Immune-mediated genetic pathways resulting in pulmonary function impairment increase lung cancer susceptibility

Linda Kachuri[1,2], Mattias Johansson[§ 3], Sara R. Rashkin[1], Rebecca E. Graff[1], Yohan Bossé[4], Venkata Manem[4], Neil E. Caporaso[5], Maria Teresa Landi[5], David C. Christiani[6], Paolo Vineis[7], Geoffrey Liu[8], Ghislaine Scelo[§ 3], David Zaridze[9], Sanjay S. Shete[10], Demetrius Albanes[5], Melinda C. Aldrich[11], Adonina Tardón[12], Gad Rennert[13], Chu Chen[14], Gary E. Goodman[14], Jennifer A. Doherty[15], Heike Bickeböller[16], John K. Field[17], Michael P. Davies[17], M. Dawn Teare[18], Lambertus A. Kiemeney[19], Stig E. Bojesen[20], Aage Haugen[21], Shanbeh Zienolddiny[21], Stephen Lam[22], Loïc Le Marchand[23], Iona Cheng[1], Matthew B. Schabath[24], Eric J. Duell[25], Angeline S. Andrew[26], Jonas Manjer[27], Philip Lazarus[28], Susanne Arnold[29], James D. McKay[§ 3], Nima C. Emami[1], Matthew T. Warkentin[2,30], Yonathan Brhane[2], Ma'en Obeidat[31], Richard M. Martin[32-34], Caroline Relton[32,33], George Davey Smith[32,33], Phillip C. Haycock[32,33], Christopher I. Amos[35], Paul Brennan[§ 2], John S. Witte*[1], Rayjean J. Hung*[2,30]

1. Department of Epidemiology & Biostatistics, University of California San Francisco, San Francisco, CA, USA
2. Prosserman Centre for Population Health Research, Lunenfeld-Tanenbaum Research Institute, Sinai Health System, Toronto, ON, Canada
3. International Agency for Research on Cancer, Lyon, France
4. Institut universitaire de cardiologie et de pneumologie de Québec – Université Laval, Quebec City, Canada
5. Division of Cancer Epidemiology & Genetics, US NCI, Bethesda, MD, USA
6. Department of Epidemiology, Harvard TH Chan School of Public Health, Boston, MA, USA
7. Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, London, UK
8. Princess Margaret Cancer Center, University Health Network, Toronto, ON, Canada
9. Russian N.N. Blokhin Cancer Research Centre, Moscow, Russian Federation
10. Department of Biostatistics, Division of Basic Sciences, MD Anderson Cancer Center, Houston, TX, USA
11. Department of Thoracic Surgery and Division of Epidemiology, Vanderbilt University Medical Center, Nashville, TN, USA
12. Faculty of Medicine, University of Oviedo and CIBERESP, Campus del Cristo, Oviedo, Spain
13. Clalit National Cancer Control Center, Technion Faculty of Medicine, Haifa, Israel
14. Fred Hutchinson Cancer Research Center, Seattle, WA, USA
15. Department of Population Health Sciences, Huntsman Cancer Institute, Salt Lake City, UT, USA
16. Department of Genetic Epidemiology, University Medical Center, Georg-August-Universität Göttingen, Göttingen, Germany
17. Roy Castle Lung Cancer Research Programme, Department of Molecular and Clinical Cancer Medicine, The University of Liverpool, UK
18. Biostatistics Research Group, Institute of Health and Society, Newcastle University, Newcastle upon Tyne, UK
19. Radboud Institute for Health Sciences, Radboud University Medical Centre, Nijmegen, The Netherlands
20. Department of Clinical Biochemistry, Herlev and Gentofte Hospital, Copenhagen University Hospital, Herlev, Denmark
21. The National Institute of Occupational Health, Oslo, Norway
22. BC Cancer Agency, Vancouver, BC, Canada
23. Epidemiology Program, University of Hawaii Cancer Center, Honolulu, HI, USA
24. Department of Cancer Epidemiology, H. Lee Moffitt Cancer Center & Research Institute, Tampa, FL, USA
25. Catalan Institute of Oncology (ICO-IDIBELL), Barcelona, Spain
26. Department of Epidemiology, Geisel School of Medicine, Dartmouth College, Hanover, NH, USA
27. Skåne University Hospital, Lund University, Lund, Sweden
28. College of Pharmacy, Washington State University, Spokane, WA, USA
29. Markey Cancer Center, University of Kentucky, Lexington, KY, USA
30. Epidemiology Division, Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada
31. University of British Columbia, Centre for Heart Lung Innovation, Vancouver, BC, Canada
32. MRC Integrative Epidemiology Unit, University of Bristol, Bristol, UK
33. Bristol Medical School, Population Health Sciences, University of Bristol, Bristol, United Kingdom

34. National Institute for Health Research (NIHR) Bristol Biomedical Research Centre, University Hospitals Bristol NHS Foundation Trust and the University of Bristol, Bristol, UK
35. Institute for Clinical and Translational Research, Baylor College of Medicine, Houston, TX, USA

§ Disclaimer: Where authors are identified as personnel of the International Agency for Research on Cancer / World Health Organization, the authors alone are responsible for the views expressed in this article and they do not necessarily represent the decisions, policy or views of the International Agency for Research on Cancer / World Health Organization

*Co-senior / corresponding authors:

Rayjean J. Hung
Lunenfeld-Tanenbaum Research Institute of Sinai Health System
60 Murray Street, Room L5-215, Box 18
Toronto, ON M5T 3L9 Canada
E-mail: rayjean.hung@lunenfeld.ca

John S. Witte
Department of Epidemiology and Biostatistics
University of California, San Francisco
1450 3rd Street, Room 388
San Francisco, CA 94158 USA
Email: jwitte@ucsf.edu

ABSTRACT

Impaired lung function is an indicator of obstructive pulmonary disease and may be a consequence of cigarette smoking, making it challenging to disentangle its role in lung cancer etiology. We investigated the shared genetic basis of pulmonary dysfunction and lung cancer susceptibility using genome-wide data from the UK Biobank (>370,000 individuals) and the International Lung Cancer Consortium (29,266 cases, 56,450 controls). We observed strong genetic correlations between lung cancer and reduced forced expiratory volume in one second (FEV_1: r_g=0.098, p=2.3×10^{-8}) and the ratio of FEV_1 to forced vital capacity (FEV_1/FVC: r_g=0.137, p=2.0×10^{-12}). In Mendelian randomization analyses reduced FEV_1 was associated with risk of squamous cell carcinoma (odds ratio (OR)=1.51, 95% confidence intervals: 1.21-1.88), while reduced FEV_1/FVC increased risk of adenocarcinoma (OR=1.17, 1.01-1.35) and lung cancer in never smokers (OR=1.56, 1.05-2.30). These results support a causal role of lung impairment in lung cancer etiology. Integrative analyses of pulmonary function instruments, including 73 newly discovered variants, revealed significant effects on lung tissue gene expression and implicated immune-related pathways in mediating the effects on lung cancer susceptibility.

INTRODUCTION

Lung cancer is the most commonly diagnosed cancer worldwide and the leading cause of cancer mortality[1]. Although tobacco smoking remains the predominant risk factor for lung cancer, clinical observations and epidemiological studies have consistently shown that individuals with airflow limitation, particularly those with chronic obstructive pulmonary disease (COPD), have a significantly higher risk of developing lung cancer[2-7]. Several lines of evidence suggest that biological processes resulting in pulmonary impairment warrant consideration as independent lung cancer risk factors, including observations that previous lung diseases influence lung cancer risk independently of tobacco use[6,8-10], and overlap in genetic susceptibility loci for lung cancer and chronic obstructive pulmonary disease (COPD) on 4q24 (*FAM13A*), 4q31 (*HHIP*), 5q.32 (*HTR4*), the 6p21 region, and 15q25 (*CHRNA3*/*CHRNA5*)[11-14]. Inflammation and oxidative stress have been proposed as key mechanisms promoting lung carcinogenesis in individuals affected by COPD or other non-neoplastic lung pathologies[9,11,15].

Despite an accumulation of observational findings, previous epidemiological studies have been unable to conclusively establish a causal link between indicators of impaired pulmonary function and lung cancer risk due to the interrelated nature of these conditions[7]. Lung cancer and obstructive pulmonary disease share multiple etiological factors, such as cigarette smoking, occupational inhalation hazards, and air pollution, and 50-70% of lung cancer patients present with co-existing COPD or airflow obstruction[6]. Furthermore, reverse causality remains a concern since pulmonary symptoms may be early manifestations of lung cancer or acquired lung diseases in patients whose immune system has already been compromised by undiagnosed cancer.

Disentangling the role of pulmonary impairment in lung cancer development is important from an etiological perspective, for refining disease susceptibility mechanisms, and for informing precision prevention and risk stratification strategies. In this study we comprehensively assessed the shared genetic basis of impaired lung function and lung cancer risk by conducting genome-wide association analyses in the UK Biobank cohort to identify genetic determinants of three pulmonary phenotypes, forced expiratory volume in 1 second (FEV_1), forced vital capacity (FVC), and FEV_1/FVC. We examined the genetic correlation between pulmonary function phenotypes and lung cancer, followed by Mendelian randomization (MR) using novel genetic instruments to formally test the causal relevance of impaired pulmonary function, using the largest available dataset of 29,266 lung cancer cases and 56,450 controls from the OncoArray lung cancer collaboration[16].

RESULTS

Heritability and Genetic Correlation

Array-based, or narrow-sense, heritability (h_g) estimates for all lung phenotypes were obtained using LD score regression[17] based on summary statistics from our GWAS of the UKB cohort (n=372,750 for FEV_1, n=370,638 for FVC, n=368,817 for FEV_1/FVC; Supplementary Figure 1) are presented in Table 1. Heritability estimates based on UKB-specific LD scores (n=7,567,036 variants), were consistently lower but more precise than those based on the 1000 Genomes (1000 G) Phase 3 reference population (n=1,095,408 variants). For FEV_1, h_g = 0.163 (SE=0.006) and h_g = 0.201 (SE=0.008), based on UKB and 1000G LD scores, respectively. Estimates for FVC were h_g = 0.175 (SE=0.007) and h_g = 0.214 (SE=0.010). Heritability was lower for FEV_1/FVC: h_g = 0.128 (SE=0.006) and 0.157 (SE=0.010), based on internal and 1000G reference panels, respectively. For all phenotypes, h_g did not differ by smoking status and estimates were not affected by excluding the major histocompatibility complex (MHC) region.

Partitioning heritability by functional annotation identified large and statistically significant ($p<8.5\times10^{-4}$) enrichments for multiple categories (Figure 1; Supplementary Tables 1-3). A total of 35 categories, corresponding to 22 distinct annotations, were significantly enriched for all three pulmonary phenotypes,

including annotations that were not previously reported[18]. Large enrichment, defined as the proportion of heritability accounted for by a specific category relative to the proportion of SNPs in that category, was observed for elements conserved in primates[19,20] (17.6% of SNPs, 54.7-58.5% of h_g), McVicker background selection statistic[21,22] (17.8% of SNPs, 22.6-25.1% of h_g), flanking bivalent transcription starting sites (TSS)/enhancers from Roadmap[20,23] (1.4% of SNPs, 11.1-13.2% of h_g), and super enhancers (16.7% of SNPs, 33.9-38.6% of h_g). We also replicated previously reported significant enrichments for histone methylation and acetylation marks H3K4me1, H3K9Ac, and H3K27Ac[18,24].

Substantial genetic correlation was observed for pulmonary phenotypes with body composition and smoking traits, mirroring phenotypic correlations in epidemiologic studies (Figure 2). Large positive correlations with height were observed for FEV_1 (r_g=0.568, p=2.5×10^{-567}) and FVC (r_g=0.652, p=1.8×10^{-864}). Higher adiposity was negatively correlated with FEV_1 (BMI: r_g=-0.216, p=4.2×10^{-74}; percent body fat: r_g=-0.221, p=1.7×10^{-66}), FVC (BMI: r_g=-0.262, p=1.6×10^{-114}; percent body fat: r_g=-0.254, p=1.2×10^{-88}). Smoking status (ever vs. never) was significantly correlated with all lung function phenotypes (FEV_1 r_g=-0.221, p=8.1×10^{-78}; FVC r_g=-0.091, 1.0×10^{-16}; FEV_1/FVC r_g=-0.360, p=7.5×10^{-130}). Cigarette pack years and impaired lung function in smokers were also significantly genetically correlated with FEV_1 (r_g=-0.287 p=1.1×10^{-35}), FVC (r_g=-0.253, p=1.9×10^{-30}), and FEV_1/FVC (r_g=-0.108, p=3.0×10^{-4}). As a positive control, we verified that FEV_1 and FVC were genetically correlated with each other (r_g=0.922) and with FEV_1/FVC (FEV_1: r_g=0.232 p=4.1×10^{-32}; FVC: r_g= -0.167 p=1.0×10^{-19}).

Genetic correlations between lung function phenotypes and lung cancer are presented in Figure 3. For simplicity of interpretation coefficients were rescaled to represent genetic correlation with impaired (decreasing) lung function. Impaired FEV_1 was positively correlated with lung cancer overall (r_g=0.098, p=2.3×10^{-8}), squamous cell carcinoma (r_g=0.137, p=7.6×10^{-9}), and lung cancer in smokers (r_g=0.140, p=1.2×10^{-7}). Genetic correlations were attenuated for adenocarcinoma histology (r_g=0.041, p=0.044) and null for never-smokers (r_g=-0.002, p=0.96). A similar pattern of associations was observed for FVC. Reduced FEV_1/FVC was positively correlated with all lung cancer subgroups (overall: r_g=0.137, p=2.0×10^{-12}; squamous carcinoma: r_g=0.137, p=4.3×10^{-8}; adenocarcinoma: r_g=0.125, p=7.2×10^{-9}; smokers: r_g=0.185, p=1.4×10^{-10}), except for never smokers (r_g=0.031, p=0.51).

Exploring the functional underpinnings of these genetic correlations revealed three functional categories that were significantly enriched for lung cancer (Supplementary Table 4), and have not been previously reported[25]. All these categories were also significantly enriched for pulmonary traits. CpG dinucleotide content[22] included only 1% of SNPs, but had a strong enrichment signal for lung cancer (p=2.1×10^{-7}), FEV_1 (p=7.7×10^{-24}), FVC (p=2.3×10^{-23}, and FEV_1/FVC (p=3.8×10^{-17}). Other shared features included background selection (lung cancer: p=1.0×10^{-6}, FEV_1: p=1.9×10^{-20}, FVC: p=6.9×10^{-23}, FEV_1/FVC: p=1.5×10^{-15}) and super enhancers (lung cancer: p=4.4×10^{-6}, FEV_1: p=3.4×10^{-24}, FVC: p=5.1×10^{-20}, FEV_1/FVC: p=9.6×10^{-22}).

Genome-Wide Association Analysis for Instrument Development

Based on the results of our GWAS in the UK Biobank, we identified 207 independent instruments for FEV_1 ($P<5\times10^{-8}$, replication $P<0.05$; LD $r^2<0.05$ within 10,000 kb), 162 for FVC, and 297 for FEV_1/FVC. We confirmed that our findings were not affected by spirometry performance quality, with a nearly perfect correlation between effect sizes (R^2=0.995, p=2.5×10^{-196}) in the main discovery analysis and after excluding individuals with potential blow acceptability issues (Field 3061≠0; n=60,299). After applying these variants to the lung cancer OncoArray dataset and selecting LD proxies (r^2>0.90) for unavailable variants, the final set of instruments consisted of 193 variants for FEV_1, 144 for FVC, and 264 SNPs for FEV_1/FVC (Supplementary Files 1-3), for a total of 601 instruments. The proportion of trait variation accounted for by each set of instruments was estimated in the UKB replication sample consisting of over 110,00 individuals (Supplementary Figure 1), and corresponded to 3.13% for FEV_1, 2.27% for FVC, and 5.83% for FEV_1/FVC.

We also developed instruments specifically for never smokers based on a separate GWAS of this population, which yielded 76 instruments for FEV_1, 112 for FEV_1/FVC, and 57 for FVC, accounting for 2.06%, 4.21%, and 1.36% of phenotype variation, respectively (Supplementary Files 4-6).

After removing overlapping instruments between pulmonary phenotypes and LD-filtering (r^2<0.05) across the three traits, 447 of the 601 variants were associated with at least one of FEV_1, FVC, or FEV_1/FVC ($P<5\times10^{-8}$, replication P<0.05). We compared these 447 independent variants to the 279 lung function variants recently reported by Shrine et al.[18] based on an analysis of the UK Biobank and SpiroMeta consortium, by performing clumping with respect to these index variants (LD r^2<0.05 within 10,000 kb). Our set of instruments included an additional 73 independent variants, 69 outside the MHC region (Supplementary Table 5), that achieved replication at the Bonferroni-corrected threshold for each trait (maximum replication $P = 2.0\times10^{-4}$).

Our instruments included additional independent signals in known lung function loci and variants in genes newly linked to lung function, such as *HORMAD2* at 22q12.1 (rs6006399: $P_{FEV1}=1.9\times10^{-18}$), which is involved in synapsis surveillance in meiotic prophase, and *RIPOR1* at 16q22.1 (rs7196853: $P_{FEV1/FVC}=1.3\times10^{-16}$), which plays a role in cell polarity and directional migration. Several new variants further support the importance of the tumor growth factor beta (TGF-β) signaling pathway, including *CRIM1* (rs1179500: $P_{FEV1/FVC}=3.6\times10^{-17}$) and *FGF18* (rs11745375: $P_{FEV1/FVC}=1.6\times10^{-11}$). Another novel gene, *PIEZO1* (rs750739: $P_{FEV1}=1.8\times10^{-10}$), encodes a mechano-sensory ion channel, supports adaptation to stretch of the lung epithelium and endothelium, and promotes repair after alveolar injury[26,27]. In never smokers a signal was identified at 6q15 in *BACH2* (rs58453446: $P_{FEV1/FVC\text{-}nvsmk}=8.9\times10^{-10}$), a gene required for pulmonary surfactant homeostasis. Lastly, two novel lung function variants mapped to genes somatically mutated in lung cancer: *EML4* (rs12466981: $P_{FEV1/FVC}=2.7\times10^{-14}$) and *BRAF* (rs13227429: $P_{FVC}=5.6\times10^{-9}$).

Mendelian Randomization

The causal relevance of impaired pulmonary function was investigated by applying genetic instruments developed in the UK Biobank to the OncoArray lung cancer dataset, comprised of 29,266 lung cancer cases and 56,450 controls (Supplementary Table 6). Primary analyses were based on the maximum likelihood (ML) and inverse variance weighted (IVW) multiplicative random effects estimators[28,29]. Sensitivity analyses were conducted using the weighted median (WM) and robust adjusted profile score (RAPS) estimators[30,31]. A genetically predicted decrease in FEV_1 was significantly associated with increased risk of lung cancer overall (OR_{ML}=1.28, 95% CI: 1.12-1.47, $p=3.4\times10^{-4}$) and squamous carcinoma (OR_{ML}=2.04, 1.64-2.54, $p=1.2\times10^{-10}$), but not adenocarcinoma (OR_{ML}=0.99, 0.83-1.19, p=0.96) (Figure 4; Supplementary Table 7). The association with lung cancer was not significant across all estimators (OR_{WM}=1.06, p=0.57; OR_{RAPS}=1.13, p=0.26). There was no evidence of directional pleiotropy based on the MR Egger intercept test ($\beta_{0\ Egger} \neq 0$, p<0.05), but significant heterogeneity among SNP-specific causal effect estimates was observed, which may be indicative of balanced horizontal pleiotropy (lung cancer: $P_Q=2.1\times10^{-41}$; adenocarcinoma: $P_Q=3.4\times10^{-9}$; squamous carcinoma: $P_Q=1.1\times10^{-30}$). After excluding outlier variants contributing to this heterogeneity, 36 for lung cancer and 34 for squamous carcinoma, the association with FEV_1 diminished for both phenotypes (lung cancer: $OR_{ML}=OR_{IVW}$=1.12, p=0.13), but remained statistically significant for squamous carcinoma (OR_{IVW}=1.51, 1.21-1.88, $p=2.2\times10^{-4}$), with comparable effects observed using other estimators (OR_{ML}=1.50, $p=6.7\times10^{-4}$; ORRAPS=1.48, $p=1.7\times10^{-3}$; OR_{WM}=1.44, p=0.040).

Genetic predisposition to reduced FVC was inconsistently associated with squamous carcinoma risk (OR_{ML}=1.68, $p=1.8\times10^{-4}$; OR_{WM}=1.19, p=0.38). Effects became attenuated and more similar after removing outliers (OR_{ML}=1.27, p=0.10; OR_{RAPS}=1.25, p=0.14) (Figure 4; Supplementary Table 8). A genetically predicted 10% decrease in FEV_1/FVC was associated with an elevated risk of lung cancer in some models (OR_{ML}=1.18, 1.07-1.31, $p=1.6\times10^{-3}$), but not others (OR_{WM}=1.10, p=0.30; OR_{RAPS}=1.11, p=0.14) (Figure 4;

Supplementary Table 9). The association with squamous carcinoma was also inconsistent across estimators. After removing outliers contributing to significant effect heterogeneity (lung cancer: $P_Q=1.2\times10^{-28}$; adenocarcinoma: $P_Q=3.4\times10^{-9}$; squamous carcinoma: $P_Q=5.3\times10^{-15}$), the association with adenocarcinoma strengthened (OR_{ML}=1.17, 1.01-1.35; OR_{RAPS}=1.18, 1.02-1.38), while associations for lung cancer and squamous carcinoma became attenuated.

We examined the cancer risk in never smokers, by applying genetic instruments developed specifically in this population, to 2355 cases and 7504 controls (Figure 5; Supplementary Table 10). A genetically predicted 1-SD decrease in FEV_1 and FVC was not associated with lung cancer risk in never smokers. However, a 10% reduction in FEV_1/FVC was associated with a 61% increased risk (OR_{ML}=1.61, 1.10-2.35, p=0.014; OR_{IVW}=1.60, p=0.030). Outlier filtering did not have an appreciable impact on the results (OR_{ML}=1.56, 1.05-2.30, p=0.027; OR_{IVW}=1.55, 1.05-2.28, p=0.028). A sensitivity analysis applied to 264 FEV_1/FVC instruments not specific to never smokers yielded an attenuated estimate (OR_{IVW}=1.35, 1.03-1.75, p=0.027), but confirmed the impact of FEV_1/FVC reduction on lung cancer risk.

For completeness, we also present MR estimates for the effect of impaired pulmonary function on lung cancer risk in smokers (Supplementary Table 11). Despite the larger sample size (23,223 cases and 16,964 controls) compared to never smokers, a genetically predicted 10% reduction in FEV_1/FVC was weakly and inconsistently associated with lung cancer risk (OR_{IVW}=1.15, p=0.038; OR_{RAPS}=1.08, p=0.488). Genetic predisposition to FEV_1 and FVC impairment did not appear to confer an increased risk among smokers.

