ORIGINAL ARTICLE

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

Teri-Louise North,[1] Yoav Ben-Shlomo,[1] Cyrus Cooper,[2,3,4] Ian J Deary,[5,6] John Gallacher,[7] Mika Kivimaki,[8] Meena Kumari,[8,9] Richard M Martin,[1,10] Alison Pattie,[5] Avan Aihie Sayer,[2] John M Starr,[6] Andrew Wong,[11] Diana Kuh,[11] Santiago Rodriguez,[1] Ian N M Day[1]

▸ Additional material is published online only. To view please visit the journal online (http://dx.doi.org/10.1136/jmedgenet-2015-103342).

For numbered affiliations see end of article.

Correspondence to
Professor Ian N M Day, School of Social and Community Medicine, University of Bristol, Oakfield House, Oakfield Grove, Bristol BS8 2BN, UK; ian.day@bristol.ac.uk

Received 26 June 2015
Revised 4 December 2015
Accepted 6 December 2015
Published Online First
1 February 2016

To cite: North T-L, Ben-Shlomo Y, Cooper C, *et al*. *J Med Genet* 2016;**53**: 280–288.

ABSTRACT

Background Several recessive Mendelian disorders are common in Europeans, including cystic fibrosis (*CFTR*), medium-chain-acyl-Co-A-dehydrogenase deficiency (*ACADM*), phenylketonuria (*PAH*) and alpha 1-antitrypsin deficiency (*SERPINA1*).

Methods In a multicohort study of >19 000 older individuals, we investigated the relevant phenotypes in heterozygotes for these genes: lung function (forced expiratory volume in 1 second (FEV1), forced vital capacity (FVC)) for *CFTR* and *SERPINA1*; cognitive measures for *ACADM* and *PAH*; and physical capability for *ACADM*, *PAH* and *SERPINA1*.

Results Findings were mostly negative but lung function in *SERPINA1* (protease inhibitor (PI) Z allele, rs28929474) showed enhanced FEV1 and FVC (0.13 z-score increase in FEV1 ($p=1.7\times10^{-5}$) and 0.16 z-score increase in FVC ($p=5.2\times10^{-8}$)) in PI-MZ individuals. Height adjustment (a known, strong correlate of FEV1 and FVC) revealed strong positive height associations of the Z allele (1.50 cm increase in height ($p=3.6\times10^{-10}$)).

Conclusions The PI-MZ rare (2%) SNP effect is nearly four times greater than the 'top' common height SNP in *HMGA2*. However, height only partially attenuates the *SERPINA1*-FEV1 or FVC association (around 50%) and vice versa. Height SNP variants have recently been shown to be positively selected collectively in North versus South Europeans, while the Z allele high frequency is localised to North Europe. Although PI-ZZ is clinically disadvantageous to lung function, PI-MZ increases both height and respiratory function; potentially a balanced polymorphism. Partial blockade of PI could conceivably form part of a future poly-therapeutic approach in very short children. The notion that elastase inhibition should benefit patients with chronic obstructive pulmonary disease may also merit re-evaluation. PI is already a therapeutic target: our findings invite a reconsideration of the optimum level in respiratory care and novel pathway potential for development of agents for the management of growth disorders.

INTRODUCTION

Heterozygote carriers for recessive Mendelian (monogenic) disorders such as cystic fibrosis (MIM: 219700), medium-chain-acyl-Co-A-dehydrogenase deficiency (MIM: 201450), phenylketonuria (MIM: 261600) and alpha 1-antitrypsin (AAT) deficiency (MIM: 613490) are relatively common in the UK population (1.5% (*ACADM*) to ~10% (protease inhibitor (PI)-MS)). Unlike in homozygote carriers, no clinical features are evident in heterozygotes although biochemical phenotype may be detectable (eg, phenylalanine level after aspartame[1] (*PAH*)). The AAT deficiency phenotype is continuous across the six genotypes of the S and Z alleles (MM wildtype, MS, MZ, SS, SZ and ZZ). However, only individuals of ZZ genotype are notable clinically; the condition results in early-onset lung emphysema with a penetrance of 60% for ZZ individuals.[2] There is no clear association in the literature between lung disease and individuals with either MZ[3] or SZ genotype.[4]

The prefix PI is added to the allele or genotype name. According to this, the normal (most common) allele is PI-M and the most common pathogenic allele is PI-Z. Mendelian disease alleles such as PI-Z may be prevalent in a population through new mutation and chance, with insufficient time for fitness and selection to take effect, or through balancing selection where heterozygote advantage outweighs homozygote disadvantage. A textbook example of the latter is sickle cell anaemia where resistance to malaria confers a heterozygote advantage.

Within the Healthy Ageing across the Life Course (HALCyon) collaboration[5 6] of UK observational cohorts, we tested whether heterozygote carriers for these four Mendelian diseases exhibit phenotypic differences from non-carriers in later life. In eight studies, we genotyped the deltaF508 mutation for cystic fibrosis, the K340E mutation for medium-chain-acyl-Co-A-dehydrogenase deficiency, the three most common phenylketonuria mutations in the UK (rs5030861, rs5030858 and rs75193786 (T to C mutation)) and lastly rs28929474 and rs17580 representing PI-Z and PI-S alleles, respectively, to infer AAT PI genotypes. Lung function, cognitive capability and physical capability are complex traits that have each been shown to predict mortality.[7–9] For homozygotes or

compound heterozygotes of these four Mendelian diseases, large differences in earlier life are seen for lung function (*CFTR*, *SERPINA1*) and cognitive function (*ACADM*, *PAH*). We tested heterozygotes against equivalent later life traits accordingly, with an additional analysis of physical capability (*ACADM*, *PAH* and *SERPINA1*[10]). To generate estimates using all of the individual participant data (IPD), we pooled IPD into a single data set and conducted one-step meta-analyses of the harmonised outcomes. This is superior to a conventional two-step approach (analyses performed within each cohort and study-specific estimates pooled in a meta-analysis) when the exposure is rare.[11 12]

A well-known signature of recent selection in humans is the very fast increase in frequency of the favoured allele (or haplotype) in a population.[13] Two haplotype-based tests can detect it: the extended haplotype homozygosity (EHH) test[14] and the integrated test iHS.[13] Rare haplotypes are also informative. It has been suggested[15] that reduced decay of EHH of haplotypes that are both rare and extended is informative to identify signatures of natural selection. These signatures could reflect either residual levels of an older selection phenomenon that is being diluted or an active process of natural selection.[15] We performed an EHH analysis of rs28929474 and rs17580 using genome-wide association study (GWAS) data and PI genotype status in a UK cohort, ALSPAC.[16] We also tested selection related to common variation around *SERPINA1* from Haplotter[13] and estimated allele age based both on allele frequency[17] and on local recombination between the Z locus and other SNPs in the ALSPAC data.

MATERIALS AND METHODS

A list of acronyms used in this article is shown in table 1.

HALCyon

Study participants

Individuals included in this analysis belonged to the HALCyon collaboration.[5] We meta-analysed IPD from eight UK cohorts: the Boyd Orr Cohort, the Caerphilly Prospective Study (CaPS), the English Longitudinal Study of Ageing (ELSA), the Hertfordshire Ageing Study (HAS), the Hertfordshire Cohort Study, the Lothian Birth Cohort 1921 (LBC1921), the MRC National Survey of Health and Development (NSHD) and the Whitehall II Study (WHII). Further information about the HALCyon cohorts can be found in earlier publications.[18]

Table 1 List of acronyms

Gene acronyms	
ACADM	Acyl-CoA dehydrogenase, C-4 to C-12 straight chain
CFTR	Cystic fibrosis transmembrane conductance regulator
HMGA2	High mobility group AT-hook 2
PAH	Phenylalanine hydroxylase
SERPINA1	Serpin peptidase inhibitor, clade A (alpha-1 antiproteinase, antitrypsin), member 1
Outcome acronyms	
FEV1	Forced expiratory volume in 1 second
FVC	Forced vital capacity
FCRT	Four choice reaction time
Disease acronyms	
PKU	Phenylketonuria
MCADD	Medium-chain-acyl-Co-A-dehydrogenase deficiency
Cohort acronyms	
BO	Boyd Orr
CaPS	Caerphilly Prospective Study
ELSA	English Longitudinal Study of Ageing
LBC1921	Lothian Birth Cohort 1921
HAS	Hertfordshire Ageing Study
HALCyon	Healthy Ageing across the Life Course
HCS	Hertfordshire Cohort Study
NSHD	MRC National Survey of Health and Development
WHII	Whitehall II Study

Mutation selection

We selected the most common causal mutation to genotype for medium-chain acyl Co-A dehydrogenase deficiency (rs77931234, otherwise known as K304E or c.985A>G[19]) and cystic fibrosis (the deltaF508 mutation, rs113993960).

With the exception of the NSHD cohort, we inferred AAT PI status using the genotypes from rs28929474 and rs17580. PI-MM corresponds to an individual who is wildtype for both rs28929474 and rs17580. PI-MS individuals are wildtype for rs28929474 and heterozygous for rs17580, while PI-MZ individuals are the converse. PI-SS individuals are homozygous for rs17580 and wildtype for rs28929474, while PI-SZ individuals are heterozygous for both SNPs. PI-ZZ individuals are wildtype for rs17580 and homozygous for rs28929474. Due to their rarity, age and very close recombination distance, other genotypic combinations of rs28929474 and rs17580 would be vanishingly rare. In the NSHD, we analysed PI status measured from isoelectric focusing.[20]

Mutation selection was more complex for phenylketonuria because several hundred causal mutations have been identified to date. We selected rs5030861 (IVS12+1 G>A), rs5030858 (R408W) and rs75193786 [T to C mutation] (I65T) after consulting a review of PKU mutations in Europe[21] and the PAH database[22] (http://www.pahdb.mcgill.ca) and considering mutations with highest frequency in UK populations.

Genotyping

Genotyping was performed by LGC Genomics (http://www.lgcgenomics.com/), with the exception of rs17580 and rs28929474 in ELSA and WHII for which genotype data were already available. We inferred rs17580 and rs28929474 genotypes in the NSHD using PI classes from isoelectric focusing.[20] Further information on the genotyping quality is provided in online supplementary table S1.

Harmonisation of outcomes and exposures by cohort

Wave of outcome assessment is detailed in online supplementary appendix S2. All core continuous outcomes (lung function, cognitive capability and physical capability) were transformed to z-scores by subtracting the mean and dividing by the SD of the measure within cohorts using all data available. All outcomes were further harmonised across cohorts before z-scoring, as detailed in online supplementary appendix S3.

Chronic obstructive pulmonary disease (COPD) status was determined using the Global Lungs Initiative ERS Task Force 2012 regression equations, which derive the lower limit of normal (LLN, 5th centile) values for forced expiratory volume in 1 second (FEV1) and FEV1/forced vital capacity (FVC) ratio given an individual's age, sex and height.[23] These specify that age should be to at least one decimal place. This was not possible in ELSA, and thus, this may have introduced some error into the prediction equation. In addition, COPD status is derived in this analysis based on absolute FEV1 and FVC values rather than standardised values. Recent studies[24] have confirmed that different apparatus are likely to result in systematic differences in lung function readings, which our categorisation of

cases and non-cases for COPD has not taken into account. An individual was classed as having COPD if their FEV1/FVC ratio and their FEV1 were below the sex, height and age-specific LLN. This identified approximately 8% of individuals as having COPD, which indicated false positives as we would expect 5%.

Carrier status was defined as a binary variable in all analyses and was coded as [0] non-carrier and [1] carrier. The three *PAH* mutations were combined so that a non-carrier was homozygous for all three SNPs and a carrier was heterozygous for at least one SNP. In the analysis of PI status, separate analyses were conducted for PI-MS, PI-MZ, PI-SS, PI-SZ and PI-ZZ versus PI-MM (with PI-MM coded as 0).

Several of the outcomes were transformed prior to z-scoring to improve the normality of the residual distributions. Four choice reaction time in CaPS was inverse transformed, search speed was natural log transformed (NSHD and ELSA) and Mill Hill was squared in WHII.

Analyses of FVC were repeated with a square-root transformation and of FEV1/FVC ratio with a cube transformation. Analyses of weight and body mass index (BMI) were repeated with a natural log transformation, although these anthropometric outcomes were not z-scored.

Prior to analysis, individuals of non-European ancestry (self-reported or detected from genome-wide data) and related individuals were removed from the data set.

Statistical analyses

All analyses were conducted using Stata v.13.1[25] and basic covariates were age in years and sex. Analyses considering additional covariates or conducted within strata were restricted to individuals with these covariates/information available.

The analysis of lung function by AAT PI status tested for a linear association between binary PI status (PI-MS, MZ, SS, SZ, ZZ vs PI-MM) and (1) FEV1, (2) FVC and (3) FEV1/FVC ratio. Analyses were repeated in current, ex and never smokers and in individuals classified as having COPD. Associations in all individuals were repeated with adjustment for (1) height and height-squared and (2) height, height-squared and height-cubed. Associations in COPD cases were also repeated with simultaneous adjustment for height, height-squared and smoking status. The analysis of physical capability by AAT PI status tested for association of binary PI-status with continuous or binary outcome, adjusted for age and sex.

To explore the change in effect of PI status on lung function following height adjustment, we tested for association of PI status with height (cm), weight (kg) and BMI (kg/m^2). Associations with height were repeated with simultaneous adjustment for FEV1 and FVC.

The analysis of lung function for *CFTR* tested for an association of deltaF508 carrier status with FEV1, FVC and FEV1/FVC ratio in all individuals adjusted for age and sex, and stratified by smoking status. We also repeated the analysis in individuals classified as cases for COPD. The analysis in all individuals was repeated with simultaneous adjustment for height and height-squared.

We also tested for association of PI status (in the usual approach of PI-MS, MZ, SS, SZ, ZZ vs PI-MM) or deltaF508 carrier status with COPD case status.

The analysis of physical and cognitive capability outcomes for *PAH* and *ACADM* tested for an association of mutation carrier status with continuous or binary outcome, adjusted for age and sex.

Within-cohort analyses

To produce estimates by cohort, linear regression was implemented for continuous outcomes and logistic regression for binary outcomes.

One-step meta-analysis

A one-step meta-analysis approach using the IPD from all eight cohorts was used to derive estimates of effect sizes across all studies. This approach was adopted rather than the two-step method because the mutations are rare and thus the exposure of interest (carrier status) was often a rare event in the cohorts. One-step meta-analyses are based on the exact likelihood for the data, do not assume a normal distribution of effect estimates and do not assume that the SE of the effect estimate is exact; they are thus more appropriate in this instance.[11 12] A fixed effects (FE) or a random effects (RE) meta-analysis can be implemented within the one-step framework. We first implemented an RE meta-analysis (as described below) in all associations due to the heterogeneity in study characteristics (age, sex, geographical location). An RE model assumes that the true effect of interest differs across the populations from which the studies are sampled and estimates the average effect.

To implement a one-step RE meta-analysis for continuous outcomes, we used the following command in Stata

mixed outcome binary_genetic_exposure i.study study#c.age study#sex || study: binary_genetic_exposure, noconstant residuals(independent, by(study)).

This mixed model tests for an RE of carrier status by cohort. The fixed portion of the model includes adjustment within cohorts for age and sex, and an intercept by cohort. Residuals are modelled to have study-specific distributions. A random intercept is not assumed.

To implement a one-step RE meta-analysis for binary outcomes, we used the following command in Stata

meqrlogit outcome binary_genetic_exposure i.study study#c.age study#sex || study: binary_genetic_exposure, noconstant.

This similarly tests for a random carrier effect by cohort, with covariate adjustment within cohorts in the fixed part of the model.

The corresponding mathematical model for the continuous outcomes, with β coefficients for FEs and u coefficients for REs, as per the nomenclature in the Stata Reference Manual[26] for mixed is

$$\begin{aligned}\text{Outcome}_{ij} = {} & \beta_0 + \beta_{1j} + \beta_{2j}\,\text{age}_{ij} + \beta_{3j}\,\text{sex}_{ij} + \beta_4\,\text{carrier}_{ij} \\ & + u_{5j}\,\text{carrier}_{ij} + \varepsilon_{ij}\end{aligned}$$

where ε_{ij} is the normally distributed residual term with mean 0 and cohort specific variance and u_{5j} is the random carrier effect by cohort with mean 0 and variance estimated by the model. The corresponding mathematical model[26] for the binary outcomes is

$$\begin{aligned}\text{logit}\,(\Pr(\text{Outcome}_{ij} = 1)) & \\ = {} & \beta_0 + \beta_{1j} + \beta_{2j}\,\text{age}_{ij} + \beta_{3j}\,\text{sex}_{ij} + \beta_4\,\text{carrier}_{ij} \\ & + u_{5j}\,\text{carrier}_{ij}\end{aligned}$$

In practice, we generally found that the estimated variance of the random component of the carrier status effect (the additional effect by cohort) was negligibly small. An FE model was, therefore, more appropriate. The results presented in the main tables also include an FE model using linear regression for

continuous outcomes and logistic regression for binary outcomes, pooling all of the data across cohorts, and including a dummy variable for study. In all FE models, the covariates were again included as factor variables to adjust for effects by cohort (as would be the approach in a standard two-step meta-analysis). For completeness, all tables provide the RE and the FE estimates in addition to the estimated variance of the random carrier effect for interpretation. While the variance of the RE is informative as to whether the genotypic effect was the same across cohorts, it should also be noted that the RE model for continuous outcomes assumed heteroscedastic residuals (by cohort) while the FE model used a simplification of homoscedastic residuals. In a two-step framework, heteroscedastic residuals are modelled because associations are implemented within studies before meta-analysis of the effect estimates. Our main results were robust to either implementation. For the binary outcomes of COPD status and ability to balance, we make the simplifying assumption of independent and identically distributed residuals across cohorts.

The within-cohort estimates are provided for completeness, but these often analyse a rather small number of heterozygotes (or PI-MS, MZ, SS, SZ, ZZ). The meta-analysed estimates are the most reliable as these pool the data to maximise the sample size of the carriers. Online supplementary table S2, which details sample size for the meta-analyses by outcome, should be taken into account when interpreting the coefficients.

Selection analysis

Genotyping

In total, 9912 ALSPAC children were genotyped using the Illumina HumanHap550 quad genome-wide SNP genotyping platform by Sample Logistics and Genotyping Facilities at the Wellcome Trust Sanger Institute and LabCorp supported by 23andMe. Complete data for linkage disequilibrium (LD) analysis were available for 7583 unrelated individuals.

Statistical analyses

EHH was analysed as previously described.[14] EHH measures the decay of homozygosity at a core haplotype of interest. Phased haplotypes involving rs28929474 and rs17580 plus 120 other SNPs (spanning ~100 kb either side from rs28929474 and rs17580) were obtained by the software fastPHASE v1.2[27] from 7583 ALSPAC individuals. We used the Sweep program for the identification of core haplotypes involving the two SNPs using the block definition from Gabriel *et al*.[28]

We used the Haplotter program[13] to explore signatures of selection in the *SERPINA1* gene and surrounding genomic region (of 1 Mb either side). To this end, Haplotter considers data available for ~800 000 common SNPs and 309 unrelated individuals from three populations. This web tool displays the results of selection from HapMap data by computing iHS, Fay and Wu's H, Tajima's D and F_{st}.[13]

RESULTS

Meta-analysis of HALCyon cohorts

For completeness, we show both RE and FE analyses. All analyses are given as online supporting information in the order: *SERPINA1* (see online supplementary tables S3–S35), *CFTR* (see online supplementary tables S36–S40), *ACADM* (see online supplementary tables S41–S43) and *PAH* (see online supplementary tables S44–S46).

The genotype frequencies are provided in online supplementary tables S3–S5, S36, S41 and S44. There were no mutant homozygote calls for *CFTR*, *ACADM* or *PAH*. There was limited evidence for any carrier effect of K304E or the three *PAH* mutations combined. There was weak evidence for a negative effect of deltaF508 heterozygosity on height-adjusted FVC (see online supplementary table S37). The individual cohort and meta-analysed effect estimates for *CFTR*, *ACADM* and *PAH* are provided in online supplementary tables S37–S46. Overall for *SERPINA1*, there was no compelling evidence of an association between PI status and physical capability (see online supplementary table S24). However, there was consistent evidence across the cohorts for a respiratory difference of PI-MZ individuals compared with MM individuals (table 2). No effect was observed in PI-MS individuals. The estimated variance of the RE of carrier status on FEV1 and FVC in the RE one-step meta-analysis was very small, suggesting a fixed carrier effect across cohorts. The FEs estimate was a 0.13 SD increase in FEV1 ($p=1.7\times10^{-5}$) and a 0.16 SD increase in FVC ($p=5.2\times10^{-8}$) using IPD data from all eight cohorts. Taking the study SDs and multiplying by these coefficients, this corresponds to a difference of approximately 81–108 mL (FEV1) and 115–170 mL (FVC). There was no association with FEV1/FVC ratio (see online supplementary table S6). Our analysis of the possible effect of smoking is shown in figure 1 (see online supplementary tables S10–S15). Stratifying as current (N=2430), ex (N=6422) and never (N=5473) smokers, there was no evidence for a difference in PI-MZ effect by smoking status.

Considering the well-known correlation of lung function with height,[23] additional models adjusted for height were run for the AAT variants (see online supplementary table S17). We initially adjusted for height and height-squared (theoretically considering respiratory surface area), with additional adjustments for height-cubed (theoretically considering total respiring cell mass; see online supplementary table S22). Empirically, FEV1 and FVC depend on powers of height in the range 2.1–2.4 (Global Lung Function Initiative prediction equations[23]). The association of PI-MZ status with FEV1 and FVC was attenuated after adjustment for powers of height (height and height-squared, table 3), but approximately half of the effect remained, suggestive that height and lung function are partially related covariates of PI-MZ. Including height-cubed did not further attenuate the genotypic association. We also considered the unadjusted PI-MZ association (FE meta-analysis) with percentage of predicted FEV1 or FVC using the Global Lungs Initiative ERS Task Force 2012 regression equations[23] used in the COPD classification. This resulted in a slight attenuation of the association with FEV1 (1.3% increase, p=0.09) and FVC (1.6% increase, p=0.02). While the prediction equations could be accounting for height in a purer way to covariate adjustment, they produced percentage of predicted values lower than 100% in HALCyon never smokers, which indicates that prediction equations specific to this sample of British ageing individuals of European ancestry may be required. The question of whether PI-MZ exerts a pleiotropic effect of enhanced respiratory capacity independently of its height association thus requires further investigation.

The linear association between PI-MZ and height, adjusted for age and sex (table 4), was notable ($p=3.6\times10^{-10}$, FE analysis) but was not observed for PI-MS. MZ individuals averaged approximately 1.5 cm taller than MM individuals. The FE and RE meta-analyses were repeated in individuals <55 years of age. The coefficient was reduced slightly (1.3 cm increase in MZ, p=0.005, n=4552 FE analysis of four cohorts), but contained the CI including all eight cohorts. We, therefore, concluded that the PI-MZ effect on height represents a growth not age-related shrinkage effect. There is also some hint (see online supplementary table S27) that mean height may increase across genotypes

Table 2 Association of alpha 1-antitrypsin protease inhibitor (PI) status with standardised lung function adjusted for age and sex

Outcome	Cohort	Regression coefficient (95% CI)	
		MS vs MM	MZ vs MM†
Maximum FEV1	BO	−0.05 (−0.36 to 0.27)	0.45 (−0.14 to 1.04)
	CaPS	0.07 (−0.10 to 0.24)	0.11 (−0.12 to 0.33)
	ELSA	0.05 (−0.03 to 0.12)	0.12* (0.01 to 0.23)
	HAS	−0.49* (−0.94 to −0.05)	0.08 (−0.39 to 0.55)
	HCS	−0.01 (−0.10 to 0.09)	0.05 (−0.10 to 0.19)
	LBC1921	−0.06 (−0.31 to 0.19)	0.16 (−0.22 to 0.54)
	NSHD	0.02 (−0.09 to 0.13)	0.18* (0.01 to 0.35)
	WHII	0.05 (−0.03 to 0.12)	0.18** (0.06 to 0.29)
	Combined FE	0.03 (−0.01 to 0.07)	0.13**** (0.07 to 0.19)
	Combined RE	0.03 (−0.01 to 0.07)	0.13**** (0.07 to 0.19)
	Estimated var‡	1.20e−13 (0.00e+00 to .)	1.09e−19 (0.00e+00 to .)
Maximum FVC	BO	0.03 (−0.26 to 0.32)	0.43 (−0.11 to 0.97)
	CaPS	0.08 (−0.10 to 0.26)	0.15 (−0.09 to 0.39)
	ELSA	0.02 (−0.05 to 0.09)	0.12* (0.01 to 0.22)
	HAS	−0.35 (−0.73 to 0.03)	0.33 (−0.09 to 0.74)
	HCS	−0.02 (−0.11 to 0.07)	0.14* (0.01 to 0.27)
	LBC1921	−0.07 (−0.32 to 0.17)	0.20 (−0.17 to 0.56)
	NSHD	0.02 (−0.09 to 0.13)	0.20* (0.04 to 0.37)
	WHII	−0.01 (−0.08 to 0.07)	0.19** (0.07 to 0.30)
	Combined FE	0.01 (−0.03 to 0.04)	0.16**** (0.10 to 0.22)
	Combined RE	0.00 (−0.04 to 0.04)	0.16**** (0.10 to 0.22)
	Estimated var‡	3.76e−15 (1.01e−28 to 1.40e−01)	3.53e−19 (0.00e+00 to .)

Estimates for PI-SS, SZ and ZZ are provided in the online supplement.
*p<0.05, **p<0.01, ***p<0.001, ****p<0.0001.
†Exact p values for PI-MZ-FEV1 were 1.7×10^{-5} (FE) and 1.1×10^{-5} (RE); for PI-MZ-FVC were 5.2×10^{-8} (FE) and 3.2×10^{-8} (RE).
‡Estimated variance of the random slope on carrier status modelled by the RE model.
BO, Boyd Orr; CaPS, Caerphilly Prospective Study; ELSA, English Longitudinal Study of Ageing; FE, fixed effect; FEV1, forced expiratory volume in 1 second; FVC, forced vital capacity; HAS; Hertfordshire Ageing Study; HCS, Hertfordshire Cohort Study; LBC1921, Lothian Birth Cohort 1921; NSHD, MRC National Survey of Health and Development; RE, random effect; WHII, Whitehall II Study.

