

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*, *WVVOX* 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 airways¹. 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, age², sex, height, height² 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 control¹⁹ 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

A full list of authors and affiliations appears at the end of the paper.

Received 15 December 2013; accepted 22 May 2014; published online 15 June 2014; doi:10.1038/ng.3011

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.

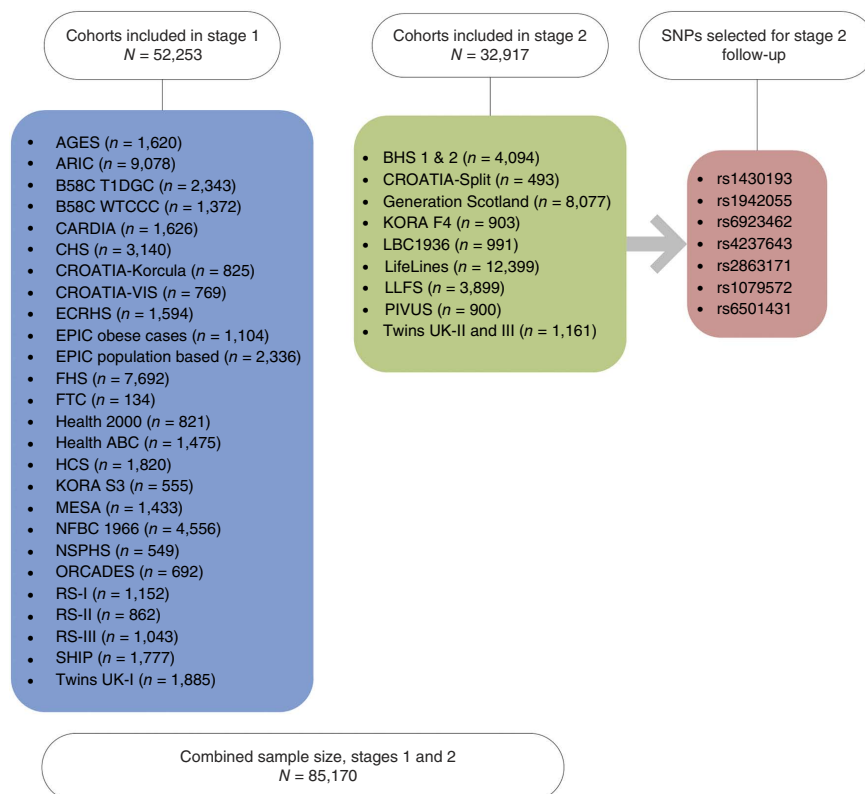
cells (PBMCs) (**Supplementary Note**). We assessed whether these SNPs were associated with gene expression in whole-blood cells (**Supplementary Table 3** and **Supplementary Note**) and queried databases to determine whether these SNPs were located within known or predicted regulatory regions.

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 (FEV₁ and FEV₁/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 panel²⁰. 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 Project²³. 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