Diagnostics and Sensitivity Analyses

Extensive MR diagnostics are summarized in Supplementary Table 12. All analyses used strong instruments (F-statistic > 40) and did not appear to be weakened by violations of the no measurement error (NOME) assumption (I^2_{GX} statistic >0.97). MR Steiger test[32] was used to orient the causal effects and confirmed that instruments for pulmonary function were affecting lung cancer susceptibility, not the reverse, and this direction of effect was highly robust. No instruments were removed based on Steiger filtering. We also confirmed that none of the genetic instruments were associated with nicotine dependence phenotypes ($P<1\times10^{-5}$), such as time to first cigarette, difficulty in quitting smoking, and number of quit attempts, which were available for a subset of individuals in the UKB. All MR analyses were adequately powered, with >80% power to detect a minimum OR of 1.25 for FEV_1 and FEV_1/FVC (Supplementary Figure 2). For never smokers, we had 80% power to detect a minimum OR of 1.40 for FEV_1/FVC and 1.60 for FEV_1.

Given the genetic correlation observed for pulmonary phenotypes cigarette smoking and adiposity, we conducted several sensitivity analyses to further address any potential confounding by these phenotypes. The finding for squamous carcinoma and FEV_1 was further interrogated using multivariable MR (MVMR) by incorporating genetic instruments for BMI[33] and smoking behavior[34] to estimate the direct effect of FEV_1 on squamous carcinoma risk. MVMR using all instruments yielded an OR of 1.95 (95% CI: 1.36-2.80, $p=2.8\times10^{-4}$) per 1-SD decrease in FEV_1 and an OR of 1.63 (95% CI: 1.20-2.23, $p=1.8\times10^{-3}$) after filtering outlier instruments.

We confirmed that none of the genetic instruments were associated with smoking status (ever/never), cigarette pack-years (continuous), or adiposity (body fat percentage) at the $P < 5\times10^{-8}$ level. However, several variants were associated based on a $P < 1\times10^{-5}$ threshold (25 for FEV_1 and 18 for FEV_1/FVC). We repeated MR analyses after removing these variants (Supplementary Table 13) and confirmed that our results remained robust for FEV_1 and squamous cell carcinoma (OR_{IVW}=2.02, 1.40-2.92, $p=1.9\times10^{-4}$) and FEV_1/FVC and adenocarcinoma (OR_{IVW}=1.19, 1.01-1.40, p=0.04). However, there was still significant heterogeneity among the causal effect estimates. After filtering the remaining outliers, the effect of a 10% decrease in FEV_1/FVC on adenocarcinoma strengthened (OR_{IVW}=1.24, 1.08-1.43, $p=2.4\times10^{-3}$), while estimates attenuated slightly for FEV1 and squamous carcinoma (OR_{IVW}=1.46, 1.14-1.87, $p=2.7\times10^{-3}$).

We also considered the possibility of residual confounding in our GWAS due to insufficient adjustment for smoking-related factors. We thus re-estimated SNP effects on FEV_1, FVC, and FEV_1/FVC with adjustment for continuous cigarette pack-years and years since quitting. The distribution of effect sizes did not differ between the two analyses (p>0.05), and the correlation with our original instrument weights was strong for all phenotypes (Pearson's $r \geq 0.87$, $p<1\times10^{-40}$) (Supplementary Figure 3).

Lastly, we examined the association between FEV_1 and FEV_1/FVC genetic instruments and COPD, defined as FEV_1/FVC<0.70. Among FEV_1 instruments, 64% (123 variants) were associated with COPD at p<0.05 and 16% (31 variants) at $p<5\times10^{-8}$ (Supplementary Figure 4). All instruments for FEV_1/FVC were associated with COPD at the nominal level, and 40% (105 variants) reached genome-wide significance. Using weights from estimated associations between the 105 instruments and COPD log(OR), we observed a modestly increased risk of lung adenocarcinoma (OR_{IVW}=1.08, 1.01-1.15, p=0.015), which parallels our findings based on instruments developed for the continuous FEV_1/FVC phenotype.

Functional Characterization of Lung Function Instruments

To gain insight into biological mechanisms mediating the observed effects of impaired pulmonary function on lung cancer risk, we conducted in-silico analyses of functional features associated with the genetic instruments for each lung phenotype.

Lung Tissue Gene Expression

We identified 185 statistically significant (Bonferroni p<0.05) *cis*-eQTLs for 101 genes among the genetic instruments for FEV_1 and FEV_1/FVC based on lung tissue gene expression data from the Laval biobank[35] (Supplementary File 7). Predicted expression of 7 genes was significantly ($p<5.0\times10^{-4}$) associated with lung cancer risk: *SECISBP2L*, *HLA-L*, *DISP2*, *MAPT*, *KANSL1-AS1*, *LRRC37A4P*, and *PLEKHM1* (Supplementary Figure 5). Of these, *SECISBP2L* (OR=0.80, $p=5.2\times10^{-8}$), *HLA-L* (OR=0.84, $p=1.6\times10^{-6}$), and *DISP2* (OR=1.25, $p=1.6\times10^{-4}$) displayed consistent directions of effect for pulmonary function and lung cancer risk, whereby alleles associated with increased expression were associated with impaired FEV_1 or FEV_1/FVC and increased cancer risk (or conversely, positively associated with pulmonary function and inversely associated with cancer risk). Gene expression associations with inconsistent effects are more likely to indicate pleiotropic pathways not operating primarily through pulmonary impairment. Differences by histology were observed for *SECISBP2L*, which was associated with adenocarcinoma (OR=0.54, $p=3.1\times10^{-14}$), but not squamous cell carcinoma (OR=1.05, p=0.44). Effects observed for *DISP2* (OR=1.21, p=0.021) and *HLA-L* (OR=0.90, p=0.034) were attenuated for adenocarcinoma, but not for squamous carcinoma (*DISP2*: OR=1.30, $p=6.2\times10^{-3}$; *HLA-L*: OR=0.75, $p=1.6\times10^{-6}$).

Plasma Protein Levels

A total of 70 lung function instruments were mapped to genome-wide significant ($p<5.0\times10^{-8}$) protein quantitative trait loci (pQTL) affecting the plasma levels of 64 different proteins (Supplementary File 8), based on data from the Human Plasma Proteome Atlas[36,37]. Many of these pQTL targets are involved in regulation of immune and inflammatory responses, such as interleukins (IL21, IL1R1, IL17RD, IL18R1), MHC class I polypeptide-related sequences, transmembrane glycoproteins expressed by natural killer cells, and members of the tumor necrosis receptor superfamily (TNFSF12, TNFRSF6B, TR19L). Other notable associations include NAD(P)H dehydrogenase [quinone] 1 (NQO1) a detoxification enzyme involved in protecting lung tissues in response to reactive oxidative stress (ROS) and promoting p53 stability[38]. NQO1 is a target of the NFE2-related factor 2 (NRF2), a master regulator of cellular antioxidant response that has generated considerable interest as a chemoprevention target[39,40].

Pathway Enrichment Analysis

We analyzed genes where the lung function instruments were localized using curated pathways from the Reactome database[41]. Significant enrichment (FDR q<0.05) was observed only for FEV_1/FVC instruments in never smokers, with an over-representation of pathways involved in adaptive immunity and cytokine signaling (Supplementary Figure 6). Top-ranking pathways with $q=2.2\times10^{-6}$ included translocation of ZAP-70 to immunological synapse, phosphorylation of CD3 and TCR zeta chains, and PD-1 signaling. These findings are in line with the predominance of immune-related pQTL associations. Examining all instruments for FEV1 and FEV1/FVC identified significant over-representation (FDR q<0.05) of six immunologic signatures from the ImmuneSigDB collection[42], including pathways implicated in host response to infection and immunization (Supplementary Figure 7).

DISCUSSION

Despite a substantial body of observational literature demonstrating an increased risk of lung cancer in individuals with pulmonary dysfunction[2-7,43], confounding by shared environmental risk factors and high co-occurrence of lung cancer and airflow obstruction created uncertainty regarding the causal nature of this relationship. We comprehensively investigated this by characterizing shared genetic profiles between lung cancer and lung function, and interrogated causal hypotheses using Mendelian randomization, which overcomes many limitations of observational studies. We also provide insight into biological pathways underlying the observed associations by incorporating functional annotations into heritability analyses, assessing eQTL and pQTL effects of lung function instruments, and conducting pathway enrichment analyses.

The large sample size of the UK Biobank allowed us to successfully create instruments for three pulmonary function phenotypes, FEV_1, FEV_1/FVC, and FVC. Although these phenotypes are closely related, they capture different aspects of pulmonary impairment, with FEV_1 and FEV_1/FVC used for diagnostic purposes in clinical setting. Our genetic instruments captured known and novel mechanisms involved in pulmonary function. Of the 73 novel variants identified here, many were in loci implicated in immune-related functions and pathologies. Examples include *HORMAD2*, which has been previously linked to inflammatory bowel disease[44,45] and tonsillitis[46], and *RIPOR1* (also known as *FAM65A*), which is part of a gene expression signature for atopy[47]. *PIEZO1* is primarily involved in mechano-transduction and tissue differentiation during embryonic development[48-50], however recent evidence has emerged delineating its role in optimal T-cell receptor activation and immune regulation[51]. *BACH2*, the new signal for FEV_1/FVC in never smokers, is involved in alveolar macrophage function[52], as well as selection-mediated *TP53* regulation and checkpoint control[53]. The lead variant identified here is independent (r^2<0.05) of *BACH2* loci nominally associated with lung function decline in a candidate gene study of COPD patients[54], suggesting there may be differences in the genetic architecture of pulmonary traits in never smokers.

Our genetic correlation analyses indicate shared genetic determinants between pulmonary function with anthropometric traits and cigarette smoking. Our results are in contrast with the recent findings of Wyss et al.[24], who did not observe statistically significant genetic correlations for any pulmonary function phenotypes with height and smoking, as well FVC and FEV_1/FVC, using publicly available summary statistics from the UKB and other studies of European ancestry individuals. In this respect, assessing genetic correlation within a single well-characterized population provides improved power while minimizing potential for bias and heterogeneity when combining data from multiple sources.

We observed statistically significant genetic correlations between pulmonary function impairment and lung cancer susceptibility for all lung cancer subtypes, except for never smokers. Reduced FEV_1/FVC was significantly correlated with increased risk of lung cancer overall, squamous cell carcinoma, and

adenocarcinoma. Significant genetic correlations with FEV_1 and FVC were observed for lung cancer overall, in smokers, and for tumors with squamous cell histology, but not adenocarcinoma. Jiang et al.[25] reported a similar magnitude of genetic correlation with FEV_1/FVC, but did not observe an association with FVC, and did not assess FEV_1. Differences in our results may be attributable to their use of GWAS summary statistics for pulmonary phenotypes from the interim UK Biobank release. Our findings demonstrate substantial overlap in the genetic architecture of obstructive and neoplastic lung disease, particularly for highly conserved variants that are likely to be subject to natural selection, and super enhancers. However, genetic correlations do not support a causal interpretation, especially considering the shared heritability with potentially confounding traits, such as smoking and obesity.

On the other hand, Mendelian randomization analyses revealed histology-specific effects of reduced FEV_1 and FEV_1/FVC on lung cancer susceptibility, suggesting that these indicators of impaired pulmonary function may be causal risk factors. Genetic predisposition to FEV_1 impairment conferred an increased risk of lung cancer overall, particularly for squamous carcinoma. This relationship persisted after filtering potentially pleiotropic instruments and performing other sensitivity analyses, including multivariable Mendelian randomization and manual filtering of variants associated with smoking or adiposity. FEV_1/FVC reduction appeared to increase the risk of lung adenocarcinoma, as well as lung cancer among never-smokers. The latter finding is particularly compelling since it precludes confounding by smoking-related factors and demonstrates an association with the most clinically relevant pulmonary phenotype. The increased lung cancer risk in never smokers was also observed using genetic instruments developed specifically in never smokers and in sensitivity analyses using instruments from the population that also includes smokers. We hypothesize that the effects of pulmonary obstruction are mediated by chronic inflammation and immune response, which is supported by the overrepresentation of adaptive immunity and cytokine signaling pathways and pQTL effects among FEV_1 and FEV_1/FVC instruments.

Examining lung eQTL effects of our genetic instruments identified additional relevant mechanisms, including gene expression of *SECISBP2L* and *DISP2*. *SECISBP2L* at 15q21 is essential for ciliary function[55] and has an inhibitory effect on lung tumor growth by suppressing cell proliferation and inactivation of Aurora kinase A[56]. This gene was among several novel susceptibility regions identified in the most recent lung cancer GWAS[16], and now we more conclusively establish impaired pulmonary function as the mechanism mediating *SECISBP2L* effects on risk of lung cancer overall, particularly adenocarcinoma. Less is known about *DISP2*, although it has been implicated in the conserved Hedgehog signaling pathway essential for embryonic development and cell differentiation[57].

One of the main challenges and outstanding questions in previous epidemiologic studies has been clarifying how smoking fits into the causal pathway between impaired pulmonary function and lung cancer risk. Are indicators of airway obstruction simply proxies for smoking-induced carcinogenesis? The association between reduced FEV_1/FVC and risk of adenocarcinoma and lung cancer in never smokers observed in our Mendelian randomization analysis and in previous studies[8,9], argues against this simplistic explanation and points to alternative pathways. Chronic airway inflammation fosters a lung microenvironment with altered signaling pathways, aberrant expression of cytokines, chemokines, growth factors, and DNA damage-promoting agents, all of which promote cancer initaiton[15]. This mechanism may be particularly relevant for adenocarcinoma, which is the most common lung cancer histology in never smokers, arising from the peripheral alveolar epithelium that has less direct contact with inhaled carcinogens.

Dysregulated immune function is a hallmark of lung cancer and COPD, with both diseases sharing similar inflammatory cell profiles characterized by macrophages, neutrophils, and CD4+ and CD8+ lymphocytes. Immune cells in COPD and emphysema exhibit T helper 1 (Th1)/Th17 polarization, decreased programmed death ligand-1 (PD-L1) expression in alveolar macrophages, and increased production of interferon (IFN)-γ by CD8+ T cells[58], a phenotype believed to prevail at tumor initiation, whereas established tumors are dominated by Th2/M2-like macrophages[59]. These putative mechanisms were highlighted in our pathway

analysis, with an enrichment of genes involved in INF-γ, PD-1 and IL-1 signaling among FEV_1/FVC genetic instruments, and over-representation of pQTL targets in these pathways. Furthermore, a study of trans-thoracically implanted tumors in an emphysema mouse model demonstrates how this pulmonary phenotype results in impaired antitumor T cell responses at a critical point when nascent cancer cells evade detection and elimination by the immune system resulting in enhanced tumor growth[60].

Other relevant pathways implicating pulmonary dysfunction in lung cancer development include lung tissue destruction via matrix degrading enzymes and increased genotoxic and apoptotic stress resulting from cigarette smoke in conjunction with macrophage- and neutrophil-derived ROS[15,61]. This may explain our findings for FEV_1 and squamous carcinoma, for which cigarette smoking is a particularly dominant risk factor. Genetic predisposition to impaired FEV_1 may create a milieu that promotes malignant transformation and susceptibility to external carcinogens and tissue damage, rather than increasing the likelihood of cigarette smoking. In our analysis we attempted to isolate the former pathway from the latter by carefully instrumenting pulmonary phenotypes and confirming that they are not associated with behavioral aspects of nicotine dependence. However, residual confounding by smoking cannot be entirely precluded, given its high genetic and phenotypic correlation with FEV_1.

The causal interpretation of our results critically depends on the validity of fundamental Mendelian randomization assumptions. We employed a range of estimation techniques with different underlying assumptions, as well as diagnostic tests, to interrogate the robustness of our results with respect to confounding, horizontal pleiotropy, and weak instrument bias. However, despite these efforts, residual confounding by related phenotypes, such as smoking, or subtle effects of population structure cannot be ruled out. In evaluating the contribution of our findings, several limitations should be acknowledged. Our approach to outlier removal based on Cochran's Q statistic with modified second order weights may have been overly stringent; however, manually pruning based on such a large set of genetic instruments may not be feasible and may introduce additional bias, thus we feel this systematic conservative approach is justified. Furthermore, outlier removal did not have an adverse impact on instrument strength and precision of the MR analysis.

In addition to pleiotropy, selection bias may also undermine the validity of a Mendelian Randomization study, particularly in the form of collider bias, if selection is a function of the exposure or outcome. In the context of the UKB, low participation (5-10%) may have resulted in an unrepresentative study population[62-64]. Although enrolment in the cohort was not explicitly contingent on cancer status or pulmonary function, it is likely that individuals who did not complete a spirometry assessment were more likely to be smokers and have poor lung function. Simulations by Gkatzionis & Burgess[64] demonstrate that when the effect of a risk factor on selection is mild to moderate (odds of selection: 0.82 to 0.61), the type I error rate remains reasonable at 5.0-6.6%. The direction of the resulting bias depends on the direction and strength of the exposure (lung function) – confounder (smoking) relationship. In the context of our study, the causal effect may be underestimated since the confounder and exposure are both likely to increase non-participation or result in missing spirometry data.

Another limitation is that we did not assess the relationship between the velocity of lung function decline and lung cancer risk, which may also prove to be a risk factor and capture a different dimension of pulmonary dysfunction. Furthermore, since our study includes the largest GWAS of lung cancer cases in never smokers, this precludes a well-powered replication study in an independent European ancestry population. In addition, dichotomous stratification by smoking status does not permit an evaluation of the relationship between pulmonary impairment and lung cancer risk across more granular levels of smoking. Lastly, in our efforts to present the most comprehensive assessment of pulmonary function impairment and lung cancer risk, a number of analyses were conducted, and it may be possible that some inconsistently observed associations were due to chance.

Despite these limitations, important strengths of this work include the large sample size for instrument development and causal hypothesis testing. Our Mendelian randomization approach leveraged a large number of novel genetic instruments, including variants specifically associated with lung function in never smokers, while balancing the concerns related to genetic confounding and pleiotropy. By triangulating evidence from gene expression and plasma protein levels, we also provide a more enriched interpretation of the genetic effects of pulmonary function loci on lung cancer risk, which implicate immune-mediated pathways. Despite the small individual SNP effect sizes, combining multiple instruments revealed meaningful increases in lung cancer risk. A genetically predicted 10% reduction in FEV_1/FVC confers an approximately 55% increased risk of lung cancer in never smokers, and a similar magnitude of effect was observed for FEV_1 and squamous carcinoma. However, effects of FEV_1/FVC on adenocarcinoma were more modest (16-23% increase). Taken together, these findings provide more robust etiological insight than previous studies that relied on using observed lung function phenotypes directly as putatively casual factors.

As our understanding of the shared genetic and molecular pathways between lung cancer and pulmonary disease continues to evolve, identification of new susceptibility loci for pulmonary function and lung cancer risk may have important implications for future precision prevention and screening endeavors. Multiple genetic determinants of lung function are in pathways that contain druggable targets, based on our pQTL findings and previous reports[18], which may open new avenues for chemoprevention or targeted therapies for lung cancers with an obstructive pulmonary etiology. In addition, with accumulating evidence supporting the effectiveness of low-dose computed tomography for lung cancer[65,66], impairment in FEV_1 and FEV_1/FVC and their genetic determinants may provide additional information for refining risk-stratification and screening eligibility criteria.

METHODS

Study Populations

The UK Biobank (UKB) is a population-based prospective cohort of over 500,000 individuals aged 40-69 years at enrollment in 2006-2010 who completed extensive questionnaires on health-related factors, physical assessments, and provided blood samples[67]. Participants were genotyped on the UK Biobank Axiom array (89%) or the UK BiLEVE array (11%)[67]. Genotype imputation was performed using the Haplotype Reference Consortium data as the main reference panel as well as using the merged UK10K and 1000 Genomes (1000G) phase 3 reference panels[67]. Our analyses were restricted to individuals of predominantly European ancestry based on self-report and after excluding samples with either of the first two genetic ancestry principal components (PCs) outside of 5 standard deviations (SD) of the population mean. Samples with discordant self-reported and genetic sex were removed. Using a subset of genotyped autosomal variants with minor allele frequency (MAF)≥0.01 and call rate ≥97%, we filtered samples with call rates <97% or heterozygosity >5 standard deviations (SD) from the mean. First-degree relatives were identified using KING[68] and one sample from each pair was excluded, leaving at total of 413,810 individuals available for analysis.

We further excluded 36,461 individuals without spirometry data, 207 individuals who only completed one blow (n=207), for whom reproducibility could not be assessed (Supplementary Figure 1). For the remaining subjects, we examined the difference between the maximum value per individual (referred to as the best measure) and all other blows. Values differing by more than 0.15L were considered non-reproducible, based on standard spirometry guidelines[69], and were excluded. Our analyses thus included 372,750 and 370,638 individuals for of FEV_1 and FVC, respectively. The best per individual measure among the reproducible blows was used to derive FEV_1/FVC, resulting in 368,817 individuals. FEV_1 and FVC values were then converted to standardized Z-scores with a mean of 0 and standard deviation (SD) of 1.

The OncoArray Lung Cancer study has been previously described[16]. Briefly, this dataset consists of genome-wide summary statistics based on 29,266 lung cancer cases (11,273 adenocarcinoma, 7426 squamous carcinoma) and 56,450 controls of predominantly European ancestry (≥80%) assembled from studies part of the International Lung Cancer Consortium. Summary statistics from the lung cancer GWAS were adjusted for appropriate covariates, including genetic ancestry PCs, and showed no signs of genomic inflation for lung cancer overall (λ_{GC}=1.0035) or for any subtypes, including adenocarcinoma (λ_{GC}=1.0050), squamous carcinoma (λ_{GC}=1.0051), and lung cancer in never smokers (λ_{GC}=1.0060).

Genome-Wide Association Analysis

Genome-wide association analyses of pulmonary function phenotypes in the UK Biobank cohort were conducted using PLINK 2.0. We excluded variants out of with Hardy-Weinberg equilibrium at $p<1\times10^{-5}$ in cancer-free individuals, call rate <95% (alternate allele dosage required to be within 0.1 of the nearest hard call to be non-missing), imputation quality INFO<0.30, and MAF<0.005. To minimize potential for reverse causation, prevalent lung cancer cases, defined as diagnoses occurring up to 5 years before cohort entry and incident cases occurring within 2 years of enrollment, were excluded (n=738). Linear regression models for pulmonary function phenotypes (standardized Z-scores for FEV_1 and FVC; untransformed FEV_1/FVC ratio bounded by 0 and 1) were adjusted for age, age^2, sex, genotyping array and 15 PCs to permit an assessment of heritability (h_g) and genetic correlation (r_g) with height, smoking (status and pack-years), and anthropometric traits[70].

