

University of Dundee

Sixteen new lung function signals identified through 1000 Genomes Project reference panel imputation

Artigas, María Soler; Wain, Louise V.; Miller, Suzanne; Kheirallah, Abdul Kader; Huffman, Jennifer E.; Ntalla, Ioanna

Published in:
Nature Communications

DOI:
[10.1038/ncomms9658](https://doi.org/10.1038/ncomms9658)

Publication date:
2015

Licence:
CC BY

Document Version
Publisher's PDF, also known as Version of record

[Link to publication in Discovery Research Portal](#)

Citation for published version (APA):

Artigas, M. S., Wain, L. V., Miller, S., Kheirallah, A. K., Huffman, J. E., Ntalla, I., Shrine, N., Obeidat, M., Trochet, H., McArdle, W. L., Alves, A. C., Hui, J., Zhao, J. H., Joshi, P. K., Teumer, A., Albrecht, E., Imboden, M., Rawal, R., Lopez, L. M., ... UK BiLEVE (2015). Sixteen new lung function signals identified through 1000 Genomes Project reference panel imputation. *Nature Communications*, 6, [89658]. <https://doi.org/10.1038/ncomms9658>

General rights

Copyright and moral rights for the publications made accessible in Discovery Research Portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from Discovery Research Portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain.
- You may freely distribute the URL identifying the publication in the public portal.

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

ARTICLE

Received 5 Feb 2015 | Accepted 17 Sep 2015 | Published 4 Dec 2015

DOI: 10.1038/ncomms9658

OPEN

Sixteen new lung function signals identified through 1000 Genomes Project reference panel imputation

María Soler Artigas *et al.*[#]

Lung function measures are used in the diagnosis of chronic obstructive pulmonary disease. In 38,199 European ancestry individuals, we studied genome-wide association of forced expiratory volume in 1s (FEV₁), forced vital capacity (FVC) and FEV₁/FVC with 1000 Genomes Project (phase 1)-imputed genotypes and followed up top associations in 54,550 Europeans. We identify 14 novel loci ($P < 5 \times 10^{-8}$) in or near *ENSA*, *RNU5F-1*, *KCNS3*, *AK097794*, *ASTN2*, *LHX3*, *CCDC91*, *TBX3*, *TRIP11*, *RIN3*, *TEKT5*, *LTBP4*, *MN1* and *AP1S2*, and two novel signals at known loci *NPNT* and *GPR126*, providing a basis for new understanding of the genetic determinants of these traits and pulmonary diseases in which they are altered.

Correspondence and requests for materials should be addressed to I.P.H. (email: Ian.Hall@nottingham.ac.uk) or to M.D.T. (email: mt47@leicester.ac.uk).
[#]A full list of authors and their affiliations appears at the end of the paper.

Lung function, as measured by spirometry, predicts morbidity and mortality^{1,2}. Altered lung function is a key criterion for the diagnosis of chronic obstructive pulmonary disease (COPD), a leading cause of death worldwide^{3,4}. The ratio of forced expiratory volume in 1 s (FEV₁) over forced vital capacity (FVC) defines patients with airflow obstruction, while FEV₁ is used to assess the severity of the obstruction. Reduced FVC values are seen in restrictive lung diseases such as pulmonary fibrosis⁵. While environmental risk factors, such as tobacco smoking or air pollution, play a significant role in determining lung function^{6,7}, genetic factors are also important contributors, with estimates of heritability ranging between 39 and 54% (refs 8,9).

Genome-wide association studies (GWAS) of around 2.5 million common (minor allele frequency (MAF) > 5%) single-nucleotide polymorphisms (SNPs) in Europeans have identified 32 loci associated with lung function at genome-wide significance level ($P < 5 \times 10^{-8}$)¹⁰⁻¹⁴. However, as for other complex traits^{15,16}, these loci only explain a limited proportion of the heritability^{11,13}. Among explanations for the ‘missing heritability’ are a large number of, as yet, undetected common variants with modest effect sizes, in addition to low-frequency (1% < MAF ≤ 5%) and rare (MAF ≤ 1%) variants with larger effect sizes^{16,17}. Of particular relevance to low-frequency variants, phase 1 of the 1000 Genomes Project¹⁸ sequenced 1,092 individuals from 14 populations, providing an imputation reference panel of ~38 million SNPs and 1.4 million indels, including autosomal and X chromosome variants.

The aim of the current study, undertaken within the SpiroMeta consortium, was to improve coverage of low-frequency variants and detect novel loci associated with lung function by undertaking imputation of GWAS data to the 1000 Genomes Project¹⁸ Phase-1 reference panel in 38,199 individuals of European ancestry. We meta-analysed GWAS results across 17 studies and followed up the most significant associations with *in silico* data in up to 54,550 Europeans. We identify 14 new loci associated with lung function at genome-wide significance level, and novel distinct signals at two previously reported loci. These include two low-frequency variant association signals, which seem to be explained by non-synonymous SNPs. The results of these analyses implicate both previously considered and novel mechanisms influencing lung function.

Results

We undertook a meta-analysis of 17 GWAS imputed using the 1000 Genomes Project¹⁸ Phase-1 reference panel in a study of 38,199 individuals of European ancestry in stage 1 (Fig. 1), of which 19,532 were individuals not included in the discovery stage of previous meta-analyses of lung GWAS¹⁰⁻¹³. Characteristics of cohort participants, genotyping and imputation are shown in Supplementary Table 1. Each study adjusted FEV₁, FEV₁/FVC and FVC, for age, age², sex, height and principal components for population structure, separately for never and ever smokers. Fourteen studies additionally undertook analyses for X chromosome variants (33,009 individuals, Supplementary Fig. 1 and Methods). Inverse normally transformed residuals were then used for association testing within each smoking stratum, assuming an additive genetic effect. Within each study, we combined smoking strata association summary statistics using inverse variance-weighted fixed-effects meta-analysis, and applied genomic control¹⁹ to account for residual population structure not accounted for by principal components. We subsequently combined study-specific estimates across studies using inverse variance weighting, and applied genomic control¹⁹ after fixed-effects meta-analysis. The genomic inflation factor across autosomal variants was 1.03 for each of the three traits, and

STAGE 1 (genome-wide association studies)

n=38,199

B58C (*n*=5,934)
 BHS1&2 (*n*=4,355)
 CROATIA-Korcula (*n*=826)
 CROATIA-Split (*n*=493)
 CROATIA-Vis (*n*=925)
 EPIC population based (*n*=2,339)
 GS:SFHS (*n*=8,093)
 H2000 (*n*=821)
 KORA F4 (*n*=1,474)
 KORA S3 (*n*=1,147)
 LBC1936 (*n*=991)
 NFBC1966 (*n*=4,563)
 NSPHS (*n*=871)
 ORCADES (*n*=1,802)
 SAPALDIA (*n*=1,378)
 SHIP (*n*=1,768)
 YFS (*n*=419)

STAGE 2 (follow-up studies)

n=54,550

ECRHS (*n*=1,747)
 PIVUS (*n*=837)
 TwinsUK (*n*=3,023)
 UK BiLEVE (*n*=48,943)



Figure 1 | Study design for autosomal chromosome analyses. The discovery stage (stage 1) included 17 studies and 38,199 individuals. Fifty-five variants were followed up in stage 2, which comprised four studies and 54,550 individuals.

across X chromosome variants was 1.04 for FEV₁ and 1.00 for FEV₁/FVC and FVC. Quantile–quantile plots are presented in Supplementary Fig. 2a. Variants with effective sample sizes (N effective, product of sample size and imputation quality summed across studies) < 70% were filtered out, and a total of 8,694,268 variants were included in this genome-wide study.

Forty-eight SNPs and seven indels in independent autosomal chromosome regions (± 500 kb either side of sentinel variant) with stage 1 $P < 5 \times 10^{-6}$ were followed up in stage 2 using *in silico* data from four studies comprising 54,550 individuals (Fig. 1; Supplementary Table 2). One SNP on the X chromosome also met these criteria and was followed up in a subset of three studies comprising 52,359 individuals (Supplementary Fig. 1; Supplementary Table 2). Characteristics of follow-up (stage 2) cohort participants, genotyping and imputation are shown in Supplementary Table 1. Stage-2 studies adjusted the traits for age, age², sex, height and principal components to account for population structure and ever-smoking status, and also undertook association testing on the inverse normally transformed residuals assuming additive genetic effects. Stage-2 estimates were combined across studies, and then with stage-1 estimates, using inverse variance-weighted fixed-effects meta-analysis. Thirteen SNPs and three indels, each representing new signals of association, met a genome-wide significance threshold corrected for multiple testing ($P < 5 \times 10^{-8}$) after combining stage-1 and stage-2 results (Table 1; Fig. 2), of which 10 SNPs and three indels achieved independent replication meeting a Bonferroni-corrected threshold for 56 tests ($P < 8.93 \times 10^{-4}$) in stage 2 alone.

Sixteen novel association signals for FEV₁, FEV₁/FVC and FVC.

Of the 16 novel signals reaching genome-wide significance, two represent distinct new signals for FEV₁/FVC in previously reported loci^{10,12} (stage-1 P value conditioned on previously reported variant $< 5 \times 10^{-6}$). Among the remaining 14, five new loci were identified for FEV₁, six new loci for FEV₁/FVC and three new loci for FVC (Table 1). The sentinel variants at the 16 loci were in or near the following genes: *MCL1-ENSA* (1q21.3), *LYPLAL1-RNU5F-1* (1q41), *KCNS3-NT5C1B* (2p24.2), *AK097794* (3q25.32), *NPNT* (4q24), *GPR126-LOC153910* (6q24.1), *ASTN2* (9q33.1), *LHX3* (9q33.1), *PTHLH-CCDC91* (12p11.22), *TBX3* (12q24.21), *TRIP11* (14q32.12), *RIN3* (14q32.12), *EMP2-TEKT5* (16p13.13),

Table 1 | Variants associated with FEV₁, FEV₁/FVC or FVC.

