Multi-ancestry genome-wide association study improves resolution of genes, pathways and pleiotropy for lung function and chronic obstructive pulmonary disease

Nick Shrine[*1], Abril G Izquierdo[*1], Jing Chen[*1], Richard Packer[*1], Robert J Hall[2], Anna L Guyatt[1], Chiara Batini[1], Rebecca J Thompson[2], Chandan Pavuluri[3], Vidhi Malik[3], Brian D Hobbs[3,4], Matthew Moll[3], Wonji Kim[3], Ruth Tal-Singer[5], Per Bakke[6], Katherine A Fawcett[1], Catherine John[1], Kayesha Coley[1], Noemi Nicole Piga[1], Alfred Pozarickij[7], Kuang Lin[7], Iona Y Millwood[7,8], Zhengming Chen[7,8], Liming Li[9], Sara RA Wijnant[10,11,12], Lies Lahousse[11,12], Guy Brusselle[10,12], Andre G Uitterlinden[13], Ani Manichaikul[14], Elizabeth C Oelsner[15], Stephen S Rich[14], R. Graham Barr[15], Shona M Kerr[16], Veronique Vitart[17], Michael R Brown[18], Matthias Wielscher[19], Medea Imboden[20,21], Ayoung Jeong[20,21], Traci M Bartz[22], Sina A Gharib[23], Claudia Flexeder[24,25,26], Stefan Karrasch[24,25,26], Christian Gieger[26,27], Annette Peters[26,28], Beate Stubbe[29], Xiaowei Hu[14], Victor E Ortega[30], Deborah A Meyers[31], Eugene R Bleecker[31], Stacey B Gabriel[32], Namrata Gupta[32], Albert Vernon Smith[33,34], Jian'an Luan[35], Jing-Hua Zhao[36], Ailin F Hansen[37], Arnulf Langhammer[38,39], Cristen Willer[40,41,42], Laxmi Bhatta[37], David Porteous[43,44], Blair H Smith[45], Archie Campbell[46], Tamar Sofer[47,48,49], Jiwon Lee[47], Martha L Daviglus[50], Bing Yu[51], Elise Lim[52], Hanfei Xu[52], George T O'Connor[53], Gaurav Thareja[54], Omar M E Albagha[55,56], Hamdi Mbarek[57], Karsten Suhre[54,58], Raquel Granell[59], Tariq O Faquih[60], Pieter S Hiemstra[61], Annelies M Slats[61], Benjamin H Mullin[62,63], Jennie Hui[64,65,66], Alan James[64], John Beilby[64,63], Karina Patasova[67,68], Pirro Hysi[67,69], Jukka T Koskela[70], Annah B Wyss[71], Jianping Jin[72], Sinjini Sikdar[73,71], Mikyeong Lee[71], Sebastian May-Wilson[74], Nicola Pirastu[74], Katherine A Kentistou[74,75], Peter K Joshi[74], Paul RHJ Timmers[74], Alexander T Williams[1], Robert C Free[76,77], Xueyang Wang[78,77], John L Morrison[79], Frank D Gilliland[79], Zhanghua Chen[79], Carol A Wang[80,81], Rachel E Foong[82,83], Sarah E Harris[84], Adele Taylor[84], Paul Redmond[84], James P Cook[85], Anubha Mahajan[86,87], Lars Lind[88], Teemu Palviainen[89], Terho Lehtimäki[90], Olli T Raitakari[91,92], Jaakko Kaprio[89], Taina Rantanen[93], Kirsi H Pietiläinen[94,95], Simon R Cox[84], Craig E Pennell[80,81], Graham L Hall[82,83], W. James Gauderman[79], Chris Brightling[96,97], James F Wilson[74,98], Tuula Vasankari[99,100], Tarja Laitinen[101], Veikko Salomaa[102], Dennis O Mook-Kanamori[60,103], Nicholas J Timpson[104,59], Eleftheria Zeggini[105,106,107], Josée Dupuis[52], Caroline Hayward[108], Ben Brumpton[38,109], Claudia Langenberg[35], Stefan Weiss[110], Georg Homuth[110], Carsten Oliver Schmidt[111], Nicole Probst-Hensch[20,21], Marjo-Riitta Jarvelin[112,113,114,115], Alanna C Morrison[18], Ozren Polasek[116], Igor Rudan[117], Joo-Hyeon Lee[118,119], Ian Sayers[2], Emma L Rawlins[120], Frank Dudbridge[1], Edwin K Silverman[3], David P Strachan[121], Robin G Walters[7,8], Andrew P Morris[122], Stephanie J London[71], Michael H Cho[3], Louise V Wain[1,123], Ian P Hall[*2], Martin D Tobin[*1,123]

[1]Department of Health Sciences, University of Leicester, Leicester, LE1 7RH, UK.
[2]Division of Respiratory Medicine and NIHR-Nottingham Biomedical Research Centre, University of Nottingham.
[3]Channing Division of Network Medicine, Division of Pulmonary and Critical Care Medicine, Department of Medicine, Brigham and Women's Hospital, Boston, MA 02115.
[4]Harvard Medical School, Boston, MA 02115.
[5]COPD Foundation, Washington, DC.
[6]Department of Clinical Science, Unversity of Bergen, Norway.
[7]Nuffield Department of Population Health, University of Oxford, Oxford OX3 7LH, UK.
[8]MRC Population Health Research Unit, University of Oxford, Oxford OX3 7LH, UK.
[9]Department of Epidemiology and Biostatistics, School of Public Health, Peking University Health Science Center, Beijing, China.
[10]Department of Respiratory Diseases, Ghent Universital Hospital, Ghent (9000), Belgium.

[11]Department of Bioanalysis, Faculty of Pharmaceutical Sciences, Ghent University, Ghent (9000), Belgium.
[12]Department of Epidemiology, Eramus Medical Center, Rotterdam, the Netherlands.
[13]Department of Internal Medicine, Eramus Medical Center, Rotterdam, the Netherlands.
[14]Center for Public Health Genomics, University of Virginia, Charlottesville, VA, 22908, USA.
[15]Department of Medicine, Columbia University Medical Center, New York, NY, 10032, USA.
[16]Medical Research Council Human Genetics Unit, Institute of Genetics and Cancer, University of Edinburgh, Edinburgh EH4 2XU, UK.
[17]Medical Research Council Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh EH4 2XU, UK.
[18]Human Genetics Center, Department of Epidemiology, Human Genetics, and Environmental Sciences, School of Public Health, The University of Texas Health Science Center at Houston, Houston, TX 77030, USA.
[19]Department of Epidemiology and Biostatistics, MRC-PHE Centre for Environment & Health, School of Public Health, Imperial College London, London, W2 1PG, UK.
[20]Department of Epidemiology and Public Health, Swiss Tropical and Public Health Institute, Basel, Switzerland (CH-4123).
[21]Department of Public Health, University of Basel, Basel, Switzerland (CH-4001).
[22]Cardiovascular Health Research Unit, Departments of Medicine and Biostatistics, University of Washington, Seattle, WA, 98101.
[23]Computational Medicine Core, Center for Lung Biology, Division of Pulmonary, Critical Care and Sleep Medicine, Department of Medicine, University of Washington, Seattle, WA, USA, 98109.
[24]Institute and Clinic for Occupational, Social and Environmental Medicine, University Hospital, LMU Munich, Munich, 80336, Germany.
[25]Comprehensive Pneumology Center Munich (CPC-M), Member of the German Center for Lung Research (DZL), Munich, 81377, Germany.
[26]Institute of Epidemiology, Helmholtz Zentrum München - German Research Center for Environmental Health, Neuherberg, 85764, Germany.
[27]Research Unit of Molecular Epidemiology, Helmholtz Zentrum München - German Research Center for Environmental Health, Neuherberg, 85764, Germany.
[28]Chair of Epidemiology, Institute for Medical Information Processing, Biometry and Epidemiology, Medical Faculty, Ludwig-Maximilians-University, Munich, 81377, Germany.
[29]Department of Internal Medicine B - Cardiology, Intensive Care, Pulmonary Medicine and Infectious Diseases, University Medicine Greifswald, 17475 Greifswald, Germany.
[30]Department of Internal Medicine, Division of Respiratory Medicine, Center for Individualized Medicine, Mayo Clinic, Scottsdale, AZ, 85259, USA.
[31]Department of Medicine, University of Arizona, Tucson, Arizona, 85724, USA.
[32]Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA.
[33]Department of Biostatistics, University of Michigan School of Public Health, Ann Arbor, MI 48109, USA.
[34]Center for Statistical Genetics, University of Michigan School of Public Health, Ann Arbor, MI 48109, USA.
[35]MRC Epidemiology Unit, Institute of Metabolic Science, University of Cambridge School of Clinical Medicine, Cambridge, CB2 0QQ, UK.
[36]Department of Public Health and Primary Care, School of Clinical Medicine, University of Cambridge, Cambridge, UK.
[37]K.G. Jebsen Center for Genetic Epidemiology, Department of Public Health and Nursing, NTNU, Norwegian University of Science and Technology, Trondheim, 7030, Norway.
[38]HUNT Research Centre, Department of Public Health and Nursing, NTNU, Norwegian University of Science and Technology, Levanger, 7600 Norway.
[39]Levanger Hospital, Nord-Trøndelag Hospital Trust, Levanger, Norway.
[40]Department of Internal Medicine, Division of Cardiology, University of Michigan, Ann Arbor, MI 48109, USA.
[41]Department of Biostatistics and Center for Statistical Genetics, University of Michigan, Ann Arbor, MI, USA.

[42]Department of Human Genetics, University of Michigan, Ann Arbor, MI, 48109, USA.
[43]Centre for Genomic and Experimental Medicine, Institute of Genetics & Molecular Medicine, University of Edinburgh, Western General Hospital, Edinburgh EH4 2XU, United Kingdom.
[44]Centre for Cognitive Ageing and Cognitive Epidemiology, Department of Psychology, The University of Edinburgh, 7 George Square, Edinburgh, EH8 9JZ, United Kingdom.
[45]Division of Population Health Sciences, Ninewells Hospital and Medical School, University of Dundee, Dundee DD1 9SY, United Kingdom.
[46]Centre for Genomic and Experimental Medicine, Institute of Genetics & Cancer, University of Edinburgh, Western General Hospital, Edinburgh EH4 2XU, United.
[47]Division of Sleep and Circadian Disorders, Brigham and Women's Hospital, Boston, MA, USA, 02115.
[48]Department of Medicine, Harvard Medical School, Boston, MA, USA, 02115.
[49]Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA, USA, 02115.
[50]Institute for Minority Health Research, University of Illinois at Chicago, Chicago, IL, 60612.
[51]Human Genetics Center, Department of Epidemiology, Human Genetics and Environmental Sciences, School of Public Health, University of Texas Health Science Center at Houston, Houston, TX, USA, 77030.
[52]Department of Biostatistics, Boston University School of Public Health, Boston, MA, USA, 02118.
[53]Pulmonary Center, School of Medicine, Boston University, Boston, MA, USA, 02118.
[54]Bioinformatics Core, Weill Cornell Medicine-Qatar, Education City, Doha, Qatar.
[55]College of Health and Life Sciences, Hamad Bin Khalifa University, Doha, Qatar.
[56]Center for Genomic and Experimental Medicine, Institute of Genetics and Cancer, University of Edinburgh, Edinburgh, UK.
[57]The Qatar Genome Program Research (QGPR) Consortium, Qatar Genome Program, Qatar Foundation, Doha, Qatar.
[58]Department of Biophysics and Physiology, Weill Cornell Medicine, New York, NY, USA.
[59]MRC Integrative Epidemiology Unit (IEU), Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, BS8 2BN, UK..
[60]Department of Clinical Epidemiology, Leiden University Medical Center, Leiden, The Netherlands.
[61]Department of Pulmonology, Leiden University Medical Center, Leiden, The Netherlands, 2300RC.
[62]Dept. of Endocrinology & Diabetes, Sir Charles Gairdner Hospital, Nedlands, Western Australia, Australia, 6009.
[63]School of Biomedical Sciences, University of Western Australia, Crawley, Western Australia, Australia, 6009.
[64]Busselton Population Medical Research Institute, QEII Medical Centre, Nedlands, Western Australia, Australia, 6009.
[65]School of Population and Global Health, University of Western Australia, Crawley, Western Australia, Australia, 6009.
[66]PathWest Laboratory Medicine of WA, Nedlands, Western Australia, Australia, 6009.
[67]Department of Twin Research and Genetic Epidemiology, King's College London School of Medicine, London, UK.
[68]Karolinska Institutet.
[69]University College London.
[70]Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, Finland, postcode 00014.
[71]Epidemiology Branch, National Institute of Environmental Health Sciences, National Institutes of Health, Department of Health and Human Services, Research Triangle Park, NC, USA, 27709.
[72]Westat, Durham, NC, USA, 27703.
[73]Old Dominion University, Department of Mathematics and Statistics, Norfolk, VA, USA 23529.
[74]Centre for Global Health Research, Usher Institute for Population Health Sciences and Informatics, University of Edinburgh, Edinburgh, EH8 9AG, UK.
[75]Centre for Cardiovascular Sciences, Queen's Medical Research Institute, University of Edinburgh, Edinburgh, EH16 4TJ, UK.
[76]NIHR Leicester Biomedical Research Centre, University of Leicester, Leicester, UK..

