






# Pleiotropic associations of heterozygosity for the *SERPINA1* Z allele in the UK Biobank

Katherine A. Fawcett<sup>1</sup>, Kijoung Song<sup>2</sup>, Guoqing Qian<sup>3,4</sup>, Aliko-Eleni Farmaki <sup>5</sup>, Richard Packer<sup>1</sup>, Catherine John<sup>1</sup>, Nick Shrine <sup>1</sup>, Raquel Granell<sup>6</sup>, Sue Ring<sup>6</sup>, Nicholas J. Timpson<sup>6,7</sup>, Laura M. Yerges-Armstrong<sup>2</sup>, Richard Eastell<sup>8</sup>, Louise V. Wain<sup>1,9</sup>, Robert A. Scott<sup>10</sup>, Martin D. Tobin <sup>1,9,11</sup> and Ian P. Hall<sup>4,11</sup>

**Affiliations:** <sup>1</sup>Dept of Health Sciences, University of Leicester, Leicester, UK. <sup>2</sup>Human Genetics, GlaxoSmithKline, Collegeville, PA, USA. <sup>3</sup>Dept of General Internal Medicine, Ningbo First Hospital, Ningbo City, Zhejiang Province, China. <sup>4</sup>Division of Respiratory Medicine, University of Nottingham, and NIHR Nottingham BRC, NUH NHS Trust, Nottingham, UK. <sup>5</sup>MRC Unit for Lifelong Health and Ageing, Institute of Cardiovascular Science, University College London, London, UK. <sup>6</sup>Medical Research Council Integrative Epidemiology Unit, Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK. <sup>7</sup>Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK. <sup>8</sup>Dept of Oncology and Metabolism, University of Sheffield, Sheffield, UK. <sup>9</sup>National Institute for Health Research, Leicester Respiratory Biomedical Research Centre, Glenfield Hospital, Leicester, UK. <sup>10</sup>Human Genetics – R&D, GSK Medicines Research Centre, Stevenage, UK. <sup>11</sup>These authors contributed equally.

**Correspondence:** Katherine A. Fawcett, George Davies Centre, Dept of Health Sciences, University of Leicester, University Road, Leicester LE1 7RH, UK. E-mail: kaf19@leicester.ac.uk

**ABSTRACT** Homozygosity for the *SERPINA1* Z allele causes  $\alpha_1$ -antitrypsin deficiency, a rare condition that can cause lung and liver disease. However, the effects of Z allele heterozygosity on nonrespiratory phenotypes, and on lung function in the general population, remain unclear.

We conducted a large, population-based study to determine Z allele effects on >2400 phenotypes in the UK Biobank (N=303 353).

Z allele heterozygosity was strongly associated with increased height ( $\beta=1.02$  cm,  $p=3.91\times 10^{-68}$ ), and with other nonrespiratory phenotypes including increased risk of gall bladder disease, reduced risk of heart disease and lower blood pressure, reduced risk of osteoarthritis and reduced bone mineral density, increased risk of headache and enlarged prostate, as well as with blood biomarkers of liver function. Heterozygosity was associated with higher height-adjusted forced expiratory volume in 1 s (FEV<sub>1</sub>) ( $\beta=19.36$  mL,  $p=9.21\times 10^{-4}$ ) and FEV<sub>1</sub>/forced vital capacity ( $\beta=0.0031$ ,  $p=1.22\times 10^{-5}$ ) in nonsmokers, whereas in smokers, this protective effect was abolished. Furthermore, we show for the first time that sex modifies the association of the Z allele on lung function.

We conclude that Z allele heterozygosity and homozygosity exhibit opposing effects on lung function in the UK population, and that these associations are modified by smoking and sex. In exploratory analyses, heterozygosity for the Z allele also showed pleiotropic associations with nonrespiratory health-related traits and disease risk.



@ERSpublications

**Heterozygosity for the *SERPINA1* Z allele is associated with higher lung function in nonsmokers (but this advantage is abolished by smoking) as well as a variety of nonrespiratory diseases and traits, including liver-, heart- and bone-related conditions** <https://bit.ly/37XC0Yg>

**Cite this article as:** Fawcett KA, Song K, Qian G, *et al.* Pleiotropic associations of heterozygosity for the *SERPINA1* Z allele in the UK Biobank. *ERJ Open Res* 2021; 7: 00049-2021 [<https://doi.org/10.1183/23120541.00049-2021>].

This article has supplementary material available from [openres.ersjournals.com](https://openres.ersjournals.com)

Received: 22 Jan 2021 | Accepted: 20 Feb 2021

Copyright ©The authors 2021. This article is open access and distributed under the terms of the Creative Commons Attribution Licence 4.0.



## Introduction

Homozygosity for the *SERPINA1* Z allele (rs28929474(T)) is the commonest cause of severe  $\alpha_1$ -antitrypsin deficiency (AATD) and is a well-established genetic risk factor for lung diseases such as chronic obstructive pulmonary disease (COPD). However, the health consequences of heterozygosity for the Z allele are not as well-understood [1]. Given that approximately one in 30 Europeans is heterozygous for the Z allele, the phenotypic consequences of carriage of this allele could have important public health implications.

Some previous studies have sought to characterise the effect of Z allele heterozygosity on nonrespiratory traits, particularly liver diseases [2–6]. However, these have often been carried out in small sample sizes and/or clinical subgroups. The recent development of phenome-wide association studies (PheWASs) and the availability of the well-phenotyped UK Biobank population-based cohort provides a platform for systematic investigation of the effects of heterozygosity for the Z allele on nonrespiratory traits. PheWASs test the association between genetic variants and a large number of phenotypic traits, including diseases and their subtypes, and potential intermediate phenotypes [7]. This differs from genome-wide association studies, which test a large number of variants across the genome for association with only one trait.

The effect of Z allele heterozygosity on lung function traits and lung disease has been the subject of many studies. Recent COPD case–control and family-based studies have shown reduced lung function and increased risk of COPD in heterozygous current and former smokers [8–11]. However, a population-based study demonstrated no significant reductions in lung function in heterozygous smokers, despite having greater numbers of heterozygous smokers compared to previous studies [12]. It also showed enhanced lung function in heterozygous individuals overall, partially explained by strong association of the Z allele with increased height. This discrepancy may be due, in part, to the fact that identifying and recruiting study participants based on their health status (as in case–control studies [8, 9]) or based on the health status of a family member [10, 11] can lead to causal estimates that are subject to ascertainment bias [13], whereas population-based studies [12, 14] overcome these biases. The effects of the Z allele on lung function in relation to smoking status therefore remain uncertain.

Finally, despite evidence for sex-differential effects of Z allele homozygosity on lung function [15, 16], we are not aware of any studies that have compared the effect of Z allele heterozygosity on lung function in males *versus* females.

We therefore systematically evaluated the effects of Z allele heterozygosity in the UK biobank population, which in total includes >18000 Z allele heterozygotes. We aimed: 1) to undertake the most extensive PheWAS to date, including blood biomarkers, for Z allele heterozygosity and homozygosity to identify effects beyond the respiratory system; and 2) to fully define the effects of Z allele heterozygosity on lung function measures (forced expiratory volume in 1 s (FEV<sub>1</sub>) and forced vital capacity (FVC)) in smokers and nonsmokers and in males and females.