Heritability and Genetic Correlation

LD Score regression[17] was used to estimate h_g for each lung phenotype and r_g with lung cancer and other traits. To better capture LD patterns present in the UKB data, we generated LD scores for all variants with MAF>0.0001 (n=7,567,036 variants) using a random sample of 10,000 UKB participants. UKB LD scores were used to estimate h_g for each lung phenotype and r_g with other non-cancer traits. Genetic correlation with lung cancer was estimated using publicly available LD scores based on the 1000G phase 3 reference population (n=1,095,408 variants).

To assess the importance of specific functional annotations in SNP-heritability, we partitioned trait-specific heritability using stratified-LDSC[71]. The analysis was performed using 86 annotations (baseline-LD model v2.1), which incorporated MAF-adjustment and other LD-related annotations, such as predicted allele age and recombination rate[20,22]. The MHC region was excluded from partitioned heritability analyses. Enrichment was considered statistically significant if $p<8.5\times10^{-4}$, which reflects Bonferroni correction for 59 annotations (functional categories with and without a 500 bp window around it were considered as the same annotation).

Development of Genetic Instruments for Pulmonary Function

For the purpose of instrument development, a two-stage genome-wide analysis was employed, with a randomly sampled 70% of the cohort used for discovery and the remaining 30% reserved for replication. In addition to age, age^2, sex, genotyping array and 15 PC's, models were adjusted for covariates that explain a substantial proportion of variation in pulmonary phenotypes, such as smoking and height, in order to decrease the residual variance and help isolate the relevant genetic signals. Specifically, we adjusted for height, height^2, and cigarette pack-year categories (0, corresponding to never-smokers, >0-10, >10-20, >20-30, >30-40, and >40). Other covariates, such as UKB assessment center (Field 54), use of an inhaler prior to spirometry (Field 3090), and blow acceptability (Field 3061) were considered. However, these covariates did not explain a substantial proportion of phenotype variation and had low variable importance metrics (Img<0.01), and thus were not included in our final models. Instruments were selected from independent associated variants (LD r^2<0.05 in a clumping window of 10,000 kb) with $P<5\times10^{-8}$ in the

discovery stage and $P<0.05$ and consistent direction of effect in the replication stage. Since the primary goal of our GWAS was to develop a comprehensive set of genetic instruments we applied a less stringent replication threshold in anticipation of subsequent filtering based on potential violation of Mendelian randomization assumptions.

Mendelian Randomization

Mendelian randomization (MR) analyses were carried out to investigate the potential causal relationship between impaired pulmonary function and lung cancer risk. Genetic instruments excluded multi-allelic and non-inferable palindromic variants with intermediate allele frequencies (MAF>0.42). Odds ratios (OR) and corresponding 95% confidence intervals were obtained using the maximum likelihood and inverse variance weighted multiplicative random-effects (IVW-RE) estimators[28,29]. Effects for FEV_1 and FVC were estimated for a genetically predicted 1-SD decrease in the standardized Z-score. For FEV_1/FVC, we modeled cancer risk corresponding to a 10% decrease in the ratio. Sensitivity analyses included the weighted median (WM) estimator[30], which provides unbiased estimates when up to 50% of the weights are from invalid instruments, and MR RAPS (Robust Adjusted Profile Score), which incorporates random effect and robust loss functions to limit the influence of potentially pleiotropic instruments. MR RAPS assumes balanced (mean 0) horizontal pleiotropy. In contrast to IVW-RE, MR RAPS models idiosyncratic and systematic pleiotropy effects as additive, rather than multiplicative[31]. Using MR estimation techniques with different underlying statistical models allows for a more comprehensive assessment of the robustness of our results with respect to violations of MR assumptions. We also applied the following diagnostic tests: i) significant ($p<0.05$) deviation of the MR Egger intercept ($\beta_{0\ Egger}$) from 0, as a test for directional pleiotropy[72]; ii) I^2_{GX} statistic <0.90 indicative of regression dilution bias and inflation in the MR Egger pleiotropy test due to violation of the no measurement error (NOME) assumption[72]; iii) Cochran's Q-statistic with modified second order weights to asses heterogeneity (p-value<0.05) indicative of (balanced) horizontal pleiotropy[73].

Functional Characterization of Lung Function Instruments

In order to characterize functional pathways that are represented by the genetic instruments for FEV_1 and FEV_1/FVC, we examined effects on gene expression in lung tissues from 409 subjects from the Laval eQTL study[35]. Lung function instruments with significant (Bonferroni p-value<0.05) eQTL effects were used as instruments to estimate the effect of the gene expression on lung cancer risk. For genes with multiple eQTLs, independent variants (LD $r^2<0.05$) were used to obtain IVW estimates of the predicted effects of increased gene expression on lung cancer risk. For genes with a single eQTL, OR estimates were obtained using the Wald method. Next, we examined data from the genetic atlas of the human plasma proteome[36,37], queried using PhenoScanner[74], to assess whether any of the genetic instruments for FEV_1 and FEV_1/FVC had significant ($p<5\times10^{-8}$) effects on intracellular protein levels. Lastly, we summarized the pathways represented by the genes where the lung function instruments were localized using pathway enrichment analysis via the Reactome database[41] and ImmuneSigDB (collection C7 from MSigDB)[42].

URLs

PLINK 2.0: https://www.cog-genomics.org/plink/2.0/

LDSC functional annotations available from:

https://data.broadinstitute.org/alkesgroup/LDSCORE/1000G_Phase3_EUR_baselineLD_v2.1_ldscores.tgz

R package for Mendelian Randomization: https://github.com/MRCIEU/TwoSampleMR

R package for PhenoScanner: https://github.com/phenoscanner/phenoscanner

R packages for pathway analysis: https://bioconductor.org/packages/release/bioc/html/ReactomePA.html and https://bioconductor.org/packages/release/bioc/html/clusterProfiler.html

DATA AVAILABILITY

The datasets generated during and/or analyzed during the current study are available from the authors on request. Lung cancer: genotype data for lung cancer are available at the database of Genotypes and Phenotypes (dbGaP) under accession phs001273.v1.p1. Readers interested in obtaining a copy of the original data can do so by completing the proposal request form at http://oncoarray.dartmouth.edu/. The UK Biobank in an open access resource. This research was conducted with approved access to UK Biobank data under applications number 14105 and 23261.

ACKNOWEDGEMENTS

This research was supported by funding from the National Institutes of Health (US NCI R25T CA112355 and R01 CA201358; PI: Witte and Canada Research Chair (PI: Hung).

The OncoArray project was supported by NIH U19 CA203654 (MPI: Hung, Amos, Brennan).

The lung eQTL study at Laval University was supported by the Fondation de l'Institut universitaire de cardiologie et de pneumologie de Québec and the Canadian Institutes of Health Research (MOP - 123369). Y.B. holds a Canada Research Chair in Genomics of Heart and Lung Diseases.

Richard M. Martin is supported by a CRUK programme grant, the Integrative Cancer Epidemiology Programme (C18281/A19169), and by the National Institute for Health Research (NIHR) Bristol Biomedical Research Centre based at University Hospitals Bristol NHS Foundation Trust and the University of Bristol. The views expressed in this publication are those of the authors and not necessarily those of the NHS, the National Institute for Health Research or the Department of Health.

The CARET study was supported by the National Institutes of Health / National Cancer Institute: UM1 CA167462 (PI: Goodman), U01 CA6367307 (PIs: Omen, Goodman); R01 CA111703 (PI: Chen), 5R01 CA151989-01A1 (PI: Doherty) and U01 CA167462 (PI: Chen).

Michael P. Davies is supported by the Roy Castle Lung Cancer Foundation UK.

References

1. Ferlay, J. *et al.* Estimating the global cancer incidence and mortality in 2018: GLOBOCAN sources and methods. *Int J Cancer* **144**, 1941-1953 (2019).
2. Wasswa-Kintu, S., Gan, W.Q., Man, S.F., Pare, P.D. & Sin, D.D. Relationship between reduced forced expiratory volume in one second and the risk of lung cancer: a systematic review and meta-analysis. *Thorax* **60**, 570-5 (2005).
3. Calabro, E. *et al.* Lung function predicts lung cancer risk in smokers: a tool for targeting screening programmes. *Eur Respir J* **35**, 146-51 (2010).
4. Fry, J.S., Hamling, J.S. & Lee, P.N. Systematic review with meta-analysis of the epidemiological evidence relating FEV1 decline to lung cancer risk. *BMC Cancer* **12**, 498 (2012).
5. Mannino, D.M., Aguayo, S.M., Petty, T.L. & Redd, S.C. Low lung function and incident lung cancer in the United States: data From the First National Health and Nutrition Examination Survey follow-up. *Arch Intern Med* **163**, 1475-80 (2003).
6. Young, R.P. *et al.* COPD prevalence is increased in lung cancer, independent of age, sex and smoking history. *Eur Respir J* **34**, 380-6 (2009).
7. Zhai, R., Yu, X., Wei, Y., Su, L. & Christiani, D.C. Smoking and smoking cessation in relation to the development of co-existing non-small cell lung cancer with chronic obstructive pulmonary disease. *Int J Cancer* **134**, 961-70 (2014).
8. Brenner, D.R., McLaughlin, J.R. & Hung, R.J. Previous lung diseases and lung cancer risk: a systematic review and meta-analysis. *PLoS One* **6**, e17479 (2011).
9. Brenner, D.R. *et al.* Previous lung diseases and lung cancer risk: a pooled analysis from the International Lung Cancer Consortium. *Am J Epidemiol* **176**, 573-85 (2012).
10. Denholm, R. *et al.* Is previous respiratory disease a risk factor for lung cancer? *Am J Respir Crit Care Med* **190**, 549-59 (2014).
11. Durham, A.L. & Adcock, I.M. The relationship between COPD and lung cancer. *Lung Cancer* **90**, 121-7 (2015).
12. Yang, I.A., Holloway, J.W. & Fong, K.M. Genetic susceptibility to lung cancer and co-morbidities. *J Thorac Dis* **5 Suppl 5**, S454-62 (2013).
13. Young, R.P. *et al.* Chromosome 4q31 locus in COPD is also associated with lung cancer. *Eur Respir J* **36**, 1375-82 (2010).
14. 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).
15. Houghton, A.M. Mechanistic links between COPD and lung cancer. *Nat Rev Cancer* **13**, 233-45 (2013).
16. McKay, J.D. *et al.* Large-scale association analysis identifies new lung cancer susceptibility loci and heterogeneity in genetic susceptibility across histological subtypes. *Nat Genet* **49**, 1126-1132 (2017).
17. Bulik-Sullivan, B.K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet* **47**, 291-5 (2015).
18. Shrine, N. *et al.* New genetic signals for lung function highlight pathways and chronic obstructive pulmonary disease associations across multiple ancestries. *Nat Genet* **51**, 481-493 (2019).
19. Siepel, A. *et al.* Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res* **15**, 1034-50 (2005).
20. Gazal, S. *et al.* Functional architecture of low-frequency variants highlights strength of negative selection across coding and non-coding annotations. *Nat Genet* **50**, 1600-1607 (2018).
21. McVicker, G., Gordon, D., Davis, C. & Green, P. Widespread genomic signatures of natural selection in hominid evolution. *PLoS Genet* **5**, e1000471 (2009).
22. Gazal, S. *et al.* Linkage disequilibrium-dependent architecture of human complex traits shows action of negative selection. *Nat Genet* **49**, 1421-1427 (2017).
23. Roadmap Epigenomics, C. *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317-30 (2015).
24. Wyss, A.B. *et al.* Multiethnic meta-analysis identifies ancestry-specific and cross-ancestry loci for pulmonary function. *Nat Commun* **9**, 2976 (2018).
25. Jiang, X. *et al.* Shared heritability and functional enrichment across six solid cancers. *Nat Commun* **10**, 431 (2019).

26. Gudipaty, S.A. *et al.* Mechanical stretch triggers rapid epithelial cell division through Piezo1. *Nature* **543**, 118-121 (2017).
27. Zhong, M., Komarova, Y., Rehman, J. & Malik, A.B. Mechanosensing Piezo channels in tissue homeostasis including their role in lungs. *Pulm Circ* **8**, 2045894018767393 (2018).
28. Burgess, S., Butterworth, A. & Thompson, S.G. Mendelian randomization analysis with multiple genetic variants using summarized data. *Genet Epidemiol* **37**, 658-65 (2013).
29. Bowden, J. *et al.* A framework for the investigation of pleiotropy in two-sample summary data Mendelian randomization. *Stat Med* **36**, 1783-1802 (2017).
30. Bowden, J., Davey Smith, G., Haycock, P.C. & Burgess, S. Consistent Estimation in Mendelian Randomization with Some Invalid Instruments Using a Weighted Median Estimator. *Genet Epidemiol* **40**, 304-14 (2016).
31. Zhao, Q., Wang, J., Hemani, G., Bowden, J. & Small, D.S. Statistical inference in two-sample summary-data Mendelian randomization using robust adjusted profile score. *Ann Stat*, arXiv:1804.07371v3 (2018).
32. Hemani, G., Tilling, K. & Davey Smith, G. Orienting the causal relationship between imprecisely measured traits using GWAS summary data. *PLoS Genet* **13**, e1007081 (2017).
33. Locke, A.E. *et al.* Genetic studies of body mass index yield new insights for obesity biology. *Nature* **518**, 197-206 (2015).
34. Tobacco & Genetics, C. Genome-wide meta-analyses identify multiple loci associated with smoking behavior. *Nat Genet* **42**, 441-7 (2010).
35. Hao, K. *et al.* Lung eQTLs to help reveal the molecular underpinnings of asthma. *PLoS Genet* **8**, e1003029 (2012).
36. Sun, B.B. *et al.* Genomic atlas of the human plasma proteome. *Nature* **558**, 73-79 (2018).
37. Yao, C. *et al.* Genome-wide mapping of plasma protein QTLs identifies putatively causal genes and pathways for cardiovascular disease. *Nat Commun* **9**, 3268 (2018).
38. Asher, G., Lotem, J., Cohen, B., Sachs, L. & Shaul, Y. Regulation of p53 stability and p53-dependent apoptosis by NADH quinone oxidoreductase 1. *Proc Natl Acad Sci U S A* **98**, 1188-93 (2001).
39. Sporn, M.B. & Liby, K.T. NRF2 and cancer: the good, the bad and the importance of context. *Nat Rev Cancer* **12**, 564-71 (2012).
40. Rojo de la Vega, M., Chapman, E. & Zhang, D.D. NRF2 and the Hallmarks of Cancer. *Cancer Cell* **34**, 21-43 (2018).
41. Yu, G. & He, Q.Y. ReactomePA: an R/Bioconductor package for reactome pathway analysis and visualization. *Mol Biosyst* **12**, 477-9 (2016).
42. Godec, J. *et al.* Compendium of Immune Signatures Identifies Conserved and Species-Specific Biology in Response to Inflammation. *Immunity* **44**, 194-206 (2016).
43. Wilson, D.O. *et al.* Association of radiographic emphysema and airflow obstruction with lung cancer. *Am J Respir Crit Care Med* **178**, 738-44 (2008).
44. Franke, A. *et al.* Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci. *Nat Genet* **42**, 1118-25 (2010).
45. Imielinski, M. *et al.* Common variants at five new loci associated with early-onset inflammatory bowel disease. *Nat Genet* **41**, 1335-40 (2009).
46. Feenstra, B. *et al.* Genome-wide association study identifies variants in HORMAD2 associated with tonsillectomy. *J Med Genet* **54**, 358-364 (2017).
47. Howrylak, J.A. *et al.* Gene expression profiling of asthma phenotypes demonstrates molecular signatures of atopy and asthma control. *J Allergy Clin Immunol* **137**, 1390-1397 e6 (2016).
48. Li, J. *et al.* Piezo1 integration of vascular architecture with physiological force. *Nature* **515**, 279-282 (2014).
49. Lewis, A.H., Cui, A.F., McDonald, M.F. & Grandl, J. Transduction of Repetitive Mechanical Stimuli by Piezo1 and Piezo2 Ion Channels. *Cell Rep* **19**, 2572-2585 (2017).
50. Andolfo, I. *et al.* Multiple clinical forms of dehydrated hereditary stomatocytosis arise from mutations in PIEZO1. *Blood* **121**, 3925-35, S1-12 (2013).
51. Liu, C.S.C. *et al.* Cutting Edge: Piezo1 Mechanosensors Optimize Human T Cell Activation. *J Immunol* **200**, 1255-1260 (2018).
52. Nakamura, A. *et al.* Transcription repressor Bach2 is required for pulmonary surfactant homeostasis and alveolar macrophage function. *J Exp Med* **210**, 2191-204 (2013).

53. Swaminathan, S. *et al.* BACH2 mediates negative selection and p53-dependent tumor suppression at the pre-B cell receptor checkpoint. *Nat Med* **19**, 1014-22 (2013).
54. Sandford, A.J. *et al.* NFE2L2 pathway polymorphisms and lung function decline in chronic obstructive pulmonary disease. *Physiol Genomics* **44**, 754-63 (2012).
55. Boldt, K. *et al.* An organelle-specific protein landscape identifies novel diseases and molecular mechanisms. *Nat Commun* **7**, 11491 (2016).
56. Yu, C.T. *et al.* The novel protein suppressed in lung cancer down-regulated in lung cancer tissues retards cell proliferation and inhibits the oncokinase Aurora-A. *J Thorac Oncol* **6**, 988-97 (2011).
57. Katoh, Y. & Katoh, M. Hedgehog signaling pathway and gastric cancer. *Cancer Biol Ther* **4**, 1050-4 (2005).
58. Grumelli, S. *et al.* An immune basis for lung parenchymal destruction in chronic obstructive pulmonary disease and emphysema. *PLoS Med* **1**, e8 (2004).
59. Conway, E.M. *et al.* Macrophages, Inflammation, and Lung Cancer. *Am J Respir Crit Care Med* **193**, 116-30 (2016).
60. Kerdidani, D. *et al.* Cigarette Smoke-Induced Emphysema Exhausts Early Cytotoxic CD8(+) T Cell Responses against Nascent Lung Cancer Cells. *J Immunol* **201**, 1558-1569 (2018).
61. Haqqani, A.S., Sandhu, J.K. & Birnboim, H.C. Expression of interleukin-8 promotes neutrophil infiltration and genetic instability in mutatect tumors. *Neoplasia* **2**, 561-8 (2000).
62. Watts, G. UK Biobank gets 10% response rate as it starts recruiting volunteers. *BMJ* **334**, 659 (2007).
63. Manolio, T.A. *et al.* New models for large prospective studies: is there a better way? *Am J Epidemiol* **175**, 859-66 (2012).
64. Gkatzionis, A. & Burgess, S. Contextualizing selection bias in Mendelian randomization: how bad is it likely to be? *Int J Epidemiol* (2018).
65. National Lung Screening Trial Research, T. *et al.* Reduced lung-cancer mortality with low-dose computed tomographic screening. *N Engl J Med* **365**, 395-409 (2011).
66. De Koning, H., Van Der Aalst, C., Ten Haaf, K. & Oudkerk, M. PL02.05 Effects of Volume CT Lung Cancer Screening: Mortality Results of the NELSON Randomised-Controlled Population Based Trial. *Journal of Thoracic Oncology* **13**, S185 (2018).
67. Bycroft, C. *et al.* The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203-209 (2018).
68. Manichaikul, A. *et al.* Robust relationship inference in genome-wide association studies. *Bioinformatics* **26**, 2867-73 (2010).
69. Miller, M.R. *et al.* Standardisation of spirometry. *Eur Respir J* **26**, 319-38 (2005).
70. Aschard, H., Vilhjalmsson, B.J., Joshi, A.D., Price, A.L. & Kraft, P. Adjusting for heritable covariates can bias effect estimates in genome-wide association studies. *Am J Hum Genet* **96**, 329-39 (2015).
71. Finucane, H.K. *et al.* Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat Genet* **47**, 1228-35 (2015).
72. Bowden, J. *et al.* Assessing the suitability of summary data for two-sample Mendelian randomization analyses using MR-Egger regression: the role of the I2 statistic. *Int J Epidemiol* **45**, 1961-1974 (2016).
73. Bowden, J. *et al.* Improving the accuracy of two-sample summary-data Mendelian randomization: moving beyond the NOME assumption. *Int J Epidemiol* (2018).
74. Kamat, M.A. *et al.* PhenoScanner V2: an expanded tool for searching human genotype-phenotype associations. *Bioinformatics* (2019).

Figure 1: Functional partitioning of array-based heritability for each pulmonary function phenotype, based on genome-wide summary statistics from the UK Biobank cohort. The upper panel depicts the magnitude of category-specific enrichment and corresponding -$\log_{10}$(p-value) for 22 distinct functional annotations that were significantly enriched for all three phenotypes (FEV_1, FVC, FEV_1/FVC). The lower panel shows the proportion of heritability, Pr(h_g), accounted for by each functional annotation. Functional annotation categories are not mutually exclusive.