[77]Department of Respiratory Sciences, University of Leicester, Leicester, UK.
[78]NIHR Leicester Biomedical Research Centre, University of Leicester, Leicester, UK.
[79]Department of Population and Public Health Sciences, Keck School of Medicine,University of Southern California, Los Angeles CA, USA 90089.
[80]School of Medicine and Public Health, College of Health, Medicine and Wellbeing, University of Newcastle, New South Wales, Australia, NSW 2308.
[81]Hunter Medical Research Institute, New South Wales, Australia, NSW 2305.
[82]Wal-yan Respiratory Research Centre, Telethon Kids Institute, Perth, Western Australia, Australia, WA 6009.
[83]School of Allied Health, Faculty of Health Sciences, Curtin University, Perth, Western Australia, Australia, WA 6102.
[84]Lothian Birth Cohorts studies, Department of Psychology, University of Edinburgh, Edinburgh, EH8 9JZ, UK.
[85]Department of Health Data Science, University of Liverpool, Liverpool, UK.
[86]Wellcome Centre for Human Genetics, University of Oxford, Oxford, UK.
[87]Genentech, South San Francisco, CA, USA.
[88]Department of Medical Sciences, Uppsala University, Uppsala, Sweden.
[89]Institute for Molecular Medicine Finland - FIMM, University of Helsinki.
[90]Department of Clinical Chemistry, Fimlab Laboratories, and Finnish Cardiovascular Research Center - Tampere, Faculty of Medicine and Life Sciences, University of Tampere, Tampere 33520, Finland.
[91]Department of Clinical Physiology and Nuclear Medicine, Turku University Hospital, Turku 20521, Finland.
[92]Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku, Turku 20520, Finland.
[93]Faculty of Sport and Health Sciences, University of Jyvaskyla, POBox 35, 40014 University of Jyvaskyla, Jyvaskyla, Finland.
[94]Obesity Research Unit, Research Program for Clinical and Molecular Metabolism, Faculty of Medicine, University of Helsinki.
[95]Obesity Center, Abdominal Center, Helsinki University Hospital and University of Helsinki.
[96]Department of Infection, Inflammation and Immunity, Institute for Lung Health, University of Leicester, Leicester, UK.
[97]National Institute for Health Research, Leicester Respiratory Biomedical Research Unit, Glenfield Hospital, Leicester, UK.
[98]MRC Human Genetics Unit, Institute of Genetic and Cancer, University of Edinburgh, Western General Hospital, Crewe Road, Edinburgh, EH4 2XU, Scotland.
[99]Filha ry, Helsinki, Finland.
[100]University of Turku, Department of Respiratory diseases and allergology, Turku, Finland.
[101]Administration Center, Tampere University Hospital and University of Tampere, Tampere, Finland, postcode 33521.
[102]Dept. Of Public Health and Welfare, Finnish Institute for Health and Welfare, Helsinki, Finland.
[103]Department of Public Health and Primary Care, Leiden University Medical Center, Leiden, The Netherlands.
[104]ALSPAC, Department of Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, BS8 2BN, UK..
[105]Wellcome Sanger Institute, Hinxton, CB10 1SA, UK.
[106]Institute of Translational Genomics, Helmholtz Zentrum Muenchen - German Research Center for Environmental Health, Neuherberg, Germany.
[107]Technical University of Munich (TUM) and Klinikum Rechts der Isar, TUM School of Medicine, Munich, Germany.
[108]The Institute of Medical Sciences, Aberdeen Biomedical Imaging Centre, University of Aberdeen, Aberdeen AB25 2ZD, United Kingdom.
[109]Clinic of Medicine, St. Olavs Hospital, Trondheim University Hospital, Trondheim, 7030, Norway.

[110]Interfaculty Institute for Genetics and Functional Genomics, Department of Functional Genomics, University Medicine Greifswald, 17475 Greifswald, Germany.
[111]Institute for Community Medicine, SHIP/Clinical-Epidemiological Research, University Medicine Greifswald, 17475 Greifswald, Germany..
[112]Center for Life Course Health Research, Faculty of Medicine, University of Oulu, PO Box 8000, FI-90014 Oulun yliopisto, Finland.
[113]Biocenter Oulu, University of Oulu, Aapistie 5, 90220 Oulu, Finland.
[114]Unit of Primary Health Care, Oulu University Hospital, OYS, Kajaanintie 50, 90220 Oulu, Finland.
[115]Department of Epidemiology and Biostatistics, MRC–PHE Centre for Environment & Health, School of Public Health, Imperial College London, London, W2 1PG, UK.
[116]University of Split School of Medicine, Split, Croatia.
[117]Centre for Global Health, Usher Institute, University of Edinburgh, Edinburgh, UK.
[118]Jeffrey Cheah Biomedical Centre, Wellcome - MRC Cambridge Stem Cell Institute, University of Cambridge, Cambridge, UK.
[119]Department of Physiology, Development and Neuroscience, University of Cambridge, Cambridge, UK.
[120]Wellcome Trust/CRUK Gurdon Institute.
[121]Population Health Research Institute, St George's, University of London, Cranmer Terrace, London SW17 0RE, UK.
[122]Centre for Genetics and Genomics Versus Arthritis, Centre for Musculoskeletal Research, Division of Musculoskeletal and Dermatological Sciences, The University of Manchester, Manchester, UK.
[123]National Institute for Health Research, Leicester Respiratory Biomedical Research Centre, Glenfield Hospital, Leicester, LE3 9QP, UK.

*Contributed equally

Abstract

Lung function impairment underlies chronic obstructive pulmonary disease (COPD) and predicts mortality. In the largest multi-ancestry GWAS meta-analysis of lung function to date, comprising 580,869 participants, 1020 independent association signals identified 559 genes supported by ≥2 criteria from a systematic variant-to-gene mapping framework. These genes were enriched in 29 pathways. Individual variants showed heterogeneity across ancestries, age and smoking groups, and collectively as a genetic risk score (GRS) showed strong association with COPD across ancestry groups. We undertook phenome-wide association studies (PheWAS) for selected associated variants, and trait and pathway-specific GRS to infer possible consequences of intervening in pathways underlying lung function. We highlight new putative causal variants, genes, proteins and pathways, including those targeted by existing drugs. These findings bring us closer to understanding the mechanisms underlying lung function and COPD, and should inform functional genomics experiments and potentially future COPD therapies.

Introduction

Lung function - even within the normal range - predicts mortality and is a key diagnostic criterion for COPD[1], which has the highest prevalence of respiratory diseases globally[2] and lacks disease-modifying treatments. Whilst smoking and environmental risk factors for COPD are well known, and genetic susceptibility (heritability) is recognised, the molecular pathways underlying COPD are incompletely understood. In common with many other complex traits, there has been under-representation of diverse ancestries in genome-wide association studies (GWAS)[3] of lung function[4-6]. Multi-ancestry studies improve the power and fine-mapping resolution of GWAS, and ultimately the prospects for prediction, prevention, diagnosis and treatment in diverse populations[3,4,7].

Understanding of genes, proteins and pathways involved in diseases and disease-related traits underpins modern drug development. A high yield of genetic association signals, improved signal resolution and integration with functional evidence are all required to confidently identify causal genes and the variants and pathways that impact

gene function and regulation. Although datasets and *in-silico* tools to connect GWAS signals to causal genes are improving, the findings from different datasets and tools have lacked consensus[8,9], highlighting a need for frameworks to integrate functional evidence types and to compare findings[10].

Aggregation of genetic variants associated with lung function into a genetic risk score (GRS) provides a tool for COPD prediction[5]. When a GRS comprises a sufficient number of variants, partitioning the GRS according to the biological pathways the variants influence could provide a tool to explore their aggregated consequences across a wide range of traits through phenome-wide association studies (PheWAS). Just as PheWAS of individual genetic variants can predict the consequences of perturbation of specific protein targets, informing assessment of drug efficacy, drug safety and drug repurposing opportunities[11], PheWAS of pathway-partitioned GRS could inform the understanding of consequences of perturbing specific pathways.

Through the largest global assembly of lung function genomics studies to date we: (i) undertook a multi-ancestry meta-analysis of GWAS of lung function traits in 580,869 individuals to detect novel signals, improve fine-mapping and estimate the extent of heterogeneity in allelic effects attributable to ancestry; (ii) tested whether lung function signals were age-dependent or smoking-dependent, and assessed their relationship to height; (iii) investigated cell type and functional specificity of lung function association signals; (iv) fine-mapped signals through annotation-informed credible sets, integrating functional data such as respiratory cell-specific chromatin accessibility signatures; (v) applied a consensus-based framework to systematically investigate and identify putative causal genes, integrating eight locus-based or similarity-based criteria; (vi) developed and applied a GRS for the ratio of forced expiratory volume in 1 second to forced vital capacity (FEV_1/FVC) in different ancestries in UK Biobank and in COPD case-control studies; (vii) applied PheWAS to individual variants, GRS for each lung function trait, and GRS partitioned by pathway. Through these approaches we aimed to detect novel lung function signals and novel putative causal genes, and provide new insights into the mechanistic pathways underlying lung function, some of which may be amenable to drug therapy.

Results

We undertook genome-wide association analyses of forced expired volume in 1 second (FEV_1), forced vital capacity (FVC), FEV_1/FVC, and peak expiratory flow rate (PEF) from 49 cohorts (**Methods**, **Supplementary Table 1, Supplementary Table 2**). Our sample of up to 580,869 participants comprised individuals of African (AFR: N=8,590), American/Hispanic (AMR: N=14,668), East Asian (EAS: N=85,279), South Asian (SAS: N=10,093) and European ancestry (EUR: N=462,239, **Supplementary Figure 1a, 1b**). Adjustments were made for age, age^2, sex, and height in association testing within cohorts, and we accounted for population structure and, where appropriate, relatedness (**Methods** and **Supplementary Tables 2-4**). Genomic control was applied to each cohort before meta-analysis using the linkage disequilibrium (LD) score intercept[12]. After filtering and meta-analysis across multi-ancestry cohorts, 66.8M variants were available genome-wide for signal selection in each of 4 lung function traits, with genomic inflation factors λ of 1.025, 1.022, 0.984 and 0.996 for FEV_1, FVC, FEV_1/FVC and PEF respectively (**Supplementary Figures 2-3, Supplementary Table 5**).

1020 signals for lung function

After excluding 8 signals associated with smoking behaviour (**Supplementary Note**), and combining signals that colocalised across multiple traits, we identified 1020 distinct signals for lung function using a stringent threshold of $P<5\times10^{-9}$ (ref.[13], Figure 1a). Of these, 713 are novel with respect to the signals and studies described in the Supplementary Note (**Supplementary Table 6)**. These 1020 signals explain 33.0% of FEV_1/FVC heritability (21.3% for FEV_1, 17.3% for FVC, 21.4% for PEF, **Methods**).

To facilitate fine-mapping, we included larger and more diverse populations than previous lung function GWAS. We performed multi-ancestry meta-regression with MR-MEGA[7] , which incorporates axes of genetic ancestry as covariates to model heterogeneity (**Methods**). We then incorporated functional annotation for chromatin accessibility and transcription factor binding sites in respiratory-relevant cells and tissues, and enriched genomic annotations[14] to weight prior causal probabilities of association for putative causal variants (**Methods**). Overall reductions in credible set size and higher maximum posterior probabilities of association for the most likely causal variant in the credible set were evident after multi-ancestry meta-regression was employed and after functional annotation was incorporated (**Supplementary Figure 4**). Following fine-mapping, 438 (43%) signals had a single

putative causal variant (posterior probability >50%) and the median credible set size was 9 variants (**Supplementary Note**).

