Genome-wide association analysis identifies six new loci associated with forced vital capacity

Forced vital capacity (FVC), a spirometric measure of pulmonary function, reflects lung volume and is used to diagnose and monitor lung diseases. We performed genome-wide association study meta-analysis of FVC in 52,253 individuals from 26 studies and followed up the top associations in 32,917 additional individuals of European ancestry. We found six new regions associated at genome-wide significance ($P < 5 \times 10^{-8}$) with FVC in or near *EFEMP1*, *BMP6*, *MIR129-2–HSD17B12*, *PRDM11*, *WWOX* and *KCNJ2*. Two loci previously associated with spirometric measures (*GSTCD* and *PTCH1*) were related to FVC. Newly implicated regions were followed up in samples from African-American, Korean, Chinese and Hispanic individuals. We detected transcripts for all six newly implicated genes in human lung tissue. The new loci may inform mechanisms involved in lung development and the pathogenesis of restrictive lung disease.

Pulmonary function is a heritable trait that can be reliably measured by spirometry and reflects the physiological state of the lungs and airways[1]. FVC, one of the most widely used measures of pulmonary function, approximates vital capacity. In conjunction with forced expiratory volume in 1 s (FEV_1), FVC is used to diagnose various respiratory diseases. A reduced ratio of FEV_1 to FVC (FEV_1/FVC) indicates airflow obstruction when FEV_1 is reduced disproportionately relative to FVC. In contrast, a decreased FVC measure in the context of a normal to elevated FEV_1/FVC ratio suggests a restrictive ventilatory defect. In clinical practice, FVC is often used as a surrogate measure of disease progression in patients with established restrictive lung disorders, such as idiopathic pulmonary fibrosis[2,3]. Reduced FVC is a strong predictor of mortality in the general population, independent of FEV_1 and standard risk factors such as age and cigarette smoking[4–8].

Measures of pulmonary function show familial aggregation, with evidence for genetic effects in twin and family studies[9,10]. We previously reported associations between FEV_1 or FEV_1/FVC and at least 27 genetic loci using large-scale meta-analyses of genome-wide association studies (GWAS)[11–14]. Thus far, the genetic determinants of FVC have not been studied using GWAS methods. We conducted a comprehensive GWAS meta-analysis across 2 large consortia with cohorts of European ancestry—the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) and SpiroMeta—to identify common genetic variants associated with cross-sectional measures of FVC in 52,253 subjects of European descent. In each of 6 newly associated loci, we confirmed that the genes closest to the new variants were expressed in lung tissue and performed expression quantitative trait locus (eQTL) analyses in 762 whole-blood samples. We evaluated the new loci in 6,070 African Americans in the National Heart, Lung, and Blood Institute (NHLBI)-sponsored Candidate Gene Association Resource (CARe) Project, in a Chinese subset ($n = 563$) of the Multi-Ethnic Study of Atherosclerosis (MESA) Lung study, in a Hispanic subset ($n = 849$) of MESA Lung

and in Koreans ($n = 8,074$) from 2 cohort studies; Healthy Twin[15–17] and Korea Association Resource 3 (KARE3)[18].

RESULTS

The study consisted of two stages. Stage 1 was a meta-analysis of study-specific genome-wide analyses of FVC conducted in 26 studies with a total of 52,253 individuals of European ancestry. Study characteristics are shown in **Supplementary Table 1**. Individual cohorts performed GWAS analysis using linear regression models with FVC (in ml) as the outcome, stratified by never- or ever-smoking status. Adjustment factors included age, age^2, sex, height, $height^2$ and weight. If applicable, cohorts adjusted for center, cohort or principal components to adjust for population stratification. Stage 1 results are shown in **Supplementary Data Set 1**. In stage 2, we followed up SNPs showing association with FVC ($P < 5 \times 10^{-7}$) and performed meta-analysis of β regression coefficients (effect estimates) and standard errors across stages 1 and 2 (**Fig. 1**). Stage 2 encompassed 32,917 subjects of European ancestry from 9 independent cohorts. Study characteristics are shown in **Supplementary Table 2**. The follow-up studies used the same models as in stage 1. Test statistics for each study were corrected using genomic control[19] separately within smoking strata and after meta-analysis of ever- and never-smokers. The test statistic inflation factor (λ_{GC}) before applying genomic control at the meta-analysis level was 1.12 (**Supplementary Fig. 1**). The test statistic inflation factor standardized for a sample size of 1,000 individuals was 1.002. Study-specific λ_{GC} estimates are shown in **Supplementary Table 1**.

Regions encompassing SNPs with associations reaching genome-wide significance after meta-analysis of the two stages were followed up in diverse-ancestry samples from African Americans, Koreans, Chinese and Hispanics. Characteristics for these multi-ancestry follow-up studies are shown in **Supplementary Table 2**. After the meta-analysis, we investigated the mRNA expression of the nearest gene for each of the new SNPs in human lung tissue, human airway smooth muscle (HASM) cells, human bronchial epithelial cells (HEBCs) and peripheral blood mononuclear

---

Figure 1 Overview of our staged analysis to identify new variants influencing FVC. After a large-scale meta-analysis of GWAS data for stage 1 cohorts ($n = 52,253$ subjects), we followed up a total of 7 SNPs showing evidence of association with FVC ($P < 5 \times 10^{-7}$) in stage 2. The studies included in stage 2, encompassing a total sample size of $n = 32,917$ subjects, undertook *in silico* testing of the 7 loci, which were not previously associated with any pulmonary phenotype. See **Supplementary Tables 1** and **2** for definitions of all study abbreviations.

Cohorts included in stage 1
$N = 52,253$

- AGES ($n = 1,620$)
- ARIC ($n = 9,078$)
- B58C T1DGC ($n = 2,343$)
- B58C WTCCC ($n = 1,372$)
- CARDIA ($n = 1,626$)
- CHS ($n = 3,140$)
- CROATIA-Korcula ($n = 825$)
- CROATIA-VIS ($n = 769$)
- ECRHS ($n = 1,594$)
- EPIC obese cases ($n = 1,104$)
- EPIC population based ($n = 2,336$)
- FHS ($n = 7,692$)
- FTC ($n = 134$)
- Health 2000 ($n = 821$)
- Health ABC ($n = 1,475$)
- HCS ($n = 1,820$)
- KORA S3 ($n = 555$)
- MESA ($n = 1,433$)
- NFBC 1966 ($n = 4,556$)
- NSPHS ($n = 549$)
- ORCADES ($n = 692$)
- RS-I ($n = 1,152$)
- RS-II ($n = 862$)
- RS-III ($n = 1,043$)
- SHIP ($n = 1,777$)
- Twins UK-I ($n = 1,885$)

Cohorts included in stage 2
$N = 32,917$

- BHS 1 & 2 ($n = 4,094$)
- CROATIA-Split ($n = 493$)
- Generation Scotland ($n = 8,077$)
- KORA F4 ($n = 903$)
- LBC1936 ($n = 991$)
- LifeLines ($n = 12,399$)
- LLFS ($n = 3,899$)
- PIVUS ($n = 900$)
- Twins UK-II and III ($n = 1,161$)

SNPs selected for stage 2 follow-up

- rs1430193
- rs1942055
- rs6923462
- rs4237643
- rs2863171
- rs1079572
- rs6501431

Combined sample size, stages 1 and 2
$N = 85,170$

cells (PBMCs) (**Supplementary Note**). We assessed whether these SNPs were associated with gene expression in whole-blood cells (**Supplementary Table 3** and **Supplementary Note**) and queried databases to determine whether these SNPs were located within known or predicted regulatory regions.

Stage 1 and 2 results

There were nine regions containing at least one SNP associated with FVC at $P < 5 \times 10^{-7}$ in stage 1. Of these regions, two (*GSTCD* and *PTCH1*) had previously been reported in GWAS[12,13] of spirometric traits (FEV$_1$ and FEV$_1$/FVC) and were thus not evaluated further, leaving seven SNPs in seven loci for follow-up in stage 2.

Six loci had associations that reached genome-wide significance ($P < 5 \times 10^{-8}$) for FVC in the meta-analysis of stages 1 and 2 (**Fig. 2** and **Table 1**). These loci were in or near the following genes: *BMP6* (rs6923462, 6p24, intronic), *EFEMP1* (rs1430193, 2p16.1, intronic), *MIR129-2–HSD17B12* (rs4237643, 11p11.2, 54 kb upstream), *PRDM11* (rs2863171, 11p11.2, 3 kb downstream), *WWOX* (rs1079572, 16q23.1, intronic) and *KCNJ2* (rs6501431, 17q24.3, 800 kb downstream) (**Supplementary Fig. 2a–g**). Effect sizes were generally consistent across studies (**Supplementary Figs. 3a–g** and **4a–g**). Three of these regions (*BMP6*, *EFEMP1* and *PRDM11*) also showed independent replication in stage 2 samples of European ancestry, with association P values below a Bonferroni-corrected threshold for seven tests ($P < 7.14 \times 10^{-3}$) (**Table 1**). The lowest P value (5.89×10^{-13}) for the meta-analysis effect estimate across stages 1 and 2 was found for SNP rs6923462 (intronic SNP in *BMP6*).

