

Epidemiology and Genetics of Preserved Ratio Impaired Spirometry (PRISm): An Analysis of UK Biobank

Daniel H Higbee MBBS^{1,2}, Raquel Granell PhD¹, George Davey Smith FRS¹, James W Dodd MB ChB, PhD^{1,2}

1 MRC Integrative Epidemiology Unit (IEU), University of Bristol, Oakfield Grove, Bristol, BS8 2BN, UK.

2 Academic Respiratory Unit, University of Bristol, Southmead Hospital, Bristol, BS10 5NB UK

Corresponding Author: Dr James W. Dodd

james.dodd@bristol.ac.uk

Tel +44117 4142012

Fax +44117 4149496

ORCID [0000-00034805-5759](https://orcid.org/0000-00034805-5759)

JD & DH conceived the study and initial draft, all authors made substantial contributions to the analysis, drafting and final approval of the paper.

Funding: This work was supported by the Medical Research Council and the University of Bristol Integrative Epidemiology Unit (MC_UU_00011). MRC CARP Fellowship (Grant ref: MR/T005114/1)

Abstract

Rationale

Preserved ratio impaired spirometry (PRISm) is defined as Forced Expiratory Volume in one second <80% predicted with a FEV₁/Forced Vital Capacity of >0.70. It may be a precursor of COPD and has been associated with increased respiratory symptoms and comorbidities. Current understanding of PRISm is based on small selective cohorts and genetic determinants remain unknown.

Objectives

Identify the prevalence, risk factors, co-morbidity, and genetics of PRISm in a large cohort to improve power and generalizability.

Methods

The UKBiobank, provides a large cohort of genotyped individuals. Regression analysis determined risk factors for and associated co-morbidities of PRISm. Genome wide association studies, conditional, and gene-based analyses were performed with a PheWAS and literature search conducted on the results.

Measurements and Main Results

PRISm has a similar prevalence to airflow obstruction (8% vs 7.8%) and is associated with increased cardiovascular co-morbidity, OR of 1.26 (95%CI 1.20-1.31) after adjusting for BMI, smoking status, age, diabetes, and sex. Genetic analysis discovered four associated single nucleotide polymorphisms and eight genes. Two SNPs have not been described with respiratory disease before. Other genetic loci identified are known to increase risk of asthma and COPD, and influence response to inhaled steroids.

Conclusions

PRISm affects 8% of adults of whom ~40% report cardiovascular comorbidity. Detailed analysis reveals that PRISm has shared genetic associations with asthma and COPD. These findings highlight the need for research into the potential therapeutic strategies and screening for individuals with PRISm, in particular whether this may reduce transition to COPD and reduce cardiovascular disease.

Introduction

Preserved Ratio Impaired Spirometry (PRISm), also referred to as ‘restrictive pattern’ or ‘unclassified’ spirometry, is defined as Forced Expiratory Volume in one second (FEV₁) percent predicted <80%, despite a normal FEV₁/Forced Vital Capacity (FVC) ratio >0.70. The true prevalence of PRISm is unknown with estimates ranging from just 4% up to 48% amongst adult varying by gender, ancestry and smoking history.^(1, 2) Interest in the potential clinical implications of PRISm come from longitudinal data which suggest that over 5 years, up to 50% of people with PRISm transition to COPD but also that 15% may return to ‘normal’ spirometry.⁽³⁾ PRISm is understudied, but available data suggests that it is associated with increased respiratory symptoms, important co-morbidities such as obesity, diabetes, cardiac disease and an increased overall mortality.^(3, 4) However, the epidemiology and genetics of PRISm remains unclear. Research is needed to clarify and better understand the nature of this abnormality in lung function.

Detailed epidemiological and genetic understanding of PRISm has to date been limited by analysis restricted to insufficient population numbers in cohorts rarely containing more than 1000 cases of PRISm. Generalisability of data has also been an issue with some cohorts restricted to smokers.^(2, 4) This may introduce selection bias and confounding, limiting conclusions regarding PRISm as a potential pre-COPD state and as an independent risk for cardiovascular disease.⁽⁵⁾

There has been just one attempted Genome Wide Association Study (GWAS) in PRISm but numbers again were small (N=1257), restricted to ever smokers, and failed to find any significant single nucleotide polymorphisms (SNPs, p value <5x10⁻⁸).⁽²⁾ Genetic markers are critical to our understanding of PRISm since they provide insight into disease susceptibility and pathogenesis. Lung function trajectories determined by genetic and early life factors are now

known to influence risk of COPD,^(6, 7) and the same may be true of PRISm. Genetic variants could therefore provide therapeutic targets for PRISm and associated co-morbidities.

Our first objective was to utilise the UKBiobank to examine a cohort of individuals with PRISm, obstructive and normal spirometry.⁽⁸⁾ The large sample size of UKBiobank and broad recruitment based on age, not smoking history, increases power and reduces selection bias allowing improved generalizability to the wider population. Differences in demographics, symptom burden and associated co-morbidity in PRISm were compared to other spirometry states.

Our second objective was to use genetic epidemiology to determine associations of PRISm. A GWAS, conditional analysis, and gene-based analysis were performed to discover genetic associations of PRISm and compare to known genetic causes of other respiratory diseases.

Method

Objective One - Observational Epidemiology

UKBiobank includes 502,543 individuals aged between 40 and 69 at recruitment across the UK.⁽⁸⁾ Participants completed detailed health questionnaires and blood samples were taken for genotyping. 151,973 participants had pre-bronchodilator lung function testing that included FEV₁ percent predicted. Obstructive spirometry was defined as FEV₁ <80% predicted and FEV₁/FVC <0.70 (Stage II-IV obstruction),⁽⁹⁾ and controls as FEV₁ ≥80% with FEV₁/FVC >0.70. Statistical analysis was performed using Stata 15.⁽¹⁰⁾ We used logistic regression analysis to examine for associations between risk factors for PRISm, and clinically relevant correlates of PRISm.