rs number (chr:position)	Gene (function)	eQTL gene	Coded allele/ other allele	Measure	Stage 1				Stage 2				Meta-analysis	
					Coded allele freq.	N	Beta (s.e.)	P	Coded allele freq.	N	Beta (s.e.)	P	Beta (s.e.)	P
rs6681426 (chr1:150586971)	<i>MCL1</i> (dist. = 34,757), <i>ENSA</i> (dist. = 7,628)	<i>ARNT</i>	G/A	FEV ₁	0.36	37,944	0.042 (0.008)	1.07 × 10 ⁻⁷	0.35	5,4301	0.021 (0.006)	1.13 × 10 ⁻³	0.029 (0.005)	4.35 × 10 ⁻⁹
rs201204531 (chr1:219963090)	<i>LYPLAL1</i> (dist. = 576,883), <i>RNUSF-1</i> (dist. = 83,529)	—	A/ATG	FEV ₁ /FVC	0.41	34,866	-0.038 (0.008)	1.98 × 10 ⁻⁶	0.38	53,760	-0.027 (0.006)	1.77 × 10⁻⁵	-0.031 (0.005)	2.68 × 10 ⁻¹⁰
rs61067109 (chr2:18292452)	<i>KCNK53</i> (dist. = 178,227), <i>NTSC1B</i> (dist. = 443,537)	—	G/A	FEV ₁ /FVC	0.77	37,416	-0.050 (0.009)	3.09 × 10 ⁻⁸	0.77	54,341	-0.042 (0.007)	6.56 × 10⁻⁹	-0.045 (0.006)	1.40 × 10 ⁻¹⁵
rs6441207 (chr3:158282459)	<i>AK097794</i> (ncRNA_intronic)	<i>MLF1</i>	C/T	FVC	0.59	36,173	0.036 (0.008)	4.54 × 10 ⁻⁶	0.59	53,174	0.036 (0.006)	5.85 × 10⁻⁹	0.036 (0.005)	1.27 × 10 ⁻¹³
rs6856422 (chr4:106841962)	<i>NPNT</i> (intronic)	—	G/T	FEV ₁ /FVC	0.53	31,446	-0.044 (0.008)	1.30 × 10 ⁻⁷	0.56	49,026	-0.055 (0.006)	1.15 × 10⁻¹⁷	-0.051 (0.005)	1.51 × 10 ⁻²³
rs148274477 (chr6:142838173)	<i>GPR126</i> (dist. = 70,770), <i>LOC153910</i> (dist. = 9,419)	—	C/T	FEV ₁ /FVC	0.98	30,398	-0.161 (0.027)	2.68 × 10 ⁻⁹	0.97	50,047	-0.162 (0.019)	5.63 × 10⁻¹⁸	-0.162 (0.015)	9.58 × 10 ⁻²⁶
rs34886460 (chr9:119359372)	<i>ASTN2</i> (intronic)	—	T/TA	FEV ₁ /FVC	0.53	37,567	0.041 (0.008)	6.75 × 10 ⁻⁸	0.54	53,920	0.025 (0.006)	4.07 × 10⁻⁵	0.031 (0.005)	4.72 × 10 ⁻¹¹
rs2274116 (chr9:139094805)	<i>LHX3</i> (exonic)	<i>QSOX2</i>	C/T	FVC	0.67	32,004	0.041 (0.009)	2.68 × 10 ⁻⁶	0.66	51,730	0.038 (0.007)	4.20 × 10⁻⁹	0.039 (0.005)	5.55 × 10 ⁻¹⁴
rs11383346 (chr12:28283187)	<i>PTHLH</i> (dist. = 158,271), <i>CCDC91</i> (dist. = 126,946)	<i>CCDC91</i>	A/AT	FVC	0.41	37,254	-0.043 (0.008)	3.83 × 10 ⁻⁸	0.41	53,931	-0.041 (0.006)	4.42 × 10⁻¹¹	-0.042 (0.005)	9.52 × 10 ⁻¹⁸
rs10850377 (chr12:115201436)	<i>TBX3</i> (dist. = 79,467), <i>MED13L</i> (dist. = 1,194,945)	—	G/A	FEV ₁	0.67	37,268	-0.047 (0.008)	6.68 × 10 ⁻⁹	0.65	52,722	-0.028 (0.006)	1.55 × 10⁻⁵	-0.035 (0.005)	2.50 × 10 ⁻¹²
rs7155279 (chr14:92485881)	<i>TRIP11</i> (intronic)	<i>ATXN3</i>	G/T	FEV ₁	0.64	37,691	-0.041 (0.008)	1.39 × 10 ⁻⁷	0.64	54,471	-0.022 (0.006)	4.17 × 10⁻⁴	-0.030 (0.005)	1.41 × 10 ⁻⁹
rs117068593 (chr14:93118229)	<i>RIN3</i> (exonic)	—	C/T	FEV ₁	0.82	34,496	-0.048 (0.010)	2.72 × 10 ⁻⁶	0.81	52,572	-0.027 (0.008)	5.68 × 10⁻⁴	-0.035 (0.006)	2.25 × 10 ⁻⁸
rs12149828 (chr16:10706328)	<i>EMP2</i> (dist. = 31,789), <i>TEKT5</i> (dist. = 15,033)	—	G/A	FEV ₁ /FVC	0.83	33,999	0.049 (0.010)	3.16 × 10 ⁻⁶	0.83	50,807	0.035 (0.008)	3.33 × 10⁻⁵	0.040 (0.007)	7.65 × 10 ⁻¹⁰
rs113473882 (chr19:41124155)	<i>LTBP4</i> (intronic)	—	T/C	FEV ₁ /FVC	0.98	32,207	-0.174 (0.033)	1.48 × 10 ⁻⁷	0.99	52,907	-0.138 (0.028)	1.00 × 10⁻⁶	-0.153 (0.021)	9.95 × 10 ⁻¹³
rs134041 (chr22:28056338)	<i>MIAT</i> (dist. = 983,898), <i>MNI</i> (dist. = 87,927)	—	T/C	FEV ₁	0.43	37,669	-0.045 (0.008)	4.19 × 10 ⁻⁹	0.44	52,770	-0.018 (0.006)	4.19 × 10 ⁻³	-0.028 (0.005)	3.03 × 10 ⁻⁹
rs7050036 (chrX:15964845)	<i>APIS2</i> (dist. = 91,708), <i>GRPR</i> (dist. = 176,579)	—	T/A	FEV ₁ /FVC	0.38	32,285	-0.041 (0.007)	4.37 × 10 ⁻⁹	0.39	50,530	-0.013 (0.005)	1.60 × 10 ⁻²	-0.023 (0.004)	4.14 × 10 ⁻⁸

Chr., chromosome; dist., distance; eQTL, expression quantitative trait loci; FDR, false discovery rate; FEV₁, forced expiratory volume in 1s; freq., frequency; FVC, forced vital capacity. Results are shown for each sentinel variant associated ($P < 5 \times 10^{-8}$) with FEV₁, FEV₁/FVC or FVC in a joint analysis of up to 92,744 individuals of European ancestry, ordered by chromosome and position. The allele frequency presented in the table corresponds to the coded allele (Coded allele freq.), which was chosen as the reference allele in the 1000 Genomes Project reference panel, and in most instances (but not always) is the major allele. Two-sided *P* values are given for stage 1, stage 2 and the meta-analysis of both stages. Variants reaching independent replication in stage 2 ($P = 0.05/56 = 8.93 \times 10^{-4}$) are indicated with their stage-2 *P* value in bold. The sample sizes (N) shown are the effective sample sizes. The effective sample size within each study is the product of sample size and the imputation quality metric. Beta values reflect effect-size estimates on an inverse-normal transformed scale after adjustments for age, age², sex, height and ancestry principal components, and stratified by ever-smoking status. 'eQTL gene' presents a gene expressed in the lung, whose expression is associated with the highest ranking proxy for the sentinel variant (ranking on *r*² with sentinel first, all with *r*² > 0.8, and on eQTL *P*-value second, all with $P < 5 \times 10^{-5}$) with FDR < 10%.

LTBP4 (19q13.2), *MIAT-MNI* (22q12.1) and on chromosome X, *APIS2-GRPR* (Xp22.2) (Supplementary Fig. 2b,c). To gain further insight into the associated variants, we assessed whether the novel sentinel variants, or their proxies, were associated with gene expression in lung tissues²⁰ and blood²¹ (Methods, Supplementary Methods; Supplementary Table 3a) or were in DNase hypersensitivity sites²² in relevant cell types (Methods; Supplementary Table 3b). For relevant genes, we investigated RNA-seq splice isoforms in human bronchial epithelial cells (Supplementary Methods; Supplementary Fig. 3), searched for evidence of protein expression in the respiratory system²³ (Supplementary Table 3c), assessed differential expression across the pseudoglandular and canalicular stages of fetal human lung development (Methods; Supplementary Table 3d) and assessed evidence for differences in gene expression in bronchial epithelial brush samples from COPD cases and smoking controls (Methods; Supplementary Table 3e).

The two novel signals in known loci were the strongest ($P < 5 \times 10^{-23}$) association signals after meta-analysing

stage 1 and 2. The strongest signal was for a low-frequency SNP near *GPR126* (rs148274477, MAF = 2.4%, intergenic on chromosome 6) associated with FEV₁/FVC ($P = 9.6 \times 10^{-26}$, Table 1) and in high linkage disequilibrium (LD, *r*² = 0.85) with a missense variant (rs17280293 (Ser123Gly), Supplementary Table 3f) in *GPR126*, but distinct from the previously reported signal for FEV₁/FVC in this region^{10,13} (stage-1 *P* value for rs148274477 conditioning on rs3817928 (ref. 10) and rs262129 (ref. 13) = 1.86×10^{-7} , unconditional stage-1 $P = 2.68 \times 10^{-9}$). *GPR126* encodes a G-protein-coupled receptor and is expressed in adult and fetal lung tissue^{24,25} (Supplementary Table 3d). Other studies have shown that *GPR126* is required for mice embryonic viability and cardiovascular development²⁶, and that *GPR126* is expressed in adult mice lung²⁷. More recently, *GPR126* has been shown to bind type-IV collagen, a major collagen in the lung, leading to cAMP signalling²⁸.

The second strongest signal ($P = 1.5 \times 10^{-23}$, Table 1) was an intronic SNP (rs6856422) in *NPNT* on chromosome 4 associated with FEV₁/FVC, distinct from the previously discovered signal for

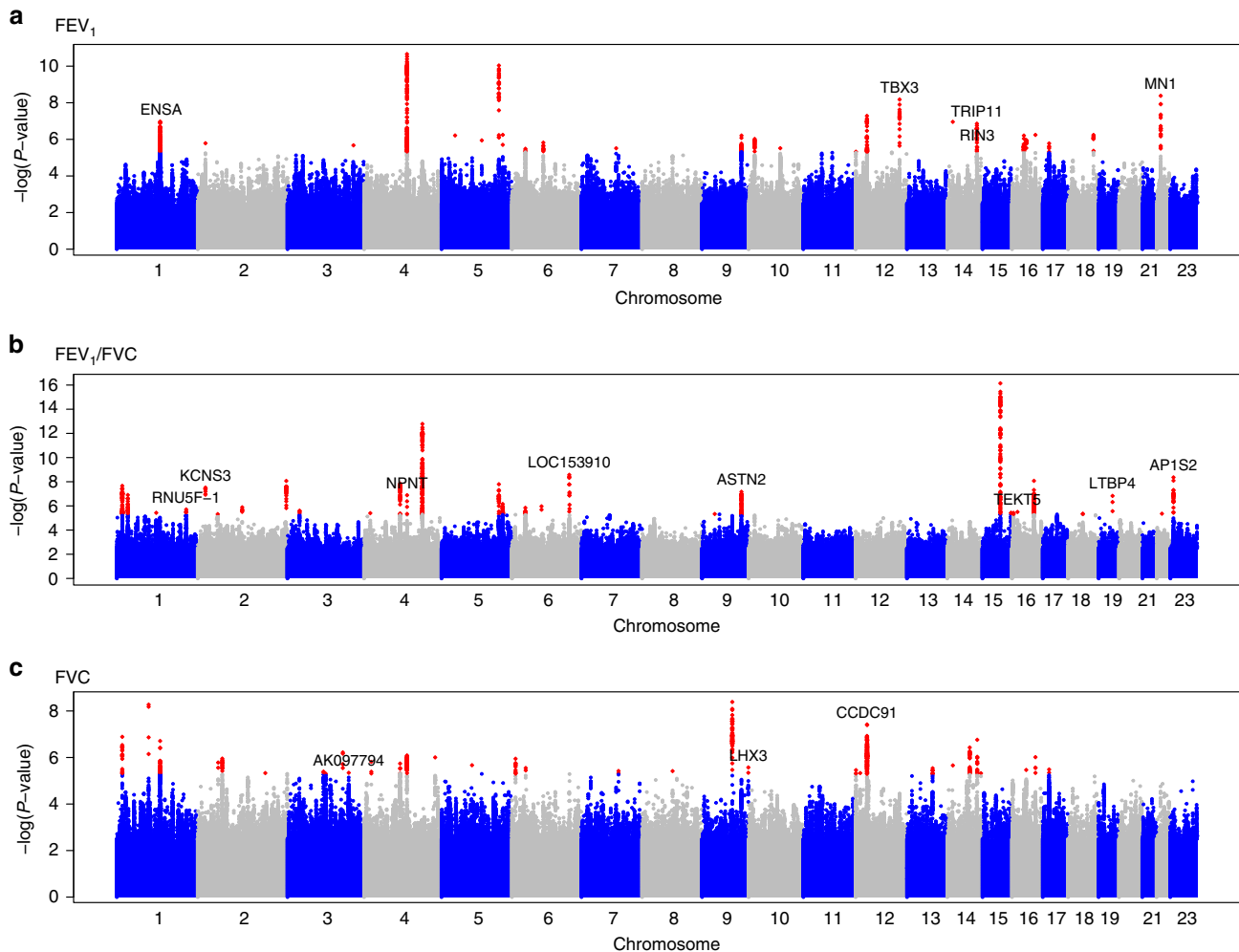


Figure 2 | Manhattan plots for association results. (a) FEV₁, (b) FEV₁/FVC and (c) FVC. Manhattan plots ordered by chromosome and position for stage-1 results. Variants with $P < 5 \times 10^{-6}$ are indicated in red. Novel signals that reached genome-wide significance after meta-analysing stage 1 and stage 2 are labelled with the nearest gene. Only variants with N effective $\geq 70\%$ are presented here.