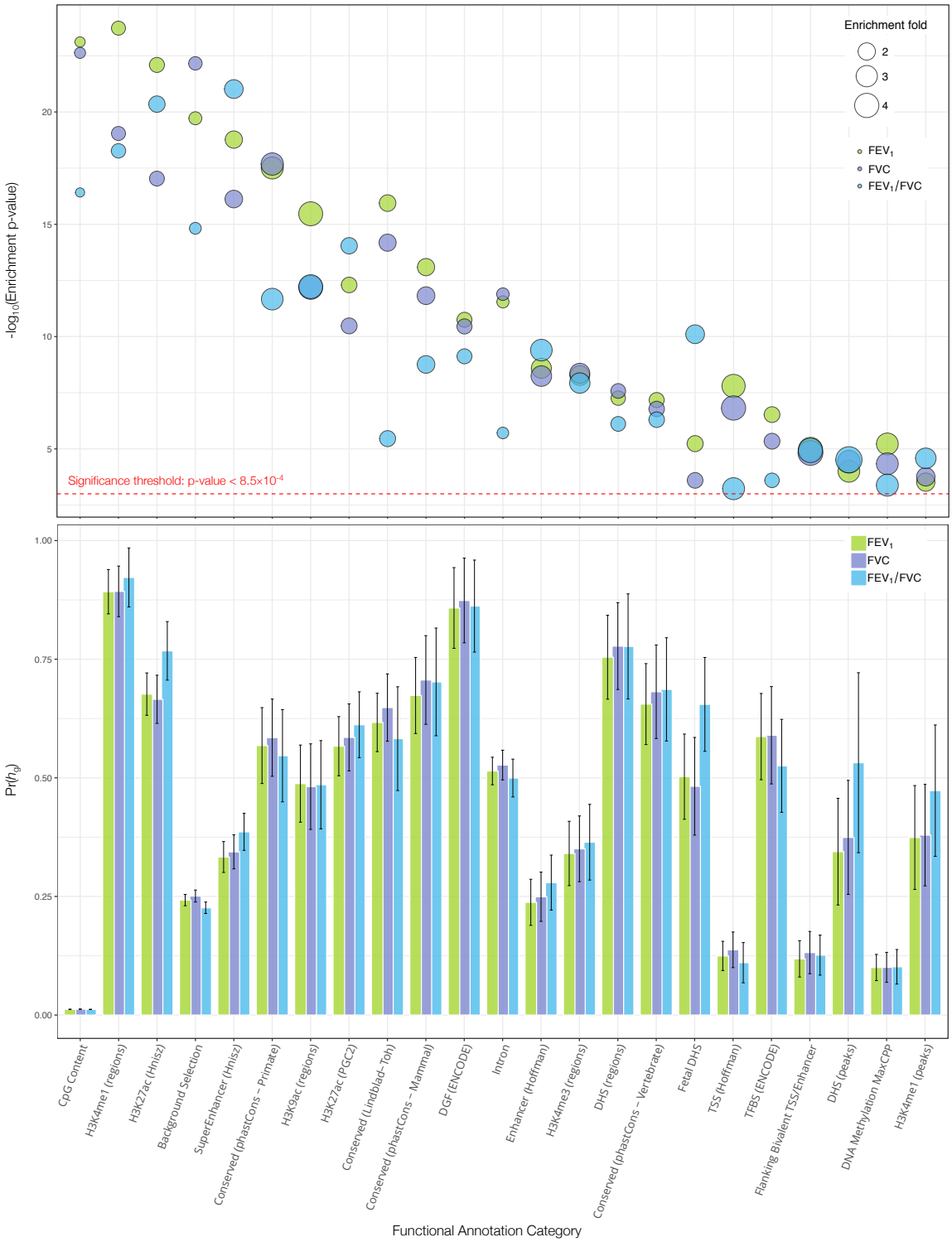
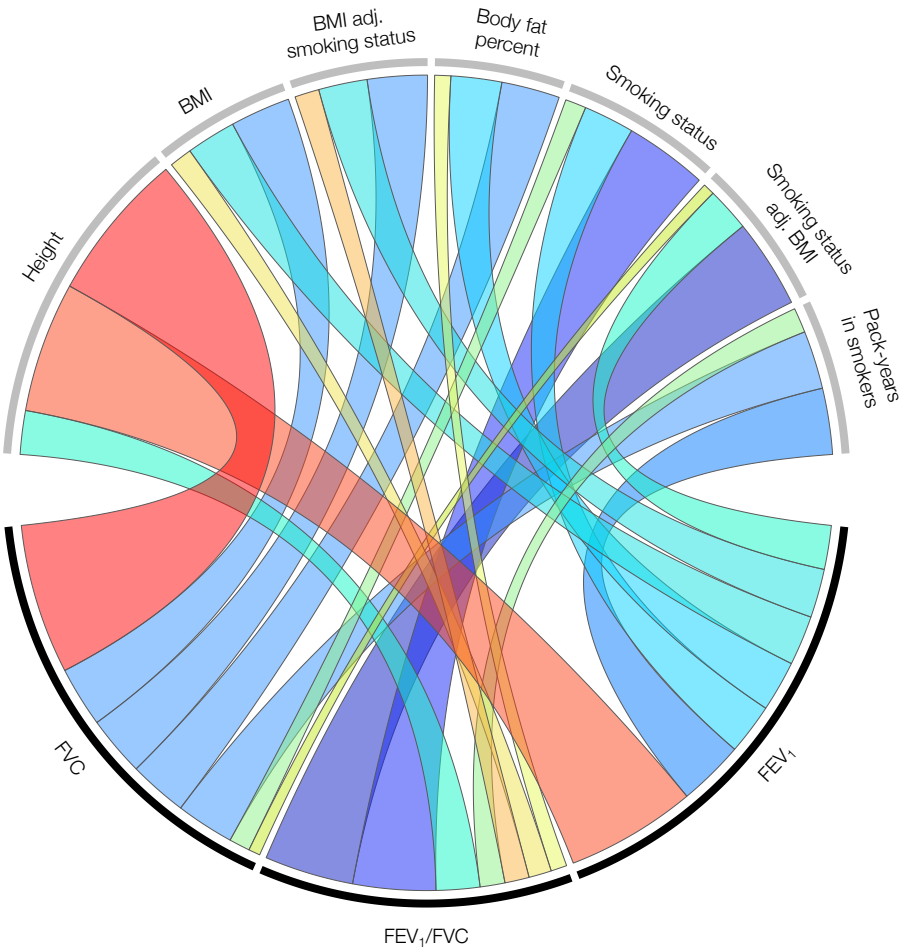
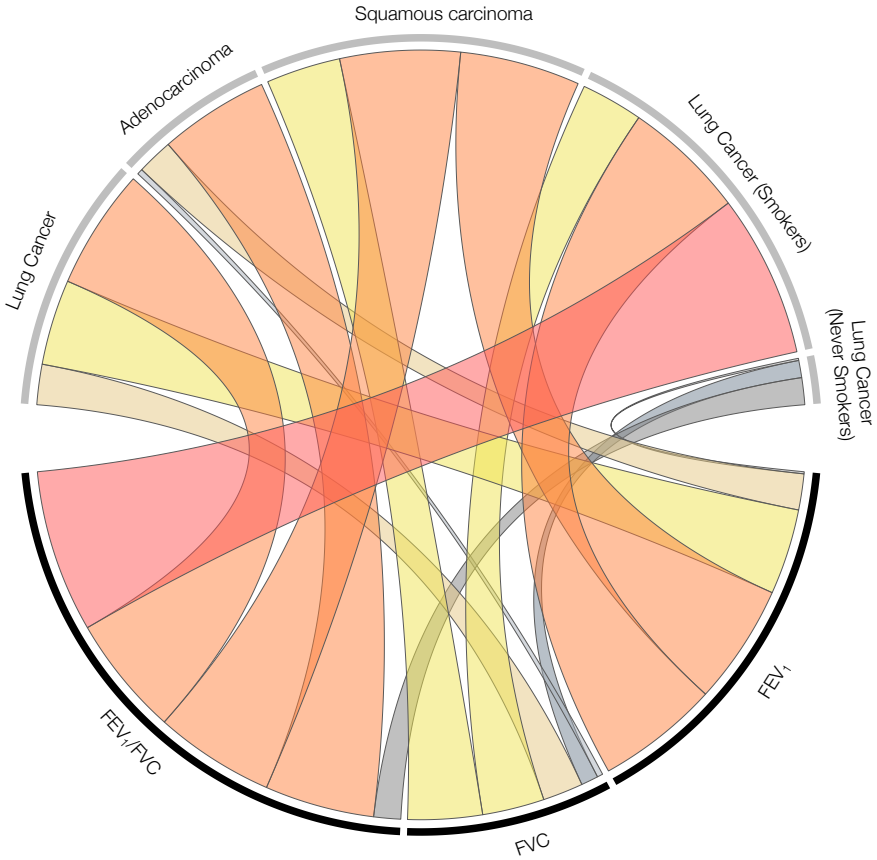


Figure 2: Genetic correlation (r_g) between pulmonary function phenotypes (FEV_1, FVC, FEV_1/FVC) and anthropometric and smoking phenotypes in the UK Biobank cohort. Estimates for r_g are based on UK Biobank-specific LD scores. The colors in the circos plot correspond to the direction of genetic correlation, with warm shades denoting positive relationships and cool tones depicting negative correlations. The width of each band in the circos plot is proportional to the magnitude of the absolute value of the r_g estimate.



	FVC		FEV_1/FVC		FEV_1	
	r_g	p-value	r_g	p-value	r_g	p-value
Height	0.652	1.8×10^{-864}	-0.190	1.9×10^{-34}	0.568	2.5×10^{-567}
BMI	-0.262	1.6×10^{-114}	0.099	4.2×10^{-14}	-0.216	4.2×10^{-74}
BMI adjusted for smoking status	-0.261	3.3×10^{-112}	0.106	9.1×10^{-16}	-0.212	1.7×10^{-70}
Body fat percent	-0.254	1.2×10^{-88}	0.070	4.4×10^{-8}	-0.221	1.7×10^{-66}
Smoking status	-0.091	1.0×10^{-16}	-0.360	7.5×10^{-130}	-0.221	8.1×10^{-78}
Smoking status adjusted for BMI	-0.051	1.7×10^{-6}	-0.388	5.8×10^{-145}	-0.191	7.0×10^{-60}
Cigarette pack-years in smokers	-0.253	1.9×10^{-30}	-0.108	3.0×10^{-4}	-0.287	1.1×10^{-35}

Figure 3: Genetic correlation (r_g) between pulmonary function phenotypes (FEV_1, FVC, FEV_1/FVC) and lung cancer subtypes. Estimates of r_g are based on genome-wide summary statistics from the UK Biobank cohort for pulmonary traits, and the International Lung Cancer Consortium OncoArray study for lung cancer. Genetic correlations have been re-scaled to depict associations between impaired (reduced) pulmonary function and lung cancer risk. The colors in the circos plot correspond to the direction of genetic correlation, with warm shades depicting positive correlations between impaired pulmonary function and lung cancer risk, and gray tones corresponding to inverse and null correlations. The width of each band in the circos plot is proportional to the magnitude of the absolute value of the r_g estimate.



	FEV_1/FVC		FVC		FEV_1	
	r_g	p-value	r_g	p-value	r_g	p-value
Lung Cancer	0.137	2.0×10^{-12}	0.046	5.3×10^{-3}	0.098	2.3×10^{-8}
Adenocarcinoma	0.125	7.2×10^{-9}	-0.006	0.75	0.041	0.044
Squamous cell carcinoma	0.137	4.3×10^{-8}	0.085	7.9×10^{-5}	0.137	7.6×10^{-9}
Lung Cancer (Smokers)	0.185	1.4×10^{-10}	0.071	2.6×10^{-3}	0.140	1.2×10^{-7}
Lung Cancer (Never Smokers)	0.031	0.51	-0.020	0.65	0.002	0.96

Figure 4: Odds ratios (OR) and 95% confidence intervals depicting the effect of impaired pulmonary function on lung cancer risk, estimated using Mendelian Randomization (MR). Multiple MR estimators were applied to the International Lung Cancer Consortium OncoArray dataset comprised of 29,266 lung cancer cases (11,273 adenocarcinoma and 7,426 squamous cell carcinoma cases) and 56,450 controls. Effect estimates based on the full set of genetic instruments are compared to OR estimates based on a subset of instruments, excluding outliers suspected of violating MR assumptions. Associations with p-values less than 0.25 are labeled. Proportion of variation explained by the set of genetic instruments for each pulmonary phenotype was estimated in a separate replication sample of over 110,000 individuals from the UK Biobank.

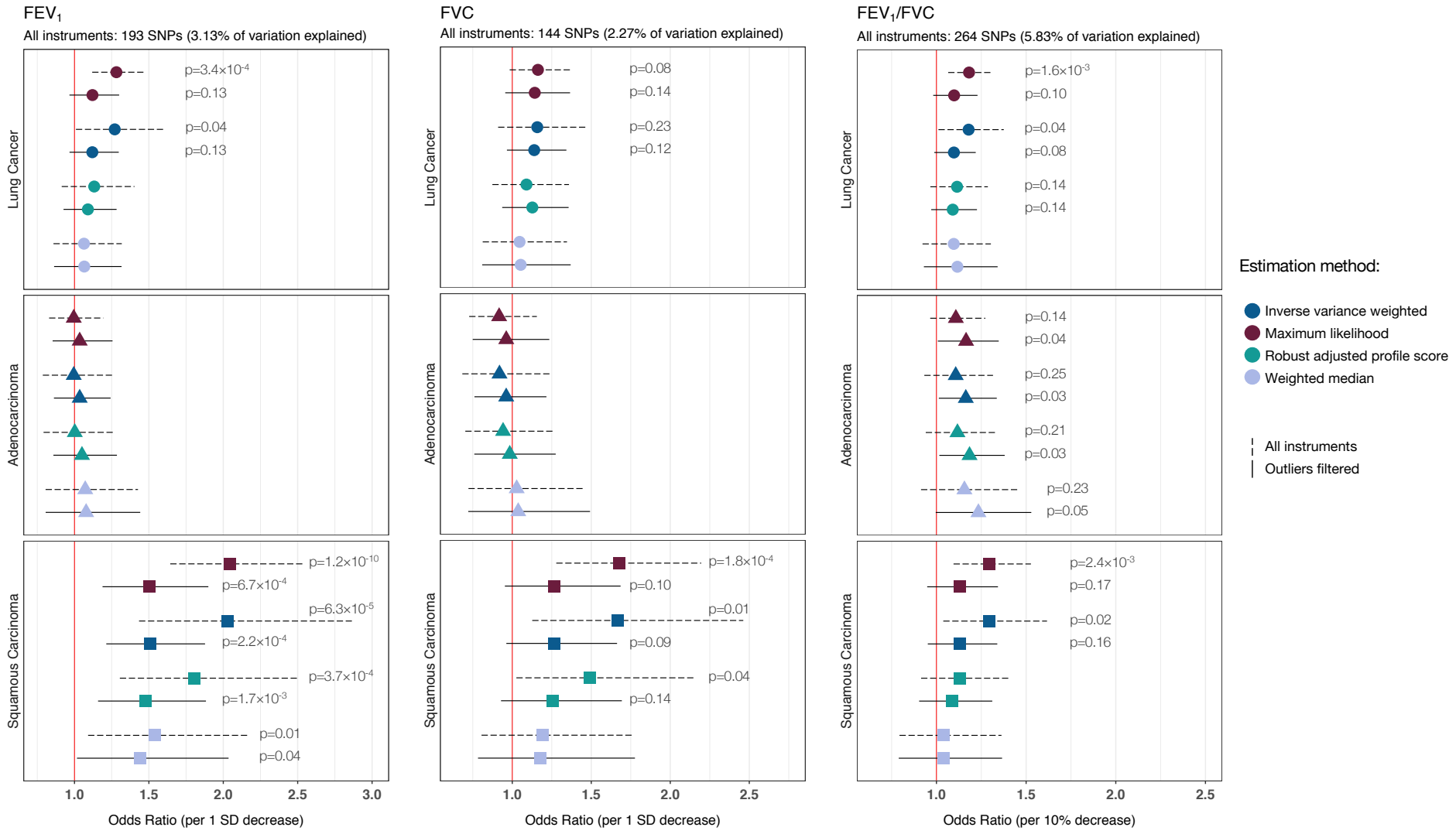
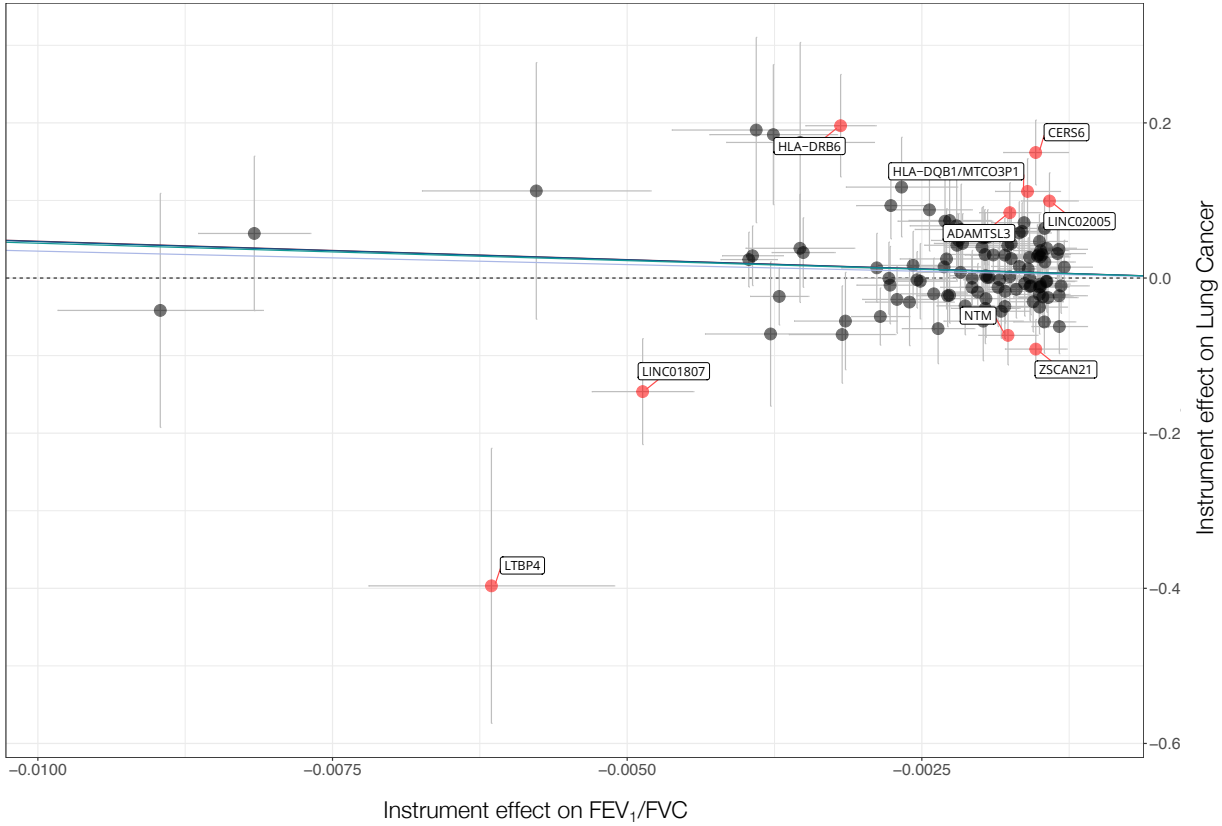


Figure 5: Scatterplot depicting the Mendelian randomization results for FEV_1/FVC in never smokers and lung cancer in never smokers (2355 cases, 7504 controls). The scatterplot illustrates the effects of individual instruments on FEV_1/FVC and lung cancer risk, highlighting potentially invalid outlier instruments that were filtered. The log(OR) estimated using each Mendelian randomization estimation method corresponds to the slope of the lines in scatterplot. Individual SNP effects in the scatterplot correspond to a 1-unit decrease in FEV_1/FVC, but the summary odds ratios for lung cancer have been rescaled to correspond to a 10% decrease in FEV_1/FVC.



Estimator	OR	(95% CI)	p-value
Inverse Variance Weighted	1.60	(1.05 – 2.45)	0.030
Maximum Likelihood	1.61	(1.10 – 2.35)	0.014
Robust adjusted profile score	1.57	(1.03 – 2.37)	0.035
Weighted Median	1.41	(0.78 – 2.57)	0.254
Outliers filtered:			
Inverse Variance Weighted	1.55	(1.09 – 2.19)	0.014
Maximum Likelihood	1.56	(1.05 – 2.30)	0.027
Robust adjusted profile score	1.54	(1.03 – 2.32)	0.035
Weighted Median	1.47	(0.83 – 2.61)	0.186

Table 1: Array-based heritability for FEV_1, FVC, and FEV_1/FVC estimated using LD score regression, comparing UK Biobank-specific LD scores with pre-computed LD scores based on the 1000 Genomes Phase 3 reference population.

	FEV_1		FVC		FEV_1/FVC	
UKB LD Scores	h_g	(SE)	h_g	(SE)	h_g	(SE)
Overall	0.163	(0.006)	0.175	(0.007)	0.128	(0.006)
Never smokers	0.163	(0.007)	0.169	(0.007)	0.126	(0.008)
Smokers	0.159	(0.007)	0.172	(0.009)	0.129	(0.008)
Overall no MHC	0.162	(0.006)	0.175	(0.007)	0.125	(0.006)
1000G LD Scores						
Overall	0.201	(0.008)	0.214	(0.010)	0.157	(0.010)
Never smokers	0.209	(0.010)	0.215	(0.011)	0.159	(0.011)
Smokers	0.208	(0.010)	0.221	(0.011)	0.166	(0.010)

Supplementary Table 1: Functional partitioning of array-based heritability for FEV_1, based on genome-wide summary statistics from the UK Biobank cohort and UK Biobank-specific LD scores.