## Methods

### Cohorts

#### Ethics statement

This study used anonymised data from UK Biobank, which has ethical approval from the UK National Health Service National Research Ethics Service (ref. 11/NW/0382). All participants provided written informed consent. Ethical approval for the Avon Longitudinal Study of Parents and Children (ALSPAC) was obtained from the ALSPAC Ethics and Law Committee and the Local Research Ethics Committees. Written informed consent for the use of data collected *via* questionnaires and clinics was obtained from participants, parents or guardians following the recommendations of the ALSPAC Ethics and Law Committee at the time.

#### UK Biobank

The UK Biobank data resource is described elsewhere ([www.ukbiobank.ac.uk](http://www.ukbiobank.ac.uk)). Genotyping was undertaken using the Affymetrix Axiom UK BiLEVE and UK Biobank arrays [17]. Genotypes were imputed based on the Human Reference Consortium panel as described elsewhere [18]. Genotyping quality control was performed as described previously [19]. UK Biobank individuals were selected for inclusion in PheWAS analyses if they met the following criteria: 1) they had no missing data for sex or age; 2) they had genome-wide imputed genetic data for rs28929474; 3) they were of genetically determined European ancestry (see [19] for details); and 4) they were not first- or second-degree relatives. For lung function analyses, individuals were additionally required to have spirometry data that passed quality control, as described previously [19], and to have full data for height and smoking status as well as data derived from direct genotyping for rs28929474 (N=303 353).

### *The Avon Longitudinal Study of Parents and Children*

The ALSPAC cohort was invited to participate in this study after the main analyses in UK Biobank had been completed, in order to explore the effects of the Z allele on height and lung function in younger age groups. ALSPAC children were genotyped using the Illumina HumanHap550 quad genome-wide single-nucleotide polymorphism genotyping platform (Illumina Inc., San Diego, CA, USA) by the Wellcome Trust Sanger Institute (Cambridge, UK) and the Laboratory Corporation of America (Burlington, NC, USA), using support from 23andMe. Further details of ALSPAC are available in the supplementary material. Participants were excluded from this study if they had incorrectly recorded sex, minimal or excessive heterozygosity, disproportionate levels of individual missingness (>3%), evidence of cryptic relatedness or non-European ancestry.

### *Phenome-wide association studies*

To identify whether *SERPINA1* Z allele heterozygosity or homozygosity was associated with respiratory and nonrespiratory traits and diseases, two PheWASs across all available traits were performed. The first compared traits in individuals heterozygous for the Z allele *versus* wild type, and the second compared individuals homozygous for the Z allele *versus* wild type and heterozygotes combined (recessive genetic model). Up to 379 101 individuals were available for these analyses. Traits included UK Biobank baseline measures (from both questionnaires and physical measures), self-reported medication usage and operative procedures, as well as those captured in Office of Population Censuses and Surveys codes from the electronic health record. We also included self-reported disease variables and those from hospital episode statistics (HES) (International Classification of Disease, 10th revision (ICD-10) codes truncated to three-character codes, and combined in block and chapter groups) as well as combining both self-report and hospital-diagnosed diseases, where possible, to maximise power. The analysis included 2411 traits (traits with >200 cases were included [20]). Analyses were conducted in unrelated European-ancestry individuals (KING kinship coefficient of <0.0442), and were adjusted for age, age<sup>2</sup>, sex, genotyping array and 10 ancestry principal components (PCs). Logistic models were fitted for binary outcomes and linear models were fitted for quantitative outcomes (rank transformed to normality). False discovery rates (FDRs) were calculated according to the number of the traits in the analysis.

### *Statistical analyses for biomarkers, lung function, COPD and height*

For continuous traits, linear regression models adjusted for sex, age, age<sup>2</sup>, the first 10 ancestry-based PCs and genotyping array were tested in R. *SERPINA1* Z genotype (derived from direct genotyping) was coded according to the genetic model tested: either heterozygous (*i.e.* heterozygous *versus* wild type) or recessive. For lung function analyses, ever-smoking status and standing height were also included in the models unless indicated. To test for interaction of Z allele hetero- or homozygosity with ever-smoking status and sex, interaction terms for sex and smoking were added to the model. Association testing with moderate-severe COPD (defined as FEV<sub>1</sub>/FVC <0.7 and FEV<sub>1</sub><80% predicted) or COPD of Global Initiative for Chronic Obstructive Lung Disease (GOLD) grade 1–4 (defined as FEV<sub>1</sub>/FVC <0.7) was carried out using logistic regression in R adjusting for covariates sex, age, age<sup>2</sup>, ever-smoking status, height, the first 10 ancestry-based PCs and genotyping array. In addition to the biomarkers available from UK Biobank, we calculated estimated glomerular filtration rate (eGFR) using the following formula: if cystatin C (cys) ≤ 0.8, then eGFR = 133 \* ((cys/0.8)\*\*-0.499) \* (0.996\*\*age) \* [0.932 if female]; whereas if cys > 0.8, then eGFR = 133 \* ((cys/0.8)\*\*-1.328) \* (0.996\*\*age) \* [0.932 if female].

In ALSPAC, the association of Z allele heterozygosity with height and lung function z-scores were examined using linear regression models including age and gender (and including height in the lung function analyses).

### *Polygenic risk scores*

Polygenic risk scores for lower FEV<sub>1</sub>/FVC were generated in UK Biobank using Plink version 1.90b3. These were based on 60 730 autosomal variants with minor allele frequency >1% and p < 0.05 (for the association with FEV<sub>1</sub>/FVC in UK Biobank), and linkage disequilibrium pruned to r<sup>2</sup> < 0.1 in 250-kb windows. Variants were weighted by FEV<sub>1</sub>/FVC β-coefficients from the SpiroMeta consortium cohorts [19].

## **Results**

### *Frequency of the SERPINA1 Z in UK Biobank*

Amongst 411 002 unrelated, European UK Biobank participants with full sex, age, height, smoking status and genotyping data, 16 199 (3.94%) were heterozygous and 129 (0.0314%) were homozygous for the Z allele. There were fewer ZZ homozygotes than expected under Hardy–Weinberg equilibrium in all samples (p = 0.004), ever-smokers (p = 0.003), individuals >60 years of age (p = 0.007) and females (p = 0.007) (table S1), consistent with reduced survival.