Objective Two – GWAS

A GWAS of PRISm and controls was performed using BOLT-LMM via the IEU GWAS pipeline adjusting for sex, body mass index (BMI), year of birth and smoking status.⁽¹¹⁾ This uses a linear mixed model to account for both cryptic relatedness and population stratification. Only participants of European ancestry were included. SNPs used all met the criteria: Mean Allele Frequency >0.01, Genotyping rate >0.015, Hardy-Weinberg equilibrium p-value<0.0001, r^2 threshold of 0.10.^(11, 12) Linkage Disequilibrium(LD)-clumping, which retains the most significant SNP at a locus to ensure all signals are independent, was performed with criteria kb = 1000 and $r^2 = 0.001$. LD-regression analysis was performed to assess correlation of PRISm with a GWAS examining continuous lung function measures.⁽¹³⁻¹⁵⁾

Objective Two – Conditional analysis

GWAS methodology assumes the top SNP found explains the maximum variance in the region, and that close SNPs showing association only do so because they are in high LD. This assumption may be incorrect leading to SNPs not being detected. Conditional analysis mutually adjusts on the top SNP to test whether the other SNPs at the locus are significantly associated.⁽¹⁶⁾ We used Genome-wide Complex Trait Analysis software for a Conditional and Joint analysis (GCTA-COJO) on our GWAS results.⁽¹⁷⁾ A reference sample of individually genotyped participants of the Avon Longitudinal Study of Parent and Children (ALSPAC) was used to estimate LD patterns.⁽¹⁸⁾

Objective Two – Gene-based analysis

Gene-based analysis was performed using Versatile Gene-based Association Study software (VEGAS2).⁽¹⁹⁾ Gene based analysis reduces the risk of type 1 error as ~20,000 genes are tested rather than millions of SNPs. It has greater power to detect genes that may contain multiple SNPs of marginal significance, which may be missed by SNP based GWAS. 21212 genes were tested so a Bonferroni corrected p-value threshold of $<2.36 \times 10^{-6}$ ($=0.05/21212$) was applied

to correct for multiple testing. The gene-based analysis determines the SNP with the highest p value in each gene, and reports it as Top SNP.

Genes found were cross referenced with online resources to determine their chromosomal location and function.⁽²⁰⁻²²⁾ We performed a literature search for reported lung disease associations of significant genes found.^(23, 24) A Phenome Wide Association Study (PheWAS) was performed on SNPs from all three objectives to determine if they have associations with other lung function measures or disease.^(21, 25) Full PheWAS results are in the online additional material.

Objective Two methods were repeated after excluding participants who have asthma/COPD. This is reported in the online data supplement.

Results

Objective One – Observational Epidemiology

12,238 of participants in UKBiobank meet the criteria for PRISm, 11904 participants had stage II-IV obstruction and there are 115,890 controls. Pearson's chi-squared test shows that the prevalence of PRISm (8.0%) is marginally higher than obstruction stage II-IV pattern spirometry (7.8%, p-value 0.025). PRISm is strongly associated with current smoking (p value <0.001), the absolute percentage of current smokers in PRISm is double that of controls, and doubles again between PRISm and obstruction (**Table 1**). Current smoking is higher in PRISm with an Odds Ratio (OR) of 2.37 (95%CI 2.23-2.51) vs. control after adjustment for age, sex and BMI. Mean BMI values are 2.2kg/m² higher amongst the PRISm group than controls. Regression analysis shows a small OR of 1.08 (95%CI 1.07-1.08) per SD increase in BMI for PRISm compared to controls after adjustment for age, sex and smoking.

Self-reported rates of asthma were higher in PRISm than controls (18% vs 10%, p value <0.001). Rates of self-reported COPD equivalent between PRISm and obstruction (4.97% vs 4.51%, p value 0.1).

PRISm participants report a higher burden of respiratory symptoms than controls, with 7.1% reporting shortness of breath walking on ground level compared to 2.7% controls, OR 2.03 (95%CI 1.88-2.26) after adjustment by age, sex and BMI. This was measured by participants being asked, "Do you get short of breath walking with people of your own age on level ground?". Co-morbidities are more frequent in PRISm with >10% diagnosed with diabetes, OR 1.67 (95%CI 1.56-1.79) in PRISm vs. controls after adjustment for age, sex and BMI. Cardiovascular co-morbidity in UKBiobank is defined as patients self-reporting any diagnosis of hypertension, angina, stroke, or myocardial infarction. 39.4% of PRISm reported cardiovascular disease vs. 26.7% of controls, OR of 1.26 (95%CI 1.20-1.31) after adjusting for BMI, smoking status, age, diabetes, and sex. There does not appear to be a difference between the prevalence of cardiovascular co-morbidity between PRISm vs. participants with obstructive pattern spirometry despite the difference in smoking rates, sex, and age (cardiovascular co-morbidity 39.4% PRISm, 38.9% Obstruction, p value 0.59).

Table 1. Characteristics of PRISm, Obstructive and Control participants in UK Biobank

	PRISm N=12,238 (8.0%)	Control N=115,807 (76.2%)	Pval* PRISm vs. Control	Stage II-IV Obstruction N=11,904 (7.8%)	Pval* PRISm vs. Obstruction
Age (Years) Mean (SD)	63.7 (7)	62.5 (7)	<0.001	66.5(7)	<0.001
BMI (kg/m ²) Mean (SD)	29.6 (5)	27.4 (5)	<0.001	27.5 (5)	<0.001
Female (%)	58.6	60.4	<0.001	47.4	<0.001
FEV ₁ % predicted Mean (SD)	72.1 (7)	101.4 (12)	<0.001	64.9 (12)	<0.001
Never smoker (%)	59.7	74.4	<0.001	37.5	<0.001
Ex-smoker (%)	25.8	17.4	<0.001	33.7	<0.001
Current smoker (%)	14.3	7.7	<0.001	28.6	<0.001
Pack/years [†] Mean (SD)	37.4 (19)	31.3 (15)	<0.001	42.0 (20)	<0.001
SOB walking on ground (%)	7.1	2.7	<0.001	9.1	<0.001
Self-reported asthma	18.4%	10.4%	<0.001	35.3%	<0.001
Self-reported COPD	4.97%	0.08%	<0.001	4.51%	0.105
Diabetes (%)	10.7	4.1	<0.001	6.9	<0.001
Any cardiac history [‡] (%)	39.4	26.7	<0.001	38.9	0.599
Weight loss in last year (%)	14.2	15.1	0.014	15.2	0.050

*Z-score from regression for continuous outcomes, Pearson's Chi squared for categorical outcomes; [†] (ex & current smokers); [‡]angina, MI, CVE, HTN

Objective Two – GWAS

12,321,875 imputed SNPs were tested in the GWAS. LD score regression showed very strong negative correlations with continuous lung function measures from a large GWAS,⁽¹⁵⁾ FEV₁ (rg -0.93, p value 2.39⁻³⁷), FVC (rg -0.89, p value 1.13⁻³⁵), with strong negative correlation for FEV₁/FVC ratio (-0.19, p value 0.004).