FEV₁ in this region^{10,12,13}. The stage-1 P value for this variant conditioned on the previously reported sentinel SNPs (rs17036341 (ref. 10) and rs10516526 (refs 12,13)) and on the sentinel SNP for FEV₁ in this analysis (rs12374256, *INTS12* intron) was 4.7×10^{-6} (unconditional stage-1 $P = 1.30 \times 10^{-7}$). Proxies of the sentinel SNP were associated with expression of *INTS12* and *GSTCD* in blood (Supplementary Table 3a). *INTS12*, *GSTCD* and *NPNT* are contiguously positioned at 4q24, and are all expressed in adult and fetal lung tissues (Supplementary Table 3c,d). Our previous work characterizing *GSTCD* and *INTS12* showed that they are oppositely transcribed genes that are to some extent co-ordinately regulated, although while *GSTCD* expression in human lung tissue is ubiquitous, *INTS12* expression was predominantly in the nucleus of epithelial cells and pneumocytes²⁹.

Among the 14 novel loci, six novel loci were associated with FEV₁/FVC. One of them was a low-frequency variant (rs113473882, intronic in *LTBP4* on chromosome 19, MAF = 1.5%, Table 1) in almost complete LD ($r^2 = 0.99$) with a missense variant (rs34093919, Asp752Asn, Supplementary Table 3f) in *LTBP4*, which encodes a protein that binds transforming growth factor beta (TGF β) as it is secreted and targeted to the extracellular matrix. Mice deficient in *ltbp4* displayed defects in lung septation and elastogenesis, which may be TGF β 2 and fibulin-5 dependent³⁰, and disruption of this gene in mice led to abnormal lung development, cardiomyopathy and

colorectal cancer³¹. Variants near *LTBP4*, uncorrelated ($r^2 < 0.05$) with the sentinel SNP we report here, have been associated with COPD³² and smoking behaviour³³. A further novel FEV₁/FVC locus mapping near *AP1S2* is the first to be reported for lung function on the X chromosome; sentinel SNP (rs7050036, intergenic) proxies were associated with the expression of *AP1S2* and *ZRSR2* in lung tissue (Supplementary Table 3a). Other new loci for FEV₁/FVC were in or near *KCNS3* (2p24.2), *ASTN2* (9q33.1), *RNU5F-1* (1q41) and *TEKT5* (16p13.13).

The strongest signal for FEV₁ in a novel locus was upstream of *TBX3* on chromosome 12 (Table 1); *TBX3* is involved in the TGF β 1 signalling pathway³⁴. At a second novel locus for FEV₁ (rs7155279, *TRIP11* intron on chromosome 14, Table 1), proxies of the sentinel variant were associated with lung and blood expression of *TRIP11*. *TRIP11* encodes a protein associated with the Golgi apparatus³⁵. In the lung, rs7155279 showed strongest association with expression of *ATXN3* (Supplementary Table 3a), which encodes ataxin 3, a deubiquitinating enzyme. Expanded trinucleotide repeats in *ATXN3* cause spinocerebellar ataxia-3 (ref. 36). In blood, a proxy ($r^2 = 0.94$) for rs7155279 showed strong association ($P = 3 \times 10^{-34}$, Supplementary Table 3a) with the expression of *FBLN5*. Fibulin-5 was shown to be implicated in tissue repair in COPD³⁷ and elastogenesis and lung development³⁰. A third signal for FEV₁ was a missense variant (rs117068593, Arg279Cys, Supplementary Table 3f) in *RIN3* on

chromosome 14 (Table 1), which was ~ 632 kb from the *TRIP11* sentinel SNP (rs7155279) and independent from it ($r^2 = 8.84 \times 10^{-5}$). Although this is the first report of association of a *RIN3* variant with lung function, a correlated variant (rs754388, $r^2 = 0.99$) was recently associated with moderate to severe COPD, although the association did not replicate in an independent study³⁸. In a fourth novel region for FEV₁, on chromosome 1, a sentinel SNP, rs6681426, ~ 8 kb downstream of *ENSA* (Table 1) and a second signal ~ 700 kb apart (rs4926386, Supplementary Table 2a) were both associated with *ARNT* expression in lung (Supplementary Table 3a). *ARNT* is differentially expressed during fetal lung development (Supplementary Table 3d) and acts as a co-factor for transcriptional regulation by hypoxia-inducible factor 1 during lung development³⁹ and may regulate cytokine responses⁴⁰. SNP rs6681426 was also associated with the expression of *LASS2* (also known as *CERS2*) in lung tissue (Supplementary Table 3a); *lass2* knock-out mice develop lung inflammation and airway obstruction⁴¹. The other new locus for FEV₁ was near *MN1* (22q12.1).

All three novel loci for FVC had sentinel variants or close proxies associated with expression of a nearby gene in lung, implicating *CCDC91*, *MLF1* and *QSOX2*, located on chromosomes 12, 3 and 9, respectively. The putative function of the key genes in each of the two known and 14 novel loci for FEV₁, FEV₁/FVC and FVC are summarized in Supplementary Table 4.

Functional characterization of novel signals. The protein products of genes nearest to the sentinel variant of novel signals for lung function were expressed in bronchial epithelial cells, pneumocytes or lung macrophages (Supplementary Table 3c). Among the 16 novel signals of association with lung function, sentinel variants or close proxies were *cis* expression quantitative trait loci (eQTLs) in lung for *ARNT*, *MLF1*, *QSOX2*, *CCDC91* and *ATXN3* (Table 1; Supplementary Table 3a), and in eight loci the sentinel variant or at least one strong proxy ($r^2 > 0.8$) was in a DNase hypersensitivity site in a cell type potentially relevant to lung function (in or near *ENSA*, *RNU5F-1*, *ASTN2*, *CCDC91*, *TBX3*, *RIN3*, *TEKT5* and *MN1*, Supplementary Table 3b). The sentinel variant association was explained (conditional $P > 0.01$) by a missense variant in each of the two novel signals in which we detected a low-frequency sentinel variant (near *GPR126* and in *LTBP4*), and was explained in four of the remaining novel signals by a putatively functional variant (in or near *ENSA*, *AK097794*, *TEKT5* and *MN1*, Supplementary Table 3f and Methods). Genes in four of the novel loci showed differential expression across the pseudoglandular and canalicular stages of fetal lung development, particularly *EMP2* (Supplementary Table 3d). *MLF1* and *ATXN3* showed differences in expression levels in bronchial brushings between COPD cases and controls (Supplementary Table 3e). We detected novel splice isoforms of $> 20\%$ abundance for *GFM1*, *TRIM32*, *LTBP4* and *MN1* in human bronchial epithelial cells (Supplementary Fig. 3; Supplementary Methods).

Association in children. To assess whether the 16 new sentinel variants associated with lung function in adults may act through an effect on lung development, we assessed their association in the ALSPAC study⁴² that includes 5,062 children (Supplementary Table 5a). Eleven of the 16 sentinel variants showed consistent directions of effect in adults and children. The association with FVC of variant rs6441207 on chromosome 3 in the noncoding RNA *AK097794* exceeded a Bonferroni-corrected threshold for 16 tests (Supplementary Table 5a).

Association with smoking and gene by smoking interaction. The 16 new variants had consistent effect sizes in never smokers

and ever smokers, and no gene–smoking interaction ($P > 0.05$) in stage 1 (Supplementary Table 5b). We found no evidence that any of these signals were driven by smoking behaviour. Only the two-base-pair insertion on chromosome 1 (rs201204531) revealed an association ($P = 1.5 \times 10^{-3}$) with smoking behaviour (heavy-versus never-smoking status) that met a Bonferroni-corrected threshold for 16 tests (Supplementary Table 5c). However, this variant also showed an association with FEV₁/FVC in never smokers, and the allele associated with higher likelihood of being a smoker was associated with increased FEV₁/FVC (Supplementary Table 5b,c).

Associations with other traits. We queried the GWAS catalog⁴³ for variants in 2-Mb regions centred on the sentinel variant for the 16 loci (Supplementary Table 5d). Five loci contained variants associated with height^{44–46} (Supplementary Table 5d). In the *GPR126* and *LHX3* loci, the previously reported height variants were not correlated ($r^2 < 0.2$) with the lung function variants reported here. In the *AK097794*, *CCDC91* and *TRIP11* loci, the variants associated with height were correlated ($r^2 > 0.3$) with the lung function sentinel variants, but the alleles associated with reduced height were associated with increased FEV₁ or FVC. Associations with other traits have been reported for variants in LD ($r^2 > 0.3$) with sentinel variants in regions of *RIN3* (Paget's disease⁴⁷ and bone mineral density⁴⁸), *ENSA* (body fat mass⁴⁹ and melanoma⁵⁰) and *LHX3* (thyroid hormone levels⁵¹). None of the novel signals relate to known asthma loci, and the association findings were consistent after removing individuals with asthma (Supplementary Fig. 4).

Genetic architecture of lung function traits. The proportion of the additive polygenic variance explained by the 49 signals discovered to date (Supplementary Table 6), including new and previously reported signals^{10–14} is 4.0% for FEV₁, 5.4% for FEV₁/FVC and 3.20% for FVC (Supplementary Table 7). These estimates are likely upper bounds on the proportion of the variance explained due to the winner's curse bias. Across the 49 signals, we observed larger effect sizes for associations with lower-frequency variants (Fig. 3), supporting the hypothesis that lower-frequency variants will contribute to explaining the missing heritability¹⁶.

We examined the increase in coverage of low-frequency and common variants by the 1000 Genomes Project reference panel, compared with the HapMap imputation reference panel, at both the novel and previously reported loci (Supplementary Fig. 5a). The two association signals where the 1000 Genomes sentinel variants had low MAF ($< 5\%$), were not present when restricting the results only to variants that could be imputed using the HapMap imputation panel (rs113473882 and rs148274477 in Supplementary Fig. 5a).

For each of the 32 previously discovered regions^{10–14}, we identified the most strongly associated variant present on the 1000 Genomes Project¹⁸ reference panel and the most strongly associated variant present on the HapMap reference panel using stage-1 results, and compared the stage-1 MAFs between these two groups of variants. The 1000 Genomes sentinel variants in or near *GPR126* (rs148274477), *TGFB2* (rs147187942) and *MMP15* (rs150232756) had MAFs that were more than twofold lower than the HapMap sentinel variant MAFs (Supplementary Fig. 5b) and were statistically independent ($r^2 \leq 0.06$) from the previously discovered HapMap-imputed sentinel variants¹³. The *GPR126* 1000 Genomes-imputed sentinel was described above as one of the 16 new signals. We tested the association of the 1000 Genomes-imputed sentinel variants near *TGFB2* and *MMP15* in UK BiLEVE (Supplementary Table 8), and found supportive

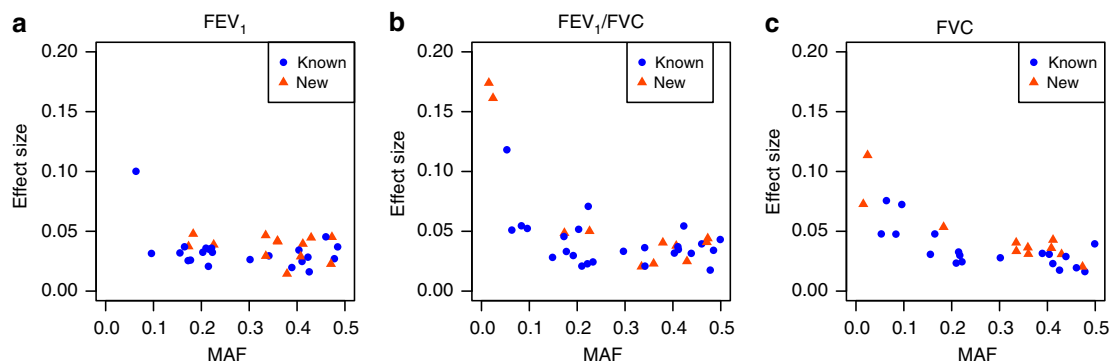


Figure 3 | Minor allele frequency against effect-size plots (a) FEV_1 , (b) FEV_1/FVC and (c) FVC. MAF is plotted against stage-1 effect sizes for variants within the 33 known^{10–14} and the 16 new signals, which had stage-1 $P < 0.05$ for association with FEV_1 , FEV_1/FVC and FVC separately. Known signals are represented with blue circles and new signals are represented with orange triangles.

evidence of association for the signal near *TGFB2* (rs147187942, MAF = 9%, $P = 5.7 \times 10^{-3}$).