### *PheWASs of the Z allele*

To explore the effect of Z allele heterozygosity and homozygosity on a broad range of phenotypes, we performed two PheWASs. The phenotypes most strongly associated with Z allele heterozygosity and Z allele homozygosity (FDR <0.01) are shown in figure 1a and b respectively (full results are in tables S2 and S3). Standing height, AATD and various respiratory phenotypes (lung function, reported emphysema, chronic bronchitis and COPD) were strongly associated with the Z allele under either genetic model. However, we also detected strong associations between Z allele heterozygosity and a number of other phenotypes, including increased risk of gall stones (OR 1.35,  $p=1.69\times 10^{-16}$ ), gall bladder removal (OR 1.51,  $p=1.85\times 10^{-29}$ ) and bile duct disease (OR 1.25,  $p=5.42\times 10^{-8}$ ); cardiovascular phenotypes such as lower risk of heart disease (OR 0.84,  $p=6.66\times 10^{-6}$ ) and family history of heart disease, and lower blood pressure ( $p=1.02\times 10^{-10}$ ); musculoskeletal phenotypes such as a higher grip strength ( $p=5.67\times 10^{-7}$ ), lower risk of osteoarthritis, particularly of knee or hip (OR 0.85,  $p=1.18\times 10^{-6}$ ), lower bone mineral density ( $p=1.40\times 10^{-5}$ ) and higher risk of osteoporosis (OR 1.17,  $p=0.001$ ); and higher risk of headache (OR 1.12,  $p=5.66\times 10^{-9}$ ) and enlarged prostate (OR 1.26,  $p=1.95\times 10^{-7}$ ). The PheWAS also included blood counts for a variety of different cell types, showing reduced reticulocytes in individuals heterozygous for the Z allele. Z allele homozygosity was associated with increased haemoglobin ( $p=7.17\times 10^{-10}$ ), haematocrit ( $p=6.17\times 10^{-8}$ ) and red blood cells ( $p=5.91\times 10^{-5}$ ), as well as optic neuritis (OR 26.9,  $p=4.32\times 10^{-6}$ ) and pancreatitis (OR 6.07,  $p=7.76\times 10^{-5}$ ).

To extend these analyses to traits not previously available for the PheWAS, we investigated the effect of Z allele heterozygosity and homozygosity on 30 key blood biochemistry markers in UK Biobank data (table 1). Heterozygosity for the Z allele was associated with a variety of liver function markers including higher alanine aminotransferase (ALT) ( $\beta=1.16\text{ U}\cdot\text{L}^{-1}$ ,  $p=2.95\times 10^{-25}$ ), higher albumin ( $\beta=0.71\text{ g}\cdot\text{L}^{-1}$ ,  $p=3.11\times 10^{-229}$ ), higher alkaline phosphatase (ALP) ( $\beta=2.97\text{ U}\cdot\text{L}^{-1}$ ,  $p=5.81\times 10^{-44}$ ), higher aspartate aminotransferase (AST) ( $\beta=0.71\text{ U}\cdot\text{L}^{-1}$ ,  $p=2.27\times 10^{-16}$ ) and higher direct bilirubin ( $\beta=0.041\text{ }\mu\text{mol}\cdot\text{L}^{-1}$ ,  $p=1.64\times 10^{-8}$ ). There were also strong associations with calcium ( $\beta=0.012\text{ mmol}\cdot\text{L}^{-1}$ ,  $p=4.01\times 10^{-50}$ ), cystatin C ( $0.0090\text{ mg}\cdot\text{L}^{-1}$ ,  $p=8.91\times 10^{-12}$ ), C-reactive protein ( $\beta=-0.17\text{ mg}\cdot\text{L}^{-1}$ ,  $p=1.11\times 10^{-6}$ ) and insulin-like growth factor (IGF)-1 ( $\beta=-0.21\text{ nmol}\cdot\text{L}^{-1}$ ,  $p=2.35\times 10^{-6}$ ). Heterozygotes exhibited higher sex hormone-binding globulin (SHBG) ( $\beta=4.53\text{ nmol}\cdot\text{L}^{-1}$ ,  $p=1.06\times 10^{-95}$ ) and testosterone ( $\beta=0.37\text{ nmol}\cdot\text{L}^{-1}$ ,  $p=1.84\times 10^{-57}$ ). The association with higher testosterone was driven by heterozygosity in males ( $\beta\pm\text{SE}=0.76\pm 0.045\text{ nmol}\cdot\text{L}^{-1}$ ,  $p=3.38\times 10^{-64}$ ) but not females ( $\beta\pm\text{SE}=-0.030\pm 0.0079\text{ nmol}\cdot\text{L}^{-1}$ ,  $p=0.00015$ ).

### *Z allele heterozygosity is associated with increased lung function in nonsmokers*

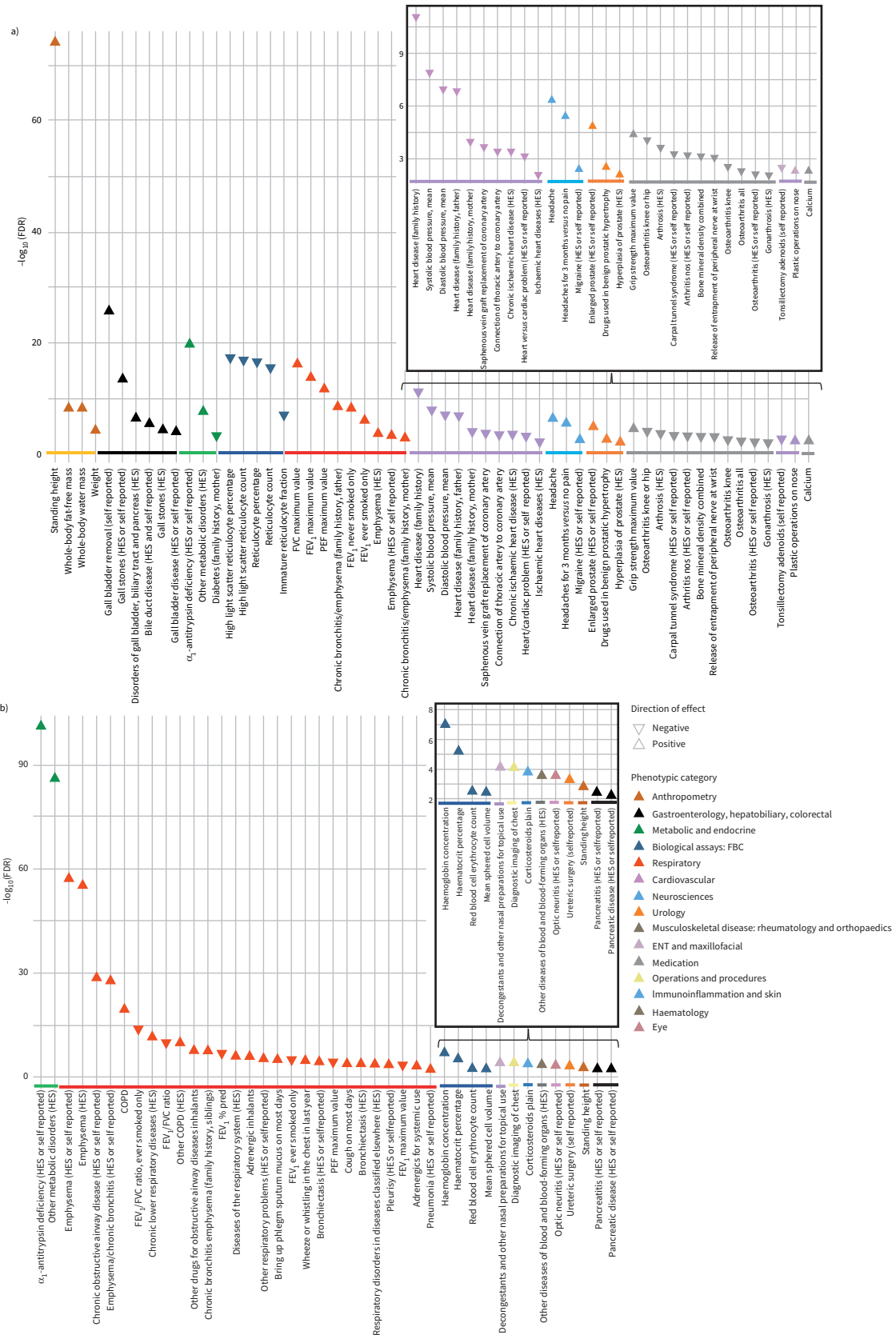
We tested association of heterozygosity for the *SERPINA1* Z allele with lung function traits in unrelated, European UK Biobank participants with full sex, age, height, smoking status, lung function and genotyping data (table 2). Individuals heterozygous for the Z allele exhibited higher FEV<sub>1</sub> ( $\beta=9.26\text{ mL}$ ,  $p=0.041$ ) but no association with FEV<sub>1</sub>/FVC or FVC compared to wild type (table 2). However, in stratified analyses of UK Biobank never- and ever-smokers, we found that heterozygosity for the Z allele was associated with a large increase in FEV<sub>1</sub> ( $\beta=19.36\text{ mL}$ ,  $p=9.21\times 10^{-4}$ ) and increased FEV<sub>1</sub>/FVC ( $\beta=0.0031$ ,  $p=1.22\times 10^{-5}$ ) in never-smokers, but not in ever-smokers (table 2). The FEV<sub>1</sub> association in never-smokers was only slightly attenuated by including a quadratic height term in the regression ( $\beta=18.85\text{ mL}$ ,  $p=1.23\times 10^{-3}$ ). Statistical tests of Z allele\*ever-smoking interactions showed interactions for FEV<sub>1</sub> ( $p=0.022$ ) and FEV<sub>1</sub>/FVC ( $p=1.06\times 10^{-4}$ ) (table S4). Furthermore, heterozygous ever-smokers exhibited a small increased risk of COPD (both moderate-severe COPD (OR 1.16,  $p=0.005$ ) and COPD GOLD 1-4 (OR 1.12,  $p=0.001$ )), whereas heterozygous never-smokers did not (table 2).