Category	Pr(SNPs)	Pr(h_g)	SE	Enrichment	SE	Enrichment P
H3K4me1 Trynka (500bp window)	0.606	0.892	0.024	1.473	0.039	1.86E-24
SuperEnhancer Hnisz (500bp window)	0.170	0.339	0.014	1.989	0.084	3.35E-24
CpG Content 50kb	0.010	0.012	0.0002	1.201	0.017	7.64E-24
H3K27ac Hnisz (500bp window)	0.420	0.676	0.023	1.609	0.054	8.05E-23
H3K27ac Hnisz	0.389	0.666	0.023	1.712	0.060	1.02E-22
Background Selection Statistic	0.178	0.242	0.006	1.364	0.034	1.91E-20
SuperEnhancer Hnisz	0.167	0.333	0.017	1.992	0.099	1.70E-19
Conserved Primate phastCons (500bp window)	0.176	0.568	0.041	3.230	0.232	3.12E-18
Conserved LindbladToh (500bp window)	0.330	0.617	0.031	1.866	0.095	1.14E-16
Conserved Primate phastCons	0.019	0.251	0.023	13.022	1.207	1.58E-16
H3K9ac Trynka	0.125	0.488	0.041	3.889	0.331	3.41E-16
Conserved Mammal phastCons (500bp window)	0.339	0.674	0.041	1.986	0.121	8.07E-14
H3K27ac PGC2 (500bp window)	0.335	0.567	0.032	1.690	0.095	5.03E-13
H3K9ac Trynka (500bp window)	0.230	0.518	0.040	2.256	0.173	5.44E-13
Intron UCSC (500bp window)	0.397	0.514	0.015	1.296	0.037	2.85E-12
DGF ENCODE (500bp window)	0.538	0.858	0.043	1.593	0.081	1.78E-11
Bivariate Flanking TSS/Enhancer	0.014	0.111	0.015	8.183	1.083	1.33E-10
Conserved LindbladToh	0.026	0.306	0.040	11.923	1.575	1.98E-10
H3K4me1 Trynka	0.424	0.760	0.053	1.793	0.125	5.71E-10
Conserved Vertebrate phastCons	0.029	0.261	0.035	8.861	1.181	7.37E-10
H3K4me3 Trynka (500bp window)	0.255	0.524	0.043	2.053	0.169	1.16E-09
Enhancer Hoffman (500bp window)	0.090	0.238	0.025	2.644	0.276	2.57E-09
Conserved Mammal phastCons	0.021	0.234	0.034	10.893	1.596	4.66E-09
H3K4me3 Trynka	0.133	0.341	0.035	2.561	0.260	5.42E-09
H3K27ac PGC2	0.269	0.520	0.043	1.935	0.160	1.32E-08
TSS Hoffman (500bp window)	0.034	0.125	0.016	3.629	0.458	1.58E-08
DHS Trynka (500bp window)	0.496	0.754	0.045	1.521	0.091	5.49E-08
Conserved Vertebrate phastCons (500bp window)	0.407	0.655	0.044	1.610	0.107	6.69E-08
TFBS ENCODE (500bp window)	0.341	0.587	0.046	1.720	0.136	2.97E-07
DGF ENCODE	0.136	0.449	0.057	3.302	0.422	3.70E-07
Fetal DHS Trynka (500bp window)	0.283	0.502	0.046	1.773	0.162	5.83E-06
BLUEPRINT DNA methylation MaxCPP	0.032	0.100	0.014	3.158	0.444	6.00E-06
Bivariate Flanking TSS/Enhancer (500bp window)	0.031	0.118	0.019	3.797	0.625	1.03E-05
TSS Hoffman	0.018	0.094	0.017	5.287	0.967	1.48E-05
Coding UCSC	0.014	0.098	0.018	6.873	1.286	1.63E-05
Intron UCSC	0.387	0.459	0.017	1.185	0.043	2.19E-05
DHS Trynka	0.166	0.438	0.060	2.636	0.363	2.29E-05

DHS peaks Trynka	0.111	0.344	0.057	3.114	0.519	9.96E-05
Non-synonymous	0.003	0.037	0.008	13.481	3.069	1.07E-04
TFBS ENCODE	0.131	0.331	0.051	2.524	0.392	1.10E-04
Enhancer Hoffman	0.042	0.153	0.030	3.649	0.707	2.36E-04
UTR 3 UCSC	0.011	0.078	0.018	7.006	1.613	2.79E-04
H3K4me1 peaks Trynka	0.170	0.374	0.056	2.204	0.329	3.01E-04
UTR 3 UCSC (500bp window)	0.026	0.085	0.017	3.229	0.626	4.75E-04
Coding UCSC (500bp window)	0.064	0.123	0.017	1.934	0.261	5.29E-04
Fetal DHS Trynka	0.084	0.256	0.049	3.051	0.582	6.17E-04
H3K4me3 peaks Trynka	0.042	0.144	0.033	3.468	0.800	1.99E-03
GERP.RSsup4	0.008	0.080	0.024	9.874	2.933	3.07E-03
H3K9ac peaks Trynka	0.038	0.147	0.037	3.824	0.955	3.34E-03
GTEx eQTL MaxCPP	0.010	0.040	0.010	3.889	1.010	5.75E-03
BLUEPRINT H3K27acQTL MaxCPP	0.017	0.048	0.013	2.898	0.815	0.022
BLUEPRINT H3K4me1QTL MaxCPP	0.013	0.035	0.010	2.645	0.768	0.032
UTR 5 UCSC	0.005	0.027	0.010	4.881	1.761	0.033
Promoter UCSC (500bp window)	0.057	0.084	0.017	1.482	0.304	0.113
Promoter UCSC	0.046	0.074	0.019	1.607	0.404	0.132
Enhancer Andersson	0.004	0.019	0.010	4.472	2.365	0.144
Promoter Flanking Hoffman (500bp window)	0.033	0.060	0.019	1.812	0.583	0.163
CTCF Hoffman (500bp window)	0.071	0.099	0.026	1.392	0.368	0.289
Enhancer Andersson (500bp window)	0.019	0.030	0.012	1.566	0.617	0.358
Transcr Hoffman	0.346	0.389	0.047	1.123	0.136	0.369
Synonymous	0.003	0.012	0.011	3.791	3.561	0.433
UTR 5 UCSC (500bp window)	0.027	0.034	0.014	1.275	0.515	0.594
Weak Enhancer Hoffman	0.021	0.018	0.020	0.837	0.975	0.867
CTCF Hoffman	0.024	0.026	0.022	1.110	0.928	0.906
Weak Enhancer Hoffman (500bp window)	0.089	0.089	0.026	1.001	0.296	0.998

Supplementary Table 2: Functional partitioning of array-based heritability for FEV_1/FVC, based on genome-wide summary statistics from the UK Biobank cohort and UK Biobank-specific LD scores.

Category	Pr(SNPs)	Pr(h_g)	SE	Enrichment	SE	Enrichment P
H3K27ac Hnisz	0.389	0.800	0.030	2.055	0.077	4.18E-23
SuperEnhancer Hnisz	0.167	0.386	0.020	2.309	0.119	9.57E-22
SuperEnhancer Hnisz (500bp window)	0.170	0.397	0.019	2.330	0.111	1.31E-21
H3K27ac Hnisz (500bp window)	0.420	0.768	0.031	1.826	0.075	4.48E-21
H3K4me1 Trynka (500bp window)	0.606	0.922	0.032	1.522	0.052	5.38E-19
CpG Content 50kb	0.010	0.012	0.0002	1.187	0.020	3.84E-17
Background Selection Statistic	0.178	0.226	0.006	1.272	0.035	1.51E-15
H3K27ac PGC2 (500bp window)	0.335	0.612	0.035	1.825	0.106	9.04E-15
H3K9ac Trynka	0.125	0.485	0.047	3.871	0.378	5.98E-13
H3K9ac Trynka (500bp window)	0.230	0.607	0.051	2.639	0.221	2.00E-12
Conserved Primate phastCons (500bp window)	0.176	0.547	0.050	3.109	0.282	2.15E-12
H3K4me3 Trynka (500bp window)	0.255	0.616	0.050	2.413	0.195	3.83E-12
Fetal DHS Trynka (500bp window)	0.283	0.655	0.050	2.311	0.178	7.85E-11
Enhancer Hoffman (500bp window)	0.090	0.279	0.030	3.107	0.329	4.02E-10
DGF ENCODE (500bp window)	0.538	0.862	0.049	1.601	0.092	7.56E-10
Conserved Mammal phastCons (500bp window)	0.339	0.702	0.058	2.070	0.171	1.74E-09
H3K4me1 Trynka	0.424	0.800	0.059	1.888	0.139	4.28E-09
H3K4me3 Trynka	0.133	0.364	0.041	2.739	0.306	1.17E-08
Fetal DHS Trynka	0.084	0.420	0.059	5.010	0.698	1.33E-07
Bivariate Flanking TSS/Enhancer	0.014	0.131	0.022	9.685	1.638	2.47E-07
Conserved Vertebrate phastCons (500bp window)	0.407	0.686	0.056	1.686	0.136	5.00E-07
DHS Trynka (500bp window)	0.496	0.777	0.056	1.566	0.114	7.71E-07
Conserved LindbladToh	0.026	0.212	0.038	8.239	1.473	9.52E-07
Enhancer Hoffman	0.042	0.211	0.033	5.014	0.776	1.02E-06
Intron UCSC (500bp window)	0.397	0.500	0.020	1.259	0.051	1.95E-06
Conserved LindbladToh (500bp window)	0.330	0.583	0.056	1.763	0.169	3.49E-06
H3K27ac PGC2	0.269	0.512	0.053	1.905	0.199	4.17E-06
Conserved Primate phastCons	0.019	0.165	0.031	8.548	1.610	4.91E-06
Bivariate Flanking TSS/Enhancer (500bp window)	0.031	0.126	0.022	4.050	0.691	1.17E-05
Conserved Mammal phastCons	0.021	0.151	0.032	7.057	1.480	2.48E-05
H3K4me1 peaks Trynka	0.170	0.473	0.071	2.785	0.416	2.62E-05
DHS peaks Trynka	0.111	0.532	0.097	4.809	0.875	3.06E-05
DHS Trynka	0.166	0.584	0.095	3.515	0.574	3.35E-05
DGF ENCODE	0.136	0.462	0.079	3.398	0.582	9.07E-05
Conserved Vertebrate phastCons	0.029	0.148	0.033	5.028	1.108	2.27E-04
TFBS ENCODE (500bp window)	0.341	0.525	0.050	1.539	0.146	2.51E-04
GTEx eQTL MaxCPP	0.010	0.054	0.013	5.239	1.216	3.94E-04

BLUEPRINT DNA methylation MaxCPP	0.032	0.102	0.019	3.206	0.585	4.10E-04
TSS Hoffman (500bp window)	0.034	0.110	0.022	3.213	0.630	5.89E-04
H3K9ac peaks Trynka	0.038	0.199	0.049	5.168	1.279	8.69E-04
H3K4me3 peaks Trynka	0.042	0.170	0.042	4.093	1.011	1.82E-03
Intron UCSC	0.387	0.465	0.027	1.200	0.069	4.87E-03
Coding UCSC (500bp window)	0.064	0.147	0.030	2.301	0.475	6.36E-03
TFBS ENCODE	0.131	0.311	0.067	2.373	0.514	7.18E-03
UTR 3 UCSC	0.011	0.048	0.014	4.325	1.226	7.51E-03
TSS Hoffman	0.018	0.079	0.024	4.407	1.326	9.21E-03
Coding UCSC	0.014	0.059	0.019	4.125	1.332	0.018
BLUEPRINT H3K4me1QTL MaxCPP	0.013	0.037	0.010	2.759	0.743	0.019
GERP.RSsup4	0.008	0.061	0.025	7.448	3.036	0.031
Enhancer Andersson	0.004	0.031	0.013	7.128	2.967	0.043
UTR 3 UCSC (500bp window)	0.026	0.066	0.020	2.482	0.766	0.052
BLUEPRINT H3K27acQTL MaxCPP	0.017	0.045	0.015	2.735	0.888	0.053
Non-synonymous	0.003	0.016	0.008	5.847	3.040	0.117
UTR 5 UCSC (500bp window)	0.027	0.045	0.017	1.669	0.619	0.274
Transcr Hoffman	0.346	0.408	0.058	1.179	0.168	0.293
WeakEnhancer Hoffman (500bp window)	0.089	0.120	0.036	1.352	0.403	0.379
WeakEnhancer Hoffman	0.021	0.041	0.024	1.960	1.145	0.398
Transcr Hoffman (500bp window)	0.762	0.729	0.041	0.956	0.054	0.419
CTCF Hoffman	0.024	0.005	0.028	0.206	1.192	0.504
PromoterFlanking Hoffman (500bp window)	0.033	0.049	0.024	1.467	0.718	0.514
Promoter UCSC (500bp window)	0.057	0.068	0.022	1.192	0.381	0.614
CTCF Hoffman (500bp window)	0.071	0.058	0.031	0.812	0.443	0.670
Enhancer Andersson (500bp window)	0.019	0.026	0.019	1.342	1.006	0.733
UTR 5 UCSC	0.005	0.009	0.014	1.558	2.638	0.832
Synonymous	0.003	0.001	0.014	0.457	4.364	0.901
Promoter UCSC	0.046	0.048	0.026	1.035	0.550	0.950

Supplementary Table 3: Functional partitioning of array-based heritability for FVC, based on genome-wide summary statistics from the UK Biobank cohort and UK Biobank-specific LD scores.

Category	Pr(SNPs)	Pr(h2)	SE	Enrichment	SE	Enrichment P
CpG Content 50kb	0.010	0.012	0.0002	1.230	0.019	2.34E-23
Background Selection Statistic	0.178	0.251	0.006	1.411	0.036	6.87E-23
SuperEnhancer Hnisz (500bp window)	0.170	0.347	0.016	2.037	0.095	5.12E-20
H3K4me1 Trynka (500bp window)	0.606	0.893	0.027	1.474	0.045	9.09E-20
Conserved Primate phastCons (500bp window)	0.176	0.585	0.042	3.325	0.236	2.08E-18
H3K27ac Hnisz	0.389	0.652	0.027	1.675	0.069	2.80E-18
H3K27ac Hnisz (500bp window)	0.420	0.666	0.026	1.583	0.062	9.25E-18
SuperEnhancer Hnisz	0.167	0.344	0.018	2.057	0.109	7.54E-17
Conserved Primate phastCons	0.019	0.253	0.025	13.125	1.290	1.91E-15
H3K9ac Trynka (500bp window)	0.230	0.575	0.040	2.503	0.175	2.37E-15
Conserved LindbladToh (500bp window)	0.330	0.648	0.036	1.961	0.110	6.60E-15
H3K9ac Trynka	0.125	0.481	0.046	3.839	0.367	6.48E-13
Intron UCSC (500bp window)	0.397	0.527	0.016	1.328	0.040	1.29E-12
Conserved Mammal phastCons (500bp window)	0.339	0.706	0.048	2.082	0.140	1.51E-12
H3K27ac PGC2 (500bp window)	0.335	0.585	0.036	1.746	0.107	3.34E-11
DGF ENCODE (500bp window)	0.538	0.874	0.046	1.623	0.085	3.53E-11
Conserved LindbladToh	0.026	0.297	0.039	11.556	1.517	1.51E-10
H3K4me3 Trynka (500bp window)	0.255	0.538	0.043	2.110	0.169	3.76E-10
Conserved Mammal phastCons	0.021	0.242	0.032	11.282	1.506	3.80E-10
H3K4me1 Trynka	0.424	0.769	0.052	1.816	0.123	5.46E-10
Conserved Vertebrate phastCons	0.029	0.254	0.034	8.630	1.148	9.43E-10
H3K4me3 Trynka	0.133	0.350	0.035	2.634	0.266	4.30E-09
Enhancer Hoffman (500bp window)	0.090	0.249	0.026	2.776	0.295	5.72E-09
Bivariate Flanking TSS/Enhancer	0.014	0.132	0.020	9.766	1.447	8.44E-09
DGF ENCODE	0.136	0.506	0.060	3.724	0.443	1.58E-08
DHS Trynka (500bp window)	0.496	0.778	0.047	1.568	0.094	2.63E-08
TSS Hoffman (500bp window)	0.034	0.137	0.019	4.002	0.557	1.51E-07
Conserved Vertebrate phastCons (500bp window)	0.407	0.681	0.050	1.674	0.124	1.67E-07
H3K27ac PGC2	0.269	0.513	0.045	1.909	0.168	2.47E-07
TSS Hoffman	0.018	0.110	0.019	6.152	1.093	4.03E-06
TFBS ENCODE (500bp window)	0.341	0.590	0.052	1.728	0.153	4.54E-06
TFBS ENCODE	0.131	0.392	0.056	2.989	0.428	8.29E-06
DHS Trynka	0.166	0.466	0.064	2.805	0.386	1.14E-05
Bivariate Flanking TSS/Enhancer (500bp window)	0.031	0.132	0.023	4.223	0.729	1.51E-05
DHS peaks Trynka	0.111	0.374	0.061	3.386	0.554	3.68E-05
BLUEPRINT DNA methylation MaxCPP	0.032	0.101	0.016	3.169	0.508	4.61E-05
Coding UCSC	0.014	0.102	0.021	7.153	1.450	5.50E-05

Intron UCSC	0.387	0.461	0.019	1.191	0.048	1.09E-04
H3K4me1 peaks Trynka	0.170	0.379	0.055	2.233	0.321	1.75E-04
Coding UCSC (500bp window)	0.064	0.143	0.021	2.241	0.323	2.21E-04
FetalDHS Trynka (500bp window)	0.283	0.482	0.053	1.702	0.185	2.52E-04
Enhancer Hoffman	0.042	0.160	0.034	3.822	0.804	6.03E-04
UTR 3 UCSC	0.011	0.082	0.020	7.327	1.823	7.48E-04
Non-synonymous	0.003	0.033	0.009	12.121	3.398	1.42E-03
FetalDHS Trynka	0.084	0.243	0.049	2.894	0.587	1.51E-03
UTR 3 UCSC (500bp window)	0.026	0.081	0.018	3.076	0.675	2.41E-03
GERP.RSsup4	0.008	0.077	0.022	9.487	2.753	2.63E-03
GTEx eQTL MaxCPP	0.010	0.045	0.012	4.395	1.140	3.69E-03
BLUEPRINT H3K27acQTL MaxCPP	0.017	0.061	0.015	3.658	0.899	4.00E-03
H3K4me3 peaks Trynka	0.042	0.137	0.036	3.291	0.861	8.58E-03
UTR 5 UCSC	0.005	0.033	0.012	5.952	2.178	0.025
BLUEPRINT H3K4me1QTL MaxCPP	0.013	0.037	0.011	2.767	0.811	0.030
H3K9ac peaks Trynka	0.038	0.125	0.044	3.248	1.131	0.049
PromoterFlanking Hoffman (500bp window)	0.033	0.074	0.022	2.230	0.669	0.068
Promoter UCSC (500bp window)	0.057	0.088	0.019	1.547	0.331	0.099
Promoter UCSC	0.046	0.076	0.020	1.639	0.429	0.137
Synonymous	0.003	0.023	0.014	7.340	4.368	0.150
Transcr Hoffman	0.346	0.419	0.052	1.211	0.150	0.162
Enhancer Andersson (500bp window)	0.019	0.037	0.013	1.966	0.699	0.165
UTR 5 UCSC (500bp window)	0.027	0.046	0.014	1.727	0.536	0.179
Enhancer Andersson	0.004	0.019	0.011	4.303	2.560	0.198
WeakEnhancer Hoffman	0.021	0.037	0.024	1.762	1.127	0.498
CTCF Hoffman	0.024	0.028	0.026	1.173	1.071	0.872
WeakEnhancer Hoffman (500bp window)	0.089	0.093	0.028	1.050	0.311	0.872
CTCF Hoffman (500bp window)	0.071	0.073	0.027	1.026	0.379	0.946

Supplementary Table 4: Partitioned heritability for estimates for lung cancer

Category	Pr(SNPs)	Pr(h2)	SE	Enrichment	SE	Enrichment P
CpG Content 50kb	0.010	0.014	0.001	1.351	0.061	2.07E-07
Background Selection Statistic	0.178	0.251	0.019	1.412	0.108	1.03E-06
SuperEnhancer Hnisz	0.167	0.367	0.048	2.194	0.285	4.40E-06
SuperEnhancer Hnisz (500bp window)	0.170	0.350	0.051	2.052	0.298	7.61E-05
Coding UCSC (500bp window)	0.064	0.364	0.082	5.710	1.295	4.12E-04
Intron UCSC (500bp window)	0.397	0.574	0.055	1.447	0.137	1.11E-03
H3K27ac PGC2	0.269	0.700	0.128	2.605	0.477	1.28E-03
Conserved Vertebrate phastCons (500bp window)	0.407	0.857	0.147	2.106	0.362	3.60E-03
Conserved Mammal phastCons (500bp window)	0.339	0.791	0.149	2.332	0.438	3.93E-03
Conserved Primate phastCons (500bp window)	0.176	0.524	0.130	2.981	0.742	5.34E-03
Conserved LindbladToh (500bp window)	0.330	0.761	0.144	2.302	0.436	5.36E-03
Conserved LindbladToh	0.026	0.322	0.110	12.552	4.268	0.011
H3K27ac Hnisz (500bp window)	0.420	0.670	0.095	1.594	0.227	0.012
DHS Trynka (500bp window)	0.496	0.901	0.157	1.817	0.316	0.013
Conserved Primate phastCons	0.019	0.228	0.086	11.826	4.454	0.013
UTR 3 UCSC (500bp window)	0.026	0.144	0.051	5.452	1.912	0.015
TFBS ENCODE (500bp window)	0.341	0.736	0.163	2.156	0.479	0.018
Conserved Mammal phastCons	0.021	0.447	0.172	20.855	8.021	0.022
Intron UCSC	0.387	0.506	0.062	1.307	0.160	0.041
H3K9ac Trynka (500bp window)	0.230	0.572	0.163	2.489	0.711	0.046
Conserved Vertebrate phastCons	0.029	0.400	0.174	13.592	5.905	0.046
TSS Hoffman (500bp window)	0.034	0.187	0.076	5.451	2.201	0.050
H3K27ac Hnisz	0.389	0.536	0.073	1.379	0.188	0.055
DGF ENCODE (500bp window)	0.538	0.813	0.140	1.510	0.259	0.058
TFBS ENCODE	0.131	0.489	0.188	3.728	1.430	0.064
H3K4me1 Trynka	0.424	0.705	0.158	1.664	0.372	0.084
UTR 5 UCSC (500bp window)	0.027	0.166	0.081	6.185	3.006	0.091
H3K4me1 Trynka (500bp window)	0.606	0.772	0.098	1.274	0.161	0.106
Enhancer Hoffman (500bp window)	0.090	0.225	0.086	2.502	0.955	0.114
Fetal DHS Trynka (500bp window)	0.283	0.557	0.172	1.967	0.606	0.121
Transcr Hoffman	0.346	0.571	0.159	1.651	0.460	0.125
H3K27ac PGC2 (500bp window)	0.335	0.507	0.119	1.512	0.356	0.126
Promoter UCSC (500bp window)	0.057	0.170	0.076	2.984	1.342	0.149
BLUEPRINT DNA methylation MaxCPP	0.032	0.137	0.072	4.306	2.271	0.150
H3K4me3 Trynka (500bp window)	0.255	0.441	0.135	1.729	0.530	0.175
H3K4me3 Trynka	0.133	0.293	0.120	2.204	0.901	0.181
Synonymous	0.003	0.068	0.049	21.735	15.804	0.190
BLUEPRINT H3K27acQTL MaxCPP	0.017	0.071	0.043	4.315	2.573	0.201

GTEx eQTL MaxCPP	0.010	0.046	0.030	4.483	2.908	0.214
Bivariate Flanking TSS/Enhancer	0.014	0.096	0.070	7.050	5.164	0.243
H3K9ac Trynka	0.125	0.261	0.123	2.079	0.983	0.275
GERP.RSsup4	0.008	0.068	0.056	8.287	6.909	0.299
PromoterFlanking Hoffman	0.008	0.070	0.060	8.483	7.258	0.306
BLUEPRINT H3K4me1QTL MaxCPP	0.013	0.047	0.034	3.482	2.515	0.320
H3K4me3 peaks Trynka	0.042	0.163	0.127	3.902	3.049	0.348
Coding UCSC	0.014	0.069	0.060	4.839	4.228	0.367
Promoter UCSC	0.046	0.154	0.124	3.329	2.680	0.387
Enhancer Hoffman	0.042	0.134	0.105	3.182	2.506	0.389
WeakEnhancer Hoffman (500bp window)	0.089	0.202	0.130	2.273	1.461	0.392
Bivariate Flanking TSS/Enhancer (500bp window)	0.031	0.084	0.065	2.682	2.091	0.416
Enhancer Andersson (500bp window)	0.019	0.055	0.057	2.882	2.991	0.534
UTR 3 UCSC	0.011	0.035	0.038	3.106	3.357	0.534
TSS Hoffman	0.018	0.058	0.071	3.267	4.006	0.573
DGF ENCODE	0.136	0.014	0.233	0.100	1.713	0.599
PromoterFlanking Hoffman (500bp window)	0.033	0.068	0.070	2.055	2.122	0.620
FetalDHS Trynka	0.084	0.004	0.172	0.042	2.049	0.633
DHS Trynka	0.166	0.074	0.223	0.445	1.342	0.669
CTCF Hoffman (500bp window)	0.071	0.143	0.179	2.017	2.522	0.681
H3K4me1 peaks Trynka	0.170	0.242	0.214	1.427	1.257	0.733
H3K9ac peaks Trynka	0.038	0.002	0.119	0.047	3.085	0.754
Transcr Hoffman (500bp window)	0.762	0.750	0.099	0.984	0.130	0.900

Supplementary Table 5: Additional 73 independent variants associated with pulmonary function identified and replicated in the UK Biobank cohort and used as genetic instruments in Mendelian randomization analyses of lung cancer risk