We evaluated the effect of the six new loci separately in ever-smokers and in never-smokers, and effect sizes were consistent across smoking strata for all the variants (**Supplementary Table 4**).
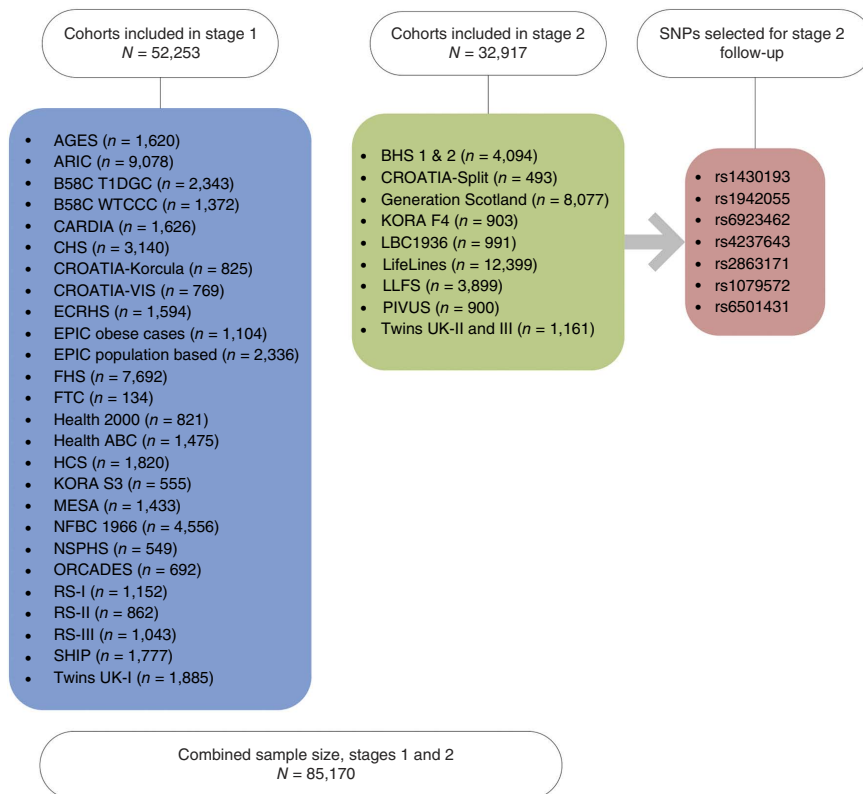
Multi-ancestry follow-up

To examine the portability of the identified loci to other ancestry groups, we evaluated the regions of our newly identified SNPs in African Americans, Hispanics, Chinese and Koreans. For three of these samples (African Americans, Hispanics and Chinese), we looked up the SNPs with minor allele frequency (MAF) of ≥0.05 within 200 kb in either direction of the sentinel SNP in individuals of European ancestry in the 1000 Genomes Project all-ancestries imputation panel[20]. For Koreans, we looked up SNPs with MAF of ≥0.05 within 200 kb of the sentinel SNP according to imputation to HapMap and the Korean panel. To determine the appropriate Bonferroni-corrected P-value threshold for declaring statistical significance in each ancestry group, we used the Nyholt method to calculate the effective number of independent variants, on the basis of pairwise linkage disequilibrium (LD) among the follow-up SNPs[21,22].

African Americans were participants of the Candidate CARe Project[23]. Baseline characteristics of the 6,070 African Americans in CARe are shown by cohort in **Supplementary Table 2b**. We performed regional meta-analyses of the 7,470 SNPs (MAF ≥ 0.05) within 200 kb of the sentinel SNPs in individuals of European ancestry using 1000
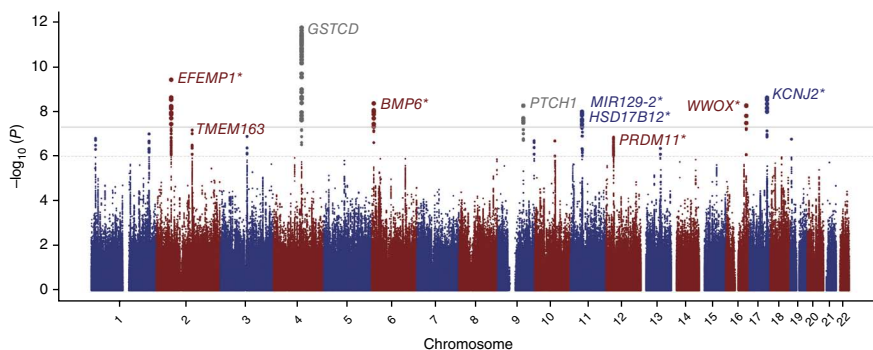

Figure 2 Manhattan plot for the association results for FVC. The plot shows all the loci analyzed in stage 1, where the two loci previously associated with either FEV$_1$ or FEV$_1$/FVC are indicated in gray. Newly associated loci were in or near the indicated adjacent genes. The loci reaching genome-wide significance after the combined analysis of stages 1 and 2 are marked with an asterisk. The solid line shows the threshold $P < 5 \times 10^{-8}$. The dashed line shows the level $P < 1 \times 10^{-6}$.

Table 1 Main results from stage 1, stage 2 and the meta-analysis of stages 1 and 2 for loci associated with FVC

SNP ID	Chr.	NCBI Build 36 position	Nearest gene	Coded allele	Analysis stage	β (ml)	SE	P value	Coded allele freq.	N effective
rs1430193	2	55,974,357	*EFEMP1* (intronic)	T	Stage 1	−23.75	4.022	3.52×10^{-9}	0.370	45,852
					Stage 2	−17.839	4.500	**7.36×10^{-5}**	0.361	28,103
					Joint meta-analysis	−21.125	2.999	**1.86×10^{-12}**		
rs1942055	2	135,215,394	*TMEM163* (50 kb upstream)	G	Stage 1	−19.943	3.919	3.60×10^{-7}	0.458	46,365
					Stage 2	−3.13	4.447	0.482	0.470	24,984
					Joint meta-analysis	−12.594	2.940	1.84×10^{-5}		
rs6923462	6	7,746,111	*BMP6* (intronic)	T	Stage 1	28.828	5.208	3.11×10^{-8}	0.843	48,680
					Stage 2	35.204	7.552	**3.14×10^{-6}**	0.846	17,271
					Joint meta-analysis	30.883	4.288	**5.89×10^{-13}**		
rs4237643	11	43,604,944	*HSD17B12* (54 kb upstream)	T	Stage 1	−21.366	3.957	6.66×10^{-8}	0.311	51,977
					Stage 2	−10.073	4.686	0.032	0.305	30,119
					Joint meta-analysis	−16.666	3.023	**3.53×10^{-8}**		
rs2863171	11	45,207,308	*PRDM11* (3 kb downstream)	C	Stage 1	25.343	5.015	4.33×10^{-7}	0.158	51,758
					Stage 2	21.755	6.23	**4.79×10^{-3}**	0.160	25,121
					Joint meta-analysis	23.924	3.906	**8.97×10^{-10}**		
rs1079572	16	76,744,639	*WWOX* (intronic)	G	Stage 1	20.539	3.733	3.76×10^{-8}	0.417	51,049
					Stage 2	10.41	4.364	0.017	0.419	28,103
					Joint meta-analysis	16.258	2.837	**9.95×10^{-9}**		
rs6501431	17	66,488,010	*KCNJ2* (800 kb downstream)	T	Stage 1	26.729	4.751	1.84×10^{-8}	0.798	47,576
					Stage 2	15.641	6.746	0.020	0.789	17,694
					Joint meta-analysis	23.053	3.884	**2.94×10^{-9}**		

Shown are FVC results for the leading SNPs, ordered by chromosome and position, for each independent locus associated ($P < 5 \times 10^{-8}$) with FVC in a joint analysis of up to 85,170 individuals of European ancestry from the CHARGE-SpiroMeta GWAS (stage 1) and follow-up (stage 2). Two-sided *P* values are given for stage 1, stage 2 and the joint meta-analysis of both stages. *P* values corresponding to genome-wide significance ($P < 5 \times 10^{-8}$) in the joint meta-analysis of all stages are indicated in bold. SNPs showing independent replication in stage 2 ($P = 0.05/7 = 7.14 \times 10^{-3}$) are indicated with their stage 2 *P* value shown in bold. The sample sizes (*N*) shown are the effective sample sizes. The effective sample size is the product of sample size and imputation quality summed across studies. β values reflect effect size estimates with standard error (SE) in ml. Chr., chromosome.

Genomes Project imputed data. Using the *P*-value threshold of 4.42×10^{-5} (based on 1,132 independent tests), 78 SNPs in the region of *EFEMP1* were significantly associated with FVC (lowest $P = 1.63 \times 10^{-7}$) in African Americans. Our top hit at the *EFEMP1* locus in samples from individuals of European ancestry (SNP rs1430193, T allele frequency = 0.37; **Table 1**) had a very different allele frequency in our replication African-American samples (T allele frequency = 0.71) and did not show statistically significant association ($P = 0.13$). As has recently been noted regarding extension of GWAS SNPs discovered in European populations to African-American populations[24], the effect size was in the same direction but was attenuated in the African-American cohort ($\beta = -15.79$ ml in African Americans versus −23.75 ml in individuals of European ancestry). The top hit in African Americans for the *EFEMP1* region (rs62164511, $P = 1.63 \times 10^{-7}$) showed a decrease in FVC of 84.7 ml for each copy of the A allele (A allele frequency = 0.9) (**Supplementary Table 5a**). LD (r^2) between the most significantly associated SNP in African Americans (rs62164511) and the most significantly associated SNP in individuals of European ancestry (rs1430193) was low (0.16), providing further evidence of allelic heterogeneity at this locus (**Supplementary Fig. 5a,b**).

Baseline characteristics of the 563 Chinese subjects from MESA Lung are shown in **Supplementary Table 2c**. We performed regional analyses (±200 kb, MAF of ≥0.05) around the sentinel SNPs in individuals of European ancestry using 1000 Genomes Project imputation. The *P*-value threshold was set at 4.41×10^{-5} (on the basis of 1,133 independent tests). None of the 7,436 investigated SNPs had associations that reached a statistical significance level below the predefined threshold.

Baseline characteristics of the 849 Hispanics from the MESA Lung study are shown in **Supplementary Table 2d**. We performed regional analyses of SNPs (MAF ≥ 0.05) within 200 kb of the sentinel SNPs in individuals of European ancestry using 1000 Genomes Project imputation.

None of the 7,473 investigated SNPs reached a statistical significance level below the predefined threshold (4.41×10^{-5}, based on 1,133 independent tests).

Baseline characteristics of the 8,074 Koreans from the Healthy Twin[15–17] and KARE3 (ref. 18) studies are shown in **Supplementary Table 2e**. In this sample, only HapMap (HapMap 3 Phase 2 and Korean HapMap) imputed data were available. There were 72 SNPs (MAF ≥ 0.05) within 200 kb of the sentinel SNPs in individuals of European ancestry. Using the threshold *P* value of 1.52×10^{-3} (based on 26 independent tests), 2 SNPs (rs12449659 and rs4793331, both located approximately 700 kb upstream of *KCNJ2*) were associated with FVC in Koreans with decreases of 32.5 ml (rs12449659, per each copy of the T allele, $P = 7.92 \times 10^{-4}$) and 22.5 ml (rs4793331, per each copy of the A allele, $P = 1.22 \times 10^{-3}$). These SNPs did not show a significant association with FVC in individuals of European ancestry ($P = 0.17$ and 0.34, respectively; **Supplementary Table 5b**).

