M2W_kidney_stoneureter_stonebladder_stone
HES_Glomerular_disorders_in_diseases_classified_elsewhere
M2W_polycystic_kidney
M2W_diabetic_nephropathy
M2W_kidney_nephropathy
HES_Unspecified_nephritic_syndrome
HES_Calculus_of_kidney_and_ureter
OPC_Excision_of_other_fascia
MED_other_diagnostic_radiopharmaceuticals
OPC_Connection_of_organ_NOC
OPC_Minimal_access_to_other_body_cavity
NIOPC_biopsy
OPC_Other_lymph_node
OPC_Approach_to_organ_under_image_control
OPC_Diagnostic_endoscopic_examination_of_mediastinum
CREG_mal_neo_breast
CREGBL_mal_neoplasm_of_breast
OPC_Other_excision_of_breast
NIOPC_lumpectomy
NI_breast_cancer
OPC_Other_plastic_operations_on_breast
OPC_Reconstruction_of_breast_using_abdominal_flap
PROXYS_Breast_cancer
rs3844313
rs7713065
rs17577877
rs3743609
rs7838717
rs1416685
rs803923
rs62070631
rs10059996
rs72699866

Supplementary Tables

Supplementary Table 1: UK Biobank demographics

Demographic information for UK Biobank samples of European ancestry used in discovery.

N Total	321,047
N male	142,558
N female	178,489
Age range (y) at lung function measurement	39-72
Mean age, y (s.d.)	56.44 (8.02)
Mean height, cm (s.d.)	168.57 (9.13)
Mean FEV_1, L (s.d.)	2.84 (0.76)
Mean FVC, L (s.d.)	3.74 (0.96)
Mean FEV_1/FVC (s.d.)	0.76 (0.06)
Mean PEF, L/min (s.d.)	406.19 (117.55)
N never smokers	173,658
N ever smokers	147,389
UK BiLEVE array	49,107
UK Biobank array	271,940

Supplementary Table 2: SpiroMeta Studies

B58C (B58C-T1DGC, British 1958 Birth Cohort–Type 1 Diabetes Genetics Consortium; B58C-GABRIEL British 1958 Birth Cohort–GABRIEL consortium; B58C-WTCCC, British 1958 Birth Cohort–Wellcome Trust Case Control Consortium); BHS1&2, Busselton Health Study 1 and 2; the CROATIA- Korcula study; the CROATIA-Split study; the CROATIA-Vis study; EPIC population based, European Prospective Investigation into Cancer and Nutrition Cohort; GS:SFHS, Generation Scotland: Scottish Family Health Study; H2000, Finnish Health 2000 survey; KORA F4, Cooperative Health Research in the Region of Augsburg; KORA S3, Cooperative Health Research in the Region of Augsburg; LBC1936, Lothian Birth Cohort 1936; NFBC1966, Northern Finland Birth Cohort of 1966; NFBC1966, Northern Finland Birth Cohort of 1986; NSPHS, Northern Sweden Population Health Study; ORCADES, Orkney Complex Disease Study; PIVUS, Prospective Investigation of the Vasculature in Uppsala Seniors; SHIP, Study of Health in Pomerania; SHIP-TREND; UKHLS; VIKING; YFS, the Young Finish Study. The total size in this table is not exactly equal to the maximum sample size given in the main text, since some studies had subtly different subsets of individuals entering each of the four lung function trait GWAS.

Study name	**N Total**	**N male**	**N female**	**Age range (y) at lung function measurement**	**Mean age, y (s.d.)**	**Mean height, cm (s.d.)**	**Mean FEV_1, L (s.d.)**	**Mean FVC, L (s.d.)**	**Mean FEV_1/FVC (s.d.)**	**Mean PEF, L/min (s.d.)**	**N never smokers**	**N ever smokers**	**Genotyping Platform**	**Imputation Panel**
B58C	5934	2955	2979	44-45	45.12 (0.38)	169.43 (9.29)	3.30 (0.76)	4.19 (0.98)	0.79 (0.08)	--	1709	4225	Illumina 550k/610k	1000G
BHS1&2	4355	1922	2433	17-97	51.21 (17.00)	168.00 (9.39)	3.01 (0.96)	3.88 (1.16)	0.77 (0.07)	--	2301	2054	Illumina 610-Quad (N=1,168) & Illumina 660W-Quad (N=3,428)	1000G
CROATIA-Korcula	826	302	524	18-90	55.63 (13.50)	168.10 (9.20)	2.72 (0.83)	3.29 (0.96)	0.83 (0.1)	--	403	423	Illumina HumanHap370CNV duo chip	1000G
CROATIA-Split	493	210	283	18-85	49.08 (14.63)	172.60 (9.49)	3.19 (0.91)	3.80 (1.06)	0.84 (0.08)	--	239	254	Illumina HumanHap370CNV quad chip	1000G
CROATIA-Vis	925	390	535	18-88	55.90 (15.51)	167.80 (9.88)	3.42 (1.21)	4.41 (1.42)	0.77 (0.09)	--	388	537	Illumina Infinium HumanHap300 BeadChip	1000G
EPIC population based	20771	9664	11107	39-79	59.1 (9.27)	167.1 (9.08)	2.51 (0.74)	3.06 (0.93)	0.83 (0.11)	364.07 (123.16)	9532	11239	Affymetrix UKBioBank Axiom	HRC
GS:SFHS	16048	6633	10415	18-99	46.87 (14.6)	168.4 (9.50)	2.97 (0.88)	3.88 (1.00)	0.76 (0.11)	--	8581	7467	Illumina OmniExpress+Exome	HRC
H2000	821	394	427	30-75	50.47 (10.91)	169.10 (9.14)	3.29 (0.9)	4.16 (1.07)	0.79 (0.07)	--	249	572	Illumina HumanHap 610K	1000G
KORA F4	1474	717	757	41-84	55.08 (9.90)	169.15 (9.42)	3.23 (0.85)	4.19 (1.05)	0.77 (0.07)	--	556	918	Affymetrix Axiom	1000G
KORA S3	1147	551	596	28-89	50.82 (15.23)	169.22 (9.32)	3.34 (0.90)	4.10 (1.06)	0.81 (0.08)	--	520	627	Illumina Omni 2.5/ Illumina Omni Express	1000G
LBC1936	991	501	490	68-71	69.55 (0.84)	166.44 (8.93)	2.38 (0.67)	3.04 (0.87)	0.79 (0.10)	--	437	554	Illumina 610-Quadv1	1000G
NFBC1966	5078	2417	2661	30-32	31.15 (0.35)	171.24 (9.09)	3.95 (0.79)	4.72 (0.99)	0.84 (0.06)	--	2478	2600	Illumina HumanCNV-370DUO Analysis BeadChip	HRC

Study name	N Total	N male	N female	Age range (y) at lung function measurement	Mean age, y (s.d.)	Mean height, cm (s.d.)	Mean FEV_1, L (s.d.)	Mean FVC, L (s.d.)	Mean FEV_1/FVC (s.d.)	Mean PEF, L/min (s.d.)	N never smokers	N ever smokers	Genotyping Platform	Imputation Panel
NFBC1986	3210	1516	1694	14-16	16.01 (0.37)	169.34 (8.43)	3.78 (0.70)	4.31 (0.85)	0.88 (0.08)	--	2476	734	Illumina Human Omni Express Exome 8v1.2	HRC
NSPHS	871	400	471	14-91	49.20 (20.00)	164.00 (10.10)	2.92 (0.90)	3.53 (1.06)	0.83 (0.09)	--	750	121	Illumina Infinum HapMap 300 v2 & Illumina Human OmniExpress	1000G
ORCADES	1802	719	1083	16-91	54.00 (15.00)	166.00 (9.20)	2.89 (0.83)	3.61 (0.99)	0.79 (0.08)	--	1022	780	Illumina Hap300, Illumina Omni1 & Illumina OmniX	1000G
PIVUS	806	395	411	69-72	70.20 (0.176)	169.09 (9.208)	2.45 (0.680)	3.23 (0.869)	0.764 (0.103)	--	393	413	Illumina OmniExpress and Metabochip	HRC
SAPALDIA	1378	665	713	18-61	41.30 (11.20)	169.47 (9.12)	3.53 (0.86)	4.50 (1.04)	0.78 (0.08)	--	631	747	Illumina 610k quad	1000G
SHIP	1759	860	899	20-80	47.17 (13.67)	169.7 (9.13)	3.28 (0.89)	3.87 (1.03)	0.85 (0.06)	437.58 (125.17)	818	941	Affymetrix SNP 6.0	HRC
SHIP-TREND	804	363	441	21-81	51.24 (13.34)	169.9 (9.00)	3.29 (0.87)	4.14 (1.06)	0.80 (0.06)	392.82 (125.91)	342	462	Illumina Human Omni 2.5	HRC
UKHLS	7442	3290	4152	16-99	53.11 (15.94)	167.7 (9.45)	2.84 (0.90)	3.83 (1.09)	0.75 (0.09)	--	2938	4504	Illumina CoreExome v1.0	HRC
VIKING	1701	672	1029	18-91	50.72 (14.97)	168 (0.09)	3.07 (0.81)	4.02 (0.96)	0.76 (0.09)	450.08 (130.14)	943	757	Illumina OmniExpress Exome	HRC
YFS	419	198	221	30-47	38.88 (5.07)	172.25 (8.90)	3.73 (0.75)	4.68 (0.99)	0.8 (0.06)	--	233	186	Illumina 670k custom	1000G