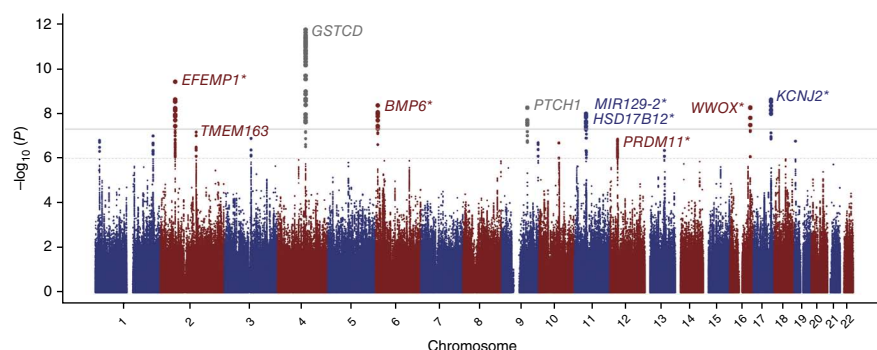


Figure 2 Manhattan plot for the association results for FVC. The plot shows all the loci analyzed in stage 1, where the two loci previously associated with either FEV₁ or FEV₁/FVC are indicated in gray. Newly associated loci were in or near the indicated adjacent genes. The loci reaching genome-wide 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	<i>P</i> value	Coded allele freq.	<i>N</i> effective
rs1430193	2	55,974,357	<i>EFEMP1</i> (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	<i>TMEM163</i> (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	<i>BMP6</i> (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	<i>HSD17B12</i> (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	<i>PRDM11</i> (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	<i>WWOX</i> (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	<i>KCNJ2</i> (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 populations²⁴, 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 analysis²⁵ 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	<i>EFEMP1</i>	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	<i>BMP6</i>	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	<i>MIR129-2</i> (downstream)– <i>HSD17B12</i>	This gene encodes a key 17 β -hydroxysteroid dehydrogenase (17 β -HSD).	+	+	+	+
rs2863171 (downstream)	11	<i>PRDM11</i>	This gene encodes PR domain-containing protein 11.	+	+	+	+
rs1079572 (intronic)	16	<i>WWOX</i>	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	<i>KCNJ2</i>	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 publicly 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 RegulomeDB³⁰ 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 (FEV₁ or FEV₁/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 population³¹. 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 FEV₁/FVC (*PTCH1*) or FEV₁ (*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 traits³². Unexplained heritability has become a well-known phenomenon in genetic epidemiology³³, 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 FEV₁, FEV₁/FVC or both^{11–14}. Although FEV₁ and FVC are statistically correlated ($r = 0.83$ in the Rotterdam Study, adjusted for age, sex, height and height²), these measures represent clinically different entities. FVC is used for the evaluation of restrictive ventilatory defects and is a predictor of mortality independent of FEV₁, standard risk factors and even previous cardiovascular disease^{5,6}. In contrast, FVC and FEV₁/FVC have a very low correlation ($r = -0.08$ in the Rotterdam Study, adjusted for age, sex, height and height²). 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 FEV₁ or FEV₁/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 FEV₁ and FVC (**Supplementary Table 8**). Among our FVC-associated SNPs, the smallest *P* value for FEV₁ (9.43×10^{-7}) was observed for rs1079572 (*WVVOX*). An intragenic SNP (rs11654749) in the region between *KCNJ2* and *SOX9* was associated with FEV₁ at genome-wide significance in our previous meta-analysis of SNP and SNP-by-smoking effects³⁴. To assess whether the variant identified in that analysis of FEV₁ (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 GCTA³⁵ 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 GWAS³⁶. The FVC-associated SNPs in or near these two genes show modest to weak correlation with the top SNPs from the GWAS of height³⁶. 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 injury³⁷. *BMP6* expression in bronchial epithelial cells has been reported to increase in experimental models of allergic airway inflammation³⁸. *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 mice³⁹. Expression of other members of the fibulin family (fibulin-1 and fibulin-5) seems to be influenced by *TGFB1* (encoding transforming growth factor- β 1), a gene previously linked to inflammation in individuals with chronic obstructive pulmonary disease (COPD)^{40,41}. *WVVOX* might influence protein-induced apoptosis and behaves as a tumor suppressor gene in various types of neoplasms, including small-cell lung cancer⁴². 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 Atlas⁴³. 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 function⁴⁴. 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/daledN/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.

ACKNOWLEDGMENTS

For all studies, information on funding and acknowledgments can be found in the [Supplementary Note](#).

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.E. 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., J.B.W., N.H., M.F.P., K.M.B., 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.E., 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, J.B. 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.

COMPETING FINANCIAL INTERESTS

The authors declare competing financial interests: details are available in the [online version of the paper](#).

Reprints and permissions information is available online at <http://www.nature.com/reprints/index.html>.