Gene set enrichment analysis

To identify plausible pathways associated with FVC, we broadened our focus beyond genome-wide significant variants by performing gene set enrichment analysis[25] on the entire set of GWAS variants that underwent meta-analysis. We queried approximately 2,000 gene sets, including canonical pathways and Gene Ontology (GO) functional categories. Using a false discovery rate (FDR) of <0.01, we identified 65 enriched pathways (**Supplementary Table 6**). Although the over-represented gene sets encompassed diverse functions, many involved processes critical to organ development and tissue remodeling, including epithelial morphogenesis, cell proliferation, extracellular matrix development and remodeling, Notch signaling and cell adhesion. Other prominent pathways included acetylcholine binding and channel activity, smooth muscle contraction, glutamate receptor activity, immunity and transcriptional or DNA repair processes.

Table 2 Expression profiling of candidate genes in the lung and periphery

Sentinel SNP (relationship to gene)	Chr.	Gene	Putative function of encoded protein	Tissue			
				Lung	HASM	HBECs	PBMCs
rs1430193 (intronic)	2	*EFEMP1*	Binds EGFR, the EGF receptor, inducing EGFR autophosphorylation and the activation of downstream signaling pathways. May have a role in cell adhesion and migration. May function as a negative regulator of chondrocyte differentiation.	+	+	+	−
rs6923462 (intronic)	6	*BMP6*	BMPs are a family of secreted signaling molecules that can induce ectopic bone growth. Many BMPs are part of the TGF-β superfamily.	+	+	+	+
rs4237643 (intergenic)	11	*MIR129-2* (downstream)–*HSD17B12*	This gene encodes a key 17β-hydroxysteroid dehydrogenase (17β-HSD).	+	+	+	+
rs2863171 (downstream)	11	*PRDM11*	This gene encodes PR domain–containing protein 11.	+	+	+	+
rs1079572 (intronic)	16	*WWOX*	WW domain–containing proteins are found in all eukaryotes and have an important role in the regulation of a wide variety of cellular functions, such as protein degradation, transcription and RNA splicing. This gene encodes a protein that contains 2 WW domains and a short-chain dehydrogenase/reductase domain (SRD).	+	+	+	+
rs6501431 (downstream)	17	*KCNJ2*	The protein encoded by this gene is an integral membrane protein and inward-rectifier type potassium channel. The encoded protein, which has a greater tendency to allow potassium to flow into rather than out of a cell, probably participates in establishing action potential waveform and excitability of neuronal and muscle tissues.	+	+	+	+

A plus sign indicates that the gene is expressed in the cell type used, and a minus sign indicates that we did not detect gene expression at the mRNA level after 70 cycles of PCR. Amplification was followed in real time using gene-specific TaqMan probes, and final PCR products were visualized by gel electrophoresis. We used *GAPDH* (encoding glyceraldehyde-3-phosphate dehydrogenase) as a positive control for the cDNA, and this gene was expressed in all tissues. Chr., chromosome; HASM, human airway smooth muscle; HBECs, human bronchial epithelial cells; PBMCs, peripheral blood mononuclear cells.

Gene expression

Expression profiles of genes from the six loci that were significantly associated in the meta-analysis of stages 1 and 2 (*EFEMP1*, *BMP6*, *WWOX*, *KCNJ2*, *PRDM11* and *HSD17B12*) and the housekeeping gene *GAPDH* were examined in human lung tissue and primary cell samples using RT-PCR. We detected transcripts for all six newly implicated genes in lung tissue, HBECs and HASM cells. Transcripts for five of the six genes (excluding *EFEMP1*) were present in PBMCs (**Table 2** and **Supplementary Figs. 6** and **7**).

Expression quantitative trait locus analysis in peripheral blood cells

We investigated whether the top SNPs or their proxies ($r^2 \geq 0.7$) in the six newly implicated loci for FVC were associated with gene expression using eQTL data (Online Methods). Multiple SNPs in or near *HSD17B12* showed significant *cis*-eQTL associations ($P < 1 \times 10^{-4}$) in peripheral blood, with the strongest association represented by rs11037676 (a proxy of rs4237643, $r^2 = 0.7$) at a P value of 8.42×10^{-81} (**Supplementary Table 3**). The sentinel SNP associated with FVC in this region (rs4237643) also exhibited a strong *cis* effect on *HSD17B12* expression ($P = 1.82 \times 10^{-35}$) (**Supplementary Fig. 8**). Furthermore, this SNP showed a significant *cis*-eQTL association in lymphoblastoid cell lines ($P = 6.7 \times 10^{-11}$)[26] and brain tissue ($P = 1.2 \times 10^{-8}$) (ref. 27). Another FVC-associated SNP located in the intronic region of *EFEMP1* (rs1430189) demonstrated local effects on expression of this gene in eQTL data from human fibroblasts ($P = 4.8 \times 10^{-6}$) (ref. 28). We did not find statistically significant *cis*-eQTLs for the other FVC-associated variants and loci.

eQTL analysis in lung tissue

To better assess the relevance of *cis*-eQTLs to lung biology, we queried a publically available database that included lung tissue (Genotype-Tissue Expression project, GTEx)[29] to further investigate the top SNPs and their proxies. Multiple SNPs that mapped to *HSD17B12*, including rs11037676 and the sentinel SNP rs4237643,

were highly significant *cis*-eQTLs in human lung samples ($P = 2.8 \times 10^{-26}$ and 7.2×10^{-14}, respectively).

Fetal lung mRNA expression for genes associated with FVC

We investigated whether the genes we identified were differentially expressed relative to gestational age during two phases of normal human fetal lung development (**Supplementary Table 7**). There was strong evidence (P value controlling for FDR = 6.7×10^{-6}) for differential expression of *PRDM11*, suggesting that this gene might have an important role *in utero* in lung development. One probe for *WWOX* also showed correlation between lung expression and fetal age, although this correlation was not seen with other probes for the gene.

Putative regulatory variants

We queried the RegulomeDB[30] database to assess whether any of the newly identified FVC-associated SNPs ($P < 1 \times 10^{-7}$, $n = 150$ SNPs) were located in known or predicted regulatory elements, including regions of DNase I hypersensitivity, binding sites for transcription factors and promoter regions that have been biochemically characterized to regulate transcription. Five SNPs received high likelihood scores (based on the amount of supporting data) for mapping to regulatory regions and affecting gene expression; these included four variants upstream of *HSD17B12* (rs9783304, rs2862996, rs10768966 and rs6485443) and one variant downstream of *EFEMP1* (rs1430189). These variants showed evidence of eQTL, transcription factor binding and/or DNase I hypersensitivity peaks.

DISCUSSION

In a 2-stage meta-analysis across 35 cohorts encompassing 85,170 individuals of European ancestry, we found 6 new loci associated with FVC that had not been identified in previous GWAS of spirometric measures of airflow obstruction (FEV$_1$ or FEV$_1$/FVC). The six new loci showed consistent associations across the studies of individuals of European ancestry in the discovery and replication stages. Effect estimates in the meta-analysis range from 13 to 31 ml per allele, values

similar to the annual rate of decline in FVC ranging from 12 to 47 ml in the general population[31]. Expression analyses showed that all the top candidate genes at these loci were expressed in lung tissue and primary lung cells (HBECs and HASM cells). Two additional loci associated with FVC at genome-wide significance in the stage 1 analysis were previously associated with FEV_1/FVC (*PTCH1*) or FEV_1 (*GSTCD*)[12–14] at genome-wide levels of significance in GWAS.

The six new associations found in this analysis explain only a modest proportion of the additive polygenic variance in FVC (0.74%). Stage 2 effect size estimates were used to calculate the proportion of the variance explained by the six new loci, to avoid the effect of bias from winner's curse. When we take the other known loci for pulmonary function into account, the proportion of the additive polygenic variance explained is 1.78%, a finding that is comparable to those with many other complex traits[32]. Unexplained heritability has become a well-known phenomenon in genetic epidemiology[33], and possible explanations include multiple effects of common variants, rare variant effects, gene-by-environment interactions, gene-gene interactions and epigenetic regulation—mechanisms that are not captured by existing GWAS platforms.

We and others previously identified 27 regions associated at genome-wide significance with FEV_1, FEV_1/FVC or both[11–14]. Although FEV_1 and FVC are statistically correlated ($r = 0.83$ in the Rotterdam Study, adjusted for age, sex, height and height[2]), these measures represent clinically different entities. FVC is used for the evaluation of restrictive ventilatory defects and is a predictor of mortality independent of FEV_1, standard risk factors and even previous cardiovascular disease[5,6]. In contrast, FVC and FEV_1/FVC have a very low correlation ($r = −0.08$ in the Rotterdam Study, adjusted for age, sex, height and height[2]). In this analysis, we were able to identify six new loci that are associated with FVC at genome-wide significance. Only two of the loci that were previously associated with FEV_1 or FEV_1/FVC showed genome-wide significant association with FVC in our study (*GSTCD* and *PTCH1*)[12,13]. The sentinel SNPs at each of the six newly associated loci showed consistent directions of effect on both FEV_1 and FVC (**Supplementary Table 8**). Among our FVC-associated SNPs, the smallest P value for FEV_1 (9.43×10^{-7}) was observed for rs1079572 (*WWOX*). An intragenic SNP (rs11654749) in the region between *KCNJ2* and *SOX9* was associated with FEV_1 at genome-wide significance in our previous meta-analysis of SNP and SNP-by-smoking effects[34]. To assess whether the variant identified in that analysis of FEV_1 (rs11654749) and the sentinel SNP from our current FVC analysis (rs6501431) are representative of the same signal, we fitted both variants together in the model using GCTA[35] software. The effect sizes of these SNPs increased slightly when they were fitted together, as expected given that the marginal correlation ($r = 0.03$) of their alleles was positive and the effects of these alleles on lung function were in opposite directions. Thus, these SNPs appear to represent independent signals.