953 SNPs met the GWAS significance threshold of p value <5x10⁻⁸, all of which were on chromosome 17 (**Figure 1**). After LD-clumping two SNPs remained (p value <5x10⁻⁸): rs378392 (chr17:43687542) and rs553682853 (chr17:44759766) with OR of PRISm 0.992 (0.990-0.994) and 0.990 (0.988-0.992) with C and G allele respectively.

PheWAS did not show that either SNP has previously been associated with lung function or lung disease. They have been described in association with other traits, see supplementary tables, including circulating cells. Rs378392 is an intron in the ribosomal protein S26

pseudogene 8 gene (*RPS26P8*), a processed pseudogene in the 17q21 locus.⁽²¹⁾ This gene has been associated with FEV₁ and FVC in previous studies of non-UK Biobank populations.⁽²⁶⁾

Figure 1. Each point represents one of the 755,702 SNPs that reached p value <0.05 in our GWAS

Objective Two – Conditional analysis

Conditional analysis found one p value <5x10⁻⁸ SNP, rs1991556 in chromosome 17 (17:44083402). Per G allele it confers OR (95%CI) 0.91 (0.88–0.94) of PRISm. PheWAS showed that this SNP has a very strong association with FEV₁ (p value 4.3x10⁻²⁵) and FVC (p value 7.5x10⁻¹⁸) percent predicted, some evidence of association with self-reported emphysema/chronic bronchitis (p value 7.8x10⁻⁴) and age of asthma diagnosis (p value 9.7x10⁻⁴). It is found in the *MAPT* gene discussed below.

Objective Two – Gene-based analysis

Gene-based analysis found one p value <5x10⁻⁸ SNP, rs7213474 (17:79962718), an intron variant in the Alveolar Soft Part Sarcoma Gene (*ASPSCR1*) gene on chromosome 17.⁽²⁰⁾ Extracting from the GWAS it gives OR (95% CI) 1.006 (1.004 – 1.008) per T allele. PheWAS showed it has previously been strongly associated with FEV₁, FVC and PEF_R (all p value <0.5x10⁻⁸).

8 genes reached significance threshold, (Table 2) the 6 most significant are all on chromosome 17. All genes discovered are protein coding.

Table 2. Results of Gene based analysis. Genes p-value <2.36x10⁻⁶

Chr	Gene	nSNPs	Pvalue for Gene	Top SNP	Top SNP Pvalue
17	<i>CRHR1</i>	281	9.9 x 10 ⁻⁰⁷	rs113173628	2.40 x 10 ⁻⁰⁷
17	<i>KANSL1</i>	88	9.9 x 10 ⁻⁰⁷	rs112333322	1.60 x 10 ⁻⁰⁷
17	<i>MAPT</i>	73	9.9 x 10 ⁻⁰⁷	rs118087478	1.60 x 10 ⁻⁰⁷
17	<i>MAPT-ASI</i>	26	9.9 x 10 ⁻⁰⁷	rs111962225	3.30 x 10 ⁻⁰⁷
17	<i>MAPT-ITI</i>	5	9.9 x 10 ⁻⁰⁷	rs62056785	3.90 x 10 ⁻⁰⁷
17	<i>NSF</i>	69	9.9 x 10 ⁻⁰⁷	rs1378358	8.70 x 10 ⁻⁰⁷
20	<i>ZGPAT</i>	81	1.9 x 10 ⁻⁰⁶	rs4809327	6.10 x 10 ⁻⁰⁸
6	<i>ZKSCAN4</i>	19	1.9 x 10 ⁻⁰⁶	rs9986596	9.30 x 10 ⁻⁰⁷

Three of the genes with the strongest signal; *CRHR1* (Corticotrophin-Releasing Hormone Receptor 1), *KANSL1* (KAT8 Regulatory NSL Complex Subunit 1) and *MAPT* (Microtubule Associated Protein Tau) are all in the same 17q21 locus. Variants in this locus are the strongest known genetic determinants of early-onset asthma and are associated with frequent exacerbation⁽²⁷⁾ and persistence of symptoms.⁽²⁸⁾ Variants in *CRHR1* and *MAPT* have both been shown to affect lung function response to corticosteroid.⁽²⁹⁾ An integrative analysis combining results of a COPD GWAS with an expression quantitative trait locus analysis from lung tissue samples reported *MAPT* as the top candidate gene for causing COPD.⁽³⁰⁾ *KANSL1* has been associated with extremes of FEV₁, with the strongest signal seen in never smokers.⁽³¹⁾

NSF gene (N-Ethylmaleimide Sensitive Factor) also reached significance in chromosome 17. It contains rs553682853 which was discovered in our GWAS. *NSF* has been associated with FEV₁ and FVC in UKBB previously.⁽²¹⁾

A PheWAS was conducted on the top SNP in the *ZGPAT* gene (Zinc Finger CCCH-Type And G-Patch Domain Containing), p value 6.10 x 10⁻⁰⁸. PheWAS showed strong associations for rs4809327 with FEV₁ percent predicted, FVC and wheeze in the last year (p value <0.5x10⁻⁸) and has weaker associations with self-reported asthma (p value 5.6x10⁻⁵). Literature review

found evidence from an epigenetic paper that CpG sites in *ZGPAT* have been shown to have an inverse causal effect on FEV₁ and COPD.⁽³²⁾

The top SNP in *ZKSCAN4* on chromosome 6, rs9986596, has been associated with FEV₁ percent predicted, FVC and PEFR (all p value <0.5x10⁻⁸) and asthma diagnosis, p value 2.7x10⁻⁶.