We assessed heterogeneity of variant associations attributable to ancestry utilising MR-MEGA. Of the 960 signals represented in ≥7 cohorts, 109 signals showed ancestry-correlated heterogeneity (P_{Het}<0.05, **Supplementary Figure 5**, **Supplementary Table 7**), more than expected by chance (binominal test, P = $3.93x10^{-15}$). Among these, five signals (rs9393688, rs28574670 (*LTBP4*), rs7183859 (*THSD4*), rs59985551 (*EFEMP1*), rs78101726 (*MECOM*)) showed significant ancestry-correlated heterogeneity (Bonferroni correction for 960 signals tested, P_{Het}< $5.21x10^{-5}$, **Supplementary Figures 6a-e**). The intronic variant rs7183859 in *THSD4*, which we previously implicated in lung function[15], showed larger effect size estimates in non-EUR ancestries, and in particular African ancestries (**P_{HET}**=$3.33x10^{-5}$, **Supplementary Figure 6c**).

We examined associations of lung function associated SNPs in children's cohorts (**Supplementary Table 8**) and tested for differences in the estimated effect sizes of lung function associated SNPs between children and adults, as well as between ever-smokers and never-smokers, in European individuals (**Methods**). Effect size estimates between children and adults were correlated (r from 0.51 for FEV_1/FVC to 0.79 for FEV_1**, Supplementary Figure 7**), although 113 signals showed nominal evidence (P <0.05) of age-dependent effects (more than expected by chance, binomial P = $2.56x10^{-13}$). Three signals showed age-dependent effects (Bonferroni-corrected P <$4.64x10^{-5}$): rs7977418 (*CCDC91*), rs34712979 (*NPNT*) and rs931794 (*HYKK*) (**Supplementary Table 9**). We observed 69 out of 1020 signals with nominal evidence (P <0.05) of smoking-dependent effects (**Supplementary Figure 8)**, more than expected by chance (binomial P= 0.0079). The intronic SNP rs7733410 in *HTR4*, a signal we previously reported for lung function[15], showed a 76.2% larger effect on FEV_1 in ever than never-smokers (P=$4.09x10^{-5}$, **Supplementary Table 10**). As height is a determinant of lung growth, we compared height and lung function associations and tested the impact of additional height adjustments for sentinel SNPs. We found no correlation between estimated effect sizes for height and lung function (**Supplementary Figure 9**), and the addition of $height^2$ and $height^3$ covariates had little impact on effect size estimates (**Supplementary Figure 10**).

Cell-type and functional specificity

We assessed whether our association signals were enriched for regulatory or functional features in specific cell types. Using stratified LD-score regression[16] we found enrichment of all histone marks we tested (H3K27ac, H3K9ac, H3K4me3, H3K4me1) in lung and smooth muscle containing cell lines (**Supplementary Table 11)**. Using GARFIELD[17] we assessed enrichment of our signals for DNaseI hypersensitivity sites (DHS) and chromatin accessibility peaks, showing enrichment in a wide variety of cell types, including higher enrichment in foetal and adult lung and blood for FEV_1, FEV_1/FVC, and PEF and fibroblast enrichment for FVC (**Supplementary Figure 11a**). Our signals were enriched for transcription factor footprints in foetal lung for FEV_1, FEV_1/FVC, and PEF, for footprints in skin for FVC, and also in blood for PEF (**Supplementary Figure 11b**). Genic annotation enrichment patterns were similar across all traits, with enrichment mainly in exonic, 3' UTR and 5' UTR regions (**Supplementary Figure 11c**). For all traits we saw enrichment for transcription start sites (TSS), weak enhancers, enhancers and promoter flanks, with cell types for weak enhancer enrichment including endothelial cells for FEV_1, FEV_1/FVC, and PEF (**Supplementary Figure 11d**). For transcription factor binding sites, we observed a similar enrichment pattern across all the lung function traits with the largest fold-enrichment in endothelial cells (**Supplementary Figure 11e**). We used ATAC-seq data for the above fine-mapping and also to describe enrichment of our signals in specific cell types. Our signals were enriched in ATAC-seq peaks (**Supplementary Note**) in matrix fibroblast 1 for FVC, matrix fibroblast 2 for FEV_1, myofibroblast for FEV_1, FEV_1/FVC, and PEF, and alveolar type 1 cells in FEV_1/FVC and genic annotations showed enrichment of exon variants for FEV_1, FEV_1/FVC, and 3' UTR variants for FEV_1 and FVC. We also found enrichment of transcription factor binding sites in lung across all phenotypes and in bronchus for FEV_1/FVC (**Supplementary Table 12**).

Identification of putative causal genes and causal variants

To systematically investigate and identify putative causal genes, we integrated orthogonal evidence, using eight locus-based or similarity-based criteria (**Supplementary Note**): (i) the nearest gene to the sentinel SNP; (ii) colocalisation of GWAS signal and eQTL or (iii) pQTL signals in relevant tissues (**Methods)**; (iv) rare variant association in whole exome sequencing in UK Biobank; (v) proximity to a gene for a Mendelian disease with a respiratory phenotype (+/-500kb); (vi) proximity to a human ortholog of a mouse knockout gene with a respiratory phenotype (+/-500kb); (vii) an annotation-informed credible set[14] containing a missense/deleterious/damaging variant with a posterior probability of association >50% and; (viii) the gene with the highest polygenic priority score (PoPS), a

method based on the assumption that causal genes on different chromosomes share similar functional characteristics[9]. We identified 559 putative causal genes satisfying at least two criteria, of which 135 were supported by at least three criteria (Figure 1b, Figure 2, **Supplementary Figure 12**). Among 20 genes supported by 4 or more criteria (**Supplementary Table 13**), six previously implicated genes (*TGFB2, NPNT, LTBP4, TNS1, SMAD3, AP3B1*) [5,15,18-20] were supported by additional criteria compared with the original reports. Fourteen of the 20 genes supported by 4 or more criteria have not been confidently implicated in lung function previously (*CYTL1, HMCN1, GATA5, ADAMTS10, IGHMBP2, SCMH1, GLI3, ABCA3, TIM1, CFH, FGFR1, LRBA, CLDN18, IGF2BP2*). These are involved in smooth muscle function (*FGFR1, GATA5, STIM1*), tissue organisation (*ADAMTS10*), alveolar and epithelial function (*ABCA3, CLDN18*), and inflammation and immune response to infection (*CFH, CYTL1, HMCN1, LRBA, STIM1*).

In order to supplement understanding of the biological pathways and range of clinical phenotypes that lung function associated variants influence, we undertook PheWAS of selected individual variants. We selected 27 putative causal genes implicated by ≥4 criteria (20 genes), or implicated by a single putative causal missense variant that was deleterious (5 genes: *ACAN, ADGRG6, SCARF2, CACNA1S, HIST1H2BE*) or rare (2 genes: *SOS2*, *ADRB2*, **Supplementary Table 14**). We interpreted the PheWAS findings (shown in full in **Supplementary Figure 13** and **Supplementary Table 15**) alongside literature reviews (**Supplementary Table 16)** for each of these 27 genes; examples are highlighted in the three paragraphs below.

The putative causal deleterious missense variant in *ABCA3* associated with reduced FEV_1/FVC, rs149989682 (A allele, frequency 0.6%), has been reported to cause paediatric interstitial lung disease[21]. ABCA3, expressed in alveolar type II cells and localised to lamellar bodies, is involved in surfactant phospholipid metabolism and several *ABCA3* mutations cause severe neonatal surfactant deficiency[22]. The putative causal deleterious missense variant rs200383755 C allele (frequency 0.6%) in *GATA5*, associated with lower FEV_1, was associated with increased risk of asthma, higher blood pressure and reduced risk of benign prostatic hyperplasia in our PheWAS. *GATA5* associations have not been previously noted in GWAS of asthma, although Gata5-deficient mice show airway hyperresponsiveness[23] (**Supplementary Figure 13i**). GATA5 is a transcription factor involved in smooth muscle cell diversity, expressed in bronchial smooth muscle, and highly expressed in bladder and prostate; a previous benign prostatic hyperplasia GWAS reported a *GATA5* signal [24 23]. *CLDN18* was implicated by 4 criteria, including a mouse knockout with abnormal pulmonary alveolar epithelium morphology[25]. Through calcium-independent cell-adhesion, *CLDN18* influences epithelial barrier function through tight junction-specific obliteration of the intercellular space[26], and its splice variant CLDN18.1 is predominantly expressed in the lung[27]. Reduced *CLDN18* expression has been reported in asthma[28]. However, our PheWAS showed no association with asthma susceptibility or other traits (CLDN18_rs182770 in **Supplementary Table 15**). *LRBA* was also implicated by 4 criteria. Mutations resulting in LRBA deficiency cause common variable immunodeficiency-8 with autoimmunity, which can include cough, respiratory infections, bronchiectasis, or interstitial lung disease [29 30]. Putative causal *LRBA* variant rs2290846 (posterior probability 56.3%) is a tolerated missense variant which showed pleiotropic associations with 31 associated traits (FDR<1%) in our PheWAS (**Supplementary Figure 13n, Supplementary Table 15**). The rs2290846 G allele, associated with lower FVC and lower FEV_1, was associated with lower neutrophils, lower risks of cholelithiasis and cholecystitis[31], and lower diverticular disease risk.

FGFR1, encoding Fibroblast Growth Factor Receptor 1, has roles in lung development and regeneration[32], and loss-of-function *FGFR1* mutations cause hypogonadotropic hypogonadism[33]. Notably, in our PheWAS, the T allele of rs881299, associated with lower FEV_1/FVC and higher FVC, is strongly associated with higher testosterone (particularly in males) and higher sex hormone binding globulin (SHBG), lower BMI, lower alanine transaminase and urate levels (**Supplementary Figure 13w-y, Supplementary Table 15**). Missense variant rs72681869 in *SOS2* also showed association with SHBG in our PheWAS. In both sexes the C allele of rs72681869, associated with higher FVC and higher FEV_1, was associated with lower SHBG, higher alanine aminotransferase (ALT) and aspartate aminotransferase (AST), higher fat mass, HbA1c and higher systolic and diastolic blood pressure, higher urate and creatinine, and in males lower testosterone, and reduced inguinal hernia risk (**Supplementary Figure 13z-bb)**. Mutations in *SOS2* have been reported in Noonan Syndrome. The A allele of rs7514261 implicating *CFH*, associated

with lower FVC, was strongly associated with reduced risk of macular degeneration[34] and also with raised albumin in our PheWAS (**Supplementary Figure 13g**).

CACNA1S is one of several genes prioritised encoding calcium voltage-gated channel subunits in skeletal muscle (*CACNA1S*, *CACNA1D*, and *CACNA2D3* supported by ≥2 criteria; *CACNA1C* was supported by PoPS). Mutations in *CACNA1S* have been reported to cause hypokalemic periodic paralysis[35] and malignant hyperthermia[36]. *CACNA1S* is strongly expressed in skeletal muscle, but at much lower levels in airway smooth muscle. The common *CACNA1S* missense variant, rs3850625 (A allele, frequency EUR 11.8%, SAS 21.4%) was associated with lower FVC, lower FEV_1, and in the PheWAS, with lower whole body fat-free mass, reduced hand grip strength, and lower aspartate aminotransferase and creatinine levels (**Supplementary Figure13f**). CACNA1S *and* CACNA1D are targeted by dihydropyridine calcium channel blockers, which have been reported to produce small improvements in lung function in asthma[37]. The low frequency missense variant rs1800888 in *ADRB2* (T, 1.49% EUR), associated with lower FEV_1 and lower FEV_1/FVC, showed strongest association in the PheWAS with increased eosinophil count.

Druggable targets

Using the Drug Gene Interaction Database (DGIDB), we surveyed 559 genes supported by ≥2 criteria. We found 292 drugs indicated by ChEMBL interactions mapping to 55 genes (**Supplementary Table 17**), including *ITGA2*, encoding Integrin Subunit Alpha 2. The reduced expression of ITGA2 in lung tissue with the C allele of rs12522114 mimics vatelizumab-induced ITGA2 inhibition; this allele is associated with higher FEV_1 and FEV_1/FVC, indicating a potential to repurpose vatelizumab, which increases T regulatory cell populations[38], for COPD.