Supplementary Table 3: SpiroMeta analysis method

Study name	**Individual call rate filter (applied before imp'n)**	**SNP call rate filter (applied before imp'n)**	**SNP HWE *P* filter (applied before imp'n)**	**SNP MAF filter (applied before imp'n)**	**Other filters**	**No of SNPs after filtering (before imp'n)**	**Imputation software and version**	**Reference panel used for imp'n**	**Genotype-phenotype association software**
B58C	None	>=95%	≥0.0001 (tested on females only for chromosome X)	≥1%	Consistent allele frequencies across data deposits (*P*≥0.0001 for pairwise comparisons) and for chrX SNPs, consistent allele frequencies between males and females (*P*≥0.0001).	500,521 (including 11,696 chrX)	MACH 1.0.18 & Minimac 2012-11-16	1000 Genomes Phase 1 March 2012	probABEL 0.1-9e
BHS1&2	0.95	0.95	1.00E-06	0.01	Individuals were removed if they had sex inconsistencies, had heterozygosity >5 s.d. from the mean, were PCA outliers, were 1 individual from a pair of duplicates or had IBD inconsistencies.	521,307	Minimac and MACH1 v1.0.18	b37; 1000 Genomes Phase 1 March 2012	ProbABEL
CROATIA-Korcula	97%	98%	1.00E-06	0.01		316,879	SHAPEIT2, IMPUTE2	b37; ALL (1000 Genomes Phase 1 integrated release v3, April 2012)	ProbABEL
CROATIA-Split	97%	98%	1.00E-06	0.01		321,727	SHAPEIT2, IMPUTE2	b37; ALL (1000 Genomes Phase 1 integrated release v3, April 2012)	ProbABEL
CROATIA-Vis	97%	98%	1.00E-06	0.01		273,671	SHAPEIT2, IMPUTE2	b37; ALL (1000 Genomes Phase 1 integrated release v3, April 2012)	ProbABEL
EPIC population based	None	95%	1.00E-08	Per-plate basis	Monomorphic SNPs; chr 23-26; INDELS; monomorphic; call rate<95%; chr-pos-allels duplicates; delta-AF > 0.2; delta_AF>0.1 if MAF<0.01. Oxford QC:) exclude SNPs if not in HRC ref (no INDEL in HRC ref); 2) exclude if don't match on chr-pos-allele; 3) strand check and flip; 4) exclude if delta-AF>0.2; 5) exclude A/T and G/C SNPs with MAF>0.4 in ref; 6) exclude if chr-pos duplicates	708,715	SHAPEIT v2.r790, Oxford	HRC v1.0, 1000 Genomes p3	BOLT-LMM v2.2
GS:SFHS	97%	98%	1.00E-06	0.01	Genetic ancestry outliers; monomorphic SNPs; high heterozygosity	602,451	SHAPEIT2 v2.r837, Sanger	HRC panel v1.1, European	REGSCAN

Study name	Individual call rate filter (applied before imp'n)	SNP call rate filter (applied before imp'n)	SNP HWE *P* filter (applied before imp'n)	SNP MAF filter (applied before imp'n)	Other filters	No of SNPs after filtering (before imp'n)	Imputation software and version	Reference panel used for imp'n	Genotype-phenotype association software
H2000	0.95	0.95 (0.99 for SNPs with MAF < 0.05)	1.00E-06	0.01		553,722	IMPUTE version 2.2.2	1,000 Genomes haplotypes -- Phase I integrated variant set release (v3) in NCBI build 37 (hg19) coordinates	SNPTest
KORA F4	0.97	0.98	5x10-6	0.01	-mismatch of phenotypic and genetic gender - 5s.d. from mean heterozygosity rate - check for European ancestry - check for population outlier	523,260 (chr 1-26) 508,532 (chr 1-22) 14,096(chrX-nonPAR) 444(chrX-PAR1) 58(chrX-PAR2)	SHAPEIT v2, IMPUTE v2.3.0	1000g phase1 all (ALL_1000G_phase1integrated_v3_impute_mac1)	SNPTEST v2.4.1
KORA S3	0.97	0.98	5x10-6	0.01	person wise: -mismatch of phenotypic and genetic gender - 5s.d. from mean heterozygosity rate - check for European ancestry - check for population outlier SNP wise: only SNPs that were genotyped with good quality on both chips	600641 (chr 1-26) 588307 (chr 1-22) 14625 (chrX-nonPAR)	SHAPEIT v2, IMPUTE v2.3.0	1000g phase1 all (ALL_1000G_phase1integrated_v3_impute_mac1)	SNPTEST v2.4.1
LBC1936	0.95	0.98	≥0.001	0.01		549,692	minimac 2012-11-16	1000 Genomes version 3, cosmopolitan	mach2qtl
NFBC1966	0.95	0.95	1.00E-04	0.01	Genetic ancestry outliers; monomorphic SNPs; high heterozygosity; Gender mismatch; 0 genetic sex; high heterozygosity; high relatedness	364,535	Eagle v2.3, Michigan	HRC r1.1 2016, European	rvtests
NFBC1986	0.99	0.99	1.00E-04	0.01	Genetic ancestry outliers; monomorphic SNPs; high heterozygosity; Gender mismatch; high heterozygosity; high relatedness	889,119	Eagle v2.3, Michigan	HRC r1.1 2016, European	rvtests
NSPHS	0.9	0.95	3.2E-08 (Infinum) & 1.4E-08 (OmniExpress)	0.01	FDR level of heterozygosity 0.01	306,086 (Infinum) & 631503 (OmniExpress)	Impute2 (v 2.2.2)	hg19, 1000 Genomes	ProbABEL
ORCADES	98%	97%	1.00E-06	1% (Hap300) & monomorphic (Omni & OmniX)	Subject Heterozygosity FDR<1%	287,208 (Hap300), 843723 (Omni) & 654651 (OmniX)	shapeit.v2.r644.+impute_v2.2.2_x86_64_static/impute2	1000G Phase I Integrated Release Version 3 Haplotypes (2010-11 data freeze, 2012-03-14 haplotypes).	probABEL v. 0.4.3

Study name	Individual call rate filter (applied before imp'n)	SNP call rate filter (applied before imp'n)	SNP HWE *P* filter (applied before imp'n)	SNP MAF filter (applied before imp'n)	Other filters	No of SNPs after filtering (before imp'n)	Imputation software and version	Reference panel used for imp'n	Genotype-phenotype association software
PIVUS	0.95	0.95 (0.99 if MAF<0.05)	1.00E-06	0.01	Genetic ancestry outliers; monomorphic SNPs; >3SD from mean for heterozygosity, pi-hat>0.125, gender discordance	738,583	SHAPEITv2, Oxford	HRC v1.1, all	SNPTEST v2.5
SAPALDIA	95%	95%	1.00E-06	0.01	none	545,131	Mach 1.0.16.a, minimac-omp RELEASE STAMP 2012-05-29 (autosomes) & MiniMac RELEASE STAMP 2012-11-16 (chr X)	build37, 1000 Genomes	probABEL
SHIP	0.92	0.95	1.00E-04	None	Genetic ancestry outliers; gender mismatch; pi-hat>0.25; monomorphic SNPS	760,787	Eagle v2.3, Michigan	HRC v1.1 reference, European	Rvtests
SHIP-TREND	0.94	0.95	1.00E-04	None	Genetic ancestry outliers; gender mismatch; pi-hat>0.25; monomorphic SNPS	1,691,610	Eagle v2.3, Michigan	HRC v1.1 reference, European	Rvtests
UKHLS	0.98	0.98	1.00E-04	None	Genetic ancestry, monomorphic SNPs, heterozygosity 3sd <>mean -visualised at 2 different MAF bins (≥1% and <1%); PI_HAT 0.2; Cluster separation score <0.4; sex check, ethnicity duplicates, withdrawn consent. Pre-imputation variants excluded that were: monomorphic, indels, differed to HRC in terms of strand, alleles, allele frequency (>0.2), A/T & G/C SNPs if MAF >0.4 and not in reference panel.	357,230	Autosomes: Eagle v2.3; ChrX: Shapeit v2.r790, Michigan	Autosomes: HRC r1.1 2016; ChrX: HRC r1.1 2017, European	SNPTEST v2.5
VIKING	0.97	0.98	1.00E-06	MAF>0.01 for OMNI markers; MAF>0.0001 for Exome Chip markers	Genetic ancestry outliers; monomorphic SNPs; Duplicates and siblings	668,762	shapeit2r837 + duohmm; PBWT Sanger	HRC v1.1, European	REGSCAN 0.4
YFS	0.95	0.95	1.00E-06	0.01	heterozygosity, relatedness	546,674	SHAPEIT v1 and IMPUTE v2.2.2	1000 Genomes Phase 1, release v3, March 2012 haplotypes	SNPTEST v.2.4.1

Supplementary Table 4: 139 novel signals

See Excel spreadsheet.

139 independent (r^2<0.1) novel signals of association with lung function (99 tier 1, 40 tier 2): tier 1 signals meet the criteria $P<5\times10^{-9}$ in UK Biobank and $P<10^{-3}$ in SpiroMeta; tier 2 signals meet the criteria $P<5\times10^{-9}$ in the meta-analysis and $P<10^{-3}$ in both UK Biobank and SpiroMeta. UK Biobank p values have genomic control applied using the LD score regression intercept as the inflation factor (**Supplementary Table 24**). No genomic control was applied to SpiroMeta and the meta-analysis as there was no significant inflation after LD score regression. The allele frequencies and individual variant sample sizes for SpiroMeta were calculated based on a working total sample size of 83,118. For two secondary signals (rs10874851 and rs4796334) the association results are from a conditional analysis (no genomic control) where the primary signal is shown in the "conditioned on" column. Direction of effect is consistent for all signals.

Supplementary Table 5: Tier 3 signals

See Excel spreadsheet.

Signals that reached $P<5\times10^{-9}$ in UK Biobank or the meta-analysis of UK Biobank and SpiroMeta, with consistent directions of effect but did not meet $P<10^{-3}$ in SpiroMeta required to qualify as a Tier 2 signal.

Supplementary Table 6: Association with smoking behaviour

See Excel spreadsheet.

Look up of smoking behaviour for 139 novel signals and 142 previously reported signals associated with lung function in this study. Also show are: lung function association results from the UK Biobank and SpiroMeta meta-analysis. Bold P value for Smoking initiation (SI) or Cigarettes per day (CPD) indicates association with smoking behaviour $P<1.8\times10^{-4}$ (Bonferroni threshold for 281 tests).