With the exception of *CRHR1*, all genes discovered are known to be expressed in human lung tissue.⁽²⁰⁾ *CRHR1* is found in mouse lung tissue and mutations in it can cause pulmonary alveolar haemorrhage, abnormal lung morphology, emphysema and respiratory distress.⁽³³⁾

Discussion

Evidence before this study

COPD is a common and heterogenous diseases, defined by airway obstructive spirometry.⁽⁹⁾

The current diagnostic criteria of COPD excludes many individuals with smoking histories and a high burden of respiratory symptoms.^(34, 35) Many of these excluded individuals have PRISm, and although they do not meet the diagnostic spirometric criteria for COPD they exhibit similar changes such as small airways disease, emphysema and gas trapping.^(4, 36, 37)

Little focus has been given to whether asthma contributes to PRISm, despite high rates of self-reported asthma in other PRISm cohorts (21% in COPDgene⁽²⁾) and in this UKBB cohort reported asthma among PRISm was 18.4%.⁽⁴⁾ There are many plausible ways asthma could contribute to both PRISm and airflow obstruction. Asthma causes small airways obstruction, CT changes of airways disease and gas trapping, and is a significant risk factor for COPD.⁽³⁸⁾

Research thus far has been produced by cohorts with small PRISm samples, with some excluding never smokers. This has limited generalizability of findings and prevented genetic variants from being discovered.

Added value of this study

The sample size and detailed data of the UKBiobank cohort enables reliable assessment of individuals with PRISm. Estimates of prevalence and associations of PRISm compared to normal and obstructive spirometry are well powered and generalizable. The increased sample size also allows for genetic discoveries.

PRISm in UKBiobank is common at 8%, and slightly more prevalent than stage II-IV obstruction. This is lower than many estimates, but is similar to another cohort that included never smokers.⁽³⁾ The importance of smoking as a risk factor was replicated in our analysis. It is unclear why some smokers develop PRISm and others COPD. The pack/years of those with COPD is higher but it does not fully explain the risk of PRISm given the transition between normal lung function, PRISm and COPD seen in longitudinal work and the non-smoking related genetic risk factors we found. BMI does not appear to be as clinically relevant as previously described, providing further evidence that PRISm is not simply due to extra-pulmonary restriction related to obesity.⁽³⁾ The high prevalence of cardiovascular disease and diabetes in PRISm even after adjustment for confounders is important. COPD may have a causal effect on extra-pulmonary disease, perhaps via systemic inflammation or oxidative stress,⁽³⁹⁾ it is conceivable this could occur in PRISm too. Even if no casual pathway between PRISm and its co-morbidity is found, future studies to determine if screening for diabetes and cardiovascular disease in PRISm are warranted.

Our study is the first to utilise conditional analysis and gene-based analysis in PRISm and the first to successfully identify genetic variants. GWAS identified two SNPs associated with PRISm that are novel for lung function and disease. Conditional analysis and gene-based analysis found genetic variants with reported associations for wheeze, asthma, COPD, and inhaled steroid response in asthma and COPD. COPD is no longer viewed as a single organ self-inflicted

smoking disease, but a result of genetic and early life factors that determine lung function trajectories.^(6, 7) Our findings that PRISm and COPD have shared genetic risk factors could partially explain the transition between them. Variants pre-disposing to asthma, especially as a child, suggest that early life factors and disease could lead to a trajectory that includes PRISm.

Limitations

UKBiobank has only conducted cross sectional lung function tests, so we are unable to add to longitudinal follow up, but our genetic findings contribute to our understanding of why PRISm can progress to COPD. Post-bronchodilator spirometry has not been performed in UKBiobank which is the gold standard for COPD diagnosis. However, post-bronchodilator spirometry is not used for diagnosis of PRISm, and the effect of bronchodilation in PRISm is not known. We used FEV_1 predicted $<80\%$ and $FEV_1/FVC <0.7$ to define obstructive spirometry. Obstructive spirometry is not the same as COPD which remains a clinical diagnosis assuming spirometric criteria are fulfilled. Using lower limits of normal (LLN), FEV_1/FVC ratio and post-bronchodilator spirometry may have reduced numbers classified as having obstructive spirometry. By eliminating stage 1 obstruction from the analysis, we are likely to have eliminated many individuals whose FEV_1/FVC ratio was below 0.7, but within the LLN. Many of our conclusions are drawn from comparing PRISm to normal spirometry, which would not be affected by bronchodilation or the use of LLN FEV_1/FVC . We do not have access to lung volumes or gas transfer, but they are not necessary for the diagnosis of PRISm. Interstitial lung diseases are very rare and comprise $<0.1\%$ of UKBiobank so are unlikely to influence results. We do not have access to a second cohort for a replication analysis of our genetic discoveries, which potentially means our results are limited by “Winners Curse”.⁽⁴⁰⁾

Future Research

Research assessing underlying structural and functional lung changes of PRISm is warranted. Particularly the frequency and severity of small airways dysfunction, perhaps via FEF₂₅₋₇₅ or measures of resistance from forced oscillation testing, may provide further insight into underlying pathology in PRISm and contribute to COPD prediction. Small airways obstruction may be amenable to treatment with bronchodilators, reversing spirometry to normal or preventing obstruction. Given our findings, research into the initiation and optimisation of inhaled steroids in PRISm may be warranted, especially in those with pre-existing asthma. Lung volume measurements may not produce targets for currently available treatments but could aid risk stratification and understanding of PRISm pathogenesis. Trials examining the utility of screening for diabetes and cardiovascular risk could help reduce morbidity.

Conclusion

PRISm affects 8% of adults, reporting a high burden of breathlessness and a ~40% cardiovascular comorbidity rate. Genetic variants that increase the risk of obstructive lung diseases in childhood and adults are strongly associated with PRISm. This provides novel biological mechanisms for the observed transition between PRISm and obstructive spirometry and has implications for prevention, diagnosis, and treatment of PRISm.