- Wilk, J.B. *et al.* Evidence for major genes influencing pulmonary function in the NHLBI family heart study. *Genet. Epidemiol.* **19**, 81–94 (2000).
- 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).
- 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).
- 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).
- 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).
- Lee, H.M., Lee, 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).
- Lange, P., Nyboe, J., Appleyard, M., Jensen, G. & Schnohr, P. Spirometric findings and mortality in never-smokers. *J. Clin. Epidemiol.* **43**, 867–873 (1990).
- 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).
- 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).
- 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).
- Wilk, J.B. *et al.* A genome-wide association study of pulmonary function measures in the Framingham Heart Study. *PLoS Genet.* **5**, e1000429 (2009).
- Repapi, E. *et al.* Genome-wide association study identifies five loci associated with lung function. *Nat. Genet.* **42**, 36–44 (2010).
- 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).
- 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).
- Sung, J. *et al.* The Korean Twin Registry—methods, current stage, and interim results. *Twin Res.* **5**, 394–400 (2002).
- Sung, J. *et al.* Healthy Twin: a twin-family study of Korea—protocols and current status. *Twin Res. Hum. Genet.* **9**, 844–848 (2006).
- Gombajav, 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 β 1 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- β 1 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 Lee¹⁶, David P Strachan¹⁷, Alan L James^{18–20}, Jennifer E Huffman²¹, Veronique Vitart²¹, Adaikalavan Ramasamy^{22,23}, Nicholas J Wareham²⁴, Jaakko Kaprio^{25–27}, Xin-Qun Wang²⁸, Holly Trochet²¹, Mika Kähönen²⁹, Claudia Flexeder³⁰, Eva Albrecht³¹, Lorna M Lopez^{32,33}, Kim de Jong^{34,35}, Bharat Thyagarajan³⁶, Alexsander Couto Alves²³, Stefan Enroth^{37,38}, Ernst Omenaas^{39,40}, Peter K Joshi⁴¹, Tove Fall^{38,42}, Ana Viñuela⁴³, Lenore J Launer⁴⁴, Laura R Loehr^{7,8}, Myriam Fornage^{45,46}, Guo Li⁴⁷, Jemma B Wilk⁴⁸, Wenbo Tang⁴⁹, Ani Manichaikul^{28,50}, Lies Lahousse^{1,51}, Tamara B Harris⁴⁴, Kari E North⁷, Alicja R Rudnicka¹⁷, Jennie Hui⁵², Xiangjun Gu^{45,46}, Thomas Lumley⁵³, Alan F Wright²¹, Nicholas D Hastie²¹, Susan Campbell²¹, Rajesh Kumar⁵⁴, Isabelle Pin^{55–57}, Robert A Scott²⁴, Kirsi H Pietiläinen^{27,58,59}, Ida Surakka^{27,60}, Yongmei Liu⁶¹, Elizabeth G Holliday^{14,15}, Holger Schulz³⁰, Joachim Heinrich^{30,62}, Gail Davies^{32,33,63,64}, Judith M Vonk^{34,35}, Mary Wojczynski⁶⁵, Anneli Pouta^{66,67}, Åsa Johansson^{37,38,68}, Sarah H Wild⁴¹, Erik Ingelsson^{38,42,69}, Fernando Rivadeneira^{70,71}, Henry Völzke⁷², Pirro G Hysi⁴³, Gudny Eiriksdottir¹¹, Alanna C Morrison⁷³, Jerome I Rotter^{74,75}, Wei Gao⁷⁶, Dirkje S Postma^{35,77}, Wendy B White⁷⁸, Stephen S Rich⁵⁰, Albert Hofman^{1,71}, Thor Aspelund^{11,12}, David Couper⁷⁹, Lewis J Smith⁵⁴, Bruce M Psaty^{6,47,80,81}, Kurt Lohman⁸², Esteban G Burchard^{83,84}, André G Uitterlinden^{1,70,71}, Melissa Garcia⁴⁴, Bonnie R Joubert⁸⁵, Wendy L McArdle⁸⁶, A Bill Musk⁸⁷, Nadia Hansel⁸⁸, Susan R Heckbert^{47,80,81}, Lina Zgaga^{89,90}, Joyce B J van Meurs^{70,71}, Pau Navarro²¹, Igor Rudan⁴¹, Yeon-Mok Oh^{91,92}, Susan Redline⁹³, Deborah L Jarvis^{22,94}, Jing Hua Zhao²⁴, Taina Rantanen⁹⁵, George T O'Connor^{96,97}, Samuli Ripatti^{27,60,98}, Rodney J Scott^{14,15}, Stefan Karrasch^{30,99,100}, Harald Grallert¹⁰¹, Nathan C Gaddis¹⁰², John M Starr^{32,103}, Cisca Wijmenga¹⁰⁴, Ryan L Minster¹⁰⁵, David J Lederer^{10,106}, Juha Pekkanen^{107,108}, Ulf Gyllenstein^{37,38}, Harry Campbell⁴¹, Andrew P Morris⁶⁹, Sven Gläser⁹, Christopher J Hammond⁴³, Kristin M Burkart¹⁰, John Beilby⁵², Stephen B Kritchevsky¹⁰⁹, Vilmundur Gudnason^{11,12}, Dana B Hancock^{85,110}, O Dale Williams¹¹¹, Ozren Polasek¹¹², Tatijana Zemunik¹¹³, Ivana Kolcic¹¹², Marcy F Petrini¹¹⁴, Matthias Wjst¹¹⁵, Woo Jin Kim^{116,117}, David J Porteous⁶³, Generation Scotland¹¹⁸, Blair H Smith¹¹⁹, Anne Viljanen⁹⁵, Markku Heliövaara²⁶, John R Attia^{14,15}, Ian Sayers¹²⁰, Regina Hampel¹²¹, Christian Gieger³¹, Ian J Deary^{32,33}, H Marike Boezen^{34,35}, Anne Newman¹²², Marjo-Riitta Jarvelin^{23,123–126}, James F Wilson⁴¹, Lars Lind¹²⁷, Bruno H Stricker^{1,2,70,71}, Alexander Teumer¹²⁸, Timothy D Spector⁴³,

Erik Melén¹²⁹, Marjolein J Peters^{70,71}, Leslie A Lange¹³, R Graham Barr^{10,106}, Ken R Bracke⁵¹, Fien M Verhamme⁵¹, Joonhoo Sung^{16,130}, Pieter S Hiemstra¹³¹, 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}