MM;MZ;SZ;ZZ. The association of PI-MZ versus MM and height was additionally simultaneously adjusted for FVC and FEV, which attenuated but did not remove the association (see online supplementary table S29). We note that both the respiratory and height associations occur in geographically confined cohorts. The by-cohort analyses show no evidence for a geographically stratified effect. The PI-MZ age-adjusted and sex-adjusted associations with height, FEV1 z-score and FVC z-score did not appear to be driven by population stratification when we ran models adjusted for principal components in four of the studies (subsamples of ELSA, WHII, CaPS and LBC1921 with principal components available), although sample size was

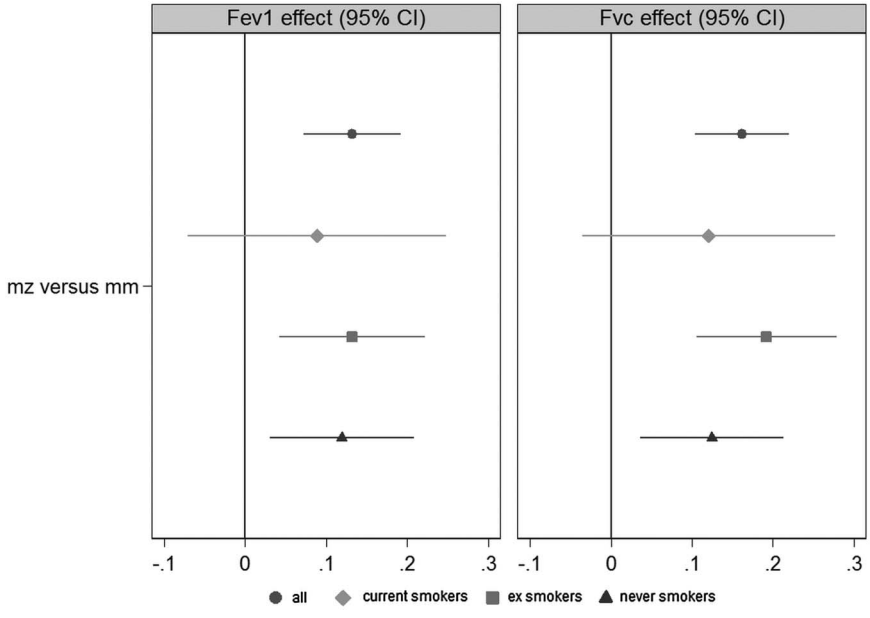


Figure 1 Regression coefficients from fixed effects one-step meta-analysis of protease inhibitor-MZ effect on lung function (z-scored within cohorts) adjusted for age and sex. Estimates from analyses stratified by smoking status are also provided (current smokers N=2430; ex smokers N=6422; never smokers N=5473).[38] FEV1, forced expiratory volume in 1 second; FVC, forced vital capacity.

Table 3 Association of alpha 1-antitrypsin protease inhibitor (PI)-MZ status and standardised lung function adjusted for age, sex, height and height-squared

Outcome	Cohort	Regression coefficient (95% CI)
Maximum FEV1	BO	0.17 (−0.40 to 0.74)
	CaPS	0.09 (−0.12 to 0.30)
	ELSA	0.06 (−0.05 to 0.16)
	HAS	0.11 (−0.34 to 0.56)
	HCS	−0.01 (−0.15 to 0.12)
	LBC1921	0.09 (−0.27 to 0.44)
	NSHD	0.08 (−0.08 to 0.23)
	WHII	0.11 (−0.00 to 0.21)
	Combined FE	0.07* (0.01 to 0.12)
	Combined RE	0.07* (0.01 to 0.12)
	Estimated var†	1.03e−17 (1.80e−32 to 5.87e−03)
Maximum FVC	BO	0.06 (−0.43 to 0.56)
	CaPS	0.13 (−0.08 to 0.34)
	ELSA	0.04 (−0.06 to 0.14)
	HAS	0.35 (−0.04 to 0.74)
	HCS	0.07 (−0.05 to 0.18)
	LBC1921	0.10 (−0.23 to 0.44)
	NSHD	0.08 (−0.06 to 0.23)
	WHII	0.10 (−0.00 to 0.20)
	Combined FE	0.08** (0.03 to 0.13)
	Combined RE	0.08** (0.03 to 0.13)
	Estimated var†	1.25e−13 (5.20e−27 to 3.01e+00)

*p<0.05, **p<0.01, ***p<0.001, ****p<0.0001.
†Estimated variance of the random slope on carrier status modelled by the RE model.
BO, Boyd Orr; CaPS, Caerphilly Prospective Study; ELSA, English Longitudinal Study of Ageing; FE, fixed effect; FEV1, forced expiratory volume in 1 second; FVC, forced vital capacity; HAS; Hertfordshire Ageing Study; HCS, Hertfordshire Cohort Study; LBC1921, Lothian Birth Cohort 1921; NSHD, MRC National Survey of Health and Development; RE, random effect; WHII, Whitehall II Study.

markedly attenuated. A homogeneity analysis (χ^2 contingency test) to test whether genotype frequencies of AAT deficiency PI status differ among cohorts did not reveal significant heterogeneity (p=0.310). Nominal but minor differences between observed and expected genotype frequencies were observed for

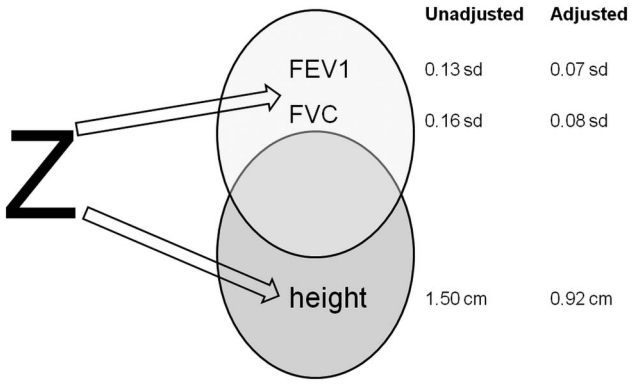


Figure 2 The pleiotropic effect of the Z allele on respiratory capacity and height. 'Unadjusted' estimates are from the (upper circle) one-step fixed effects analyses of z-scored forced expiratory volume in 1 second (FEV1) or z-scored forced vital capacity (FVC) on protease inhibitor (PI)-MZ versus PI-MM adjusted for age and sex or (lower circle) from the one-step fixed effects analysis of height (cm) on PI-MZ versus PI-MM adjusted for age and sex. The 'adjusted' estimates are additionally adjusted for (upper circle) height (cm) and (lower circle) z-scored FEV1 and z-scored FVC simultaneously.

HAS (contributions to $\chi^2>3.84$), but these are related to low numbers and may be explained as type I error. We concluded that the PI-Z allele may have pleiotropic effects on height and respiratory function (figure 2).

The association of PI-MZ with weight and BMI was assessed (see online supplementary tables S31–S35). We observed no association for BMI and an effect estimate for weight that was consistent with what is predicted given the observational correlation between height and weight.

A previous population-based study showed a lower FEV1 in PI-MZ compared with PI-MM in individuals with clinically defined COPD, adjusted for age, sex, height and smoking status.[29] Using the Global Lungs Initiative ERS Task Force 2012 regression equations,[23] we classified all individuals as either cases or non-cases for COPD and reran the age-adjusted and sex-adjusted model in COPD cases (see online supplementary tables S16, S19 and S20). We did not replicate the results of the

Table 4 Association of alpha 1-antitrypsin protease inhibitor (PI) status and height (cm) adjusted for age and sex

	Regression coefficient (95% CI)	
Cohort	MS vs MM	MZ vs MM†
BO	2.31 (−0.24 to 4.86)	7.33** (2.36 to 12.30)
CaPS	0.27 (−0.97 to 1.51)	0.55 (−1.08 to 2.18)
ELSA	0.20 (−0.40 to 0.81)	2.02**** (1.12 to 2.92)
HAS	0.24 (−3.04 to 3.51)	−0.26 (−3.69 to 3.18)
HCS	−0.10 (−0.87 to 0.67)	1.23* (0.09 to 2.37)
LBC1921	0.44 (−1.55 to 2.43)	1.06 (−1.89 to 4.00)
NSHD	0.68 (−0.11 to 1.47)	1.84** (0.64 to 3.04)
WHII	0.23 (−0.34 to 0.81)	1.24** (0.34 to 2.14)
Combined FE	0.28 (−0.03 to 0.59)	1.50**** (1.03 to 1.97)
Combined RE	0.28 (−0.03 to 0.59)	1.51**** (1.04 to 1.97)
Estimated var‡	2.77e−13 (1.06e−26 to 7.27e+00)	1.67e−12 (1.59e−25 to 1.76e+01)

Estimates for PI-SS, SZ and ZZ are provided in the online supplement.
*p<0.05, **p<0.01, ***p<0.001, ****p<0.0001.
†Exact p values for PI-MZ—height were 3.6×10^{-10} (FE) and 2.9×10^{-10} (RE).
‡Estimated variance of the random slope on carrier status modelled by the RE model.
BO, Boyd Orr; CaPS, Caerphilly Prospective Study; ELSA, English Longitudinal Study of Ageing; FE, fixed effect; HAS; Hertfordshire Ageing Study; HCS, Hertfordshire Cohort Study; LBC1921, Lothian Birth Cohort 1921; NSHD, MRC National Survey of Health and Development; RE, random effect; WHII, Whitehall II Study.

previous study, nor after adjustment for powers of height and smoking status. We meta-analysed the odds of COPD in PI classes compared with MM individuals (see online supplementary tables S25 and S26). There was no compelling evidence for an association with PI-MS or PI-MZ, while the PI-SS, SZ and PI-ZZ meta-analyses were not possible due to the low genotype frequencies. However, two out of six ZZs with available data displayed COPD (one extreme).

Online supplementary table S2 shows that overall we conducted in the region of 182 one-step meta-analyses across all outcomes and genetic variants. Many of these tests were not independent due to the outcomes (eg, FEV1/FVC ratio is derived from FEV1 and FVC) or the genetic exposures (eg, PI-MM were included in all PI analyses) or due to subgroup analyses (eg, smokers, COPD) or rerunning adjusted models. However, even with a Bonferroni adjustment based on this number (p=0.00027), our main results still produce comparatively small p values. The results of further sensitivity analyses are described in online supplementary appendix S4.

Selection analysis

EHH results involving rs28929474 and rs17580 show small decay of EHH, from 1 to 0.6, after 90 kb from 5′ (see online supplementary figure S1). This relatively small reduction is observed for two rare haplotypes (of 5% and 2% frequency, respectively) each of them including the rare allele of each SNP. The decay of EHH is more pronounced to the 3′ end, with EHH for both rare haplotypes being reduced to 0.5 at a distance of 30 kb. These results were qualitatively unchanged with the addition of neighbouring SNPs to the core region.

Results observed from Haplotter for common SNPs analysed by iHS, Fay and Wu's H, Tajima's D and F_{st} show no evidence of selection driving alleles at intermediate or high frequency in and around the *SERPINA1* gene (see online supplementary figure S2).

Recombination data from the ALSPAC sample combined with pairwise LD between SNPs around rs28929474 suggested an allele age of between 100 and 250 generations (see online supplementary figure S3). In contrast, using Z allele frequency this estimate was 1758 generations.

DISCUSSION

The PI-MZ rare (2%) SNP height effect is about fourfold greater than that for the top common SNP in *HMGA2* for height. However, PI-MZ is not represented on GWAS chips, so the largest height meta-analyses of up to 250 000 individuals[30] would not have detected it directly and apparently did not do so by imputation. Furthermore, whole-genome sequencing studies such as UK10K (http://www.UK10K.org) would not have analysed enough individuals to robustly detect the effect even if calls and imputation on low read depth were efficient. While analyses of the possible contribution of common SNPs to height suggest that they could explain the large part of this highly heritable trait,[31] our observation raises the possibility that many common SNPs might be each weakly proxying rarer causal alleles.

Our main results of interest (MZ carrier effect on lung function and height) were obtained from a large number of carriers (>600) at the meta-analysis level. Neither association explains the other, although there is partial phenotypic correlation. The enhancement (rather than reduction) of FEV1 and FVC by PI-Z allele heterozygosity was unexpected and is in apparent contrast with the suggestion of greater incidence of respiratory infections in PI-MZ children[20] and with the well-known severe deleterious effects of PI-ZZ. However, mechanisms for balancing selection on PI-MZ (rs28929474) have previously been proposed,[32] and the potential connective tissue and immunological/inflammatory effects of the Z allele[32] could plausibly lead to enhanced FEV1 and FVC with either positively or negatively correlated inflammatory or infection susceptibility. Previous studies have detected an interaction of PI-MZ with smoking such that PI-MZ ever smokers have reduced respiratory capacity compared with PI-MM.[33] Our analysis restricted to current smokers did not detect reduced respiratory capacity in this group of individuals, and we observed enhanced respiratory capacity in ex smokers. Seventeen per cent of individuals with the relevant covariates (PI status, lung function, age and sex) were current smokers in HALCyon, 97 of which were PI-MZ. Future observational studies with increased sample size should consider current or ever smoking PI-MZ individuals to consider whether there is reduced respiratory capacity in this subgroup of individuals. Alternatively, it could be that a cumulative smoke exposure of an as yet undetermined amount determines the development of respiratory disease in PI-MZ individuals; there is evidence in PI-ZZ and PI-SZ individuals that such a concept exists.[4 34] Consequently, future studies may also need to quantify relevant environmental exposures such as cigarette smoking.

Microsatellite dating of the Z allele suggests appearance 107–135 generations ago, with high prevalence in North Europe.[35] Height SNP variants have recently been shown collectively to have been positively selected in North (vs South) Europeans.[36] Using GWAS data and PI genotype status in another UK cohort, ALSPAC,[16] we analysed for EHH (see online supplementary figure S1). We also tested selection related to common variation around *SERPINA1* from Haplotter[13] (see online supplementary figure S2) and estimated allele age based both on allele frequency[17] and on local recombination between the Z locus and other SNPs. Recombination data in conjunction with pairwise LD between SNPs around rs28929474 indicate an allele age consistent with earlier microsatellite estimates (from 100 to 250 generations, see online supplementary figure S3), and even for a rare SNP, the haplotypes on which Z and S reside are extended, whereas Z allele frequency estimates an age about 10× older (1758 generations). These genomic features all point towards positive selection acting on the Z (and S) alleles. It is, therefore, possible that PI-Z, here shown to be a rarer allele for greater height, has been positively selected on height (or weight—a possible survival advantage in colder latitudes) though PI-ZZ is detrimental to respiratory health. PI-MZ may thus represent a balanced polymorphism with greater height or FEV1 or FVC being advantageous in heterozygotes but lung (and liver) disease being disadvantageous in ZZ homozygotes.

AAT is a therapeutic agent and target in relation to its respiratory importance.[37] Our findings in PI-MZ heterozygotes invite both a reconsideration of what may be an optimal level of AAT for best respiratory function and for the first time a consideration whether AAT may mark a novel aspect of height determination, which could itself become a therapeutic target for height modification in some growth deficiency disorders.

Author affiliations
[1]School of Social and Community Medicine, University of Bristol, Bristol, UK
[2]MRC Lifecourse Epidemiology Unit, University of Southampton, Southampton, UK
[3]National Institute for Health Research Nutrition Biomedical Research Centre, University of Southampton and University Hospital Southampton NHS Foundation Trust, Southampton, UK
[4]National Institute for Health Research Musculoskeletal Biomedical Research Unit, University of Oxford, Oxford, UK
[5]Department of Psychology, University of Edinburgh, Edinburgh, UK

[6]Centre for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh, Edinburgh, UK
[7]Department of Psychiatry, University of Oxford, Oxford, UK
[8]Department of Epidemiology and Public Health, UCL, London, UK
[9]ISER, University of Essex, Essex, UK
[10]University of Bristol/University Hospitals Bristol NHS Foundation Trust National Institute for Health Research Bristol Nutrition Biomedical Research Unit, University of Bristol, Bristol, UK
[11]MRC Unit for Lifelong Health and Ageing at UCL, London, UK

Acknowledgements The authors thank Rachel Cooper for providing Stata code to harmonise the physical capability outcomes in HALCyon. They thank Julian Higgins, Andrew Simpkin and Corrie Macdonald-Wallis for providing expertise during the manuscript preparation. They also thank Kate Birnie, Karen Jameson, Holly Syddall, Vanessa Cox, Nikki Graham, Jorgen Engmann, Aida Sanchez, Sarah Harris, David Evans, Hashem Shihab and Christopher Raistrick for providing data.

Contributors Substantial contributions to the conception or design of the work; or the acquisition, analysis or interpretation of data for the work: T-LN, YB-S, CC, IJD, JG, MK, MK, RMM, AP, AAS, JMS, AW, DK, SR and INMD. Drafting the work or revising it critically for important intellectual content: T-LN, YB-S, CC, IJD, JG, MK, MK, RMM, AP, AAS, JMS, AW, DK, SR and INMD. Final approval of the version to be published: T-LN, YB-S, CC, IJD, JG, MK, MK, RMM, AP, AAS, JMS, AW, DK, SR, INMD. Agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved: T-LN, SR and INMD.

Funding *The HALCyon* cohort collaboration was established by the HALCyon grant funded by the New Dynamics of Ageing (RES-353-25-0001). *BO*: The Wellcome Trust funded R.M.M. to undertake the clinical third wave of follow-up of Boyd Orr (2002–03) as part of a research training fellowship in clinical epidemiology (grant GR063779FR). *CaPS* was conducted by the former MRC Epidemiology Unit (South Wales) and funded by the Medical Research Council of the United Kingdom. The School of Social and Community Medicine, University of Bristol now maintains the archive; more information about data access can be found at http://www.bris.ac.uk/social-community-medicine/people/project/1392. *Samples from the English Longitudinal Study of Ageing (ELSA)* DNA Repository (EDNAR), received support under a grant (AG1764406S1) awarded by the NIA. ELSA was developed by a team of researchers based at the National Centre for Social Research, University College London and the Institute of Fiscal Studies. The data were collected by the National Centre for Social Research. *HCS/HAS*: The Hertfordshire studies were supported by the Medical Research Council, NIHR Southampton BRC, NIHR Musculoskeletal BRU (Oxford), Arthritis Research UK, International Osteoporosis Foundation, British Heart Foundation, and EU Framework 7 programme. *The LBC1921* was funded by the Biotechnology and Biological Sciences Research Council (15/SAG09977), the Chief Scientist Office of the Scottish Executive Health Department (CZB/4/505, ETM/55, CZH/4/213, CZG/3/2/79), a Royal Society-Wolfson Research Merit Award to IJD. The work was undertaken by The University of Edinburgh Centre for Cognitive Ageing and Cognitive Epidemiology, part of the cross council Lifelong Health and Wellbeing Initiative (MR/K026992/1). Funding from the Biotechnology and Biological Sciences Research Council (BBSRC) and Medical Research Council (MRC) is gratefully acknowledged. *The MRC National Survey of Health and Development (NSHD)* is supported by the UK Medical Research Council [MC_UU_12019/1]. Bona fide researchers can apply to access the NSHD data via a standard application procedure (further details available at: http://www.nshd.mrc.ac.uk/data.aspx). *The Whitehall II study* has been supported by grants from the UK: the Medical Research Council (K013351); Economic and Social Research Council; British Heart Foundation; Health and Safety Executive; Department of Health; and from the US: the National Institute on Aging (NIA; grant no. AG13196; R01AG034454); National Heart Lung and Blood Institute (grant no. HL36310), NIH; and the John D. and Catherine T. MacArthur Foundation Research Networks on Successful Midlife Development and Socio-economic Status and Health. *ALSPAC*: We are extremely grateful to all the families who took part in this 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. The UK Medical Research Council and the Wellcome Trust (grant ref: 102215/2/13/2) and the University of Bristol provide core support for ALSPAC. This publication is the work of the authors and TLN will serve as guarantor for the contents of this paper. Please note that the study website contains details of all the data that is available through a fully searchable data dictionary (http://www.bris.ac.uk/alspac/researchers/data-access/data-dictionary/). T-LN is the recipient of an MRC PhD studentship in the Bristol Centre for Systems Biomedicine (BCSBmed) MRC doctoral training centre. CC and DK are funded by the Medical Research Council [MC_UU_12019/1, MC_UU_12019/4]. MK is supported by an Economic and Social Research Council professorial fellowship. IJD and JMS are supported by the Centre for Cognitive Ageing and Cognitive Epidemiology, which is funded by the Biotechnology Sciences Research Council and the Medical Research Council and the University of Edinburgh as part of the cross council Lifelong Health and Wellbeing initiative (MR/K026992/1). The NIHR Bristol Nutrition Biomedical Research Unit is funded by the National Institute for Health Research (NIHR) and is a partnership between the University Hospitals Bristol NHS Foundation Trust and the University of Bristol. The funders had no role in the conduct or reporting of the study.

Competing interests None declared.

Patient consent Obtained.

Ethics approval Please see online supporting information S1 for a list of Ethics Committees that approved each study.

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES

1 Stegink LD, Koch R, Blaskovics ME, Filer LJ Jr, Baker GL, McDonnell JE. Plasma phenylalanine levels in phenylketonuric heterozygous and normal adults administered aspartame at 34 mg/kg body weight. *Toxicology* 1981;20:81–90.
2 de Serres F, Blanco I. Role of alpha-1 antitrypsin in human health and disease. *J Intern Med* 2014;276:311–35.
3 Hersh CP, Dahl M, Ly NP, Berkey CS, Nordestgaard BG, Silverman EK. Chronic obstructive pulmonary disease in alpha1-antitrypsin PI MZ heterozygotes: a meta-analysis. *Thorax* 2004;59:843–9.
4 Green CE, Vayalapra S, Hampson JA, Mukherjee D, Stockley RA, Turner AM. PiSZ alpha-1 antitrypsin deficiency (AATD): pulmonary phenotype and prognosis relative to PiZZ AATD and PiMM COPD. *Thorax* 2015;70:939–45.
5 Kuh D, Cooper R, Richards M, Gale C, von Zglinicki T, Guralnik J. A life course approach to healthy ageing: The HALCyon programme. *Public Health* 2012;126:193–5.
6 *A life course approach to healthy ageing*. Oxford: Oxford University Press, 2014.
7 Cooper R, Kuh D, Hardy R, Mortality Review Group; FALCon and HALCyon Study Teams. Objectively measured physical capability levels and mortality: systematic review and meta-analysis. *BMJ* 2010;341:c4467.
8 Leivseth L, Nilsen TIL, Mai X-M, Johnsen R, Langhammer A. Lung function and respiratory symptoms in association with mortality: the HUNT study. *COPD: Journal of Chronic Obstructive Pulmonary Disease* 2014;11:59–80.
9 Sabia S, Guéguen A, Marmot MG, Shipley MJ, Ankri J, Singh-Manoux A. Does cognition predict mortality in midlife? Results from the Whitehall II cohort study. *Neurobiol Aging* 2010;31:688–95.
10 Buisseret PD, Pembrey ME, Lessof MH. α1-antitrypsin phenotypes in rheumatoid arthritis and ankylosing spondylitis. *Lancet* 1977;310:1358–9.
11 Simmonds MC, Higgins JP. A general framework for the use of logistic regression models in meta-analysis. *Stat Methods Med Res* 2014.
12 Higgins JPT, Whitehead A, Turner RM, Omar RZ, Thompson SG. Meta-analysis of continuous outcome data from individual patients. *Stat Med* 2001;20:2219–41.
13 Voight BF, Kudaravalli S, Wen X, Pritchard JK. A map of recent positive selection in the human genome. *PLoS Biol* 2006;4:e72.
14 Sabeti PC, Reich DE, Higgins JM, Levine HZP, Richter DJ, Schaffner SF, Gabriel SB, Platko JV, Patterson NJ, McDonald GJ, Ackerman HC, Campbell SJ, Altshuler D, Cooper R, Kwiatkowski D, Ward R, Lander ES. Detecting recent positive selection in the human genome from haplotype structure. *Nature* 2002;419:832–7.
15 Rodriguez S, Williams DM, Guthrie PAI, McArdle WL, Smith GD, Evans DM, Gaunt TR, Day INM. Molecular and population analysis of natural selection on the human haptoglobin duplication. *Ann Hum Genet* 2012;76:352–62.
16 Boyd A, Golding J, Macleod J, Lawlor DA, Fraser A, Henderson J, Molloy L, Ness A, Ring S, Davey Smith G. Cohort Profile: the 'Children of the 90s'—the index offspring of the Avon Longitudinal Study of Parents and Children. *Int J Epidemiol* 2013;42:111–27.
17 Slatkin M, Rannala B. Estimating allele age. *Annu Rev Genomics Hum Genet* 2000;1:225–49.
18 Alfred T, Ben-Shlomo Y, Cooper R, Hardy R, Cooper C, Deary IJ, Elliott J, Gunnell D, Harris SE, Kivimaki M, Kumari M, Martin RM, Power C, Sayer AA, Starr JM, Kuh D, Day INM, HALCyon Study Team. Absence of association of a single-nucleotide polymorphism in the *TERT-CLPTM1L* locus with age-related phenotypes in a large multicohort study: the HALCyon programme. *Aging Cell* 2011;10:520–32.
19 Khalid JM, Oerton J, Cortina-Borja M, Andresen BS, Besley G, Dalton RN, Downing M, Green A, Henderson M, Leonard J, Dezateux C. Ethnicity of children with homozygous c.985A>G medium-chain acyl-CoA dehydrogenase deficiency: findings from screening approximately 1.1 million newborn infants. *J Med Screen* 2008;15:112–7.
20 Wadsworth MEJ, Vinall LE, Jones AL, Hardy RJ, Whitehouse DB, Butterworth SL, Hilder WS, Lovegrove JU, Swallow DM. Alpha1-Antitrypsin as a Risk for Infant and