Acknowledgements

This research was conducted using UKBiobank resource (project number 55521). We would like to thank all UKBiobank participants and all staff involved in UKBiobank.

References

1. Mannino DM, McBurnie MA, Tan W, Kocabas A, Anto J, Vollmer WM, Buist AS. Restricted spirometry in the Burden of Lung Disease Study. *The international journal of tuberculosis and lung disease : the official journal of the International Union against Tuberculosis and Lung Disease* 2012; 16: 1405-1411.
2. Wan ES, Castaldi PJ, Cho MH, Hokanson JE, Regan EA, Make BJ, Beaty TH, Han MK, Curtis JL, Curran-Everett D, Lynch DA, DeMeo DL, Crapo JD, Silverman EK, Investigators CO. Epidemiology, genetics, and subtyping of preserved ratio impaired spirometry (PRISm) in COPDGene. *Respiratory research* 2014; 15: 89-89.

3. Wijnant SRA, De Roos E, Kavousi M, Stricker BH, Terzikhan N, Lahousse L, Brusselle GG. Trajectory and mortality of preserved ratio impaired spirometry: the Rotterdam Study. *The European respiratory journal* 2020; 55: 1901217.
4. Wan ES, Fortis S, Regan EA, Hokanson J, Han MK, Casaburi R, Make BJ, Crapo JD, DeMeo DL, Silverman EK, Investigators CO. Longitudinal Phenotypes and Mortality in Preserved Ratio Impaired Spirometry in the COPD Gene Study. *American journal of respiratory and critical care medicine* 2018; 198: 1397-1405.
5. Lundbäck B, Backman H, Calverley PMA. Lung Function through the PRISm. Spreading Light or Creating Confusion? *American journal of respiratory and critical care medicine* 2018; 198: 1358-1360.
6. Agusti A, Noell G, Brugada J, Faner R. Lung function in early adulthood and health in later life: a transgenerational cohort analysis. *The Lancet Respiratory medicine* 2017; 5: 935-945.
7. Shrine N, Guyatt AL, Erzurumluoglu AM, Jackson VE, Hobbs BD, Melbourne CA, Batini C, Fawcett KA, Song K, Sakornsakolpat P, Li X, Boxall R, Reeve NF, Obeidat M, Zhao JH, Wielscher M, Weiss S, Kentistou KA, Cook JP, Sun BB, Zhou J, Hui J, Karrasch S, Imboden M, Harris SE, Marten J, Enroth S, Kerr SM, Surakka I, Vitart V, Lehtimäki T, Allen RJ, Bakke PS, Beaty TH, Bleecker ER, Bossé Y, Brandsma C-A, Chen Z, Crapo JD, Danesh J, DeMeo DL, Dudbridge F, Ewert R, Gieger C, Gulsvik A, Hansell AL, Hao K, Hoffman JD, Hokanson JE, Homuth G, Joshi PK, Joubert P, Langenberg C, Li X, Li L, Lin K, Lind L, Locantore N, Luan J, Mahajan A, Maranville JC, Murray A, Nickle DC, Packer R, Parker MM, Paynton ML, Porteous DJ, Prokopenko D, Qiao D, Rawal R, Runz H, Sayers I, Sin DD, Smith BH, Soler Artigas M, Sparrow D, Tal-Singer R, Timmers PRHJ, Van den Berge M, Whittaker JC, Woodruff PG, Yerges-Armstrong LM, Troyanskaya OG, Raitakari OT, Kähönen M, Polašek O, Gyllenstein U, Rudan I, Deary IJ, Probst-Hensch NM, Schulz H, James AL, Wilson JF, Stubbe B, Zeggini E, Jarvelin M-R, Wareham N, Silverman EK, Hayward C, Morris AP, Butterworth AS, Scott RA, Walters RG, Meyers DA, Cho MH, Strachan DP, Hall IP, Tobin MD, Wain LV, Understanding Society Scientific G. New genetic signals for lung function highlight pathways and chronic obstructive pulmonary disease associations across multiple ancestries. *Nature Genetics* 2019; 51: 481-493.
8. <https://www.ukbiobank.ac.uk/>.
9. Global Initiative for Chronic Obstructive Pulmonary Lung Disease; 2019.
10. STATA Corp. Stata Statistical Software: Release 15. 2017.
11. Loh PR, Kichaev G, Gazal S, Schoech AP, Price AL. Mixed-model association for biobank-scale datasets. *Nat Genet* 2018; 50: 906-908.
12. Ruth Mitchell GH, Tom Dudding, Laura Corbin, Sean Harrison, Lavinia Paternoster. UK BioBank Genetic Data: MRC-IEU Quality Control, version 2. 2019.
13. Zheng J, Erzurumluoglu AM, Elsworth BL, Kemp JP, Howe L, Haycock PC, Hemani G, Tansey K, Laurin C, Early G, Lifecourse Epidemiology Eczema C, Pourcain BS, Warrington NM, Finucane HK, Price AL, Bulik-Sullivan BK, Anttila V, Paternoster L, Gaunt TR, Evans DM, Neale BM. LD Hub: a centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. *Bioinformatics (Oxford, England)* 2017; 33: 272-279.
14. Bulik-Sullivan B, Finucane HK, Anttila V, Gusev A, Day FR, Loh P-R, Duncan L, Perry JRB, Patterson N, Robinson EB, Daly MJ, Price AL, Neale BM, ReproGen C, Psychiatric Genomics C, Genetic Consortium for Anorexia Nervosa of the Wellcome Trust Case Control C. An atlas of genetic correlations across human diseases and traits. *Nature Genetics* 2015; 47: 1236-1241.
15. Soler Artigas M, Wain LV, Miller S, Kheirallah AK, Huffman JE, Ntalla I, Shrine N, Obeidat M, Trochet H, McArdle WL, Alves AC, Hui J, Zhao JH, Joshi PK, Teumer A, Albrecht E, Imboden M, Rawal R, Lopez LM, Marten J, Enroth S, Surakka I, Polasek O, Lytikainen LP, Granell R, Hysi PG, Flexeder C, Mahajan A, Beilby J, Bosse Y, Brandsma CA, Campbell H, Gieger C, Glaser S, Gonzalez JR, Grallert H, Hammond CJ, Harris SE, Hartikainen AL, Heliövaara M, Henderson J, Hocking L, Horikoshi M, Hutri-Kahonen N, Ingelsson E, Johansson A, Kemp JP, Kolcic I, Kumar