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

Research Triangle Institute International, Research Triangle Park, North Carolina, USA. ¹¹¹Florida International University, Miami, Florida, USA. ¹¹²Department of Public Health, Medical School, University of Split, Split, Croatia. ¹¹³Department of Medical Biology, Medical School, University of Split, Split, Croatia. ¹¹⁴Division of Pulmonary, Critical Care and Sleep Medicine, Department of Medicine, University of Mississippi Medical Center, Jackson, Mississippi, USA. ¹¹⁵Comprehensive Pneumology Center (CPC), Helmholtz Zentrum München (HMGU), Munich, Germany. ¹¹⁶Department of Internal Medicine, Kangwon National University Hospital, School of Medicine, Kangwon National University, Chuncheon, South Korea. ¹¹⁷Environmental Health Center, Kangwon National University Hospital, School of Medicine, Kangwon National University, Chuncheon, South Korea. ¹¹⁸A collaboration between the University Medical Schools and National Health Service (NHS) in Aberdeen, Dundee, Edinburgh and Glasgow, UK. ¹¹⁹Medical Research Institute, University of Dundee, Dundee, UK. ¹²⁰Division of Therapeutics and Molecular Medicine, University of Nottingham, Nottingham, UK. ¹²¹Institute of Epidemiology II, Helmholtz Zentrum München—German Research Center for Environmental Health, Neuherberg, Germany. ¹²²Department of Epidemiology, Center for Aging and Population Health, University of Pittsburgh, Pittsburgh, Pennsylvania, USA. ¹²³Institute of Health Sciences, University of Oulu, Oulu, Finland. ¹²⁴Biocenter Oulu, University of Oulu, Oulu, Finland. ¹²⁵Unit of Primary Care, Oulu University Hospital, Oulu, Finland. ¹²⁶Department of Children and Young People and Families, National Institute for Health and Welfare, Oulu, Finland. ¹²⁷Department of Medical Sciences, Uppsala University, Uppsala, Sweden. ¹²⁸Department for Genetics and Functional Genomics, Interfaculty Institute for Genetics and Functional Genomics, University Medicine Greifswald, Greifswald, Germany. ¹²⁹Institute of Environmental Medicine, Karolinska Institutet and Sachs' Children's Hospital, Stockholm, Sweden. ¹³⁰Complex Disease and Genetic Epidemiology Branch, Department of Epidemiology, Seoul National University School of Public Health, Seoul, South Korea. ¹³¹Department of Pulmonology, Leiden University Medical Center, Leiden, the Netherlands. ¹³²Department of Public Health, Division of Biostatistics and Epidemiology, Weill Cornell Medical College, New York, New York, USA. ¹³³University of New Mexico Health Sciences Center School of Medicine, Albuquerque, New Mexico, USA. ¹³⁴Department of Respiratory Medicine, Erasmus MC, Rotterdam, the Netherlands. ¹³⁵These authors contributed equally to this work. ¹³⁶These 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 MACH⁴⁷, IMPUTE⁴⁸, Beagle^{49,50} or BAMBAM⁵¹ 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 (BAMBAM), 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, age², sex and height (plus height² 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 control¹⁹ 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 METAL⁵² 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 $\geq 80\%$ (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 METAL⁵².

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 FEV₁ or FEV₁/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 METAL⁵² 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 METAL⁵². 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, age², sex, height, height², 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 Ji²², which is a modification of the technique originally proposed by Cheverud⁵⁵ and implemented by Nyholt²¹. This calculation was performed using *matSpD*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 categories²⁵. 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 μ g 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 MegaQTL⁵⁷. 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 described⁵⁸ 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 post-conception) 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 Environment 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).