Pathway analyses. We undertook a pathway analysis using MAGENTA v2 (ref. 52) and stage-1 genome-wide results for FEV_1 , FEV_1/FVC and FVC (Supplementary Methods). For FVC, the platelet-derived growth factor signalling, and the chromatin-packaging and -remodelling pathways were significant ($P = 2 \times 10^{-4}$, false discovery rate (FDR) < 0.3% and $P = 1.82 \times 10^{-4}$, FDR < 4%, respectively) (Supplementary Table 9).

Discussion

In this study, we aimed to improve coverage of low-frequency variants and detect novel loci associated with lung function, by undertaking imputation of GWAS data in 17 studies and 38,199 individuals to the 1000 Genomes Project¹⁸ reference panel, and by following up the most significant signals in an additional 54,550 individuals. Overall, 16 new association signals attained a genome-wide significance threshold corrected for multiple testing ($P < 5 \times 10^{-8}$) after meta-analysing stage 1 and stage 2, including 15 autosomal and one X chromosome signal. While two of the new findings relate to novel signals for FEV_1/FVC in previously reported regions^{10,12}, five new loci were identified for FEV_1 , six new loci for FEV_1/FVC and three new loci for FVC. Including the 16 signals discovered in these analyses, the number of lung function signals discovered to date is 49 (refs 10–14), and they jointly explain a modest proportion of the additive polygenic variance (4.0% for FEV_1 , 5.4% for FEV_1/FVC and 3.2% for FVC).

Some of the 49 distinct lung function signals^{10–14} seem to cluster close to each other. If we define regions as 500 kb either side of the sentinel variants, there are three regions that each include two distinct signals (in or near *INTS12-GSTCD-NPNT*, *GPR126* and *PTCH1* (refs 10,12)), so that the 49 signals would map to 46 loci. If we use a wider definition of region (1,000 kb either side of the sentinel), there are four regions that each include two distinct signals (in or near *INTS12-GSTCD-NPNT*, *GPR126*, *PTCH1* (refs 10,12) and *TRIP11-RIN3*). In addition, the human leukocyte antigen region on chromosome 6 includes three distinct signals (in or near *ZKSCAN3-NCR3-AGER*^{10,12,13}) within 3.8 Mb. Furthermore, we have shown evidence of an additional signal in the *TGFB2* region, and the new lung function signal in *LTBP4* lies 179 kb away from a known COPD signal³². These findings are consistent with reports from very large studies of height and lipids^{53,54}, which report multiple signals in associated regions, and highlight the importance of taking into account LD between variants to improve our understanding of

known regions. Multiple signals within known regions are likely to explain some of the hidden heritability of these traits.

To identify pathways relevant to lung function, we undertook additional analyses using MAGENTA, which have implicated pathways for platelet-derived growth factor signalling and chromatin-packaging and -remodelling. Independent analyses undertaken in a concurrent study by the UK BiLEVE consortium, which focused on the extremes of the lung function distribution⁵⁵, highlight the histone subset of the chromatin-packaging and -remodelling pathway. The *TGF β* signalling pathway has now been implicated by three independent loci: an FEV_1/FVC signal explained by a missense variant in *LTBP4*, which encodes a protein that binds *TGF β* ; an FEV_1 signal upstream of *TBX3*, which is involved in the *TGF β* 1 signalling pathway³⁴, and a previously reported signal downstream of *TGFB2* (ref. 17). In addition, a pathway involving fibulin-5 has been implicated by two of the novel loci (*LTBP4* and *TRIP11*). The identification, through different approaches, of pathways which appear to be involved in determining lung function should help focus future functional studies.

Pathways affecting lung function also have the potential to affect COPD risk, since lung function measures are used to diagnose the disease. Currently, 13 signals (in or near *TGFB2*, *TNS1*, *RARB*, *FAM13A*, *GSTCD*, *HHIP*, *HTRA4*, *ADAM19*, *AGER*, *LOC153910*, *C10orf11*, *RIN3* and *THSD4*) out of the 49 lung function signals discovered to date^{10–14} have also shown association with some definition of COPD^{38,56–60}. This illustrates that the study of lung function measures is a powerful approach to bring insights into the genetics of COPD.

In agreement with previous findings for other lung function loci^{12,13}, none of the 16 new associations seem to be driven by either smoking behaviour or by a gene–smoking interaction. One variant showed association with smoking behaviour that met a Bonferroni correction for 16 tests in UK BiLEVE. This variant also had an effect in never smokers in stage 1, and the allele associated with increased lung function was also associated with increased risk of smoking, which does not suggest an association with lung function mediated by smoking behaviour. Variants in five out of the 16 loci associated with lung function in this study have also shown associations with height^{44–46}. However, the variants associated with height were either independent of those associated with lung function, or if they were correlated, the alleles associated with increased height, were associated with decreased FEV_1 or FVC. If the association with lung function was driven by an effect on height, we would expect consistent direction of effect between these two traits. Therefore, the associations identified for lung function in these regions are not likely to be driven by associations with height.

This study had a large follow-up stage, which included 54,550 individuals, of which 48,943 were contributed by the UK BiLEVE

study. UK BiLEVE is a particularly powerful study since it has sampled UK Biobank individuals from the extremes of the lung function distribution, and it has spirometry performed in a uniform way across individuals. Had these data been available when we undertook the discovery stage of this study, their addition would have greatly improved the discovery power. Nevertheless, incorporating these data into the follow-up stage improved power to provide replication and deal with potential winners' curse bias. Another strength of the current study design was the increased coverage of common and low-frequency variants obtained through the imputation to 1000 Genomes Project¹⁸ reference panel. This enabled us to detect two low-allele-frequency variants (with MAF of 1.5 and 2.4% and stage-1 effect sizes of 0.17 and 0.16 s.d. units, respectively) that have an effect on lung function. No associations with lower allele frequency variants have been detected in this study, despite having power > 80% in discovery to detect associations ($P < 5 \times 10^{-6}$) for variants with MAF of 0.5 and 1%, and effect sizes above 0.3 and 0.2 s.d. units, respectively. The poorer imputation quality for low-allele-frequency variants coupled with the strict criteria we used to select variants for follow-up (N effective $\geq 70\%$) have probably affected our ability to detect rare variants. For instance, a variant representing an additional signal for FEV₁/FVC in the *GSTCD-INTS12-NPNT* region, reported by the UK BiLEVE study, where it was directly genotyped⁵⁵, would have been detected in this analysis had we used a more lenient threshold (N effective > 60%). Imputation quality for rare variants will improve as larger imputation reference panels become available.

In summary, 16 new association signals for lung function have been identified in this study, including two signals explained by non-synonymous low-frequency variants. These findings highlight new loci not previously connected with lung function or COPD, and bring new insights into previously detected loci. This study also highlights the added value of imputing to new reference panels as they become available. Understanding the molecular pathways that connect the newly identified loci with lung function and COPD risk has the potential to point to new targets for therapeutic intervention.

Methods

Study design. The study consisted of two stages. Stage 1 was a meta-analysis of 17 GWAS in a total of 38,199 individuals of European ancestry. Supplementary Table 1 gives the details of these studies. Fifty-six variants selected according to the results in stage 1 were followed up in stage 2 in 54,550 European individuals.

Stage-1 samples. Stage 1 comprised 17 studies: B58C (T1DGC and WTCCC), BHS1 and -2, EPIC (obese cases and population-based studies), the EUROSPAN studies (CROATIA-Korcula, ORCADES, CROATIA-Split and CROATIA-Vis), GS:SFHS, Health 2000, KORA F4, KORA S3, LBC1936, NFBC1966, NSPHS, SAPALDIA, SHIP and YFS (see Supplementary Table 1a for the definitions of all abbreviations). All participants provided written informed consent and studies were approved by local Research Ethics Committees and/or Institutional Review Boards. Measurements of spirometry for each study are described in the Supplementary Note. The genotyping platforms and quality-control criteria implemented by each study are described in Supplementary Table 1b.

Imputation. Imputation to the all ancestries 1000 Genomes Project¹⁸ Phase-1 reference panel released in March 2012 was undertaken using MACH⁶¹ and minimac⁶² or IMPUTE2 (ref. 63) with pre-imputation filters and parameters as shown in Supplementary Table 1b. Specific software guidelines were used to impute the non-pseudoautosomal part of the X chromosome. The pseudoautosomal part of the X chromosome was not included in these analyses. Variants were excluded if the imputation information, assessed using r2.hat (MACH and minimac) or .info (IMPUTE2), was < 0.3.

Data transformation and association testing in stage 1. Linear regression of age, age², sex, height and principal components for population structure was undertaken on FEV₁, FEV₁/FVC and FVC separately for ever smokers and never smokers. The residuals were transformed to ranks and then transformed to normally distributed z-scores. These transformed residuals were then used as the

phenotype for association testing under an additive genetic model, separately for ever smokers and never smokers. For X chromosome analyses, residuals for males and females were analysed separately and dosages for males were coded 0 for 0 copies of the coded allele and 2 for 1 copy of the coded allele. The software used was specified in Supplementary Table 1b. Studies with related individuals analysed ever smokers and never smokers together adjusting the regression for ever-smoking status and used appropriate tests for association in related individuals, as described in the Supplementary Note.

Meta-analysis of stage-1 data. Quality-control checks on the stage-1 data were undertaken using GWAToolbox⁶⁴ and R version 3.0.2 (see URLs). All meta-analysis steps were undertaken using inverse variance-weighted fixed-effects meta-analysis. Effect estimates were flipped across studies so that the coded allele was the reference allele in the 1000 Genomes Project¹⁸ reference panel. For each study with unrelated individuals, autosomal chromosomes results were meta-analysed between ever smokers and never smokers. After that, all study-specific standard errors were corrected using genomic control¹⁹. Study-specific genomic inflation factor estimates are shown in Supplementary Table 1a. Finally, effect-size estimates and s.e. were combined across studies, and genomic control¹⁹ was applied again at the meta-analysis level. For the X chromosome, studies of unrelated individuals meta-analysed smoking strata estimates within sex strata and then meta-analysed pooled sex strata estimates. After that, genomic control¹⁹ was applied to each study and results were meta-analysed across studies. Genomic control¹⁹ was applied again after the meta-analysis. To describe the effect of imperfect imputation on power, for each variant we report the effective sample size (N effective), which is the sum of the study-specific products of the sample size and the imputation quality metric. Meta-analysis statistics and figures were produced using R version 3.0.2 (see URLs).

Selection of variants for stage 2. Variants with N effective < 70% were filtered out before selecting variants for follow-up (8,916,621 variants remained after filtering). Independent regions (± 500 kb from the sentinel variant) were selected for FEV₁, FEV₁/FVC and FVC if the sentinel SNP or indel had $P < 5 \times 10^{-6}$. If the same variant was selected for different traits, it was followed up for all the traits. If two different variants were selected for different traits within the same region, or if any of the regions selected had already been identified in previous GWAS¹⁰⁻¹⁴ but the sentinel variant was different from that previously reported, conditional analyses were undertaken to assess whether the signals within the same regions were distinct. If previously reported sentinel SNPs for a region were strongly correlated ($r^2 > 0.9$), we only conditioned on the SNP that had shown the strongest association. If two variants were selected for different traits within the same new region, both variants were taken forward if their P -value conditioning on the other variant was $< 5 \times 10^{-6}$; if not, only the variant with the most significant P value was taken forward. Variants within known regions were only taken forward if their P value conditioned on the previously reported variant was $< 5 \times 10^{-6}$. Conditional analyses were undertaken using GCTA⁶⁵, and B58C data were used to estimate LD. In total, 56 variants (49 SNPs and seven indels) were taken forward for follow-up, two of which were distinct signals within previously reported regions^{10,12,13}. These variants are listed in Supplementary Table 2. Previously reported signals¹⁰⁻¹⁴ were not followed up.

Stage-2 samples. The 48 SNPs and seven indels on autosomal chromosomes were followed up in up to 54,550 individuals from four studies with *in silico* data: ECRHS, PIVUS, TwinsUK and UK BiLEVE (see Supplementary Table 1a for the definitions of all abbreviations). All participants provided written informed consent and studies were approved by local Research Ethics Committees and/or Institutional Review Boards. One SNP in the chromosome X was followed up in 52,359 individuals from PIVUS, TwinsUK and UK BiLEVE. Measurements of spirometry for each study are described in the Supplementary Note.