Adult Respiratory Outcomes in a National Birth Cohort. *Am J Respir Cell Mol Biol* 2004;31:559–64.
21 Zschocke J. Phenylketonuria mutations in Europe. *Hum Mutat* 2003;21:345–56.
22 Scriver CR, Hurtubise M, Konecki D, Phommarinh M, Prevost L, Erlandsen H, Stevens R, Waters PJ, Ryan S, McDonald D, Sarkissian C. PAHdb 2003: What a locus-specific knowledgebase can do. *Hum Mutat* 2003;21:333–44.
23 Quanjer PH, Stanojevic S, Cole TJ, Baur X, Hall GL, Culver BH, Enright PL, Hankinson JL, Ip MSM, Zheng J, Stocks J, ERS Global Lung Function Initiative. Multi-ethnic reference values for spirometry for the 3–95-yr age range: the global lung function 2012 equations. *Eur Respir J* 2012;40:1324–43.
24 Orfei L, Strachan DP, Rudnicka AR, Wadsworth MEJ. Early influences on adult lung function in two national British cohorts. *Arch Dis Child* 2008;93:570–4.
25 *Stata Statistical Software: Release 13[program]*. College Station, TX: StataCorp LP, 2013.
26 StataCorp. *Stata 13 Base Reference Manual*. College Station, TX: Stata Press, 2013.
27 Scheet P, Stephens M. A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase. *Am J Hum Genet* 2006;78:629–44.
28 Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, Higgins J, DeFelice M, Lochner A, Faggart M, Liu-Cordero SN, Rotimi C, Adeyemo A, Cooper R, Ward R, Lander ES, Daly MJ, Altshuler D. The structure of haplotype blocks in the human genome. *Science* 2002;296:2225–9.
29 Dahl M, Nordestgaard BG, Lange P, Vestbo J, Tybjærg-Hansen A. Molecular diagnosis of intermediate and severe α1-antitrypsin deficiency: MZ individuals with chronic obstructive pulmonary disease may have lower lung function than MM individuals. *Clin Chem* 2001;47:56–62.
30 Wood AR, Esko T, Yang J, Vedantam S, Pers TH, Gustafsson S, Chu AY, Estrada K, Luan JA, Kutalik Z, Amin N, Buchkovich ML, Croteau-Chonka DC, Day FR, Duan Y, Fall T, Fehrmann R, Ferreira T, Jackson AU, Karjalainen J, Lo KS, Locke AE, Magi R, Mihailov E, Porcu E, Randall JC, Scherag A, Vinkhuyzen AAE, Westra H-J, Winkler TW, Workalemahu T, Zhao JH, Absher D, Albrecht E, Anderson D, Baron J, Beekman M, Demirkan A, Ehret GB, Feenstra B, Feitosa MF, Fischer K, Fraser RM, Goel A, Gong J, Justice AE, Kanoni S, Kleber ME, Kristiansson K, Lim U, Lotay V, Lui JC, Mangino M, Leach IM, Medina-Gomez C, Nalls MA, Nyholt DR, Palmer CD, Pasko D, Pechlivanis S, Prokopenko I, Ried JS, Ripke S, Shungin D, Stancakova A, Strawbridge RJ, Sung YJ, Tanaka T, Teumer A, Trompet S, van der Laan SW, van Setten J, Van Vliet-Ostaptchouk JV, Wang Z, Yengo L, Zhang W, Afzal U, Arnlov J, Arscott GM, Bandinelli S, Barrett A, Bellis C, Bennett AJ, Berne C, Bluher M, Bolton JL, Bottcher Y, Boyd HA, Bruinenberg M, Buckley BM, Buyske S, Caspersen IH, Chines PS, Clarke R, Claudi-Boehm S, Cooper M, Daw EW, De Jong PA, Deelen J, Delgado G, Denny JC, Dhonukshe-Rutten R, Dimitriou M, Doney ASF, Dorr M, Eklund N, Eury E, Folkersen L, Garcia ME, Geller F, Giedraitis V, Go AS, Grallert H, Grammer TB, Graszler J, Gronberg H, de Groot LCPGM, Groves CJ, Haessler J, Hall P, Haller T, Hallmans G, Hannemann A, Hartman CA, Hassinen M, Hayward C, Heard-Costa NL, Helmer Q, Hemani G, Henders AK, Hillege HL, Hlatky MA, Hoffmann W, Hoffmann P, Holmen O, Houwing-Duistermaat JJ, Illig T, Isaacs A, James AL, Jeff J, Johansen B, Johansson A, Jolley J, Juliusdottir T, Junttila J, Kho AN, Kinnunen L, Klopp N, Kocher T, Kratzer W, Lichtner P, Lind L, Lindstrom J, Lobbens S, Lorentzon M, Lu Y, Lyssenko V, Magnusson PKE, Mahajan A, Maillard M, McArdle WL, McKenzie CA, McLachlan S, McLaren PJ, Menni C, Merger S, Milani L, Moayyeri A, Monda KL, Morken MA, Muller G, Muller-Nurasyid M, Musk AW, Narisu N, Nauck M, Nolte IM, Nothen MM, Oozageer L, Pilz S, Rayner NW, Renstrom F, Robertson NR, Rose LM, Roussel R, Sanna S, Scharnagl H, Scholtens S, Schumacher FR, Schunkert H, Scott RA, Sehmi J, Seufferlein T, Shi J, Silventoinen K, Smit JH, Smith AV, Smolonska J, Stanton AV, Stirrups K, Stott DJ, Stringham HM, Sundstrom J, Swertz MA, Syvanen A-C, Tayo BO, Thorleifsson G, Tyrer JP, van Dijk S, van Schoor NM, van der Velde N, van Heemst D, van Oort FVA, Vermeulen SH, Verweij N, Vonk JM, Waite LL, Waldenberger M, Wennauer R, Wilkens LR, Willenborg C, Wilsgaard T, Wojczynski MK, Wong A, Wright AF, Zhang Q, Arveiler D, Bakker SJL, Beilby J, Bergman RN, Bergmann S, Biffar R, Blangero J, Boomsma DI, Bornstein SR, Bovet P, Brambilla P, Brown MJ, Campbell H, Caulfield MJ, Chakravarti A, Collins R, Collins FS, Crawford DC, Cupples LA, Danesh J, de Faire U, den Ruijter HM, Erbel R, Erdmann J, Eriksson JG, Farrall M, Ferrannini E, Ferrieres J, Ford I, Forouhi NG, Forrester T, Gansevoort RT, Gejman PV, Gieger C, Golay A, Gottesman O, Gudnason V, Gyllensten U, Haas DW, Hall AS, Harris TB, Hattersley AT, Heath AC, Hengstenberg C, Hicks AA, Hindorff LA, Hingorani AD, Hofman A, Hovingh GK, Humphries SE, Hunt SC, Hypponen E, Jacobs KB, Jarvelin M-R, Jousilahti P, Jula AM, Kaprio J, Kastelein JJP, Kayser M, Kee F, Keinanen-Kiukaanniemi SM, Kiemeney LA, Kooner JS, Kooperberg C, Koskinen S, Kovacs P, Kraja AT, Kumari M, Kuusisto J, Lakka TA, Langenberg C, Le Marchand L, Lehtimaki T, Lupoli S, Madden PAF, Mannisto S, Manunta P, Marette A, Matise TC, McKnight B, Meitinger T, Moll FL, Montgomery GW, Morris AD, Morris AP, Murray JC, Nelis M, Ohlsson C, Oldehinkel AJ, Ong KK, Ouwehand WH, Pasterkamp G, Peters A, Pramstaller PP, Price JF, Qi L, Raitakari OT, Rankinen T, Rao DC, Rice TK, Ritchie M, Rudan I, Salomaa V, Samani NJ, Saramies J, Sarzynski MA, Schwarz PEH, Sebert S, Sever P, Shuldiner AR, Sinisalo J, Steinthorsdottir V, Stolk RP, Tardif J-C, Tonjes A, Tremblay A, Tremoli E, Virtamo J, Vohl M-C, Electronic Medical Records and Genomics (eMEMERGEGE) Consortium; MIGen Consortium; PAGEGE Consortium; LifeLines Cohort Study, Amouyel P, Asselbergs FW, Assimes TL, Bochud M, Boehm BO, Boerwinkle E, Bottinger EP, Bouchard C, Cauchi S, Chambers JC, Chanock SJ, Cooper RS, de Bakker PIW, Dedoussis G, Ferrucci L, Franks PW, Froguel P, Groop LC, Haiman CA, Hamsten A, Hayes MG, Hui J, Hunter DJ, Hveem K, Jukema JW, Kaplan RC, Kivimaki M, Kuh D, Laakso M, Liu Y, Martin NG, Marz W, Melbye M, Moebus S, Munroe PB, Njolstad I, Oostra BA, Palmer CNA, Pedersen NL, Perola M, Perusse L, Peters U, Powell JE, Power C, Quertermous T, Rauramaa R, Reinmaa E, Ridker PM, Rivadeneira F, Rotter JI, Saaristo TE, Saleheen D, Schlessinger D, Slagboom PE, Snieder H, Spector TD, Strauch K, Stumvoll M, Tuomilehto J, Uusitupa M, van der Harst P, Volzke H, Walker M, Wareham NJ, Watkins H, Wichmann HE, Wilson JF, Zanen P, Deloukas P, Heid IM, Lindgren CM, Mohlke KL, Speliotes EK, Thorsteinsdottir U, Barroso I, Fox CS, North KE, Strachan DP, Beckmann JS, Berndt SI, Boehnke M, Borecki IB, McCarthy MI, Metspalu A, Stefansson K, Uitterlinden AG, van Duijn CM, Franke L, Willer CJ, Price AL, Lettre G, Loos RJF, Weedon MN, Ingelsson E, O'Connell JR, Abecasis GR, Chasman DI, Goddard ME, Visscher PM, Hirschhorn JN, Frayling TM. Defining the role of common variation in the genomic and biological architecture of adult human height. *Nat Genet* 2014;46:1173–86.
31 Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, Nyholt DR, Madden PA, Heath AC, Martin NG, Montgomery GW, Goddard ME, Visscher PM. Common SNPs explain a large proportion of the heritability for human height. *Nat Genet* 2010;42:565–9.
32 Lomas DA. The Selective Advantage of α1-Antitrypsin Deficiency. *Am J Respir Crit Care Med* 2006;173:1072–7.
33 Molloy K, Hersh CP, Morris VB, Carroll TP, O'Connor CA, Lasky-Su JA, Greene CM, O'Neill SJ, Silverman EK, McElvaney NG. Clarification of the Risk of Chronic Obstructive Pulmonary Disease in α1-Antitrypsin Deficiency PiMZ Heterozygotes. *Am J Respir Crit Care Med* 2014;189:419–27.
34 Castaldi PJ, DeMeo DL, Kent DM, Campbell EJ, Barker AF, Brantly ML, Eden E, McElvaney NG, Rennard SI, Stocks JM, Stoller JK, Strange C, Turino G, Sandhaus RA, Griffith JL, Silverman EK. Development of predictive models for airflow obstruction in alpha-1-antitrypsin deficiency. *Am J Epidemiol* 2009;170:1005–13.
35 Seixas S, Garcia O, Trovoada J, Santos T, Amorim A, Rocha J. Patterns of haplotype diversity within the serpin gene cluster at 14q32.1: insights into the natural history of the α1-antitrypsin polymorphism. *Hum Genet* 2001;108:20–30.
36 Turchin MC, Chiang CWK, Palmer CD, Sankararaman S, Reich D, Hirschhorn JN. Evidence of widespread selection on standing variation in Europe at height-associated SNPs. *Nat Genet* 2012;44:1015–19.
37 Sorrells S, Camprubi S, Griffin R, Chen J, Ayguasanosa J. SPARTA clinical trial design: exploring the efficacy and safety of two dose regimens of alpha1-proteinase inhibitor augmentation therapy in alpha1-antitrypsin deficiency. *Respir Med* 2015;109:490–9.
38 Jann B. Plotting regression coefficients and other estimates. *Stata Journal* 2014;14:708–37.

S1 Supporting Information

S1 Appendix. Ethical Approval

BO: Ethical approval for the clinical third wave of follow-up of Boyd Orr (2002-03) was obtained from Multi-centre Research Ethics Committee Scotland. All participants gave informed consent.

CaPS: Ethical approval for genotypic analyses was provided by South East Wales Local Research Ethics Committee Panel B (05/WSE02/131). The original CaPS project received ethical approval from the former South Glamorgan Area Health Authority.

English Longitudinal Study of Ageing (ELSA): ELSA has been approved by the National Research Ethics Service and all participants have given informed consent.

HCS/HAS: Ethical approval for the Hertfordshire studies was obtained from the Hertfordshire Local Research Ethics Committee.

LBC1921: Ethical approval for the Lothian Birth Cohort 1921 study was given by the Lothian Research Ethics Committee.

NSHD: Ethical approval for the NSHD data collection at 53 years was approved by the North Thames Multi-Centre Research Ethics Committee (ref. MREC 98/1/121). At 60–64 years ethical approval was obtained from the Central Manchester Local Research Ethics Committee (ref. 07/H1008/245) and the Scotland A Research Ethics Committee (ref. 08/MRE00/12). Written informed consent was obtained from study members at each stage of data collection.

Whitehall II: All participants provided written consent and the University College London ethics committee approved the study.

ALSPAC: Ethical approval for the study was obtained from the ALSPAC Ethics and Law Committee and the Local Research Ethics Committees. Written informed consent for the study was obtained for genetic analysis.

S2 Appendix. Wave of outcome assessment

Boyd Orr (BO)(2): Physical capability and lung function were assessed at the third wave (2002-03).

Caerphilly Prospective Study (CaPS): Cognitive function measures were assessed at phase III. Physical capability measures were assessed at phase V. Lung function measures were analysed from phase I where available, but were substituted with measures from phase II for those individuals who did not have valid measures at phase I but did at phase II.

English Longitudinal Study of Ageing (ELSA)(3): Physical capability, cognitive capability and lung function were assessed at wave 2 (2004/5).

Hertfordshire Ageing Study (HAS) (4): Cognitive function was assessed at wave 1 (1994/5). Grip strength was assessed at wave 1 and all other physical capability measures at wave 2. Lung function was measured at wave 2 (2003/05).

Hertfordshire Cohort Study (HCS) (5): Grip strength and lung function were assessed at wave 1 (1999-2004) while TUG (Timed-Up and Go) speed, walk speed, balance ability and chair rise speed were assessed at both waves 1 (1999-2004) and 2 (2004/05) with partial overlap in some tests and no overlap in others. These latter measures were combined across waves, with priority given to wave 1, and the covariates tailored as such.

Lothian Birth Cohort 1921 (LBC1921) (6): Physical and cognitive capability and lung function were assessed at wave 1 age 79 years.

MRC National Survey of Health and Development (NSHD) (7): All cognitive capability and lung function measures were taken from the 1999 wave when the study members were 53

years. All physical capability measures were taken from the 1999 wave with the exception of TUG speed, which was analysed when the study members were 60-64 years.

Whitehall II Study (WHII) (8): Walking speed and lung function were analysed at phase 7 (2002-04), while all cognitive outcomes were analysed at phase 5 (1997-99).

S3 Appendix. The derivation and harmonization of variables

Physical capability

Details of the ascertainment and harmonisation of the five measures of physical capability used in analyses are described in detail elsewhere(9) and are summarised here. The approach to harmonise chair rise times (5 or 10 rises) was to calculate chair rise speed in the current study.

Grip strength was tested in ELSA, HAS, HCS, LBC1921 AND NSHD using handheld dynamometers (the specific devices used in each study are described elsewhere (9)). The maximum measure was used in each study (extracted from 3 measures of each hand in ELSA, HAS and HCS, 3 measures of the dominant hand in LBC1921 and 2 measures in each hand in NSHD). If repeat measures were missing the existing measures were used to derive the maximum.

Standing balance was assessed in BO, CaPS, ELSA, HAS, HCS and NSHD. Owing to the heterogeneity in the way the test was administered across cohorts, the outcome used in analyses was a derived binary variable for inability to balance on one leg with eyes open for five seconds. In ELSA the tests administered were more complex as described by Cooper *et al.*(9) and we derived the outcome in the same way, namely, inability to balance in full tandem with eyes open for 5 seconds with individuals who were not progressed to the next phase of testing classed as unable. Individuals who did not complete the balance test for health reasons were classed as unable in all analyses. If tests were conducted more than once the best performance was used to derive the outcome variable.

The timed walk test was conducted in LBC1921 (6 metres as fast as possible), HAS and HCS (3 metres at normal pace), ELSA (8 feet at normal pace with 2 trials) and WHII (8 feet at

normal pace with 3 trials). To normalise the distribution and to make a higher outcome a healthier outcome, times were converted to speeds in metres per second and then averaged where repeat trials were available.

The timed get up and go test was performed in BO, HAS, HCS, CaPS and NSHD. In all cohorts, study members had to rise from a chair, walk 3 metres at a normal pace and return to a seated position in the chair. The test was repeated in BO and CaPS. Again all times were converted to speeds in metres per second and then averaged where the trial was conducted more than once.

Timed chair rises were assessed in HAS, HCS, ELSA and NSHD. All times were converted to chair rise speed in stands per second. The cohorts measured time to complete 5 or 10 chair rises as fast as possible. In ELSA, individuals under 69 years performed 10 rises while those aged 70 and over performed 5 rises. Time to complete 5 rises was measured in both age groups and this was used to derive chair rise speed.

Some physical performance measures were conducted in part of the HCS cohort in one wave and in the remaining cohort in a later wave. To maximise sample size, measures were pooled across waves and covariates were tailored according to the wave at which the outcome had been performed.

Cognitive capability

The measures of cognition across the HALCyon cohorts were categorised into measures of crystallised ability and measures of fluid cognition.

Measures of crystallised cognitive function

The National Adult Reading Test (NART)(10) was available in LBC1921, CaPS and NSHD. This requires study members to read aloud 50 words with irregular pronunciation and the number of words pronounced correctly is used in analyses here. NART should reflect pre-morbid IQ.

The Mill Hill vocabulary test(11) was administered in HAS and WHII. Study members had to choose the correct synonym for 33 words out of 6 multiple choice answers with increasing difficulty. The number of correct answers is used in analyses.

Measures of fluid cognitive function

Semantic fluency was tested in ELSA, NSHD and WHII via a verbal or written test where study members were asked to name as many animals as possible in 1 minute. The number of unique animals named was used in analyses.

Verbal memory was tested in ELSA, NSHD and WHII via a word recall test. The numbers of words correctly recalled was used in analyses. In NSHD, we summed the total score for remembering the same 15 words in writing over three consecutive trials. The sum of two trials with a delay for the second trial for remembering 10 words verbally was analysed in ELSA. 20 words were recalled in writing in WHII.

Phonemic fluency was analysed in LBC1921 and WHII. In LBC1921, study members were given three 1 minute trials to name as many words as possible beginning with F, L and C. The total number of words is used in analyses. In WHII, study members wrote as many words as possible in 1 minute beginning with S.

Search speed was tested in ELSA (780 letters) and NSHD (600 letters) whereby participants crossed out particular letters in a large grid of letters. The number of letters searched per minute was used in analyses.

The Alice Heim 4-I test (AH4)(12) was available for analyses in CaPS, HAS and WHII. This involves 65 verbal and mathematical questions. The total score achieved in 10 minutes was used in analyses here.

Choice reaction time (FCRT) was assessed in CaPS via a computer test in which the study members had to press one of four key pads depending on which box a stimulus appeared in on screen.

Wechsler logical memory(13) was tested in LBC1921. The participants were asked to recall two stories immediately and following a delay for each. The total sum of the scores for each story were progressed to analysis.

Raven's Progressive Matrices(14) were used in LBC1921, in which study members were given 20 minutes to complete 60 multiple choice "complete the pattern" questions. The total score was used in the analysis.

Lung function

We analysed Forced Expiratory Volume in 1 second (FEV1) and Forced Vital Capacity (FVC). All measures for which an unsatisfactory technique had been recorded were removed. All measures were cleaned so that values <0.3 litres and >9 litres were removed. Any instances for which FVC was less than FEV1 were changed to missing (both values excluded from the analysis). If a cohort provided the individual repeat trial data, the cleaning was applied to each trial; if a cohort provided a cleaned summary measure across trials, this was applied to these values.

If a study provided data from individual trials, the maximum FEV1 and the maximum FVC were derived. These could come from different trials. The FEV1/FVC ratio was derived by taking the ratio of these maxima. Individuals were only included in the analysis of lung function if they had both a maximum FEV1 and a maximum FVC. The FEV1, FVC and FEV1/FVC ratio values (derived from the maxima of repeat trials) were z-scored within cohorts to have a mean of 0 and a standard deviation of 1 using all data available.

In the Boyd Orr cohort, lung function was assessed using a compact II Vitalograph Spirometer. Up to 5 blows were conducted per study member and provided for analysis.

In CaPS, the values provided were the maximum FEV1 from 3 trials and the maximum FVC from 3 trials when the highest two valid FEV1 (and separately the highest two valid FVC) were within 100ml of each other. These were derived by Bolton *et al*. and spirometry was performed in the standing position using a McDermott spirometer(15). Values from phase 1 were preferentially used but values from phase 2 were substituted into the analysis if the phase 1 measures were missing or removed by cleaning. Covariates were tailored as such.

ELSA provided the highest technically satisfactory FEV1 reading and the highest technically satisfactory FVC reading (both in litres). These had been derived using data across 3 blows using a Vitalograph Escort spirometer.

In LBC1921, lung function was measured with 3 blows of a MicroMedical Spirometer in the sitting position without nose clips. The best FEV1 and FVC of the three blows were provided by the cohort for analysis.

In HAS, FEV1 and FVC were provided for 2 blows using a MicroMedical Micro Spirometer(4).

Lung function in HCS was measured using a MicroMedical Micro Spirometer(5). Three trials were conducted and provided for analysis.

In NSHD, lung function was measured using a MicroMedical Plus Spirometer(16), the cohort provided values as the maximum of two blows (for FEV1 and FVC separately) when the difference between trials was less than or equal to 0.30 L. If the study member only had one valid measure they were excluded from analyses. Biologically unfeasible values (<0.30L or >9L) and individuals regarded as having an unsatisfactory technique were removed before deriving the maxima.

Lung function was measured in WHII using a MicroMedical Micro Plus Spirometer. Each study member attempted three blows which were provided for analysis.

Height, weight and BMI

Standing height (cm), weight (kg) and Body Mass Index (BMI, kg/m^2) were included in analyses. In CaPS, height was taken from phase 1 and replaced with height from phase 2 if the phase 1 measure was missing. For the PI-height analyses, a tailored age variable for the wave at which height was assessed was derived. In Boyd Orr, height, weight and BMI were analysed from the clinic sample.

Age and smoking covariates

Individuals aged 90 years or over are not assigned an exact age in ELSA data releases. As such, we estimated the age of these individuals using a representative estimate of the mean age of individuals aged 90 and over in England and Wales in 2005 (the year of wave 2 assessment). To calculate this estimate, we used the England and Wales Mid-Year Population Estimates of the Very Elderly, 2002-2010, demographic table "Mid-2010 Estimates of the very elderly (including centenarians) England and Wales; estimated resident population"

which was part of the Population Estimates of the Very Elderly, 2010 Office for National Statistics release (release date 29 September 2011, date accessed 5 February 2014 from http://www.ons.gov.uk/ons/publications/re-reference-tables.html?edition=tcm%3A77-223697). The estimated age used in analyses was 92.62 years. These individuals were not included in analyses of lung function or anthropometry.

Smoking status was defined as current, ex or never smoker; pipe and cigar smoking was included if this information was available.

S4 Appendix. Sensitivity analyses

The association between PI-MZ and lung function was repeated (basic model, height adjusted model) with FEV1 and FVC values outside +/- 3 standard deviations removed within cohorts before z-scoring to explore the robustness of the coefficient estimates to extreme values, S8 and S21 Tables. This was also implemented for the PI-MZ and height associations (basic model, lung function adjusted model) and the PI-MZ and weight association (S28, S30 and S32 Tables).

We repeated the basic PI-MZ vs PI-MM random effects models against FEV1, FVC and height using restricted maximum likelihood rather than the default maximum likelihood. This made negligible difference to the fixed effects estimates and p-values, while the variance of the random carrier effect was still immaterial.

The distribution of the raw, standardised and studentised residuals were reviewed from the fixed-effects analyses to examine the normality of the distribution. The residuals versus the fitted values were examined for all models suggestive of an association to check for independence. The fixed effects linear model of PI-MZ and z-scored FEV1 suggested that the variance of the residuals increased slightly with the fitted values, but a transformation of the outcome would not remediate this. The distribution of the residuals of FEV1 tended to have a fat tail for negative residuals, for which there was no appropriate transformation.

We repeated the fixed effects analysis of PI-MZ vs MM against z-scored FEV1 and FVC (age and sex adjusted) with robust standard errors, which do not assume that the residuals in the model are identically distributed. The regression coefficients do not change with this approach but the updated 95% confidence intervals were (0.07, 0.19) for FEV1 and (0.10, 0.22) for FVC. The p-values were very slightly attenuated.

We considered whether the variance in lung function (z-scored FEV1 and FVC, square-root transformed and z-scored FVC) differed across PI classes and in deltaF508 carriers vs non-

carriers (FVC), in addition to whether the variance in height (cm) differed across PI classes by pooling the data across cohorts and using Levene's test (S47 Table). This revealed that there was evidence for a difference in the variance of FEV1 and FVC in PI-MZ vs MM (greater variance in PI-MZ). However, the difference was reasonably small (difference of 0.06 in the standard deviation of pooled z-scored FEV1 and difference of 0.08 in the standard deviation of pooled z-scored FVC).

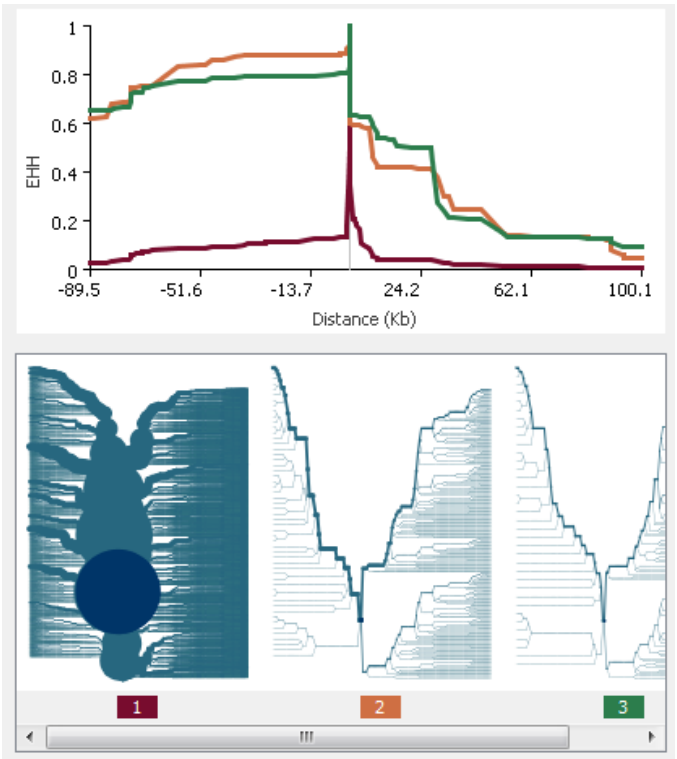
To account for the unequal variance in PI-MZ vs PI-MM, we repeated the basic fixed effects model (age and sex adjusted) with z-scored FEV1 and FVC within a mixed effects framework, enforcing heteroskedastic residuals by genotype rather than by cohort (so the only random component of the model is the clustering of residuals by genotype). The effect estimates were unchanged but the p-values were slightly attenuated ($p=2.6 \times 10^{-5}$ (FEV1) and $p=1.3 \times 10^{-7}$ (FVC)).