### *The Z allele is strongly associated with height*

Heterozygosity for the Z allele was strongly associated with height ( $\beta=1.02\text{ cm}$ ,  $p=3.91\times 10^{-68}$ ) (table 2) and, when lung function measures were not adjusted for height, the Z allele heterozygosity had much larger effect estimates for FEV<sub>1</sub> ( $\beta\pm\text{SE}=44.72\pm 4.97\text{ mL}$ ,  $p=2.20\times 10^{-19}$ ) and FVC ( $\beta\pm\text{SE}=61.06\pm 6.18\text{ mL}$ ,  $p=5.18\times 10^{-23}$ ) but not FEV<sub>1</sub>/FVC. These results suggest that carrying one copy of the Z allele confers an advantage primarily (but not entirely) driven by increased height. To investigate the developmental stage at which Z allele heterozygosity influences height, we tested association with height in the ALPSAC cohort at ages 8, 15 and 24 years. The findings were consistent with Z allele heterozygosity influencing height from adolescence and, at all the ages, directions of effect on height and lung function were consistent with those in UK Biobank (table S5).

### *Sex modifies the effect of Z allele heterozygosity on lung function*

We also assessed whether sex can modify the effect of the Z allele on lung function or height and found that heterozygosity for the Z allele was associated with higher height-adjusted FEV<sub>1</sub>/FVC in females



**FIGURE 1** Traits associated with the *SERPINA1 Z* allele in UK Biobank. Traits displayed are those with a false discovery rate (FDR) <0.01 from phenome-wide association study of 2411 phenotypes across 379 101 UK Biobank participants using the a) heterozygous model for the Z allele (wild-type versus heterozygous individuals) and b) recessive model for the Z allele (wild-type and heterozygous individuals versus individuals homozygous for the Z allele). HES: hospital episode statistics; FBC: full blood count; FVC: forced vital capacity; FEV<sub>1</sub>: forced expiratory volume in 1 s; PEF: peak expiratory flow; ENT: ear, nose and throat.



TABLE 1 Association between the *SERPINA1* Z allele and blood biomarkers in unrelated, European UK Biobank participants

Biomarker	Heterozygous genetic model		Recessive genetic model	
	$\beta \pm \text{SE}$	p-value	$\beta \pm \text{SE}$	p-value
Alanine aminotransferase U·L <sup>-1</sup>	1.163±0.112	2.95×10 <sup>-25</sup>	3.037±1.246	0.015
Albumin g·L <sup>-1</sup>	0.712±0.022	3.11×10 <sup>-229</sup>	0.306±0.243	0.208
Alkaline phosphatase U·L <sup>-1</sup>	2.968±0.213	5.81×10 <sup>-44</sup>	2.379±2.373	0.316
Apolipoprotein B g·L <sup>-1</sup>	-0.0037±0.0021	0.082	0.042±0.024	0.086
Apolipoprotein A g·L <sup>-1</sup>	-0.0045±0.0019	0.020	0.019±0.022	0.375
Aspartate aminotransferase U·L <sup>-1</sup>	0.707±0.086	2.27×10 <sup>-16</sup>	6.353±0.956	3.04×10 <sup>-11</sup>
C-reactive protein mg·L <sup>-1</sup>	-0.174±0.036	1.11×10 <sup>-6</sup>	0.192±0.396	0.628
Calcium mmol·L <sup>-1</sup>	0.012±0.00080	4.01×10 <sup>-50</sup>	0.01±0.009	0.265
Cholesterol mmol·L <sup>-1</sup>	0.0051±0.0092	0.576	0.19±0.102	0.063
Creatinine $\mu\text{mol}\cdot\text{L}^{-1}$	-0.067±0.129	0.601	-0.94±1.435	0.512
Cystatin C mg·L <sup>-1</sup>	0.0090±0.0013	8.91×10 <sup>-12</sup>	0.019±0.015	0.204
Direct bilirubin $\mu\text{mol}\cdot\text{L}^{-1}$	0.041±0.0073	1.64×10 <sup>-8</sup>	0.158±0.079	0.046
eGFR <sup>#</sup>	-0.241±0.157	0.125	-1.810±1.732	0.296
$\gamma$ -Glutamyltransferase U·L <sup>-1</sup>	0.253±0.342	0.460	8.045±3.818	0.035
Glucose mmol·L <sup>-1</sup>	-0.0032±0.010	0.758	-0.133±0.114	0.244
HbA1c mmol·mol <sup>-1</sup>	-0.187±0.052	3.40×10 <sup>-4</sup>	-1.781±0.588	2.46×10 <sup>-3</sup>
HDL-cholesterol mmol·L <sup>-1</sup>	0.0015±0.0030	0.605	0.142±0.033	2.01×10 <sup>-5</sup>
IGF-1 nmol·L <sup>-1</sup>	-0.212±0.045	2.35×10 <sup>-6</sup>	-2.245±0.5	7.19×10 <sup>-6</sup>
LDL direct mmol·L <sup>-1</sup>	0.011±0.007	0.113	0.115±0.079	0.145
Lipoprotein A nmol·L <sup>-1</sup>	-0.324±0.452	0.474	-0.151±5.115	0.976
Oestradiol pmol·L <sup>-1</sup>	9.577±8.441	0.257	88.051±96.144	0.360
Phosphate mmol·L <sup>-1</sup>	0.0027±0.0013	0.042	-0.014±0.015	0.352
Rheumatoid factor IU·mL <sup>-1</sup>	0.113±0.529	0.832	-3.27±5.7	0.566
SHBG nmol·L <sup>-1</sup>	4.531±0.218	1.06×10 <sup>-95</sup>	40.573±2.402	5.33×10 <sup>-64</sup>
Testosterone nmol·L <sup>-1</sup>	0.37±0.023	1.84×10 <sup>-57</sup>	2.986±0.258	6.96×10 <sup>-31</sup>
Total bilirubin $\mu\text{mol}\cdot\text{L}^{-1}$	0.146±0.035	3.54×10 <sup>-5</sup>	0.52±0.391	0.184
Total protein g·L <sup>-1</sup>	0.578±0.035	1.06×10 <sup>-62</sup>	0.459±0.382	0.229
Triglycerides mmol·L <sup>-1</sup>	-0.023±0.0082	4.69×10 <sup>-3</sup>	-0.324±0.092	4.21×10 <sup>-4</sup>
Urate $\mu\text{mol}\cdot\text{L}^{-1}$	1.096±0.56	0.050	-1.325±6.252	0.832
Urea mmol·L <sup>-1</sup>	0.051±0.011	2.69×10 <sup>-6</sup>	0.06±0.122	0.620
Vitamin D nmol·L <sup>-1</sup>	-0.041±0.175	0.816	-2.547±1.955	0.193