Pathway analysis

Using ConsensusPathDB[39], we tested whether specific biological pathways were enriched for the 559 causal genes supported by 2 or more criteria, highlighting multiple pathways consistent with developmental pathways, tissue integrity and remodelling (**Supplementary Table 18**). These include pathways not previously implicated in pathway enrichment analyses for lung function such as PI3K-Akt signalling, integrin pathways, endochondral ossification, calcium signalling, hypertrophic cardiomyopathy, and dilated cardiomyopathy, as well as those previously implicated via individual genes[5] such as TNF signalling, actin cytoskeleton, AGE-RAGE signalling, Hedgehog signalling and cancers. We also show strengthened enrichment by newly identified genes in pathways we previously described, such as extracellular matrix organisation (34 new genes), elastic fibre formation (8 genes), and TGF-Core (4 new genes). Consistent with our ConsensusPathDB findings, Ingenuity Pathway Analysis (https://digitalinsights.qiagen.com/IPA)[40] highlighted enrichment of cardiac hypertrophy signalling and osteoarthritis pathways, and additionally implicated pulmonary and hepatic fibrosis signalling pathways, axonal guidance and PTEN signalling, and upstream regulators TGFB1 and IGF1 (**Supplementary Table 19**).

Multi-ancestry genetic risk score associations with FEV_1/FVC and COPD

We built multi-ancestry and ancestry-specific genetic risk scores weighted by FEV_1/FVC effect sizes and tested for association with FEV_1/FVC and COPD (GOLD stage 2-4) using independent testing datasets in different ancestry groups in UK Biobank (**Methods**). Our new GRS noticeably improved the predictive power for quantitative lung function and COPD compared with our previous GRS based only on European ancestry samples[5] (Figure 3a and 3b, **Supplementary Table 20**) and the multi-ancestry GRS outperformed the ancestry-specific GRS in all ancestry groups in UK Biobank. We then tested the association of the multi-ancestry GRS with COPD susceptibility in five independent COPD case-control studies (**Supplementary Table 21**, **Methods**). Improved association results were observed across all the five European ancestry studies compared with previous GRS[5] (Figure 3c, **Supplementary Table 22**). The odds ratio for COPD per standard deviation of the weighted GRS was 1.63 (95% CI: [1.56, 1.71], $P=7.1\times10^{-93}$) in the meta-analysis of these EUR studies compared to 1.55 (95% CI: [1.48, 1.62], $P=2.9\times10^{-75}$) using the previous GRS[5]. In SPIROMICS African ancestry individuals, results were comparable to UK Biobank African ancestry individuals, but of a lower magnitude in the COPDGene African ancestry population (Figure 3c).

To aid clinical interpretation, we divided individuals in each of the five European ancestry COPD case-control studies into ten deciles according to their values of the multi-ancestry GRS. The odds ratio for COPD in members of the highest GRS decile compared to the lowest GRS decile was 5.16 (95% CI: [4.14, 6.42], $P=1.0\times10^{-48}$, **Supplementary Table 23**). Across the multi-ancestry GRS deciles, the odds ratio for COPD showed a greater monotonic upward trend than across the previous GRS deciles (Figure 3d).

Phenome-wide associations of trait-specific genetic risk scores

To study the aggregate effects of genetic variants associated with each specific lung function trait on a wide range of diseases and disease-relevant traits, we created genetic risk scores (GRS) for each of , one for each trait FEV_1, FVC, FEV_1/FVC and PEF, and used each of these GRS in PheWAS. To construct each GRS, we included all sentinel variants associated with the trait ($P < 5 \times 10^{-9}$), using the weights estimated from the multi-ancestry meta-regression (**Methods**), for a total of 425, 372, 442 and 194 variants in each trait-specific GRS respectively.

GRS constructed from the four lung function traits showed distinct patterns of associations with a range of respiratory and non-respiratory phenotypes in our PheWAS (Figure 4**, Supplementary Table 24**). A GRS for lower FEV_1 was most strongly associated with increased risk of asthma and COPD, as well as family history of chronic bronchitis/emphysema, lower hand grip strength, increased fat mass, increased HbA1c and type 2 diabetes risk, and elevated C-reactive protein (CRP). Additionally, associations were seen with increased asthma exacerbations and lower age of onset for COPD (Figure 4a). The GRS for lower FEV_1/FVC was associated with key respiratory phenotypes: increased risk of COPD and asthma, increased family history of chronic bronchitis/emphysema, increased emphysema risk, and increased risk of respiratory insufficiency or respiratory failure, younger age of onset for COPD but a slightly lower risk of COPD exacerbations (Figure 4b). In contrast, the GRS for lower FVC was strongly associated with many traits – among the strongest associations were with high CRP, increased fat mass, raised HbA1c and type 2 diabetes, raised systolic blood pressure, lower hand grip strength and raised alanine aminotransferase, as well as showing increased risk of clinical codes for asthma and COPD (Figure 4c). Whilst the GRS for lower FEV_1/FVC was associated with increased standing height and sitting height, the GRS for lower FEV_1 and FVC were associated with increased standing height but reduced sitting height. Broadly similar phenome-wide associations were seen for the PEF GRS as for the FEV_1 GRS (Figure 4d).

Phenome-wide associations of genetic risk scores partitioned by pathway

Finally, we hypothesised that partitioning our lung function GRS into pathway-specific GRSs according to the biological pathways the variants influence could inform understanding of mechanisms underlying lung function and COPD, and the likely consequences of perturbing specific pathways. Informed by the above prioritisation of putative causal genes and classification of these genes by pathway (**Pathway analysis**, above), we undertook PheWAS for FEV_1/FVC GRSs partitioned by each of the 29 pathways enriched (FDR $<10^{-5}$) for the 559 genes implicated by ≥2 criteria. In each case, we weighted the GRSs using the FEV_1/FVC multi-ancestry meta-regression results (**Methods**). Partitioning GRSs in this way highlighted marked differences in patterns of phenome-wide associations (full results in **Supplementary Figure 14** and **Supplementary Table 25**). We highlight four examples in Figure 5; whilst all four pathway-specific GRSs illustrated showed association with COPD clinical codes and with a family history of chronic bronchitis/emphysema, associations with other traits varied. The GRS for lower FEV_1/FVC specific to elastic fibre formation was associated with increased risk of inguinal, abdominal, diaphragmatic and femoral hernia, diverticulosis, arthropathies, hallux valgus and genital prolapse, but reduced risk of carpal tunnel syndrome, as well as reduced BMI, and increased asthma risk (Figure 5a). In contrast, the GRS for lower FEV_1/FVC specific to PI3K-Akt signalling was associated with increased asthma risk, lower IGF-1, liver enzymes (ALT, AST, gamma glutamyltransferase (GGT)) and lower lymphocyte count, raised eosinophils, lower fat mass and BMI and reduced diabetes risk (Figure 5b). The GRS for lower FEV_1/FVC specific to the hypertrophic cardiomyopathy pathway was associated with reduced liver enzymes (ALT, GGT), lower apolipoprotein B and lower LDL, lower IGF-1 and lower mean platelet volume (Figure 5c). The GRS associations for lower FEV_1/FVC partitioned to signal transduction were specific to respiratory traits, including asthma and emphysema (Figure 5d). Variable height associations were evident: the GRS for lower FEV_1/FVC showed association with increased height when partitioned to elastic fibre formation or hypertrophic cardiomyopathy (Figure 5a,c), reduced height when partitioned to ESC pluripotency (**Supplementary Figure 14g**), and no association with height when partitioned to PI3K-Akt signalling or signal transduction (Figure 5b,d).

We hypothesised that individuals may have high GRS for one or more pathways and low GRS for other pathways. Comparing individuals' GRS across pairs of pathways for each of 29 pathways (**Supplementary Figure 15a**) and in detail for the elastic fibre, PI3K-Akt signalling, hypertrophic cardiomyopathy and signal transduction pathways (**Supplementary Figure 15b**) show how GRS profiles may be concordant or discordant across pathways, which could have implications for choice of therapies.

Discussion

Our study represents the largest and most ancestrally diverse GWAS of lung function to date and the most comprehensive initiative to relate lung function and COPD associated variants to functional annotations, cell types, genes and pathways. It is also the first to investigate possible phenotypic consequences of intervening in relevant pathways through PheWAS studies, utilising pathway-partitioned GRS.

The 1020 signals identified to date were enriched in functionally active regions in alveolar type 1 cells, fibroblasts and myofibroblasts, bronchial epithelial cells, adult and fetal lung. We showed effect heterogeneity attributable to ancestry for 109 signals (including *LTBP4*, *THSD4*, *EFEMP1*, *MECOM*), between ever-smokers and never-smokers (*HTR4*), and differences in effects between adults and children (including *CCDC91* and *NPNT*). We mapped lung function signals to 559 genes putatively inferred as causal based on meeting at least two independent criteria. Exemplar genes supported by ≥4 criteria or by deleterious or rare putative causal missense variants implicated surfactant phospholipid metabolism, smooth muscle function, epithelial morphology and barrier function, innate immunity, calcium signalling, adrenoceptor signalling, lung development and regeneration. Among the pathways enriched for the putative causal genes, were PI3K-Akt signalling, integrin pathways, endochondral ossification, calcium signalling, hypertrophic cardiomyopathy, and dilated cardiomyopathy that have not been previously implicated in lung function using GWAS approaches.

Combined as a genetic risk score weighted by FEV_1/FVC effect size, the 1020 variants strongly predicted COPD in UK Biobank and in COPD case-control studies, with a more than five-fold change in risk between highest and lowest GRS deciles, illustrating the clinical relevance of our findings. This GRS more strongly predicted FEV_1/FVC and COPD across all ancestries than a previously constructed risk score[5]. Partitioning this lung function GRS by the pathways defined by specific variants, informed by detailed, systematic variant-to-gene mapping and pathway analyses and using our new Deep-PheWAS platform[41], illustrated unique patterns of phenotype associations for each pathway GRS. These patterns of PheWAS findings are relevant to the potential efficacy and potential side-effects of intervening in these pathways. As a proof-of-concept, the GRS associated with lower FEV_1/FVC specific to PI3K-Akt signalling was associated with increased risk of COPD but a lower risk of diabetes; PI3K inhibition impairs glucose uptake in muscle and increases hepatic gluconeogenesis, contributing to glucose intolerance and diabetes[42]. The PheWAS and druggability analyses we conducted have potential to identify drug repurposing opportunities for COPD.

The patterns of pleiotropy we show through PheWAS for individual variants, for trait-specific GRS and pathway-partitioned GRS may help to explain variants and pathways that increase susceptibility to more than one disease, and thereby predispose to particular patterns of multimorbidity. For example, the elastic fibre pathway GRS was associated with increased risk of muscular (e.g. herniae) and musculoskeletal conditions related to connective tissue laxity. Our findings also help to further elucidate the complex relationship between height, body mass index or obesity, and lung function, and their genetic determinants[5,43]. We saw no overall correlation between the magnitude of lung function and height associations, and relationships differed between GRS for different lung function traits, and even between sitting and standing height for the same trait. The pathway-partitioned GRS indicate that the relationship between genetic variants, height and lung function traits depends on the pathways through which the variants act.

Our discovery effort was enabled by the largest worldwide collaboration to bring together multi-ancestry populations with curated lung function and genomic data, and to map these signals to putative causal genes. The last comprehensive attempt to map lung function associated variants to genes identified 107 putative causal genes, mostly through eQTLs only, and only eight genes were then implicated by ≥2 criteria[5]. In contrast, we implicated 559 causal genes meeting at least two criteria, through drawing upon new data and methodologies, such as single cell epigenome data, rare variant associations identified in sequencing data in UK Biobank and similarity-based approach PoPs[9]. Nevertheless, our study has limitations. Our focus was on discovery of multi-ancestry rather than ancestry-specific signals, as sample sizes for lung function genomics studies in all non-European ancestry groups fall far short of those in European ancestries, particularly in African ancestry populations[4]. Indeed, non-European ancestries are under-represented generally in genomic studies[3], constraining both genome-wide and especially phenome-wide approaches in these populations. Correcting this will require substantial global investment in studies with suitably

phenotyped and genotyped individuals, coupled with appropriate models of community participation and workforce development. Improved sample sizes across all ancestries would improve power for discovery in for ancestry-specific studies[43], and for fine-mapping genetic associations detected from multi-ancestry meta-analyses.

Strategies for in-silico mapping of association signals to causal genes are constantly evolving and difficult to evaluate until a reference set of fully functionally characterised lung function-associated variants and causal genes is developed. The framework we used to map signals to genes parallels one recently adopted[10], showing the consensus between approaches in implicating each putative causal gene. Such frameworks, whilst guiding prioritisation of genes for functional experiments, do not provide definitive guidance on how variant-to-gene criteria should be weighted or on a minimal number of variant-to-gene criteria required. Such in-silico evidence cannot firmly demonstrate causality, and confirmation of mechanism will require functional genomics experiments such as gene editing in suitable organoids with appropriate readouts. An additional limitation is that classifications of pathways may be imperfect; we used multiple pathway classifications as it is unclear which is superior across all component pathways, and we present the pathway-partitioned PheWAS results as a resource to others.