Supplementary Table 7: Smoking interaction

See Excel spreadsheet.

Results from stratified analyses of ever / never smokers in UK Biobank, SpiroMeta and a fixed-effects meta-analysis of the two. Evidence of interaction between ever and never smokers was sought by conducting a Welch test on the results of the stratified meta-analysis. A Bonferroni threshold of $1.79\text{x}10^{-4}$ (p=0.05/279 tests) was used.

Supplementary Table 8: Previously reported signals

See Excel spreadsheet.

185 signals previously reported for lung function or COPD. For inclusion with our 139 novel signals in downstream analyses we first removed 24 non-independent (r^2>0.1) signals from a recent GWAS of lung function. We also removed 3/6 HLA signals that were not independent, as established in one of our previous publications[45]. We then selected a subset of 142 signals that showed evidence of association in this study (UK Biobank P in bold): $P<5\times10^{-5}$ in 321,047 UK Biobank samples for any lung function phenotype, either for the reported sentinel or a proxy with r^2>0.5, or if one of our Tier 1 or 2 signals is in LD r^2>0.1 with the previously reported sentinel. In downstream analyses, we excluded two signals (15q25 and *CYP2A6*, which have previously been reported to be associated with smoking behaviour, but are not associated with lung function in never smokers in the present study. This left 140 previously reported signals for inclusion in our final set of signals for downstream analyses Where PubMed ID (PMID) is missing, the variants are currently reported on bioRxiv[46].

Supplementary Table 9: Results for 279 lung function signals for all 4 traits

See Excel spreadsheet.

Results from the meta-analysis of UK Biobank and SpiroMeta for all 279 reported lung function signals for each of the lung function quantitative traits FEV_1, FVC, FEV_1/FVC and PEF.

Supplementary Table 10: Bayesian 99% credible sets

See Excel spreadsheet.

Bayesian 99% credible sets calculated using Wakefield's method[50] for 276 signals: 139 novel and 137 of 140 previously reported showing significant association in this study (3 HLA signals excluded; **Supplementary Table 8**). Effect sizes and standard errors for the credible set calculation are from the meta-analysis of UK Biobank and SpiroMeta; variants with $r^2>0.4$ with the sentinel and $P<10^{-4}$ are included in the calculation the prior probability parameter W is 0.04. For previously reported signals we used the sentinel variant from the meta-analysis of UK Biobank and SpiroMeta in this study. 182 signals have the sentinel with the (joint) highest posterior probability (109 novel, 73 previous); 20 signals have only the sentinel in the credible set (12 novel, 8 previous); 8 signals do not contain the sentinel in the credible set. Individual regions for all of these 276 signals are available to download as a separate file.

Supplementary Table 11: Functional annotation of coding variants in the 99% credible sets

See Excel spreadsheet.

Variants that entered the functional annotation were those annotated as "exonic", "splicing", "ncRNA_exonic", "5' UTR" or "3' UTR" (untranslated region) by ANNOVAR. Annotation software used: SIFT, PolyPhen-2 and FATHMM all annotate missense variants, and CADD annotates non-coding variation. Variant annotated as deleterious (1) versus not (0) if the variant was labelled 'deleterious' by SIFT, 'probably damaging' or 'possibly damaging' by PolyPhen-2, if it had a CADD scaled score ≥20, or was annotated as "damaging" by FATHMM. See also **Online Methods**. Column explanations: All=harmful according to at least one of CADD, SIFT, PolyPhen-2, FATHMM; Post Prob=posterior probability for sentinel variant; Highest PP SNP(s)=SNP(s) with highest posterior probabilities for a credible set; Highest PP=value of highest posterior probability for top SNP for a credible set; Highest Flag=annotated SNP is also top SNP for credible set

Supplementary Table 12: Z scores and P values for eQTL look up in lung tissue resources

Variants in the 99% credible sets that are associated with gene expression at FDR<5% in an eQTL resource (n=1,111) of lung tissue from Laval University[51], Canada, Groningen University[52], Netherlands and University of British Columbia (UBC)[53], Canada. The sentinel SNP out of our 279 lung function associated SNPs is given, the SNP most highly associated with expression in the 99% credible set of the lung function sentinel, the posterior probability of this SNP within the credible set, the gene expression Z score and P value and the eQTL SNP most highly associated with gene expression for that gene (eQTL sentinel).

This table includes the eQTL data for all genes where there was a variant in the credible set with FDR<5% for association with expression. Only genes where the eQTL sentinel is in the credible set were added to our list of putative causal genes for downstream analysis.

Supplementary Table 13: Genes implicated by eQTL or pQTL associations or deleterious variants

(-): COPD risk allele decreases gene expression or protein level. (+): COPD risk allele increases gene expression or protein level. *Nine GTEx tissues were screened (n up to 388): Artery Aorta (n=267), Artery Coronary (n=152), Artery Tibial (n=388), Colon Sigmoid (n=203), Colon Transverse (n=246), Esophagus Gastroesophageal Junction (n=213), Esophagus Muscularis (n=335), Small Intestine Terminal Ileum (n=122), and Stomach (n=237) – note direction of gene expression not provided for the genes implicated by these tissues as >1 tissue is screened. 88 genes were implicated where the eQTL sentinel was in our lung function 99% credible set for 58 sentinel SNPs; 5 genes were implicated where the pQTL sentinel was in our lung function 99% credible set for 5 SNPs; 21 genes were implicated by a coding deleterious variant in the 99% credible set for 20 sentinels, giving a union across all 3 look ups of 107 unique putative causal genes. Z scores and P values for the Lung eQTL look up are in **Supplementary Table 12**.

Novelty	Nearest_Gene	Phenotype	Tier	Sentinel SNP	Chr	Pos	COPD risk allele	Alt allele	Lung eQTL	NESDA-NTR Blood eQTL	GTEx Lung	GTEx Whole Blood	*Genes that appear >1 in smooth muscle containing GTEx tissues	pQTL plasma proteins	Genes that contain a coding deleterious variant
Previous	*MFAP2*	FEV_1/FVC	Previous	rs9435733	1	17308254	C	T			*MFAP2*(+)				
Novel	*DHDDS*	FVC	2	rs9438626	1	26775367	G	C				*DHDDS*(+), *DRAM2*(-)	*DHDDS*		
Novel	*DHDDS*	FEV1	1	rs12096239	1	26796922	C	G	*HMGN2*(-)				*DHDDS*		
Previous	*LOC101929516*	FEV1/FVC	Previous	rs755249	1	39995074	T	C			*PABPC4*(-)		*PABPC4*		
Novel	*NEXN*	FEV1/FVC	1	rs9661687	1	78387270	T	C	*NEXN*(+)						
Novel	*DENND2D*	FEV1/FVC	1	rs9970286	1	111737398	G	A		*CEPT1*(-)	*CHI3L2*(-)	*CEPT1*(-)			
Novel	*C1orf54*	PEF	1	rs11205354	1	150249101	C	A	*MRPS21*(+)	*RPRD2*(-)	*RPRD2*(-)			*ECM1*(+)	
Novel	*KRTCAP2*	FEV1/FVC	1	rs141942982	1	155153537	T	G						*THBS4*(+)	
Novel	*RALGPS2*	FEV1	1	rs4651005	1	178719306	C	T			*ANGPTL1*(-)				
Novel	*LMOD1*	FEV1/FVC	2	rs4309038	1	201884647	G	C			*SHISA4*(-)				
Previous	*TGFB2*	PEF	Previous	rs6604614	1	218631452	C	G	*TGFB2*(+)						
Novel	*ATAD2B*	FVC	2	rs13009582	2	24018480	G	A			*UBXN2A*(+)	*UBXN2A*(+)	*UBXN2A*		
Novel	*PKDCC*	FVC	1	rs4952564	2	42243850	A	G			*PKDCC*(+)				
Novel	*ITGAV*	FEV1/FVC	1	rs2084448	2	187530520	C	T		*ITGAV*(+)					
Novel	*SPATS2L*	FEV1/FVC	2	rs985256	2	201208692	C	A					*SPATS2L*		
Previous	*TRAF3IP1*	FEV1	Previous	rs6710301	2	239441308	C	A							*ASB1*
Novel	*C2orf54*	FVC	1	rs6437219	2	241844033	C	T	*C2orf54*(-)		*C2orf54*(-)		*C2orf54*		*C2orf54*
Previous	*SLMAP*	FEV1	Previous	rs6445932	3	57879611	T	G					*SLMAP*		
Novel	*MIR548G*	FVC	1	rs1610265	3	99420192	T	C				*FILIP1L*(-)			
Previous	*RSRC1*	FVC	Previous	rs12634907	3	158226886	G	A			*RSRC1*(-)				
Novel	*BCHE*	FEV1/FVC	1	rs1799807	3	165548529	C	T							*BCHE*
Novel	*BTC*	FEV1/FVC	1	rs62316310	4	75676529	G	A							*BTC*
Previous	*GSTCD*	FEV1	Previous	rs11722225	4	106766430	T	C			*INTS12*(+)		*INTS12*		
Previous	*NPNT*	FEV1/FVC	Previous	rs34712979	4	106819053	A	G					*NPNT*	*NPNT*(-)	
Novel	*LOC100996325*	FEV1	1	rs11739847	5	609661	A	G							*CEP72*
Previous	*AP3B1*	FVC	Previous	rs425102	5	77396400	G	T			*AP3B1*(+)				
Previous	*SPATA9*	FEV1/FVC	Previous	rs987068	5	95025146	C	G	*RHOBTB3*(-)						
Previous	*P4HA2-AS1*	FVC	Previous	rs3843503	5	131466629	A	T	*SLC22A5*(+)	*SLC22A5*(-)			*P4HA2*	*C1QTNF5*(-)	
Previous	*CYFIP2*	FEV1/FVC	Previous	rs11134766	5	156908317	T	C				*ADAM19*(+)			
Previous	*ADAM19*	FEV1/FVC	Previous	rs11134789	5	156944199	A	C			*ADAM19*(+)		*ADAM19*		*ADAM19*
Previous	*DSP*	FEV1/FVC	Previous	rs2076295	6	7563232	T	G	*DSP*(+)		*DSP*(+)				
Novel	*RNU6-71P*	FEV1	1	rs2894837	6	56336406	G	A							*DST*
Previous	*MIR588*	FVC	Previous	rs6918725	6	126990392	T	G					*CENPW*		