We considered influential observations for the associations of interest. These were (1) PI-MZ vs MM and FEV1 (basic and height, height-squared adjusted models) (2) PI-MZ vs MM and FVC (basic and height, height-squared adjusted models) (3) PI-MZ vs MM and height (basic and lung function adjusted models) (4) PI-MZ vs MM and natural log transformed weight (5) deltaF508 carrier status (height, height-squared adjusted) and FVC and (6) PI-MZ vs MM and grip strength. We calculated the dfbeta statistics from the fixed effects meta-analyses. These calculate by how many standard errors the regression coefficient would change if a single observation were omitted. This confirmed that no observations had a dfbeta value with magnitude greater than or equal to 1, such that no single observation affected the coefficient estimate by more than one standard error. However, we did note that the usual threshold of 2/sqrt(sample size)(1) was extremely sensitive and identified >60% of the mutation carriers for each regression. Owing to the large sample size (>14,000 observations for all analyses except the grip strength analysis which included >10,000 individuals), this

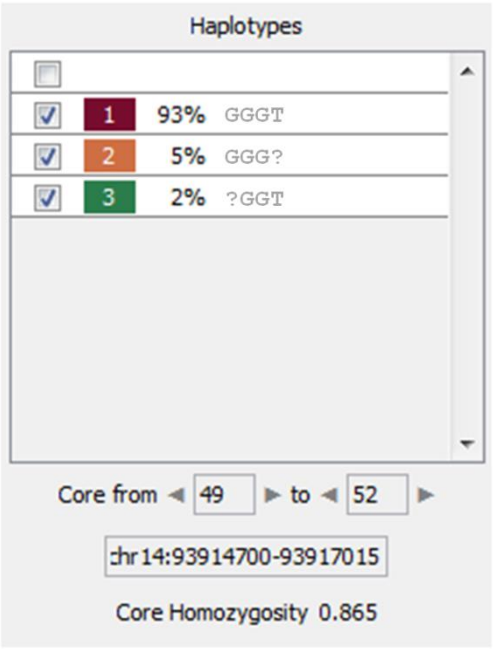
identified observations which changed the coefficients by approximately 0.017-0.02 standard errors. Given the small standard errors in these associations, this identified observations which changed the regression coefficient by only a small amount. A summary of the range of dfbeta values by analysis is provided in S48 Table. We repeated these associations removing the top and bottom 30 individuals with the most extreme dfbeta values. This suggested an increase in the magnitude of the association of PI-MZ with FEV1 and height.

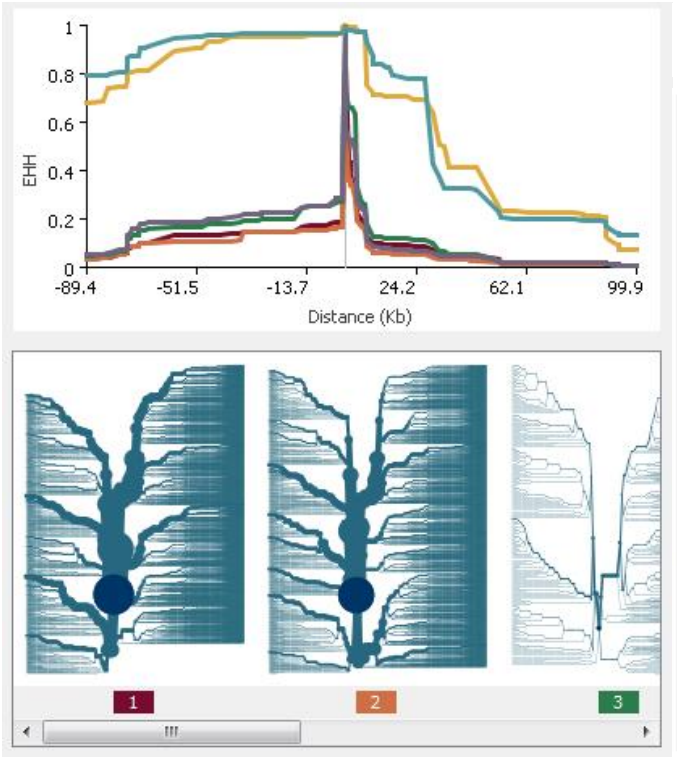
It was noted on examination of the residual versus fitted plots from the fixed effects meta-analyses that when FEV1 or (FEV1/FVC ratio)^3 were analysed, on some occasions there was a bias in the residuals from individuals in ELSA who have COPD (i.e. a very small number of individuals, but this could be detected from the plots). The basic analysis of PI-MZ vs PI-MM against FEV1 was repeated excluding ELSA individuals with COPD. The updated regression coefficient was 0.13 (95% CI: 0.08, 0.19) and was 0.07 (95% CI: 0.01, 0.12) for the height, height-squared adjusted model. We repeated the fixed effects analysis of PI-MZ vs PI-MM with height, excluding Boyd Orr due to the comparatively large effect estimate in this cohort. The regression coefficient was 1.45 cm (0.98cm, 1.92cm).

S1 Fig. EHH analysis of rs28929474 and rs17580 and neighbouring SNPs. A) Core haplotype involves rs28929474, rs1802959, rs28929471 and rs17580. B) Core haplotype involves rs1303, rs28929473, rs28929474, rs28929474, rs1802959, rs28929471, rs17580, rs28929472 and rs28929470.

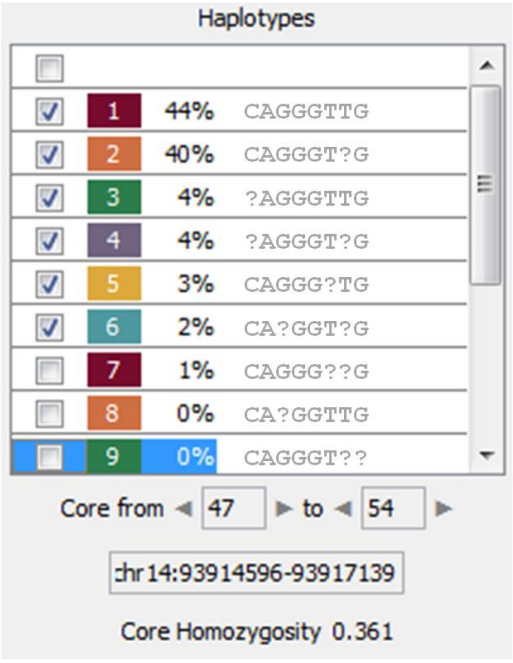


A)





B)



S2 Fig. Results observed from the web tool Haplotter http://haplotter.uchicago.edu/.

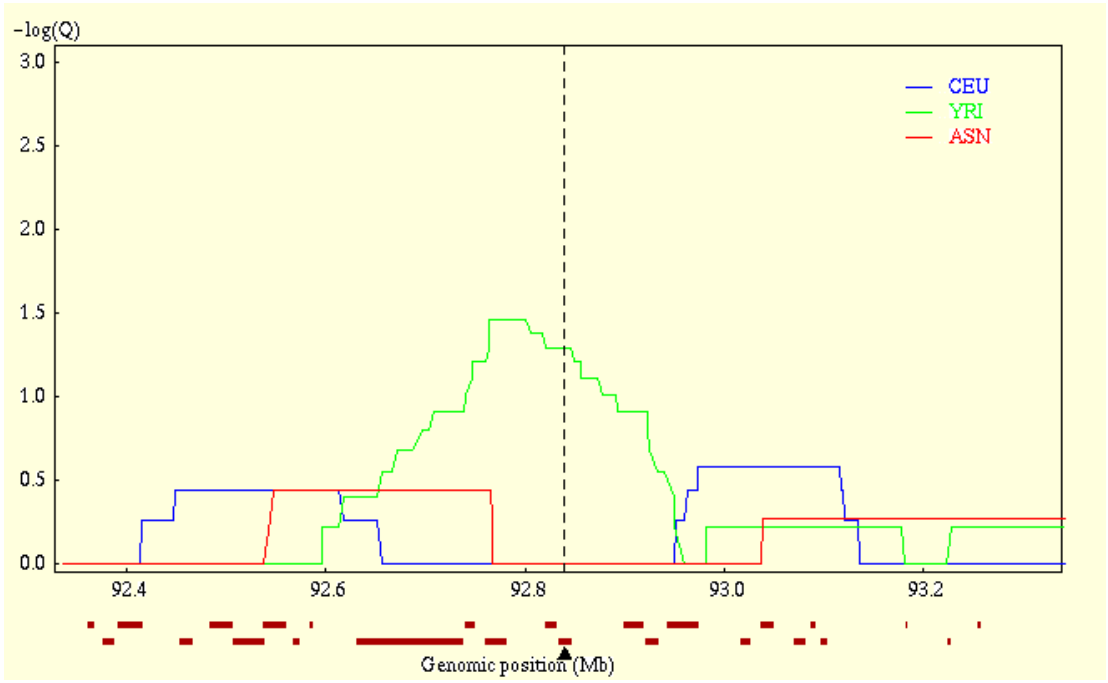
Signatures of recent selection for the SERPINA1 gene and surrounding region (2Mb) have been tested using four different approaches: iHS, Fay and Wu's H, Tajima's D and F_{st}. According to the thresholds described in the literature for each test, there was no evidence of selection involving common SNPs in any case.

Gene name:SERPINA1 ID:5265

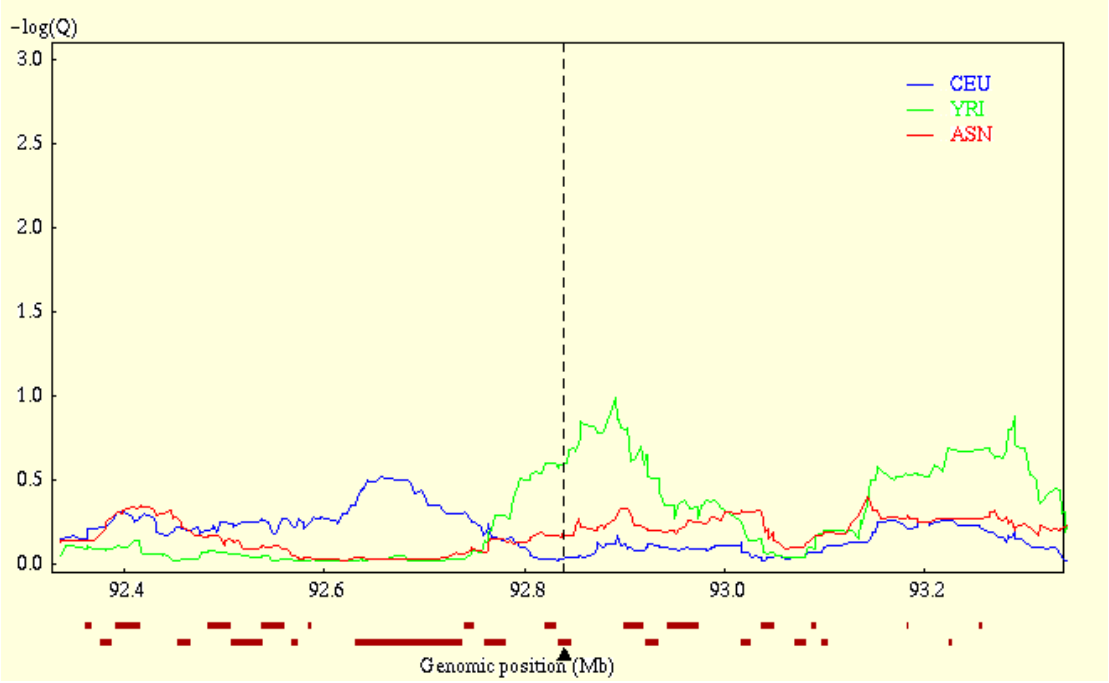
chromosome 14 [92834751:92845165]

-1 Mb | +1 Mb

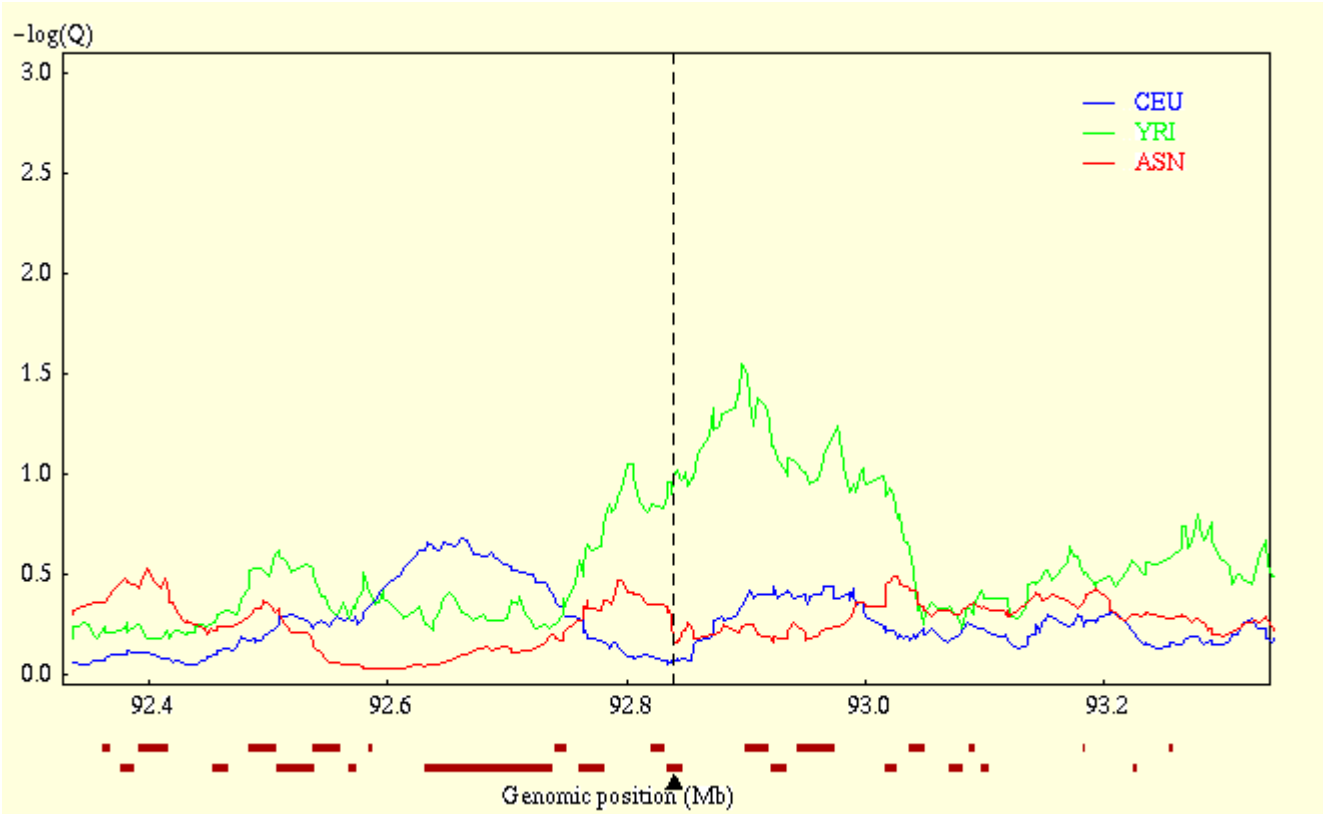
iHS



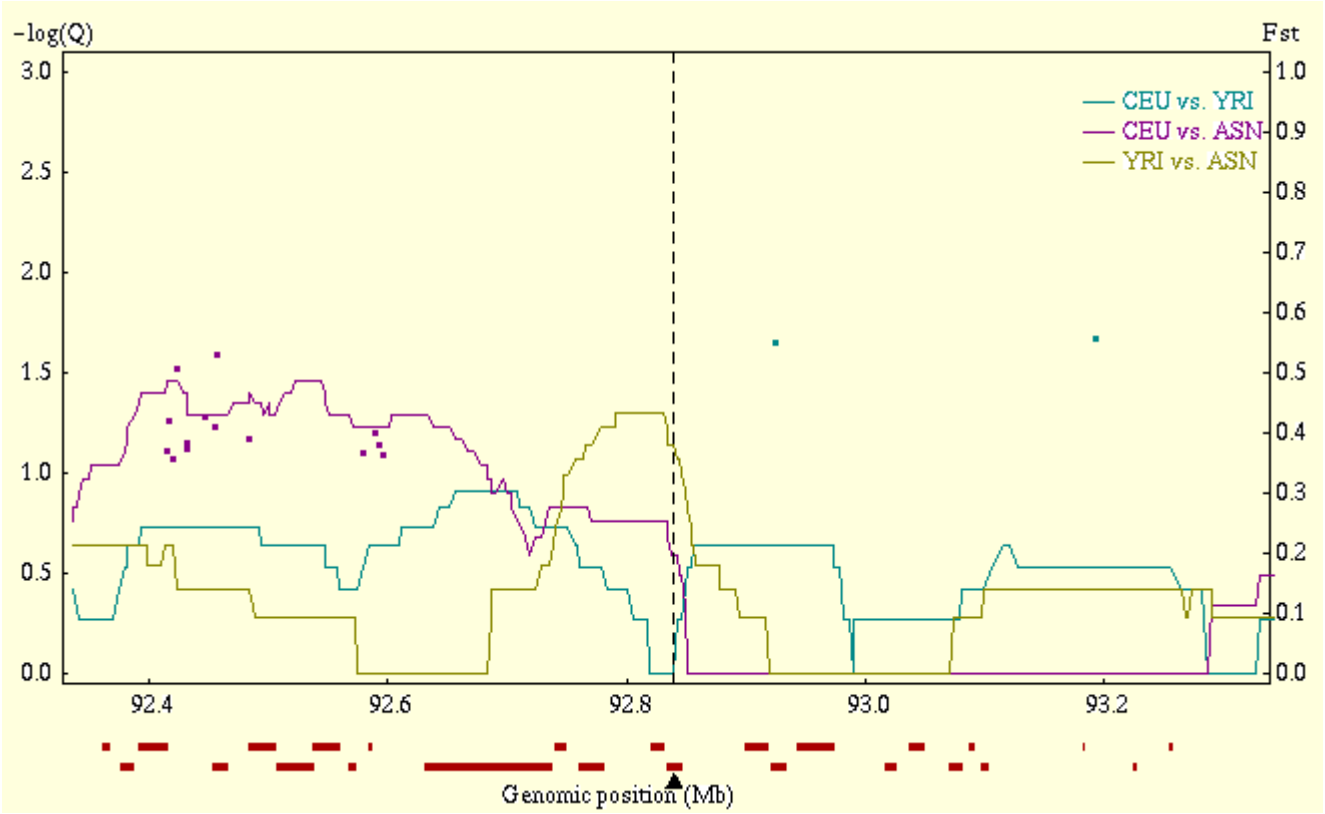
H



D


-log(Q)
3.0
2.5
2.0
1.5
1.0
0.5
0.0
CEU
YRI
ASN
92.4
92.6
92.8
93.0
93.2
Genomic position (Mb)

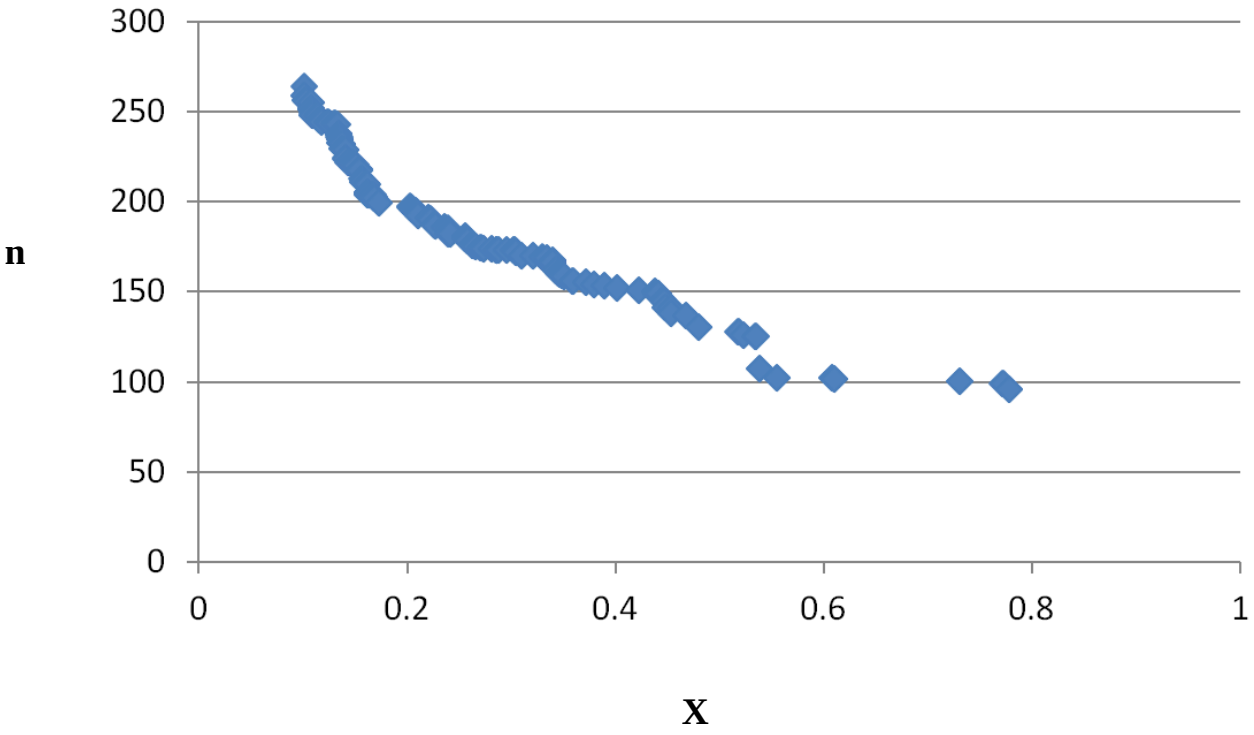
Fst


-log(Q)
Fst
3.0
2.5
2.0
1.5
1.0
0.5
0.0
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
CEU vs. YRI
CEU vs. ASN
YRI vs. ASN
92.4
92.6
92.8
93.0
93.2
Genomic position (Mb)

Genes in the region

Gene ID	Name	Region	CEU	YRI	ASN
90050	C14orf152	92375299 - 92385985	0.999955	0.999955	0.999954
51676	ASB2	92390554 - 92413808	0.999955	0.999955	0.999954
256369	C14orf48	92453683 - 92465250	0.351631	0.999955	0.999954
78990	OTUB2	92482765 - 92505317	0.351631	0.999955	0.999954
57062	DDX24	92507309 - 92537599	0.351631	0.999955	0.999954
122509	FAM14B	92537716 - 92559096	0.351631	0.999955	0.359281
3429	IFI27	92568017 - 92573073	0.351631	0.999955	0.359281
83982	FAM14A	92584162 - 92585964	0.351631	-	0.359281
57718	KIAA1622	92630690 - 92736113	0.999955	0.225666	0.359281
51156	SERPINA10	92739691 - 92749402	0.999955	0.063964	0.359281
866	SERPINA6	92760627 - 92779714	0.999955	0.036685	0.999954
390502	SERPINA2	92820692 - 92830320	0.999955	0.052676	0.999954
5265	SERPINA1	92834751 - 92845165	0.999955	0.052676	0.999954
256394	SERPINA11	92898844 - 92917877	0.999955	0.132766	0.999954
327657	SERPINA9	92921004 - 92932711	0.999955	0.170705	0.999954
145264	SERPINA12	92943661 - 92974222	0.541838	0.999955	0.999954
5267	SERPINA4	93017824 - 93026284	0.252749	0.607928	0.999954
5104	SERPINA5	93037852 - 93049493	0.252749	0.607928	0.539616
12	SERPINA3	93070812 - 93080432	0.252749	0.607928	0.539616
390503	LOC390503	93088276 - 93090906	0.252749	0.607928	0.539616
388007	SERPINA13	93097103 - 93103372	0.252749	0.607928	0.539616
145259	LAMR1P4	93182590 - 93183619	0.999955	0.999955	0.539616

S3 Fig. Estimation of the number of generations (n) computed from the proportion of haplotypes not recombined (X). Each point corresponds to a pairwise combination involving rs28929474 and neighbouring SNPs at a distance ranging from 104 bp to 224 Kb.