CHR	Position (GRCh37)	SNP	Alleles Effect/Other	EAF	Discovery Beta	Discovery SE	Discovery P-value	Replication P-value	Lead Phenotype[1]	INFO	Nearest Gene and Function	
6	7146350	rs4960289	A/G	0.427	-0.0017	0.0002	9.6E-20	3.7E-14	FEV_1/FVC	0.986	*RREB1*	intronic
22	30598516	rs6006399	T/G	0.884	-0.0245	0.0028	1.9E-18	6.3E-05	FEV1	1	*HORMAD2*	intronic
2	36724282	rs1179500	A/C	0.721	-0.0017	0.0002	3.6E-17	6.8E-05	FEV_1/FVC	0.988	*CRIM1*	intronic
16	67560613	rs7196853	T/C	0.897	-0.0025	0.0003	1.3E-16	1.4E-05	FEV_1/FVC	0.994	*FAM65A*	intronic
4	146087979	rs116125427	G/A	0.922	0.0276	0.0034	2.3E-16	1.7E-06	FEV_1	0.987	*OTUD4*	intronic
10	124230612	rs12571363	C/A	0.893	0.0228	0.0028	3.2E-16	1.3E-06	FVC	0.995	*HTRA1*	intronic
8	103131300	rs659398	T/C	0.269	-0.0017	0.0002	4.9E-16	6.4E-05	FEV_1/FVC	0.983	*NCALD*	intronic
16	58062081	rs41418849	T/G	0.940	-0.0031	0.0004	5.2E-16	4.1E-06	FEV_1/FVC	0.949	*MMP15*	intronic
22	30254175	rs9614084	C/T	0.521	-0.0015	0.0002	7.2E-16	4.9E-11	FEV_1/FVC	0.996	*ASCC2*	intergenic
5	157022475	rs72811372	G/A	0.895	-0.0023	0.0003	2.3E-15	8.3E-06	FEV_1/FVC	0.998	*AC008694.2*	intergenic
1	221473248	rs11118683	C/T	0.579	-0.0140	0.0018	2.5E-15	2.7E-06	FVC	0.975	*RP11-421L10.1*	intergenic
14	54410919	rs4444235	T/C	0.536	0.0014	0.0002	8.9E-15	4.4E-06	FEV_1/FVC	1	*MIR5580*	intergenic
5	120078424	rs1125578	A/C	0.461	-0.0014	0.0002	1.1E-14	1.3E-04	FEV_1/FVC	0.991	*RNU4-69P*	intergenic
19	46309203	rs1548029	A/C	0.644	-0.0015	0.0002	1.3E-14	5.5E-11	FEV_1/FVC	0.998	*RSPH6A*	intronic
2	42433247	rs12466981	C/T	0.728	-0.0015	0.0002	2.7E-14	5.9E-06	FEV_1/FVC	0.996	*EML4*	intronic
1	3425867	rs2794359	C/A	0.895	-0.0023	0.0003	2.8E-14	2.0E-10	FEV_1/FVC	0.986	*MEGF6*	intronic
5	158368797	rs7443323	T/A	0.748	0.0151	0.0020	4.6E-14	1.6E-05	FVC	0.992	*EBF1*	intronic
5	43484257	rs79904209	C/T	0.906	0.0229	0.0030	5.6E-14	2.2E-05	FEV_1	1	*TMEM267*	5' UTR
2	178210268	rs6740092	T/A	0.145	-0.0181	0.0025	1.9E-13	4.3E-05	FVC	0.992	*LOC100130691*	ncRNA_intronic
6	32652486	rs9275156	T/C	0.766	0.0016	0.0002	2.4E-13	1.0E-05	FEV_1/FVC	0.998	*HLA-DQB1*	intergenic
18	20032854	rs4800410	A/C	0.596	-0.0134	0.0018	2.6E-13	1.3E-04	FEV_1	0.988	*RP11-863N1.4*	intergenic
2	100568631	rs11887136	G/A	0.868	0.0189	0.0026	7.6E-13	8.9E-05	FEV_1	0.996	*AFF3*	intronic
6	152593102	rs6904757	A/G	0.636	0.0129	0.0018	9.4E-13	1.1E-04	FVC	0.982	*SYNE1*	intronic
11	13166565	rs11022690	T/C	0.545	-0.0013	0.0002	1.2E-12	8.2E-06	FEV_1/FVC	0.995	*RP11-413N13.1*	intergenic
2	69407720	rs10173269	T/G	0.539	-0.0013	0.0002	1.3E-12	1.0E-04	FEV_1/FVC	1	*ANTXR1*	intronic

6	33025953	rs6457709	A/G	0.705	0.0014	0.0002	1.4E-12	5.3E-06	FEV_1/FVC	1	*HLA-DPA1*	intergenic
6	26207174	rs9393688	A/T	0.738	-0.0139	0.0020	1.7E-12	2.9E-05	FVC	0.992	*HIST1H4E*	downstream
6	19837774	rs11759102	C/T	0.840	-0.0171	0.0024	2.3E-12	5.2E-05	FEV_1	0.989	*ID4*	5' UTR
6	22072425	rs196025	A/G	0.368	-0.0130	0.0019	2.4E-12	9.0E-06	FEV_1	0.985	*CASC15*	ncRNA_intronic
8	69570332	rs17387279	T/G	0.819	0.0156	0.0023	5.0E-12	1.8E-04	FVC	0.990	*C8orf34*	intronic
1	92164100	rs4233430	C/T	0.814	-0.0016	0.0002	9.9E-12	1.4E-08	FEV_1/FVC	0.992	*TGFBR3*	intronic
15	77354847	rs4886509	C/A	0.329	-0.0013	0.0002	1.3E-11	1.0E-06	FEV_1/FVC	0.992	*TSPAN3*	intronic
17	60720058	rs12452590	T/G	0.638	0.0013	0.0002	1.5E-11	5.8E-08	FEV_1/FVC	0.974	*MRC2*	intronic
5	170909410	rs11745375	C/T	0.527	-0.0012	0.0002	1.6E-11	1.3E-04	FEV_1/FVC	1	*FGF18*	intergenic
7	19050020	rs28719767	G/C	0.711	-0.0130	0.0020	5.1E-11	7.0E-05	FEV_1	0.983	*HDAC9*	intergenic
22	18357509	rs72490631	T/C	0.773	0.0140	0.0021	5.5E-11	2.0E-06	FEV_1	0.994	*MICAL3*	intronic
2	230258493	rs207672	T/G	0.340	0.0012	0.0002	7.2E-11	1.3E-07	FEV_1/FVC	0.992	*DNER*	intronic
11	12660894	rs7927422	T/C	0.593	0.0115	0.0018	7.5E-11	2.1E-05	FVC	0.995	*TEAD1*	intergenic
3	73554922	rs4677294	T/A	0.641	0.0121	0.0019	9.3E-11	1.6E-04	FEV_1	0.989	*PDZRN3*	intronic
12	65962636	rs10878300	T/G	0.198	-0.0140	0.0022	1.3E-10	8.5E-05	FVC	0.992	*LOC100507065*	ncRNA_intronic
17	38489170	rs2715554	A/G	0.849	0.0016	0.0003	1.5E-10	4.4E-05	FEV_1/FVC	1	*RARA*	intronic
16	88807608	rs750739	A/G	0.833	0.0153	0.0024	1.8E-10	1.0E-04	FEV_1	0.993	*PIEZO1*	intronic
1	92031492	rs2125126	G/A	0.850	-0.0016	0.0003	2.1E-10	8.3E-07	FEV_1/FVC	0.983	*RP11-47K11.2*	intergenic
5	132375622	rs113638840	A/G	0.744	0.0130	0.0020	2.5E-10	2.0E-04	FEV_1	0.996	*HSPA4*	intergenic
14	93507197	rs1956028	T/C	0.874	0.0168	0.0027	4.3E-10	3.4E-07	FEV_1	0.990	*ITPK1*	intronic
1	9498113	rs9660890	T/C	0.803	-0.0014	0.0002	5.8E-10	1.2E-04	FEV_1/FVC	0.992	*RNA5SP40*	upstream
6	90948093	rs58453446	G/C	0.652	-0.0015	0.0002	8.9E-10	3.5E-05	FEV_1/FVC nvsmk	0.998	*BACH2*	intronic
2	169479763	rs10184235	A/G	0.753	-0.0013	0.0002	1.1E-09	1.1E-06	FEV_1/FVC	0.999	*CERS6*	intronic
5	173303392	rs55993676	G/T	0.713	-0.0012	0.0002	1.4E-09	7.4E-05	FEV_1/FVC	0.998	*CPEB4*	intergenic
4	174582067	rs10005540	C/T	0.386	-0.0107	0.0018	2.6E-09	1.7E-04	FVC	0.978	*RANP6*	intergenic
1	22653424	rs4233284	C/G	0.672	0.0109	0.0018	3.2E-09	1.3E-04	FVC	0.993	*RP11-415K20.1*	intergenic
12	28840892	rs12313454	A/G	0.882	-0.0158	0.0027	3.5E-09	9.2E-05	FVC	0.996	*CCDC91*	intergenic
8	64806567	rs1425794	G/T	0.565	-0.0103	0.0017	4.2E-09	1.5E-04	FVC	0.995	*RP11-32K4.1*	ncRNA_intronic

7	140560023	rs13227429	T/C	0.436	0.0102	0.0017	5.6E-09	6.6E-05	FVC	0.994	*BRAF*	intronic
2	67052289	rs17032590	A/G	0.791	-0.0013	0.0002	8.1E-09	8.0E-06	FEV_1/FVC	0.996	*AC009474.1*	intergenic
8	13194983	rs1528624	T/G	0.505	0.0010	0.0002	8.4E-09	7.6E-09	FEV_1/FVC	0.990	*DLC1*	intronic
10	63839417	rs4948502	T/C	0.573	0.0101	0.0018	8.9E-09	3.0E-07	FVC	0.995	*ARID5B*	intronic
5	148203236	rs35684381	T/C	0.854	-0.0015	0.0003	8.9E-09	3.4E-06	FEV_1/FVC	0.979	*ADRB2*	intergenic
4	145557467	rs35797611	T/C	0.968	-0.0029	0.0005	9.6E-09	1.9E-05	FEV_1/FVC	1	*HHIP-AS1*	intergenic
20	3660789	rs603112	T/C	0.157	-0.0014	0.0002	1.2E-08	1.3E-04	FEV_1/FVC	1	*ADAM33*	intronic
3	11642114	rs1561073	T/A	0.262	-0.0012	0.0002	1.6E-08	2.2E-05	FEV_1/FVC	0.992	*VGLL4*	intronic
15	67464291	rs72743477	A/G	0.783	0.0012	0.0002	1.7E-08	1.1E-04	FEV_1/FVC	0.985	*SMAD3*	intronic
7	134567104	rs28517513	G/T	0.751	0.0012	0.0002	2.1E-08	5.7E-05	FEV_1/FVC	0.990	*CALD1*	intronic
11	65324276	rs11227223	C/T	0.945	-0.0022	0.0004	2.2E-08	3.4E-07	FEV_1/FVC	0.995	*LTBP3*	intronic
17	13416372	rs1978218	G/T	0.404	-0.0132	0.0024	2.3E-08	1.4E-04	FVC-nvsmk	0.986	*HS3ST3A1*	intronic
4	90036240	rs17821105	A/T	0.830	0.0013	0.0002	2.5E-08	2.3E-06	FEV_1/FVC	0.997	*TIGD2*	downstream
16	31030344	rs4889526	C/A	0.625	0.0102	0.0018	2.9E-08	9.5E-09	FEV_1	0.998	*STX1B*	intergenic
3	53672471	rs9819463	T/C	0.795	0.0123	0.0022	2.9E-08	2.5E-07	FEV_1	0.994	*CACNA1D*	intronic
9	129416317	rs10987386	C/T	0.813	0.0013	0.0002	3.3E-08	1.2E-08	FEV_1/FVC	0.982	*LMX1B*	intronic
2	43762112	rs77972916	G/A	0.924	-0.0019	0.0003	3.3E-08	3.0E-05	FEV_1/FVC	0.995	*THADA*	intronic
18	38023468	rs4636990	G/A	0.457	-0.0096	0.0017	3.8E-08	7.3E-05	FVC	0.990	*RNU7-145P*	intergenic
8	109446828	rs10089406	T/C	0.404	0.0010	0.0002	4.4E-08	5.0E-06	FEV_1/FVC	0.987	*EIF3E*	intergenic

[1.] Lead phenotype refers to the lung function trait for which this variant was selected to be a genetic instrument. For each SNP the lead phenotype has the smallest p-value.

Supplementary Table 6: Descriptive characteristics of the OncoArray lung cancer population

		Lung Cancer Cases		Controls	
		Number	(%)	Number	(%)
Age	≤50 years	3112	(12)	6,032	(12)
	>50 years	23025	(88)	44,075	(88)
Sex	Male	18208	(62)	27,178	(53)
	Female	11059	(38)	24,069	(47)
Smoking status	Never smokers	2355	(9)	7,504	(31)
	Ever smokers	23223	(91)	16,964	(69)
	Former smokers	9037	(35)	8,554	(35)
	Current smokers	13356	(52)	7,477	(31)
Histology	Adenocarcinoma	11273	(39)		
	Squamous cell carcinoma	7426	(35)		
	Small cell carcinoma	2664	(9)		
Total		29266	(100)	56,450	(100)

Supplementary Table 7: Mendelian randomization (MR) odds ratio (OR) estimates for lung cancer overall and by histology for a genetically predicted 1-SD decrease in a standardized FEV_1 z-score

Outcome	MR Estimator	N_{SNPs}	OR	(95% CI)	P
Lung Cancer (cases: n=29,266 controls: n=56,450)	IVW-RE	193	1.27	(1.01 – 1.60)	0.041
	Maximum likelihood	193	1.28	(1.12 – 1.47)	3.4×10^{-4}
	Weighted median	193	1.06	(0.86 – 1.32)	0.572
	MR RAPS	193	1.13	(0.91 – 1.40)	0.255
	MR Egger slope	193	1.65	(0.70 – 3.87)	0.251
	Outliers filtered:				
	IVW-RE	157	1.12	(0.97 – 1.30)	0.132
	Maximum likelihood	157	1.12	(0.97 – 1.30)	0.132
	Weighted median	157	1.07	(0.86 – 1.32)	0.554
	MR RAPS	157	1.09	(0.93 – 1.28)	0.299
	MR Egger slope	157	0.84	(0.47 – 1.50)	0.555
Adenocarcinoma (cases: n=11,273 controls: n=56,450)	IVW-RE	192	0.99	(0.79 – 1.26)	0.965
	Maximum likelihood	192	0.99	(0.83 – 1.19)	0.956
	Weighted median	192	1.07	(0.81 – 1.43)	0.633
	MR RAPS	192	1.00	(0.79 – 1.27)	0.987
	MR Egger slope	192	0.91	(0.39 – 2.10)	0.827
	Outliers filtered:				
	IVW-RE	169	1.03	(0.86 – 1.24)	0.714
	Maximum likelihood	169	1.03	(0.85 – 1.26)	0.728
	Weighted median	169	1.08	(0.81 – 1.44)	0.612
	MR RAPS	169	1.05	(0.86 – 1.29)	0.634
	MR Egger slope	169	0.80	(0.39 – 1.66)	0.557
Squamous carcinoma (cases: n=7426) controls: n=56,450)	IVW-RE	190	2.03	(1.43 – 2.86)	6.3×10^{-5}
	Maximum likelihood	190	2.04	(1.64 – 2.54)	1.2×10^{-10}
	Weighted median	190	1.54	(1.09 – 2.17)	0.014
	MR RAPS	190	1.81	(1.30 – 2.50)	3.7×10^{-4}
	MR Egger slope	190	5.43	(1.50 – 19.63)	0.011
	Outliers filtered:				
	IVW-RE	156	1.51	(1.21 – 1.88)	2.2×10^{-4}
	Maximum likelihood	156	1.50	(1.19 – 1.90)	6.7×10^{-4}
	Weighted median	156	1.44	(1.02 – 2.04)	0.040
	MR RAPS	156	1.48	(1.16 – 1.88)	1.7×10^{-3}
	MR Egger slope	156	2.11	(0.89 – 5.02)	0.092

IVW – RE : Inverse variance weighted – multiplicative random effects
RAPS : Robust adjusted profile score – Huber loss function
PRESSO : Randomization Pleiotropy RESidual Sum and Outlier

Supplementary Table 8: Mendelian randomization (MR) odds ratio (OR) estimates for the risk of lung cancer overall and by histology for a genetically predicted 1-SD decrease in a standardized FVC z-score.

Outcome	MR Estimator	N_{SNPs}	OR	(95% CI)	P
Lung Cancer (cases: n=29,266 controls: n=56,450)	IVW-RE	144	1.16	(0.91 – 1.47)	0.232
	Maximum likelihood	144	1.16	(0.98 – 1.37)	0.078
	Weighted median	144	1.05	(0.81 – 1.35)	0.726
	MR RAPS	144	1.09	(0.87 – 1.36)	0.445
	MR Egger slope	144	0.75	(0.28 – 2.04)	0.576
	Outliers filtered:				
	IVW-RE	124	1.14	(0.97 – 1.34)	0.121
	Maximum likelihood	124	1.14	(0.96 – 1.37)	0.144
	Weighted median	124	1.05	(0.81 – 1.37)	0.699
	MR RAPS	124	1.13	(0.94 – 1.36)	0.206
	MR Egger slope	124	0.45	(0.20 – 1.05)	0.067
Adenocarcinoma (cases: n=11,273 controls: n=56,450)	IVW-RE	143	0.92	(0.68 – 1.24)	0.576
	Maximum likelihood	143	0.92	(0.73 – 1.16)	0.462
	Weighted median	143	1.03	(0.72 – 1.46)	0.878
	MR RAPS	143	0.94	(0.70 – 1.26)	0.689
	MR Egger slope	143	0.64	(0.19 – 2.18)	0.478
	Outliers filtered:				
	IVW-RE	124	0.96	(0.76 – 1.22)	0.744
	Maximum likelihood	124	0.96	(0.75 – 1.23)	0.760
	Weighted median	124	1.04	(0.72 – 1.49)	0.842
	MR RAPS	124	0.98	(0.76 – 1.27)	0.904
	MR Egger slope	124	0.92	(0.29 – 2.91)	0.884
Squamous carcinoma (cases: n=7426) controls: n=56,450)	IVW-RE	140	1.67	(1.13 – 2.46)	0.011
	Maximum likelihood	140	1.68	(1.28 – 2.20)	1.8×10^{-4}
	Weighted median	140	1.19	(0.81 – 1.76)	0.381
	MR RAPS	140	1.46	(1.11 – 1.92)	6.5×10^{-3}
	MR Egger slope	140	0.73	(0.14 – 3.70)	0.700
	Outliers filtered:				
	IVW	122	1.27	(0.96 – 1.66)	0.092
	Maximum likelihood	122	1.27	(0.95 – 1.69)	0.103
	Weighted median	122	1.18	(0.78 – 1.78)	0.429
	MR RAPS	122	1.25	(0.93 – 1.69)	0.139
	MR Egger slope	122	0.43	(0.13 – 1.41)	0.167

IVW – RE : Inverse variance weighted – multiplicative random effects
RAPS : Robust adjusted profile score – Huber loss function

Supplementary Table 9: Mendelian randomization (MR) odds ratio (OR) estimates for the risk of lung cancer overall and by histology for a genetically predicted 10% decrease in FEV_1/FVC.

Outcome	MR Estimator	N_{SNPs}	OR	(95% CI)	P
Lung Cancer (cases: n=29,266 controls: n=56,450)	IVW-RE	264	1.18	(1.01 – 1.38)	0.035
	Maximum likelihood	264	1.18	(1.07 – 1.31)	1.6×10^{-3}
	Weighted median	264	1.10	(0.92 – 1.30)	0.296
	MR RAPS	264	1.11	(0.97 – 1.29)	0.137
	MR Egger slope	264	1.15	(0.77 – 1.71)	0.485
	Outliers filtered:				
	IVW-RE	226	1.10	(0.99 – 1.22)	0.082
	Maximum likelihood	226	1.10	(0.98 – 1.23)	0.100
	Weighted median	226	1.12	(0.93 – 1.34)	0.237
	MR RAPS	226	1.09	(0.97 – 1.23)	0.143
	MR Egger slope	255	1.22	(0.91 – 1.64)	0.184
Adenocarcinoma (cases: n=11,273 controls: n=56,450)	IVW-RE	265	1.11	(0.93 – 1.32)	0.245
	Maximum likelihood	265	1.11	(0.97 – 1.27)	0.145
	Weighted median	265	1.16	(0.91 – 1.46)	0.226
	MR RAPS	265	1.12	(0.94 – 1.33)	0.207
	MR Egger slope	265	1.29	(0.84 – 1.98)	0.240
	Outliers filtered:				
	IVW-RE	239	1.16	(1.01 – 1.34)	0.032
	Maximum likelihood	239	1.17	(1.01 – 1.35)	0.039
	Weighted median	239	1.23	(1.00 – 1.53)	0.054
	MR RAPS	239	1.18	(1.02 – 1.38)	0.030
	MR Egger slope	239	1.43	(1.00 – 2.05)	0.052
Squamous carcinoma (cases: n=7426) controls: n=56,450)	IVW-RE	263	1.30	(1.04 – 1.62)	0.022
	Maximum likelihood	263	1.29	(1.10 – 1.53)	2.4×10^{-3}
	Weighted median	263	1.04	(0.79 – 1.36)	0.775
	MR RAPS	263	1.13	(0.91 – 1.40)	0.255
	MR Egger slope	263	1.25	(0.70 – 2.24)	0.446
	Outliers filtered:				
	IVW-RE	239	1.13	(0.95 – 1.34)	0.164
	Maximum likelihood	239	1.13	(0.95 – 1.34)	0.173
	Weighted median	239	1.04	(0.79 – 1.37)	0.785
	MR RAPS	239	1.09	(0.90 – 1.31)	0.372
	MR Egger slope	239	1.03	(0.65 – 1.62)	0.912

IVW – RE : Inverse variance weighted – multiplicative random effects
RAPS : Robust adjusted profile score – Huber loss function

Supplementary Table 10: Mendelian randomization (MR) odds ratio (OR) estimates for risk of lung cancer in never smokers (2355 cases, 7504 controls) corresponding to a 1-SD decrease in FEV_1 and FVC z-scores and 10% decrease in FEV_1/FVC. Genetic instruments for pulmonary function were developed in a separate GWAS of never smokers in the UK Biobank.