- A, Lind L, Melen E, Musk AW, Navarro P, Nickle DC, Padmanabhan S, Raitakari OT, Ried JS, Ripatti S, Schulz H, Scott RA, Sin DD, Starr JM, Vinuela A, Volzke H, Wild SH, Wright AF, Zemunik T, Jarvis DL, Spector TD, Evans DM, Lehtimäki T, Vitart V, Kahonen M, Gyllensten U, Rudan I, Deary IJ, Karrasch S, Probst-Hensch NM, Heinrich J, Stubbe B, Wilson JF, Wareham NJ, James AL, Morris AP, Jarvelin MR, Hayward C, Sayers I, Strachan DP, Hall IP, Tobin MD. Sixteen new lung function signals identified through 1000 Genomes Project reference panel imputation. *Nature communications* 2015; 6: 8658.
16. Yang J, Ferreira T, Morris AP, Medland SE, Madden PA, Heath AC, Martin NG, Montgomery GW, Weedon MN, Loos RJ, Frayling TM, McCarthy MI, Hirschhorn JN, Goddard ME, Visscher PM. Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat Genet* 2012; 44: 369-375, s361-363.
17. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet* 2011; 88: 76-82.
18. Fraser A, Macdonald-Wallis C, Tilling K, Boyd A, Golding J, Davey Smith G, Henderson J, Macleod J, Molloy L, Ness A, Ring S, Nelson SM, Lawlor DA. Cohort Profile: the Avon Longitudinal Study of Parents and Children: ALSPAC mothers cohort. *International journal of epidemiology* 2013; 42: 97-110.
19. Mishra A, Macgregor S. VEGAS2: Software for More Flexible Gene-Based Testing. *Twin Research and Human Genetics* 2015; 18: 86-91.
20. NCBI. www.ncbi.nlm.nih.gov/gene. 2020 20/07/2020].
21. Staley JR, Blackshaw J, Kamat MA, Ellis S, Surendran P, Sun BB, Paul DS, Freitag D, Burgess S, Danesh J, Young R, Butterworth AS. PhenoScanner: a database of human genotype-phenotype associations. *Bioinformatics (Oxford, England)* 2016; 32: 3207-3209.
22. Science Wlo. www.genecards.org. 20/07/2020].
23. Yu W, Gwinn M, Clyne M, Yesupriya A, Khoury MJ. A navigator for human genome epidemiology. *Nat Genet* 2008; 40: 124-125.
24. CDC. <https://phgkb.cdc.gov/PHGKB/hNHome.action>.
25. Hemani G, Zheng J, Elsworth B, Wade KH, Haberland V, Baird D, Laurin C, Burgess S, Bowden J, Langdon R, Tan VY, Yarmolinsky J, Shihab HA, Timpson NJ, Evans DM, Relton C, Martin RM, Davey Smith G, Gaunt TR, Haycock PC. The MR-Base platform supports systematic causal inference across the human phenome. *eLife* 2018; 7.
26. NHGRI-EBI. <https://www.ebi.ac.uk/gwas/>.
27. Farzan N, Vijverberg SJ, Hernandez-Pacheco N, Bel EHD, Berce V, Bønnelykke K, Bisgaard H, Burchard EG, Canino G, Celedón JC, Chew FT, Chiang WC, Cloutier MM, Forno E, Francis B, Hawcutt DB, Herrera-Luis E, Kabesch M, Karimi L, Melén E, Mukhopadhyay S, Merid SK, Palmer CN, Pino-Yanes M, Pirmohamed M, Potočnik U, Repnik K, Schieck M, Sevelsted A, Sio YY, Smyth RL, Soares P, Söderhäll C, Tantisira KG, Tavendale R, Tse SM, Turner S, Verhamme KM, Maitland-van der Zee AH. 17q21 variant increases the risk of exacerbations in asthmatic children despite inhaled corticosteroids use. *Allergy* 2018; 73: 2083-2088.
28. Granell R, Henderson AJ, Timpson N, St Pourcain B, Kemp JP, Ring SM, Ho K, Montgomery SB, Dermizakis ET, Evans DM, Sterne JA. Examination of the relationship between variation at 17q21 and childhood wheeze phenotypes. *The Journal of allergy and clinical immunology* 2013; 131: 685-694.
29. Tantisira KG, Lazarus R, Litonjua AA, Klanderma B, Weiss ST. Chromosome 17: association of a large inversion polymorphism with corticosteroid response in asthma. 2008; 18: 733-737.
30. Morrow JD, Cho MH, Platig J, Zhou X, DeMeo DL, Qiu W, Celli B, Marchetti N, Criner GJ, Bueno R, Washko GR, Glass K, Quackenbush J, Silverman EK, Hersh CP. Ensemble genomic analysis in human lung tissue identifies novel genes for chronic obstructive pulmonary disease. *Human genomics* 2018; 12: 1.
31. Wain LV, Shrine N, Miller S, Jackson VE, Ntalla I, Soler Artigas M, Billington CK, Kheirallah AK, Allen R, Cook JP, Probert K, Obeidat M, Bosse Y, Hao K, Postma DS, Pare PD, Ramasamy A, Magi R, Mihailov E, Reinmaa E, Melen E, O'Connell J, Frangou E, Delaneau O, Freeman C, Petkova D,