Two of the top loci for FVC (*BMP6* and *EFEMP1*) have been associated with height in a previous GWAS[36]. The FVC-associated SNPs in or near these two genes show modest to weak correlation with the top SNPs from the GWAS of height[36]. For *EFEMP1*, the r^2 value between rs1430193 (for FVC) and rs3791675 (for height) was 0.45. For *BMP6*, the correlation between rs6923462 (for FVC) and two SNPs associated with height (rs3812163 and rs1219896) was low. For rs6923462 and rs3812163, the r^2 value was 0.01, and, for rs6923462 and rs1219896, the r^2 value was 0.02. Because we adjusted for height in our analysis, our findings are likely to be independent of height but might reflect genetic effects on body or organ size.

To assess whether our identified loci are associated with FVC across populations of different ancestry, we examined the associations of our main findings in African-American, Korean, Hispanic and Chinese subjects. Despite the limited size of these samples, 78 SNPs in the region of *EFEMP1* had associations that reached the significance threshold of $P < 4.42 \times 10^{-5}$ in African Americans. These results support the involvement of this locus in lung function in individuals of both European and African descent, although there was evidence for allelic heterogeneity between these populations. In the Korean data set, we found two SNPs in the region of our locus near *KCNJ2* to be significantly associated with FVC. In the smaller samples of Chinese and Hispanic participants, none of the investigated SNPs were significantly associated with FVC, which may not be surprising given the greatly reduced power. In summary, despite smaller sample sizes, we were able to show significant evidence of association with FVC for the *EFEMP1* locus in African Americans and for the locus downstream of *KCNJ2* in Koreans.

A literature review identified candidate genes within the newly associated loci plausibly involved in lung growth and pathogenesis. For example, *BMP6* is a member of the bone morphogenetic protein (BMP) family that constitutes a key canonical signaling pathway in the regulation of lung development, repair and response to injury[37]. *BMP6* expression in bronchial epithelial cells has been reported to increase in experimental models of allergic airway inflammation[38]. *EFEMP1* is part of the fibulin family of extracellular matrix glycoproteins and encodes fibulin-3. Targeted disruption of a member of this family, *Efemp2* (also known as *fibulin-4*), has been shown to cause reduced elasticity and emphysematous morphology in the lungs of mice[39]. Expression of other members of the fibulin family (fibulin-1 and fibulin-5) seems to be influenced by *TGFB1* (encoding transforming growth factor-β1), a gene previously linked to inflammation in individuals with chronic obstructive pulmonary disease (COPD)[40,41]. *WWOX* might influence protein-induced apoptosis and behaves as a tumor suppressor gene in various types of neoplasms, including small-cell lung cancer[42]. Our analysis in fetal lung showed strong evidence of differential expression of *PRDM11*, suggesting that this gene has a role in development *in utero*. In addition, nuclear expression of *PRDM11* has been shown in respiratory epithelial cells of the human bronchus in three subjects in the Human Protein Atlas[43]. Interestingly, our eQTL analysis showed highly significant *cis* effects of FVC-associated variants on *HSD17B12* expression in multiple tissues, including lung. Furthermore, several of these SNPs were located within regulatory sites upstream of *HSD17B12*, suggesting that these variants might have a regulatory role in gene expression. Exploratory pathway analysis using all SNPs in the meta-analysis of FVC implicated multiple processes, including several involved in tissue development and remodeling. These findings suggest that distinct and identifiable biological pathways underlie the genetic basis of lung vital capacity in the general population.

There are some limitations to our analysis. With cross-sectional measures of FVC, we cannot determine whether the identified signals are due to influence on lung growth or age-related decline in lung function[44]. The primary analyses were not adjusted for pack-years (the duration of smoking in years multiplied by the number of cigarettes per day, divided by 20 cigarettes per pack) to avoid attrition in sample size, but, within the CHARGE cohorts, estimates from meta-analysis with and without adjustment for pack years were very similar (data not shown).

A key strength of our study is its considerable sample size of European-ancestry individuals. Our application of genomic control

at the three stages is likely to be overly conservative because it has recently been shown that, in large meta-analyses, test statistics are expected to be elevated under polygenic inheritance, even when there is no population structure. Estimates of genomic inflation increase with sample size, as has been shown for other traits[14,36,45,46]. Following the two-stage meta-analysis, we were able to test the associated SNPs and their regions for tissue-specific expression.

In conclusion, using a large-scale staged meta-analysis, we report six new loci associated with FVC and show that all are expressed in lung tissue and primary lung cells. Our findings point to previously unexplored pathways and mechanisms underlying lung function. Improvement of the understanding of the role that these genes have in normal lung development and pathogenesis could lead to the identification of novel therapeutic targets for lung diseases.

URLs. Korean HapMap Project, http://www.khapmap.org/; RegulomeDB http://regulome.stanford.edu/; Human Protein Atlas, http://www.proteinatlas.org/; METAL, http://www.sph.umich.edu/csg/abecasis/metal/; matSpD, http://gump.qimr.edu.au/general/daleN/matSpD/; Gene Expression Omnibus, http://www.ncbi.nlm.nih.gov/geo/.

METHODS
Methods and any associated references are available in the online version of the paper.

Accession codes. Data from transcriptomic analyses of human lung development are at the NCBI Gene Expression Omnibus under accession GSE14334.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

AUTHOR CONTRIBUTIONS
Project conception, design and management. *Stage 1 GWAS. Ages:* G.E., M.G., V.G., T.B.H., L.J.L. *ARIC:* S.J.L., N.F., D.C., D.B.H., B.R.J., A.C.M., K.E.N. *B58C T1DGC:* D.P.S. *B58C WTCCC:* D.P.S. *CARDIA:* A.S., L.J.S. *CHS:* S.A.G., S.R.H., B.M.P. *CROATIA-Korcula:* O.P., A.F.W. *CROATIA-Vis:* I.R. *ECRHS:* D.J., E.O., I.P., M. Wjst *EPIC:* R.A.S., N.J.W., J.H.Z. *FHS:* J.B.W., G.T.O., J.D. *FTC:* J.K., K.H.P., T.R., A. Viljanen *Health ABC:* P.A.C., T.B.H., S.B.K., Y.L. *Health 2000:* M.H., M.K. *HCS:* J.R.A., R.J.S. *KORA S3:* C.G., J. Heinrich *MESA-Lung:* R.G.B. *NFBC1966:* M.-R.J., A.P. *NSPHS:* U.G. *ORCADES:* H.C., J.F.W., S.H.W. *RS I, II and III:* A.H., B.H. Stricker, G.G.B. *SHIP:* S.G., B.K., H.V. *Twins UK-I:* C.J.H., T.D.S. *Stage 2 follow-up studies. BHS 1 and 2:* A.L.J., A.B.M., J.B. *CROATIA-Split:* N.D.H., C.H. *Generation Scotland:* D.J.P., B.H. Smith. *KORA F4:* H.S. *LBC1936:* I.J.D., J.M.S. *LifeLines:* H.M.B., D.S.P., J.M.V., C.W. *LLFS:* A.N., B.T., M. Wojczynski, R.L.M. *PIVUS:* E.I., L. Lind *Twins UK-II and III:* C.J.H., T.D.S. *Multi-ancestry follow-up studies. CARe:* R.G.B., K.M.B., D.J.L., R.K., L.J.S., S. Redline, E.G.B., G.T.O., L.R.L., W.B.W. *KARE3 and Healthy Twin Study:* J.S., W.J.K., Y.-M.O. *Gene expression analyses. RT-PCR:* K.R.B., G.G.B. *eQTL analysis:* J.B.J.v.M., A.G.U. *Fetal lung expression analysis:* I.P.H., I. Sayers, E.M.

Phenotype collection and data management. *Stage 1 GWAS. Ages:* T.A. *ARIC:* D.C., N.F., A.C.M., K.E.N. *B58C T1DGC:* A.R.R., D.P.S. *B58C WTCCC:* A.R.R., D.P.S. *CARDIA:* L.J.S., O.D.W. *CHS:* S.A.G., S.R.H., B.M.P., T.L. *CROATIA-Korcula:* O.P., L.Z. *CROATIA-Vis:* S.C., I.K. *ECRHS:* D.L.J., E.O., I.P., M. Wjst *EPIC:* J.H.Z. *FHS:* J.B.W., G.T.O., J.D. *FTC:* J.K., K.H.P., T.R., A. Viljanen *Health ABC:* P.A.C., W.T. *Health 2000:* M.H., M.K. *HCS:* J.R.A. *KORA S3:* J. Heinrich *MESA-Lung:* R.G.B. *NFBC1966:* M.-R.J., J.P., A.P. *NSPHS:* Å.J., S.E., U.G. *ORCADES:* S.H.W., J.F.W. *RS I, II and III:* G.G.B., L. Lahousse, D.W.L., B.H. Stricker *SHIP:* S.G., B.K., H.V. *Twins UK-I:* P.G.H., A. Viñuela. *Stage 2 follow-up studies. BHS 1 and 2:* A.L.J., A.B.M., J.B. *CROATIA-Split:* C.H., T.Z. *Generation Scotland:* D.J.P., B.H. Smith. *KORA F4:* R.H., S.K., H.S. *LBC1936:* I.J.D., J.M.S. *LifeLines:* D.S.P., J.M.V. *LLFS:* A.N., B.T., R.L.M. *PIVUS:* E.I., L. Lind *Twins UK-II and III:* P.G.H., A. Viñuela

Multi-ancestry follow-up studies. CARe: R.G.B., T.D.P., K.M.B., D.J.L., R.K., L.J.S., J.B.W., N.H., M.F.P., K.M.B., S. Ripatti, E.G.B., G.T.O., L.R.L., W.B.W. *KARE3 and Healthy Twin Study:* J.S., W.J.K., Y.-M.O. *Gene expression analyses. RT-PCR:* K.R.B., G.G.B., F.M.V., P.S.H. *eQTL analysis:* J.B.J.v.M., M.J.P. *Fetal lung expression analysis:* I.P.H., I. Sayers, E.M.