Results are based on linear regression adjusting for sex, age, age<sup>2</sup>, ancestry-based principal components and genotyping array. The heterozygous or recessive genetic model was also included in the regression as shown. eGFR: estimated glomerular filtration rate; HbA1c: glycated haemoglobin; HDL: high-density lipoprotein; IGF: insulin-like growth factor; LDL: low-density lipoprotein; SHBG: sex hormone-binding globulin. #: calculated using the following formula: if cystatin C (cys)≤0.8, then eGFR=133\*[(cys/0.8)\*\*-0.499]\*(0.996\*\*age)\*[0.932 if female]; whereas if cys>0.8, then eGFR=133\*[(cys/0.8)\*\*-1.328]\*(0.996\*\*age)\*[0.932 if female].

( $\beta \pm \text{SE} = 0.0028 \pm 0.00071$ ,  $p = 5.89 \times 10^{-5}$ ) but not in males ( $\beta \pm \text{SE} = -0.0015 \pm 0.0009$ ,  $p = 0.093$ ) (Z allele\*sex interaction:  $p = 0.00021$  for FEV<sub>1</sub>/FVC) (tables S4 and S6).

#### Association between Z allele homozygosity and lung function

In contrast to the lung function-raising effects of carrying one copy of the Z allele, homozygosity for the Z allele has been widely reported to reduce lung function and increase risk of COPD. Similarly, UK Biobank individuals homozygous for the Z allele had significantly lower FEV<sub>1</sub> ( $\beta = -249.21$  mL,  $p = 6.96 \times 10^{-7}$ ), FEV<sub>1</sub>/FVC ( $\beta = -0.051$ ,  $p = 3.54 \times 10^{-16}$ ) and FVC ( $\beta = -132.16$ ,  $p = 0.028$ ) compared to wild type and heterozygous individuals, and greater risk of COPD (both moderate-severe COPD (OR 7.42,  $p = 5.67 \times 10^{-14}$ ) and COPD GOLD 1-4 (OR 4.58,  $p = 8.67 \times 10^{-12}$ )) (table 2). The associations between homozygosity for the Z allele and reduced lung function and increased risk of COPD were stronger in smokers compared to nonsmokers (table 2). Given the strength of these associations, it is interesting to note that 58 out of 93 individuals homozygous for the Z allele did not have spirometrically defined COPD (FEV<sub>1</sub>/FVC <0.7) and four out of 15 homozygous ever-smokers aged >60 years did not have spirometrically defined COPD. We generated polygenic risk scores for lower FEV<sub>1</sub>/FVC by weighting variants according to  $\beta$ -coefficients in the SpiroMeta consortium cohorts [19]. The 58 individuals homozygous for the Z allele but without COPD have an average score percentile of 47.6 compared to 57.0

TABLE 2 Association between the *SERPINA1* Z allele and height, lung function traits, and COPD in unrelated, European UK Biobank participants

Genetic model	Smoking status	Test/comparison n	Height cm		FEV <sub>1</sub> mL		FEV <sub>1</sub> /FVC		FVC mL		COPD <sup>#</sup>		COPD GOLD 1–4 <sup>¶</sup>	
			β±SE	p-value	β±SE	p-value	β±SE	p-value	β±SE	p-value	OR (95% CI)	p-value	OR (95% CI)	p-value
<b>Het</b>	All	11877/291383	1.023±0.059	3.91×10 <sup>-68</sup>	9.26±4.53	0.041	0.00091±0.00057	0.108	9.38±5.43	0.084	1.05 (0.96–1.14)	0.268	1.04 (0.98–1.09)	0.171
	Ever-smokers	5357/133779	1.027±0.087	8.53×10 <sup>-32</sup>	-2.63±7.05	0.709	-0.0017±0.00091	0.062	5.47±8.28	0.509	1.16 (1.05–1.29)	0.005	1.12 (1.04–1.20)	0.001
	Never-smokers	6520/157604	1.022±0.079	3.03×10 <sup>-38</sup>	19.36±5.84	9.21×10 <sup>-4</sup>	0.0031±0.00071	1.22×10 <sup>-5</sup>	12.77±7.15	0.074	0.87 (0.75–1.01)	0.070	0.93 (0.86–1.02)	0.113
<b>Rec</b>	All	93/303260	1.644±0.650	0.011	-249.21±50.22	6.96×10 <sup>-7</sup>	-0.051±0.0063	3.54×10 <sup>-16</sup>	-132.16±60.09	0.028	7.42 (4.29–12.27)	5.67×10 <sup>-14</sup>	4.58 (2.93–7.05)	8.67×10 <sup>-12</sup>
	Ever-smokers	35/139136	1.242±1.062	0.242	-379.21±85.54	9.30×10 <sup>-6</sup>	-0.081±0.011	2.00×10 <sup>-13</sup>	-177.95±100.35	0.076	9.19 (4.19–19.00)	6.27×10 <sup>-9</sup>	6.13 (3.07–12.29)	2.30×10 <sup>-7</sup>
	Never-smokers	58/164124	1.873±0.821	0.023	-174.68±60.68	0.004	-0.034±0.0074	4.95×10 <sup>-6</sup>	-107.69±74.23	0.147	5.98 (2.66–12.03)	2.69×10 <sup>-6</sup>	3.68 (2.01–6.47)	1.11×10 <sup>-5</sup>

Results are based on linear (or logistic, in the case of COPD) regression adjusting for sex, age, age<sup>2</sup>, ancestry-based principal components, genotyping array and, in the case of lung function and COPD, standing height. The heterozygote (Het) or recessive (Rec) genetic model was also included in the regression as shown. FEV<sub>1</sub>: forced expiratory volume in 1 s; FVC: forced vital capacity; GOLD: Global Initiative for Chronic Obstructive Lung Disease. <sup>#</sup>: spirometrically defined as FEV<sub>1</sub>/FVC <0.7 and FEV<sub>1</sub><80% predicted; <sup>¶</sup>: defined as FEV<sub>1</sub>/FVC <0.7.

in those with COPD. This suggests that individuals homozygous for the Z allele but without COPD may have a more protective genetic profile across other genomic loci.