- McCarthy M, Sayers I, Deloukas P, Hubbard R, Pavord I, Hansell AL, Thomson NC, Zeggini E, Morris AP, Marchini J, Strachan DP, Tobin MD, Hall IP. Novel insights into the genetics of smoking behaviour, lung function, and chronic obstructive pulmonary disease (UK BiLEVE): a genetic association study in UK Biobank. *The Lancet Respiratory medicine* 2015; 3: 769-781.
32. Jamieson E, Korologou-Linden R, Wootton RE, Guyatt AL, Battram T, Burrows K, Gaunt TR, Tobin MD, Munafò M, Davey Smith G, Tilling K, Relton C, Richardson TG, Richmond RC. Smoking, DNA Methylation, and Lung Function: a Mendelian Randomization Analysis to Investigate Causal Pathways. *Am J Hum Genet* 2020; 106: 315-326.
33. <http://www.informatics.jax.org/>. 20/07/2020].
34. Adibi A, Sadatsafavi M. Looking at the COPD spectrum through "PRISm". 2020; 55: 1902217.
35. Barnes PJ, Vestbo J, Calverley PM. The Pressing Need to Redefine "COPD". *Chronic Obstr Pulm Dis* 2019; 6: 380-383.
36. Lowe KE, Regan EA, Anzueto A, Austin E, Austin JHM, Beaty TH, Benos PV, Benway CJ, Bhatt SP, Bleecker ER, Bodduluri S, Bon J, Boriak AM, Boueiz AR, Bowler RP, Budoff M, Casaburi R, Castaldi PJ, Charbonnier JP, Cho MH, Comellas A, Conrad D, Costa Davis C, Criner GJ, Curran-Everett D, Curtis JL, DeMeo DL, Diaz AA, Dransfield MT, Dy JG, Fawzy A, Fleming M, Flenaugh EL, Foreman MG, Fortis S, Gebrekristos H, Grant S, Grenier PA, Gu T, Gupta A, Han MK, Hanania NA, Hansel NN, Hayden LP, Hersh CP, Hobbs BD, Hoffman EA, Hogg JC, Hokanson JE, Hoth KF, Hsiao A, Humphries S, Jacobs K, Jacobson FL, Kazerooni EA, Kim V, Kim WJ, Kinney GL, Koegler H, Lutz SM, Lynch DA, MacIntyre NR, Jr., Make BJ, Marchetti N, Martinez FJ, Maselli DJ, Mathews AM, McCormack MC, McDonald MN, McEvoy CE, Moll M, Molye SS, Murray S, Nath H, Newell JD, Jr., Occhipinti M, Paoletti M, Parekh T, Pistolesi M, Pratte KA, Putcha N, Ragland M, Reinhardt JM, Rennard SI, Rosiello RA, Ross JC, Rossiter HB, Ruczinski I, San Jose Estepar R, Sciruba FC, Sieren JC, Singh H, Soler X, Steiner RM, Strand MJ, Stringer WW, Tal-Singer R, Thomashow B, Vegas Sánchez-Ferrero G, Walsh JW, Wan ES, Washko GR, Michael Wells J, Wendt CH, Westney G, Wilson A, Wise RA, Yen A, Young K, Yun J, Silverman EK, Crapo JD. COPDGene[®] 2019: Redefining the Diagnosis of Chronic Obstructive Pulmonary Disease. *Chronic Obstr Pulm Dis* 2019; 6: 384-399.
37. Wei X, Ding Q, Yu N, Mi J, Ren J, Li J, Xu S, Gao Y, Guo Y. Imaging Features of Chronic Bronchitis with Preserved Ratio and Impaired Spirometry (PRISm). *Lung* 2018; 196: 649-658.
38. Silva GE, Sherrill DL, Guerra S, Barbee RA. Asthma as a risk factor for COPD in a longitudinal study. *Chest* 2004; 126: 59-65.
39. MacLay JD, MacNee W. Cardiovascular disease in COPD: mechanisms. *Chest* 2013; 143: 798-807.
40. Kraft P. Curses--winner's and otherwise--in genetic epidemiology. *Epidemiology (Cambridge, Mass)* 2008; 19: 649-651; discussion 657-648.

Online Data Supplement

Epidemiology and Genetics of Preserved Ratio Impaired Spirometry (PRISm): An Analysis of UK Biobank

Daniel H Higbee, Raquel Granell PhD, George Davey Smith FRS, James W Dodd

WITHDRAWN
see manuscript DOI for details

Contents

Appendix 1. UKBiobank data fields used

Appendix 2. Flow chart of analysis

Appendix 3. Details of asthma and COPD diagnosis in UKBiobank and results of genetic analysis done excluded those with asthma and COPD