Novelty	Nearest_Gene	Phenotype	Tier	Sentinel SNP	Chr	Pos	COPD risk allele	Alt allele	Lung eQTL	NESDA-NTR Blood eQTL	GTEx Lung	GTEx Whole Blood	*Genes that appear >1 in smooth muscle containing GTEx tissues	pQTL plasma proteins	Genes that contain a coding deleterious variant
Previous	*GPR126*	FEV1/FVC	Previous	rs17280293	6	142688969	A	G							*GPR126*
Previous	*C1GALT1*	FEV1/FVC	Previous	rs4318980	7	7256490	A	G					*C1GALT1*		
Novel	*JAZF1*	FEV1	1	rs1513272	7	28200097	C	T					*JAZF1*		
Novel	*MET*	FEV1/FVC	2	rs193686	7	116431427	T	C					*MET*		
Novel	*IER5L*	FEV1	2	rs967497	9	131943843	G	A	*CRAT*(-)	*CRAT*(+), *PPP2R4*(+)			*CRAT*		*IER5L*
Previous	*QSOX2*	FVC	Previous	rs7024579	9	139100413	T	C	*QSOX2*(+)						
Previous	*DNLZ*	FVC	Previous	rs4073153	9	139259349	G	A	*SNAPC4*(-)	*CARD9*(+)	*CARD9*(+), *INPP5E*(-)	*CARD9*(+)			
Previous	*CDC123*	FEV1/FVC	Previous	rs7090277	10	12278021	T	A				*NUDT5*(+)			
Previous	*MYPN*	FVC	Previous	rs10998018	10	69962954	A	G							*MYPN*
Previous	*EML3*	FEV1	Previous	rs71490394	11	62370155	G	A			*EEF1G*(+), *ROM1*(+), *EML3*(-)	*EML3*(-)	*EEF1G*, *EML3*, *ROM1*		*EML3*, *ROM1*
Previous	*ARHGEF17*	FEV1/FVC	Previous	rs2027761	11	73036179	C	T		*FAM168A*(-)		*ARHGEF17*(-)			*ARHGEF17*
Previous	*RAB5B*	FEV1	Previous	rs1689510	12	56396768	C	G	*CDK2*(-)						
Previous	*LRP1*	FEV1/FVC	Previous	rs11172113	12	57527283	T	C					*LRP1*		
Previous	*FGD6*	FEV1/FVC	Previous	rs113745635	12	95554771	T	C	*FGD6*(+)						
Novel	*DOCK9*	FEV1/FVC	1	rs11620380	13	99665512	A	C							*DOCK9*
Novel	*CHAC1*	FVC	1	rs4924525	15	41255396	A	C		*INO80*(+)	*CHP1*(+)	*RAD51*(-)			
Previous	*RPAP1*	FEV1/FVC	Previous	rs2012453	15	41840238	G	A	*TYRO3*(+), ITPKA(+), LTK(+)		*ITPKA*(-), *LTK*(-), *TYRO3*(-)	*RPAP1*(+)			
Previous	*AAGAB*	FVC	Previous	rs12917612	15	67491274	A	C		*AAGAB*(-), *SMAD3*(+)			*IQCH*		
Previous	*THSD4*	FEV1/FVC	Previous	rs1441358	15	71612514	G	T	*THSD4*(+)						
Previous	*IL27*	FEV1	Previous	rs12446589	16	28870962	A	G	*SBK1*(+), *TUFM*(-)	*CCDC101*(-)	*SULT1A1*(-), *TUFM*(+), *SH2B1*(-)	*NPIPB7*(-), *CLN3*(+), *SULT1A2*(+), *ATXN2L*(-), *TUFM*(+)	*CCDC101*, *EIF3C*, *SH2B1*, *SULT1A1*, *SULT1A2*, *TUFM*		*SULT1A2*
Previous	*MMP15*	FEV1/FVC	Previous	rs11648508	16	58063513	G	T	*MMP15*(+)		*MMP15*(-)				
Novel	*ATP2A3*	FEV1/FVC	1	rs8082036	17	3882613	G	C	*ATP2A3*(-)		*ATP2A3*(-)				
Novel	*PITPNM3*	FEV1	2	rs4796334	17	6469793	A	G		*KIAA0753*(-), *TXNDC17*(-)	*KIAA0753*(-)	*PITPNM3*(+), *KIAA0753*(-)	*KIAA0753*, *PITPNM3*		*KIAA0753*
Novel	*TNFSF12-TNFSF13*	FEV1	2	rs4968200	17	7448457	C	G	*SENP3*(-)			*TNFSF13*(-)			
Novel	*NCOR1*	FVC	2	rs34351630	17	16030520	C	T				*ADORA2B*(-), *TTC19*(+)	*ADORA2B*, *TTC19*		
Previous	*SSH2*	FEV1/FVC	Previous	rs2244592	17	28072327	A	G					*EFCAB5*		
Previous	*FBXL20*	FVC	Previous	rs8069451	17	37504933	C	T	*CRKRS*(-)	*FBXL20*(-)					
Previous	*MAPT-AS1*	FEV1	Previous	rs79412431	17	43940021	A	G	*LRRC37A4*(+)						*MAPT*
Previous	*TSEN54*	FEV1	Previous	rs9892893	17	73525670	G	T				*CASKIN2*(-)			*TSEN54*
Novel	*ASPSCR1*	FVC	1	rs59606152	17	79952944	C	T							*LRRC45*
Novel	*C18orf8*	FVC	1	rs303752	18	21074255	A	G					*C18ORF8*		

Novelty	**Nearest_Gene**	**Phenotype**	**Tier**	**Sentinel SNP**	**Chr**	**Pos**	**COPD risk allele**	**Alt allele**	**Lung eQTL**	**NESDA-NTR Blood eQTL**	**GTEx Lung**	**GTEx Whole Blood**	***Genes that appear >1 in smooth muscle containing GTEx tissues**	**pQTL plasma proteins**	**Genes that contain a coding deleterious variant**
Novel	*ZFP82*	FVC	2	rs2967516	19	36881643	A	G	*ZFP14*(-)	*ZFP82*(+)		*ZFP82*(+)			
Previous	*LTBP4*	FEV1/FVC	Previous	rs34093919	19	41117300	G	A							*LTBP4*
Previous	*ABHD12*	FEV1	Previous	rs2236180	20	25282608	C	T			*PYGB*(+)		*PYGB*		*PYGB*
Previous	*UQCC1*	FVC	Previous	rs143384	20	34025756	G	A	*UQCC1*(-)		*UQCC1*(-), *GDF5*(-)	*UQCC1*(-)	*UQCC1*		
Previous	*SLC2A4RG*	FVC	Previous	rs4809221	20	62372706	A	G				*LIME1*(-)	*LIME1*		
Previous	*SCARF2*	FEV1	Previous	rs9610955	22	20790723	C	G						*SCARF2*(+)	*SCARF2*

Footnote:

Genes implicated by a new signal: ADORA2B, ANGPTL1, ATP2A3, BCHE, BTC, C18ORF8, C2orf54, CEP72, CEPT1, CHI3L2, CHP1, CRAT, DHDDS, DRAM2, DST, ECM1, FILIP1L, HMGN2, IER5L, INO80, ITGAV, JAZF1, KIAA0753, LRRC45, MET, MRPS21, NEXN, PITPNM3, PKDCC, PPP2R4, RAD51, RPRD2, SENP3, SHISA4, SPATS2L, THBS4, TNFSF13, TTC19, TXNDC17, UBXN2A, ZFP14, ZFP82.

Genes implicated by a previously reported signal that were not previously implicated[45]: AAGAB, AP3B1, ARHGEF17, ATXN2L, C1QTNF5, CCDC101, CDK2, CENPW, CLN3, CRKRS, DSP, EEF1G, EIF3C, FAM168A, FBXL20, GDF5, IQCH, ITPKA, LTK, NPIPB7, PYGB, RPAP1, SBK1, SCARF2, SH2B1, SLMAP, SMAD3, SULT1A1, SULT1A2, TUFM, TYRO3, UQCC1

Supplementary Table 14: Proteins implicated by pQTL analysis

Lung function sentinel SNPs where one of the SNPs in the 99% credible set is the most highly associated SNP for a protein (top pSNP) in Sun *et al.* protein expression dataset[54] and the association with protein levels is $P<5.03 \times 10^{-8}$ (5% Bonferroni-adjusted threshold for 276 independent sentinel SNPs x 3,600 plasma protein levels tested).