In summary, our multi-ancestry study highlights new putative causal variants, genes and pathways, some of which are targeted by existing drug compounds. These findings bring us closer to understanding mechanisms underlying lung function and COPD and will inform functional genomics experiments to confirm mechanisms and consequently guide the development of therapies for impaired lung function and COPD.

References

1. Young, R.P., Hopkins, R. & Eaton, T.E. Forced expiratory volume in one second: not just a lung function test but a marker of premature death from all causes. *Eur Respir J* **30**, 616-22 (2007).
2. GBD Chronic Respiratory Disease Collaborators. Prevalence and attributable health burden of chronic respiratory diseases, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet Respir Med* **8**, 585-596 (2020).
3. Sirugo, G., Williams, S.M. & Tishkoff, S.A. The Missing Diversity in Human Genetic Studies. *Cell* **177**, 26-31 (2019).
4. Tobin, M.D. & Izquierdo, A.G. Improving ethnic diversity in respiratory genomics research. *Eur Respir J* **58**(2021).
5. Shrine, N. *et al.* New genetic signals for lung function highlight pathways and chronic obstructive pulmonary disease associations across multiple ancestries. *Nature Genetics* **51**, 481-493 (2019).
6. Wain, L.V. *et al.* Genome-wide association analyses for lung function and chronic obstructive pulmonary disease identify new loci and potential druggable targets. *Nat Genet* **49**, 416-425 (2017).
7. Mägi, R. *et al.* Trans-ethnic meta-regression of genome-wide association studies accounting for ancestry increases power for discovery and improves fine-mapping resolution. *Human Molecular Genetics* **26**, 3639-3650 (2017).
8. Barbeira, A.N. *et al.* Exploiting the GTEx resources to decipher the mechanisms at GWAS loci. *Genome biology* **22**, 1-24 (2021).
9. Weeks, E.M. *et al.* Leveraging polygenic enrichments of gene features to predict genes underlying complex traits and diseases. *medRxiv*, 2020.09.08.20190561 (2020).
10. Aragam, K.G. *et al.* Discovery and systematic characterization of risk variants and genes for coronary artery disease in over a million participants. *medRxiv*, 2021.05.24.21257377 (2021).
11. Wang, L. *et al.* Methodology in phenome-wide association studies: a systematic review. *J Med Genet* **58**, 720-728 (2021).
12. Finucane, H.K. *et al.* Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nature Genetics* **47**, 1228-1235 (2015).
13. Pulit, S.L., de With, S.A. & de Bakker, P.I. Resetting the bar: Statistical significance in whole-genome sequencing-based association studies of global populations. *Genet Epidemiol* **41**, 145-151 (2017).
14. Pickrell, J.K. Joint analysis of functional genomic data and genome-wide association studies of 18 human traits. *Am J Hum Genet* **94**, 559-73 (2014).
15. Repapi, E. *et al.* Genome-wide association study identifies five loci associated with lung function. *Nature genetics* **42**, 36-44 (2010).
16. Bulik-Sullivan, B.K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nature genetics* **47**, 291-295 (2015).
17. Iotchkova, V. *et al.* GARFIELD classifies disease-relevant genomic features through integration of functional annotations with association signals. *Nature Genetics* **51**, 343-353 (2019).
18. Cho, M.H. *et al.* Risk loci for chronic obstructive pulmonary disease: a genome-wide association study and meta-analysis. *Lancet Respir Med* **2**, 214-25 (2014).
19. Soler Artigas, M. *et al.* Sixteen new lung function signals identified through 1000 Genomes Project reference panel imputation. *Nat Commun* **6**, 8658 (2015).
20. Wyss, A.B. *et al.* Multiethnic meta-analysis identifies ancestry-specific and cross-ancestry loci for pulmonary function. *Nature Communications* **9**, 2976 (2018).
21. Bullard, J.E., Wert, S.E., Whitsett, J.A., Dean, M. & Nogee, L.M. ABCA3 mutations associated with pediatric interstitial lung disease. *American journal of respiratory and critical care medicine* **172**, 1026-1031 (2005).
22. Yengo, L. *et al.* Meta-analysis of genome-wide association studies for height and body mass index in∼ 700000 individuals of European ancestry. *Human molecular genetics* **27**, 3641-3649 (2018).
23. Gudmundsson, J. *et al.* Genome-wide associations for benign prostatic hyperplasia reveal a genetic correlation with serum levels of PSA. *Nature communications* **9**, 1-8 (2018).
24. Morrisey, E.E., Ip, H.S., Tang, Z., Lu, M.M. & Parmacek, M.S. GATA-5: a transcriptional activator expressed in a novel temporally and spatially-restricted pattern during embryonic development. *Developmental biology* **183**, 21-36 (1997).
25. LaFemina, M.J. *et al.* Claudin-18 deficiency results in alveolar barrier dysfunction and impaired alveologenesis in mice. *American journal of respiratory cell and molecular biology* **51**, 550-558 (2014).
26. Sweerus, K. *et al.* Claudin-18 deficiency is associated with airway epithelial barrier dysfunction and asthma. *Journal of Allergy and Clinical Immunology* **139**, 72-81. e1 (2017).

27. Türeci, Ö., Mitnacht-Kraus, R., Wöll, S., Yamada, T. & Sahin, U. Characterization of zolbetuximab in pancreatic cancer models. *Oncoimmunology* **8**, e1523096 (2019).
28. Sweerus, K. *et al.* Claudin-18 deficiency is associated with airway epithelial barrier dysfunction and asthma. *J Allergy Clin Immunol* **139**, 72-81.e1 (2017).
29. Krone, K.A. *et al.* Pulmonary manifestations of immune dysregulation in CTLA-4 haploinsufficiency and LRBA deficiency. *Pediatric Pulmonology* **56**, 2232-2241 (2021).
30. Shamriz, O. *et al.* Respiratory manifestations in LPS-responsive beige-like anchor (LRBA) protein-deficient patients. *European journal of pediatrics* **177**, 1163-1172 (2018).
31. Ferkingstad, E. *et al.* Genome-wide association meta-analysis yields 20 loci associated with gallstone disease. in *Nature communications* Vol. 9 5101 (2018).
32. Yuan, T. *et al.* FGF10-FGFR2B signaling generates basal cells and drives alveolar epithelial regeneration by bronchial epithelial stem cells after lung injury. *Stem Cell Reports* **12**, 1041-1055 (2019).
33. Akkuş, G. *et al.* Hypogonadotropic Hypogonadism due to Novel FGFR1 Mutations. *Journal of clinical research in pediatric endocrinology* **9**, 95-100 (2017).
34. Klein, R.J. *et al.* Complement factor H polymorphism in age-related macular degeneration. *Science* **308**, 385-389 (2005).
35. Miller, T.M. *et al.* Correlating phenotype and genotype in the periodic paralyses. *Neurology* **63**, 1647-55 (2004).
36. Monnier, N., Procaccio, V., Stieglitz, P. & Lunardi, J. Malignant-hyperthermia susceptibility is associated with a mutation of the alpha 1-subunit of the human dihydropyridine-sensitive L-type voltage-dependent calcium-channel receptor in skeletal muscle. *Am J Hum Genet* **60**, 1316-25 (1997).
37. Chiu, K.Y., Li, J.G. & Lin, Y. Calcium channel blockers for lung function improvement in asthma: A systematic review and meta-analysis. *Annals of Allergy, Asthma & Immunology* **119**, 518-523. e3 (2017).
38. Breuer, J. *et al.* VLA-2 blockade in vivo by vatelizumab induces CD4+FoxP3+ regulatory T cells. *Int Immunol* **31**, 407-412 (2019).
39. Herwig, R., Hardt, C., Lienhard, M. & Kamburov, A. Analyzing and interpreting genome data at the network level with ConsensusPathDB. *Nature Protocols* **11**, 1889-1907 (2016).
40. Krämer, A., Green, J., Pollard Jr, J. & Tugendreich, S. Causal analysis approaches in ingenuity pathway analysis. *Bioinformatics* **30**, 523-530 (2014).
41. Packer, R.J. *et al.* Deep-PheWAS: a pipeline for phenotype generation and association analysis for phenome-wide association studies. *medRxiv*, 2022.05.05.22274419 (2022).
42. Sahakian, N. *et al.* SGLT2 inhibitors as potentially helpful drugs in PI3K inhibitor-induced diabetes: a case report. *Clinical Diabetes and Endocrinology* **7**, 1-4 (2021).
43. Zhu, Z. *et al.* A large-scale genome-wide association analysis of lung function in the Chinese population identifies novel loci and highlights shared genetic aetiology with obesity. *European Respiratory Journal* **58**, 2100199 (2021).

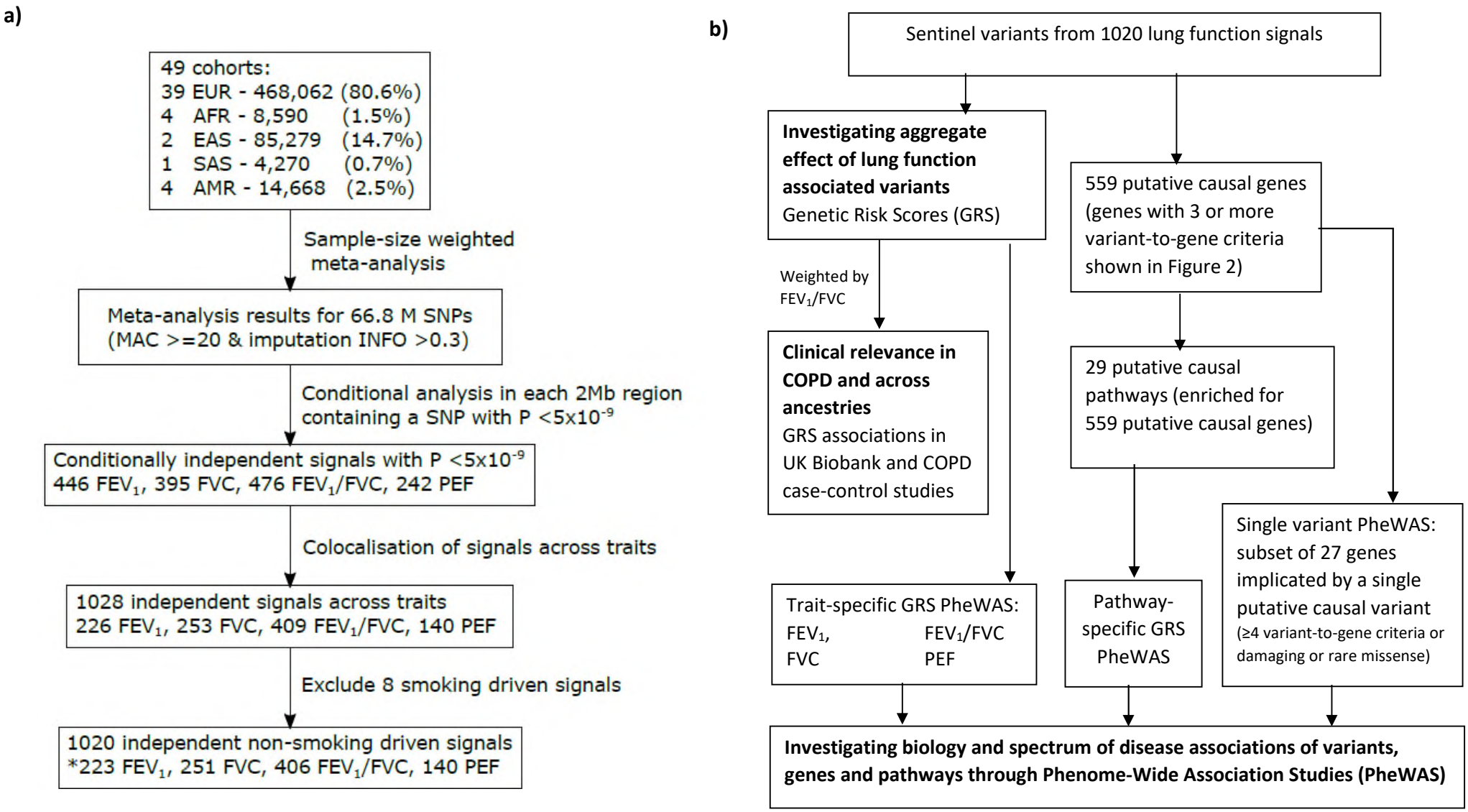


Figure 1: Overview of (a) Discovery meta-analysis; Ancestry abbreviations: EUR – European, AFR – African, EAS – East Asian, SAS – South Asian, AMR – Admixed American/Hispanic. *For signals present in more than one trait, the signal is only counted once (for the most significant trait); **(b) pathway analyses, genetic risk score (GRS) analyses and phenome-wide association studies (PheWAS)**