Genotyping. *Stage 1 GWAS. B58C T1DGC:* W.L.M. *B58C WTCCC:* W.L.M. *CARDIA:* M.F., X.G. *CHS:* J.I.R., B.M.P. *CROATIA-Korcula:* J.E.H. *CROATIA-Vis:* S.C. *ECRHS:* M. Wjst *EPIC:* J.H.Z. *FTC:* J.K. *Health ABC:* Y.L., K.L. *Health 2000:* S. Ripatti, I. Surakka. *HCS:* R.J.S. *KORA S3:* H.G. *MESA-Lung:* S.S.R. *NFBC1966:* M.-R.J. *NSPHS:* Å.J., S.E., U.G. *Orcades:* H.C., J.F.W. *RS I, II and III:* F.R., A.G.U. *SHIP:* S.G., B.K., A.T., H.V. *Twins UK-I:* C.J.H., T.D.S. *Stage 2 follow-up studies. BHS 1 and 2:* J. Hui. *CROATIA-Split:* C.H., P.N., T.Z. *Generation Scotland:* D.J.P., B.H. Smith, H.T. *LBC1936:* G.D. *LifeLines:* C.W. *LLFS:* A.N., B.T. *PIVUS:* E.I., A.P.M. *Twins UK-II and III:* C.J.H., T.D.S.

Data analysis. *Stage 1 GWAS. Ages:* A.V.S. *ARIC:* N.F., D.B.H. *B58C T1DGC:* A.R.R., D.P.S. *B58C WTCCC:* A.R.R., D.P.S. *CARDIA:* X.G. *CHS:* S.A.G., G.L., S.R.H., T.L. *CROATIA-Korcula:* J.E.H. *CROATIA-Vis:* V.V. *ECRHS:* D.L.J., A.R. *EPIC:* J.H.Z. *FHS:* J.B.W., J.D., W.G. *Health ABC:* P.A.C., Y.L., K.L., W.T. *Health 2000:* M.K., S. Ripatti, I. Surakka. *HCS:* C.O., E.G.H. *KORA S3:* E.A. *MESA-Lung:* A.M., S.S.R. *NFBC1966:* A.C.A. *NSPHS:* S.E. *ORCADES:* P.K.J. *RS I, II and III:* L. Lahousse, D.W.L. *SHIP:* A.T. *Twins UK-I:* P.G.H. *Stage 2 follow-up studies. KORA F4:* C.F., R.H. *LBC1936:* L.M.L. *LifeLines:* K.d.J., H.M.B. *LLFS:* M. Wojczynski, B.T. *PIVUS:* T.F. *Twins UK-II and III:* P.G.H. *Multi-ancestry follow-up studies.* N.C.G. *CARe:* T.D.P., Q.D., L.A.L., X.-Q.W. *KARE3 and Healthy Twin Study:* M.K.L. *Gene expression analyses. RT-PCR:* K.R.B., F.M.V. *eQTL analysis:* M.J.P. *Fetal lung expression analysis:* I.P.H., I. Sayers, E.M.

Analysis group: *CHARGE Consortium:* D.W.L., S.A.G., S.J.L., J.D., N.F., A.V.S. *CARe:* T.D.P., Q.D. *SpiroMeta Consortium:* M.S.A., L.V.W., B.K., I.P.H., M.D.T.

Writing group: *CHARGE Consortium:* S.J.L., D.W.L., S.A.G., N.F., J.D., G.G.B., A.S. *SpiroMeta Consortium:* I.P.H., M.S.A., M.D.T., L.V.W., C.H.

1. Wilk, J.B. *et al.* Evidence for major genes influencing pulmonary function in the NHLBI family heart study. *Genet. Epidemiol.* **19**, 81–94 (2000).
2. 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).
3. Raghu, G. *et al.* An official ATS/ERS/JRS/ALAT statement: idiopathic pulmonary fibrosis: evidence-based guidelines for diagnosis and management. *Am. J. Respir. Crit. Care Med.* **183**, 788–824 (2011).
4. Ashley, F., Kannel, W.B., Sorlie, P.D. & Masson, R. Pulmonary function: relation to aging, cigarette habit, and mortality. *Ann. Intern. Med.* **82**, 739–745 (1975).
5. Burney, P.G. & Hooper, R. Forced vital capacity, airway obstruction and survival in a general population sample from the USA. *Thorax* **66**, 49–54 (2011).
6. Lee, H.M., Le, H., Lee, B.T., Lopez, V.A. & Wong, N.D. Forced vital capacity paired with Framingham Risk Score for prediction of all-cause mortality. *Eur. Respir. J.* **36**, 1002–1006 (2010).
7. Lange, P., Nyboe, J., Appleyard, M., Jensen, G. & Schnohr, P. Spirometric findings and mortality in never-smokers. *J. Clin. Epidemiol.* **43**, 867–873 (1990).
8. Kannel, W.B., Hubert, H. & Lew, E.A. Vital capacity as a predictor of cardiovascular disease: the Framingham study. *Am. Heart J.* **105**, 311–315 (1983).
9. 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).
10. van Putte-Katier, N. *et al.* Relationship between parental lung function and their children's lung function early in life. *Eur. Respir. J.* **38**, 664–671 (2011).
11. Wilk, J.B. *et al.* A genome-wide association study of pulmonary function measures in the Framingham Heart Study. *PLoS Genet.* **5**, e1000429 (2009).
12. Repapi, E. *et al.* Genome-wide association study identifies five loci associated with lung function. *Nat. Genet.* **42**, 36–44 (2010).
13. 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).
14. 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).
15. Sung, J. *et al.* The Korean Twin Registry—methods, current stage, and interim results. *Twin Res.* **5**, 394–400 (2002).
16. Sung, J. *et al.* Healthy Twin: a twin-family study of Korea—protocols and current status. *Twin Res. Hum. Genet.* **9**, 844–848 (2006).
17. Gombojav, B. *et al.* The Healthy Twin Study, Korea updates: resources for omics and genome epidemiology studies. *Twin Res. Hum. Genet.* **16**, 241–245 (2013).

18. Cho, Y.S. *et al.* A large-scale genome-wide association study of Asian populations uncovers genetic factors influencing eight quantitative traits. *Nat. Genet.* **41**, 527–534 (2009).
19. Devlin, B. & Roeder, K. Genomic control for association studies. *Biometrics* **55**, 997–1004 (1999).
20. 1000 Genomes Project Consortium. A map of human genome variation from population-scale sequencing. *Nature* **467**, 1061–1073 (2010).
21. Nyholt, D.R. A simple correction for multiple testing for single-nucleotide polymorphisms in linkage disequilibrium with each other. *Am. J. Hum. Genet.* **74**, 765–769 (2004).
22. Li, J. & Ji, L. Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix. *Heredity (Edinb.)* **95**, 221–227 (2005).
23. Musunuru, K. *et al.* Candidate gene association resource (CARe): design, methods, and proof of concept. *Circ Cardiovasc Genet* **3**, 267–275 (2010).
24. Carlson, C.S. *et al.* Generalization and dilution of association results from European GWAS in populations of non-European ancestry: the PAGE study. *PLoS Biol.* **11**, e1001661 (2013).
25. Zhang, K., Cui, S., Chang, S., Zhang, L. & Wang, J. i-GSEA4GWAS: a web server for identification of pathways/gene sets associated with traits by applying an improved gene set enrichment analysis to genome-wide association study. *Nucleic Acids Res.* **38**, W90–W95 (2010).
26. Dixon, A.L. *et al.* A genome-wide association study of global gene expression. *Nat. Genet.* **39**, 1202–1207 (2007).
27. Gibbs, J.R. *et al.* Abundant quantitative trait loci exist for DNA methylation and gene expression in human brain. *PLoS Genet.* **6**, e1000952 (2010).
28. Dimas, A.S. *et al.* Common regulatory variation impacts gene expression in a cell type–dependent manner. *Science* **325**, 1246–1250 (2009).
29. Genotype-Tissue Expression Consortium. The Genotype-Tissue Expression (GTEx) project. *Nat. Genet.* **45**, 580–585 (2013).
30. Boyle, A.P. *et al.* Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res.* **22**, 1790–1797 (2012).
31. Gottlieb, D.J. *et al.* Heritability of longitudinal change in lung function. The Framingham study. *Am. J. Respir. Crit. Care Med.* **164**, 1655–1659 (2001).
32. Park, J.H. *et al.* Estimation of effect size distribution from genome-wide association studies and implications for future discoveries. *Nat. Genet.* **42**, 570–575 (2010).
33. Eichler, E.E. *et al.* Missing heritability and strategies for finding the underlying causes of complex disease. *Nat. Rev. Genet.* **11**, 446–450 (2010).
34. Hancock, D.B. *et al.* Genome-wide joint meta-analysis of SNP and SNP-by-smoking interaction identifies novel loci for pulmonary function. *PLoS Genet.* **8**, e1003098 (2012).
35. Yang, J. *et al.* Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat. Genet.* **44**, 369–375 (2012).
36. Lango Allen, H. *et al.* Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature* **467**, 832–838 (2010).
37. Sountoulidis, A. *et al.* Activation of the canonical Bone Morphogenetic Protein (BMP) pathway during lung morphogenesis and adult lung tissue repair. *PLoS ONE* **7**, e41460 (2012).
38. Rosendahl, A. *et al.* Activation of bone morphogenetic protein/Smad signaling in bronchial epithelial cells during airway inflammation. *Am. J. Respir. Cell Mol. Biol.* **27**, 160–169 (2002).
39. McLaughlin, P.J. *et al.* Targeted disruption of fibulin-4 abolishes elastogenesis and causes perinatal lethality in mice. *Mol. Cell. Biol.* **26**, 1700–1709 (2006).
40. de Boer, W.I. *et al.* Transforming growth factor β1 and recruitment of macrophages and mast cells in airways in chronic obstructive pulmonary disease. *Am. J. Respir. Crit. Care Med.* **158**, 1951–1957 (1998).
41. Takizawa, H. *et al.* Increased expression of transforming growth factor-β1 in small airway epithelium from tobacco smokers and patients with chronic obstructive pulmonary disease (COPD). *Am. J. Respir. Crit. Care Med.* **163**, 1476–1483 (2001).
42. Kimura, M. *et al.* Bmi1 regulates cell fate via tumor suppressor WWOX repression in small-cell lung cancer cells. *Cancer Sci.* **102**, 983–990 (2011).
43. Uhlen, M. *et al.* Towards a knowledge-based Human Protein Atlas. *Nat. Biotechnol.* **28**, 1248–1250 (2010).
44. Weiss, S.T. Lung function and airway diseases. *Nat. Genet.* **42**, 14–16 (2010).
45. Elks, C.E. *et al.* Thirty new loci for age at menarche identified by a meta-analysis of genome-wide association studies. *Nat. Genet.* **42**, 1077–1085 (2010).
46. Lindgren, C.M. *et al.* Genome-wide association scan meta-analysis identifies three loci influencing adiposity and fat distribution. *PLoS Genet.* **5**, e1000508 (2009).