Meta-analysis of stage-2 data. All stage-2 studies undertook linear regression of age, age², sex, height, ever-smoking status and principal components for population structure, if available, on FEV₁, FEV₁/FVC and FVC, then the residuals were transformed to ranks and to normally distributed Z-scores. These transformed residuals were then used as the phenotype for association testing under an additive genetic model. For the X chromosome analyses, allele dosages for hemizygous males were coded as 2. Effect sizes were flipped to be consistent with the stage-1 estimates, using the reference allele in 1000 Genomes Project¹⁸ as the coded allele. Genomic control¹⁹ was applied for studies that undertook the analysis genome-wide. Effect estimates and s.e. were combined across the stage-2 studies using an inverse variance-weighted meta-analysis.

Combination of stage 1 and 2 and multiple testing correction. A meta-analysis of stage-1 and stage-2 results was undertaken using inverse variance-weighted meta-analysis. We take into account the multiple tests undertaken by describing an association as genome-wide significant if it has $P < 5 \times 10^{-8}$. In addition, we assessed whether any of the findings achieved independent replication in stage 2 using a threshold corrected for the number of variants followed up ($0.05/56 = 8.93 \times 10^{-4}$).

Functional characterization of novel loci. A series of analyses were undertaken to provide insights into the expression of genes within the 16 loci (defined as ± 1 Mb either side of the sentinel variant) represented here. Blood²¹ and lung tissue²⁰ eQTL analyses were undertaken for variants in these loci that were in LD ($r^2 > 0.3$) with the sentinel variant in the region. We assessed whether variants within these loci that were strongly correlated with the sentinel variants ($r^2 > 0.8$) were in DNase hypersensitivity sites as defined by ENCODE²² for cells potentially relevant to lung function. We also carried out conditional analyses, using GCTA⁶⁵, of sentinel variants conditioning on functional variants within the loci to assess whether the association signals were explained by functional variants (P value of the sentinel variant conditioned on the functional variant, conditional P , > 0.01). Functional variants were defined using SIFT⁶⁶, PolyPhen-2 (ref. 67), CADD⁶⁸ and GWAVA⁶⁹ databases. Additional analyses were undertaken for a subset of priority genes within the 16 loci (description of the selection is given in the Supplementary Methods). These included RNA-seq analyses to confirm messenger RNA expression in a lung-relevant cell (bronchial epithelium) and detect novel splice isoforms; assessment of differential expression across pseudoglandular and canalicular stages of human fetal lung development using gestational age as a continuous variable in linear regression²⁵, and assessment of differences in expression levels in bronchial brushings between COPD cases and smoking controls⁷⁰. Details for all these analyses are provided in the Supplementary Methods.

Associations with other traits. The association of the 16 sentinel variants with the following traits was assessed: lung function in children undertaking the same analysis as for adults in the ALSPAC data set⁴²; gene by smoking interaction by undertaking a Z-test comparing the effect of a given variant in ever smokers and in never smokers using stage-1 results; smoking behaviour by undertaking a logistic regression analysis with heavy- versus never-smoking status as an outcome in the UK BiLEVE data set. In addition, the GWAS catalog⁴³ was queried for variants in 2-Mb regions centred on the sentinel variant for the 16 loci. Variants that were genome-wide significant ($P < 5 \times 10^{-8}$) in the GWAS catalog⁴³ and were in LD ($r^2 > 0.3$) with the sentinel variants, or were in genes that contained at least one variant in LD ($r^2 > 0.3$) with the sentinel variants were selected.

Pathway analyses. Stage-1 GWAS results were tested for enrichment of known biological pathways using MAGENTA v2 (ref. 52). Six databases of biological pathways, including Ingenuity Pathway (June 2008, number of pathways $n = 81$), KEGG (2010, $n = 186$), PANTHER Molecular Function (January 2010, $n = 216$), PANTHER Biological Processes (January 2010, $n = 217$), PANTHER Pathways (January 2010, $n = 94$) and Gene Ontology (April 2010, $n = 1778$), were tested. An FDR threshold of 5% was used and significance thresholds were Bonferroni corrected for each database. Genes within 500 kb either side from the sentinel variants were flagged in the analysis. Sensitivity analyses were run after removing genes in the human leukocyte antigen region on chromosome 6. More details on the method are provided in the Supplementary Methods.

Additional analyses. Heterogeneity tests were undertaken for the 16 sentinel variants in stage 1. We undertook stepwise conditional analyses as performed by GCTA⁶⁵ in each locus to identify additional signals. Full methods and results are described in the Supplementary Notes.

References

- Hole, D. J. *et al.* Impaired lung function and mortality risk in men and women: findings from the Renfrew and Paisley prospective population study. *BMJ* **313**, 711–715 discussion 715–716 (1996).
- Young, R. P., Hopkins, R. & Eaton, T. E. Forced expiratory volume in one second: not just a lung function test but a marker of premature death from all causes. *Eur. Respir. J.* **30**, 616–622 (2007).
- Lopez, A. D. *et al.* Chronic obstructive pulmonary disease: current burden and future projections. *Eur. Respir. J.* **27**, 397–412 (2006).
- Lozano, R. *et al.* Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet* **380**, 2095–2128 (2012).
- Zappala, C. J. *et al.* Marginal decline in forced vital capacity is associated with a poor outcome in idiopathic pulmonary fibrosis. *Eur. Respir. J.* **35**, 830–836 (2010).
- Abbey, D. E. *et al.* Long-term particulate and other air pollutants and lung function in nonsmokers. *Am. J. Respir. Crit. Care Med.* **158**, 289–298 (1998).
- Global Initiative for Chronic Obstructive Lung Disease. Global Strategy for the Diagnosis Management and Prevention of COPD. <http://www.goldcopd.org/> (2014).
- Wilk, J. B. *et al.* Evidence for major genes influencing pulmonary function in the NHLBI family heart study. *Genet. Epidemiol.* **19**, 81–94 (2000).
- Palmer, L. J. *et al.* Familial aggregation and heritability of adult lung function: results from the Busselton Health Study. *Eur. Respir. J.* **17**, 696–702 (2001).
- Hancock, D. B. *et al.* Meta-analyses of genome-wide association studies identify multiple loci associated with pulmonary function. *Nat. Genet.* **42**, 45–52 (2010).
- Loth, D. W. *et al.* Genome-wide association analysis identifies six new loci associated with forced vital capacity. *Nat. Genet.* **46**, 669–677 (2014).
- Repapi, E. *et al.* Genome-wide association study identifies five loci associated with lung function. *Nat. Genet.* **42**, 36–44 (2010).
- Soler Artigas, M. *et al.* Genome-wide association and large-scale follow up identifies 16 new loci influencing lung function. *Nat. Genet.* **43**, 1082–1090 (2011).
- Wilk, J. B. *et al.* A genome-wide association study of pulmonary function measures in the Framingham Heart Study. *PLoS Genet.* **5**, e1000429 (2009).
- Maher, B. Personal genomes: the case of the missing heritability. *Nature* **456**, 18–21 (2008).
- Manolio, T. A. *et al.* Finding the missing heritability of complex diseases. *Nature* **461**, 747–753 (2009).
- Gibson, G. Rare and common variants: twenty arguments. *Nat. Rev. Genet.* **13**, 135–145 (2011).
- 1000 Genomes Project Consortium. An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56–65 (2012).
- Devlin, B. & Roeder, K. Genomic control for association studies. *Biometrics* **55**, 997–1004 (1999).
- Lamontagne, M. *et al.* Refining susceptibility loci of chronic obstructive pulmonary disease with lung eqtls. *PLoS ONE* **8**, e70220 (2013).
- Westra, H. J. *et al.* Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nat. Genet.* **45**, 1238–1243 (2013).
- Rosenbloom, K. R. *et al.* ENCODE data in the UCSC Genome Browser: year 5 update. *Nucleic Acids Res.* **41**, D56–D63 (2013).
- Uhlen, M. *et al.* Towards a knowledge-based Human Protein Atlas. *Nat. Biotechnol.* **28**, 1248–1250 (2010).
- GTEX Consortium. The Genotype-Tissue Expression (GTEx) project. *Nat. Genet.* **45**, 580–585 (2013).
- Melen, E. *et al.* Expression analysis of asthma candidate genes during human and murine lung development. *Respir. Res.* **12**, 86 (2011).
- Waller-Evans, H. *et al.* The orphan adhesion-GPCR GPR126 is required for embryonic development in the mouse. *PLoS ONE* **5**, e14047 (2010).
- Moriguchi, T. *et al.* DREG, a developmentally regulated G protein-coupled receptor containing two conserved proteolytic cleavage sites. *Genes Cells* **9**, 549–560 (2004).
- Paavola, K. J., Sidik, H., Zuchero, J. B., Eckart, M. & Talbot, W. S. Type IV collagen is an activating ligand for the adhesion G protein-coupled receptor GPR126. *Sci. Signal.* **7**, ra76 (2014).
- Obeidat, M. *et al.* GSTCD and INTS12 regulation and expression in the human lung. *PLoS ONE* **8**, e74630 (2013).
- Dabovic, B. *et al.* Function of latent TGFbeta binding protein 4 and fibulin 5 in elastogenesis and lung development. *J. Cell. Physiol.* **230**, 226–236 (2015).
- Sterner-Kock, A. *et al.* Disruption of the gene encoding the latent transforming growth factor-beta binding protein 4 (LTBP-4) causes abnormal lung development, cardiomyopathy, and colorectal cancer. *Genes Dev.* **16**, 2264–2273 (2002).
- Cho, M. H. *et al.* A genome-wide association study of COPD identifies a susceptibility locus on chromosome 19q13. *Hum. Mol. Genet.* **21**, 947–957 (2012).
- Thorgeirsson, T. E. *et al.* Sequence variants at CHRN3-CHRNA6 and CYP2A6 affect smoking behavior. *Nat. Genet.* **42**, 448–453 (2010).
- Li, J. *et al.* The anti-proliferative function of the TGF-beta1 signalling pathway involves the repression of the oncogenic TBX2 by its homologue TBX3. *J. Biol. Chem.* **289**, 35633–35643 (2014).
- Follit, J. A. *et al.* The Golgin GMAP210/TRIP11 anchors IFT20 to the Golgi complex. *PLoS Genet.* **4**, e1000315 (2008).
- Kawaguchi, Y. *et al.* CAG expansions in a novel gene for Machado-Joseph disease at chromosome 14q32.1. *Nat. Genet.* **8**, 221–228 (1994).
- Brandsma, C. A. *et al.* A large lung gene expression study identifying fibulin-5 as a novel player in tissue repair in COPD. *Thorax* **70**, 21–32 (2015).
- Cho, M. H. *et al.* Risk loci for chronic obstructive pulmonary disease: a genome-wide association study and meta-analysis. *Lancet. Respir. Med.* **2**, 214–225 (2014).
- Groenman, F., Rutter, M., Caniggia, I., Tibboel, D. & Post, M. Hypoxia-inducible factors in the first trimester human lung. *J. Histochem. Cytochem.* **55**, 355–363 (2007).
- Ovrevik, J. *et al.* AhR and Arnt differentially regulate NF-kappaB signaling and chemokine responses in human bronchial epithelial cells. *Cell Commun. Signal.* **12**, 48 (2014).
- Petrache, I. *et al.* Ceramide synthases expression and role of ceramide synthase-2 in the lung: insight from human lung cells and mouse models. *PLoS ONE* **8**, e62968 (2013).
- Boyd, A. *et al.* Cohort Profile: the ‘children of the 90s’—the index offspring of the Avon Longitudinal Study of Parents and Children. *Int. J. Epidemiol.* **42**, 111–127 (2013).
- Welter, D. *et al.* The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Research* **42**, D1001–D1006 (2014).
- Lango Allen, H. *et al.* Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature* **467**, 832–838 (2010).
- Lette, G. *et al.* Identification of ten loci associated with height highlights new biological pathways in human growth. *Nat. Genet.* **40**, 584–591 (2008).