Novelty	Nearest Gene	Trait	Sentinel SNP ID	Sentinel SNP chrom: Position (b37)	Sentinel SNP Coded/ Noncoded	Top pSNP	Top pSNP chrom: Position (b37)	Top pSNP Noncoded	Top pSNP Coded	Top pSNP Beta (SE)	P-value	Protein
Novel (Tier 1)	*C1orf54*	PEF	rs11205354	1:150,476,516	A/G	rs11205385	1:150,476,516	G	A	-0.3153 (0.0245)	8.91E-38	ECM1
Novel (Tier 1)	*KRTCAP2*	FEV_1/FVC	rs141942982	1:155,153,537	T/C	rs111508230	1:155,153,537	C	T	0.2170 (0.0388)	2.19E-08	THBS4
Previous	*NPNT*	FEV_1	rs34712979	4:106,819,053	A/G	rs34712979	4:106,819,053	G	A	-0.2274 (0.0280)	4.57E-16	NPNT
Previous	*P4HA2-AS1*	FVC	rs3843503	5:131,567,924	A/G	rs11955347	5:131,567,924	G	A	-0.2464 (0.0245)	9.12E-24	C1QTNF5
Previous	*SCARF2*	FEV_1/FVC	rs9610955	22:20,775,556	T/G	rs738086	22:20,775,556	G	T	0.2997 (0.0318)	4.79E-21	SCARF2

Supplementary Table 15: Pathway analysis

Gene-based pathway enrichment analyses. Summary of gene-sets overrepresented in known biological pathways and gene ontology (GO) terms. 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. FDR: False discovery rate. Novel genes from novel and previous signals are marked with a dagger (†) and double dagger (‡), respectively.

Genes that contain an eQTL that is in our 99% credible sets (thirteen tissues/datasets) and/or 'deleterious' coding variant (n=104 genes)				
Enriched biological pathways				
P value	**FDR**	**pathway**	**Genes associated with biological pathway**	**Total size of pathway gene-set**
4.26E-07	9.33E-05	Molecules associated with elastic fibres	*ITGAV†; TGFB2; LTBP4; MFAP2; GDF5‡*	30
9.51E-07	0.000104	Elastic fibre formation	*ITGAV†; TGFB2; LTBP4; MFAP2; GDF5‡*	35
3.51E-05	0.00241	Extracellular matrix organization	*MMP15; TGFB2; LTBP4; DST†; ITGAV†; P4HA2; MFAP2; GDF5‡; ADAM19*	294
0.000158	0.00812	Malaria - Homo sapiens (human)	*MET†; TGFB2; LRP1; THBS4†*	49
0.000497	0.017	Extracellular vesicle-mediated signaling in recipient cells	*MET†; TGFB2; SMAD3‡*	30
0.00059	0.018468	Alpha6Beta4Integrin	*MET†; DST†; DSP‡; SMAD3‡*	74
0.001345	0.036822	TGF-Core	*TGFB2; GDF5†; SMAD3‡*	42
Enriched gene ontology terms				
P value	**FDR**	**Name of GO term (GO term category/level)**	***Genes associated with GO term***	**Total size of pathway gene-set**
2.39E-05	0.007332	cytoskeleton organization (b/4)	*TGFB2; LRP1; CEP72†; ARHGEF17‡; DST†; CHP1†; INO80†; DSP‡; SMAD3‡; ITPKA‡; PTPA†; CDK2‡; FGD6; MYPN; NEXN†; MAPT; KIAA0753†*	1075
0.000271	0.034539	regulation of cartilage development (b/5)	*TGFB2; PKDCC†; GDF5‡; SMAD3‡*	62
0.000319	0.025007	ammonium ion metabolic process (b/3)	*CRAT†; TGFB2; CEPT1†; SLC22A5; CLN3‡; BCHE†*	182
0.00036	0.025007	organelle organization (b/3)	*TGFB2; CHP1†; INO80†; BTC†; SMAD3‡; AP3B1‡; GDF5‡; ITPKA‡; CLN3‡; RAD51†; UBXN2A†; MYPN; NEXN†; FGD6; CEP72†; ARHGEF17; DST†; DSP‡; SH2B1‡; CENPW‡; KIAA0753†; ATXN2L‡; LRP1; UQCC1‡; MRPS21†; PTPA†; CDK2‡; MAPT; TTC19†; TUFM‡*	3207
0.000371	0.025007	centriole replication (b/3)	*CDK2‡; KIAA0753†; CEP72†*	28
0.000405	0.004661	protein kinase activity (m/5)	*TGFB2; LTBP4; CDK12‡; PKDCC†; ITPKA‡; MET†; CDK2‡; BTC†; LTK‡; SBK1‡; TYRO3‡*	646

0.000413	0.034539	positive regulation of cartilage development (b/5)	*PKDCC*†; *GDF5*‡; *SMAD3*‡	29
0.000691	0.034539	transforming growth factor beta2 production (b/5)	*TGFB2*; *SMAD3*‡	8
0.000691	0.034539	mitochondrial respiratory chain complex III assembly (b/5)	*UQCC1*‡; *TTC19*†	8
0.000786	0.022497	transmembrane receptor protein kinase activity (m/4)	*LTK*‡; *MET*†; *LTBP4*; *TYRO3*‡	82
0.000939	0.037577	positive regulation of ossification (b/5)	*NPNT*; *TGFB2*; *PKDCC*†; *SMAD3*‡	86
0.001016	0.045883	microtubule-based process (b/3)	*AP3B1*‡; *CEP72*†; *DST*†; *CHP1*†; *INO80*†; *PTPA*†; *CDK2*‡; *CLN3*‡; *MAPT*; *KIAA0753*†	611
0.001136	0.045883	extracellular structure organization (b/3)	*MMP15*; *TGFB2*; *THSD4*; *ITGAV*†; *SMAD3*‡; *NPNT*; *MFAP2*	318
0.001481	0.049872	phosphorus metabolic process (b/3)	*DHDDS*†; *CHP1*†; *TGFB2*; *BTC*†; *SMAD3*‡; *PKDCC*†; *ADORA2B*†; *MET*†; *CDK12*‡; *GDF5*‡; *ITPKA*‡; *CARD9*; *CLN3*‡; *RAD51*†; *SULT1A1*‡; *NUDT5*; *SULT1A2*‡; *LTK*‡; *SBK1*‡; *TYRO3*‡; *INPP5E*; *RSRC1*; *CEPT1*†; *ITGAV*†; *NPNT*; *PTPA*†; *CDK2*‡; *PITPNM3*†	3164
0.00149	0.022497	phosphatase binding (m/4)	*AP3B1*‡; *MET*†; *PTPA*†; *MAPT*; *SMAD3*‡	165
0.001512	0.04837	regulation of chondrocyte differentiation (b/5)	*PKDCC*†; *GDF5*‡; *SMAD3*‡	45
0.001824	0.022497	phosphotransferase activity, alcohol group as acceptor (m/4)	*TGFB2*; *LTBP4*; *CDK12*‡; *PKDCC*†; *ITPKA*‡; *MET*†; *CDK2*‡; *BTC*†; *LTK*‡; *SBK1*‡; *TYRO3*‡	777
0.001846	0.04837	regulation of neuron death (b/5)	*TGFB2*; *LRP1*; *CHP1*†; *GDF5*‡; *CLN3*‡; *TYRO3*‡	255
0.001935	0.04837	catechol-containing compound metabolic process (b/5)	*TGFB2*; *SULT1A1*‡; *SULT1A2*‡	49
0.002171	0.012484	transforming growth factor beta receptor binding (m/5)	*TGFB2*; *GDF5*‡; *SMAD3*‡	51
0.002884	0.026674	transforming growth factor beta binding (m/4)	*LTBP4*; *ITGAV*†	16
0.003486	0.016037	protein phosphatase binding (m/5)	*AP3B1*‡; *MET*†; *MAPT*; *PTPA*†	123
0.004122	0.030499	kinase activity (m/4)	*TGFB2*; *LTBP4*; *CDK12*‡; *PKDCC*†; *ITPKA*‡; *MET*†; *CDK2*‡; *BTC*†; *LTK*‡; *SBK1*‡; *TYRO3*‡	864
0.004328	0.016592	transmembrane receptor protein tyrosine kinase activity (m/5)	*LTK*‡; *MET*†; *TYRO3*‡	65

Supplementary Table 16: DeepSEA prediction of functional effect

See Excel spreadsheet.

DeepSEA predictions for the SNPs in the 99% credible sets (total n=9446 SNPs) in lung-related cell lines from the RoadMap Epigenome and ENCODE projects. We queried four lung-related cell lines (foetal lung, foetal lung fibroblasts [IMR90], human lung fibroblasts [NHLF] and adenocarcinomic human alveolar basal epithelial cells [A549]) for which 55 chromatin features and transcription factor binding sites were measured. The absolute difference between reference and alternative allele is shown. Only the results for the 161 SNPs with a predicted functional effect (i.e. absolute difference ≥0.1) in ≥1 cell line are presented. SNPs which have the highest posterior probability in their respective credible sets are coloured in red. Non-significant results (i.e. absolute difference between reference and alternative allele <0.1) are replaced with a "-" for clarity. E-values (i.e. the expected proportion of SNPs with larger predicted effect for this chromatin feature based on empirical distributions of predicted effects for 1000 Genomes SNPs) for each result are presented in brackets. E-values<0.05 and <0.01 are highlighted in red and green, respectively.

Supplementary Table 17: Druggability analysis

See Excel spreadsheet. Please note that it is possible to filter this table using the drop-down arrows at the top of each column.

Table showing drugs interacting with either high-priority genes that were identified in eQTL or pQTL analysis or annotated as deleterious (N=107) (**Supplementary Table 13**).

The 107 genes were queried against gene-drug interactions within the Drug-Gene Interactions Database (DGIDB) (http://www.dgidb.org/data/). The 68 drugs (identified from CHEMBL interactions) that mapped to these genes were mapped to CHEMBL IDs and indications (as Medical Subject Headings, or 'MeSH' terms, https://www.ebi.ac.uk/chembl/drug/indications). For each gene, the sentinel SNP that implicated this gene is given. Drug names associated with each gene, plus CHEMBL IDs, and drug indications (with maximum development phase in brackets) are also shown.

In addition to the above analysis, we also undertook a STRING analysis of protein-protein interactions, using the 107 high priority genes described above as input. These results are described in **Supplementary Table 26**.