S1 Table. Genotyping Quality

Analysis	Cohort	SNP	Call rate (%)	Duplicate concordance rate (%)
Alpha 1-antitrypsin deficiency	BO	rs28929474	99.9	100
	CaPS		99.2	100
	ELSA		98.8	100
	LBC1921		98.6	100
	HAS		94.0	Not available
	HCS		98.1	Not available
	NSHD		N/A	N/A
	WHII		99.1	100
	BO	rs17580	97.2	100
	CaPS		94.5	100
	ELSA		98.0	99.9
	LBC1921		96.3	100
	HAS/HCS		98.1	Not available
	NSHD		N/A	N/A

	WHII		99.2	100
Cystic fibrosis	BO	rs113993960	99.6	100
	CaPS		98.3	100
	ELSA		99.8	100
	LBC1921		98.3	100
	HAS/HCS		98.4	Not available
	NSHD		99.0	100
	WHII		99.4	100
Phenylketonuria	BO	rs5030858	99.5	100
	CaPS		99.7	100
	ELSA		99.2	100
	LBC1921		99.3	100
	HAS/HCS		98.5	Not available
	NSHD		99.2	100
	WHII		99.1	100
	BO	rs5030861	99.9	100
	CaPS		99.6	100

	ELSA		99.2	100
	LBC1921		98.6	100
	HAS/HCS		97.8	Not available
	NSHD		99.0	N/A
	WHII		98.6	100
	BO	rs75193786	99.2	100
	CaPS		98.4	100
	ELSA		99.4	100
	LBC1921		98.6	100
	HAS/HCS		97.5	Not available
	NSHD		97.8	N/A
	WHII		99.1	100
Medium chain acyl-coA dehydrogenase deficiency	BO	rs77931234	99.6	100
	CaPS		98.8	100
	ELSA		99.4	100

	LBC1921		97.0	100
	HAS/HCS		97.8	Not available
	NSHD		98.9	100
	WHII		98.4	100

S2 Table. Sample size by analysis

Outcome	Covariates	Genetic exposure	Number of carriers	Number of non-carriers	Total sample size analysed
weight	age,sex	PI-MS vs PI-MM	1746	15873	17619
height	age,sex	PI-MS vs PI-MM	1739	15902	17641
BMI	age,sex	PI-MS vs PI-MM	1732	15782	17514
FEV1	age,sex	PI-MS vs PI-MM	1490	13721	15211
FVC	age,sex	PI-MS vs PI-MM	1490	13721	15211
FEV1/FVC ratio	age,sex	PI-MS vs PI-MM	1490	13721	15211
FEV1	age,sex, hadj	PI-MS vs PI-MM	1471	13579	15050
FVC	age,sex, hadj	PI-MS vs PI-MM	1471	13579	15050
FEV1/FVC ratio	age,sex, hadj	PI-MS vs PI-MM	1471	13579	15050
FEV1	age,sex, hadj3	PI-MS vs PI-MM	1471	13579	15050
FVC	age,sex, hadj3	PI-MS vs PI-MM	1471	13579	15050
FEV1/FVC ratio	age,sex, hadj3	PI-MS vs PI-MM	1471	13579	15050
COPD	age,sex	PI-MS vs PI-MM	1471	13579	15050
height	age,sex - lfadj	PI-MS vs PI-MM	1471	13579	15050
grip strength	age,sex	PI-MS vs PI-MM	1114	10025	11139
walk speed	age,sex	PI-MS vs PI-MM	1073	10088	11161
ability to balance	age,sex	PI-MS vs PI-MM	994	9209	10203
chair rise speed	age,sex	PI-MS vs PI-MM	804	7547	8351
height	age,sex	PI-MZ vs PI-MM	725	15902	16627
weight	age,sex	PI-MZ vs PI-MM	725	15873	16598
height_trim	age,sex	PI-MZ vs PI-MM	724	15871	16595
BMI	age,sex	PI-MZ vs PI-MM	722	15782	16504
weight_trim	age,sex	PI-MZ vs PI-MM	719	15763	16482
FEV1	exsmok	PI-MS vs PI-MM	665	6135	6800
FVC	exsmok	PI-MS vs PI-MM	665	6135	6800
FEV1/FVC ratio	exsmok	PI-MS vs PI-MM	665	6135	6800
FEV1	age,sex	PI-MZ vs PI-MM	640	13721	14361
FEV1_trim	age,sex	PI-MZ vs PI-MM	640	13697	14337
FVC	age,sex	PI-MZ vs PI-MM	640	13721	14361
FEV1/FVC ratio	age,sex	PI-MZ vs PI-MM	640	13721	14361
FVC_trim	age,sex	PI-MZ vs PI-MM	637	13682	14319
FEV1	age,sex, hadj	PI-MZ vs PI-MM	633	13579	14212
FEV1_trim	age,sex, hadj	PI-MZ vs PI-MM	633	13556	14189
FVC	age,sex, hadj	PI-MZ vs PI-MM	633	13579	14212
FEV1/FVC ratio	age,sex, hadj	PI-MZ vs PI-MM	633	13579	14212
FEV1	age,sex, hadj3	PI-MZ vs PI-MM	633	13579	14212
FVC	age,sex, hadj3	PI-MZ vs PI-MM	633	13579	14212
FEV1/FVC ratio	age,sex, hadj3	PI-MZ vs PI-MM	633	13579	14212
COPD	age,sex	PI-MZ vs PI-MM	633	13579	14212
height	age,sex - lfadj	PI-MZ vs PI-MM	633	13579	14212
height_trim	age,sex - lfadj	PI-MZ vs PI-MM	632	13555	14187
FVC_trim	age,sex, hadj	PI-MZ vs PI-MM	630	13540	14170
FEV1	neversmok	PI-MS vs PI-MM	558	5219	5777
FVC	neversmok	PI-MS vs PI-MM	558	5219	5777
FEV1/FVC ratio	neversmok	PI-MS vs PI-MM	558	5219	5777
TUG speed	age,sex	PI-MS vs PI-MM	481	4381	4862
grip strength	age,sex	PI-MZ vs PI-MM	467	10025	10492
FEV1	age,sex	DeltaF508 carrier	452	14854	15306
FVC	age,sex	DeltaF508 carrier	452	14854	15306
FEV1/FVC ratio	age,sex	DeltaF508 carrier	452	14854	15306
walk speed	age,sex	PI-MZ vs PI-MM	450	10088	10538
FEV1	hadj	DeltaF508 carrier	442	14697	15139
FVC	hadj	DeltaF508 carrier	442	14697	15139
FEV1/FVC ratio	hadj	DeltaF508 carrier	442	14697	15139
COPD	age,sex	DeltaF508 carrier	442	14697	15139
ability to balance	age,sex	PI-MZ vs PI-MM	422	9209	9631
chair rise speed	age,sex	PI-MZ vs PI-MM	348	7547	7895
FEV1	exsmok	PI-MZ vs PI-MM	287	6135	6422
FVC	exsmok	PI-MZ vs PI-MM	287	6135	6422
FEV1/FVC ratio	exsmok	PI-MZ vs PI-MM	287	6135	6422
FEV1	currsmok	PI-MS vs PI-MM	263	2333	2596
FVC	currsmok	PI-MS vs PI-MM	263	2333	2596
FEV1/FVC ratio	currsmok	PI-MS vs PI-MM	263	2333	2596
FEV1	neversmok	PI-MZ vs PI-MM	254	5219	5473
FVC	neversmok	PI-MZ vs PI-MM	254	5219	5473

FEV1/FVC ratio	neversmok	PI-MZ vs PI-MM	254	5219	5473
FEV1	exsmok	DeltaF508 carrier	221	6656	6877
FVC	exsmok	DeltaF508 carrier	221	6656	6877
FEV1/FVC ratio	exsmok	DeltaF508 carrier	221	6656	6877
TUG speed	age,sex	PI-MZ vs PI-MM	218	4381	4599
semantic fluency	age,sex	K304E carrier	208	12753	12961
height	age,sex - under 55	PI-MZ vs PI-MM	204	4348	4552
word recall ability	age,sex	K304E carrier	193	11505	11698
walk speed	age,sex	K304E carrier	177	10541	10718
grip strength	age,sex	K304E carrier	171	11651	11822
ability to balance	age,sex	K304E carrier	156	10650	10806
FEV1	neversmok	DeltaF508 carrier	151	5603	5754
FVC	neversmok	DeltaF508 carrier	151	5603	5754
FEV1/FVC ratio	neversmok	DeltaF508 carrier	151	5603	5754
chair rise speed	age,sex	K304E carrier	134	8720	8854
FEV1	age,sex, in COPD	PI-MS vs PI-MM	118	1118	1236
FEV1	age,sex,hadj,SSadj in COPD	PI-MS vs PI-MM	118	1116	1234
FVC	age,sex, in COPD	PI-MS vs PI-MM	118	1118	1236
FVC	age,sex,hadj,SSadj in COPD	PI-MS vs PI-MM	118	1116	1234
FEV1/FVC ratio	age,sex, in COPD	PI-MS vs PI-MM	118	1118	1236
FEV1/FVC ratio	age,sex,hadj,SSadj in COPD	PI-MS vs PI-MM	118	1116	1234
FEV1	currsmok	PI-MZ vs PI-MM	97	2333	2430
FVC	currsmok	PI-MZ vs PI-MM	97	2333	2430
FEV1/FVC ratio	currsmok	PI-MZ vs PI-MM	97	2333	2430
AH4	age,sex	K304E carrier	94	5232	5326
semantic fluency	age,sex	PAH mutation carrier	81	12708	12789
FEV1	currsmok	DeltaF508 carrier	79	2559	2638
FVC	currsmok	DeltaF508 carrier	79	2559	2638
FEV1/FVC ratio	currsmok	DeltaF508 carrier	79	2559	2638
word recall ability	age,sex	PAH mutation carrier	74	11456	11530
TUG speed	age,sex	K304E carrier	69	5053	5122
grip strength	age,sex	PAH mutation carrier	69	11587	11656
walk speed	age,sex	PAH mutation carrier	67	10525	10592
ability to balance	age,sex	PAH mutation carrier	62	10589	10651
chair rise speed	age,sex	PAH mutation carrier	54	8661	8715
FEV1	age,sex, in COPD	PI-MZ vs PI-MM	51	1118	1169
FEV1	age,sex,hadj,SSadj in COPD	PI-MZ vs PI-MM	51	1116	1167
FVC	age,sex, in COPD	PI-MZ vs PI-MM	51	1118	1169
FVC	age,sex,hadj,SSadj in COPD	PI-MZ vs PI-MM	51	1116	1167
FEV1/FVC ratio	age,sex, in COPD	PI-MZ vs PI-MM	51	1118	1169
FEV1/FVC ratio	age,sex,hadj,SSadj in COPD	PI-MZ vs PI-MM	51	1116	1167
height	age,sex	PI-SZ vs PI-MM	42	15902	15944
FEV1	COPD	DeltaF508 carrier	41	1212	1253
FVC	COPD	DeltaF508 carrier	41	1212	1253
FEV1/FVC ratio	COPD	DeltaF508 carrier	41	1212	1253
height	age,sex	PI-SS vs PI-MM	41	15902	15943
weight	age,sex	PI-SZ vs PI-MM	41	15873	15914
weight	age,sex	PI-SS vs PI-MM	41	15873	15914
BMI	age,sex	PI-SZ vs PI-MM	41	15782	15823
BMI	age,sex	PI-SS vs PI-MM	41	15782	15823
AH4	age,sex	PAH mutation carrier	38	5225	5263
FEV1	age,sex, hadj	PI-SZ vs PI-MM	37	13579	13616
FEV1	age,sex, hadj	PI-SS vs PI-MM	37	13579	13616
FEV1	age,sex	PI-SZ vs PI-MM	37	13721	13758
FVC	age,sex	PI-SZ vs PI-MM	37	13721	13758
FEV1/FVC ratio	age,sex	PI-SZ vs PI-MM	37	13721	13758
FVC	age,sex, hadj	PI-SZ vs PI-MM	37	13579	13616
FEV1/FVC ratio	age,sex, hadj	PI-SZ vs PI-MM	37	13579	13616

FEV1	age,sex, hadj3	PI-SZ vs PI-MM	37	13579	13616
FVC	age,sex, hadj3	PI-SZ vs PI-MM	37	13579	13616
FEV1/FVC ratio	age,sex, hadj3	PI-SZ vs PI-MM	37	13579	13616
height	age,sex - lfadj	PI-SZ vs PI-MM	37	13579	13616
height	age,sex - lfadj	PI-SS vs PI-MM	37	13579	13616
FEV1	age,sex	PI-SS vs PI-MM	37	13721	13758
FVC	age,sex	PI-SS vs PI-MM	37	13721	13758
FEV1/FVC ratio	age,sex	PI-SS vs PI-MM	37	13721	13758
FVC	age,sex, hadj	PI-SS vs PI-MM	37	13579	13616
FEV1/FVC ratio	age,sex, hadj	PI-SS vs PI-MM	37	13579	13616
FEV1	age,sex, hadj3	PI-SS vs PI-MM	37	13579	13616
FVC	age,sex, hadj3	PI-SS vs PI-MM	37	13579	13616
FEV1/FVC ratio	age,sex, hadj3	PI-SS vs PI-MM	37	13579	13616
walk speed	age,sex	PI-SS vs PI-MM	32	10088	10120
grip strength	age,sex	PI-SZ vs PI-MM	31	10025	10056
TUG speed	age,sex	PAH mutation carrier	30	5028	5058
NART	age,sex	PAH mutation carrier	29	4203	4232
walk speed	age,sex	PI-SZ vs PI-MM	28	10088	10116
grip strength	age,sex	PI-SS vs PI-MM	28	10025	10053
chair rise speed	age,sex	PI-SZ vs PI-MM	25	7547	7572
chair rise speed	age,sex	PI-SS vs PI-MM	19	7547	7566
FEV1	exsmok	PI-SZ vs PI-MM	18	6135	6153
FVC	exsmok	PI-SZ vs PI-MM	18	6135	6153
FEV1/FVC ratio	exsmok	PI-SZ vs PI-MM	18	6135	6153
FEV1	neversmok	PI-SZ vs PI-MM	17	5219	5236
FVC	neversmok	PI-SZ vs PI-MM	17	5219	5236
FEV1/FVC ratio	neversmok	PI-SZ vs PI-MM	17	5219	5236
FEV1	exsmok	PI-SS vs PI-MM	17	6135	6152
FVC	exsmok	PI-SS vs PI-MM	17	6135	6152
FEV1/FVC ratio	exsmok	PI-SS vs PI-MM	17	6135	6152
FEV1	neversmok	PI-SS vs PI-MM	16	5219	5235
FVC	neversmok	PI-SS vs PI-MM	16	5219	5235
FEV1/FVC ratio	neversmok	PI-SS vs PI-MM	16	5219	5235
TUG speed	age,sex	PI-SZ vs PI-MM	14	4381	4395
TUG speed	age,sex	PI-SS vs PI-MM	12	4381	4393
height	age,sex	PI-ZZ vs PI-MM	8	15902	15910
weight	age,sex	PI-ZZ vs PI-MM	8	15873	15881
BMI	age,sex	PI-ZZ vs PI-MM	8	15782	15790
FEV1	age,sex, hadj	PI-ZZ vs PI-MM	6	13579	13585
FEV1	age,sex	PI-ZZ vs PI-MM	6	13721	13727
FVC	age,sex	PI-ZZ vs PI-MM	6	13721	13727
height	age,sex - lfadj	PI-ZZ vs PI-MM	6	13579	13585
FEV1/FVC ratio	age,sex	PI-ZZ vs PI-MM	6	13721	13727
FVC	age,sex, hadj	PI-ZZ vs PI-MM	6	13579	13585
FEV1/FVC ratio	age,sex, hadj	PI-ZZ vs PI-MM	6	13579	13585
FEV1	age,sex, hadj3	PI-ZZ vs PI-MM	6	13579	13585
FVC	age,sex, hadj3	PI-ZZ vs PI-MM	6	13579	13585
FEV1/FVC ratio	age,sex, hadj3	PI-ZZ vs PI-MM	6	13579	13585
FEV1	currsmok	PI-SS vs PI-MM	4	2333	2337
FVC	currsmok	PI-SS vs PI-MM	4	2333	2337
FEV1/FVC ratio	currsmok	PI-SS vs PI-MM	4	2333	2337
FEV1	neversmok	PI-ZZ vs PI-MM	3	5219	5222
FVC	neversmok	PI-ZZ vs PI-MM	3	5219	5222
FEV1/FVC ratio	neversmok	PI-ZZ vs PI-MM	3	5219	5222

Hadj – height + height^2 adjusted
Hadj3 – height + height^2 + height^3 adjusted
SSadj – smoking status adjusted
Lfadj – FEV1 and FVC adjusted
Exsmok – analysis in ex smokers
Neversmok-analysis in never smokers
Currsmok – analysis in current smokers
COPD – analysis in individuals classed as having COPD

Alpha 1-antitrypsin deficiency

S3 Table. Alpha 1-antitrypsin deficiency PI status frequencies

Cohort	MM	MS	MZ	SS	SZ	ZZ	Total, N[a]	% female	Mean age[b] (SE)
BO	232	24	6	0	0	0	262	54.96	69.61(0.26)
CaPS	1112	112	61	0	3	1	1289	0	53.79(0.14)
ELSA	4279	450	195	11	16	1	4952	53.92	65.59(0.13)
HAS	155	15	13	1	2	0	186	35.48	76.38(0.17)
HCS	2133	256	111	6	7	0	2513	47	66.09(0.06)
LBC1921	443	40	17	3	1	0	504	58.73	79.06(0.03)
NSHD	1704	191	81	6	2	1	1985	52.34	53.45(0.00)
WHII	3663	402	156	10	6	3	4240	29.41	60.74(0.09)
Combined	13721	1490	640	37	37	6	15931	41.7	62.53(0.07)
Combined (%)	86.13	9.35	4.02	0.23	0.23	0.04	100		

[a] Total number of individuals is total number with a valid PI status, sex, age and lung function
[b] mean age at wave of lung function outcomes

S4 Table. HWE of PI-S (rs17580)

Cohort	T/T	T/A	A/A	Total	HWE P-value
BO	238	24	0	262	0.437
CaPS	1174	115	0	1289	0.094
ELSA	4475	466	11	4952	0.756
HAS	168	17	1	186	0.436
HCS	2244	263	6	2513	0.557
LBC1921	460	41	3	504	0.056
NSHD[a]	1786	193	6	1985	0.746
WHII	3822	408	10	4240	0.797
TOTAL	14367	1527	37	15931	0.593

Based on all individuals with a valid PI status, age, sex and lung function outcomes

[a] Derived from PI classes from isoelectric focusing(17)

S5 Table. HWE of PI-Z (rs28929474)

Cohort	G/G	G/A	A/A	Total	HWE P-value
BO	256	6	0	262	0.851
CaPS	1224	64	1	1289	0.862
ELSA	4740	211	1	4952	0.384
HAS	171	15	0	186	0.567
HCS	2395	118	0	2513	0.228
LBC1921	486	18	0	504	0.683
NSHD[a]	1901	83	1	1985	0.923
WHII	4075	162	3	4240	0.291
TOTAL	15248	677	6	15931	0.587
Based on all individuals with a valid PI status, age, sex and lung function outcomes [a] Derived from PI classes from isoelectric focusing(17)					

S6 Table. Association of PI status with lung function, adjusted for age and sex

Outcome	Cohort	Regression Coefficient (95% CI)				
		MS VS. MM	MZ VS. MM	SS VS. MM	SZ VS. MM	ZZ VS. MM
Maximum FEV1	BO	-0.05(-0.36,0.27)	0.45(-0.14,1.04)	[a]		
	CaPS	0.07(-0.10,0.24)	0.11(-0.12,0.33)			
	ELSA	0.05(-0.03,0.12)	0.12*(0.01,0.23)			
	HAS	-0.49*(-0.94,-0.05)	0.08(-0.39,0.55)			
	HCS	-0.01(-0.10,0.09)	0.05(-0.10,0.19)			
	LBC1921	-0.06(-0.31,0.19)	0.16(-0.22,0.54)			
	NSHD	0.02(-0.09,0.13)	0.18*(0.01,0.35)			
	WHII	0.05(-0.03,0.12)	0.18**(0.06,0.29)			
	Combined Fixed Effect	0.03(-0.01,0.07)	0.13****(0.07,0.19)	0.01(-0.23,0.26)	0.06(-0.18,0.31)	-0.29(-0.89,0.32)
	Combined Random Effect	0.03(-0.01,0.07)	0.13****(0.07,0.19)	0.02(-0.22,0.26)	0.07(-0.18,0.31)	-0.35(-0.95,0.25)
	Estimated variance of random effect	1.20e-13(0.00e+00,.)	1.09e-19(0.00e+00,.)	7.37e-12(4.15e-26,1.31e+03)	2.04e-14(5.46e-28,7.65e-01)	5.65e-10(0.00e+00,.)
Maximum FVC	BO	0.03(-0.26,0.32)	0.43(-0.11,0.97)	[a]		
	CaPS	0.08(-0.10,0.26)	0.15(-0.09,0.39)			
	ELSA	0.02(-0.05,0.09)	0.12*(0.01,0.22)			
	HAS	-0.35(-0.73,0.03)	0.33(-0.09,0.74)			
	HCS	-0.02(-0.11,0.07)	0.14*(0.01,0.27)			
	LBC1921	-0.07(-0.32,0.17)	0.20(-0.17,0.56)			
	NSHD	0.02(-0.09,0.13)	0.20*(0.04,0.37)			
	WHII	-0.01(-0.08,0.07)	0.19**(0.07,0.30)			
	Combined Fixed Effect	0.01(-0.03,0.04)	0.16****(0.10,0.22)	-0.12(-0.35,0.12)	0.07(-0.17,0.31)	-0.13(-0.71,0.46)
	Combined Random Effect	0.00(-0.04,0.04)	0.16****(0.10,0.22)	-0.11(-0.34,0.12)	0.07(-0.17,0.30)	-0.18(-0.77,0.41)
	Estimated variance of random effect	3.76e-15(1.01e-28,1.40e-01)	3.53e-19(0.00e+00,.)	4.07e-14(8.72e-29,1.90e+01)	9.77e-13(1.11e-31,8.59e+06)	3.87e-14(2.43e-32,6.18e+04)

FEV1/FVC ratio	BO	-0.31(-0.73,0.11)	0.15(-0.64,0.95)	[a]		
	CaPS	0.02(-0.15,0.18)	0.03(-0.20,0.25)			
	ELSA	0.04(-0.06,0.13)	0.01(-0.14,0.15)			
	HAS	-0.34(-0.89,0.22)	-0.29(-0.85,0.27)			
	HCS	0.03(-0.10,0.15)	-0.13(-0.32,0.06)			
	LBC1921	-0.02(-0.34,0.31)	-0.15(-0.65,0.34)			
	NSHD	-0.03(-0.18,0.12)	-0.10(-0.32,0.12)			
	WHII	0.11*(0.01,0.21)	0.02(-0.14,0.18)			
	Combined Fixed Effect	0.03(-0.02,0.09)	-0.04(-0.11,0.04)	0.23(-0.09,0.55)	0.01(-0.31,0.32)	-0.69(-1.47,0.10)
	Combined Random Effect	0.03(-0.02,0.09)	-0.03(-0.11,0.04)	0.23(-0.09,0.55)	0.01(-0.31,0.32)	-0.55(-1.43,0.34)
	Estimated variance of random effect	3.41e-17(4.23e-30,2.75e-04)	1.34e-18(9.21e-33,1.95e-04)	4.77e-17(7.55e-32,3.02e-02)	2.73e-15(2.55e-32,2.92e+02)	1.60e-01(2.99e-04,8.50e+01)

Outcomes z-scored within cohorts
*p<0.05, **p<0.01, ***p<0.001, ****p<0.0001
[a] Cohort specific estimates are suppressed for PI-SS, SZ and ZZ due to the small number of carriers

S7 Table. Association of PI status with transformed lung function, adjusted for age and sex

Outcome	Cohort	Regression Coefficient (95% CI)				
		MS VS. MM	MZ VS. MM	SS VS. MM	SZ VS. MM	ZZ VS. MM
Sqrt(Maximum FVC)	BO	-0.02(-0.32,0.27)	0.43(-0.12,0.97)	[a]		
	CaPS	0.08(-0.10,0.25)	0.13(-0.10,0.36)			
	ELSA	0.02(-0.05,0.09)	0.10(-0.00,0.21)			
	HAS	-0.35(-0.73,0.03)	0.31(-0.10,0.72)			
	HCS	-0.01(-0.10,0.07)	0.12(-0.00,0.25)			
	LBC1921	-0.08(-0.33,0.17)	0.18(-0.19,0.55)			
	NSHD	0.02(-0.09,0.13)	0.20*(0.04,0.37)			
	WHII	-0.01(-0.08,0.06)	0.18**(0.07,0.29)			
	Combined Fixed Effect	0.00(-0.04,0.04)	0.15****(0.09,0.21)	-0.11(-0.34,0.13)	0.07(-0.16,0.31)	-0.11(-0.69,0.46)
	Combined Random Effect	0.00(-0.04,0.04)	0.15****(0.09,0.21)	-0.10(-0.33,0.13)	0.07(-0.16,0.31)	-0.17(-0.75,0.42)
	Estimated variance of random effect	1.03e-14(5.17e-28,2.07e-01)	4.11e-17(9.66e-31,1.75e-03)	1.13e-14(1.77e-30,7.23e+01)	1.80e-12(0.00e+00,.)	4.18e-09(3.76e-27,4.66e+09)
(FEV1/FVC ratio)^3	BO	-0.30(-0.72,0.13)	0.07(-0.75,0.88)	[a]		
	CaPS	-0.01(-0.18,0.17)	0.02(-0.22,0.25)			
	ELSA	0.03(-0.07,0.12)	0.02(-0.12,0.16)			
	HAS	-0.27(-0.81,0.27)	-0.37(-0.94,0.19)			
	HCS	0.02(-0.11,0.14)	-0.08(-0.27,0.11)			
	LBC1921	-0.06(-0.38,0.26)	-0.14(-0.64,0.35)			
	NSHD	-0.03(-0.18,0.12)	-0.14(-0.36,0.08)			
	WHII	0.10*(0.00,0.21)	0.03(-0.13,0.19)			
	Combined Fixed Effect	0.03(-0.03,0.08)	-0.03(-0.11,0.05)	0.19(-0.13,0.50)	-0.07(-0.39,0.25)	-0.61(-1.40,0.17)
	Combined Random Effect	0.03(-0.03,0.08)	-0.03(-0.10,0.05)	0.19(-0.13,0.50)	-0.07(-0.38,0.25)	-0.57(-1.35,0.21)
	Estimated variance of random effect	1.74e-17(5.35e-31,5.65e-04)	2.18e-17(2.82e-30,1.69e-04)	8.96e-20(7.43e-34,1.08e-05)	4.48e-06(3.06e-19,6.56e+07)	7.56e-10(3.04e-32,1.88e+13)

Outcomes z-scored within cohorts
*p<0.05, **p<0.01, ***p<0.001, ****p<0.0001
[a] Cohort specific estimates are suppressed for PI-SS, SZ and ZZ due to the small number of carriers

S8 Table. Association of PI-MZ with lung function with +/-3SDs removed from outcome before standardisation, adjusted for age and sex

Outcome	Cohort	Regression Coefficient (95% CI)
Maximum FEV1	Combined Fixed Effect	0.14****(0.08,0.20)
	Combined Random Effect	0.14****(0.08,0.20)
	Estimated variance of random effect	2.19e-17(2.33e-29,2.05e-05)
Maximum FVC	Combined Fixed Effect	0.16****(0.10,0.22)
	Combined Random Effect	0.16****(0.10,0.22)
	Estimated variance of random effect	1.84e-17(9.28e-33,3.66e-02)
Outcomes z-scored within cohorts *p<0.05, **p<0.01, ***p<0.001, ****p<0.0001		

S9 Table. Association of PI-MZ with lung function, raw values without z-scoring, adjusted for age and sex

Outcome	Cohort	Regression Coefficient (95% CI) (Litres)
Maximum FEV1 (L)	BO	0.33(-0.10,0.75)
	CaPS	0.08(-0.09,0.25)
	ELSA	0.10*(0.01,0.19)
	HAS	0.05(-0.25,0.35)
	HCS	0.03(-0.07,0.13)
	LBC1921	0.10(-0.14,0.34)
	NSHD	0.12*(0.01,0.24)
	WHII	0.14**(0.05,0.23)
	Combined Fixed Effect	0.10****(0.06,0.15)
	Combined Random Effect	0.10****(0.05,0.14)
	Estimated variance of random effect	6.77e-21(1.15e-32,3.97e-09)
Maximum FVC (L)	BO	0.39(-0.10,0.89)
	CaPS	0.12(-0.07,0.30)
	ELSA	0.13*(0.02,0.24)
	HAS	0.26(-0.07,0.58)
	HCS	0.13*(0.01,0.25)
	LBC1921	0.14(-0.13,0.42)
	NSHD	0.18*(0.04,0.32)
	WHII	0.18**(0.07,0.29)
	Combined Fixed Effect	0.15****(0.10,0.21)
	Combined Random Effect	0.15****(0.10,0.21)
	Estimated variance of random effect	4.75e-21(5.59e-36,4.04e-06)
*p<0.05, **p<0.01, ***p<0.001, ****p<0.0001		

S10 Table. Association of PI status with lung function, adjusted for age and sex, in current smokers

Outcome	Cohort	Regression Coefficient (95% CI)		
		MS VS. MM	MZ VS. MM	SS VS. MM
Maximum FEV1	Combined Fixed Effect	0.04(-0.06,0.14)	0.09(-0.07,0.25)	-0.30(-1.08,0.48)
	Combined Random Effect	0.03(-0.07,0.13)	0.09(-0.07,0.25)	-0.31(-1.30,0.69)
	Estimated variance of random effect	1.58e-24(7.90e-45,3.15e-04)	7.96e-22(2.05e-41,3.10e-02)	5.09e-01(2.86e-02,9.04e+00)
Maximum FVC	Combined Fixed Effect	0.00(-0.09,0.10)	0.12(-0.04,0.28)	-0.66(-1.42,0.11)
	Combined Random Effect	-0.01(-0.11,0.08)	0.14(-0.02,0.29)	-0.67(-1.69,0.35)
	Estimated variance of random effect	8.57e-23(1.06e-47,6.95e+02)	3.94e-22[a]	5.89e-01(4.30e-02,8.08e+00)
FEV1/FVC ratio	Combined Fixed Effect	0.07(-0.07,0.20)	-0.06(-0.28,0.15)	0.77(-0.29,1.83)
	Combined Random Effect	0.06(-0.07,0.19)	-0.02(-0.23,0.18)	0.74(-0.28,1.76)
	Estimated variance of random effect	2.75e-11(0.00e+00,.)	1.42e-13(5.60e-33,3.59e+06)	9.39e-20(5.69e-39,1.55e+00)