46. Berndt, S. I. *et al.* Genome-wide meta-analysis identifies 11 new loci for anthropometric traits and provides insights into genetic architecture. *Nat. Genet.* **45**, 501–512 (2013).
47. Albagha, O. M. *et al.* Genome-wide association identifies three new susceptibility loci for Paget's disease of bone. *Nat. Genet.* **43**, 685–689 (2011).
48. Kemp, J. P. *et al.* Phenotypic dissection of bone mineral density reveals skeletal site specificity and facilitates the identification of novel loci in the genetic regulation of bone mass attainment. *PLoS Genet.* **10**, e1004423 (2014).
49. Pei, Y. F. *et al.* Meta-analysis of genome-wide association data identifies novel susceptibility loci for obesity. *Hum. Mol. Genet.* **23**, 820–830 (2014).
50. Macgregor, S. *et al.* Genome-wide association study identifies a new melanoma susceptibility locus at 1q21.3. *Nat. Genet.* **43**, 1114–1118 (2011).
51. Porcu, E. *et al.* A meta-analysis of thyroid-related traits reveals novel loci and gender-specific differences in the regulation of thyroid function. *PLoS Genet.* **9**, e1003266 (2013).
52. Segrè, A. V. *et al.* Common inherited variation in mitochondrial genes is not enriched for associations with type 2 diabetes or related glycemic traits. *PLoS Genet.* **6**, e1001058 (2010).
53. Wood, A. R. *et al.* Defining the role of common variation in the genomic and biological architecture of adult human height. *Nat. Genet.* **46**, 1173–1186 (2014).
54. Tada, H. *et al.* Multiple associated variants increase the heritability explained for plasma lipids and coronary artery disease. *circulation. Cardiovasc. Genet.* **7**, 583–587 (2014).
55. Wain, L. V. *et al.* Novel insights into the genetics of smoking behaviour, lung function, and chronic obstructive pulmonary disease (UK BiLEVE): a genetic association study in UK Biobank. *Lancet Respir. Med.* doi: 10.1016/S2213-2600(15)00283-0 (2015).
56. Soler Artigas, M. *et al.* Effect of five genetic variants associated with lung function on the risk of chronic obstructive lung disease, and their joint effects on lung function. *Am. J. Respir. Crit. Care Med.* **184**, 786–795 (2011).
57. Castaldi, P. J. *et al.* The association of genome-wide significant spirometric loci with chronic obstructive pulmonary disease susceptibility. *Am. J. Respir. Cell Mol. Biol.* **45**, 1147–1153 (2011).
58. Pillai, S. G. *et al.* A genome-wide association study in chronic obstructive pulmonary disease (COPD): identification of two major susceptibility loci. *PLoS Genet.* **5**, e1000421 (2009).
59. Cho, M. H. *et al.* Variants in FAM13A are associated with chronic obstructive pulmonary disease. *Nat. Genet.* **42**, 200–202 (2010).
60. Wilk, J. B. *et al.* Genome-wide association studies identify CHRNA5/3 and HTR4 in the development of airflow obstruction. *Am. J. Respir. Crit. Care Med.* **186**, 622–632 (2012).
61. Li, Y. & Abecasis, G. R. Mach 1.0: rapid haplotype construction and missing genotype inference. *Am. J. Hum. Genet.* **S79**, 2290 (2006).
62. Howie, B., Fuchsberger, C., Stephens, M., Marchini, J. & Abecasis, G. R. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nat. Genet.* **44**, 955–959 (2012).
63. Howie, B. N., Donnelly, P. & Marchini, J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet.* **5**, e1000529 (2009).
64. Fuchsberger, C., Taliun, D., Pramstaller, P. P., Pattaro, C. & consortium, C. K. GWAToolbox: an R package for fast quality control and handling of genome-wide association studies meta-analysis data. *Bioinformatics* **28**, 444–445 (2012).
65. Yang, J. *et al.* Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat. Genet.* **44**, S361–S363 (2012).
66. Kumar, P., Henikoff, S. & Ng, P. C. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat. Protoc.* **4**, 1073–1081 (2009).
67. Adzhubei, I. A. *et al.* A method and server for predicting damaging missense mutations. *Nat. Methods* **7**, 248–249 (2010).
68. Kircher, M. *et al.* A general framework for estimating the relative pathogenicity of human genetic variants. *Nat. Genet.* **46**, 310–315 (2014).
69. Ritchie, G. R., Dunham, I., Zeggini, E. & Flicek, P. Functional annotation of noncoding sequence variants. *Nat. Methods* **11**, 294–296 (2014).
70. Steiling, K. *et al.* A dynamic bronchial airway gene expression signature of chronic obstructive pulmonary disease and lung function impairment. *Am. J. Respir. Crit. Care Med.* **187**, 933–942 (2013).

Acknowledgements

The research undertaken by M.D.T., M.S.A. and L.V.W. was partly funded by the National Institute for Health Research (NIHR). The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health. M.D.T. holds a Medical Research Council Senior Clinical Fellowship (G0902313). This research used the ALICE High Performance Computing Facility at the University of Leicester. The Universities of Leicester and Nottingham acknowledge receipt of a Collaborative Research and Development grant from the Healthcare and Bioscience iNet, a project funded by the East Midlands Development Agency, part-financed by the

European Regional Development Fund and delivered by Medilink East Midlands. I.P.H. holds a Medical Research Council programme grant (G1000861). We acknowledge the use of phenotype and genotype data from the British 1958 Birth Cohort DNA collection, funded by the Medical Research Council grant G0000934 and the Wellcome Trust grant 068545/Z/02 (<http://www.b58cgene.sgul.ac.uk/>). Genotyping for the B58C-WTCCC subset was funded by the Wellcome Trust grant 076113/B/04/Z. The B58C-T1DGC genotyping utilized resources provided by the Type 1 Diabetes Genetics Consortium, a collaborative clinical study sponsored by the National Institute of Diabetes and Digestive and Kidney Diseases, National Institute of Allergy and Infectious Diseases, National Human Genome Research Institute, National Institute of Child Health and Human Development and Juvenile Diabetes Research Foundation International and supported by U01 DK062418. B58C-T1DGC GWAS data were deposited by the Diabetes and Inflammation Laboratory, Cambridge Institute for Medical Research (CIMR), University of Cambridge, which is funded by Juvenile Diabetes Research Foundation International, the Wellcome Trust and the National Institute for Health Research Cambridge Biomedical Research Centre; the CIMR is in receipt of a Wellcome Trust Strategic Award (079895). The B58C-GABRIEL genotyping was supported by a contract from the European Commission Framework Programme 6 (018996) and grants from the French Ministry of Research. The Busselton Health Study (BHS) acknowledges the generous support for the 1994/5 follow-up study from Healthway, Western Australia and the numerous Busselton community volunteers who assisted with data collection and the study participants from the Shire of Busselton. The BHS is supported by The Great Wine Estates of the Margaret River region of Western Australia. GWAS genotyping was supported by a research collaboration with Pfizer. The CROATIA study was supported through grants from the Medical Research Council UK, the Ministry of Science, Education and Sport in the Republic of Croatia (number 216-1080315-0302), Croatian Science Foundation (grant number 8875) and the European Union framework program 6 EUROSPLAN project (contract no. LSHG-CT-2006-018947). SNP genotyping for CROATIA-Vis was performed by the Wellcome Trust Clinical Research Facility (WTCRF) at the Western General Hospital, Edinburgh, UK. CROATIA-Korcula was genotyped by Helmholtz Zentrum München, GmbH, Neuherberg, Germany and CROATIA-Split by AROS Applied Biotechnology, Aarhus, Denmark. We would like to acknowledge the invaluable contributions of the recruitment teams in Croatia (including those from the Institute of Anthropological Research in Zagreb and the Croatian Centre for Global Health at the University of Split), the administrative teams in Croatia and Edinburgh and the people of Korcula, Vis and Split. The EPIC Norfolk Study is funded by program grants from the Medical Research Council UK and Cancer Research UK, and by additional support from the European Union, Stroke Association, British Heart Foundation, Department of Health, Food Standards Agency and the Wellcome Trust. GS:SFHS is funded by the Scottish Executive Health Department, Chief Scientist Office, grant number CZD/16/6. Exome array genotyping for GS:SFHS was funded by the Medical Research Council UK and performed at the Wellcome Trust Clinical Research Facility Genetics Core at Western General Hospital, Edinburgh, UK. We acknowledge the invaluable contributions of the families who took part in the Generation Scotland: Scottish Family Health Study, the general practitioners and Scottish School of Primary Care for their help in recruiting them and the whole Generation Scotland team, which includes academic researchers, IT staff, laboratory technicians, statisticians and research managers. This study was financially supported by the Medical Research Fund of the Tampere University Hospital. S.R. was supported by the Academy of Finland (251217 and 255847), Center of Excellence in Complex Disease Genetics, EU FP7 projects ENGAGE (201413) and BioShaRE (261433), the Finnish Foundation for Cardiovascular Research, Biocentrum Helsinki and the Sigrid Juselius Foundation. The KORA authors acknowledge all members of field staffs who were involved in the planning and conduction of the KORA Augsburg studies, as well as all KORA study participants. The KORA research platform (KORA, Cooperative Health Research in the Region of Augsburg) was initiated and financed by the Helmholtz Zentrum München—German Research Center for Environmental Health, which is funded by the German Federal Ministry of Education and Research and by the State of Bavaria. The KORA-Age project was financed by the German Federal Ministry of Education and Research (BMBF FKZ 01ET0713 and 01ET1003A) as part of the 'Health in old age' program. Furthermore, KORA research was supported within the Munich Center of Health Sciences (MC Health), Ludwig-Maximilians-Universität, as part of LMUinnovativ. Further support was provided by the Competence Network ASCONET, subnetwork COSYCONET (FKZ 01G10882). We thank the cohort participants who contributed to this study. Genotyping was supported by the UK's Biotechnology and Biological Sciences Research Council (BBSRC) (ref. BB/F019394/1). Phenotype collection was supported by Research Into Ageing (continues as part of Age UK's The Disconnected Mind project). 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 BBSRC and Medical Research Council (MRC) is gratefully acknowledged. We thank the late Professor Paula Rantakallio (launch of NFBC1966), and Ms Outi Tornwall and Ms MinttuJussila (DNA biobanking). NFBC1966 received financial support from the Academy of Finland (project grants 104781, 120315, 129269, 1114194 and 24300796), University Hospital Oulu, Biocenter, University of Oulu, Finland (75617), NHLBI grant 5R01HL087679-02 through the STAMPEED program (1RL1MH083268-01), NIH/NIMH (5R01MH63706:02), EU FP7 (HEALTH-F4-2007-201413), EU FP8 (277849) and Medical Research Council, UK (G0500539, G1002319 and G0600705). U.G. acknowledges Swedish Medical Research Council (K2007-66X-20270-01-3, 2012-2884), Foundation for Strategic Research (SSF)