Daan W Loth[1,2,135], María Soler Artigas[3,4,135], Sina A Gharib[5,6,135], Louise V Wain[3,4,135], Nora Franceschini[7,8,135], Beate Koch[9,135], Tess D Pottinger[10,135], Albert Vernon Smith[11,12,135], Qing Duan[13,135], Chris Oldmeadow[14,15], Mi Kyeong Lee[16], David P Strachan[17], Alan L James[18–20], Jennifer E Huffman[21], Veronique Vitart[21], Adaikalavan Ramasamy[22,23], Nicholas J Wareham[24], Jaakko Kaprio[25–27], Xin-Qun Wang[28], Holly Trochet[21], Mika Kähönen[29], Claudia Flexeder[30], Eva Albrecht[31], Lorna M Lopez[32,33], Kim de Jong[34,35], Bharat Thyagarajan[36], Alexessander Couto Alves[23], Stefan Enroth[37,38], Ernst Omenaas[39,40], Peter K Joshi[41], Tove Fall[38,42], Ana Viñuela[43], Lenore J Launer[44], Laura R Loehr[7,8], Myriam Fornage[45,46], Guo Li[47], Jemma B Wilk[48], Wenbo Tang[49], Ani Manichaikul[28,50], Lies Lahousse[1,51], Tamara B Harris[44], Kari E North[7], Alicja R Rudnicka[17], Jennie Hui[52], Xiangjun Gu[45,46], Thomas Lumley[53], Alan F Wright[21], Nicholas D Hastie[21], Susan Campbell[21], Rajesh Kumar[54], Isabelle Pin[55–57], Robert A Scott[24], Kirsi H Pietiläinen[27,58,59], Ida Surakka[27,60], Yongmei Liu[61], Elizabeth G Holliday[14,15], Holger Schulz[30], Joachim Heinrich[30,62], Gail Davies[32,33,63,64], Judith M Vonk[34,35], Mary Wojczynski[65], Anneli Pouta[66,67], Åsa Johansson[37,38,68], Sarah H Wild[41], Erik Ingelsson[38,42,69], Fernando Rivadeneira[70,71], Henry Völzke[72], Pirro G Hysi[43], Gudny Eiriksdottir[11], Alanna C Morrison[73], Jerome I Rotter[74,75], Wei Gao[76], Dirkje S Postma[35,77], Wendy B White[78], Stephen S Rich[50], Albert Hofman[1,71], Thor Aspelund[11,12], David Couper[79], Lewis J Smith[54], Bruce M Psaty[6,47,80,81], Kurt Lohman[82], Esteban G Burchard[83,84], André G Uitterlinden[1,70,71], Melissa Garcia[44], Bonnie R Joubert[85], Wendy L McArdle[86], A Bill Musk[87], Nadia Hansel[88], Susan R Heckbert[47,80,81], Lina Zgaga[89,90], Joyce B J van Meurs[70,71], Pau Navarro[21], Igor Rudan[41], Yeon-Mok Oh[91,92], Susan Redline[93], Deborah L Jarvis[22,94], Jing Hua Zhao[24], Taina Rantanen[95], George T O'Connor[96,97], Samuli Ripatti[27,60,98], Rodney J Scott[14,15], Stefan Karrasch[30,99,100], Harald Grallert[101], Nathan C Gaddis[102], John M Starr[32,103], Cisca Wijmenga[104], Ryan L Minster[105], David J Lederer[10,106], Juha Pekkanen[107,108], Ulf Gyllensten[37,38], Harry Campbell[41], Andrew P Morris[69], Sven Gläser[9], Christopher J Hammond[43], Kristin M Burkart[10], John Beilby[52], Stephen B Kritchevsky[109], Vilmundur Gudnason[11,12], Dana B Hancock[85,110], O Dale Williams[111], Ozren Polasek[112], Tatijana Zemunik[113], Ivana Kolcic[112], Marcy F Petrini[114], Matthias Wjst[115], Woo Jin Kim[116,117], David J Porteous[63], Generation Scotland[118], Blair H Smith[119], Anne Viljanen[95], Markku Heliövaara[26], John R Attia[14,15], Ian Sayers[120], Regina Hampel[121], Christian Gieger[31], Ian J Deary[32,33], H Marike Boezen[34,35], Anne Newman[122], Marjo-Riitta Jarvelin[23,123–126], James F Wilson[41], Lars Lind[127], Bruno H Stricker[1,2,70,71], Alexander Teumer[128], Timothy D Spector[43],

Erik Melén[129], Marjolein J Peters[70,71], Leslie A Lange[13], R Graham Barr[10,106], Ken R Bracke[51], Fien M Verhamme[51], Joohon Sung[16,130], Pieter S Hiemstra[131], Patricia A Cassano[49,132], Akshay Sood[133,136], Caroline Hayward[21,136], Josée Dupuis[76,97,136], Ian P Hall[120,136], Guy G Brusselle[1,51,134,136], Martin D Tobin[3,4,136] & Stephanie J London[85,136]