#### *Diagnosis of AATD in individuals homozygous for the Z allele*

We note that only 20 out of 141 individuals who were genotyped ZZ homozygotes in the full UK Biobank cohort had a recorded ICD-10 coding for AATD (E880) in HES. As this could indicate miscoding rather than misdiagnosis, we investigated further in a subset of 65 homozygotes with linked primary care data. Of these 65 ZZ homozygotes, 15 had primary care (Read) codes or ICD-10 codes for AATD and 50 did not. Of the group without AATD codes, four had Read codes or ICD-10 codes for bronchiectasis and an additional six had primary care (Read) codes or ICD-10 codes for COPD, indicating potential misdiagnosis of AATD.

## Discussion

We describe novel phenotypic associations of heterozygosity for the Z allele of *SERPINA1* from a PheWAS of UK Biobank, and present definitive evidence that heterozygosity for the Z allele is associated with greater FEV<sub>1</sub> and FEV<sub>1</sub>/FVC only in nonsmokers. Our findings have implications for individuals heterozygous for the *SERPINA1* Z allele, and more generally for the study of rare variants and gene–environment interactions.

Our PheWAS examining the consequences of carrying one copy of the *SERPINA1* Z allele revealed potentially important associations with nonrespiratory traits and diseases. Consistent with previous reports [12], we detected a very strong association between Z allele heterozygosity and height (with each additional Z allele adding ~1 cm to height). The mechanism and timing of the effect of the Z allele on height is not understood. Our findings in the ALSPAC study suggest a possible influence of Z allele heterozygosity on height that is manifest from puberty, although results should be interpreted with caution given the limitations in power in ALSPAC alone. This suggests that sex hormones are a possible mediator of the height-raising effects of the Z allele and, indeed, we detected higher levels of testosterone in individuals heterozygous for the Z allele. However, heterozygosity for the Z allele is associated with increased height in males and females, whereas increased testosterone was only seen in males. Furthermore, Z allele heterozygotes also exhibit higher SHBG, suggesting that the amount of free testosterone will be unchanged. IGF-1 levels relate to height but in our study, the magnitude of the association was very small (<1% average levels) and it was a negative association. It is therefore still not clear why heterozygosity and homozygosity for the Z allele is associated with increased height.

We detected strong associations between heterozygosity for the Z allele and increased risk of gall stones and gall bladder removal. This is consistent with recent reports that the Z allele may be a risk factor for developing gallstone disease [21, 22]. The mechanism underlying this association is unclear but previous reports have suggested it may be related to liver dysfunction in AATD and/or the composition of bile [21, 22]. We found that individuals heterozygous for the Z allele also had increased levels of ALP, which is characteristic of liver disease and biliary tract obstruction due to gall stones. Moreover, the Z allele is associated with elevation of ALT and AST. Higher than normal levels of ALT and AST are hallmarks of liver disease, and are also observed in some patients with gall bladder disease [23]. Higher bilirubin, another potential sign of liver disease, is also exhibited by Z allele carriers. Interestingly, excess bilirubin is thought to contribute to gall stone formation [24], so a causal pathway between the Z allele, liver dysfunction and gall bladder problems is possible.

Heterozygosity for the Z allele was also associated with musculoskeletal phenotypes such as reduced risk of osteoarthritis, increased risk of osteoporosis and lower bone mineral density. Calcium was increased in heterozygotes but, as albumin was also increased, free calcium may not be different in this group. A recent study showed that  $\alpha_1$ -antitrypsin (AAT) can inhibit receptor activator of NF- $\kappa$ B ligand (RANKL)-induced osteoclast formation and bone resorption [25]. As heterozygosity for the Z allele causes reduction in AAT, individuals heterozygous for the Z allele may have increased risk of osteoporosis and lower bone mineral density due to increased RANKL-induced bone resorption. Lower bone mineral density has been associated with a lower risk of osteoarthritis [26, 27].

Heterozygotes for the Z allele also had reduced risk of heart disease (even when adjusted for height) and lower blood pressure, as well as increased risk of headache, consistent with one previous report suggesting cluster headaches may be more frequent in heterozygotes for the Z allele [6]. There was also a marked association between heterozygosity and reduced reticulocyte count and reticulocyte percentage, which, as far as we are aware, has not been reported previously.

Homozygosity for the Z allele was strongly associated with self-reported/HES-recorded AATD and respiratory traits but also haemoglobin concentration and haematocrit percentage, which may be driven by



the associations with respiratory disease. Homozygosity was also associated with markedly higher risk of optic neuritis and pancreatitis. These associations should be interpreted with care, especially the association with optic neuritis, as small case numbers may have resulted in poor estimation of effect sizes.

While homozygosity for the Z allele is an established cause of respiratory disease, the effect of heterozygosity on the respiratory health of general populations has been less clear. Our study provides definitive data on the effects of heterozygosity for the Z allele on respiratory traits, showing that heterozygosity is associated with greater lung function in never-smokers (even after height adjustment) but not in smokers. Our results contrast with previous studies [8–11] that did not detect higher lung function in participants heterozygous for the Z allele compared to wild type and with studies that did not detect differential effects by smoking status [12]. There are several possible reasons for these discrepancies. First, the relatively low frequency of the Z allele meant previous studies were underpowered. Our study contained 18 times more heterozygous participants than the previous largest population-based study and 46 times more than the largest case–control study. Whilst one previous study has suggested a beneficial effect of Z allele heterozygosity on lung function [12], it nevertheless did not detect differential effects by smoking status. Again, this is probably due to power. Second, most of the previous studies were conducted in current and former smokers only [8, 9, 11], and would therefore have missed the effect of Z allele heterozygosity in nonsmokers. Third, most previous studies recruited participants based on their health status (as in case–control studies [8, 9]) or the health status of a family member [10, 11, 28]. These approaches can lead to ascertainment biases that might distort causal estimates. A major strength of using the UK Biobank for our study was that participants were not ascertained based on having lung disease, enabling a relatively unbiased assessment of the role of the heterozygosity for the Z allele in lung function. It should be noted, however, that UK Biobank participants tend to be healthier and more highly educated than the UK population as a whole, so detrimental effects of the Z allele may be less pronounced in UK Biobank due to fewer lifestyle risk factors. Fourth, the Z allele is not present on most standard genotyping arrays and therefore, high-quality imputation is required to detect its effects. Furthermore, genetic studies tend to focus on the additive genetic model rather than heterozygous or recessive models. Because we designed the Z allele into the UK Biobank array [17] and compared different genetic models, we were able to robustly examine this association in UK Biobank participants.