and European Commission FP6 STRP (LSHG-CT-2006-01947). Å.J. acknowledges Swedish Society for Medical Research. The ORCADES study was funded by the Chief Scientist Office of the Scottish Government, the Royal Society and the MRC Human Genetics Unit. DNA extraction was performed at the Wellcome Trust Clinical Research Facility in Edinburgh. Genotyping was funded by the European Union Framework Programme 6 EUROSPAN project. Study directorate: NM Probst-Hensch (PI; e/g); T. Rochat (p), C. Schindler (s), N. Künzli (e/exp), J.M. Gaspoz (c) Scientific team: J.C. Barthélémy (c), W. Berger (g), R. Bettschart (p), A. Bircher (a), C. Brombach (n), P.O. Bridevaux (p), L. Burdet (p), Felber Dietrich D. (e), M. Frey (p), U. Frey (pd), M.W. Gerbase (p), D. Gold (e), E. de Groot (c), W. Karrer (p), F. Kronenberg (g), B. Martin (pa), A. Mehta (e), D. Miedinger (o), M. Pons (p), F. Roche (c), T. Rothe (p), P. Schmid-Grendelmeyer (a), D. Stolz (p), A. Schmidt-Trucksäss (pa), J. Schwartz (e), A. Turk (p), A. von Eckardstein (cc) and E. Zemp Stutz (e). Scientific team at coordinating centers: M. Adam (e), I. Aguilera (exp), S. Brunner (s), D. Carballo (c), S. Caviezel (pa), I. Curjuric (e), A. Di Pascale (s), J. Dratva (e), R. Ducret (s), E. Dupuis Lozeron (s), M. Eeftens (exp), I. Eze (e), E. Fischer (g), M. Foraster (e), M. Germond (s), L. Grize (s), S. Hansen (e), A. Hensel (s), M. Imboden (g), A. Ineichen (exp), A. Jeong (g), D. Keidel (s), A. Kumar (g), N. Maire (s), A. Mehta (e), R. Meier (exp), E. Schaffner (s), T. Schikowski (e) and M. Tsai (exp); (a) allergology, (c) cardiology, (cc) clinical chemistry, (e) epidemiology, (exp) exposure, (g) genetic and molecular biology, (m) meteorology, (n) nutrition, (o) occupational health, (p) pneumology, (pa) physical activity, (pd) pediatrics and (s) statistic. The study could not have been done without the help of the study participants, technical and administrative support and the medical teams and field workers at the local study sites. Local field workers: Aarau: S. Brun, G. Giger, M. Sperisen and M. Stahel; Basel: C. Bürli, C. Dahler, N. Oertli, I. Harreh, F. Karrer, G. Novovic and N. Wyttjenbacher; Davos: A. Saner, P. Senn and R. Wenzler; Geneva: F. Bonfils, B. Blicharz, C. Landolt and J. Rochat; Lugano: S. Boccia, E. Gehrig, M.T. Mandia, G. Solari and B. Viscardi; Montana: A.P. Bieri, C. Darioly and M. Maire; Payerne: F. Ding and P. Danieli A. Vonnez; Wald: D. Bodmer, E. Hochstrasser, R. Kunz, C. Meier, J. Rakic, U. Schaefroth and A. Walder. Administrative staff: N. Bauer Ott, C. Gabriel, R. Gutknecht. Funding: The Swiss National Science Foundation (grants nos 33CS30-148470/1, 33CS30-134276/1, 33CS30-108796, 3247BO-104283, 3247BO-104288, 3247BO-104284, 3247-065896, 3100-059302, 3200-052720, 3200-042532, 4026-028099, PMPDP3_129021/1 and PMPDP3_141671/1), the Federal Office for the Environment, the Federal Office of Public Health, the Federal Office of Roads and Transport, the Canton's Government of Aargau, Basel-Stadt, Basel-Land, Geneva, Luzern, Ticino, Valais and Zürich, the Swiss Lung League, the Canton's Lung League of Basel-Stadt/ Basel Landschaft, Geneva, Ticino, Valais, Graubünden and Zurich, Stiftung ehemals Bündner Heilstätten, SUVA, Freiwillige Akademische Gesellschaft, UBS Wealth Foundation, Talecris Biotherapeutics GmbH, Abbott Diagnostics, European Commission 018996 (GABRIEL), and Wellcome Trust WT 084703MA. SHIP is part of the Community Medicine Research net of the University of Greifswald, Germany, which is funded by the Federal Ministry of Education and Research, the Ministry of Cultural Affairs as well as the Social Ministry of the Federal State of Mecklenburg-West Pomerania, and the network 'Greifswald Approach to Individualized Medicine (GANI_MED)' funded by the Federal Ministry of Education and Research, and the German Asthma and COPD Network (COSYCONET) (grant nos 01ZZ9603, 01ZZ0103, 01ZZ0403, 03IS2061A and BMBF 01G10883). Genome-wide data have been supported by the Federal Ministry of Education and Research and a joint grant from Siemens Healthcare, Erlangen, Germany and the Federal State of Mecklenburg-West Pomerania (grant no. 03ZIK012). The University of Greifswald is a member of the 'Center of Knowledge Interchange' program of the Siemens AG and the Caché Campus program of the InterSystems GmbH. We acknowledge the Academy of Finland (126925, 121584 and 124282), Academy of Finland (Eye) (134309), Academy of Finland (Salve) (129378), Academy of Finland (Gendi) (117787), Academy of Finland (Skidi) (41071), Social Insurance Institution of Finland, Tampere University Hospital Medical Funds (X51001 for T.L.), Kuopio University Hospital Medical Funds, Turku University Hospital Medical Funds, Juho Vainio Foundation, Paavo Nurmi Foundation, Finnish Foundation of Cardiovascular Research (T.L.), Finnish Cultural Foundation, Tuberculosis Foundation (T.L.), Emil Aaltonen Foundation (T.L.) and Yrjö Jahnsson Foundation (T.L.). We are extremely grateful to all the families who took part in the 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: 092731) and the University of Bristol provide core support for ALSPAC. This publication is the work of the authors and D.M.E. will serve as guarantor for the contents of this paper. This work was funded by a Medical Research Council (MRC) strategic award to M.D.T., I.P.H., D.P.S. and L.V.W. (MC_PC_12010). This research has been conducted using the UK Biobank Resource. M.D.T. has been supported by MRC fellowships G0501942 and G0902313. I.P.H. is supported by an MRC programme grant (G1000861). J.M. is funded by an ERC Consolidator Grant (617306). This article presents independent research funded partially by the NIHR. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health. We would like to acknowledge all members of the UK Biobank Array Design Group. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health. Peter Donnelly (chair) (University of Oxford), Jeff Barrett (Wellcome Trust Sanger Institute), Jose Bras

(University College London), Adam Butterworth (University of Cambridge), Richard Durbin (Wellcome Trust Sanger Institute), Paul Elliott (Imperial College London), Ian Hall (University of Nottingham), John Hardy (University College London), Mark McCarthy (University of Oxford), Gil McVean (University of Oxford), Tim Peakman (UK Biobank), Nazneen Rahman (The Institute of Cancer Research), Nilesh Samani (University of Leicester), Martin Tobin (University of Leicester), Hugh Watkins (University of Oxford). We acknowledge EU funding (GABRIEL Grant Number: 018996, ECRHS II Coordination Number: QLK4-CT-1999-01237). A.P.M. acknowledges the Wellcome Trust (WT098017, WT064890 and WT090532). The PIVUS study acknowledges The Swedish Foundation for Strategic Research (ICA08-0047), The Swedish Research Council (2012-1397), The Swedish Heart-Lung Foundation (20120197), The Swedish Society of Medicine and Uppsala University. The computations were performed on resources provided by SNIC through Uppsala Multidisciplinary Center for Advanced Computational Science (UPPMAX) under Project p2013056. A.P.M. is a Wellcome Trust Senior Research Fellow in Basic Biomedical Science (grant number WT098017). This cohort received funding from the Wellcome Trust; the European Community's Seventh Framework Programme (FP7/2007-13); US National Institutes of Health/National Eye Institute (1RO1EY018246); NIH Center for Inherited Disease Research; the NIHR-funded BioResource, Clinical Research Facility and Biomedical Research Centre based at Guy's and St Thomas' NHS Foundation Trust in partnership with King's College London. We thank the research staff at the Respiratory Health Network Tissue Bank of the FRQS for collecting lung specimens for the lung eQTL study at the Laval University. The Lung Tissue eQTL study was funded by Merck Research Laboratories. M.O. is a Post-doctoral Fellow of the Michael Smith Foundation for Health Research and the Canadian Institute for Health Research Integrated and Mentored Pulmonary and Cardiovascular Training program (IMPACT). Y.B. is the recipient of a Junior 2 Research Scholar award from the Fonds de recherche Québec—Santé (FRQS).

Author contributions

I.P.H., S.M., N.S., M.S.A., D.P.S., M.D.T. and L.V.W. contributed to analysis. I.P.H., A.K.K., E.M., S.M., I.N., I.S., M.S.A., M.D.T. and L.V.W. carried out bioinformatics and functional assessment. I.P.H., S.M., M.S.A., D.P.S., M.D.T. and L.V.W. wrote the manuscript. Project conception, design and management: Stage 1—B58C: D.P.S.; CROATIA-Korcula: C.H.; CROATIA-Split: I.K., O.P., V.V. and T.Z.; CROATIA-Vis: I.R. and A.F.W.; EPIC: J.H.Z., R.A.S. and N.J.W.; GS:SFHS: L.H., S.P. and G.S.; H2000: M. Heliövaara and M.K.; KORA F4: J. Heinrich; KORA S3: C.G., S.K. and H.S.; LBC1936: I.J.D. and J.M.S.; NFBC1966: M.-R.J.; NSPHS: U.G.; ORCADES: H.C., S.H.W. and J.F.W.; SAPALDIA: N.M.P.-H.; SHIP: S.G., B.K. and H.V.; YFS: T.L. and O.T.R. Stage 2—ECRHS: J. Heinrich and D.L.J.; PIVUS: L.L.; TwinsUK: T.D.S. Phenotype collection and data management: Stage 1—B58C: W.L.M. and D.P.S.; BHS1&2: J.B., J. Hui, A.L.J. and A.W.M.; CROATIA-Korcula: C.H., J.E.H. and P.N.; CROATIA-Split: I.K., O.P., V.V. and T.Z.; CROATIA-Vis: J.M.; EPIC: J.H.Z.; GS:SFHS: L.H., S.P., G.S. and H.T.; H2000: M. Heliövaara, M.K., S.R. and I. Surakka; KORA F4: H.G. and J.S.R.; KORA S3: C.G., S.K., R.R. and H.S.; LBC1936: I.J.D., S.E.H., L.M.L. and J.M.S.; NFBC1966: A.C.A., A.-L.H. and M.-R.J.; NSPHS: S.E. and A.J.; ORCADES: H.C., S.H.W. and J.F.W.; SAPALDIA: M.I., A.K. and N.M.P.-H.; SHIP: S.G., B.S., A.T. and H.V.; YFS: N.H.-K., T.L. and O.T.R. Stage 2—ALSPAC: R.G., J. Henderson and J.P.K.; ECRHS: J.R.G., J. Heinrich and D.L.J.; PIVUS: E.I., L.L., A.M. and A.P.M.; TwinsUK: C.J.H., P.G.H., T.D.S. and A.V.; Lung eQTL study: Y.B., C.-A.B., D.C.N. and D.D.S. Data analysis: Stage 1—B58C: D.P.S.; CROATIA-Korcula: J.E.H. and P.N.; CROATIA-Vis: J.M.; EPIC: J.H.Z.; H2000: M.K., S.R. and I. Surakka; KORA F4: E.A. and J.S.R.; KORA S3: R.R.; LBC1936: L.M.L.; NFBC1966: A.C.A., M. Horikoshi and M.-R.J.; NSPHS: S.E.; ORCADES: P.K.J.; SAPALDIA: M.I., A.K. and N.M.P.-H.; SHIP: A.T.; YFS: L.-P.L. Stage 2—ALSPAC: D.M.E.; ECRHS: C.F.; PIVUS: A.M. and A.P.M.; TwinsUK: P.G.H.; Lung eQTL study: M.O.

Additional information

Supplementary information accompanies this paper at <http://www.nature.com/naturecommunications>

Competing financial interests: The authors declare no competing financial interests.