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

Daan W. Loth^{1,2*}, María Soler Artigas^{3,4*}, Sina A. Gharib^{5,6*}, Louise V. Wain^{3,4*}, Nora Franceschini^{7,8*}, Beate Koch^{9*}, Tess D. Pottinger^{10*}, Albert Vernon Smith^{11,12*}, Qing Duan^{13*}, Chris Oldmeadow^{14,15}, Mi Kyeong Lee¹⁶, David P Strachan¹⁷, Alan L James¹⁸⁻²⁰, Jennifer E. Huffman²¹, Veronique Vitart²¹, Adaikalavan Ramasamy^{22,23}, Nicholas J. Wareham²⁴, Jaakko Kaprio²⁵⁻²⁷, Xin-Qun Wang²⁸, Holly Trochet²¹, Mika Kähönen²⁹, Claudia Flexeder³⁰, Eva Albrecht³¹, Lorna M. Lopez^{32,33}, Kim de Jong^{34,35}, Bharat Thyagarajan³⁶, Alexessander Couto Alves²³, Stefan Enroth^{37,38}, Ernst Omenaas^{39,40}, Peter K. Joshi⁴¹, Tove Fall^{38,42}, Ana Viñuela⁴³, Lenore J. Launer⁴⁴, Laura R. Loehr^{7,8}, Myriam Fornage^{45,46}, Guo Li⁴⁷, Jemma B. Wilk⁴⁸, Wenbo Tang⁴⁹, Ani Manichaikul^{28,50}, Lies Lahousse^{1,51}, Tamara B. Harris⁴⁴, Kari E. North⁷, Alicja R. Rudnicka¹⁷, Jennie Hui⁵², Xiangjun Gu^{45,46}, Thomas Lumley⁵³, Alan F. Wright²¹, Nicholas D. Hastie²¹, Susan Campbell²¹, Rajesh Kumar⁵⁴, Isabelle Pin⁵⁵⁻⁵⁷, Robert A. Scott²⁴, Kirsi H. Pietiläinen^{27,58,59}, Ida Surakka^{27,60}, Yongmei Liu⁶¹, Elizabeth G. Holliday^{14,15}, Holger Schulz³⁰, Joachim Heinrich^{30,62}, Gail Davies^{32,33,63,64}, Judith M. Vonk^{34,35}, Mary Wojczynski⁶⁵, Anneli Pouta^{66,67}, Åsa Johansson^{37,38,68}, Sarah H. Wild⁴¹, Erik Ingelsson^{38,42,69}, Fernando Rivadeneira^{70,71}, Henry Völzke⁷², Pirro G. Hysi⁴³, Gudny Eiriksdottir¹¹, Alanna C. Morrison⁷³, Jerome I. Rotter^{74,75}, Wei Gao⁷⁶, Dirkje S. Postma^{35,77}, Wendy B. White⁷⁸, Stephen S. Rich⁵⁰, Albert Hofman^{1,71}, Thor Aspelund^{11,12}, David Couper⁷⁹, Lewis J. Smith⁵⁴, Bruce M. Psaty^{6,47,80,81}, Kurt Lohman⁸², Esteban G. Burchard^{83,84}, André G. Uitterlinden^{1,70,71}, Melissa Garcia⁴⁴, Bonnie R. Joubert⁸⁵, Wendy L McArdle⁸⁶, A. Bill Musk⁸⁷, Nadia Hansel⁸⁸, Susan R. Heckbert^{47,80,81}, Lina Zgaga^{89,90}, Joyce B.J. van Meurs^{70,71}, Pau Navarro²¹, Igor Rudan⁴¹, Yeon-Mok Oh^{91,92}, Susan Redline⁹³, Deborah L. Jarvis^{22,94}, Jing Hua Zhao²⁴, Taina Rantanen⁹⁵, George T. O'Connor^{96,97}, Samuli Ripatti^{27,60,98}, Rodney J. Scott^{14,15}, Stefan Karrasch^{30,99,100}, Harald Grallert¹⁰¹, Nathan C. Gaddis¹⁰², John M. Starr^{32,103}, Cisca Wijmenga¹⁰⁴, Ryan L Minster¹⁰⁵, David J. Lederer^{10,106}, Juha Pekkanen^{107,108}, Ulf Gyllenstein^{37,38}, Harry Campbell⁴¹, Andrew P. Morris⁶⁹, Sven Gläser⁹, Christopher J. Hammond⁴³, Kristin M. Burkart¹⁰, John Beilby⁵², Stephen B. Kritchevsky¹¹⁰, Vilmondur Gudnason^{11,12}, Dana B. Hancock^{85,111}, O. Dale Williams¹¹², Ozren Polasek¹¹³, Tatijana Zemunik¹¹⁴, Ivana Kolcic¹¹³, Marcy F. Petrini¹¹⁵, Matthias Wjst¹¹⁶, Woo Jin Kim^{117,118}, David J. Porteous⁶³, Generation Scotland¹¹⁹, Blair H. Smith¹⁰⁹, Anne Viljanen⁹⁵, Markku Heliövaara²⁶, John R. Attia^{14,15}, Ian Sayers¹²⁰, Regina Hampel¹²¹, Christian Gieger³¹, Ian J. Deary^{32,33}, H. Marike Boezen^{34,35}, Anne Newman¹²², Marjo-Riitta Jarvelin^{23,123-126}, James F. Wilson⁴¹, Lars Lind¹²⁷, Bruno H. Stricker^{1,2,70,71}, Alexander Teumer¹²⁸, Timothy D. Spector⁴³, Erik Melén¹²⁹, Marjolein J. Peters^{70,71}, Leslie A. Lange¹³, R. Graham Barr^{10,106}, Ken R. Bracke⁵¹, Fien M. Verhamme⁵¹, Joohon Sung^{16,130}, Pieter S. Hiemstra¹³¹, Patricia A. Cassano^{49,132}, Akshay Sood^{133**}, Caroline Hayward^{21**}, Josée Dupuis^{76,97**}, Ian P. Hall^{120**}, Guy G. Brusselle^{1,51,134**}, Martin D. Tobin^{3,4**}, Stephanie J. London^{85**}

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, WA, USA.
6. Department of Medicine, University of Washington, Seattle, WA, USA.
7. Department of Epidemiology, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA.
8. Carolina Center for Genome Sciences University of North Carolina at Chapel Hill, Chapel Hill, NC, 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, NY, USA.
11. Iceland Heart Association, Kopavogur, Iceland.
12. University of Iceland, Reykjavik, Iceland.
13. Department of Genetics, University of North Carolina, Chapel Hill, NC, USA.
14. Hunter Medical Research Institute, University of Newcastle, Newcastle, NSW, Australia.
15. Faculty of Health, University of Newcastle, Newcastle, NSW, 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, Australia.
19. School of Medicine and Pharmacology, The University of Western Australia, Perth, Australia.
20. Busselton Population Medical Research Institute, Busselton, Australia.