Column explanations:

- Drug=compound/drug name;
- CHEMBL_ID=compound identification number from CHEMBL;
- OriginalGeneAndSource=The name(s) of the gene (amongst the set of 107 high priority genes) interacting with the drug;
- IndicationPhase=Drug indication (Phase). 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.
- MAB=Drug is a monoclonal antibody;
- OriginalGenesPathway=the gene given in the 'Original Gene and Source' column is a gene identified in the 'Enriched Biological Pathways' shown in **Supplementary Table 15**;
- Cancer=the drug is used to treat some form of cancer;
- Phase3or4=the drug has at least one indication annotated as Phase 3 or 4;
- AsthmaCOPD=the drug is already indicated as being used in asthma or COPD;
- Novelty=the drug is implicated for use by genes identified from novel signals in this GWAS.

Supplementary Table 18: UK Biobank and China Kadoorie Biobank COPD and FEV_1/FVC weighted genetic risk score association results (per-allele and per standard deviation) by ancestry

Individuals in UKB Biobank and China Kadoorie Biobank were included for this analysis, and UK Biobank individuals were divided into ancestry groups as described in **Supplementary Figure 1**. The weighted genetic risk score was tested for association with COPD and FEV_1/FVC. COPD was defined as FEV_1/FVC<0.7 and FEV1<80% predicted (i.e. corresponding to GOLD 2-4 standards). The COPD model (a logistic regression, with COPD coded as COPD [1] vs. no COPD [0]) was adjusted as described in the **Online Methods**. The COPD model was only fitted in ancestral groups with >100 COPD cases. For the FEV_1/FVC model, linear regression was used. The phenotype was as prepared for the main GWAS described in this paper (see **Online Methods**).

Ancestry	per Allele			per Standard Deviation			P	Total N	N Control	N Case	Mean risk score	SD risk score
	Effect size* (OR/Beta)	95LCI	95UCI	Effect size* (OR/Beta)	95LCI	95UCI						
COPD												
UK Biobank African	1.033	1.015	1.050	1.348	1.152	1.577	1.92E-04	4225	4053	172	305.95	9.26
UK Biobank South Asian	1.030	1.020	1.041	1.414	1.254	1.594	1.42E-08	6358	6046	312	308.22	11.66
UK Biobank Chinese**								1607	1558	49	302.38	11.47
UK Biobank European***	1.030	1.029	1.032	1.436	1.411	1.461	<1e-300	303570	288467	15103	307.78	12.16
UK Biobank Mixed African & European								1208	1153	55	305.70	10.67
UK Biobank Mixed Other	1.035	1.024	1.046	1.506	1.325	1.712	3.65E-10	6033	5752	281	305.58	12.04
China Kadoorie Biobank**	1.017	1.014	1.019	1.219	1.183	1.256	3.58E-40	75580	69567	6013	298.40	11.85
FEV1/FVC												
UK Biobank African	-0.009	-0.013	-0.006	-0.086	-0.116	-0.056	2.12E-08	4225			305.95	9.26
UK Biobank South Asian	-0.015	-0.018	-0.013	-0.181	-0.205	-0.156	3.72E-47	6358			308.22	11.66
UK Biobank Chinese**	-0.012	-0.017	-0.008	-0.142	-0.191	-0.093	1.44E-08	1607			302.38	11.47
UK Biobank European***	-0.018	-0.019	-0.018	-0.224	-0.227	-0.221	<1e-300	303570			307.78	12.16
UK Biobank Mixed African & European	-0.016	-0.022	-0.011	-0.176	-0.231	-0.120	7.01E-10	1208			305.70	10.67
UK Biobank Mixed Other	-0.015	-0.018	-0.013	-0.186	-0.211	-0.162	7.00E-48	6033			305.58	12.04
China Kadoorie Biobank**	-0.007	-0.007	-0.006	-0.078	-0.085	-0.071	5.09E-99	72796			298.32	11.85

*Effect sizes are odds ratios for COPD results, and change in z-score units for FEV_1/FVC results

For details on missing SNPs in UK Biobank Chinese ancestry subjects, and China Kadoorie Biobank participants, see **Online Methods

***Europeans in UK Biobank were the discovery sample for many of the variants in the risk score, which explains the very low p-values in this subgroup.

Supplementary Table 19: Demographics of COPD case-control cohorts included in risk score included in risk score analysis

Descriptive statistics for each cohort are given separately for cases and controls, for five cohorts: the COPD Gene study, the ECLIPSE study (Evaluation of COPD Longitudinally to Identify Predictive Surrogate End-points), GenKOLS (the Bergen, Norway COPD cohort), NETT/NAS (the National Emphysema Treatment Trial [NETT] and the Normative Aging Study [NAS]) and the SPIROMICS study. Abbreviation: SD=standard deviation; age is given in years, height in centimetres, FEV1 and FVC litres.

Cohort	Case-control status	Total N	% female	Age range	Mean age (SD)	Height range	Mean height (SD)	N with spirometry data	Mean FEV1 (SD)	Mean FVC (SD)	Mean FEV1/FVC (SD)	% ever smokers (N with ever smoking data available)	Pack-years range (N with pack-years data available)	Mean pack-years (SD)
COPDGene (African-American Population)	Cases	910	44.84	45-81	58.6 (8.15)	137-208	170.96 (10.1)	910	1.39 (0.63)	0.534 (0.121)	2.546 (0.879)	100 (910)	10 - 162 (910)	42.69 (23.48)
	Controls	1556	40.94	45-80	52.84 (6.01)	142-203	171.15 (9.33)	1556	2.768 (0.644)	0.785 (0.05)	3.535 (0.839)	100 (1556)	10 - 160.4 (1556)	36.11 (19.1)
COPDGene (Non-Hispanic White Population)	Cases	3068	45.14	45-81	64.38 (8.28)	134-196	169.72 (9.45)	3068	1.424 (0.659)	2.817 (0.908)	0.495 (0.134)	100 (3068)	10 - 237.6 (3068)	54.89 (27.12)
	Controls	2110	51.47	45-81	59.18 (8.64)	140-198	169.54 (9.38)	2110	2.924 (0.679)	3.817 (0.892)	0.768 (0.044)	100 (2110)	10 - 172.5 (2110)	37.34 (20.14)
ECLIPSE	Cases	1713	32.87	40-75	63.64 (7.1)	142-201	169.54 (9.02)	1713	1.213 (0.487)	2.766 (0.873)	0.441 (0.111)	100 (1713)	6 - 220 (1713)	50.5 (27.47)
	Controls	147	42.86	40-74	57.32 (9.55)	151-196	171.24 (9.69)	147	3.164 (0.779)	4.085 (1.03)	0.778 (0.054)	100 (147)	10 - 230 (147)	31.01 (25.94)
GenKOLS	Cases	836	39.23	40-90	65.44 (10.1)	146-197	170 (9.02)	836	1.477 (0.699)	2.863 (0.957)	0.502 (0.126)	100 (836)	3 - 130 (836)	31.88 (18.62)
	Controls	692	48.84	40-88	55.43 (9.74)	151-200	172.05 (8.79)	692	3.214 (0.722)	4.169 (0.935)	0.772 (0.041)	100 (692)	2.5 - 90 (692)	19.4 (13.61)
NETT/NAS	Cases	374	36.1	40-85	67.47 (5.76)	142-190	168.76 (9.53)	374	0.726 (0.236)	2.299 (0.775)	0.324 (0.064)	100 (374)	12 - 260 (371)	66.25 (30.66)
	Controls	429	0	48-89	69.86 (7.5)	156-192	174.46 (6.79)	429	3.032 (0.507)	3.83 (0.627)	0.793 (0.053)	100 (429)	10 - 185.5 (429)	40.69 (27.79)
SPIROMICS	Cases	988	44	41-89	65.74 (7.62)	141-197	170.05 (9.64)	988	1.539 (0.605)	3.194 (0.927)	0.48 (0.13)	100 (988)	20.0 - 450 (988)	56.11 (28.78)
	Controls	537	53	40-80	62.95 (9.0)	149-205	169.54 (9.62)	537	2.824 (0.705)	3.678 (0.913)	0.77 (0.04)	100 (537)	20.0 - 400 (537)	44.76 (26.36)

Supplementary Table 20: External case-control studies COPD risk score association results (per-allele and per standard deviation)

Results of the association between genetic risk scores and COPD risk are given for both weighted (top) and unweighted (bottom) risk scores (comprising 279 novel and previous signals), for five studies: the COPD Gene study, the ECLIPSE study (Evaluation of COPD Longitudinally to Identify Predictive Surrogate End-points), GenKOLS (the Bergen, Norway COPD cohort), NETT/NAS (the National Emphysema Treatment Trial [NETT] and the Normative Aging Study [NAS]) and the SPIROMICS study. COPD Gene is stratified into African-American and Non-hispanic white subgroups. Effect sizes and 95% confidence intervals are given on two scales: a per-Allele (i.e. raw) scale, and a per standard deviation (SD) scale. Standard deviations for the weighted and unweighted risk scores are given for each cohort separately. Abbreviations: AA=African-American; Nhw=Non-Hispanic white; OR=odds ratio; 95LCI/UCI=lower and upper bounds of 95% confidence intervals; P=p-value; N=sample size. A sensitivity analysis was also run, excluding SNP rs13116999 (see 'Discussion' of manuscript). The per-allele meta-analytic estimate was consistent after excluding this SNP. *The odd ratios per standard deviation increase in the risk score were estimated as: exp(logOR on the per allele scale × standard deviation of the weighted risk score). **Approximated in R as sqrt(sum(SD^2*(N-1))/sum(N-1)), where N is a vector of sample sizes, and SD is a vector of standard deviations.