Outcomes z-scored within cohorts
Cohort specific estimates are suppressed due to the small number of carriers
Meta-analysis not possible for PI-SZ and ZZ as there were fewer than three cohorts with adequate data
*p<0.05, **p<0.01, ***p<0.001, ****p<0.0001
[a] Standard error of the variance of the random effect not estimable

S11 Table. Association of PI status with transformed lung function, adjusted for age and sex, in current smokers

Outcome	Cohort	Regression Coefficient (95% CI)		
		MS VS. MM	MZ VS. MM	SS VS. MM
Sqrt(Maximum FVC)	Combined Fixed Effect	0.01(-0.09,0.10)	0.11(-0.04,0.27)	-0.72(-1.48,0.05)
	Combined Random Effect	-0.01(-0.11,0.08)	0.13(-0.02,0.29)	-0.70(-1.78,0.38)
	Estimated variance of random effect	1.05e-17(4.73e-36,2.33e+01)	4.69e-24(5.98e-42,3.67e-06)	6.98e-01(6.46e-02,7.55e+00)
(FEV1/FVC ratio)^3	Combined Fixed Effect	0.06(-0.07,0.19)	-0.03(-0.24,0.18)	0.81(-0.19,1.82)
	Combined Random Effect	0.05(-0.08,0.18)	-0.01(-0.21,0.19)	0.80(-0.21,1.81)
	Estimated variance of random effect	3.60e-15(5.28e-28,2.46e-02)	6.98e-16(7.32e-32,6.66e+00)	1.73e-12(2.54e-31,1.17e+07)
Outcomes z-scored within cohorts Cohort specific estimates are suppressed due to the small number of carriers Meta-analysis not possible for PI-SZ and ZZ as there were fewer than three cohorts with adequate data *p<0.05, **p<0.01, ***p<0.001, ****p<0.0001				

S1 Table. Association of PI status with lung function, adjusted for age and sex, in ex smokers

Outcome	Cohort	Regression Coefficient (95% CI)			
		MS VS. MM	MZ VS. MM	SS VS. MM	SZ VS. MM
Maximum FEV1	Combined Fixed Effect	-0.01(-0.07,0.05)	0.13**(0.04,0.22)	0.18(-0.18,0.54)	0.26(-0.09,0.61)
	Combined Random Effect	-0.00(-0.06,0.06)	0.15*(0.03,0.26)	0.21(-0.15,0.57)	NOCONVERGENCE[a]
	Estimated variance of random effect	1.85e-22(1.89e-36,1.81e-08)	6.60e-03(4.74e-04,9.19e-02)	3.72e-14(0.00e+00,.)	NOCONVERGENCE[a]
Maximum FVC	Combined Fixed Effect	-0.02(-0.08,0.03)	0.19****(0.11,0.28)	0.04(-0.31,0.39)	0.29(-0.04,0.63)
	Combined Random Effect	-0.02(-0.08,0.03)	0.20****(0.11,0.28)	0.06(-0.29,0.41)	0.28(-0.06,0.61)
	Estimated variance of random effect	8.94e-16(1.09e-29,7.31e-02)	3.94e-18(3.65e-34,4.25e-02)	9.69e-14(0.00e+00,.)	1.88e-12(4.93e-40,7.19e+15)
FEV1/FVC ratio	Combined Fixed Effect	0.04(-0.04,0.11)	-0.06(-0.17,0.05)	0.21(-0.25,0.67)	-0.06(-0.51,0.38)
	Combined Random Effect	0.04(-0.04,0.12)	-0.06(-0.19,0.08)	0.21(-0.25,0.68)	-0.06(-0.50,0.38)
	Estimated variance of random effect	6.41e-16(1.93e-30,2.14e-01)	7.46e-03(2.53e-04,2.19e-01)	4.70e-14(6.50e-32,3.39e+04)	8.66e-17(3.55e-35,2.11e+02)

Outcomes z-scored within cohorts
Cohort specific estimates are suppressed due to the small number of carriers
Meta-analysis not possible for PI- ZZ as there were fewer than three cohorts with adequate data
*p<0.05, **p<0.01, ***p<0.001, ****p<0.0001
[a] Random effects model failed to converge

S2 Table. Association of PI status with transformed lung function, adjusted for age and sex, in ex smokers

Outcome	Cohort	Regression Coefficient (95% CI)			
		MS VS. MM	MZ VS. MM	SS VS. MM	SZ VS. MM
Sqrt(Maximum FVC)	Combined Fixed Effect	-0.03(-0.08,0.03)	0.17****(0.08,0.25)	0.07(-0.27,0.41)	0.27(-0.06,0.60)
	Combined Random Effect	-0.03(-0.08,0.03)	0.17****(0.09,0.26)	0.10(-0.25,0.44)	0.26(-0.07,0.60)
	Estimated variance of random effect	2.89e-15(9.33e-30,8.94e-01)	5.26e-22(1.08e-39,2.57e-04)	1.01e-13(0.00e+00,.)	1.79e-14(3.31e-33,9.64e+04)
(FEV1/FVC ratio)^3	Combined Fixed Effect	0.04(-0.04,0.11)	-0.05(-0.17,0.06)	0.15(-0.31,0.61)	-0.14(-0.59,0.31)
	Combined Random Effect	0.04(-0.04,0.12)	-0.05(-0.19,0.08)	0.16(-0.31,0.63)	-0.14(-0.59,0.31)
	Estimated variance of random effect	6.21e-18(1.73e-33,2.23e-02)	5.30e-03(6.21e-05,4.52e-01)	4.43e-13(4.36e-29,4.51e+03)	9.95e-17(6.79e-33,1.46e+00)
Outcomes z-scored within cohorts Cohort specific estimates are suppressed due to the small number of carriers Meta-analysis not possible for PI- ZZ as there were fewer than three cohorts with adequate data *p<0.05, **p<0.01, ***p<0.001, ****p<0.0001					

S3 Table. Association of PI status with lung function, adjusted for age and sex, in never smokers

Outcome	Cohort	Regression Coefficient (95% CI)				
		MS VS. MM	MZ VS. MM	SS VS. MM	SZ VS. MM	ZZ VS. MM
Maximum FEV1	Combined Fixed Effect	0.07*(0.01,0.13)	0.12**(0.03,0.21)	-0.11(-0.45,0.24)	-0.17(-0.51,0.17)	0.41(-0.39,1.22)
	Combined Random Effect	0.07*(0.01,0.13)	0.11*(0.03,0.20)	-0.11(-0.46,0.23)	-0.14(-0.48,0.19)	0.40(-0.44,1.25)
	Estimated variance of random effect	7.06e-20(4.51e-35,1.10e-04)	4.35e-22(9.47e-38,2.00e-06)	4.86e-18(1.28e-36,1.85e+01)	1.01e-20(2.68e-37,3.79e-04)	1.14e-16(0.00e+00,.)
Maximum FVC	Combined Fixed Effect	0.05(-0.01,0.11)	0.12**(0.04,0.21)	-0.19(-0.53,0.16)	-0.14(-0.48,0.19)	0.45(-0.35,1.25)
	Combined Random Effect	0.05(-0.01,0.11)	0.12**(0.03,0.21)	-0.18(-0.52,0.16)	-0.13(-0.45,0.20)	0.46(-0.38,1.30)
	Estimated variance of random effect	1.07e-23(7.42e-39,1.53e-08)	7.43e-19(1.26e-30,4.38e-07)	1.14e-19(1.06e-38,1.22e+00)	7.52e-22(3.15e-42,1.80e-01)	8.56e-20(2.12e-40,3.47e+01)
FEV1/FVC ratio	Combined Fixed Effect	0.03(-0.06,0.11)	-0.02(-0.14,0.10)	0.14(-0.32,0.60)	-0.02(-0.47,0.42)	-0.05(-1.11,1.02)
	Combined Random Effect	0.01(-0.07,0.09)	-0.02(-0.14,0.10)	0.10(-0.36,0.56)	0.01(-0.41,0.43)	0.01(-1.03,1.05)
	Estimated variance of random effect	7.19e-04(1.55e-09,3.34e+02)	2.25e-03(1.28e-07,3.96e+01)	6.94e-22(4.79e-41,1.00e-02)	4.75e-12(3.46e-30,6.51e+06)	1.19e-15(6.57e-39,2.17e+08)

Outcomes z-scored within cohorts
Cohort specific estimates are suppressed due to the small number of carriers
*p<0.05, **p<0.01, ***p<0.001, ****p<0.0001

S4 Table. Association of PI status with transformed lung function, adjusted for age and sex, in never smokers

Outcome	Cohort	Regression Coefficient (95% CI)				
		MS VS. MM	MZ VS. MM	SS VS. MM	SZ VS. MM	ZZ VS. MM
Sqrt(Maximum FVC)	Combined Fixed Effect	0.05(-0.01,0.11)	0.12**(0.04,0.21)	-0.18(-0.53,0.16)	-0.12(-0.46,0.21)	0.51(-0.29,1.30)
	Combined Random Effect	0.05(-0.01,0.11)	0.12**(0.03,0.20)	-0.20(-0.53,0.14)	-0.10(-0.42,0.22)	0.52(-0.31,1.35)
	Estimated variance of random effect	1.07e-23(1.46e-38,7.87e-09)	6.42e-19(8.06e-35,5.12e-03)	1.70e-20(3.41e-38,8.53e-03)	3.74e-21(4.78e-36,2.92e-06)	9.35e-18(5.78e-38,1.51e+03)
$(FEV1/FVC\ ratio)^3$	Combined Fixed Effect	0.01(-0.08,0.09)	-0.02(-0.14,0.10)	0.09(-0.38,0.56)	-0.09(-0.55,0.36)	-0.17(-1.26,0.92)
	Combined Random Effect	-0.00(-0.09,0.08)	-0.03(-0.16,0.10)	0.05(-0.42,0.53)	-0.13(-0.63,0.37)	-0.14(-1.21,0.93)
	Estimated variance of random effect	7.63e-04(7.09e-10,8.22e+02)	3.46e-03(2.47e-06,4.85e+00)	7.76e-17(6.18e-35,9.74e+01)	5.94e-02(1.54e-04,2.28e+01)	5.75e-17(3.92e-36,8.43e+02)
Outcomes z-scored within cohorts Cohort specific estimates are suppressed due to the small number of carriers *p<0.05, **p<0.01, ***p<0.001, ****p<0.0001						

S5 Table. Association of PI status with lung function, adjusted for age and sex, in COPD cases

		Regression Coefficient (95% CI)	
Outcome	Cohort	MS VS. MM	MZ VS. MM
Maximum FEV1	Combined Fixed Effect	0.06(-0.05,0.18)	0.01(-0.15,0.18)
	Combined Random Effect	0.04(-0.07,0.15)	-0.00(-0.17,0.16)
	Estimated variance of random effect	5.22e-20(1.89e-33,1.44e-06)	1.87e-20(2.05e-37,1.71e-03)
Maximum FVC	Combined Fixed Effect	-0.04(-0.18,0.09)	0.12(-0.08,0.32)
	Combined Random Effect	-0.05(-0.18,0.08)	0.10(-0.10,0.29)
	Estimated variance of random effect	4.09e-18(5.45e-32,3.07e-04)	3.97e-15(2.48e-30,6.34e+00)
FEV1/FVC ratio	Combined Fixed Effect	0.21*(0.01,0.41)	-0.10(-0.41,0.20)
	Combined Random Effect	0.20*(0.01,0.38)	-0.06(-0.35,0.24)
	Estimated variance of random effect	1.47e-18(0.00e+00,.)	4.19e-23(2.04e-37,8.62e-09)

Outcomes z-scored within cohorts
Cohort specific estimates are suppressed due to the small number of carriers
Meta-analyses for PI-SS, SZ and ZZ were not possible as there were fewer than three cohorts with adequate data
p<0.05, **p<0.01, ***p<0.001, ****p<0.0001

S6 Table. Association of PI status with lung function adjusted for age, sex, height and height-squared

Outcome	Cohort	Regression Coefficient (95% CI)				
		MS VS. MM	MZ VS. MM	SS VS. MM	SZ VS. MM	ZZ VS. MM
Maximum FEV1	BO	-0.16(-0.46,0.14)	0.17(-0.40,0.74)	[a]		
	CaPS	0.05(-0.11,0.21)	0.09(-0.12,0.30)			
	ELSA	0.05(-0.02,0.12)	0.06(-0.05,0.16)			
	HAS	-0.41(-0.87,0.05)	0.11(-0.34,0.56)			
	HCS	0.00(-0.09,0.09)	-0.01(-0.15,0.12)			
	LBC1921	-0.08(-0.31,0.16)	0.09(-0.27,0.44)			
	NSHD	-0.02(-0.12,0.08)	0.08(-0.08,0.23)			
	WHII	0.04(-0.03,0.10)	0.11(-0.00,0.21)			
	Combined Fixed Effect	0.02(-0.02,0.06)	0.07*(0.01,0.12)	0.03(-0.20,0.25)	0.02(-0.21,0.24)	-0.31(-0.87,0.26)
	Combined Random Effect	0.02(-0.02,0.06)	0.07*(0.01,0.12)	0.03(-0.20,0.25)	0.03(-0.20,0.26)	-0.35(-0.91,0.20)
	Estimated variance of random effect	1.32e-13(1.51e-27,1.15e+01)	1.03e-17(1.80e-32,5.87e-03)	4.91e-14(7.47e-31,3.22e+03)	1.59e-23(5.06e-38,5.00e-09)	1.53e-06(8.51e-32,2.74e+19)
Maximum FVC	BO	-0.11(-0.37,0.15)	0.06(-0.43,0.56)	[a]		
	CaPS	0.07(-0.09,0.23)	0.13(-0.08,0.34)			
	ELSA	0.03(-0.04,0.09)	0.04(-0.06,0.14)			
	HAS	-0.30(-0.69,0.09)	0.35(-0.04,0.74)			
	HCS	-0.01(-0.08,0.07)	0.07(-0.05,0.18)			
	LBC1921	-0.09(-0.31,0.14)	0.10(-0.23,0.44)			
	NSHD	-0.02(-0.12,0.07)	0.08(-0.06,0.23)			
	WHII	-0.02(-0.08,0.05)	0.10(-0.00,0.20)			
	Combined Fixed Effect	-0.00(-0.04,0.03)	0.08**(0.03,0.13)	-0.11(-0.32,0.11)	0.02(-0.20,0.23)	-0.16(-0.69,0.36)
	Combined Random Effect	-0.01(-0.04,0.03)	0.08**(0.03,0.13)	-0.10(-0.30,0.11)	0.01(-0.20,0.22)	-0.19(-0.71,0.34)
	Estimated variance of random effect	6.12e-14(3.19e-29,1.17e+02)	1.25e-13(5.20e-27,3.01e+00)	9.71e-17(8.57e-33,1.10e+00)	8.32e-19(3.11e-33,2.22e-04)	1.31e-14(4.02e-33,4.26e+04)

FEV1/FVC ratio	BO	-0.28(-0.70,0.14)	0.26(-0.55,1.07)	[a]		
	CaPS	0.01(-0.16,0.18)	0.02(-0.20,0.25)			
	ELSA	0.04(-0.06,0.14)	0.02(-0.12,0.17)			
	HAS	-0.20(-0.80,0.40)	-0.28(-0.84,0.28)			
	HCS	0.03(-0.10,0.15)	-0.12(-0.31,0.07)			
	LBC1921	-0.01(-0.34,0.31)	-0.13(-0.63,0.36)			
	NSHD	-0.01(-0.16,0.13)	-0.06(-0.28,0.16)			
	WHII	0.11*(0.01,0.21)	0.03(-0.13,0.19)			
	Combined Fixed Effect	0.04(-0.01,0.09)	-0.02(-0.10,0.06)	0.23(-0.09,0.55)	0.02(-0.30,0.33)	-0.67(-1.46,0.11)
	Combined Random Effect	0.04(-0.01,0.09)	-0.02(-0.10,0.06)	0.23(-0.09,0.55)	0.02(-0.30,0.33)	-0.55(-1.41,0.32)
	Estimated variance of random effect	2.26e-21(9.99e-37,5.13e-06)	2.27e-17(1.70e-30,3.03e-04)	1.15e-17(1.90e-33,6.99e-02)	1.19e-19(1.19e-37,1.19e-01)	1.23e-01(5.10e-05,2.97e+02)

Outcomes z-scored within cohorts
*p<0.05, **p<0.01, ***p<0.001, ****p<0.0001
[a] Cohort specific estimates are suppressed for PI-SS, SZ and ZZ due to the small number of carriers

S7 Table. Association of PI status with lung function adjusted for age, sex, height and height-squared, outcomes transformed

Outcome	Cohort	Regression Coefficient (95% CI)				
		MS VS. MM	MZ VS. MM	SS VS. MM	SZ VS. MM	ZZ VS. MM
Sqrt(Maximum FVC)	BO	-0.15(-0.42,0.11)	0.07(-0.44,0.58)	[b]		
	CaPS	0.06(-0.09,0.22)	0.11(-0.10,0.32)			
	ELSA	0.02(-0.04,0.09)	0.03(-0.07,0.13)			
	HAS	-0.29(-0.68,0.10)	0.33(-0.06,0.72)			
	HCS	-0.00(-0.08,0.07)	0.06(-0.06,0.17)			
	LBC1921	-0.10(-0.33,0.13)	0.10(-0.25,0.44)			
	NSHD	-0.02(-0.11,0.08)	0.09(-0.06,0.23)			
	WHII	-0.02(-0.08,0.04)	0.09(-0.00,0.19)			
	Combined Fixed Effect	-0.00(-0.04,0.03)	0.07**(0.02,0.12)	-0.09(-0.31,0.12)	0.02(-0.19,0.23)	-0.17(-0.69,0.36)
	Combined Random Effect	-0.01(-0.04,0.03)	0.07**(0.02,0.12)	-0.09(-0.29,0.12)	0.02(-0.19,0.23)	-0.20(-0.71,0.32)
	Estimated variance of random effect	2.16e-20(3.13e-36,1.49e-04)	6.57e-18(0.00e+00,.)	8.37e-16(2.36e-30,2.97e-01)	3.40e-17(0.00e+00,.)	1.51e-16(6.00e-34,3.79e+01)
$(FEV1/FVC ratio)^3$	BO	-0.25(-0.68,0.18)	0.21(-0.62,1.04)	[b]		
	CaPS	-0.01(-0.19,0.16)	0.01(-0.22,0.25)			
	ELSA	0.04(-0.06,0.13)	0.04(-0.10,0.19)			
	HAS	-0.11(-0.69,0.47)	-0.37(-0.93,0.20)			
	HCS	0.01(-0.11,0.14)	-0.06(-0.25,0.12)			
	LBC1921	-0.06(-0.38,0.27)	-0.12(-0.61,0.37)			
	NSHD	-0.02(-0.16,0.13)	-0.10(-0.32,0.12)			
	WHII	0.10*(0.00,0.21)	0.05(-0.11,0.21)			
	Combined Fixed Effect	0.03(-0.02,0.08)	-0.01(-0.08,0.07)	0.19(-0.13,0.50)	-0.06(-0.37,0.26)	-0.60(-1.38,0.19)

	Combined Random Effect	0.03(-0.02,0.08)	-0.01(-0.08,0.07)	0.18(-0.13,0.50)	NOCONVERGENCE[a]	-0.56(-1.33,0.22)
	Estimated variance of random effect	2.70e-21(3.30e-35,2.22e-07)	1.26e-17(7.38e-31,2.15e-04)	1.82e-19(1.20e-35,2.78e-03)	NOCONVERGENCE[a]	5.89e-12(9.75e-32,3.56e+08)

Outcomes z-scored within cohorts
*p<0.05, **p<0.01, ***p<0.001, ****p<0.0001
[a] Random effects model failed to converge
[b] Cohort specific estimates are suppressed for PI-SS, SZ and ZZ due to the small number of carriers

S8 Table. Association of PI status with lung function, adjusted for age, sex, height, height-squared and smoking status in COPD cases

		Regression Coefficient (95% CI)	
Outcome	Cohort	MS VS. MM	MZ VS. MM
Maximum FEV1	Combined Fixed Effect	0.06(-0.05,0.17)	-0.02(-0.18,0.14)
	Combined Random Effect	0.04(-0.07,0.14)	-0.05(-0.20,0.11)
	Estimated variance of random effect	1.86e-20(4.01e-36,8.62e-05)	1.06e-18(1.27e-32,8.95e-05)
Maximum FVC	Combined Fixed Effect	-0.06(-0.19,0.06)	0.06(-0.13,0.25)
	Combined Random Effect	-0.06(-0.18,0.06)	0.04(-0.14,0.21)
	Estimated variance of random effect	3.06e-24(1.30e-36,7.19e-12)	6.24e-21(2.66e-37,1.47e-04)
FEV1/FVC ratio	Combined Fixed Effect	0.23*(0.02,0.43)	-0.09(-0.40,0.21)
	Combined Random Effect	0.22*(0.04,0.41)	-0.05(-0.35,0.24)
	Estimated variance of random effect	1.63e-21(6.35e-35,4.20e-08)	1.25e-18(4.60e-33,3.42e-04)

Outcomes z-scored within cohorts
Cohort specific estimates are suppressed due to the small number of carriers
Meta-analyses for PI-SS, SZ and ZZ were not possible as there were fewer than three cohorts with adequate data*p<0.05, **p<0.01, ***p<0.001, ****p<0.0001

S20 Table. Association of PI status with lung function, adjusted for age and sex, in COPD cases with outcomes transformed

		Regression Coefficient (95% CI)	
Outcome	Cohort	MS VS. MM	MZ VS. MM
Sqrt(Maximum FVC)	Combined Fixed Effect	-0.05(-0.19,0.09)	0.12(-0.09,0.33)
	Combined Random Effect	-0.05(-0.19,0.08)	0.10(-0.11,0.30)
	Estimated variance of random effect	1.50e-20(1.42e-35,1.59e-05)	6.87e-13(1.79e-31,2.64e+06)
$(FEV1/FVC\ ratio)^3$	Combined Fixed Effect	0.12*(0.02,0.22)	-0.05(-0.19,0.10)
	Combined Random Effect	0.11*(0.02,0.20)	0.00(-0.14,0.14)
	Estimated variance of random effect	1.24e-20(9.41e-49,1.63e+08)	4.31e-19(9.86e-38,1.88e+00)
Outcomes z-scored within cohorts Cohort specific estimates are suppressed due to the small number of carriers Meta-analyses for PI-SS, SZ and ZZ were not possible as there were fewer than three cohorts with adequate data *p<0.05, **p<0.01, ***p<0.001, ****p<0.0001			

S21 Table. Association of PI-MZ with lung function with +/-3SDs removed from outcome before standardisation, adjusted for age, sex, height and height-squared

Outcome	Cohort	Regression Coefficient (95% CI)
Maximum FEV1	Combined Fixed Effect	0.07*(0.01,0.13)
	Combined Random Effect	0.07*(0.01,0.13)
	Estimated variance of random effect	1.48e-18(0.00e+00,.)
Maximum FVC	Combined Fixed Effect	0.08**(0.03,0.13)
	Combined Random Effect	0.08**(0.03,0.13)
	Estimated variance of random effect	1.06e-12(1.14e-25,9.94e+00)
Outcomes z-scored within cohorts *p<0.05, **p<0.01, ***p<0.001, ****p<0.0001		

S9 Table. Association of PI status with lung function adjusted for age, sex, height, height-squared and height-cubed

Outcome	Cohort	Regression Coefficient (95% CI)				
		MS VS. MM	MZ VS. MM	SS VS. MM	SZ VS. MM	ZZ VS. MM
Maximum FEV1	BO	-0.18(-0.48,0.12)	0.16(-0.41,0.73)	[a]		
	CaPS	0.05(-0.11,0.21)	0.09(-0.12,0.30)			
	ELSA	0.05(-0.02,0.12)	0.06(-0.04,0.17)			
	HAS	-0.41(-0.87,0.05)	0.10(-0.35,0.56)			
	HCS	0.00(-0.09,0.09)	-0.01(-0.15,0.12)			
	LBC1921	-0.08(-0.31,0.16)	0.08(-0.27,0.44)			
	NSHD	-0.02(-0.12,0.08)	0.08(-0.08,0.23)			
	WHII	0.04(-0.03,0.10)	0.11(-0.00,0.21)			
	Combined Fixed Effect	0.02(-0.02,0.06)	0.07*(0.01,0.12)	0.02(-0.20,0.25)	0.02(-0.21,0.25)	-0.30(-0.87,0.26)
	Combined Random Effect	0.02(-0.02,0.06)	0.07*(0.01,0.12)	0.03(-0.20,0.25)	0.03(-0.20,0.26)	-0.35(-0.91,0.21)
	Estimated variance of random effect	9.32e-18(7.18e-30,1.21e-05)	1.66e-19(6.53e-35,4.20e-04)	2.85e-15(4.20e-32,1.93e+02)	8.24e-16(4.26e-30,1.59e-01)	3.10e-12(4.97e-33,1.93e+09)
Maximum FVC	BO	-0.12(-0.38,0.14)	0.07(-0.43,0.57)	[a]		
	CaPS	0.07(-0.09,0.23)	0.13(-0.08,0.34)			
	ELSA	0.03(-0.04,0.09)	0.04(-0.06,0.14)			
	HAS	-0.30(-0.69,0.09)	0.34(-0.06,0.73)			
	HCS	-0.01(-0.08,0.07)	0.07(-0.05,0.18)			
	LBC1921	-0.09(-0.31,0.14)	0.11(-0.23,0.44)			
	NSHD	-0.02(-0.12,0.07)	0.08(-0.06,0.23)			
	WHII	-0.02(-0.08,0.05)	0.10(-0.00,0.20)			
	Combined Fixed Effect	-0.00(-0.04,0.03)	0.08**(0.03,0.14)	-0.11(-0.32,0.11)	0.02(-0.19,0.23)	-0.16(-0.69,0.36)
	Combined Random Effect	-0.01(-0.04,0.03)	0.08**(0.03,0.13)	-0.10(-0.30,0.11)	0.02(-0.19,0.23)	-0.18(-0.71,0.34)
	Estimated variance of random effect	4.40e-16(9.54e-32,2.02e+00)	7.69e-14(4.31e-27,1.37e+00)	4.85e-15(3.28e-31,7.17e+01)	1.66e-13(7.47e-30,3.68e+03)	7.50e-16(2.11e-35,2.67e+04)