21. MRC Human Genetics Unit, MRC Institute of Genetics and Molecular Medicine (IGMM), University of Edinburgh, Edinburgh, Scotland, 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, United Kingdom.
25. Hjelt Institute, Department of Public Health, University of Helsinki, Helsinki, Finland.
26. The 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, The University of Edinburgh, Edinburgh, UK.
34. University of Groningen, University Medical Center Groningen, Department of Epidemiology, 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, MN, 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 Clinical Sciences, University of Bergen, Bergen, Norway.
41. Centre for Population Health Sciences, Medical School, University of Edinburgh, Edinburgh, Scotland, 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, National Institutes of Health, Bethesda, MD, USA.
45. Institute of Molecular Medicine, University of Texas Health Science Center at Houston, Houston, TX, USA.
46. Human Genetics Center, University of Texas Health Science Center at Houston, Houston, TX, USA.
47. Cardiovascular Health Research Unit, University of Washington, Seattle, WA, USA.
48. Precision Medicine, Pfizer Global Research and Development, Cambridge, MA, USA.
49. Division of Nutritional Sciences, Cornell University, Ithaca, NY, 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 WA, Nedlands, Australia.
53. Department of Statistics, University of Auckland, Auckland, New Zealand.
54. Northwestern University Feinberg School of Medicine, Chicago, IL, 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. Department of Medicine, Division of Internal Medicine, Helsinki University Central Hospital, Helsinki, Finland.
60. Public Health Genomics Unit, Department of Chronic Disease Prevention, The 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, NC, 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, MO, 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, United Kingdom.
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, TX, USA.
74. Biomedical Research Institute, Harbor-UCLA Medical Center, Torrance, CA, USA.
75. Department of Pediatrics, Harbor-UCLA Medical Center, Torrance, CA, USA.
76. Department of Biostatistics, Boston University School of Public Health, Boston, MA, USA.
77. University of Groningen, University Medical Center Groningen, Department of Pulmonology, Groningen, The Netherlands.
78. Tougaloo College, Jackson, MS, USA.
79. Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA.
80. Department of Epidemiology, University of Washington, Seattle, WA, USA.
81. Group Health Research Institute, Group Health Cooperative, Seattle, WA, USA.
82. Department of Biostatistical Sciences, Division of Public Health Sciences, Wake Forest School of Medicine, Winston-Salem, NC, USA.
83. Department of Bioengineering & Therapeutic Sciences, University of California, San Francisco, CA, USA.
84. Department of Medicine, University of California, San Francisco, CA, USA.
85. Epidemiology Branch, National Institute of Environmental Health Sciences, National Institutes of Health, Department of Health and Human Services, Research Triangle Park, NC, USA.
86. School of Social and Community Medicine, University of Bristol, Bristol, United Kingdom.
87. Department of Respiratory Medicine, Sir Charles Gairdner Hospital, Nedlands, Australia.
88. Department of Medicine, Johns Hopkins University, Baltimore, MD, 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 Women's Hospital, Boston, MA, USA.
94. MRC-Public Health England Centre for Environment & 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, MA, USA.
97. The National Heart, Lung, and Blood Institute's Framingham Heart Study, Framingham, MA, USA.
98. Genetic Epidemiology Group, Wellcome Trust Sanger Institute, Hinxton, Cambridge, 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, NC, 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, PA, USA.
106. Department of Epidemiology, Mailman School of Public Health, Columbia University, New York, NY, 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. Medical Research Institute, University of Dundee, Dundee, UK.
110. Sticht Center on Aging, Wake Forest School of Medicine, Winston-Salem, NC, USA.
111. Behavioral Health Epidemiology Program, Research Triangle Institute International, Research Triangle Park, NC, USA.
112. Florida International University, Miami, FL, USA.
113. Department of Public Health, Medical School, University of Split, Split, Croatia.
114. Department of Medical Biology, Medical School, University of Split, Split, Croatia.
115. Division of Pulmonary, Critical Care and Sleep Medicine, Department of Medicine, University of Mississippi Medical Center, Jackson, MS, USA.
116. Comprehensive Pneumology Center (CPC), Helmholtz Zentrum München (HMGU), München, Germany.
117. Department of Internal Medicine, Kangwon National University Hospital, School of Medicine, Kangwon National University, Chuncheon, South Korea.
118. Environmental Health Center, Kangwon National University Hospital, School of Medicine, Kangwon National University, Chuncheon, South Korea.
119. A Collaboration between the University Medical Schools and National Health Service (NHS) in Aberdeen, Dundee, Edinburgh and Glasgow.
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, PA, 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 & 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, NY, USA.
133. University of New Mexico Health Sciences Center School of Medicine, Albuquerque, NM, USA.
134. Department of Respiratory Medicine, Erasmus MC, Rotterdam, The Netherlands.