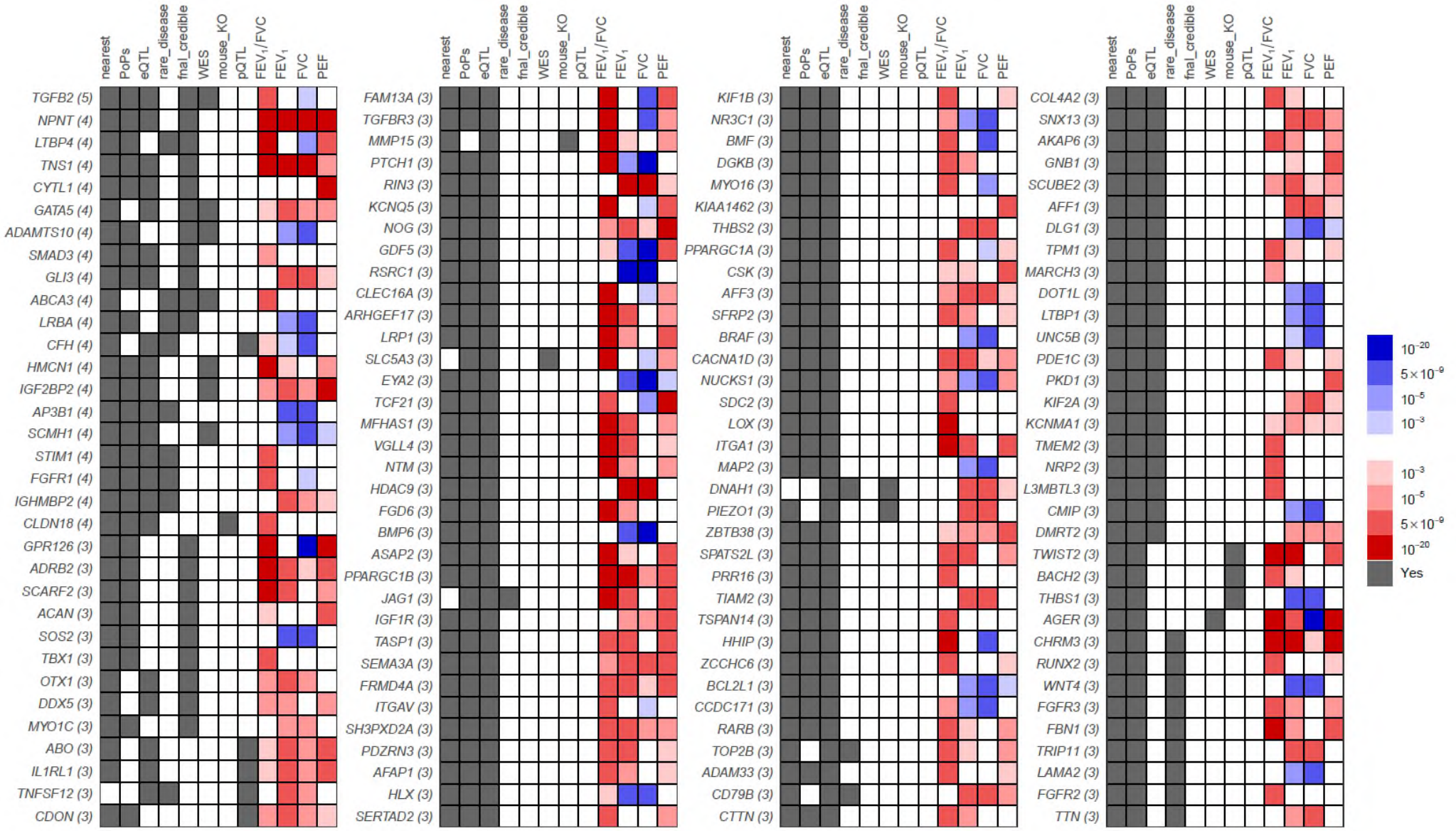
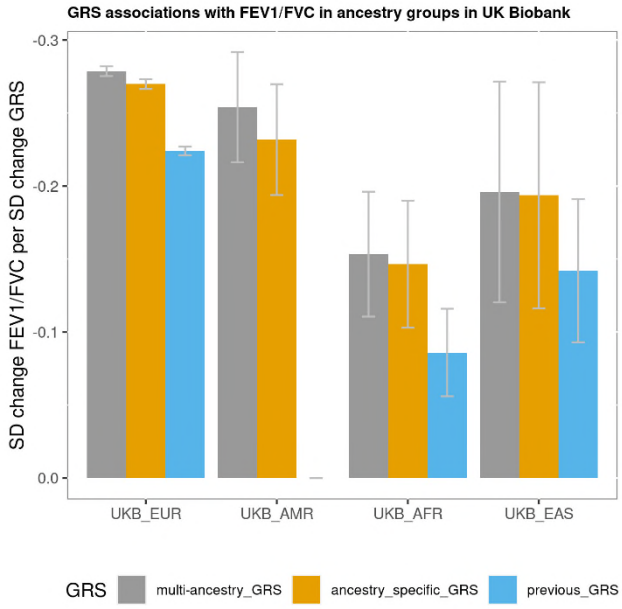


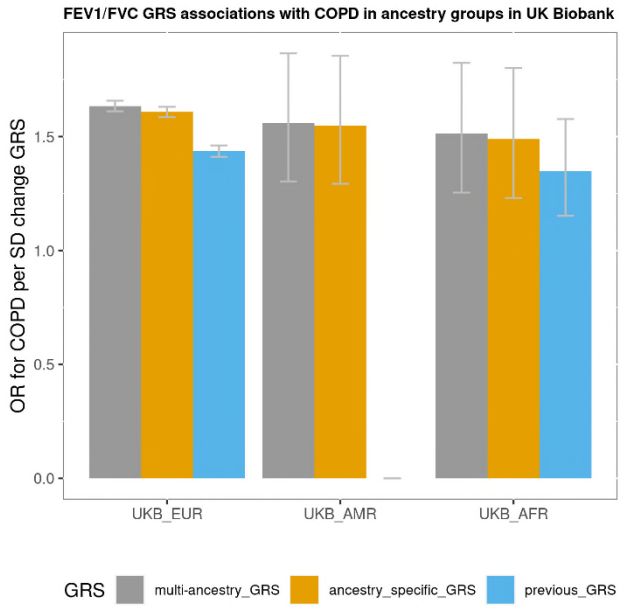
Figure 2: 135 genes prioritised with 3 or more variant-to-gene criteria.

The number of variant-to-gene criteria implicating the gene is in brackets after the gene name. The grey in the first 8 columns indicates that at least 1 variant implicates the gene as causal via the evidence for that column. The last 4 columns indicate the level of association of the most significant variant implicating the gene as causal with respect to the FEV_1/FVC decreasing allele: the same direction of effect as the FEV_1/FVC decreasing allele has red shades, the opposite direction has blue shades.

a)

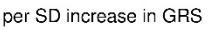


b)



c)

Ancestry	Cohort	P	Cases	Controls	OR [95%CI]
AFR	COPDGene(AFR)	8.69e-05	821	1749	1.20 [1.10, 1.32]
	SPIROMICS(AFR)	0.0037	174	142	1.47 [1.13, 1.90]
	Meta-analysis	**2.96e-06**	**995**	**1891**	1.23 [1.13, 1.34]
EUR	COPDGene(EUR)	1.43e-48	2811	2534	1.59 [1.49, 1.69]
	ECLIPSE	8.75e-09	1764	178	1.66 [1.40, 1.97]
	GenKOLS	8.44e-17	864	808	1.67 [1.48, 1.88]
	NETTNAS	5.47e-07	376	435	1.63 [1.35, 1.98]
	SPIROMICS(EUR)	2.87e-17	931	373	1.82 [1.58, 2.09]
	Meta-analysis	**7.07e-93**	**6746**	**4328**	1.63 [1.56, 1.71]
					1.53 [1.47, 1.60]



d)