Ancestry	Study group	per Allele			P	per Standard Deviation*			P*	N			Mean risk score	SD risk score
		OR	95LCI	95UCI		OR	95LCI	95UCI		Total	Cases	Controls		
Weighted														
African	COPDGene (AA)	1.023	1.014	1.032	8.36E-07	1.255	1.147	1.374	8.36E-07	2466	910	1556	306.16	10.09
European	COPDGene (NHW)	1.036	1.03	1.041	1.97E-41	1.535	1.442	1.634	1.97E-41	5178	3068	2110	307.72	12.25
European	ECLIPSE	1.039	1.023	1.055	1.42E-06	1.585	1.314	1.912	1.42E-06	1860	1713	147	309.80	12.16
European	GenKOLS	1.042	1.031	1.052	8.99E-15	1.623	1.436	1.834	8.99E-15	1528	836	692	308.05	11.89
European	NETT/NAS	1.032	1.017	1.047	3.13E-05	1.464	1.223	1.751	3.13E-05	803	374	429	307.54	12.16
European	SPIROMICS	1.037	1.027	1.046	4.47E-14	1.539	1.376	1.721	4.47E-14	1525	988	537	307.90	11.95
Meta-analysis of 5 European-ancestry study groups		1.037	1.033	1.041	1.72E-75	1.546	1.476	1.620	1.48E-75	10894	6979	3915	308.13	12.14**
Unweighted														
African	COPDGene (AA)	1.015	1.005	1.025	0.00251	1.147	1.049	1.254	0.00251	2466	910	1556	298.62	9.33
European	COPDGene (NHW)	1.034	1.028	1.04	3.03E-28	1.413	1.329	1.503	3.03E-28	5178	3068	2110	294.74	10.36
European	ECLIPSE	1.037	1.02	1.055	2.45E-05	1.476	1.232	1.769	2.45E-05	1860	1713	147	296.41	10.63
European	GenKOLS	1.046	1.033	1.059	7.58E-13	1.561	1.382	1.764	7.58E-13	1528	836	692	295.60	9.96
European	NETT/NAS	1.019	1.002	1.035	2.73E-02	1.212	1.022	1.439	2.73E-02	803	374	429	294.69	10.48
European	SPIROMICS	1.036	1.025	1.047	1.62E-10	1.435	1.284	1.603	1.62E-10	1525	988	537	294.20	10.27
Meta-analysis of 5 European-ancestry study groups		1.035	1.030	1.039	6.71E-52	1.425	1.362	1.492	4.45e-52	10894	6979	3915	295.07	10.35**

Supplementary Table 21: COPD risk score association results in external case-control studies (per-decile)

Within each study group, individuals were divided according to their value of the weighted genetic risk score. Logistic models were then fitted for each decile, comparing odds of COPD between members of each decile (2-10) and the lowest decile (1, the reference decile). Results were meta-analysed by fixed-effects across the European-ancestry subjects of COPDGene (Non-hispanic white participants), ECLIPSE, GenKOLS, NETT/NAS, SPIROMICS. Results are presented separately for African-American participants of the COPDGene study

Decile	Meta-analysis of 5 European Cohorts*				COPDGene (African-American)			
	OR	LCI	UCI	P	OR	LCI	UCI	P
1	1.000				1.000			
2	1.470	1.207	1.790	1.26E-04	0.881	0.566	1.370	0.573
3	1.572	1.289	1.918	7.97E-06	1.407	0.927	2.135	0.109
4	2.092	1.712	2.555	4.94E-13	1.281	0.838	1.961	0.253
5	2.045	1.678	2.491	1.23E-12	1.639	1.083	2.481	0.020
6	2.033	1.666	2.481	2.93E-12	1.214	0.807	1.825	0.352
7	2.520	2.054	3.091	7.21E-19	1.215	0.784	1.882	0.383
8	2.800	2.282	3.435	5.76E-23	1.376	0.902	2.101	0.139
9	3.961	3.213	4.883	5.15E-38	1.895	1.255	2.863	2.38E-03
10	4.731	3.793	5.900	3.00E-43	2.660	1.753	4.036	4.25E-06

***COPDGene (Non-hispanic white participants), ECLIPSE, GenKOLS, NETT/NAS, SPIROMICS**

Supplementary Table 22: Results for PheWAS of weighted genetic risk score

See Excel spreadsheet.

Results are given for 2,453 traits studied. The exposure was the 279-SNP weighted genetic risk score. Each trait was assigned a disease category='Final.Category'). Total sample sizes (N), as well as numbers of cases and controls are given. Odds ratios (OR) are given for binary traits, and beta coefficients are given for continuous traits. Confidence intervals (LCI95, UCI95) and P values are also provided, along with false discovery rates. 'FDR.Flag' denotes associations passing an FDR of <0.01. 'Quant.Resp.Trait' is a flag variable indicating PheWAS results for those SNPs featuring in the main GWAS. 'Figure.Name' denotes the short plain English label used in the Figure in the main text, allowing for cross reference.

Supplementary Table 23: Look-up of new and previously reported lung function signals in GRASP and GWAS catalog

See Excel spreadsheet.

Tabulated results of a lookup of sentinel variants and variants in their respective 99% credible sets against all associations $P< 5x10^{-8}$ in the EBI GWAS catalog (https://www.ebi.ac.uk/gwas/) and GRASP (https://grasp.nhlbi.nih.gov/Overview.aspx). Associations relating to methylation, expression, metabolite or protein levels, as well as associations with lung function were removed. The table first shows the ID and genomic position of the sentinel variant that was associated with the trait in question (either the sentinel variant, or one of its 99% credible set variants was associated with the trait). Next, the details of the association with lung function for this variant in the current study are shown (trait, whether the signal identified in Tier 1 or Tier 2). If this signal is not a novel signal, details of the original sentinel variant and trait are given. For retrieved studies mapping to the sentinel (or its credible set variants), all reported genes across the studies of interest are given, along with all traits, and the PUBMED IDs of the papers from which associations were retrieved.

Supplementary Table 24: LD score regression results

Results for the regression of each trait FEV1, FVC, FEV1/FVC and PEF 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 χ2 : Mean χ2 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 χ2 -1). 95% confidence intervals are shown in brackets.

UK Biobank (n=321,047)	FEV_1	FVC	FEV_1/FVC	PEF
Total Observed scale h2	0.185 (0.173, 0.198)	0.187 (0.175, 0.199)	0.211 (0.19, 0.232)	0.155 (0.14, 0.17)
Lambda GC	1.841	1.841	1.841	1.695
Mean Chi^2	2.328	2.355	2.578	2.138
Intercept	1.119 (1.096, 1.142)	1.139 (1.113, 1.164)	1.193 (1.162, 1.225)	1.133 (1.106, 1.159)
Ratio	0.09 (0.072, 0.107)	0.102 (0.083, 0.121)	0.123 (0.103, 0.142)	0.117 (0.094, 0.139)

SpiroMeta (n=79,055)	FEV_1	FVC	FEV_1/FVC	PEF
Total Observed scale h2	0.126 (0.107, 0.145)	0.116 (0.097, 0.134)	0.095 (0.077, 0.113)	0.094 (0.055, 0.134)
Lambda GC	1.146	1.146	1.114	1.017
Mean Chi^2	1.194	1.178	1.141	1.017
Intercept	0.998 (0.983, 1.013)	1.003 (0.986, 1.019)	0.993 (0.979, 1.007)	0.972 (0.959, 0.986)
Ratio	<0	0.014 (-0.078, 0.106)	<0	<0

Meta-analysis (n up to 400,102)	FEV_1	FVC	FEV_1/FVC	PEF
Total Observed scale h2	0.154 (0.144, 0.165)	0.152 (0.142, 0.161)	0.152 (0.137, 0.167)	0.131 (0.118, 0.143)
Lambda GC	1.757	1.781	1.581	1.489
Mean Chi^2	2.291	2.261	2.272	1.919
Intercept	1.041 (1.018, 1.065)	1.04 (1.015, 1.065)	1.033 (1.006, 1.061)	1.006 (0.982, 1.031)
Ratio	0.032 (0.014, 0.05)	0.032 (0.012, 0.051)	0.026 (0.005, 0.048)	0.007 (-0.02, 0.034)

Supplementary Table 25: Weights for COPD risk score

See Excel spreadsheet.

Weights for COPD risk score. Weights for each the 279 variants were selected from the FEV1/FVC ratio results for UK Biobank or SpiroMeta. The FEV1/FVC ratio decreasing (i.e. COPD risk *increasing*) allele was chosen. To minimise the risk of winner's curse bias, the study which was not used in the discovery of a given signal was used as the source of the weight. For previously reported signals, this meant that most weights were taken from UK Biobank (if UK Biobank was used in signal discovery, SpiroMeta was used to derive weights). For novel signals identified in this study, the source of weight depended on whether the signal was identified in the two-stage (Tier 1) approach, or the joint, one-stage (Tier 2) approach. SpiroMeta was the source of weights for two-stage signals, and for one-stage signals, the smallest absolute effect size from UK Biobank or SpiroMeta was chosen. Betas are the FEV1/FVC ratio effect size from the study defined in the column 'Source'. Weights were calculated as the beta for a given variant, divided by the sum of all 279 betas, multiplied by the number of variants (279), such that the sum of the weights added to 279.

Supplementary Table 26: STRING druggability analysis

See Excel spreadsheet. Please note that it is possible to filter this table using the drop-down arrows at the top of each column.

STRING was used to identify high confidence (threshold 0.9) protein-protein interactions with the products of the 107 high-priority genes (**Supplementary Table 13)**. 861 drugs were identified using this method (823 of which were not identified in the simple analysis). The drugs were then mapped to CHEMBL IDs and indications (given as Medical Subject Headings, or 'MeSH' terms, https://www.ebi.ac.uk/chembl/drug/indications). For each gene demonstrating an interaction, the sentinel SNP that implicated the gene is given. Drug names, plus CHEMBL IDs, and drug indications (with maximum development phase in brackets) are given.