The biological mechanism underlying the association between heterozygosity for the Z allele and increased lung function in nonsmokers is not clear. The *SERPINA1* gene encodes AAT, a glycoprotein with an important role in inhibiting proteases, such as neutrophil elastase, which are secreted during inflammation and cause collateral tissue damage. The Z allele produces AAT with reduced antiprotease activity and a propensity to form polymers, which accumulate in liver, thereby reducing the circulating levels of AAT. The polymers themselves have also been shown to be a proinflammatory stimulus in lung tissue [29], which further perpetuates lung inflammation and damage. Homozygosity for the Z allele is the most common cause of severe AATD, whereas heterozygosity for the Z allele has been shown to result in intermediate levels of AAT (~60% of normal) [30]. One might therefore expect heterozygosity to have, if anything, a detrimental effect on lung function. There is evidence that positive selection has acted at the Z allele locus [12, 31] and mechanisms of heterozygous advantage (through increased height [12] and protection from infectious respiratory diseases by promotion of inflammatory responses [31]) have been proposed. The amplification of the inflammatory response to cigarette smoke may therefore provide a mechanism for eliminating this advantage in smokers. It has also been shown that smoking-induced oxidative modifications to AAT reduce its ability to inhibit neutrophil elastase and turn it into a proinflammatory mediator [32]. Cigarette smoke has been found to accelerate polymerisation of Z-type AAT [33], suggesting a possible mechanism whereby smoking may modify the effect of the Z allele on lung function.

We also report a novel interaction between sex and the *SERPINA1* Z allele heterozygosity on FEV<sub>1</sub>/FVC. Women heterozygous for the Z allele exhibited greater FEV<sub>1</sub>/FVC compared to wild-type, whereas men heterozygous for the Z allele showed no significant difference to wild type. Previous studies have shown that individuals with the fastest decline in lung function amongst AATD patients are more likely to be male [15]. However, the reasons why sex would appear to modify the effect of Z allele heterozygosity on lung function are not clear.

Finally, we confirm that homozygosity for the Z allele is associated with reduced lung function and increased risk of COPD, and that these effects are exacerbated by smoking. These results are consistent with a recently published study in UK Biobank focused on homozygosity for the Z allele [34]. Also consistent with the work of NAKANISHI *et al.* [34], we show that only a minority of individuals with the ZZ genotype have been diagnosed with AATD (as has been reported elsewhere [35, 36]), and that those without a diagnosis of AATD, bronchiectasis or COPD have a protective genetic profile at other known lung function loci.

In conclusion, we have demonstrated that heterozygosity for the Z allele is associated with previously unrecognised nonrespiratory phenotypes. Further studies will be required to confirm these findings in independent populations; produce accurate effect estimates; and understand the causal pathways involved, and the potential role of genetic and environmental modifiers. We have also shown that heterozygosity for the Z allele is associated with greater lung function in nonsmokers. Our results demonstrate that large sample sizes are required to study the effects of rare variants, particularly when these effects depend on an environmental stimulus. The opposing directions of effect of Z allele heterozygosity and homozygosity on lung function demonstrate that the additive genetic model is not always appropriate for genetic studies of rarer variants such as the *SERPINA1* Z allele. Our findings suggest that while individuals who are heterozygous for the Z allele may not exhibit symptoms of AATD, they may be more susceptible to smoking-induced lung disease and have altered risk of other nonrespiratory conditions.

**Acknowledgements:** The UK Biobank analysis was conducted under approved UK Biobank data application number 648 and application number 20361. We thank UK Biobank and all the participants for generating this important health research resource. This study used the ALICE and SPECTRE High Performance Computing Facilities at the University of Leicester. We are extremely grateful to all the families who took part in the Avon Longitudinal Study of Parents and Children study, the midwives for their help in recruiting them and the whole ALSPAC team, which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists, and nurses.

**Conflict of interest:** K.A. Fawcett reports grants from Asthma UK outside the submitted work. K. Song has nothing to disclose. G. Qian has nothing to disclose. A-E. Farmaki has nothing to disclose. R. Packer reports grants from Orion Pharma outside the submitted work. C. John reports grants from the Medical Research Council during the conduct of the study. N. Shrine has nothing to disclose. R. Granell has nothing to disclose. S. Ring has nothing to disclose. N.J. Timpson has nothing to disclose. L.M. Yerges-Armstrong is an employee/stockholder of GlaxoSmithKline. R. Eastell reports grants from Amgen and Alexion, grants and personal fees from IDS, Roche and Nittobo, and personal fees from GSK Nutrition, Mereo, Sandoz, AbbVie, Samsung, Haoma Medica, Elsevier, CL Bio, FNIH, Viking, UCSF, Biocon and Lyramid, outside the submitted work. L.V. Wain reports grants from GSK outside the submitted work. R.A. Scott is an employee and shareholder of GSK. M.D. Tobin reports grants from GSK and Orion Pharma outside the submitted work. I.P. Hall reports grants from GSK, Boehringer Ingelheim and Orion, outside the submitted work.

**Support statement:** The data for this study were partially generated by Medical Research Council (MRC) strategic award to M.D. Tobin, I.P. Hall and L.V. Wain (MC\_PC\_12010). The research was partially supported by the NIHR Nottingham Biomedical Research Centre; the views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Dept of Health. M.D. Tobin is supported by a Wellcome Trust Investigator Award (WT202849/Z/16/Z). M.D. Tobin and L.V. Wain have been supported by the MRC (MR/N011317/1). L.V. Wain holds a GSK/British Lung Foundation Chair in Respiratory Research. K.A. Fawcett holds an Asthma UK fellowship award. I.P. Hall holds a NIHR Senior Investigator award. C. John holds a Medical Research Council Clinical Research Training Fellowship (MR/P00167X/1). We would like to acknowledge the support of the Health Data Research UK BREATHE Digital Innovation Hub funded by URKI (MC\_PC\_19004). The UK Medical Research Council and Wellcome (grant ref. 102215/2/13/2) and the University of Bristol provide core support for the Avon Longitudinal Study of Parents and Children (ALSPAC). This publication is the work of the authors and R. Granell will serve as guarantor for the contents of this paper. A comprehensive list of grants funding is available on the ALSPAC website ([www.bristol.ac.uk/alspac/external/documents/grant-acknowledgements.pdf](http://www.bristol.ac.uk/alspac/external/documents/grant-acknowledgements.pdf)). N.J. Timpson is a Wellcome Trust Investigator (202802/Z/16/Z), is the principal investigator of ALSPAC (MRC and WT 217065/Z/19/Z), is supported by the University of Bristol NIHR Biomedical Research Centre (BRC-1215-2001) and the MRC Integrative Epidemiology Unit (MC\_UU\_00011), and works within the CRUK Integrative Cancer Epidemiology Programme (C18281/A19169). Funding information for this article has been deposited with the Crossref Funder Registry.