FEV1/FVC ratio	BO	-0.31(-0.74,0.11)	0.23(-0.58,1.04)	[a]		
	CaPS	0.01(-0.16,0.18)	0.02(-0.20,0.25)			
	ELSA	0.04(-0.06,0.14)	0.02(-0.12,0.17)			
	HAS	-0.19(-0.79,0.41)	-0.27(-0.83,0.30)			
	HCS	0.03(-0.10,0.15)	-0.12(-0.31,0.07)			
	LBC1921	-0.01(-0.34,0.31)	-0.14(-0.64,0.36)			
	NSHD	-0.01(-0.16,0.13)	-0.06(-0.28,0.16)			
	WHII	0.11*(0.01,0.21)	0.03(-0.13,0.19)			
	Combined Fixed Effect	0.04(-0.01,0.09)	-0.02(-0.10,0.06)	0.23(-0.08,0.55)	0.02(-0.30,0.33)	-0.67(-1.46,0.11)
	Combined Random Effect	0.04(-0.01,0.09)	-0.02(-0.10,0.06)	0.23(-0.09,0.55)	0.02(-0.30,0.33)	-0.55(-1.41,0.32)
	Estimated variance of random effect	5.96e-19(1.82e-32,1.95e-05)	5.70e-14(0.00e+00,.)	1.50e-16(9.10e-31,2.48e-02)	9.05e-11(1.38e-27,5.94e+06)	1.22e-01(4.76e-05,3.13e+02)

Outcomes z-scored within cohorts
*p<0.05, **p<0.01, ***p<0.001, ****p<0.0001
[a] Cohort specific estimates are suppressed for PI-SS, SZ and ZZ due to the small number of carriers

S10 Table. Association of PI status with lung function adjusted for age, sex, height, height-squared and height-cubed, outcomes transformed

Outcome	Cohort	Regression Coefficient (95% CI)				
		MS VS. MM	MZ VS. MM	SS VS. MM	SZ VS. MM	ZZ VS. MM
Sqrt(Maximum FVC)	BO	-0.16(-0.43,0.11)	0.08(-0.43,0.59)	[a]		
	CaPS	0.06(-0.09,0.22)	0.11(-0.09,0.32)			
	ELSA	0.02(-0.04,0.09)	0.03(-0.07,0.13)			
	HAS	-0.29(-0.68,0.10)	0.32(-0.07,0.71)			
	HCS	-0.00(-0.08,0.07)	0.06(-0.06,0.17)			
	LBC1921	-0.10(-0.33,0.13)	0.10(-0.24,0.44)			
	NSHD	-0.02(-0.11,0.08)	0.09(-0.05,0.23)			
	WHII	-0.02(-0.08,0.04)	0.09(-0.00,0.19)			
	Combined Fixed Effect	-0.00(-0.04,0.03)	0.07**(0.02,0.13)	-0.10(-0.31,0.11)	0.02(-0.19,0.23)	-0.16(-0.69,0.36)
	Combined Random Effect	-0.01(-0.04,0.03)	0.07**(0.02,0.12)	-0.09(-0.29,0.12)	0.02(-0.19,0.23)	-0.19(-0.71,0.32)
	Estimated variance of random effect	1.85e-19(7.62e-32,4.47e-07)	3.78e-16(4.59e-31,3.11e-01)	3.68e-15(3.95e-33,3.43e+03)	1.22e-13(5.83e-28,2.54e+01)	2.38e-15(0.00e+00,.)
(FEV1/FVC ratio)^3	BO	-0.29(-0.71,0.14)	0.17(-0.66,1.00)	[a]		
	CaPS	-0.01(-0.19,0.16)	0.01(-0.22,0.25)			
	ELSA	0.04(-0.06,0.13)	0.04(-0.10,0.19)			
	HAS	-0.10(-0.68,0.48)	-0.35(-0.92,0.22)			
	HCS	0.01(-0.11,0.14)	-0.06(-0.25,0.12)			
	LBC1921	-0.06(-0.38,0.27)	-0.13(-0.63,0.36)			
	NSHD	-0.02(-0.16,0.13)	-0.10(-0.32,0.12)			
	WHII	0.10*(0.00,0.21)	0.05(-0.11,0.21)			
	Combined Fixed Effect	0.03(-0.02,0.08)	-0.01(-0.08,0.07)	0.19(-0.13,0.50)	-0.06(-0.37,0.26)	-0.59(-1.38,0.19)
	Combined Random Effect	0.03(-0.02,0.08)	-0.01(-0.08,0.07)	0.19(-0.13,0.50)	-0.05(-0.38,0.27)	-0.55(-1.33,0.22)
	Estimated variance of random effect	1.01e-18(8.43e-33,1.20e-04)	1.64e-16(1.28e-28,2.10e-04)	1.34e-17(1.32e-35,1.36e+01)	4.00e-03(3.93e-07,4.08e+01)	5.34e-12(2.39e-31,1.19e+08)

Outcomes z-scored within cohorts

*p<0.05, **p<0.01, ***p<0.001, ****p<0.0001

[a] Cohort specific estimates are suppressed for PI-SS, SZ and ZZ due to the small number of carriers

S11 Table. Association of PI status with physical capability, adjusted for age and sex

Outcome	Cohort	Coefficient[a] (95% CI)			
		MS VS. MM	MZ VS. MM	SS VS. MM	SZ VS. MM
Grip strength	ELSA	-0.00(-0.06,0.05)	-0.02(-0.11,0.06)	[c]	
	HAS	0.02(-0.17,0.21)	0.07(-0.18,0.33)		
	HCS	0.01(-0.07,0.08)	0.12*(0.01,0.23)		
	LBC1921	-0.12(-0.33,0.08)	0.29(-0.02,0.60)		
	NSHD	0.05(-0.04,0.15)	0.13(-0.01,0.27)		
	Combined Fixed Effect	0.01(-0.03,0.05)	0.06*(0.01,0.12)	0.00(-0.23,0.23)	0.00(-0.22,0.22)
	Combined Random Effect	0.01(-0.03,0.05)	0.08(-0.00,0.16)	0.03(-0.20,0.26)	-0.01(-0.23,0.20)
	Estimated variance of random effect	3.66e-16(4.15e-37,3.23e+05)	2.98e-03(1.75e-04,5.08e-02)	8.78e-10(1.18e-27,6.53e+08)	3.66e-16(5.67e-37,2.37e+05)
Chair rise speed	ELSA	-0.04(-0.14,0.05)	-0.03(-0.17,0.11)	[c]	
	HAS	0.21(-0.35,0.78)	0.32(-0.25,0.89)		
	HCS	-0.15(-0.31,0.02)	-0.01(-0.25,0.24)		
	NSHD	-0.01(-0.14,0.13)	0.10(-0.09,0.30)		
	Combined Fixed Effect	-0.04(-0.11,0.02)	0.03(-0.07,0.13)	0.29(-0.13,0.72)	0.03(-0.35,0.40)
	Combined Random Effect	-0.05(-0.12,0.02)	0.02(-0.08,0.12)	0.27(-0.15,0.70)	-0.00(-0.77,0.77)
	Estimated variance of random effect	9.52e-16(0.00e+00,.)	1.02e-19(4.96e-43,2.08e+04)	8.75e-19(2.29e-37,3.34e+00)	3.74e-01(3.87e-02,3.61e+00)
Walk speed	ELSA	0.06(-0.05,0.17)	0.09(-0.07,0.25)	[c]	
	HAS	-0.31(-0.81,0.19)	0.52(-0.00,1.05)		
	HCS	-0.04(-0.18,0.10)	0.05(-0.15,0.25)		
	LBC1921	0.19(-0.13,0.51)	0.20(-0.26,0.67)		
	WHII	-0.07(-0.16,0.01)	0.03(-0.10,0.17)		
	Combined Fixed Effect	-0.02(-0.08,0.03)	0.07(-0.02,0.16)	0.16(-0.17,0.48)	-0.28(-0.63,0.07)
	Combined Random Effect	-0.02(-0.09,0.05)	0.07(-0.01,0.16)	0.15(-0.17,0.48)	-0.32(-0.80,0.17)
	Estimated variance of random effect	8.25e-04(9.04e-07,7.53e-01)	1.65e-17(1.70e-32,1.59e-02)	2.26e-13(4.27e-62,1.19e+36)	1.13e-01(5.11e-03,2.51e+00)

TUG speed	BO	-0.08(-0.48,0.32)	0.70(-0.07,1.48)	[c]	
	CaPS	-0.10(-0.33,0.14)	-0.01(-0.34,0.31)		
	HAS	-0.41(-0.88,0.07)	0.42(-0.09,0.93)		
	HCS	-0.05(-0.19,0.09)	0.03(-0.17,0.24)		
	NSHD	0.09(-0.07,0.25)	-0.00(-0.24,0.23)		
	Combined Fixed Effect	-0.02(-0.11,0.07)	0.06(-0.08,0.19)	0.07(-0.49,0.62)	-0.80**(-1.31,-0.28)[d]
	Combined Random Effect	-0.02(-0.12,0.07)	0.06(-0.07,0.19)	0.05(-0.51,0.61)	-0.75*(-1.32,-0.18)[d]
	Estimated variance of random effect	6.01e-17(6.36e-46,5.68e+12)	1.29e-12(1.15e-191,1.46e+167)	1.27e-10[b]	6.60e-02(1.12e-04,3.87e+01)
Inability to balance for 5s	BO	1.28(0.46,3.54)	-	[c]	
	CaPS	0.62(0.32,1.22)	1.26(0.56,2.81)		
	ELSA	1.21(0.89,1.65)	1.21(0.77,1.90)		
	HAS	2.01(0.68,5.91)	1.07(0.31,3.71)		
	HCS	1.08(0.69,1.69)	0.52(0.22,1.22)		
	NSHD	0.96(0.49,1.87)	0.93(0.33,2.58)		
	Combined Fixed Effect	1.09(0.88,1.35)	0.98(0.72,1.35)	F3CH	F3CH
	Combined Random Effect	1.09(0.88,1.35)	0.98(0.72,1.35)	F3CH	F3CH
	Estimated variance of random effect	4.33e-12(0.00e+00,.)	2.00e-08(0.00e+00,.)	F3CH	F3CH

Continuous outcomes z-scored within cohorts
Meta-analyses for PI-ZZ were not possible as there were fewer than three cohorts with adequate data
F3CH: Fewer than 3 cohorts with adequate data to perform the meta-analysis
*p<0.05, **p<0.01, ***p<0.001, ****p<0.0001
[a] Coefficients are linear regression coefficients for continuous outcomes and odds ratios for binary outcomes
[b] Standard error of the variance of the random effect not estimable
[c] Cohort specific estimates are suppressed for PI-SS and SZ due to the small number of carriers
[d] Association driven by a single observation

S12 Table. Number of COPD cases and non-cases in PI status – COPD analysis

Cohort	Number COPD cases	Number COPD non-cases	Total
BO	35	227	262
CaPS	86	1187	1273
ELSA	486	4331	4817
HAS	22	162	184
HCS	233	2278	2511
LBC1921	28	476	504
NSHD	88	1892	1980
WHII	315	3917	4232
Total	1293	14470	15763
Total (%)	8.20	91.80	100
Numbers based on individuals with a valid PI status, age, sex, height and lung function measures			

S13 Table. Association of PI status with COPD status, adjusted for age and sex

	OR for COPD[a] (95% CI)	
Cohort	MS VS. MM	MZ VS. MM
BO	2.32(0.84,6.36)	-
CaPS	1.02(0.48,2.18)	0.94(0.33,2.67)
ELSA	1.05(0.76,1.46)	1.08(0.67,1.73)
HAS	1.39(0.28,6.82)	0.66(0.08,5.43)
HCS	0.92(0.58,1.47)	1.27(0.70,2.31)
LBC1921	0.87(0.20,3.81)	1.14(0.14,9.09)
NSHD	1.40(0.75,2.64)	-
WHII	0.65(0.41,1.02)	0.99(0.54,1.81)
Combined Fixed Effect	0.97(0.79,1.18)	0.98(0.73,1.31)
Combined Random Effect	0.97(0.79,1.18)	0.98(0.73,1.31)
Estimated variance of random effect	2.23e-07(0.00e+00,.)	3.76e-17(0.00e+00,.)
Meta-analyses for PI-SS, SZ and ZZ were not possible as there were fewer than three cohorts with adequate data		
*p<0.05, **p<0.01, ***p<0.001, ****p<0.0001		
[a] Coefficient is odds ratio for COPD in PI class versus PI-MM		

S14 Table. Association of PI status with height (cm) adjusted for age and sex

Cohort	Regression Coefficient (95% CI)				
	MS VS. MM	MZ VS. MM	SS VS. MM	SZ VS. MM	ZZ VS. MM
BO	2.31(-0.24,4.86)	7.33**(2.36,12.30)	[a]		
CaPS	0.27(-0.97,1.51)	0.55(-1.08,2.18)			
ELSA	0.20(-0.40,0.81)	2.02****(1.12,2.92)			
HAS	0.24(-3.04,3.51)	-0.26(-3.69,3.18)			
HCS	-0.10(-0.87,0.67)	1.23*(0.09,2.37)			
LBC1921	0.44(-1.55,2.43)	1.06(-1.89,4.00)			
NSHD	0.68(-0.11,1.47)	1.84**(0.64,3.04)			
WHII	0.23(-0.34,0.81)	1.24**(0.34,2.14)			
Combined Fixed Effect	0.28(-0.03,0.59)	1.50****(1.03,1.97)	-0.13(-2.07,1.80)	1.78(-0.13,3.69)	2.04(-2.33,6.41)
Combined Random Effect	0.28(-0.03,0.59)	1.51****(1.04,1.97)	-1.24(-5.15,2.67)	1.81(-0.10,3.71)	2.04(-2.36,6.43)
Estimated variance of random effect	2.77e-13(1.06e-26,7.27e+00)	1.67e-12(1.59e-25,1.76e+01)	1.57e+01(2.04e+00,1.20e+02)	2.22e-11(8.31e-27,5.92e+04)	1.13e-11(1.32e-29,9.74e+06)

*p<0.05, **p<0.01, ***p<0.001, ****p<0.0001

[a] Cohort specific estimates are suppressed for PI-SS, SZ and ZZ due to the small number of carriers

S15 Table. Association of PI-MZ with Height, +/-3 SDs removed from outcome, adjusted for age and sex

Cohort	Regression Coefficient (95% CI) (cm)
Combined Fixed Effect	1.53****(1.06,1.99)
Combined Random Effect	1.53****(1.06,1.99)
Estimated variance of random effect	6.08e-11(9.70e-25,3.82e+03)
*p<0.05, **p<0.01, ***p<0.001, ****p<0.0001	

S16 Table. Association of PI status with height, adjusted for age, sex, FEV1(z-score) and FVC(z-score)

Cohort	Regression Coefficient (95% CI) (cm)				
	MS VS. MM	MZ VS. MM	SS VS. MM	SZ VS. MM	ZZ VS. MM
BO	2.06(-0.28,4.40)	5.76*(1.22,10.29)	[a]		
CaPS	-0.04(-1.17,1.09)	-0.18(-1.66,1.31)			
ELSA	0.04(-0.54,0.63)	1.55***(0.69,2.40)			
HAS	0.94(-2.19,4.07)	-1.33(-4.49,1.83)			
HCS	-0.06(-0.75,0.63)	0.53(-0.50,1.55)			
LBC1921	0.75(-1.05,2.54)	0.62(-2.12,3.37)			
NSHD	0.61(-0.20,1.42)	1.20(-0.01,2.42)			
WHII	0.15(-0.43,0.72)	0.76(-0.14,1.66)			
Combined Fixed Effect	0.18(-0.12,0.48)	0.92****(0.47,1.37)	0.13(-1.69,1.95)	0.89(-0.93,2.71)	1.32(-3.19,5.84)
Combined Random Effect	0.19(-0.11,0.49)	0.91****(0.46,1.35)	-0.46(-3.30,2.37)	0.92(-0.91,2.74)	1.28(-3.24,5.79)
Estimated variance of random effect	1.25e-12(9.57e-31,1.64e+06)	2.70e-08(5.75e-29,1.27e+13)	6.31e+00(3.26e-01,1.22e+02)	5.25e-11(1.34e-25,2.05e+04)	3.66e-11(5.11e-31,2.62e+09)

*p<0.05, **p<0.01, ***p<0.001, ****p<0.0001

[a] Cohort specific estimates are suppressed for PI-SS, SZ and ZZ due to the small number of carriers

S30 Table. Association of PI-MZ with Height, +/-3 SDs removed from outcome, adjusted for age, sex, FEV1 (z-score) and FVC (z-score)

Cohort	Regression Coefficient (95% CI) (cm)
Combined Fixed Effect	0.96****(0.51,1.41)
Combined Random Effect	0.94****(0.49,1.38)
Estimated variance of random effect	3.42e-07(2.24e-107,5.21e+93)
*p<0.05, **p<0.01, ***p<0.001, ****p<0.0001	

S31 Table. Association of PI status with weight, adjusted for age and sex

Cohort	Regression Coefficient (95% CI) (kg)				
	MS VS. MM	MZ VS. MM	SS VS. MM	SZ VS. MM	ZZ VS. MM
BO	0.70(-4.83,6.22)	4.25(-6.21,14.72)	[a]		
CaPS	0.27(-2.05,2.59)	0.27(-2.78,3.31)			
ELSA	0.08(-1.21,1.38)	1.05(-0.89,2.99)			
HAS	-0.38(-7.76,7.00)	-6.10(-13.96,1.77)			
HCS	-0.36(-1.98,1.26)	0.84(-1.57,3.26)			
LBC1921	-2.88(-6.48,0.73)	3.20(-2.08,8.48)			
NSHD	0.13(-1.59,1.85)	3.26*(0.61,5.91)			
WHII	0.38(-0.80,1.56)	0.28(-1.58,2.13)			
Combined Fixed Effect	0.06(-0.60,0.71)	1.02*(0.04,2.01)	-2.73(-6.79,1.32)	1.97(-2.08,6.03)	-2.43(-11.60,6.74)
Combined Random Effect	0.03(-0.62,0.68)	1.01*(0.03,1.99)	-3.15(-7.19,0.88)	2.36(-2.41,7.14)	-2.10(-11.21,7.00)
Estimated variance of random effect	5.33e-13(6.25e-25,4.54e-01)	9.93e-15(2.68e-27,3.68e-02)	7.92e-08(1.05e-22,5.97e+07)	7.88e+00(8.09e-02,7.68e+02)	1.85e-07(4.06e-176,8.46e+161)

*p<0.05, **p<0.01, ***p<0.001, ****p<0.0001

[a] Cohort specific estimates are suppressed for PI-SS, SZ and ZZ due to the small number of carriers

S17 Table. Association of PI-MZ with Weight, +/-3 SDs removed from outcome, adjusted for age and sex

Cohort	Regression Coefficient (95% CI) (kg)
Combined Fixed Effect	0.92(-0.02,1.86)
Combined Random Effect	0.95*(0.01,1.88)
Estimated variance of random effect	2.80e-11(4.20e-24,1.86e+02)
*p<0.05, **p<0.01, ***p<0.001, ****p<0.0001	

S18 Table. Association of PI status with ln(weight), adjusted for age and sex

Cohort	Regression Coefficient (95% CI)				
	MS VS. MM	MZ VS. MM	SS VS. MM	SZ VS. MM	ZZ VS. MM
BO	-0.00(-0.07,0.07)	0.06(-0.08,0.20)	[a]		
CaPS	0.00(-0.03,0.03)	0.01(-0.03,0.05)			
ELSA	0.00(-0.02,0.02)	0.02(-0.01,0.04)			
HAS	-0.01(-0.11,0.09)	-0.07(-0.17,0.04)			
HCS	-0.00(-0.02,0.02)	0.01(-0.02,0.04)			
LBC1921	-0.04(-0.09,0.01)	0.05(-0.03,0.12)			
NSHD	0.00(-0.02,0.02)	0.04*(0.00,0.07)			
WHII	0.00(-0.01,0.02)	0.00(-0.02,0.03)			
Combined Fixed Effect	0.00(-0.01,0.01)	0.01*(0.00,0.03)	-0.03(-0.09,0.02)	0.03(-0.02,0.08)	-0.02(-0.14,0.10)
Combined Random Effect	0.00(-0.01,0.01)	0.01*(0.00,0.03)	-0.04(-0.09,0.02)	0.03(-0.03,0.09)	-0.02(-0.13,0.10)
Estimated variance of random effect	1.73e-19(3.95e-35,7.59e-04)	1.72e-19(6.57e-32,4.50e-07)	8.89e-05(2.82e-14,2.80e+05)	7.46e-04(5.62e-07,9.89e-01)	2.83e-18(1.88e-41,4.27e+05)

Outcome is ln(weight in kg)
*p<0.05, **p<0.01, ***p<0.001, ****p<0.0001
[a] Cohort specific estimates are suppressed for PI-SS, SZ and ZZ due to the small number of carriers

S19 Table. Association of PI status with BMI, adjusted for age and sex

Cohort	Regression Coefficient (95% CI) (kg/m^2)				
	MS VS. MM	MZ VS. MM	SS VS. MM	SZ VS. MM	ZZ VS. MM
BO	-0.66(-2.54,1.22)	-0.88(-4.50,2.73)	[a]		
CaPS	0.05(-0.65,0.76)	0.02(-0.90,0.95)			
ELSA	0.01(-0.44,0.46)	-0.26(-0.94,0.41)			
HAS	-1.74(-4.28,0.79)	-1.83(-4.47,0.81)			
HCS	-0.09(-0.63,0.46)	-0.05(-0.86,0.76)			
LBC1921	-1.25(-2.59,0.09)	0.93(-1.04,2.90)			
NSHD	-0.20(-0.79,0.39)	0.51(-0.40,1.42)			
WHII	0.04(-0.34,0.43)	-0.27(-0.87,0.34)			
Combined Fixed Effect	-0.08(-0.30,0.14)	-0.09(-0.43,0.24)	-0.94(-2.30,0.43)	0.20(-1.16,1.56)	-1.56(-4.65,1.52)
Combined Random Effect	-0.08(-0.30,0.14)	-0.09(-0.42,0.24)	-0.92(-2.29,0.45)	0.32(-1.15,1.80)	-1.48(-4.47,1.52)
Estimated variance of random effect	5.69e-18(1.13e-32,2.88e-03)	1.35e-18(2.10e-34,8.67e-03)	4.65e-14(3.71e-29,5.82e+01)	4.04e-01(8.81e-05,1.85e+03)	2.55e-10(2.47e-28,2.63e+08)

*p<0.05, **p<0.01, ***p<0.001, ****p<0.0001

[a] Cohort specific estimates are suppressed for PI-SS, SZ and ZZ due to the small number of carriers

S20 Table. Association of PI status with ln(BMI), adjusted for age and sex

Cohort	Regression Coefficient (95% CI)				
	MS VS. MM	MZ VS. MM	SS VS. MM	SZ VS. MM	ZZ VS. MM
BO	-0.03(-0.10,0.04)	-0.03(-0.16,0.10)	[a]		
CaPS	0.00(-0.02,0.03)	0.00(-0.03,0.04)			
ELSA	0.00(-0.02,0.02)	-0.01(-0.03,0.02)			
HAS	-0.07(-0.16,0.02)	-0.06(-0.15,0.03)			
HCS	-0.00(-0.02,0.02)	-0.00(-0.03,0.03)			
LBC1921	-0.05(-0.10,0.00)	0.04(-0.04,0.11)			
NSHD	-0.01(-0.03,0.01)	0.02(-0.02,0.05)			
WHII	0.00(-0.01,0.02)	-0.01(-0.03,0.01)			
Combined Fixed Effect	-0.00(-0.01,0.00)	-0.00(-0.02,0.01)	-0.03(-0.08,0.02)	0.01(-0.04,0.06)	-0.05(-0.16,0.06)
Combined Random Effect	-0.00(-0.01,0.00)	-0.00(-0.01,0.01)	-0.03(-0.08,0.02)	0.01(-0.04,0.06)	-0.05(-0.15,0.06)
Estimated variance of random effect	1.09e-20(0.00e+00,.)	2.48e-16(2.02e-28,3.03e-04)	1.62e-17(2.73e-33,9.64e-02)	2.41e-04(1.27e-11,4.57e+03)	1.95e-18(1.08e-35,3.54e-01)

Outcome is ln(BMI in kg/m^2)

*p<0.05, **p<0.01, ***p<0.001, ****p<0.0001

[a] Cohort specific estimates are suppressed for PI-SS, SZ and ZZ due to the small number of carriers

Cystic Fibrosis

S21 Table. Cystic fibrosis deltaF508 carrier frequency

Cohort	Number of carriers[a]	Number of non-carriers[a]	Total, N[a]	Minor Allele Frequency	% female	Mean age (SE)
BO	5	258	263	0.010	54.75	69.60(0.26)
CaPS	39	1254	1293	0.015	0	53.80(0.14)
ELSA	170	4931	5101	0.017	53.85	65.62(0.13)
HAS	8	184	192	0.021	34.9	76.36(0.16)
HCS	70	2495	2565	0.014	47.02	66.09(0.06)
LBC1921	12	506	518	0.012	58.88	79.06(0.03)
NSHD	49	1912	1961	0.012	52.07	53.45(0.00)
WHII	99	3314	3413	0.015	25.99	60.63(0.10)
Combined	452	14854	15306	0.015	41.66	62.69(0.07)

[a] Total number of individuals for each mutation is total number with a valid genotype, sex, age and lung function

S22 Table. Association of deltaF508 carrier status with lung function

Outcome	Cohort	Regression Coefficient (95% CI) (age and sex adjusted model)	Regression Coefficient (95% CI) (current smokers, age and sex adjusted model)	Regression Coefficient (95% CI) (ex smokers, age and sex adjusted model)	Regression Coefficient (95% CI) (never smokers, age and sex adjusted model)	Regression Coefficient (95% CI) (COPD cases, age and sex adjusted model)	Regression Coefficient (95% CI) (age, sex, height and height-squared adjusted model)
Maximum FEV1	BO	-0.33(-1.00,0.33)	[a]				-0.23(-0.86,0.39)
	CaPS	0.05(-0.23,0.33)					0.01(-0.26,0.27)
	ELSA	-0.02(-0.14,0.09)					-0.02(-0.13,0.09)
	HAS	-0.14(-0.73,0.46)					-0.19(-0.76,0.37)
	HCS	0.05(-0.13,0.23)					-0.01(-0.18,0.16)
	LBC1921	0.01(-0.44,0.45)					0.13(-0.29,0.54)
	NSHD	-0.04(-0.26,0.17)					-0.10(-0.29,0.10)
	WHII	-0.05(-0.20,0.09)					-0.05(-0.18,0.09)
	Combined Fixed Effect	-0.02(-0.09,0.05)	-0.05(-0.23,0.12)	0.02(-0.08,0.12)	-0.05(-0.16,0.07)	0.09(-0.10,0.27)	-0.03(-0.10,0.03)
	Combined Random Effect	-0.02(-0.09,0.05)	-0.06(-0.23,0.12)	0.02(-0.08,0.12)	-0.04(-0.16,0.07)	0.09(-0.08,0.27)	-0.03(-0.10,0.03)
	Estimated variance of random effect	4.07e-14(2.19e-26,7.56e-02)	5.10e-18(0.00e+00,.)	7.32e-17(5.89e-32,9.10e-02)	1.41e-22(3.51e-40,5.64e-05)	5.38e-18(1.71e-35,1.69e+00)	3.81e-17(5.16e-33,2.81e-01)
Maximum FVC	BO	-0.17(-0.79,0.45)	[a]				-0.04(-0.59,0.52)
	CaPS	-0.16(-0.45,0.14)					-0.21(-0.48,0.05)
	ELSA	-0.09(-0.20,0.02)					-0.11(-0.21,0.00)
	HAS	0.05(-0.47,0.57)					-0.01(-0.49,0.48)
	HCS	0.10(-0.06,0.26)					0.01(-0.13,0.16)
	LBC1921	-0.17(-0.61,0.27)					-0.03(-0.43,0.36)
	NSHD	0.01(-0.20,0.21)					-0.05(-0.23,0.13)
	WHII	-0.09(-0.23,0.05)					-0.08(-0.20,0.04)
	Combined Fixed Effect	-0.06(-0.12,0.01)	-0.12(-0.29,0.05)	-0.00(-0.10,0.09)	-0.08(-0.20,0.03)	-0.07(-0.29,0.16)	-0.08*(-0.14,-0.02)
	Combined Random Effect	-0.05(-0.12,0.02)	-0.12(-0.29,0.05)	0.00(-0.10,0.10)	-0.08(-0.19,0.03)	-0.08(-0.29,0.14)	-0.07*(-0.13,-0.01)