Reprints and permission information is available online at <http://npg.nature.com/reprintsandpermissions/>

How to cite this article: Artigas, M. S. *et al.* Sixteen new lung function signals identified through 1000 Genomes Project reference panel imputation. *Nat. Commun.* 6:8658 doi: 10.1038/ncomms9658 (2015).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

María Soler Artigas^{1,*}, Louise V. Wain^{1,*}, Suzanne Miller^{2,*}, Abdul Kader Kheirallah², Jennifer E. Huffman³, Ioanna Ntalla¹, Nick Shrine¹, Ma'en Obeidat⁴, Holly Trochet^{3,5}, Wendy L. McArdle⁶, Alexessander Couto Alves⁷, Jennie Hui^{8,9,10,11}, Jing Hua Zhao¹², Peter K. Joshi¹³, Alexander Teumer^{14,15}, Eva Albrecht¹⁶, Medea Imboden^{17,18}, Rajesh Rawal^{16,19,20}, Lorna M. Lopez^{21,22}, Jonathan Marten³, Stefan Enroth²³, Ida Surakka^{24,25}, Ozren Polasek^{13,26}, Leo-Pekka Lyytikäinen^{27,28}, Raquel Granell⁶, Pirro G. Hysi²⁹, Claudia Flexeder³⁰, Anubha Mahajan³¹, John Beilby^{8,9,32}, Yohan Bossé³³, Corry-Anke Brandsma³⁴, Harry Campbell¹³, Christian Gieger^{16,19,20}, Sven Gläser³⁵, Juan R. González^{36,37,38}, Harald Grallert¹⁹, Chris J. Hammond²⁹, Sarah E. Harris^{21,39}, Anna-Liisa Hartikainen⁴⁰, Markku Heliövaara²⁵, John Henderson⁶, Lynne Hocking^{5,41}, Momoko Horikoshi^{31,42}, Nina Hutri-Kähönen^{43,44}, Erik Ingelsson^{31,45,46}, Åsa Johansson^{23,47}, John P. Kemp^{6,48,49}, Ivana Kolcic²⁶, Ashish Kumar^{17,18,31,50}, Lars Lind⁵¹, Erik Melén⁵², Arthur W. Musk^{8,53,54}, Pau Navarro³, David C. Nickle⁵⁵, Sandosh Padmanabhan^{5,56}, Olli T. Raitakari^{57,58}, Janina S. Ried¹⁶, Samuli Ripatti^{24,59,60}, Holger Schulz^{30,61}, Robert A. Scott¹², Don D. Sin^{4,62}, John M. Starr^{21,63}, UK BiLEVE†, Ana Viñuela²⁹, Henry Völzke¹⁴, Sarah H. Wild¹³, Alan F. Wright³, Tatijana Zemunik⁶⁴, Deborah L. Jarvis^{65,66}, Tim D. Spector²⁹, David M. Evans^{6,48,49}, Terho Lehtimäki^{27,28}, Veronique Vitart³, Mika Kähönen⁶⁷, Ulf Gyllensten²³, Igor Rudan^{13,68}, Ian J. Deary^{21,22}, Stefan Karrasch^{30,69,70}, Nicole M. Probst-Hensch^{17,18}, Joachim Heinrich^{30,61,71}, Beate Stubbe³⁵, James F. Wilson^{3,13}, Nicholas J. Wareham¹², Alan L. James^{8,72,54}, Andrew P. Morris^{31,73,74}, Marjo-Riitta Jarvelin^{7,75,76,77}, Caroline Hayward⁵, Ian Sayers², David P. Strachan^{78,*}, Ian P. Hall^{2,*} & Martin D. Tobin^{1,79,*}

¹Genetic Epidemiology Group, Department of Health Sciences, University of Leicester, Leicester LE1 7RH, UK. ²Division of Respiratory Medicine, Queen's Medical Centre, University of Nottingham, Nottingham NG7 2RD, UK. ³MRC Human Genetics Unit, MRC Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, Scotland EH8 9AD, UK. ⁴University of British Columbia Centre for Heart Lung Innovation, St Paul's Hospital, Vancouver, British Columbia, Canada V6Z 1Y6. ⁵Generation Scotland, A Collaboration between the University Medical Schools and NHS, Aberdeen, Dundee, Edinburgh, Glasgow EH4 2XU, UK. ⁶School of Social and Community Medicine, University of Bristol, Bristol BS8 1TH, UK. ⁷Department of Epidemiology and Biostatistics, MRC -PHE Centre for Environment & Health, School of Public Health, Imperial College London, London SW7 2AZ, UK. ⁸Busselton Population Medical Research Institute, Busselton, Western Australia 6280, Australia. ⁹PathWest Laboratory Medicine WA, Sir Charles Gairdner Hospital, Western Australia 6009, Australia. ¹⁰School of Population Health, The University of Western Australia, Western Australia 6009, Australia. ¹¹School of Pathology and Laboratory Medicine, The University of Western Australia, Western Australia 6009, Australia. ¹²MRC Epidemiology Unit, University of Cambridge School of Clinical Medicine, Institute of Metabolic Science, Cambridge Biomedical Campus, Cambridge CB2 0SL, UK. ¹³Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, Teviot Place, Edinburgh EH8 9AD, Scotland, UK. ¹⁴University Medicine Greifswald, Community Medicine, SHIP—Clinical Epidemiological Research, Greifswald 17489, Germany. ¹⁵Department for Genetics and Functional Genomics, Interfaculty Institute for Genetics and Functional Genomics, University Medicine Greifswald, Greifswald 17489, Germany. ¹⁶Institute of Genetic Epidemiology, Helmholtz Zentrum München German Research Center for Environmental Health, Neuherberg D-85764, Germany. ¹⁷Swiss Tropical and Public Health Institute, Basel 4051, Switzerland. ¹⁸University of Basel, Basel 4001, Switzerland. ¹⁹Research Unit of Molecular Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg D-85764, Germany. ²⁰Institute of Epidemiology II, Helmholtz Zentrum München German Research Center for Environmental Health, Neuherberg D-85764, Germany. ²¹Centre for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh, Edinburgh EH8 9AD, UK. ²²Department of Psychology, University of Edinburgh, Edinburgh EH8 9AD, UK. ²³Department of Immunology, Genetics, and Pathology, Biomedical Center, SciLifeLab Uppsala, Uppsala University, Uppsala 751 23, Sweden. ²⁴Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki FI-00014, Finland. ²⁵The National Institute for Health and Welfare (THL), Helsinki FI-00271, Finland. ²⁶Department of Public Health, Faculty of Medicine, University of Split, Split 21000, Croatia. ²⁷Department of Clinical Chemistry, Fimlab Laboratories, Tampere FI-33101, Finland. ²⁸Department of Clinical Chemistry, University of Tampere School of Medicine, Tampere FI-33520, Finland. ²⁹KCL Department of Twins Research and Genetic Epidemiology, King's College London, London WC2R 2LS, UK. ³⁰Institute of Epidemiology I, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg D-85764, Germany. ³¹Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford OX3 7BN, UK. ³²School of Pathology and Laboratory Medicine, The University of Western Australia, Western Australia 6009, Australia. ³³Department of Molecular Medicine, Institut Universitaire de Cardiologie et de Pneumologie de Québec, Laval University, Québec, Canada G1V 0A6. ³⁴Department of Pathology and Medical Biology, University of Groningen, University Medical Center Groningen, Groningen 9700, The Netherlands. ³⁵Department of Internal Medicine B, Pneumology, Cardiology, Intensive Care, Weaning, Field of Research: Pneumological Epidemiology, University Medicine Greifswald, Greifswald 17489, Germany. ³⁶Centre for Research in Environmental Epidemiology (CREAL), Barcelona E-08003, Spain. ³⁷CIBER Epidemiología y Salud Pública (CIBERESP), Madrid 28029, Spain. ³⁸Pompeu Fabra University (UPF), Barcelona 08002, Catalonia, Spain. ³⁹Centre for Genomic and Experimental Medicine, University of Edinburgh, Edinburgh EH8 9AD, UK. ⁴⁰Department of Obstetrics and Gynecology of Oulu University Hospital, MRC of Oulu University, Oulu 90220, Finland. ⁴¹Division of Applied Health Sciences, University of Aberdeen, Aberdeen, Scotland AB24 3FX, UK. ⁴²Oxford Centre for Diabetes, Endocrinology and Metabolism, University of Oxford, Oxford OX1 2JD, UK. ⁴³Department of Pediatrics, Tampere University Hospital, Tampere 33521, Finland. ⁴⁴Department of Pediatrics, University of Tampere School of Medicine, Tampere FI-33520, Finland. ⁴⁵Department of Medical Sciences, Molecular Epidemiology and Science for Life Laboratory, Uppsala University, Uppsala 751 23, Sweden. ⁴⁶Department of Medicine, Division of Cardiovascular Medicine, Stanford University School of Medicine, Stanford, California 94305, USA. ⁴⁷Uppsala Clinical Research Centre, Uppsala University, Uppsala 751 23, Sweden. ⁴⁸Diamantina Institute, Translational Research Institute, University of Queensland, Brisbane, Queensland QLD 4072, Australia. ⁴⁹MRC Integrative Epidemiology Unit, Bristol BS8 1TH, UK. ⁵⁰Institute of Environmental Medicine, Karolinska Institutet, Stockholm SE-171 7, Sweden. ⁵¹Department of Medical Sciences, Uppsala University,

Uppsala 751 23, Sweden. ⁵²Institute of Environmental Medicine, Karolinska Institutet and Sachs' Children's Hospital, Stockholm SE-171 7, Sweden. ⁵³Department of Respiratory Medicine, Sir Charles Gairdner Hospital, Western Australia 6009, Australia. ⁵⁴School of Medicine and Pharmacology, The University of Western Australia, Western Australia 6009, Australia. ⁵⁵Genetics and Pharmacogenomics, Merck Research Labs, Boston, Massachusetts 02115, USA. ⁵⁶Division of Cardiovascular and Medical Sciences, University of Glasgow, Glasgow G12 8TA, Scotland, UK. ⁵⁷Department of Clinical Physiology and Nuclear Medicine, Turku University Hospital, Turku 20520, Finland. ⁵⁸Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku, Turku 20014, Finland. ⁵⁹Department of Public Health, University of Helsinki, Helsinki FI-00014, Finland. ⁶⁰Department of Human Genomics, Wellcome Trust Sanger Institute, Hinxton, Cambridge CB10 1SA, UK. ⁶¹Comprehensive Pneumology Center Munich (CPC-M), Member of the German Center for Lung Research, Munich 85764, Germany. ⁶²Respiratory Division, University of British Columbia, Vancouver, British Columbia, Canada V6T 1Z4. ⁶³Alzheimer Scotland Research Centre, University of Edinburgh, Edinburgh EH8 9AD, UK. ⁶⁴Department of Medical Biology, Faculty of Medicine, University of Split, Split 21000, Croatia. ⁶⁵Respiratory Epidemiology and Public Health, Imperial College London, London SW7 2AZ, UK. ⁶⁶MRC Health Protection Agency (HPA) Centre for Environment and Health, Imperial College London, London SW7 2AZ, UK. ⁶⁷Department of Clinical Physiology, University of Tampere and Tampere University Hospital, Tampere 33521, Finland. ⁶⁸Centre for Population Health Sciences, Medical School, University of Edinburgh, Edinburgh EH8 9AD, Scotland, UK. ⁶⁹Institute of General Practice, University Hospital Klinikum rechts der Isar, Technische Universität München, Munich D - 81675, Germany. ⁷⁰Institute and Outpatient Clinic for Occupational, Social and Environmental Medicine, Ludwig-Maximilians-Universität, Munich 80539, Germany. ⁷¹University Hospital Munich, Institute and Outpatient Clinic for Occupational, Social and Environmental Medicine, Ludwig-Maximilian University Munich, Munich 80539, Germany. ⁷²Department of Pulmonary Physiology and Sleep Medicine, Sir Charles Gairdner Hospital, Western Australia 6009, Australia. ⁷³Department of Biostatistics, University of Liverpool, Liverpool L69 7ZX, UK. ⁷⁴Estonian Genome Centre, University of Tartu, Tartu 50090, Estonia. ⁷⁵Center for Life Course Epidemiology, Faculty of Medicine, P.O.Box 5000, FI-90014 University of Oulu, Oulu FI-01051, Finland. ⁷⁶Biocenter Oulu, P.O.Box 5000, Aapistie 5A, FI-90014 University of Oulu, Oulu FI-01051, Finland. ⁷⁷Unit of Primary Care, Oulu University Hospital, Kajaanintie 50, P.O.Box 20, FI-90220 Oulu, 90029 OYS, Finland. ⁷⁸Population Health Research Institute, St George's, University of London, Cranmer Terrace, London WC1B 5DN, UK. ⁷⁹National Institute for Health Research (NIHR) Leicester Respiratory Biomedical Research Unit, Glenfield Hospital, Leicester LE3 9QP, UK. * These authors contributed equally to this work. † A full list of consortium members appears at the end of the paper.

UK BiLEVE

Panos Deloukas⁸⁰, Anna L. Hansell⁸¹, Richard Hubbard⁸², Victoria E Jackson⁸³, Jonathan Marchini⁸⁴, Ian Pavord⁸⁵, Neil C. Thomson⁸⁶, and Eleftheria Zeggini⁸⁷

⁸⁰William Harvey Research Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University London, London E1 4NS, UK. ⁸¹UK Small Area Health Statistics Unit, MRC-PHE Centre for Environment and Health, School of Public Health, Imperial College London, London SW7 2AZ, UK. ⁸²School of Medicine, University of Nottingham, Nottingham NG7 2RD, UK. ⁸³Department of Health Sciences, University of Leicester, Leicester LE1 7RH, UK. ⁸⁴Department of Statistics, University of Oxford, Oxford OX1 2JD, UK. ⁸⁵Respiratory Medicine, University of Oxford, Oxford OX1 2JD, UK. ⁸⁶Institute of Infection, Immunity & Inflammation, University of Glasgow, Glasgow G12 8TA, UK. ⁸⁷Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire CB10 1SA, UK.