Phenotype	MR Estimator	N_{SNPs}	OR	(95% CI)	P
	IVW-RE	76	0.99	(0.55 – 1.80)	0.984
	Maximum likelihood	76	0.99	(0.59 – 1.68)	0.982
	Weighted median	76	0.75	(0.36 – 1.58)	0.450
	MR RAPS	76	1.02	(0.55 – 1.88)	0.958
	MR Egger slope	76	0.87	(0.08 – 9.65)	0.912
FEV_1	Outliers filtered:				
	IVW-RE	68	1.11	(0.69 – 1.80)	0.669
	Maximum likelihood	68	1.11	(0.64 – 1.93)	0.704
	Weighted median	68	0.81	(0.37 – 1.79)	0.606
	MR RAPS	68	1.13	(0.64 – 2.00)	0.674
	MR Egger slope	68	1.29	(0.15 – 11.07)	0.816
	IVW-RE	57	0.79	(0.37 – 1.66)	0.527
	Maximum likelihood	57	0.78	(0.41 – 1.49)	0.452
	Weighted median	57	0.54	(0.21 – 1.41)	0.211
	MR RAPS	57	0.66	(0.30 – 1.44)	0.296
	MR Egger slope	57	0.58	(0.02 – 18.48)	0.758
FVC	Outliers filtered:				
	IVW-RE	52	0.60	(0.33 – 1.10)	0.101
	Maximum likelihood	52	0.59	(0.30 – 1.17)	0.130
	Weighted median	52	0.52	(0.20 – 1.36)	0.181
	MR RAPS	52	0.57	(0.28 – 1.15)	0.115
	MR Egger slope	52	0.42	(0.02 – 7.83)	0.565
	IVW-RE	112	1.60	(1.05 – 2.45)	0.030
	Maximum likelihood	112	1.61	(1.10 – 2.35)	0.014
	Weighted median	112	1.41	(0.78 – 2.57)	0.254
	MR RAPS	112	1.57	(1.03 – 2.37)	0.035
	MR Egger slope	112	0.86	(0.25 – 2.99)	0.809
FEV_1/FVC	Outliers filtered:				
	IVW-RE	103	1.55	(1.09 – 2.19)	0.014
	Maximum likelihood	103	1.56	(1.05 – 2.30)	0.027
	Weighted median	103	1.47	(0.83 – 2.61)	0.186
	MR RAPS	103	1.54	(1.03 – 2.32)	0.035
	MR Egger slope	103	1.50	(0.47 – 4.83)	0.494

IVW – RE : Inverse variance weighted – multiplicative random effects
RAPS : Robust adjusted profile score – Huber loss function

Supplementary Table 11: Mendelian randomization (MR) odds ratio (OR) estimates for risk of lung cancer in smokers (23223 cases, 16964 controls) corresponding to a 1-SD decrease in FEV_1 and FVC z-scores and a 10% decrease in FEV_1/FVC. Genetic instruments for pulmonary function were not specific to smokers.

Phenotype	MR Estimator	N_{SNPs}	OR	(95% CI)	P
	IVW-RE	188	1.26	(0.98 – 1.63)	0.075
	Maximum likelihood	188	1.26	(1.06 – 1.50)	9.1×10^{-3}
	Weighted median	188	1.12	(0.85 – 1.47)	0.433
	MR RAPS	188	1.16	(0.90 – 1.48)	0.246
	MR Egger slope	188	1.76	(0.69 – 4.51)	0.239
FEV_1	Outliers filtered:				
	IVW-RE	163	1.09	(0.91 – 1.31)	0.363
	Maximum likelihood	163	1.09	(0.90 – 1.31)	0.368
	Weighted median	163	1.06	(0.81 – 1.39)	0.666
	MR RAPS	163	1.05	(0.86 – 1.29)	0.600
	MR Egger slope	163	0.81	(0.41 – 1.61)	0.551
	IVW-RE	139	1.20	(0.91 – 1.58)	0.206
	Maximum likelihood	139	1.20	(0.97 – 1.48)	0.100
	Weighted median	139	1.10	(0.80 – 1.53)	0.546
	MR RAPS	139	1.10	(0.85 – 1.43)	0.472
	MR Egger slope	139	1.05	(0.33 – 3.32)	0.935
FVC	Outliers filtered:				
	IVW-RE	124	1.09	(0.88 – 1.33)	0.434
	Maximum likelihood	124	1.09	(0.87 – 1.36)	0.465
	Weighted median	124	1.07	(0.77 – 1.48)	0.685
	MR RAPS	124	1.08	(0.86 – 1.36)	0.525
	MR Egger slope	124	0.62	(0.24 – 1.61)	0.329
	IVW-RE	258	1.21	(1.01 – 1.46)	0.039
	Maximum likelihood	258	1.21	(1.06 – 1.39)	4.5×10^{-3}
	Weighted median	258	0.96	(0.78 – 1.19)	0.711
	MR RAPS	258	1.13	(0.94 – 1.35)	0.185
	MR Egger slope	258	0.92	(0.57 – 1.48)	0.722
FEV_1/FVC	Outliers filtered:				
	IVW-RE	223	1.15	(1.01 – 1.31)	0.038
	Maximum likelihood	223	1.15	(0.99 – 1.33)	0.062
	Weighted median	223	1.08	(0.87 – 1.34)	0.488
	MR RAPS	223	1.14	(0.98 – 1.32)	0.094
	MR Egger slope	223	1.01	(0.67 – 1.51)	0.972

IVW – RE : Inverse variance weighted – multiplicative random effects
RAPS : Robust adjusted profile score – Huber loss function

Supplementary Table 12: Summary of diagnostic tests conducted for Mendelian Randomization analyses

Phenotype Outcome Pair	MR Egger		Modified Cochran's Q		I^2_{GX}[1]	Mean F	MR Steiger Test[2]		Number of Outliers
	$\beta_{0\ Egger}$	p-value	Q statistic	p-value			Ratio	p-value	
FEV_1 Lung cancer	-0.004	0.53	585.58	2.1×10^{-41}	0.981	52.09	18.24	$<2\times10^{-16}$	36
FEV_1 Adenocarcinoma	0.001	0.83	326.84	3.4×10^{-9}	0.981	53.54	17.01	$<2\times10^{-16}$	23
FEV_1 Squamous	-0.015	0.12	505.33	1.1×10^{-30}	0.981	52.13	11.20	$<2\times10^{-16}$	34
FEV_1/FVC Lung cancer	4.4×10^{-4}	0.90	602.10	1.2×10^{-28}	0.984	63.25	43.25	$<2\times10^{-16}$	38
FEV_1/FVC Adenocarcinoma	-0.003	0.44	419.6	3.4×10^{-9}	0.985	66.11	38.65	$<2\times10^{-16}$	26
FEV_1/FVC Squamous	0.001	0.90	479.9	5.3×10^{-15}	0.984	62.60	28.33	$<2\times10^{-16}$	24
FVC Lung cancer	0.006	0.38	310.4	2.1×10^{-14}	0.979	48.02	13.60	$<2\times10^{-16}$	20
FVC Adenocarcinoma	0.005	0.55	240.2	5.0×10^{-7}	0.979	47.48	9.99	$<2\times10^{-16}$	19
FVC Squamous	0.012	0.30	304.4	2.2×10^{-14}	0.979	48.29	7.62	$<2\times10^{-16}$	18
FEV_1 Never smokers	0.002	0.91	98.54	0.036	0.977	45.03	3.87	3.2×10^{-7}	8
FEV_1/FVC Never smokers	0.014	0.30	142.82	0.023	0.983	59.38	10.58	$<2\times10^{-16}$	9
FVC Never smokers	0.005	0.86	78.07	0.027	0.976	40.82	2.22	4.1×10^{-4}	5
FEV_1 Smokers	-0.005	0.47	419.05	7.7×10^{-17}	0.981	52.02	12.19	$<2\times10^{-16}$	25
FEV_1/FVC Smokers	0.005	0.22	490.72	6.0×10^{-17}	0.984	62.13	27.12	$<2\times10^{-16}$	35
FVC Smokers	0.002	0.82	239.11	2.1×10^{-7}	0.979	48.77	8.76	$<2\times10^{-16}$	
After outlier filtering:									
FEV_1 Lung cancer	0.004	0.36	153.53	0.52	0.980	50.46	21.47	$<2\times10^{-16}$	
FEV_1 Adenocarcinoma	0.004	0.48	153.06	0.79	0.981	52.65	18.77	$<2\times10^{-16}$	
FEV_1 Squamous	-0.004	0.50	141.12	0.80	0.981	52.55	14.09	$<2\times10^{-16}$	
FEV_1/FVC Lung cancer	-0.002	0.44	200.22	0.88	0.984	61.29	47.97	$<2\times10^{-16}$	
FEV_1/FVC Adenocarcinoma	-0.004	0.22	219.37	0.80	0.985	65.64	40.71	$<2\times10^{-16}$	
FEV_1/FVC Squamous	0.002	0.66	228.8	0.64	0.984	61.47	30.67	$<2\times10^{-16}$	
FVC Lung cancer	0.013	0.03	106.2	0.86	0.979	46.83	16.09	$<2\times10^{-16}$	
FVC Adenocarcinoma	0.001	0.93	111.2	0.77	0.978	46.04	10.81	$<2\times10^{-16}$	
FVC Squamous	0.014	0.10	110.4	0.72	0.979	47.99	9.23	$<2\times10^{-16}$	
FEV_1 Never smokers	-0.003	0.89	52.68	0.90	0.978	45.40	4.49	2.4×10^{-12}	
FEV_1/FVC Never smokers	0.001	0.96	81.84	0.93	0.983	59.64	11.56	$<2\times10^{-16}$	
FVC Never smokers	0.001	0.97	45.45	0.73	0.976	40.40	2.45	2.7×10^{-7}	
FEV_1 Smokers	-0.005	0.39	160.75	0.52	0.980	50.70	13.92	$<2\times10^{-16}$	
FEV_1/FVC Smokers	0.002	0.49	181.48	0.98	0.983	59.27	29.55	$<2\times10^{-16}$	
FVC Smokers	0.008	0.24	105.00	0.88	0.979	48.74	10.25	$<2\times10^{-16}$	

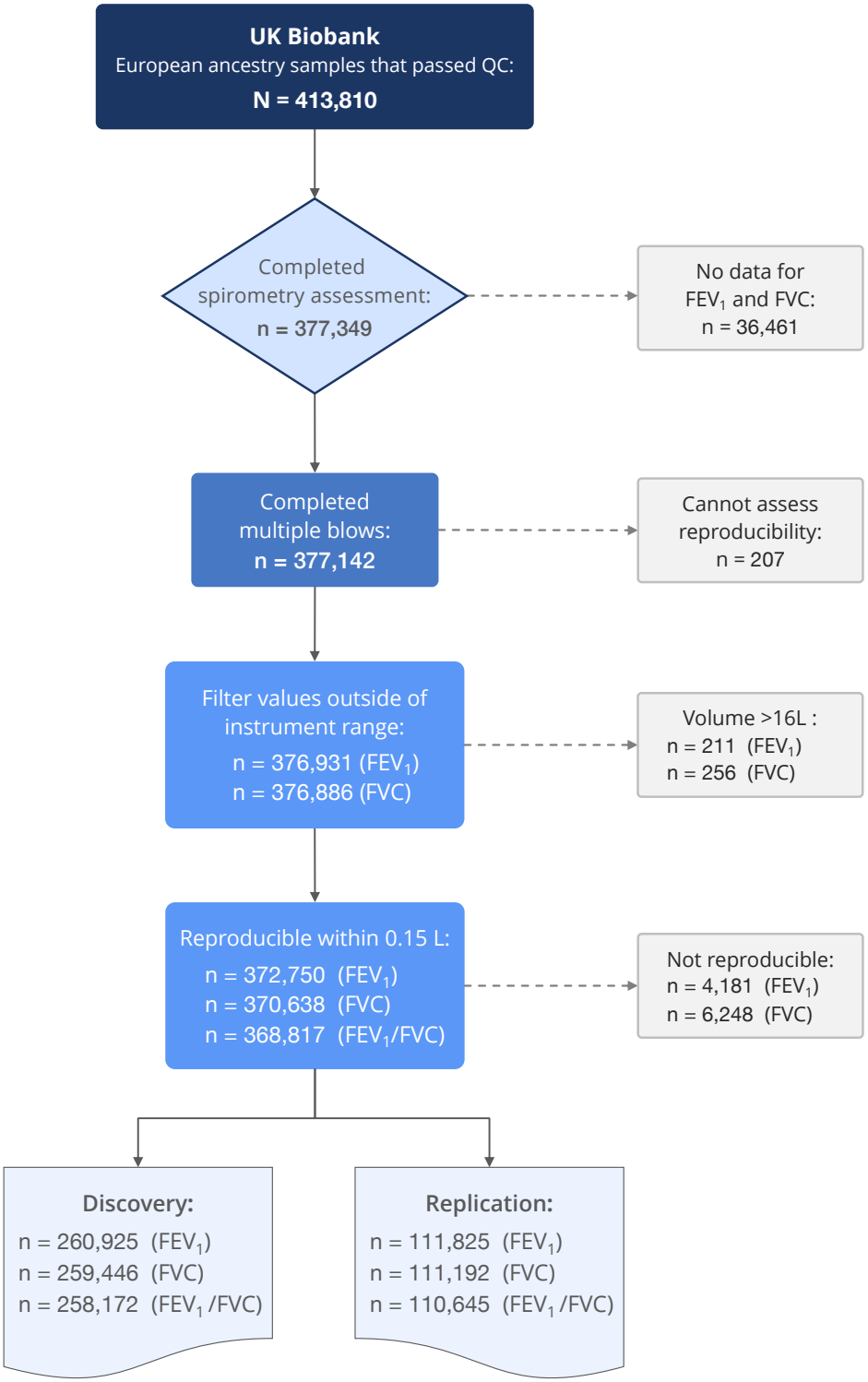
1. Values close to 1 indicate that the no measurement error (NOME) assumption is unlikely to be violated. Low I^2_{GX} estimates, less than 0.9, suggest that inferences should be interpreted with caution and adjustment methods should be considered
2. A higher ratio value indicates that the inferred direction is less susceptible to measurement error

Supplementary Table 13: Odds ratio (OR) estimates for a genetically predicted 1-SD decrease in standardized FEV_1 z-scores and 10% decrease in FEV_1/FVC after manual filtering to remove instruments associated with smoking status (ever/never), cigarette pack-years, or bodyfat percentage ($p<1\times10^{-5}$).

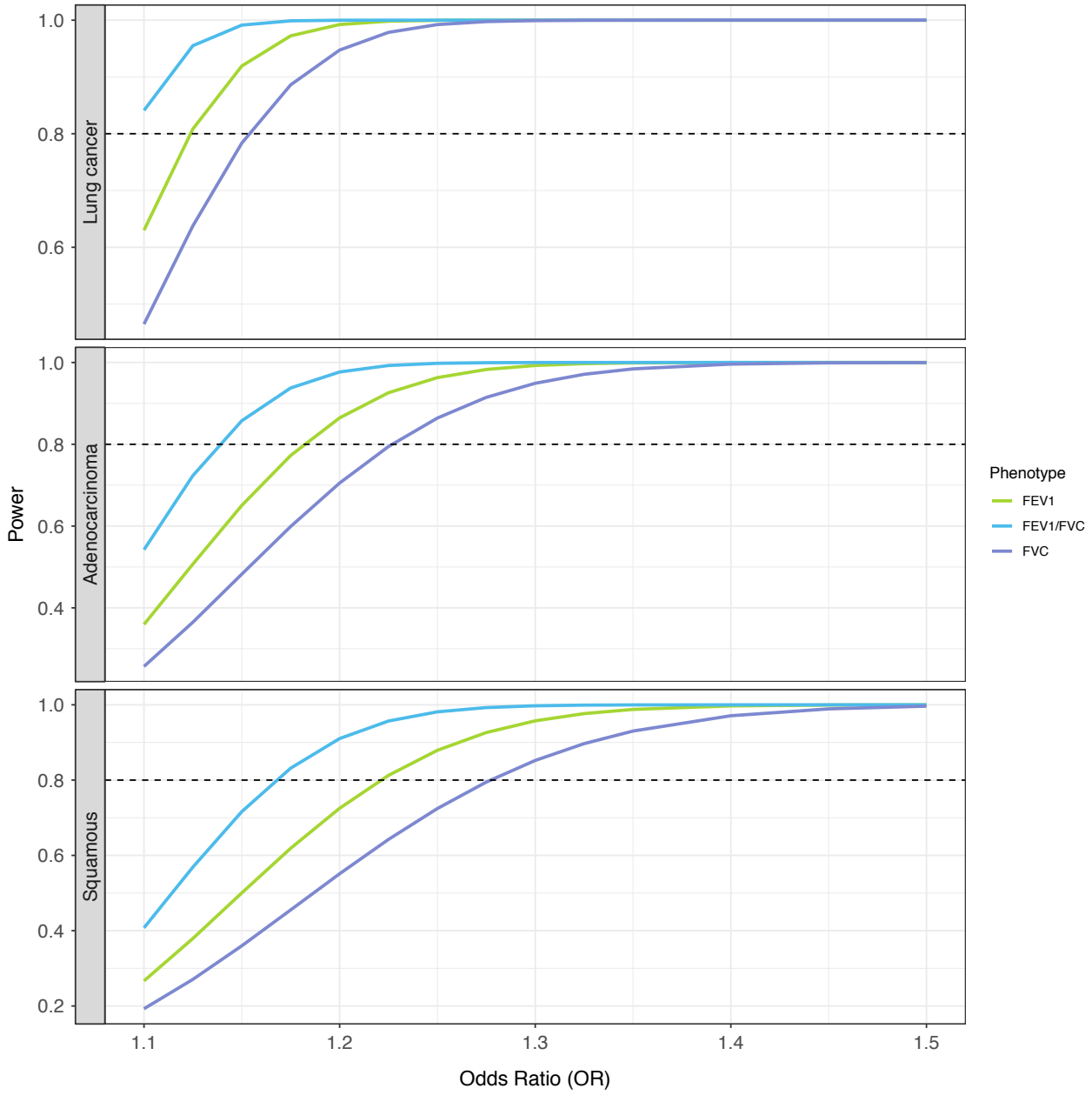
MR Estimator	FEV_1 and Lung Cancer							
	Manually filtered				Manual filter and Outlier detection			
	N_{SNPs}	OR	(95% CI)	P-value	N_{SNPs}	OR	(95% CI)	P-value
IVW-RE	168	1.25	(0.98 - 1.59)	0.067	141	1.08	(0.93 - 1.26)	0.320
Maximum likelihood	168	1.25	(1.08 - 1.45)	0.002	141	1.08	(0.93 - 1.27)	0.321
Weighted median	168	1.05	(0.84 - 1.31)	0.661	141	1.04	(0.82 - 1.32)	0.766
MR RAPS	168	1.11	(0.89 - 1.38)	0.364	141	1.05	(0.89 - 1.25)	0.564
Egger intercept	168	-0.011		0.091	141	0.002		0.699
Q statistic	168	484.7		1.0×10^{-32}	141	141.0		0.46
	FEV_1 and Squamous Carcinoma							
	Manually filtered				Manual filter and Outlier detection			
	N_{SNPs}	OR	(95% CI)	P-value	N_{SNPs}	OR	(95% CI)	P-value
IVW-RE	165	2.02	(1.40 - 2.92)	1.9×10^{-4}	137	1.46	(1.16 - 1.84)	0.001
Maximum likelihood	165	2.02	(1.60 - 2.55)	3.1×10^{-9}	137	1.46	(1.14 - 1.87)	0.003
Weighted median	165	1.47	(1.03 - 2.11)	0.035	137	1.36	(0.93 - 2.00)	0.111
MR RAPS	165	1.81	(1.29 - 2.56)	7.1×10^{-4}	137	1.41	(1.09 - 1.83)	0.009
Egger intercept	165	-0.02		0.048	137	-0.007		0.280
MR Egger slope	165	7.56	(1.97 – 29.04)	0.004	137	2.38	(0.95 – 5.94)	0.066
Q statistic	165	435.3		1.9×10^{-26}	137	120.4		0.827
	FEV_1/FVC and Lung Cancer							
	Manually filtered				Manual filter and Outlier detection			
	N_{SNPs}	OR	(95% CI)	P-value	N_{SNPs}	OR	(95% CI)	P-value
IVW-RE	246	1.19	(1.02 - 1.38)	0.03	214	1.09	(0.98 - 1.22)	0.103
Maximum likelihood	246	1.19	(1.07 - 1.32)	0.001	214	1.09	(0.97 - 1.22)	0.133
Weighted median	246	1.11	(0.93 - 1.33)	0.25	214	1.15	(0.96 - 1.39)	0.130
MR RAPS	246	1.12	(0.97 - 1.29)	0.13	214	1.09	(0.96 - 1.22)	0.175
Egger intercept	246	0.001		0.811	214	-0.002		0.444
Q statistic	246	527.1		9.8×10^{-23}	214	188.1		0.890
	FEV_1/FVC and Adenocarcinoma							
	Manually filtered				Manual filter and Outlier detection			
	N_{SNPs}	OR	(95% CI)	P-value	N_{SNPs}	OR	(95% CI)	P-value
IVW-RE	247	1.19	(1.01 - 1.40)	0.043	228	1.24	(1.08 - 1.43)	0.002
Maximum likelihood	247	1.19	(1.04 - 1.36)	0.013	228	1.24	(1.08 - 1.43)	0.002
Weighted median	247	1.34	(1.06 - 1.69)	0.013	228	1.40	(1.10 - 1.77)	0.006
MR RAPS	247	1.20	(1.02 - 1.42)	0.029	228	1.28	(1.10 - 1.48)	0.001
Egger intercept	247	-0.005		0.182	228	-0.006		0.062
Q statistic	247	387.1		3.7×10^{-8}	228	208.5		0.806

IVW – RE : Inverse variance weighted – multiplicative random effects
RAPS : Robust adjusted profile score – Huber loss function