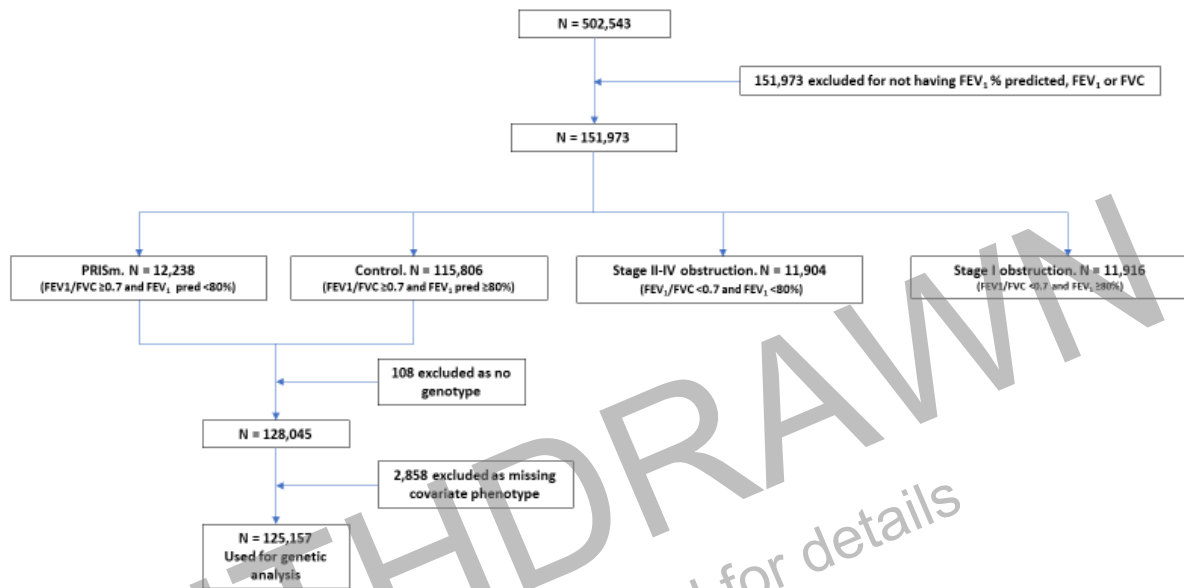
Appendix 4. Supplementary tables

Appendix 1. UKBiobank data fields used

Table E1. UKBiobank data fields used

Phenotype	UKBiobank data field (and link)
Forced expiratory volume in 1-second, predicted percentage	20154 (http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=20154)
Forced expiratory volume in 1-second (FEV1), Best measure	2010 (http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=20150)
Forced vital capacity (FVC), Best measure	20150 (http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=20151)
Smoking status	20116 (http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=20116)
Pack years of smoking	20161 (http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=20161)
Year of birth	34 (http://biobank.ndph.ox.ac.uk/showcase/field.cgi?id=34)
Age at recruitment	21102 (http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=21022)
Sex	31 (http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=31)
Body Mass Index	21001 (http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=21001)
Diabetes diagnosed by a doctor	2443 (http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=2443)
Self-reported asthma	2002 (http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=20002)
Self-reported COPD	2002 (http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=20002)
Doctor diagnosed asthma	22127 (http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=22127)
Doctor diagnosed COPD	22130 (http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=22130)
Diagnosis – ICD 10	41270 (http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=41270)
Diagnosis secondary ICD10	41204 (http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=41204)
Vascular/heart problems diagnosed by doctor	6150 (http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=6150)
Shortness of breath walking on level ground	4717 (http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=4717)
Weight change compared with 1 year ago	2306 (http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=2306)

Appendix 2. Flow chart of analysis



Appendix 3. Details of asthma and COPD diagnosis in UKBiobank and results of genetic analysis done excluded those with asthma and COPD

Asthma and COPD Diagnosis

Table E2. Asthma and COPD diagnosis in UKBiobank

	PRISm N= 12,238	Control N = 115,807	Obstruction N = 11,904
Self reported asthma*	2,262 (18.49%)	12,101 (10.45%)	4,209 (35.36%)
Self reported COPD	608 (4.97%)	92 (0.08%)	536 (4.51%)
ICD codes asthma†	1562 (12.77%)	6,299 (5.44%)	2,585 (21.72%)
ICD codes COPD	626 (5.12%)	1,019 (0.88%)	2,527 (21.23%)
	PRISm N = 2,721	Control N= 30,829	Obstruction N = 2,118
Doctor diagnosed asthma	684 (25.14%)	3,295 (10.69%)	842 (39.80%)
Doctor diagnosed COPD	103 (3.82%)	1,359 (4.41%)	377 (17.84%)

*We reported asthma diagnosis using UKBiobank variable 20002, "Non-cancer illness code, self-reported". Participants stated their illnesses, if they were uncertain of the type of illness, it was described to the interviewer (a trained nurse) who attempted to place it in the coding tree.

†ICD codes were using linked hospital ICD 10 codes.

Doctor diagnosed asthma or COPD uses UKBiobank variable 22127 and 22130 respectively. Participants were asked "Has a doctor ever told you that you have had any of the conditions below?" and asthma and COPD were listed.

Exclusion based on criteria strongly influenced by factors liable to cause an outcome can generate collider bias. Therefore, we present the results of genetic analysis after exclusion of participants with asthma and COPD here in the supplementary information.

Objective Two - GWAS

Exclusion of participants with asthma and COPD left 9408 cases and 109724 controls for analysis. GWAS produced no SNPs at p value $<5 \times 10^{-8}$, but slightly reducing the threshold to p value $<5 \times 10^{-7}$ produced 356 SNPs, pre-dominantly on chromosome 17. After clumping SNP rs189777036 remained which is in complete linkage disequilibrium ($D' = 1.0$, $r^2 = 1.0$) with rs378392, therefore the same signal was found in the GWAS despite exclusion of those with pre-existing airways disease.⁽⁴⁰⁾

Objective Two - Conditional analysis

Following exclusion of participants with asthma and COPD, repeat conditional analysis with a threshold of p value $<5 \times 10^{-7}$ found rs886249. This SNP is in near complete linkage disequilibrium with rs1991556 ($D' = 1.0$, $r^2 = 0.94$), which was found prior to exclusion of those with asthma and COPD. Therefore, the same signal was found in the GWAS despite exclusion of those with pre-existing airways disease

Objective Two - Gene-based analysis

After excluding all participants with asthma and COPD and re-running the gene-based analysis, all 6 genes on chromosome 17 still reached the significance threshold with the same top SNPs reported. No Top SNP reached p -value $<5^{-08}$.

Appendix 4.

Table E3. SNPs found via all Objective Two methods

SNP	Discovery	Chr:Pos	Gene	Function	EA	OA	EAF	beta	SE	pval
rs378392	GWAS	17:43687542	<i>RPS26P8</i>	Intron	C	G	0.72	-0.0078	0.001	1.4^{-8}
rs553682853	GWAS	17:44759766	<i>NSF</i>	Intron	G	A	0.84	-0.0098	0.001	4.1^{-8}
rs1991556	Conditional	17:44083402	<i>MAPT</i>	Intron	G	A*	0.77	-0.0917	0.016	1.0^{-8}
rs7213474	Gene-based	17:79962718	<i>ASPSCR1</i>	Intron	T*	C*	0.33*	0.0065*	0.001*	1.3^{-8}

*Extracted from GWAS as not supplied by discovery method

EA – Effect Allele. OA – Other Allele. EAF – Effect Allele Frequency. SE – Standard Error.
eQTL = expression quantitative trait loci

CHR – Chromosome. Pos – Position. Build GRCh37

Figure 1. Each point represents one of the 755,702 SNPs that reached p value <0.05 in our GWAS

WITHDRAWN
see manuscript DOI for details