## References

- 1 Silverman EK. Risk of lung disease in PI MZ heterozygotes. Current status and future research directions. *Ann Am Thorac Soc* 2016; 13: Suppl. 4, S341–S345.
- 2 Boelle PY, Debray D, Guillot L, *et al.* *SERPINA1* Z allele is associated with cystic fibrosis liver disease. *Genet Med* 2019; 21: 2151–2155.
- 3 Curjuric I, Imboden M, Bettschart R, *et al.* Alpha-1 antitrypsin deficiency: from the lung to the heart? *Atherosclerosis* 2018; 270: 166–172.
- 4 Hadzik-Blaszczyk M, Zdral A, Zielonka TM, *et al.* *SERPINA1* gene variants in granulomatosis with polyangiitis. *Adv Exp Med Biol* 2018; 1070: 9–18.
- 5 Schneider CV, Hamesch K, Gross A, *et al.* Liver phenotypes of European adults heterozygous or homozygous for Pi\*Z variant of AAT (Pi\*MZ vs Pi\*ZZ genotype) and non-carriers. *Gastroenterology* 2020; 159: 534–548.
- 6 Summ O, Gregor N, Marziniak M, *et al.* Cluster headache and alpha 1-antitrypsin deficiency. *Cephalalgia* 2010; 30: 113–117.
- 7 Pendergrass SA, Brown-Gentry K, Dudek SM, *et al.* The use of phenome-wide association studies (PheWAS) for exploration of novel genotype-phenotype relationships and pleiotropy discovery. *Genet Epidemiol* 2011; 35: 410–422.
- 8 Foreman MG, Wilson C, DeMeo DL, *et al.* Alpha-1 antitrypsin PiMZ genotype is associated with chronic obstructive pulmonary disease in two racial groups. *Ann Am Thorac Soc* 2017; 14: 1280–1287.
- 9 Li X, Ortega VE, Ampleford EJ, *et al.* Genome-wide association study of lung function and clinical implication in heavy smokers. *BMC Med Genet* 2018; 19: 134.
- 10 Molloy K, Hersh CP, Morris VB, *et al.* Clarification of the risk of chronic obstructive pulmonary disease in alpha1-antitrypsin deficiency PiMZ heterozygotes. *Am J Respir Crit Care Med* 2014; 189: 419–427.

- 11 Sorheim IC, Bakke P, Gulsvik A, *et al.*  $\alpha_1$ -Antitrypsin protease inhibitor MZ heterozygosity is associated with airflow obstruction in two large cohorts. *Chest* 2010; 138: 1125–1132.
- 12 North TL, Ben-Shlomo Y, Cooper C, *et al.* A study of common Mendelian disease carriers across ageing British cohorts: meta-analyses reveal heterozygosity for alpha 1-antitrypsin deficiency increases respiratory capacity and height. *J Med Genet* 2016; 53: 280–288.
- 13 Bowden J, Thompson JR, Burton PR. A two-stage approach to the correction of ascertainment bias in complex genetic studies involving variance components. *Ann Hum Genet* 2007; 71: 220–229.
- 14 Thun GA, Ferrarotti I, Imboden M, *et al.* SERPINA1 PiZ and PiS heterozygotes and lung function decline in the SAPALDIA cohort. *PLoS One* 2012; 7: e42728.
- 15 Dawkins PA, Dawkins CL, Wood AM, *et al.* Rate of progression of lung function impairment in  $\alpha_1$ -antitrypsin deficiency. *Eur Respir J* 2009; 33: 1338–1344.
- 16 Demeo DL, Sandhaus RA, Barker AF, *et al.* Determinants of airflow obstruction in severe alpha-1-antitrypsin deficiency. *Thorax* 2007; 62: 806–813.
- 17 Wain LV, Shrine N, Miller S, *et al.* Novel insights into the genetics of smoking behaviour, lung function, and chronic obstructive pulmonary disease (UK BiLEVE): a genetic association study in UK Biobank. *Lancet Respir Med* 2015; 3: 769–781.
- 18 Bycroft C, Freeman C, Petkova D, *et al.* Genome-wide genetic data on ~500 000 UK Biobank participants. *bioRxiv* 2017; preprint [https://doi.org/10.1101/166298].
- 19 Shrine N, Guyatt AL, Erzurumluoglu AM, *et al.* New genetic signals for lung function highlight pathways and chronic obstructive pulmonary disease associations across multiple ancestries. *Nat Genet* 2019; 51: 481–493.
- 20 Verma A, Bradford Y, Dudek S, *et al.* A simulation study investigating power estimates in phenome-wide association studies. *BMC Bioinform* 2018; 19: 120.
- 21 Ferkingstad E, Oddsson A, Gretarsdottir S, *et al.* Genome-wide association meta-analysis yields 20 loci associated with gallstone disease. *Nat Commun* 2018; 9: 5101.
- 22 Gellert-Kristensen H, Dalila N, Nielsen SF, *et al.* Identification and replication of six new loci associated with gallstone disease. *Hepatology* 2019; 70: 597–609.
- 23 Tetangco EP, Shah N, Arshad HM, *et al.* Markedly elevated liver enzymes in choledocholithiasis in the absence of hepatocellular disease: case series and literature review. *J Investig Med High Impact Case Rep* 2016; 4: 2324709616651092.
- 24 Stender S, Frikke-Schmidt R, Nordestgaard BG, *et al.* Extreme bilirubin levels as a causal risk factor for symptomatic gallstone disease. *JAMA Intern Med* 2013; 173: 1222–1228.
- 25 Akbar MA, Nardo D, Chen MJ, *et al.* Alpha-1 antitrypsin inhibits RANKL-induced osteoclast formation and functions. *Mol Med* 2017; 23: 57–69.
- 26 Hackinger S, Trajanoska K, Styrkarsdottir U, *et al.* Evaluation of shared genetic aetiology between osteoarthritis and bone mineral density identifies SMAD3 as a novel osteoarthritis risk locus. *Hum Mol Genet* 2017; 26: 3850–3858.
- 27 Johansson H, Hongslo Vala C, Oden A, *et al.* Low risk for hip fracture and high risk for hip arthroplasty due to osteoarthritis among Swedish farmers. *Osteoporos Int* 2018; 29: 741–749.
- 28 Silverman EK, Province MA, Campbell EJ, *et al.* Family study of alpha 1-antitrypsin deficiency: effects of cigarette smoking, measured genotype, and their interaction on pulmonary function and biochemical traits. *Genet Epidemiol* 1992; 9: 317–331.
- 29 Mulgrew AT, Taggart CC, Lawless MW, *et al.* Z alpha1-antitrypsin polymerizes in the lung and acts as a neutrophil chemoattractant. *Chest* 2004; 125: 1952–1957.
- 30 Ferrarotti I, Thun GA, Zorzetto M, *et al.* Serum levels and genotype distribution of  $\alpha_1$ -antitrypsin in the general population. *Thorax* 2012; 67: 669–674.
- 31 Lomas DA. The selective advantage of  $\alpha_1$ -antitrypsin deficiency. *Am J Respir Crit Care Med* 2006; 173: 1072–1077.
- 32 Taggart C, Cervantes-Laurean D, Kim G, *et al.* Oxidation of either methionine 351 or methionine 358 in alpha 1-antitrypsin causes loss of anti-neutrophil elastase activity. *J Biol Chem* 2000; 275: 27258–27265.
- 33 Alam S, Li Z, Janciauskiene S, *et al.* Oxidation of Z  $\alpha_1$ -antitrypsin by cigarette smoke induces polymerization: a novel mechanism of early-onset emphysema. *Am J Respir Cell Mol Biol* 2011; 45: 261–269.
- 34 Nakanishi T, Forgetta V, Handa T, *et al.* The undiagnosed disease burden associated with alpha-1 antitrypsin deficiency genotypes. *Eur Respir J* 2020; 56: 2001441.
- 35 Silverman EK, Miletich JP, Pierce JA, *et al.* Alpha-1-antitrypsin deficiency. High prevalence in the St. Louis area determined by direct population screening. *Am Rev Respir Dis* 1989; 140: 961–966.
- 36 Silverman EK, Pierce JA, Province MA, *et al.* Variability of pulmonary function in alpha-1-antitrypsin deficiency: clinical correlates. *Ann Intern Med* 1989; 111: 982–991.