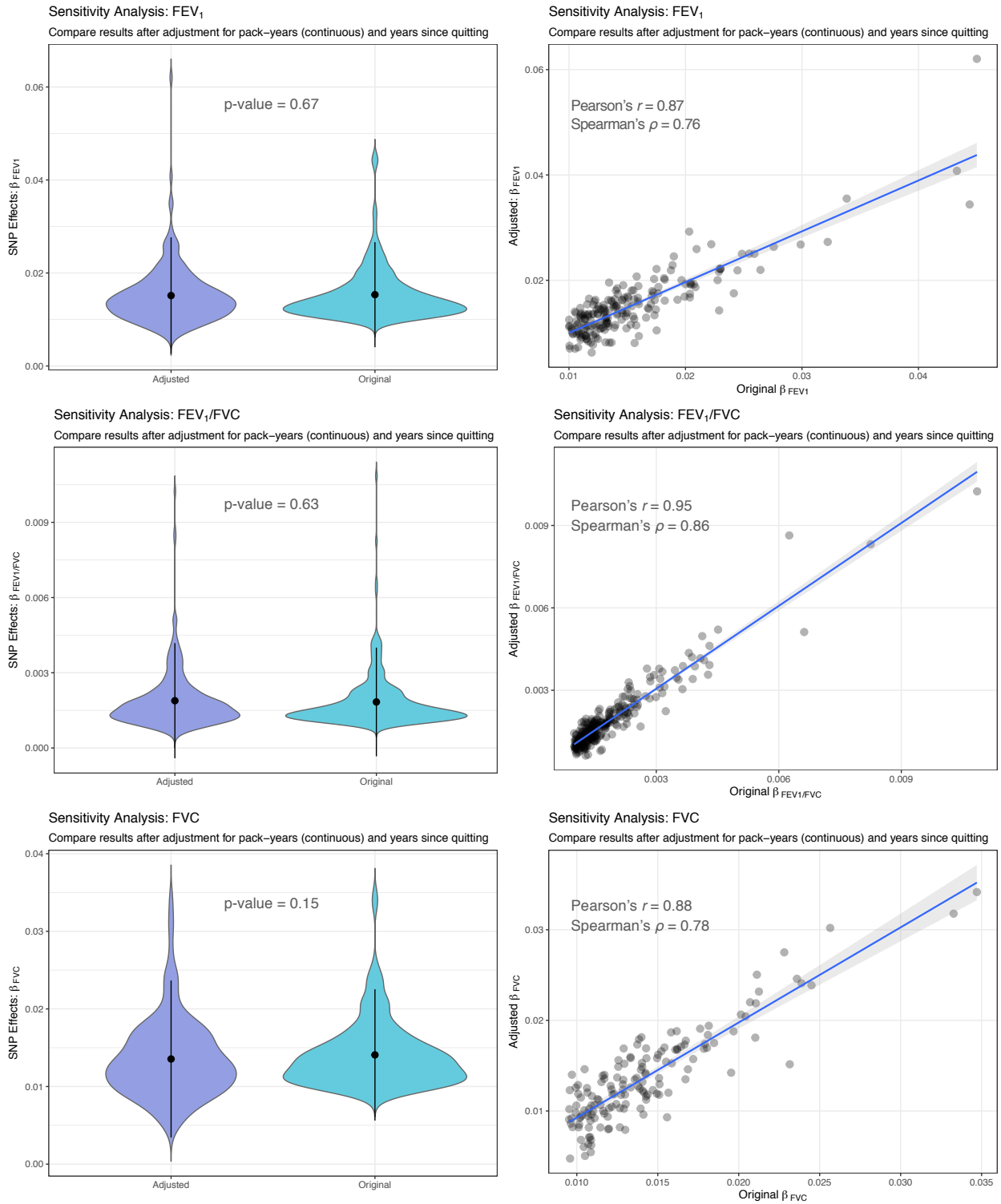
Supplementary Figure 1: Flow chart detailing the main quality control (QC) steps for the genome-wide association analyses of pulmonary phenotypes in the UK Biobank



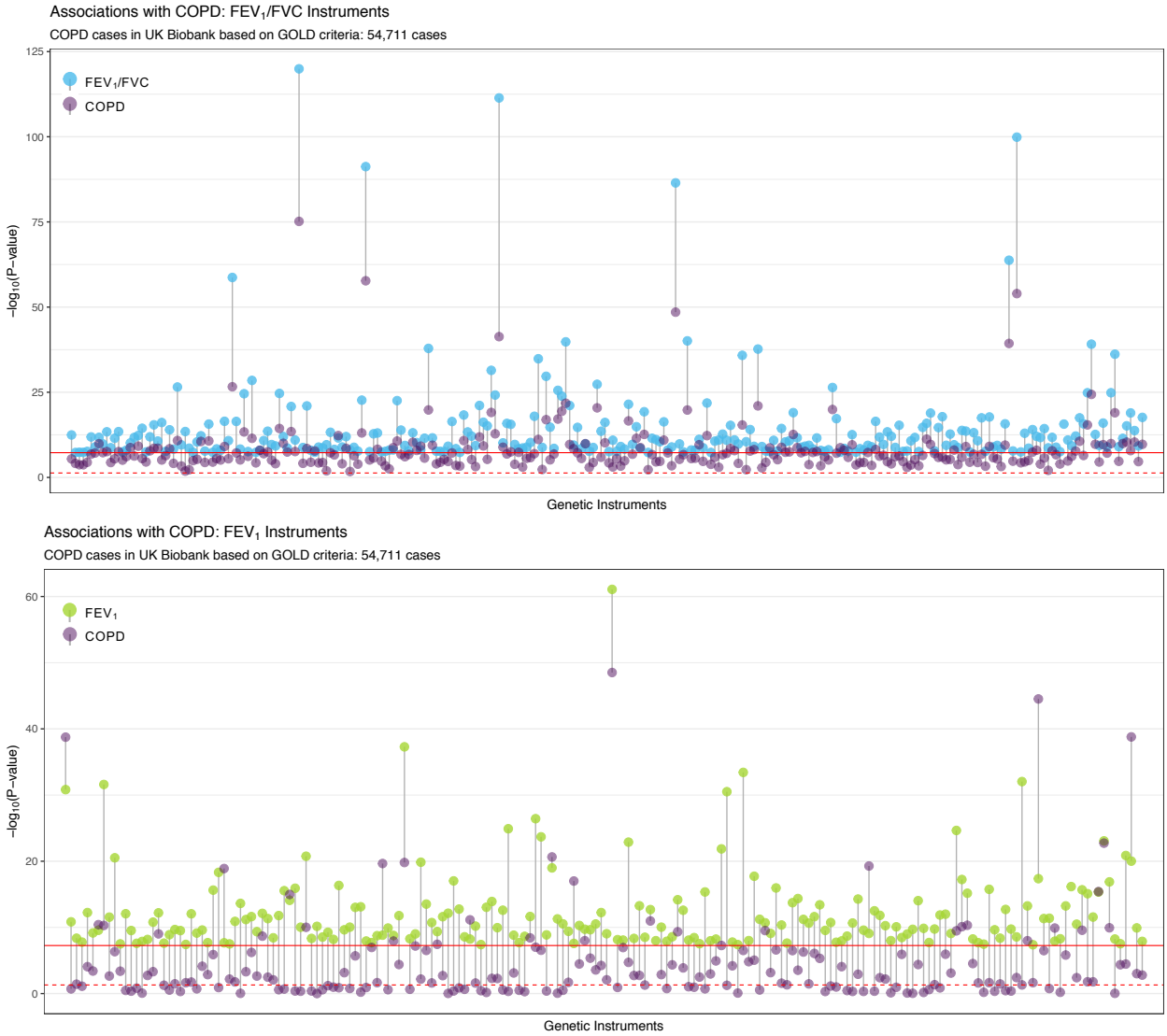
Supplementary Figure 2: Estimated power for Mendelian randomization analyses, based on the available sample size in the OncoArray dataset (29,266 lung cancers, 11,273 adenocarcinoma and 7,426 squamous cell carcinoma cases, and 56,450 controls), and assuming that the genetic instruments explain 5%, 3%, and 2% of variation in FEV_1/FVC, FEV_1, and FVC, respectively.



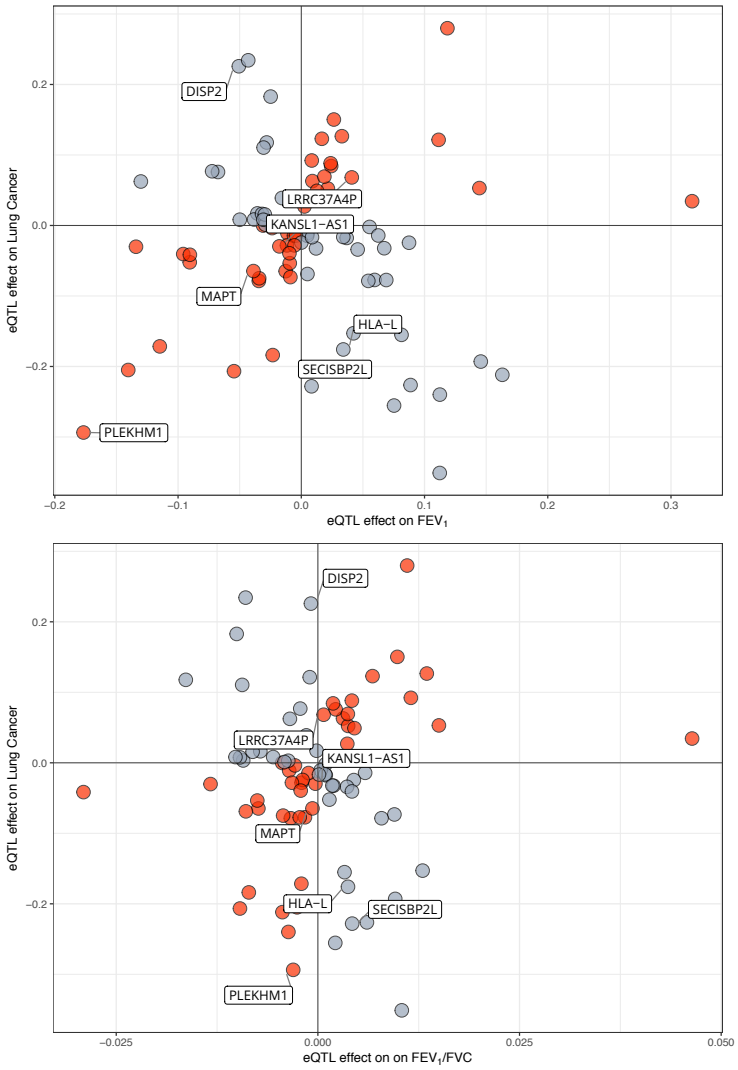
Supplementary Figure 3: Comparison of SNP effect sizes for pulmonary function genetic instruments estimated in the full UK Biobank cohort (original) and in ever smokers only (n=139,562 to 138,019) with adjustment for continuous cigarette pack-years rather than pack-year categories and years since quitting smoking (adjusted). Differences in the effect size distribution were tested using Wilcoxon rank sum test. All correlation coefficients were significant at $p<1\times10^{-40}$.



Supplementary Figure 4: Comparison of associations with chronic obstructive pulmonary disease (COPD), defined as $FEV_1/FVC<0.70$) for FEV_1/FVC and FEV_1 genetic instruments. Effect estimates for COPD were estimated in the UK Biobank dataset using logistic regression with adjustment for age, sex, height, cigarette pack-years, genotyping array, and genetic ancestry principal components (PC1-PC15). A total of 54,711 cases and 313,640 controls were included in the analysis. The red solid line corresponds to genome-wide significance ($p=5\times10^{-8}$) and the red dotted line indicates nominal significance (p=0.05).



Supplementary Figure 5: Analyses of eQTL effects based on lung tissue gene expression data from the Laval Direction of eQTL effects was considered consistent if increased expression resulted in impaired pulmonary function and increased lung cancer risk (or increased FEV_1 or FEV_1/FVC and an inverse association with lung cancer). Lung cancer associations with $p<5\times10^{-4}$ were considered statistically significant and are labeled in the scatterplot. Odds ratios (OR) for lung cancer per unit increase in predicted gene expression, and corresponding associations with FEV_1 and FEV_1/FVC, are presented for all genes achieving $p<0.05$ in the lung cancer eQTL Mendelian randomization analysis.

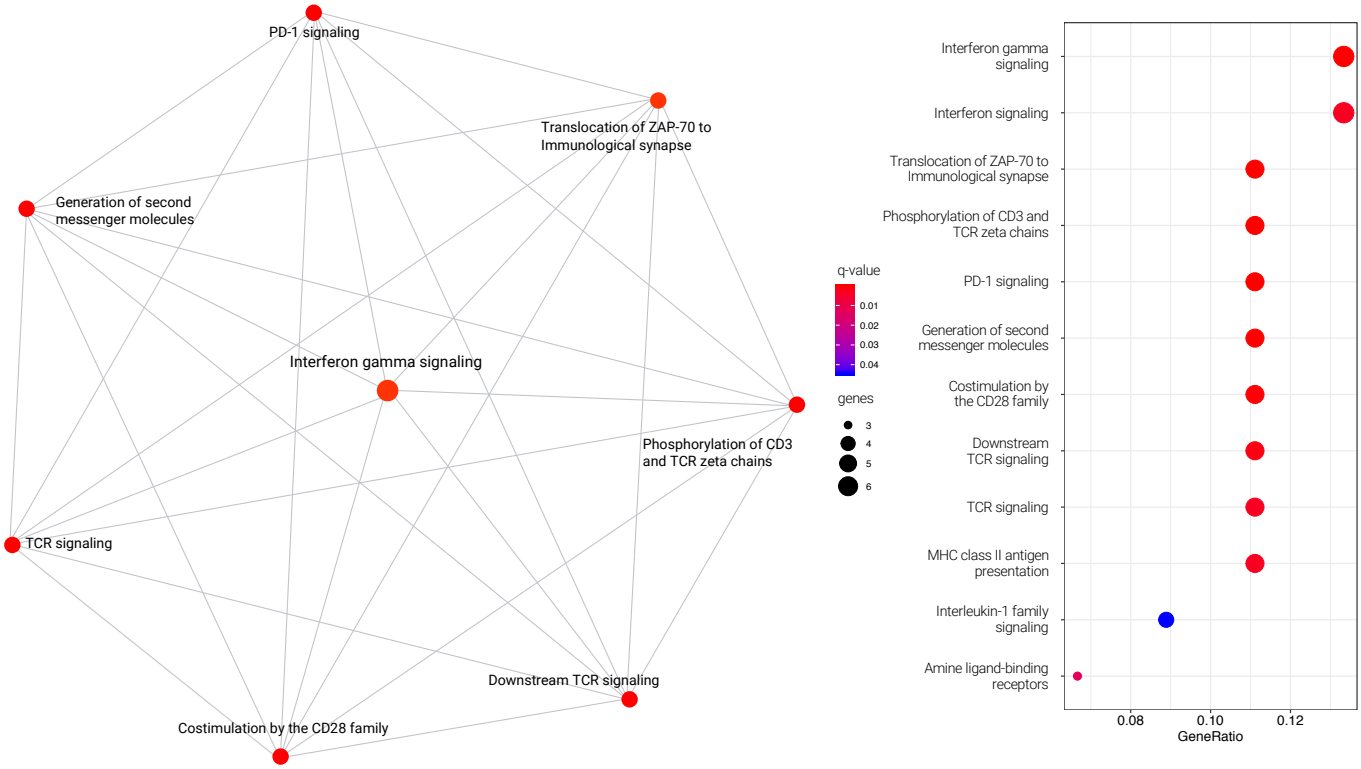


Gene	OR	(95% CI)	$P_{Lung\ cancer}$	P_{FEV_1}	$P_{FEV_1/FVC}$
SECISBP2L	0.80	(0.73-0.86)	5.2×10^{-8}	0.191	3.8×10^{-11}
HLA-L	0.84	(0.78-0.90)	1.6×10^{-6}	4.0×10^{-10}	1.2×10^{-11}
DISP2	1.25	(1.11-1.41)	1.6×10^{-4}	3.9×10^{-8}	0.354
MAPT	0.94	(0.90-0.97)	4.6×10^{-4}	3.0×10^{-62}	4.3×10^{-3}
KANSL1-AS1	0.97	(0.95-0.99)	4.6×10^{-4}	3.0×10^{-62}	4.3×10^{-3}
LRRC37A4P	1.07	(1.03-1.11)	4.6×10^{-4}	3.0×10^{-62}	4.3×10^{-3}
PLEKHM1	0.75	(0.63-0.88)	4.7×10^{-4}	2.9×10^{-62}	4.3×10^{-3}
FBXL19	0.77	(0.65-0.92)	4.4×10^{-3}	2.9×10^{-8}	0.116
AZGP1	1.06	(1.01-1.12)	0.012	0.020	1.3×10^{-15}
TRIM4	0.83	(0.72-0.96)	0.012	4.9×10^{-3}	8.1×10^{-25}
WNT3	0.84	(0.74-0.97)	0.014	9.0×10^{-33}	0.035
RBM6	0.80	(0.66-0.96)	0.016	8.2×10^{-10}	3.1×10^{-5}
HLA-DPA1	1.09	(1.01-1.18)	0.023	7.9×10^{-9}	4.9×10^{-24}
HLA-DPB1	0.81	(0.68-0.97)	0.023	7.9×10^{-9}	4.9×10^{-24}
HSPA4	1.32	(1.04-1.68)	0.023	2.5×10^{-10}	4.9×10^{-9}
UQCR10	1.20	(1.02-1.41)	0.027	0.046	1.4×10^{-15}
HHIPL1	1.26	(1.02-1.56)	0.030	7.9×10^{-3}	4.5×10^{-8}
PDXDC2P	0.98	(0.97-1.00)	0.033	8.3×10^{-25}	0.113
ITGA2	0.92	(0.86-0.99)	0.034	1.6×10^{-10}	3.0×10^{-16}
UQCC1	0.79	(0.63-0.99)	0.039	4.2×10^{-19}	4.1×10^{-3}
MDM4	0.86	(0.74-1.00)	0.044	2.2×10^{-11}	7.1×10^{-3}

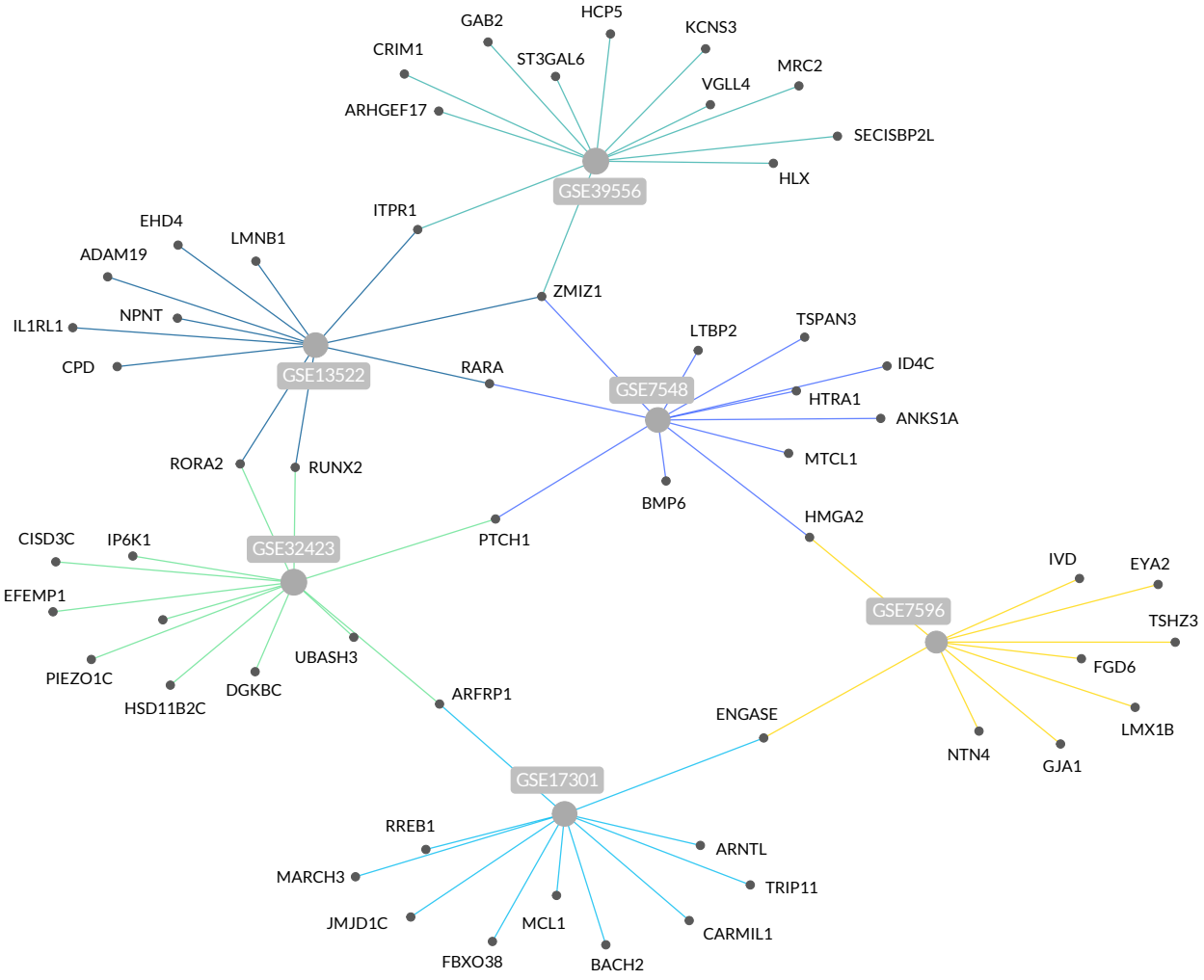
Histology-specific effects:

Gene	OR	(95% CI)	P_{Adeno}	Gene	OR	(95% CI)	P_{Squam}
SECISBP2L	0.64	(0.57-0.72)	3.1×10^{-14}	*HLA-L*	0.75	(0.67-0.84)	1.0×10^{-6}
PLEKHM1	0.64	(0.51-0.80)	8.8×10^{-5}	*DISP2*	1.30	(1.08-1.57)	6.2×10^{-3}
LRRC37A4P	1.11	(1.05-1.17)	8.9×10^{-5}	*MDM4*	0.74	(0.58-0.95)	0.017
KANSL1	0.96	(0.93-0.98)	8.9×10^{-5}	*HLA-DRB6*	1.06	(1.01-1.11)	0.022
MAPT	0.91	(0.86-0.95)	8.9×10^{-5}	*STRA13*	0.65	(0.43-0.97)	0.034
WNT3	0.74	(0.61-0.88)	1.1×10^{-3}	*LOC285835*	0.86	(0.74-0.99)	0.039
DISP2	1.21	(1.03-1.42)	0.021	*MARCH3*	0.89	(0.79-1.00)	0.048
MFAP2	0.87	(0.77-0.98)	0.026	*C5orf63*	0.76	(0.58-1.00)	0.048
HLA-L	0.90	(0.81-0.99)	0.034	*SECISBP2L*	1.05	(0.92-1.20)	0.44

Supplementary Figure 6: Visual summary of the Reactome pathway analysis depicting pathways that were significantly over-represented (FDR p<0.05) among genetic instruments for FEV_1/FVC in never smokers. The size of circles (nodes) is proportional to the number of genes included in each pathway and the colors correspond to the q-value for each pathway.



Supplementary Figure 7: Visual summary of the MSigDB pathway analysis depicting pathways that were significantly (FDR p<0.05) among genetic instruments for FEV_1 and FEV_1/FVC.



ID	Description	p-value	q-value
GSE32423	Up-regulated in memory CD8 T-cells treated with IL7 vs. IL4 and IL7	1.21×10^{-5}	0.024
GSE39556	Down-regulated in NK cells: untreated vs. pathogen-associated poly(IC)	1.21×10^{-5}	0.024
GSE13522	Down-regulated during intradermal infection: wildtype (BALB/c) vs. IFNAR1 knockout	1.91×10^{-5}	0.024
GSE7596	Down-regulated in T-cells in response to TGF-β: control vs. constantly active AKT1	2.16×10^{-5}	0.024
GSE17301	Up-regulated CD8 T-cells stimulated by INFA2 vs. IFNA5	6.19×10^{-5}	0.045
GSE7548	Up-regulated in lymph node CD4 T-cell: naïve vs. 28 days after immunization	6.19×10^{-5}	0.045