Column explanations:

- Drug=compound/drug name;
- CHEMBL_ID=compound identification number from CHEMBL;
- OriginalGeneAndSource=This is the name of the gene (or genes) amongst the 107 for which an interaction was identified using STRING analysis. The 'source' in brackets, given after the gene name, details the sentinel SNP in the GWAS that implicated this gene, and the reason for which it was implicated (i.e. location of a deleterious variant, implicated by the eQTl/pQTL analyses);
- STRINGGeneTarget=The name of a gene implicated by a protein-protein interaction identified using STRING (interacting protein targets were mapped back to genes using UniProt);
- IndicationPhase=Drug indication (Phase). 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.
- MAB=Drug is a monoclonal antibody;
- OriginalGenesPathway=the gene given in the 'Original Gene and Source' column is a gene identified in the 'Enriched Biological Pathways' shown in **Supplementary Table 15**;
- Cancer=the drug is used to treat some form of cancer;
- Phase3or4=the drug has at least one indication annotated as Phase 3 or 4;
- AsthmaCOPD=the drug is already indicated as being used in asthma or COPD;
- Novelty=the drug is implicated for use by genes identified from novel signals in this GWAS.

References

1. Miller, M.R. *et al.* Standardisation of spirometry. *Eur Respir J* **26**, 319-38 (2005).
2. 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. *The Lancet Respiratory Medicine* **3**, 769-781 (2015).
3. Bycroft, C. *et al.* Genome-wide genetic data on ~500,000 UK Biobank participants. *bioRxiv* (2017).
4. Bulik-Sullivan, B.K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nature Genetics* **47**, 291 (2015).
5. Yang, J., Lee, S.H., Goddard, M.E. & Visscher, P.M. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet* **88**, 76-82 (2011).
6. 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).
7. Strachan, D.P. *et al.* Lifecourse influences on health among British adults: effects of region of residence in childhood and adulthood. *Int J Epidemiol* **36**, 522-31 (2007).
8. Marossy, A.E., Strachan, D.P., Rudnicka, A.R. & Anderson, H.R. Childhood chest illness and the rate of decline of adult lung function between ages 35 and 45 years. *Am J Respir Crit Care Med* **175**, 355-9 (2007).
9. Vitart, V. *et al.* SLC2A9 is a newly identified urate transporter influencing serum urate concentration, urate excretion and gout. *Nat Genet* **40**, 437-42 (2008).
10. Zemunik, T. *et al.* Genome-wide association study of biochemical traits in Korcula Island, Croatia. *Croat Med J* **50**, 23-33 (2009).
11. Rudan, I. *et al.* "10001 Dalmatians:" Croatia launches its national biobank. *Croat Med J* **50**, 4-6 (2009).
12. Day, N. *et al.* EPIC-Norfolk: study design and characteristics of the cohort. European Prospective Investigation of Cancer. *Br J Cancer* **80 Suppl 1**, 95-103 (1999).
13. Smith, B.H. *et al.* Cohort Profile: Generation Scotland: Scottish Family Health Study (GS:SFHS). The study, its participants and their potential for genetic research on health and illness. *Int J Epidemiol* **42**, 689-700 (2013).
14. Heistaro, S. Methodology report. Health 2000 survey. in *Publications of National Public Health Institute* (ed. Heistaro, S.) (2000).
15. Kristiansson, K. *et al.* Genome-wide screen for metabolic syndrome susceptibility Loci reveals strong lipid gene contribution but no evidence for common genetic basis for clustering of metabolic syndrome traits. *Circ Cardiovasc Genet* **5**, 242-9 (2012).
16. Holle, R., Happich, M., Lowel, H., Wichmann, H.E. & Group, M.K.S. KORA--a research platform for population based health research. *Gesundheitswesen* **67 Suppl 1**, S19-25 (2005).
17. Wichmann, H.E., Gieger, C., Illig, T. & Group, M.K.S. KORA-gen--resource for population genetics, controls and a broad spectrum of disease phenotypes. *Gesundheitswesen* **67 Suppl 1**, S26-30 (2005).
18. Peters, A. *et al.* [Multimorbidity and successful aging: the population-based KORA-Age study]. *Z Gerontol Geriatr* **44 Suppl 2**, 41-54 (2011).
19. Burney, P.G., Luczynska, C., Chinn, S. & Jarvis, D. The European Community Respiratory Health Survey. *Eur Respir J* **7**, 954-60 (1994).
20. Main Protocol for The European Community Respiratory Health Survey (ECRHS) I, http://www.ecrhs.org/ECRHS%20I/Main%20protocol.pdf.
21. Deary, I.J. *et al.* The Lothian Birth Cohort 1936: a study to examine influences on cognitive ageing from age 11 to age 70 and beyond. *BMC Geriatr* **7**, 28 (2007).
22. Rantakallio, P. The longitudinal study of the northern Finland birth cohort of 1966. *Paediatr Perinat Epidemiol* **2**, 59-88 (1988).
23. Sovio, U. *et al.* Genetic determinants of height growth assessed longitudinally from infancy to adulthood in the northern Finland birth cohort 1966. *PLoS Genet* **5**, e1000409 (2009).
24. Jarvelin, M.R., Hartikainen-Sorri, A.L. & Rantakallio, P. Labour induction policy in hospitals of different levels of specialisation. *Br J Obstet Gynaecol* **100**, 310-5 (1993).
25. Jaaskelainen, A. *et al.* Meal frequencies modify the effect of common genetic variants on body mass index in adolescents of the northern Finland birth cohort 1986. *PLoS One* **8**, e73802 (2013).

26. Aulchenko, Y.S., Ripke, S., Isaacs, A. & van Duijn, C.M. GenABEL: an R library for genome-wide association analysis. *Bioinformatics* **23**, 1294-6 (2007).
27. Lind, L., Fors, N., Hall, J., Marttala, K. & Stenborg, A. A comparison of three different methods to evaluate endothelium-dependent vasodilation in the elderly: the Prospective Investigation of the Vasculature in Uppsala Seniors (PIVUS) study. *Arterioscler Thromb Vasc Biol* **25**, 2368-75 (2005).
28. Martin, B.W. *et al.* SAPALDIA: methods and participation in the cross-sectional part of the Swiss Study on Air Pollution and Lung Diseases in Adults. *Soz Praventivmed* **42**, 67-84 (1997).
29. Ackermann-Liebrich, U. *et al.* Follow-up of the Swiss Cohort Study on Air Pollution and Lung Diseases in Adults (SAPALDIA 2) 1991-2003: methods and characterization of participants. *Soz Praventivmed* **50**, 245-63 (2005).
30. Volzke, H. *et al.* Cohort profile: the study of health in Pomerania. *Int J Epidemiol* **40**, 294-307 (2011).
31. Nelson, S.B., Gardner, R.M., Crapo, R.O. & Jensen, R.L. Performance evaluation of contemporary spirometers. *Chest* **97**, 288-97 (1990).
32. Standardization of spirometry--1987 update. Statement of the American Thoracic Society. *Am Rev Respir Dis* **136**, 1285-98 (1987).
33. Quanjer, P.H. *et al.* Lung volumes and forced ventilatory flows. Report Working Party Standardization of Lung Function Tests, European Community for Steel and Coal. Official Statement of the European Respiratory Society. *Eur Respir J Suppl* **16**, 5-40 (1993).
34. Raitakari, O.T. *et al.* Cohort profile: the cardiovascular risk in Young Finns Study. *Int J Epidemiol* **37**, 1220-6 (2008).
35. Wilk, J.B. *et al.* A genome-wide association study of pulmonary function measures in the Framingham Heart Study. *PLoS Genet* **5**, e1000429 (2009).
36. 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).
37. Repapi, E. *et al.* Genome-wide association study identifies five loci associated with lung function. *Nat Genet* **42**, 36-44 (2010).
38. 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).
39. Cho, M.H. *et al.* A genome-wide association study of COPD identifies a susceptibility locus on chromosome 19q13. *Hum Mol Genet* **21**, 947-57 (2012).
40. Loth, D.W. *et al.* Genome-wide association analysis identifies six new loci associated with forced vital capacity. **46**, 669-77 (2014).
41. Lutz, S.M. *et al.* A genome-wide association study identifies risk loci for spirometric measures among smokers of European and African ancestry. *BMC Genet* **16**, 138 (2015).
42. Soler Artigas, M. *et al.* Sixteen new lung function signals identified through 1000 Genomes Project reference panel imputation. *Nat Commun* **6**, 8658 (2015).
43. Hobbs, B.D. *et al.* Exome Array Analysis Identifies a Common Variant in IL27 Associated with Chronic Obstructive Pulmonary Disease. **194**, 48-57 (2016).
44. Hobbs, B.D. *et al.* Genetic loci associated with chronic obstructive pulmonary disease overlap with loci for lung function and pulmonary fibrosis. *Nat Genet* **49**, 426-432 (2017).
45. Wain, L.V. *et al.* Genome-wide association analyses for lung function and chronic obstructive pulmonary disease identify new loci and potential druggable targets. *Nat Genet* **49**, 416-425 (2017).
46. Wyss, A.B. *et al.* Multiethnic Meta-analysis Identifies New Loci for Pulmonary Function. *bioRxiv* (2017).
47. Jackson, V. *et al.* Meta-analysis of exome array data identifies six novel genetic loci for lung function [version 1; referees: 1 approved with reservations]. *Wellcome Open Research* **3**(2018).
48. Yengo, L. *et al.* Meta-analysis of genome-wide association studies for height and body mass index in ~700,000 individuals of European ancestry. *bioRxiv* (2018).
49. Bulik-Sullivan, B.K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet* **47**, 291-5 (2015).
50. Wakefield, J. Reporting and interpretation in genome-wide association studies. *Int J Epidemiol* **37**, 641-53 (2008).
51. Hao, K. *et al.* Lung eQTLs to help reveal the molecular underpinnings of asthma. *PLoS Genet* **8**, e1003029 (2012).

52. Lamontagne, M. *et al.* Refining susceptibility loci of chronic obstructive pulmonary disease with lung eqtls. *PLoS One* **8**, e70220 (2013).
53. Obeidat, M. *et al.* GSTCD and INTS12 regulation and expression in the human lung. *PLoS One* **8**, e74630 (2013).
54. Sun, B.B. *et al.* Genomic atlas of the human plasma proteome. *Nature* **558**, 73-79 (2018).