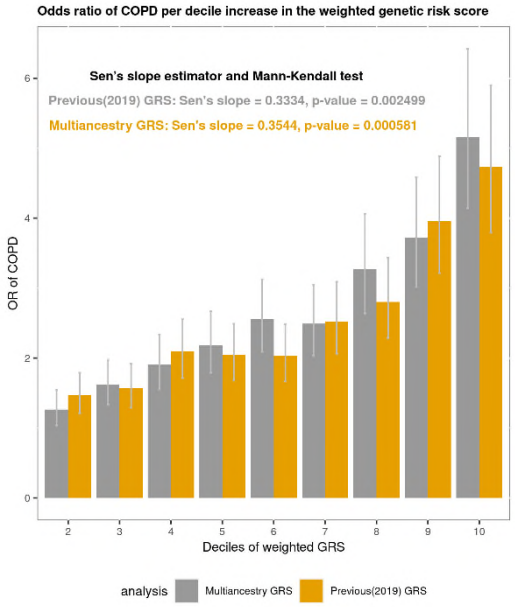
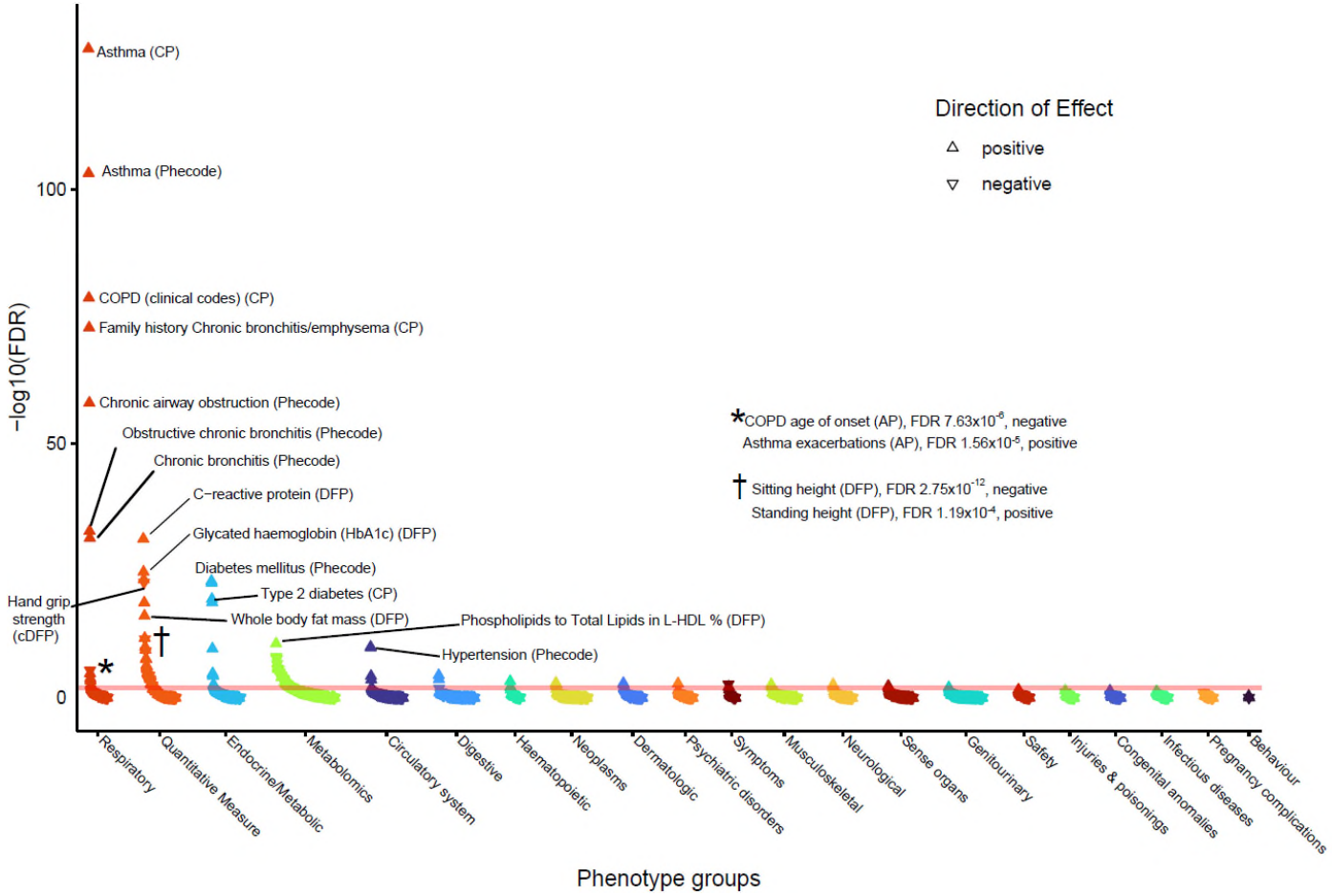
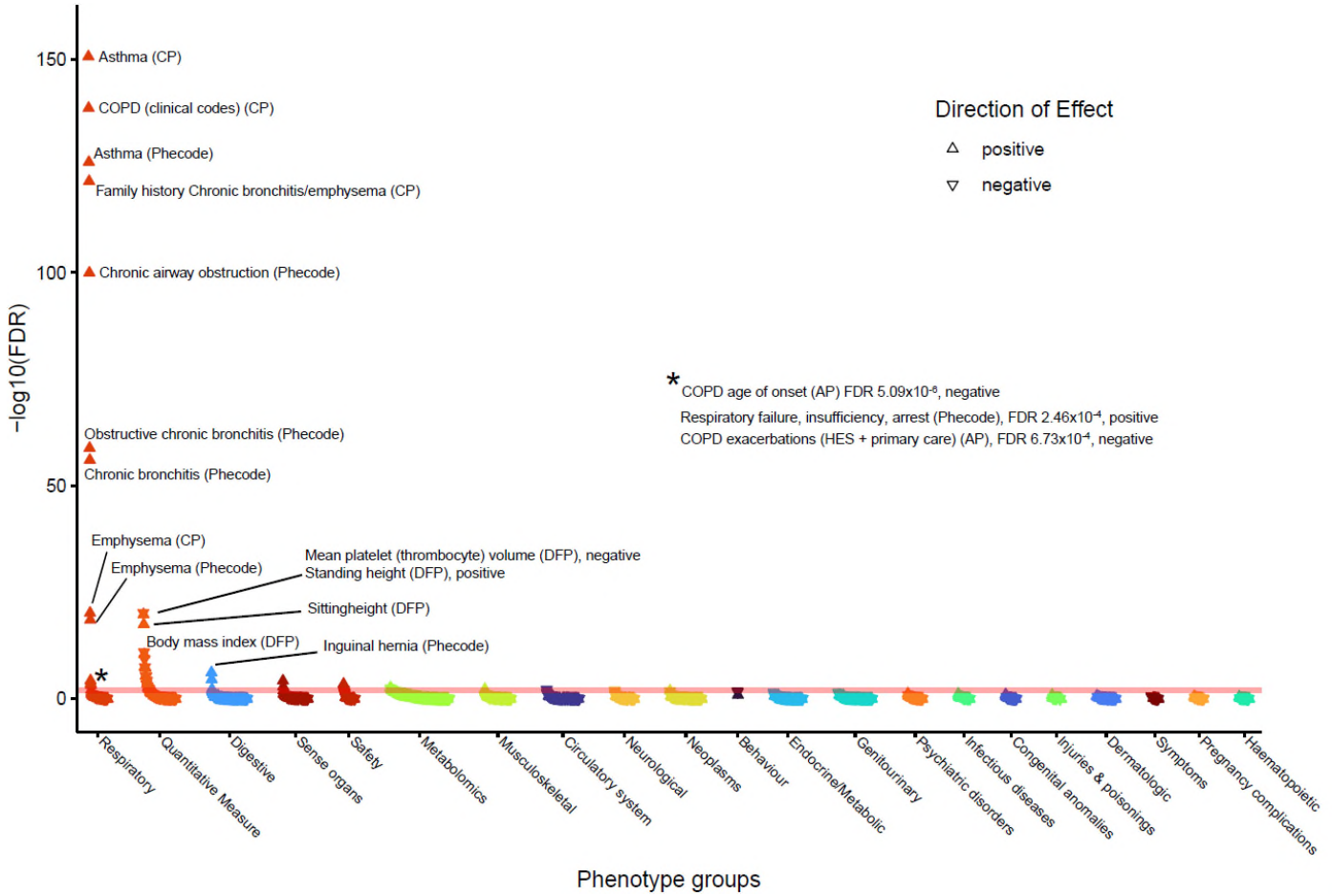


Figure 3: GRS performance: a) Prediction performance of 3 GRS across ancestry groups for FEV_1/FVC shown as standard deviation (SD) change in FEV_1/FVC per SD increase in GRS in the individuals of UK Biobank ancestry groups (whiskers represent 95% confidence intervals); **b)** Prediction performance of 3 GRS for COPD shown as COPD odds ratio per SD increase in GRS; **c)** OR for COPD per S.D. change in GRS in COPD case-control studies; **d)** decile analysis.

a) FEV_1



b) FEV_1/FVC



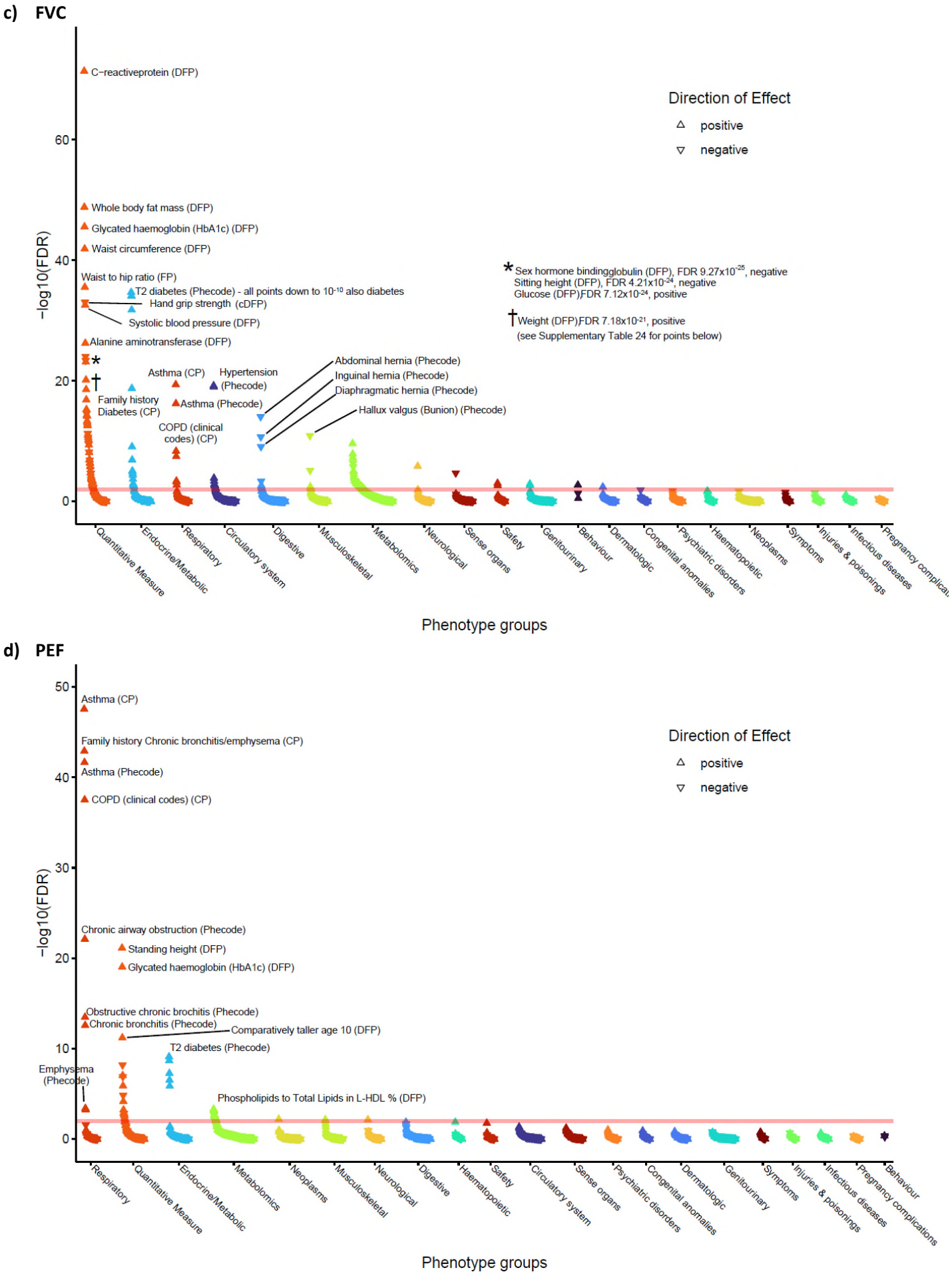
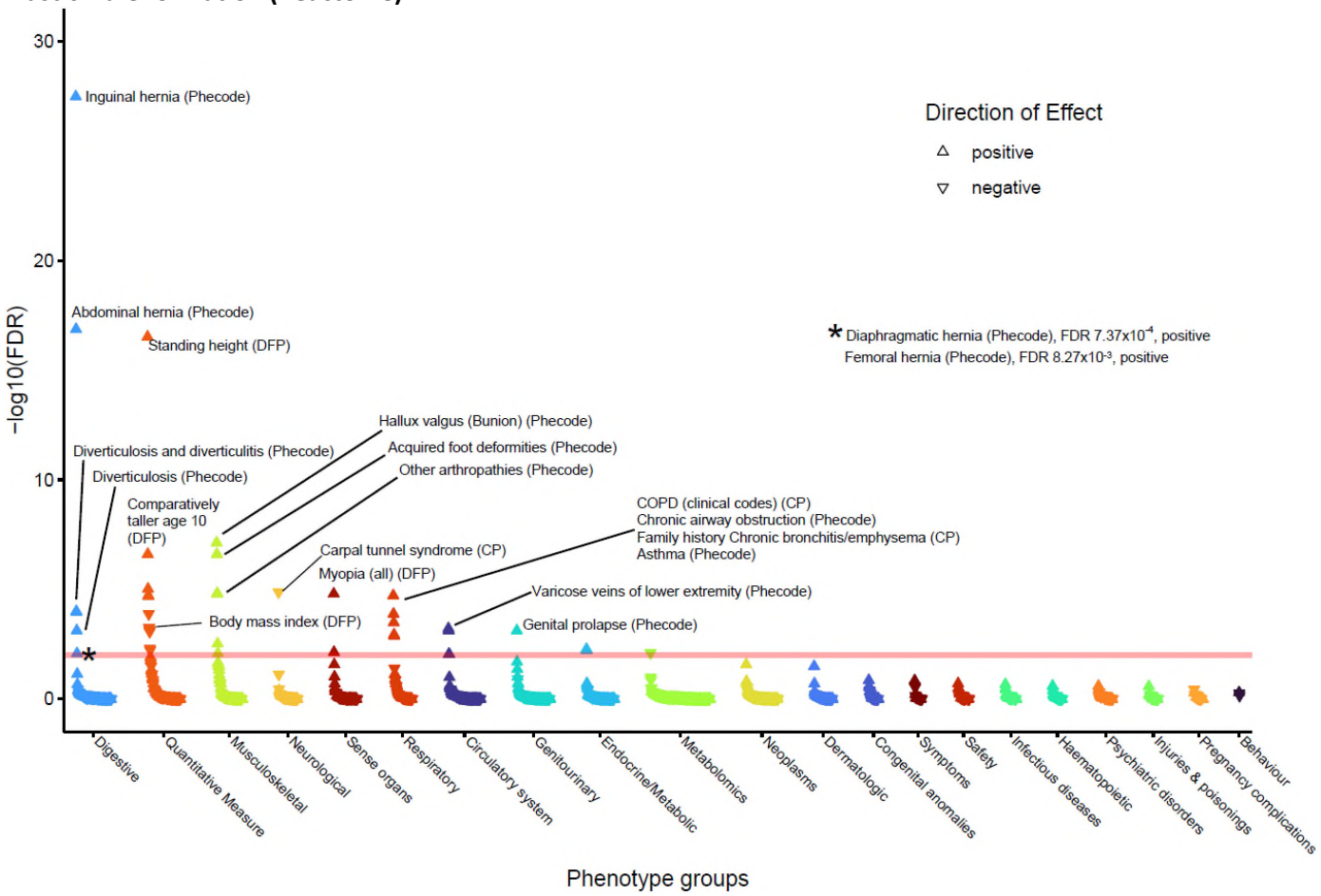
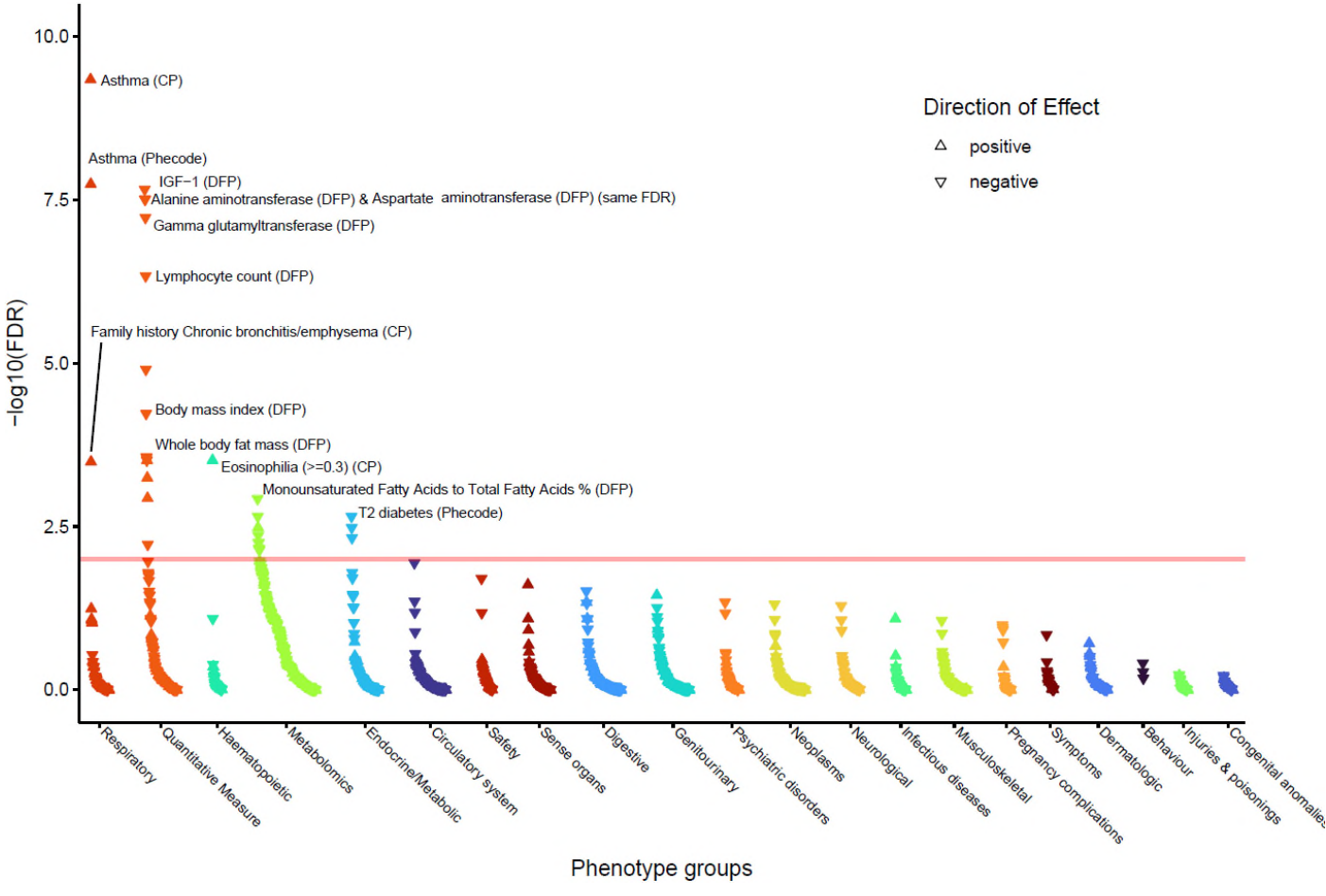