[1]Department of Epidemiology, Erasmus MC, Rotterdam, the Netherlands. [2]Netherlands Health Care Inspectorate, The Hague, the Netherlands. [3]Genetic Epidemiology Group, Department of Health Sciences, University of Leicester, Leicester, UK. [4]National Institute for Health Research (NIHR) Leicester Respiratory Biomedical Research Unit, Glenfield Hospital, Leicester, UK. [5]Computational Medicine Core, Center for Lung Biology, University of Washington, Seattle, Washington, USA. [6]Department of Medicine, University of Washington, Seattle, Washington, USA. [7]Department of Epidemiology, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA. [8]Carolina Center for Genome Sciences, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA. [9]Department of Internal Medicine B–Pneumology, Cardiology, Intensive Care and Infectious Diseases, University Hospital Greifswald, Greifswald, Germany. [10]Department of Medicine, College of Physicians and Surgeons, Columbia University, New York, New York, USA. [11]Iceland Heart Association, Kopavogur, Iceland. [12]University of Iceland, Reykjavik, Iceland. [13]Department of Genetics, University of North Carolina, Chapel Hill, North Carolina, USA. [14]Hunter Medical Research Institute, University of Newcastle, Newcastle, New South Wales, Australia. [15]Faculty of Health, University of Newcastle, Newcastle, New South Wales, Australia. [16]Institute of Health and Environment, Seoul National University, Seoul, South Korea. [17]Division of Population Health Sciences and Education, St George's, University of London, London, UK. [18]Department of Pulmonary Physiology and Sleep Medicine/West Australian Sleep Disorders Research Institute, Nedlands, Western Australia, Australia. [19]School of Medicine and Pharmacology, The University of Western Australia, Perth, Western Australia, Australia. [20]Busselton Population Medical Research Institute, Busselton, Western Australia, Australia. [21]Medical Research Council (MRC) Human Genetics Unit, MRC Institute of Genetics and Molecular Medicine (IGMM), University of Edinburgh, Edinburgh, UK. [22]Respiratory Epidemiology and Public Health Group, National Heart and Lung Institute, Imperial College London, London, UK. [23]Department of Epidemiology and Biostatistics, MRC Health Protection Agency (HPA) Centre for Environment and Health, School of Public Health, Imperial College London, London, UK. [24]MRC Epidemiology Unit, Institute of Metabolic Science, Addenbrooke's Hospital, Cambridge, UK. [25]Hjelt Institute, Department of Public Health, University of Helsinki, Helsinki, Finland. [26]National Institute for Health and Welfare (THL), Helsinki, Finland. [27]Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, Finland. [28]Division of Biostatistics and Epidemiology, Department of Public Health Sciences, University of Virginia, Charlottesville, Virginia, USA. [29]Department of Clinical Physiology, University of Tampere and Tampere University Hospital, Tampere, Finland. [30]Institute of Epidemiology I, Helmholtz Zentrum München–German Research Center for Environmental Health, Neuherberg, Germany. [31]Institute of Genetic Epidemiology, Helmholtz Zentrum München–German Research Center for Environmental Health, Neuherberg, Germany. [32]Centre for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh, Edinburgh, UK. [33]Department of Psychology, University of Edinburgh, Edinburgh, UK. [34]Department of Epidemiology, University of Groningen, University Medical Center Groningen, Groningen, the Netherlands. [35]Groningen Research Institute for Asthma and COPD (GRIAC), University of Groningen, University Medical Center Groningen, Groningen, the Netherlands. [36]Department of Laboratory Medicine and Pathology, University of Minnesota, Minneapolis, Minnesota, USA. [37]Department of Immunology, Genetics and Pathology, Rudbeck Laboratory, Uppsala University, Uppsala, Sweden. [38]Science for Life Laboratory, Uppsala University, Uppsala, Sweden. [39]Centre for Clinical Research, Haukeland University Hospital, Bergen, Norway. [40]Department of Clinical Sciences, University of Bergen, Bergen, Norway. [41]Centre for Population Health Sciences, Medical School, University of Edinburgh, Edinburgh, UK. [42]Molecular Epidemiology, Department of Medical Sciences, Uppsala University, Uppsala, Sweden. [43]Department of Twins Research and Genetic Epidemiology, King's College London, London, UK. [44]Laboratory of Epidemiology, Demography and Biometry, National Institute on Aging, US National Institutes of Health, Bethesda, Maryland, USA. [45]Institute of Molecular Medicine, University of Texas Health Science Center at Houston, Houston, Texas, USA. [46]Human Genetics Center, University of Texas Health Science Center at Houston, Houston, Texas, USA. [47]Cardiovascular Health Research Unit, University of Washington, Seattle, Washington, USA. [48]Precision Medicine, Pfizer Global Research and Development, Cambridge, Massachusetts, USA. [49]Division of Nutritional Sciences, Cornell University, Ithaca, New York, USA. [50]Center for Public Health Genomics, University of Virginia, Charlottesville, Virginia, USA. [51]Department of Respiratory Medicine, Ghent University Hospital, Ghent, Belgium. [52]PathWest Laboratory Medicine Washington, Nedlands, Western Australia, Australia. [53]Department of Statistics, University of Auckland, Auckland, New Zealand. [54]Northwestern University Feinberg School of Medicine, Chicago, Illinois, USA. [55]Centre Hospitalier Universitaire de Grenoble, Grenoble, France. [56]INSERM U823, Institut Albert Bonniot, Grenoble, France. [57]Université Joseph Fourier, Grenoble, France. [58]Obesity Research Unit, Research Programs Unit, Diabetes and Obesity, University of Helsinki, Helsinki, Finland. [59]Division of Internal Medicine, Department of Medicine, Helsinki University Central Hospital, Helsinki, Finland. [60]Public Health Genomics Unit, Department of Chronic Disease Prevention, National Institute for Health and Welfare (THL), Helsinki, Finland. [61]Department of Epidemiology and Prevention, Division of Public Health Sciences, Wake Forest School of Medicine, Winston-Salem, North Carolina, USA. [62]Comprehensive Pneumology Center Munich (CPC-M), member of the German Center for Lung Research, Munich, Germany. [63]Medical Genetics Section, Centre for Genomics and Experimental Medicine, MRC IGMM, University of Edinburgh, Edinburgh, UK. [64]MRC Institute of Genetics and Molecular Medicine, Edinburgh, UK. [65]Department of Statistical Genomics, Washington University, St. Louis, Missouri, USA. [66]National Institute for Health and Welfare, Oulu, Finland. [67]Department of Clinical Sciences/Obstetrics and Gynecology, University Hospital of Oulu, University of Oulu, Oulu, Finland. [68]Uppsala Clinical Research Center, Uppsala University, Uppsala, Sweden. [69]Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK. [70]Department of Internal Medicine, Erasmus MC, Rotterdam, the Netherlands. [71]Netherlands Genomics Initiative (NGI)-sponsored Netherlands Consortium for Healthy Aging (NCHA), Rotterdam, the Netherlands. [72]Institute for Community Medicine, University Medicine Greifswald, Greifswald, Germany. [73]School of Public Health, University of Texas Health Science Center at Houston, Houston, Texas, USA. [74]Biomedical Research Institute, Harbor–University of California, Los Angeles (UCLA) Medical Center, Torrance, California, USA. [75]Department of Pediatrics, Harbor-UCLA Medical Center, Torrance, California, USA. [76]Department of Biostatistics, Boston University School of Public Health, Boston, Massachusetts, USA. [77]Department of Pulmonology, University of Groningen, University Medical Center Groningen, Groningen, the Netherlands. [78]Tougaloo College, Jackson, Mississippi, USA. [79]Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA. [80]Department of Epidemiology, University of Washington, Seattle, Washington, USA. [81]Group Health Research Institute, Group Health Cooperative, Seattle, Washington, USA. [82]Department of Biostatistical Sciences, Division of Public Health Sciences, Wake Forest School of Medicine, Winston-Salem, North Carolina, USA. [83]Department of Bioengineering and Therapeutic Sciences, University of California, San Francisco, San Francisco, California, USA. [84]Department of Medicine, University of California, San Francisco, San Francisco, California, USA. [85]Epidemiology Branch, National Institute of Environmental Health Sciences, US National Institutes of Health, US Department of Health and Human Services, Research Triangle Park, North Carolina, USA. [86]School of Social and Community Medicine, University of Bristol, Bristol, UK. [87]Department of Respiratory Medicine, Sir Charles Gairdner Hospital, Nedlands, Western Australia, Australia. [88]Department of Medicine, Johns Hopkins University, Baltimore, Maryland, USA. [89]Department of Public Health and Primary Care, Trinity College Dublin, Dublin, Ireland. [90]Adrija Stampar School of Public Health, Medical School, University of Zagreb, Zagreb, Croatia. [91]Department of Pulmonary and Critical Care Medicine, Asan Medical Center, University of Ulsan College of Medicine, Seoul, South Korea. [92]Clinical Research Center for Chronic Obstructive Airway Diseases, Asan Medical Center, University of Ulsan College of Medicine, Seoul, South Korea. [93]Department of Medicine, Brigham and Women's Hospital, Boston, Massachusetts, USA. [94]MRC-PHE Centre for Environment and Health, Imperial College London, London, UK. [95]Gerontology Research Centre, Department of Health Sciences, University of Jyväskylä, Jyväskylä, Finland. [96]Pulmonary Center, Boston University School of Medicine, Boston, Massachusetts, USA. [97]National Heart, Lung, and Blood Institute's Framingham Heart Study, Framingham, Massachusetts, USA. [98]Genetic Epidemiology Group, Wellcome Trust Sanger Institute, Hinxton, UK. [99]Institute and Outpatient Clinic for Occupational, Social and Environmental Medicine, Ludwig Maximilians Universität, Munich, Germany. [100]Institute of General Practice, University Hospital Klinikum Rechts der Isar, Technische Universität München, Munich, Germany. [101]Research Unit of Molecular Epidemiology, Helmholtz Zentrum München–German Research Center for Environmental Health, Neuherberg, Germany. [102]Research Computing Division, Research Triangle Institute International, Research Triangle Park, North Carolina, USA. [103]Alzheimer Scotland Dementia Research Centre, University of Edinburgh, Edinburgh, UK. [104]Department of Genetics, University Medical Center Groningen, University of Groningen, Groningen, the Netherlands. [105]Department of Human Genetics, University of Pittsburgh, Pittsburgh, Pennsylvania, USA. [106]Department of Epidemiology, Mailman School of Public Health, Columbia University, New York, New York, USA. [107]Department of Environmental Health, National Institute for Health and Welfare (THL), Kuopio, Finland. [108]Public Health and Clinical Nutrition, University of Eastern Finland, Kuopio, Finland. [109]Sticht Center on Aging, Wake Forest School of Medicine, Winston-Salem, North Carolina, USA. [110]Behavioral Health Epidemiology Program,

Research Triangle Institute International, Research Triangle Park, North Carolina, USA. [111]Florida International University, Miami, Florida, USA. [112]Department of Public Health, Medical School, University of Split, Split, Croatia. [113]Department of Medical Biology, Medical School, University of Split, Split, Croatia. [114]Division of Pulmonary, Critical Care and Sleep Medicine, Department of Medicine, University of Mississippi Medical Center, Jackson, Mississippi, USA. [115]Comprehensive Pneumology Center (CPC), Helmholtz Zentrum München (HMGU), Munich, Germany. [116]Department of Internal Medicine, Kangwon National University Hospital, School of Medicine, Kangwon National University, Chuncheon, South Korea. [117]Environmental Health Center, Kangwon National University Hospital, School of Medicine, Kangwon National University, Chuncheon, South Korea. [118]A collaboration between the University Medical Schools and National Health Service (NHS) in Aberdeen, Dundee, Edinburgh and Glasgow, UK. [119]Medical Research Institute, University of Dundee, Dundee, UK. [120]Division of Therapeutics and Molecular Medicine, University of Nottingham, Nottingham, UK. [121]Institute of Epidemiology II, Helmholtz Zentrum München–German Research Center for Environmental Health, Neuherberg, Germany. [122]Department of Epidemiology, Center for Aging and Population Health, University of Pittsburgh, Pittsburgh, Pennsylvania, USA. [123]Institute of Health Sciences, University of Oulu, Oulu, Finland. [124]Biocenter Oulu, University of Oulu, Oulu, Finland. [125]Unit of Primary Care, Oulu University Hospital, Oulu, Finland. [126]Department of Children and Young People and Families, National Institute for Health and Welfare, Oulu, Finland. [127]Department of Medical Sciences, Uppsala University, Uppsala, Sweden. [128]Department for Genetics and Functional Genomics, Interfaculty Institute for Genetics and Functional Genomics, University Medicine Greifswald, Greifswald, Germany. [129]Institute of Environmental Medicine, Karolinska Institutet and Sachs' Children's Hospital, Stockholm, Sweden. [130]Complex Disease and Genetic Epidemiology Branch, Department of Epidemiology, Seoul National University School of Public Health, Seoul, South Korea. [131]Department of Pulmonology, Leiden University Medical Center, Leiden, the Netherlands. [132]Department of Public Health, Division of Biostatistics and Epidemiology, Weill Cornell Medical College, New York, New York, USA. [133]University of New Mexico Health Sciences Center School of Medicine, Albuquerque, New Mexico, USA. [134]Department of Respiratory Medicine, Erasmus MC, Rotterdam, the Netherlands. [135]These authors contributed equally to this work. [136]These authors jointly directed this work. Correspondence should be addressed to S.J.L. (london2@niehs.nih.gov) or M.D.T. (mt47@le.ac.uk).

ONLINE METHODS

Study design. The study consisted of two stages. Stage 1 was a meta-analysis of study-specific genome-wide analyses of FVC conducted in 26 studies with a total of 52,253 individuals of European ancestry. Study characteristics are shown in **Supplementary Table 1**. In stage 2, we followed up SNPs showing association with FVC ($P < 5 \times 10^{-7}$) and performed meta-analysis of β estimates and standard errors across stages 1 and 2. Stage 2 encompassed 32,917 subjects of European ancestry from 9 independent cohorts. All participants provided written informed consent. The study was approved by the institutional ethical committee of each study site and was conducted according to Declaration of Helsinki principles.

Cohorts included in stage 1. Stage 1 included a total of 26 studies, 15 from the SpiroMeta Consortium and 11 from the CHARGE Consortium: AGES, ARIC, B58C T1DGC, B58C WTCCC, CARDIA, CHS, ECRHS, EPIC (obese cases and population-based studies), the EUROSPAN studies (CROATIA-Korcula, ORCADES, CROATIA-Vis and NSPHS), FHS, FTC (incorporating the FinnTwin16 and Finnish Twin Study on Aging), Health 2000, Health ABC, HCS, KORA S3, MESA, NFBC 1966, RS-I, RS-II, RS-III, SHIP and Twins UK-I. The genotyping platforms and quality control criteria implemented by each study are described in **Supplementary Table 9**.

Cohorts included in stage 2. A total of nine studies were included in our stage 2 follow-up: BHS 1 and 2, CROATIA-Split, Generation Scotland, KORA F4, LBC1936, LifeLines, LLFS, PIVUS, and Twins UK-II and III. Study descriptions can be found in the **Supplementary Note**.