*These authors contributed equally to this work

**These authors jointly directed this work

Table of contents

Supplementary figures

Figure 1.	Quantile-quantile plot for the association results	6
Figure 2a-g.	Regional plots for the loci associated with FVC	7
Figure 3a-g.	Forest plots for the cohort specific effects in stage 1	8
Figure 4a-g	Forest plots for the cohort specific effects in stage 2	9
Figure 5a-b	Regional plots for the loci associated with FVC in African Americans	10
Figure 6.	Expression analysis of candidate genes in lung tissue and primary cells	11
Figure 7.	RT-PCR profiling of candidate genes in lung tissue	12
Figure 8.	eQTL-association	13

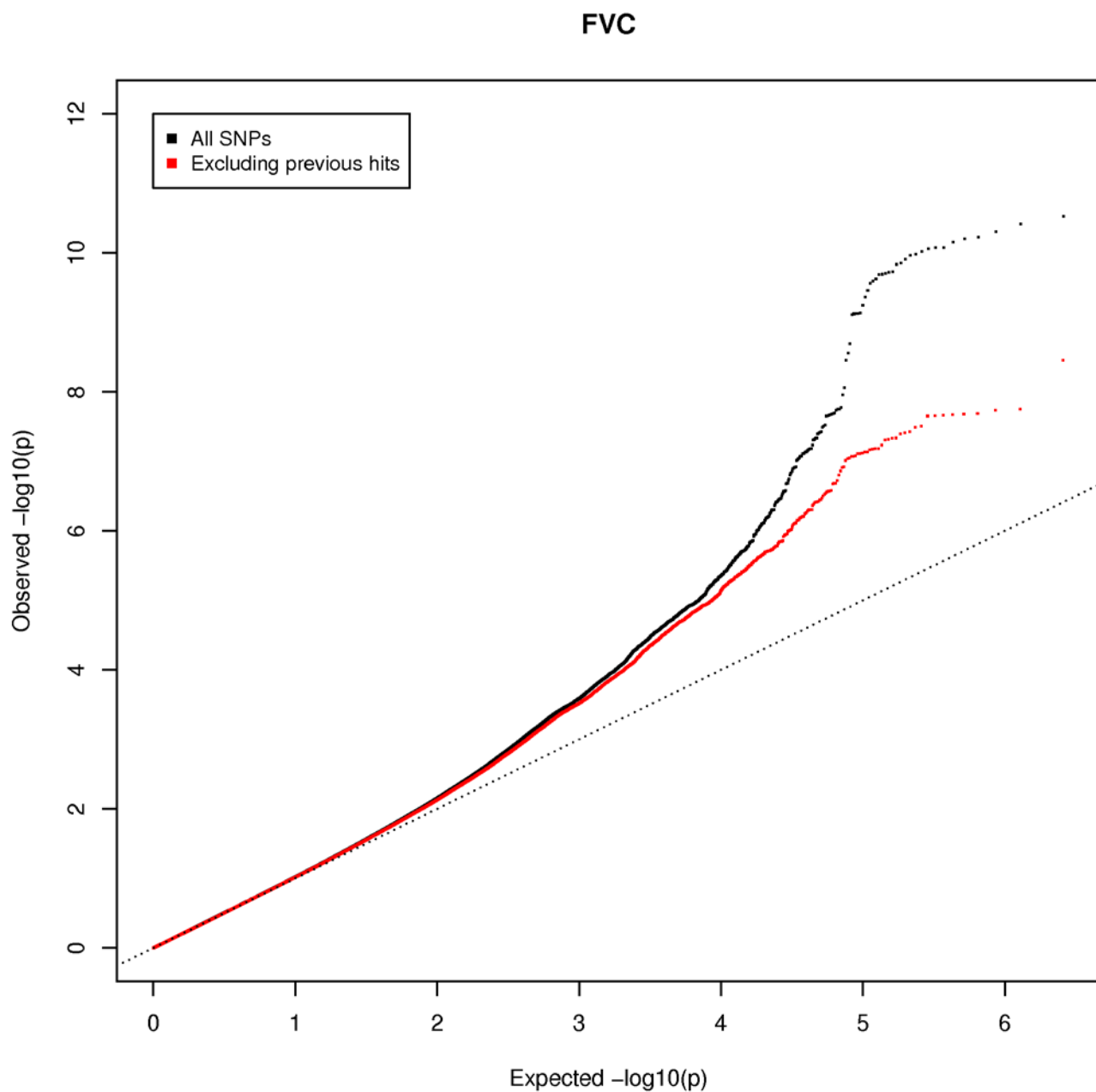
Supplementary tables

Table 1	Baseline characteristics of studies included in stage 1	14
Table 2	Baseline characteristics of studies included in stage 2 and the multi-ethnic follow-up	17
Table 3	Cis-eQTLs , <i>P</i> values for association with mRNA expression	20
Table 4	Results of top variants in ever-smokers and never-smokers separately	21
Table 5	Results for lookup in African Americans and Koreans	22
Table 6	Gene Set Enrichment Analysis, resulting pathways	27
Table 7	Fetal mRNA expression for key genes determining FVC	29
Table 8	<i>P</i> values in previously investigated spirometry phenotypes	30
Table 9	Genotyping and imputation information	31
Table 10	Imputation quality in follow-up studies for the top results	37
Table 11	Estimated number of independent tests and <i>P</i> value thresholds	38
Table 12	Gene Expression Assays for mRNA expression profiling	39