	Estimated variance of random effect	5.60e-08(7.09e-24,4.42e+08)	6.12e-15(6.97e-30,5.37e+00)	6.43e-15(7.91e-27,5.23e-03)	1.23e-22(6.83e-39,2.23e-06)	3.15e-23(4.54e-42,2.18e-04)	1.93e-17(2.22e-29,1.67e-05)
FEV1/FVC ratio	BO	-0.39(-1.27,0.49)	[a]				-0.43(-1.31,0.45)
	CaPS	0.42**(0.14,0.69)					0.42**(0.14,0.70)
	ELSA	0.08(-0.07,0.24)					0.11(-0.05,0.26)
	HAS	-0.44(-1.16,0.28)					-0.44(-1.16,0.28)
	HCS	-0.05(-0.28,0.19)					0.01(-0.23,0.24)
	LBC1921	0.35(-0.22,0.92)					0.33(-0.24,0.91)
	NSHD	-0.11(-0.39,0.17)					-0.08(-0.36,0.20)
	WHII	0.04(-0.16,0.23)					0.03(-0.17,0.22)
	Combined Fixed Effect	0.05(-0.04,0.15)	0.10(-0.13,0.34)	0.05(-0.08,0.18)	0.04(-0.11,0.19)	0.26(-0.07,0.59)	0.07(-0.02,0.16)
	Combined Random Effect	0.06(-0.05,0.17)	0.08(-0.24,0.39)	0.07(-0.06,0.19)	0.06(-0.09,0.20)	0.23(-0.09,0.55)	0.08(-0.01,0.17)
	Estimated variance of random effect	4.58e-03(1.25e-06,1.67e+01)	5.84e-02(3.72e-03,9.16e-01)	1.22e-18(3.66e-34,4.06e-03)	1.82e-07(2.94e-22,1.12e+08)	2.20e-17(5.96e-34,8.15e-01)	1.01e-11(1.73e-26,5.88e+03)

Outcomes z-scored within cohorts
*p<0.05, **p<0.01, ***p<0.001, ****p<0.0001
[a] Cohort specific estimates are suppressed for smoking and COPD strata due to the small number of carriers

S23 Table. Association of deltaF508 carrier status with lung function, outcomes transformed

Outcome	Cohort	Regression Coefficient (95% CI) (age and sex adjusted model)	Regression Coefficient (95% CI) (current smokers, age and sex adjusted model)	Regression Coefficient (95% CI) (ex smokers, age and sex adjusted model)	Regression Coefficient (95% CI) (never smokers, age and sex adjusted model)	Regression Coefficient (95% CI) (COPD cases, age and sex adjusted model)	Regression Coefficient (95% CI) (age, sex, height and height-squared adjusted model)
Sqrt(Maximum FVC)	BO	-0.16(-0.78,0.46)	[a]				-0.03(-0.60,0.53)
	CaPS	-0.14(-0.43,0.14)					-0.20(-0.46,0.06)
	ELSA	-0.08(-0.19,0.03)					-0.10(-0.20,0.01)
	HAS	0.07(-0.44,0.58)					0.01(-0.47,0.49)
	HCS	0.09(-0.07,0.25)					0.00(-0.14,0.15)
	LBC1921	-0.17(-0.61,0.28)					-0.04(-0.44,0.36)
	NSHD	0.01(-0.20,0.22)					-0.05(-0.24,0.13)
	WHII	-0.09(-0.23,0.05)					-0.08(-0.20,0.04)
	Combined Fixed Effect	-0.05(-0.12,0.02)	-0.11(-0.28,0.06)	-0.00(-0.10,0.10)	-0.09(-0.20,0.02)	-0.07(-0.31,0.17)	-0.08*(-0.14,-0.01)
	Combined Random Effect	-0.05(-0.11,0.02)	-0.11(-0.28,0.05)	0.01(-0.09,0.10)	-0.09(-0.20,0.02)	-0.08(-0.31,0.14)	-0.07*(-0.13,-0.01)
	Estimated variance of random effect	8.13e-14(6.22e-28,1.06e+01)	1.44e-10(0.00e+00,.)	1.21e-13(1.17e-26,1.25e+00)	1.35e-20(6.45e-35,2.83e-06)	2.71e-22(6.22e-37,1.18e-07)	1.63e-22(2.46e-36,1.08e-08)
(FEV1/FVC ratio)^3	BO	-0.16(-1.05,0.73)	[a]				-0.21(-1.10,0.69)
	CaPS	0.47**(0.19,0.76)					0.48**(0.19,0.77)
	ELSA	0.09(-0.06,0.24)					0.12(-0.04,0.27)
	HAS	-0.47(-1.17,0.23)					-0.48(-1.17,0.22)
	HCS	-0.02(-0.25,0.22)					0.04(-0.19,0.27)
	LBC1921	0.38(-0.20,0.95)					0.36(-0.22,0.93)
	NSHD	-0.06(-0.34,0.22)					-0.03(-0.31,0.25)
	WHII	0.01(-0.19,0.20)					0.00(-0.20,0.20)

	Combined Fixed Effect	0.07(-0.03,0.16)	0.15(-0.07,0.38)	0.06(-0.07,0.19)	0.04(-0.12,0.19)	0.13(-0.03,0.29)	0.09(-0.01,0.18)
	Combined Random Effect	0.07(-0.04,0.19)	0.09(-0.24,0.41)	0.07(-0.06,0.20)	0.06(-0.13,0.25)	0.12(-0.03,0.27)	0.09(-0.02,0.20)
	Estimated variance of random effect	7.20e-03(2.64e-05,1.96e+00)	7.61e-02(7.02e-03,8.24e-01)	7.30e-11(7.66e-27,6.95e+05)	1.47e-02(2.22e-04,9.76e-01)	6.50e-22[b]	3.68e-03(2.25e-07,6.04e+01)

Outcomes z-scored within cohorts
*p<0.05, **p<0.01, ***p<0.001, ****p<0.0001
[a] Cohort specific estimates are suppressed for smoking and COPD strata due to the small number of carriers
[b] Standard error of the variance of the random effect not estimable

S24 Table. Number of COPD cases and non-cases in deltaF508 carrier – COPD analysis

Cohort	Number COPD cases	Number COPD non-cases	Total
BO	35	228	263
CaPS	87	1189	1276
ELSA	504	4461	4965
HAS	22	168	190
HCS	240	2323	2563
LBC1921	30	488	518
NSHD	89	1867	1956
WHII	246	3162	3408
Total	1253	13886	15139
Total (%)	8.28	91.72	100
Numbers based on individuals with a valid deltaF508 genotype, age, sex, height and lung function measures			

S40 Table. Association of deltaF508 carrier status with COPD status, adjusted for age and sex

Cohort	OR for COPD[a] (95% CI)
BO	1.55(0.17,14.53)
CaPS	-
ELSA	1.01(0.61,1.69)
HAS	2.66(0.50,14.12)
HCS	1.13(0.51,2.50)
LBC1921	1.52(0.19,12.26)
NSHD	2.05(0.72,5.89)
WHII	1.39(0.69,2.82)
Combined Fixed Effect	1.14(0.82,1.58)
Combined Random Effect	1.14(0.82,1.58)
Estimated variance of random effect	5.50e-17(0.00e+00,.)
*p<0.05, **p<0.01, ***p<0.001, ****p<0.0001 [a] Coefficient is odds ratio for COPD in carriers versus non-carriers	

Medium-chain acyl Co-A dehydrogenase deficiency

S41 Table. MCADD (K304E) carrier frequency

Cohort	Number of carriers[a]	Number of non-carriers[a]	Total, N[a]	Minor Allele Frequency	% female	Mean age[b] (SE)
BO	2	263	265	0.004	55.09	69.62(0.26)
CaPS	13	1237	1250	0.005	0	73.14(0.15)
ELSA	90	5393	5483	0.008	54.37	66.10(0.13)
HAS	8	511	519	0.008	38.92	76.39(0.16)
HCS	43	2729	2772	0.008	47.19	68.25(0.07)
LBC1921	0	522	522	0.000	58.24	79.07(0.03)
NSHD	33	2649	2682	0.006	50.15	53.45(0.00)
WHII	89	4379	4468	0.010	24.17	60.90(0.09)
Combined	278	17683	17961	0.008	41.01	64.83(0.09)

[a] Total number of individuals for each mutation is total number with a valid genotype, sex, age and at least one outcome measure
[b] Mean age at walk test for LBC1921 and WHII, and at balance test for all other cohorts

S25 Table. Association of K304E carrier status with physical capability, adjusted for age and sex

Outcome	Cohort	Coefficient[a] (95% CI)
Grip strength	ELSA	[c]
	HAS	
	HCS	
	LBC1921	
	NSHD	
	Combined Fixed Effect	-0.05(-0.15,0.04)
	Combined Random Effect	-0.05(-0.15,0.04)
	Estimated variance of random effect	4.60e-18[b]
Chair rise speed	ELSA	[c]
	HAS	
	HCS	
	NSHD	
	Combined Fixed Effect	0.10(-0.06,0.26)
	Combined Random Effect	0.10(-0.06,0.26)
	Estimated variance of random effect	1.05e-16(4.35e-37,2.53e+04)
Walk speed	ELSA	[c]
	HAS	
	HCS	
	LBC1921	
	WHII	
	Combined Fixed Effect	0.12(-0.02,0.26)
	Combined Random Effect	0.12(-0.02,0.26)
	Estimated variance of random effect	5.01e-17(0.00e+00,.)

TUG speed	BO	[c]
	CaPS	
	HAS	
	HCS	
	NSHD	
	Combined Fixed Effect	0.10(-0.13,0.33)
	Combined Random Effect	0.11(-0.13,0.34)
	Estimated variance of random effect	7.56e-20[b]
Inability to balance for 5s	BO	[c]
	CaPS	
	ELSA	
	HAS	
	HCS	
	NSHD	
	Combined Fixed Effect	0.92(0.55,1.56)
	Combined Random Effect	0.92(0.55,1.56)
	Estimated variance of random effect	2.85e-10(0.00e+00,.)

Continuous outcomes z-scored within cohorts
*p<0.05, **p<0.01, ***p<0.001, ****p<0.0001
[a] Coefficients are linear regression coefficients for continuous outcomes and odds ratios for binary outcomes
[b] Standard error of the variance of the random effect not estimable
[c] Cohort specific estimates are suppressed due to the small number of carriers

S26 Table. Association of K304E carrier status with cognitive capability, adjusted for age and sex

Outcome		Cohort	Regression Coefficient (95% CI)
Crystallized ability	Mill Hill	HAS	0.27(-0.44,0.98)
		WHII	-0.39***(-0.62,-0.17)
	Mill Hill[b]	WHII	-0.35**(-0.57,-0.12)
	NART	CaPS	0.03(-0.52,0.57)
		LBC1921	NC
		NSHD	0.15(-0.22,0.51)
		Combined Fixed Effect	F3CH
		Combined Random Effect	F3CH
		Estimated variance of random effect	F3CH
Fluid ability	AH4	CaPS	0.16(-0.38,0.70)
		HAS	-0.16(-0.85,0.54)
		WHII	-0.18(-0.40,0.04)
		Combined Fixed Effect	-0.13(-0.33,0.06)
		Combined Random Effect	-0.14(-0.33,0.06)
		Estimated variance of random effect	7.24e-14(0.00e+00,.)
	Semantic fluency	CaPS	-0.30(-0.83,0.24)
		ELSA	0.10(-0.10,0.29)
		NSHD	-0.21(-0.56,0.13)
		WHII	-0.10(-0.33,0.12)
		Combined Fixed Effect	-0.05(-0.18,0.09)
		Combined Random Effect	-0.05(-0.19,0.09)
		Estimated variance of random effect	1.66e-03(7.13e-10,3.85e+03)

	Phonemic Fluency	LBC1921	NC
		WHII	-0.19(-0.42,0.04)
	Search Speed[c]	ELSA	0.01(-0.19,0.21)
		NSHD	0.15(-0.19,0.49)
	Word recall	ELSA	0.03(-0.15,0.22)
		NSHD	0.07(-0.28,0.42)
		WHII	-0.06(-0.29,0.16)
		Combined Fixed Effect	0.00(-0.13,0.14)
		Combined Random Effect	0.01(-0.13,0.14)
		Estimated variance of random effect	2.12e-20(1.12e-40,4.03e+00)
	FCRT[a]	CaPS	0.03(-0.52,0.58)
	Ravens Progressive Matrices	LBC1921	NC
	Logical Memory	LBC1921	NC

Outcomes z-scored within cohorts
F3CH: Fewer than 3 cohorts with adequate data to perform the meta-analysis
NC: No carriers
*p<0.05, **p<0.01, ***p<0.001, ****p<0.0001
[a] inverse transformed prior to z-scoring
[b] square transformed prior to z-scoring
[c] natural log transformed prior to z-scoring

Phenylketonuria

S27 Table. PKU mutation carrier status

Cohort	Number of carriers[a,c]	Number of non-carriers[a]	Total, N[a]	Minor Allele Frequency	% female	Mean age[b] (SE)
BO	1	263	264	0.002	54.92	69.64(0.26)
CaPS	7	1236	1243	0.003	0	73.11(0.15)
ELSA	30	5367	5397	0.003	54.2	66.06(0.13)
HAS	5	495	500	0.005	39.2	76.39(0.16)
HCS	13	2732	2745	0.002	47.25	68.26(0.07)
LBC1921	4	522	526	0.004	57.98	79.06(0.02)
NSHD	18	2622	2640	0.003	49.92	53.45(0.00)
WHII	33	4394	4427	0.004	24.01	60.88(0.09)
Combined	111	17631	17742	0.003	40.86	64.83(0.09)

[a] Total number of individuals is total number with a valid carrier status, sex, age and at least one outcome measure
[b] Mean age at walk test for LBC1921 and WHII, and at balance test for all other cohorts
[c] A carrier is defined as any individual who carries at least one minor allele of any of these three SNPs. A non-carrier is homozygous for the major allele at each of these three SNPs. If an individual had a missing genotype for one of the PKU SNPs, they were included in the analyses if they were a carrier but were excluded if they were a non-carrier based on the remaining PKU SNPs

S28 Table. Association of PKU mutation carrier status with physical capability, adjusted for age and sex

Outcome	Cohort	Coefficient[b] (95% CI)
Grip strength	ELSA	[a]
	HAS	
	HCS	
	LBC1921	
	NSHD	
	Combined Fixed Effect	-0.14(-0.29,0.01)
	Combined Random Effect	-0.14(-0.29,0.01)
	Estimated variance of random effect	1.80e-17(1.41e-68,2.30e+34)
Chair rise speed	ELSA	[a]
	HAS	
	HCS	
	NSHD	
	Combined Fixed Effect	-0.23(-0.49,0.02)
	Combined Random Effect	-0.33(-0.73,0.07)
	Estimated variance of random effect	6.88e-02(4.01e-04,1.18e+01)
Walk Speed	ELSA	[a]
	HAS	
	HCS	
	LBC1921	
	WHII	
	Combined Fixed Effect	-0.07(-0.29,0.16)
	Combined Random Effect	-0.07(-0.29,0.16)
	Estimated variance of random effect	1.01e-12(6.68e-28,1.54e+03)

TUG Speed	BO	[a]
	CaPS	
	HAS	
	HCS	
	NSHD	
	Combined Fixed Effect	-0.01(-0.36,0.34)
	Combined Random Effect	0.03(-0.39,0.45)
	Estimated variance of random effect	5.13e-02(5.35e-05,4.92e+01)
Inability to balance for 5s	BO	[a]
	CaPS	
	ELSA	
	HAS	
	HCS	
	NSHD	
	Combined Fixed Effect	0.76(0.31,1.87)
	Combined Random Effect	0.76(0.31,1.87)
	Estimated variance of random effect	1.32e-14(0.00e+00,.)

Continuous outcomes z-scored within cohorts
*p<0.05, **p<0.01, ***p<0.001, ****p<0.0001
[a] Cohort specific estimates are suppressed due to the small number of carriers
[b] Coefficients are linear regression coefficients for continuous outcomes and odds ratios for binary outcomes

S29 Table. Association of PKU mutation carrier status with cognitive capability, adjusted for age and sex

Outcome		Cohort	Regression Coefficient (95% CI)
Crystallized ability	Mill Hill	HAS	-0.17(-1.06,0.72)
		WHII	0.02(-0.35,0.40)
	Mill Hill [b]	WHII	0.01(-0.36,0.39)
	NART	CaPS	0.36(-0.38,1.10)
		LBC1921	0.29(-0.69,1.27)
		NSHD	0.09(-0.38,0.56)
		Combined Fixed Effect	0.18(-0.18,0.55)
		Combined Random Effect	0.18(-0.18,0.55)
		Estimated variance of random effect	3.50e-18(0.00e+00,.)
Fluid ability	AH4	CaPS	0.32(-0.41,1.06)
		HAS	0.09(-0.79,0.97)
		WHII	-0.12(-0.48,0.25)
		Combined Fixed Effect	-0.01(-0.32,0.30)
		Combined Random Effect	-0.02(-0.32,0.29)
		Estimated variance of random effect	3.83e-15(2.95e-38,4.97e+08)
	Semantic fluency	CaPS	0.10(-0.63,0.82)
		ELSA	-0.45**(-0.79,-0.11)
		NSHD	0.15(-0.31,0.61)
		WHII	-0.12(-0.49,0.25)
		Combined Fixed Effect	-0.17(-0.38,0.05)

		Combined Random Effect	-0.15(-0.41,0.10)
		Estimated variance of random effect	1.74e-02(1.35e-04,2.25e+00)
	Phonemic fluency	LBC1921	-0.34(-1.34,0.65)
		WHII	-0.08(-0.47,0.30)
	Search Speed [c]	ELSA	-0.02(-0.36,0.33)
		NSHD	-0.20(-0.66,0.26)
	Word recall	ELSA	0.04(-0.28,0.36)
		NSHD	0.09(-0.37,0.55)
		WHII	-0.07(-0.44,0.30)
		Combined Fixed Effect	0.01(-0.20,0.23)
		Combined Random Effect	0.01(-0.20,0.23)
		Estimated variance of random effect	2.03e-19(7.80e-44,5.28e+05)
	FCRT[a]	CaPS	0.25(-0.53,1.03)
	Ravens Progressive Matrices	LBC1921	0.52(-0.46,1.50)
	Logical Memory	LBC1921	-0.33(-1.32,0.66)

Outcomes z-scored within cohorts
*p<0.05, **p<0.01, ***p<0.001, ****p<0.0001
[a] inverse transformed prior to z-scoring
[b] square transformed prior to z-scoring
[c] natural log transformed prior to z-scoring

S30 Table. Tests for equality of variances

Outcome	Groups tested	p-value		
		Levene's test	Brown & Forsythe's test (median)	Brown & Forsythe's test (trimmed mean)
FVC	deltaF508 carrier vs non-carrier	0.5	0.49	0.49
FVC	All PI classes	0.06	0.05	0.07
	PI-MS vs PI-MM	0.57	0.55	0.55
	PI-MZ vs PI-MM	<0.005	<0.005	<0.005
	PI-SS vs PI-MM	0.89	0.88	0.9
	PI-SZ vs PI-MM	0.83	0.82	0.82
	PI-ZZ vs PI-MM	0.5	0.32	0.5
SQRT(FVC)	All PI classes	0.26	0.22	0.27
	PI-MS vs PI-MM	0.44	0.45	0.45
	PI-MZ vs PI-MM	0.02	0.02	0.02
	PI-SS vs PI-MM	0.99	1	0.98
	PI-SZ vs PI-MM	0.84	0.85	0.83
	PI-ZZ vs PI-MM	0.68	0.41	0.68
FEV1	All PI classes	0.31	0.28	0.3
	PI-MS vs PI-MM	0.44	0.45	0.44
	PI-MZ vs PI-MM	0.03	0.03	0.03
	PI-SS vs PI-MM	0.55	0.5	0.55
	PI-SZ vs PI-MM	0.52	0.37	0.47
	PI-ZZ vs PI-MM	0.7	0.93	0.7
Height	All PI classes	0.48	0.52	0.46
	PI-MS vs PI-MM	0.27	0.28	0.27

	PI-MZ vs PI-MM	0.78	0.78	0.79
	PI-SS vs PI-MM	0.42	0.51	0.38
	PI-SZ vs PI-MM	0.65	0.65	0.65
	PI-ZZ vs PI-MM	0.12	0.13	0.12
Lung function measures were included as z-scores, height was included in cm. Tests performed using Stata's(18) –robvar- command , pooling cohorts and restricting to individuals with sex and age variables.				

S31 Table. Detecting influential data points in associations of interest

Association of interest (fixed effects analysis)	Minimum dfbeta statistic	Maximum dfbeta statistic	Regression Coefficient (95% CI) of carrier effect with most extreme 60 values removed
Association of PI-MZ vs PI-MM with FEV1, age and sex adjusted	-0.16	0.11	0.16****(0.09,0.22)
Association of PI-MZ vs PI-MM with FVC, age and sex adjusted	-0.15	0.14	0.16****(0.10,0.22)
Association of PI-MZ vs PI-MM with FEV1, age, sex, height and height-squared adjusted	-0.18	0.12	0.09**(0.04,0.15)
Association of PI-MZ vs PI-MM with FVC, age, sex, height and height-squared adjusted	-0.15	0.19	0.08**(0.02,0.13)
Association of PI-MZ vs PI-MM with height, age and sex adjusted	-0.13	0.1	1.55****(1.06,2.03)
Association of PI-MZ vs PI-MM with height, age, sex, FEV1 and FVC adjusted	-0.13	0.15	0.92***(0.45,1.38)
Association of PI-MZ vs PI-MM with ln(weight), age and sex adjusted	-0.11	0.17	0.01(-0.00,0.02)
Association of deltaF508 carrier status with FVC, age, sex, height and height-squared adjusted	-0.22	0.18	-0.07*(-0.14,-0.01)
Association of PI-MZ vs PI-MM with grip strength, age and sex adjusted	-0.15	0.25	0.06*(0.00,0.12)
Lung function and grip strength included as z-scores. Height in cm and weight in kg *p<0.05, **p<0.01, ***p<0.001, ****p<0.0001			

Supplemental References

1. Belsley DA, Kuh E, Welsch RE. Regression Diagnostics: Identifying Influential Data and Sources of Collinearity. New York: Wiley; 1980.
2. Martin RM, Gunnell D, Pemberton J, Frankel S, Davey Smith G. Cohort Profile: The Boyd Orr cohort—an historical cohort study based on the 65 year follow-up of the Carnegie Survey of Diet and Health (1937–39). International Journal of Epidemiology. 2005;34(4):742-9.
3. Steptoe A, Breeze E, Banks J, Nazroo J. Cohort Profile: The English Longitudinal Study of Ageing. International Journal of Epidemiology. 2013;42(6):1640-8.
4. Syddall HE, Simmonds SJ, Martin HJ, Watson C, Dennison EM, Cooper C, et al. Cohort profile: The Hertfordshire Ageing Study (HAS). International Journal of Epidemiology. 2010;39(1):36-43.
5. Syddall H, Aihie Sayer A, Dennison E, Martin H, Barker D, Cooper C, et al. Cohort Profile: The Hertfordshire Cohort Study. International Journal of Epidemiology. 2005;34(6):1234-42.
6. Deary IJ, Gow AJ, Pattie A, Starr JM. Cohort Profile: The Lothian Birth Cohorts of 1921 and 1936. International Journal of Epidemiology. 2012;41(6):1576-84.
7. Kuh D, Pierce M, Adams J, Deanfield J, Ekelund U, Friberg P, et al. Cohort Profile: Updating the cohort profile for the MRC National Survey of Health and Development: a new clinic-based data collection for ageing research. International Journal of Epidemiology. 2011;40(1):e1-e9.
8. Marmot M, Brunner E. Cohort Profile: The Whitehall II study. International Journal of Epidemiology. 2005;34(2):251-6.
9. Cooper R, Hardy R, Aihie Sayer A, Ben-Shlomo Y, Birnie K, Cooper C, et al. Age and Gender Differences in Physical Capability Levels from Mid-Life Onwards: The Harmonisation and Meta-Analysis of Data from Eight UK Cohort Studies. PLoS ONE. 2011;6(11):e27899.
10. Nelson HE, Willison JR. National Adult Reading Test (NART) test manual (Part II). Windsor, England: NFER-NELSON; 1991.
11. Raven JC. Guide to using the Mill Hill vocabulary test with progressive matrices. London, England: HK Lewis; 1965.
12. Heim AW. AH4 Group Test of General Intelligence. Windsor: The NFER-Nelson Publishing Company; 1970.
13. Wechsler D. Wechsler Memory Scale-Revised. New York, NY: Psychological Corp; 1987.
14. Raven JC, Court JH, Raven J. Manual for Raven's Progressive Matrices and Vocabulary Scales. London, England: H. K. Lewis; 1977.
15. Bolton CE, Cockcroft JR, Sabit R, Munnery M, McEniery CM, Wilkinson IB, et al. Lung function in mid-life compared with later life is a stronger predictor of arterial stiffness in men: The Caerphilly Prospective Study. International Journal of Epidemiology. 2009;38(3):867-76.
16. Wadsworth M, Kuh D, Richards M, Hardy R. Cohort Profile: The 1946 National Birth Cohort (MRC National Survey of Health and Development). International Journal of Epidemiology. 2006;35(1):49-54.
17. Wadsworth MEJ, Vinall LE, Jones AL, Hardy RJ, Whitehouse DB, Butterworth SL, et al. Alpha1-Antitrypsin as a Risk for Infant and Adult Respiratory Outcomes in a National Birth Cohort. American Journal of Respiratory Cell and Molecular Biology. 2004;31(5):559-64.
18. StataCorp. Stata Statistical Software: Release 13. College Station, TX: StataCorp LP; 2013.