Imputation. Imputation to the HapMap CEU panel (Utah residents of Northern and Western European ancestry) was conducted using either MACH[47], IMPUTE[48], Beagle[49,50] or BIMBAM[51] with filters and quality control parameters as shown in **Supplementary Table 9**. SNPs were excluded for a cohort if the imputation score, assessed using r2.hat (MACH), .info (IMPUTE) or OEvar (BIMBAM), was <0.3. In total, 2,762,059 SNPs were analyzed.

Statistical analysis. Individual studies performed a GWAS analysis using linear regression models with FVC (in ml) as the outcome, stratified by never-/ever-smoking. Adjustment factors were age, age^2, sex and height (plus $height^2$ and weight for CHARGE and replication cohorts). If applicable, cohorts adjusted for center, cohort or principal components to adjust for population stratification. The follow-up studies used the same models. Effect estimates for each study were corrected using genomic control[19] separately within smoking strata. Study-specific λ estimates are shown in **Supplementary Table 1**.

Meta-analysis of stage 1 data. Variants with imputation quality below 0.3 or MAF below 0.03 were excluded from each data set before the meta-analysis. Study-specific effect estimates and standard errors for ever-smokers and never-smokers were combined using METAL[52] with fixed-effects inverse variance–weighted meta-analysis, which takes directionality into account by aligning study results according to the same effect allele. Genomic control was applied to the resulting combined (ever-smokers and never-smokers) effect estimates for each study. These combined effect estimates for each study were then combined across studies, again using fixed-effects inverse variance–weighted meta-analysis with METAL, and genomic control was applied again to the final meta-analysis estimates. Manhattan plots, quantile-quantile plots, forest plots, gene annotation and additional statistics were produced using R version 2.9.2 (ref. 53). Stage 1 results are in **Supplementary Data Set 1**.

Selection of SNPs for stage 2 and stage 2 meta-analysis. For every region containing at least one SNP showing evidence of association with FVC ($P < 5 \times 10^{-7}$), the SNP with the smallest P value that also had N effective ≥ 80% (N effective is the product of sample size and imputation quality summed across studies) of the total stage 1 sample size was followed up in a second stage using *in silico* data from nine cohorts (**Fig. 1** and **Supplementary Table 2a**). In total, seven SNPs were followed up. Results for variants with imputation quality (**Supplementary Table 10**) below 0.3 in a given study were excluded from the meta-analysis. Meta-analysis was performed on results across stage 2 studies using fixed-effects inverse variance–weighted meta-analysis with METAL[52].

Regions were defined as independent if the leading SNP from one region was >500 kb from the leading SNP of any other region. We excluded two regions (*GSTCD* and *PTCH1*) from follow-up that were previously associated with FEV_1 or FEV_1/FVC[12–14].

Combined analysis of stage 1 and stage 2. We performed an inverse variance–weighted fixed-effects meta-analysis across stages 1 and 2 using METAL[52] and obtained two-sided P values for the resulting effect estimates.

Follow-up in other ancestry groups. To evaluate these loci across ancestry groups, we studied association with FVC in four samples of non-European ancestry. We used data from the NHLBI-sponsored CARe Project[23,54], which genotyped African Americans in ARIC, MESA, CHS, CARDIA, the Jackson Heart Study and the Cleveland Family Studies.

Analyses in CARe were carried out by cohort, and meta-analysis was performed using METAL[52]. Furthermore, we assessed the SNPs in Hispanic and Chinese participants from the MESA Lung study. Lastly, we investigated the loci in Korean participants from the Healthy Twin[15–17] and KARE3 (ref. 18) studies. For the Korean studies, analyses were carried out by cohort, and meta-analysis was performed. Individual studies performed GWAS analysis using linear regression models with FVC (in ml) as the outcome. Adjustment factors were age, age^2, sex, height, $height^2$, ever-/never-smoking and weight. We assessed the sentinel SNPs from the HapMap CEU reference panel that were available and SNPs from the regions of the identified loci on the basis of the location of the sentinel SNPs ±200 kb. Only SNPs with MAF ≥ 0.05 were included. The estimated number of independent tests per population sample and corresponding Bonferroni-corrected P values are shown in **Supplementary Table 11**. The effective number of independent variants being tested in each replication population was estimated on the basis of linkage disequilibrium between SNPs using the technique of Li and Ji[22], which is a modification of the technique originally proposed by Cheverud[55] and implemented by Nyholt[21]. This calculation was performed using matSp-D[21] on the basis of the linkage disequilibrium structures of the 1000 Genomes Project all-ancestries sample.

Gene set enrichment analysis. We applied an algorithm known as improved gene set enrichment analysis for GWAS (*i*-GSEA4GWAS) to place variants associated with FVC within curated pathways and functional categories[25]. SNPs from the stage 1 and 2 GWAS meta-analysis were mapped to genes if they were within 100 kb (upstream or downstream). For a given SNP, if multiple genes were located within this range, the closest gene was selected and assigned the association P value. Because multiple SNPs can map to the same gene, SNP label permutation was used to reduce biases caused by larger loci having disproportionately higher number of SNPs. Log-transformed association P values were used to rank order the resulting gene list (18,454 genes) and to calculate gene set enrichment scores. Approximately 2,000 gene sets were used. These were limited to curated pathways derived from multiple resources such as KEGG, BioCarta, REACTOME and functional annotations extracted from the Gene Ontology database. A modified version of the GSEA procedure was performed, and P values were adjusted for multiple testing using FDR. Significant enrichment of gene sets was set at FDR < 0.01.

Gene expression analysis. After the meta-analysis, we investigated the mRNA expression of the nearest gene for each of the new SNPs in: human lung tissue, HASM cells, HEBCs and PBMCs. Lung resection specimens were obtained from individuals diagnosed with solitary pulmonary tumors at Ghent University Hospital (Ghent, Belgium). Primary human bronchial epithelial cells (HBECs) and human airway smooth muscle cells (HASM cells) were prepared from lung resection specimens obtained from anonymous donors during surgery for lung cancer at the Leiden University Medical Center (LUMC, Leiden, the Netherlands). All assays were carried out at the Ghent University Hospital. PBMCs were isolated from whole blood using Ficoll gradients. Written informed consent was obtained from all subjects according to protocols approved by the local ethics committees. Total RNA was extracted from samples using the miRNeasy Mini kit (Qiagen), and cDNA was prepared from 1 µg of RNA template using the Transcriptor Universal cDNA Master kit

(Roche) following the manufacturer's instructions. Expression of the candidate genes and the housekeeping gene *GADPH* was analyzed using TaqMan Gene Expression Assays (Applied Biosystems; assay ID numbers are given in **Supplementary Table 12**).

Expression quantitative trait loci. We assessed whether top SNPs or their proxies, identified on the basis of $r^2 > 0.7$, in the 6 new regions were associated with gene expression in whole-blood cells in a sample of 762 individuals from Rotterdam Study III (RS-III)[56]. Expression was assessed using Illumina Whole-Genome Expression BeadChips (HumanHT-12 v4). For eQTL analysis, we used the eQTL mapping pipeline called MegaQTL[57]. eQTLs were deemed *cis* when the distance between the SNP chromosomal position and the probe midpoint was less than 250 kb. eQTLs were mapped using Spearman's rank correlation, using imputation dosage values as genotypes. Resultant correlations were then converted to *P* values, and their respective *z* scores were weighted by the square root of sample size. The model was adjusted for the first 40 eigenvectors of principal-component analysis. We corrected for multiple testing by using Bonferroni correction: associations with $P < 1 \times 10^{-4}$ were considered statistically significant.

Finally, we also queried publicly available eQTL databases derived from multiple cell and tissue types (lymphoblastoid cell lines, brain tissue and human fibroblasts)[26–29]. We corrected for multiple testing by using Bonferroni correction: associations with $P < 1 \times 10^{-4}$ were considered statistically significant.

Expression in fetal lung. We used methods previously described[58] to mine publicly available data[59,60] to determine whether differential expression of relevant genes occurs during normal human lung development. Previously, human fetal lung tissues were obtained from National Institute of Child Health and Human Development tissue databases and microarray profiles used to investigate expression spanning different gestational ages. RNA samples from 38 subjects (estimated gestational age of 7–22 weeks or 53–154 d postconception) representing the pseudoglandular (gestational age, 7–16 weeks) and canalicular (17–26 weeks) stages of lung development were included in the data set. These data are available at the NCBI Gene Expression Omnibus.

47. Li, Y., Willer, C.J., Ding, J., Scheet, P. & Abecasis, G.R. MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genet. Epidemiol.* **34**, 816–834 (2010).
48. Marchini, J., Howie, B., Myers, S., McVean, G. & Donnelly, P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat. Genet.* **39**, 906–913 (2007).
49. Browning, B.L. & Browning, S.R. A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. *Am. J. Hum. Genet.* **84**, 210–223 (2009).
50. Browning, S.R. & Browning, B.L. High-resolution detection of identity by descent in unrelated individuals. *Am. J. Hum. Genet.* **86**, 526–539 (2010).
51. Servin, B. & Stephens, M. Imputation-based analysis of association studies: candidate regions and quantitative traits. *PLoS Genet.* **3**, e114 (2007).
52. Willer, C.J., Li, Y. & Abecasis, G.R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).
53. R Development Core Team. *R: A Language and Environement for Statistical Computing* (R Foundation for Statistical Computing, Vienna, 2009).
54. Lettre, G. *et al.* Genome-wide association study of coronary heart disease and its risk factors in 8,090 African Americans: the NHLBI CARe Project. *PLoS Genet.* **7**, e1001300 (2011).
55. Cheverud, J.M. A simple correction for multiple comparisons in interval mapping genome scans. *Heredity (Edinb.)* **87**, 52–58 (2001).
56. Hofman, A. *et al.* The Rotterdam Study: 2012 objectives and design update. *Eur. J. Epidemiol.* **26**, 657–686 (2011).
57. Westra, H.J. *et al.* Systematic identification of *trans* eQTLs as putative drivers of known disease associations. *Nat. Genet.* **45**, 1238–1243 (2013).
58. Hodge, E. *et al.* HTR4 gene structure and altered expression in the developing lung. *Respir. Res.* **14**, 77 (2013).
59. Melén, E. *et al.* Expression analysis of asthma candidate genes during human and murine lung development. *Respir. Res.* **12**, 86 (2011).
60. Kho, A.T. *et al.* Transcriptomic analysis of human lung development. *Am. J. Respir. Crit. Care Med.* **181**, 54–63 (2010).