Figure 4: PheWAS of lung function trait GRS: **a)** FEV_1; **b)** FEV_1/FVC; **c)** FVC; **d)** PEF
CP=composite phenotype, DFP=Data-Field ID phenotype, cDFP=combined Data-Field ID phenotype, FP=formula phenotype, AP=Added phenotype (Methods).

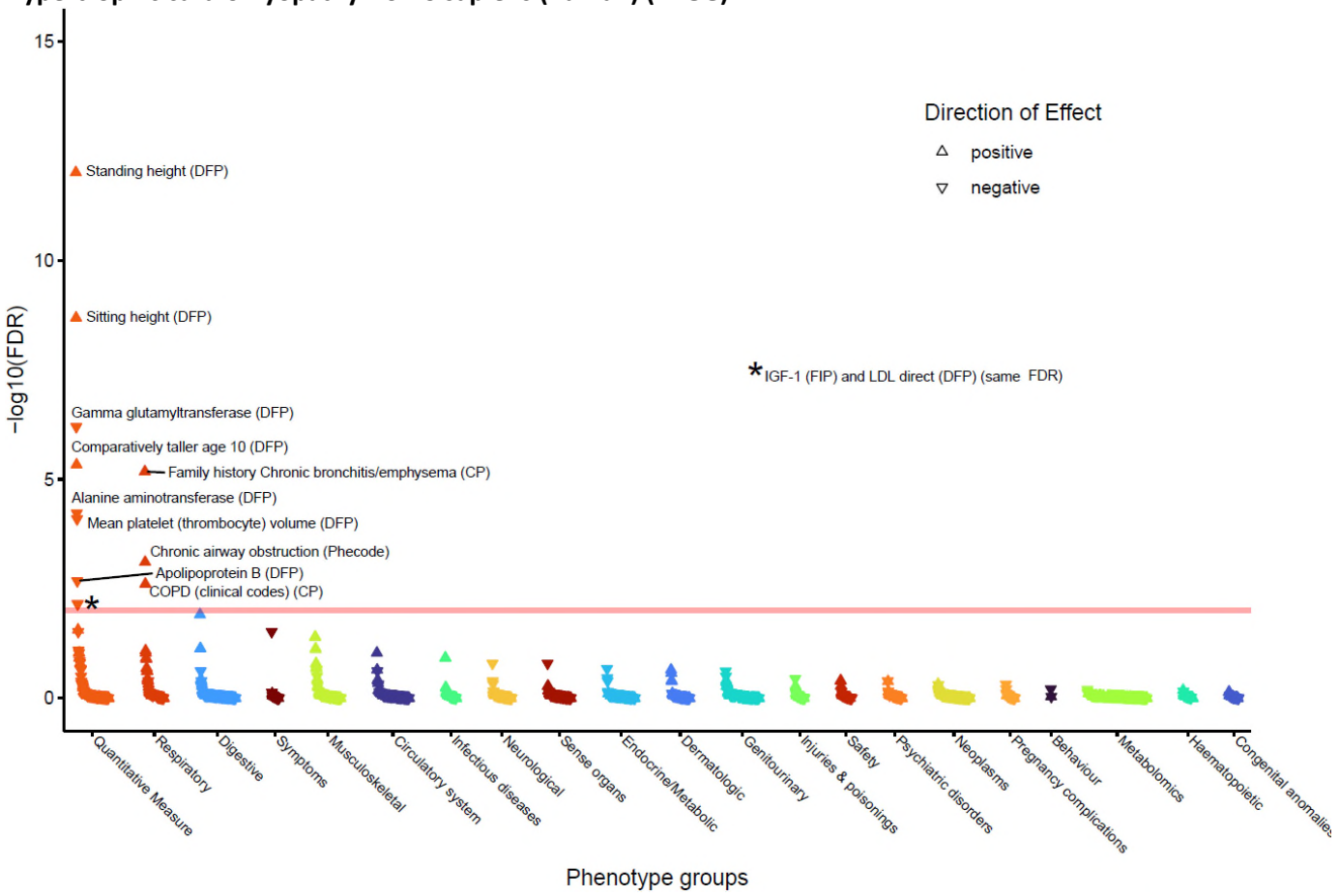
a) Elastic fibre formation (Reactome)



b) PI3K-Akt signalling pathway Homo sapiens (human) (KEGG)



c) Hypertrophic cardiomyopathy Homo sapiens (human) (KEGG)



d) Signal Transduction (Reactome)

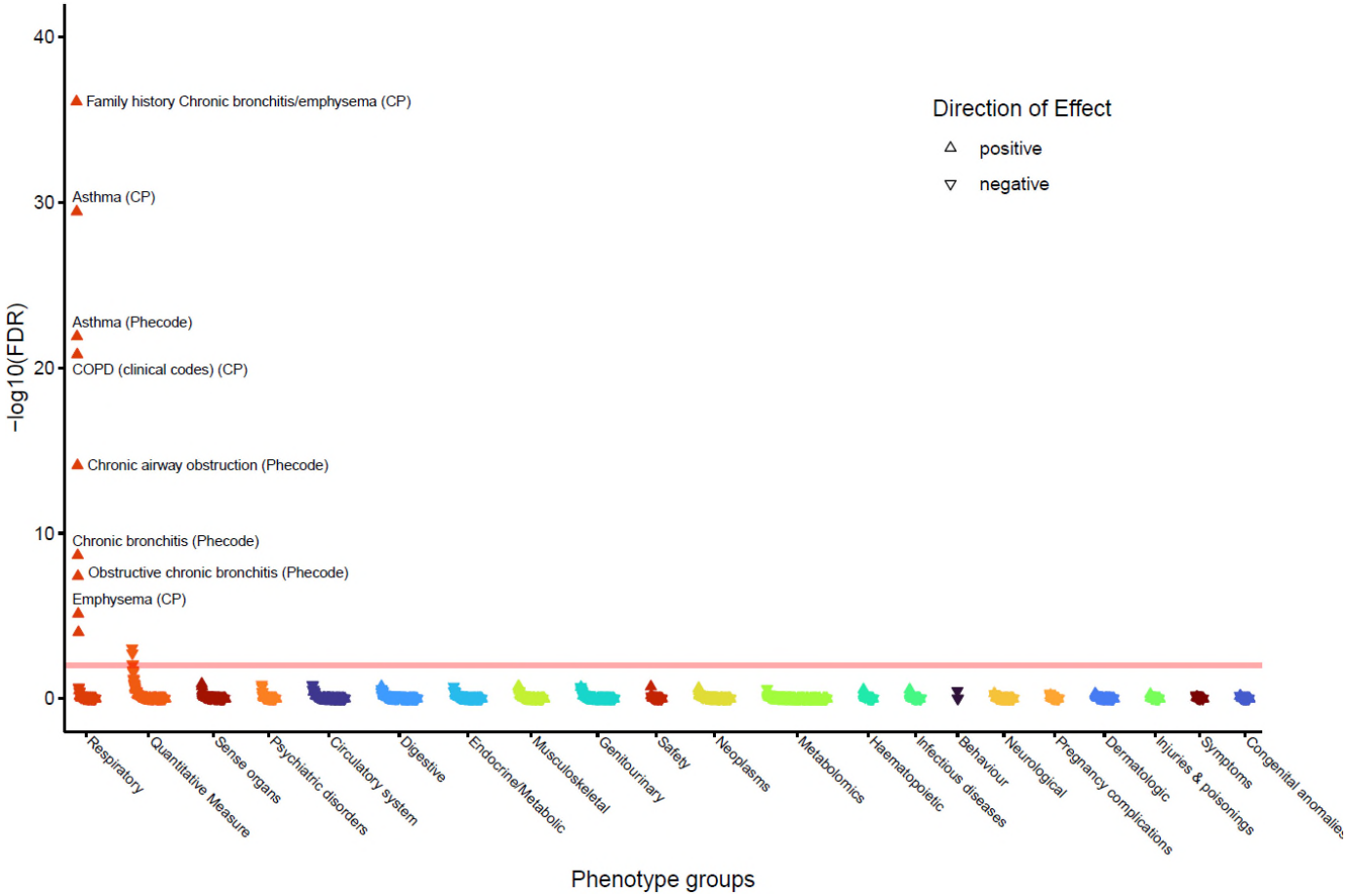


Figure 5: PheWAS for FEV_1/FVC weighted GRS partitioned by: **a)** Elastic fibre formation (Reactome); **b)** PI3K-Akt signalling pathway Homo sapiens (human) (KEGG); **c)** Hypertrophic cardiomyopathy Homo sapiens (human) (KEGG); **d)** Signal Transduction (Reactome). CP=composite phenotype, DFP=Data-Field ID phenotype, cDFP=combined Data-Field ID phenotype, FP=formula phenotype, AP=Added phenotype (Methods).