Supplementary note

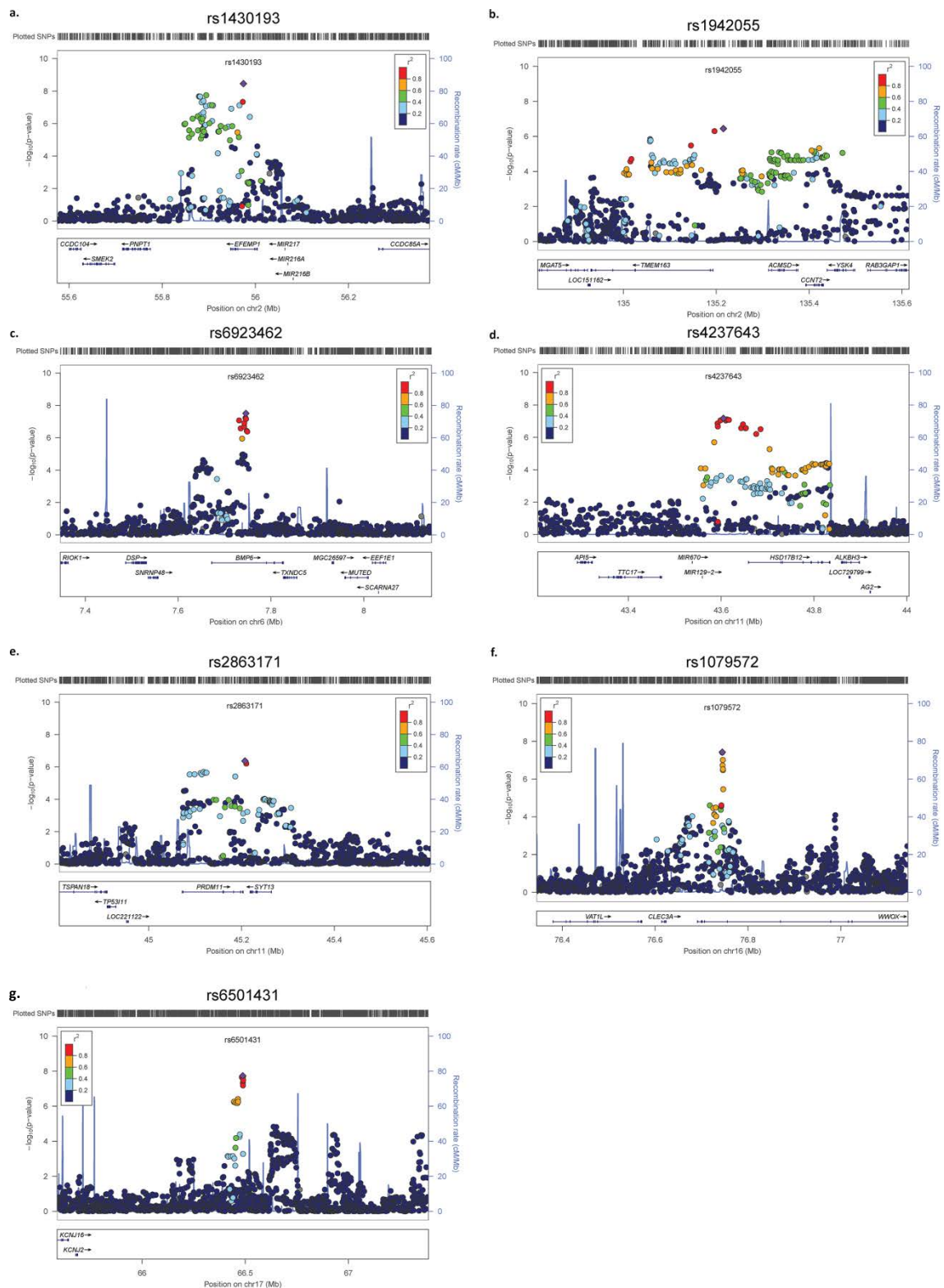
Gene expression profiling	40
Expression quantitative trait loci	40
Joint analysis for variants in the KCNJ2 locus	40
Funding and acknowledgements per study	41
Individual study descriptions	49
References	60

Supplementary Figure 1. **Quantile-Quantile (QQ) of observed genome-wide association results against expected association results for FVC.**



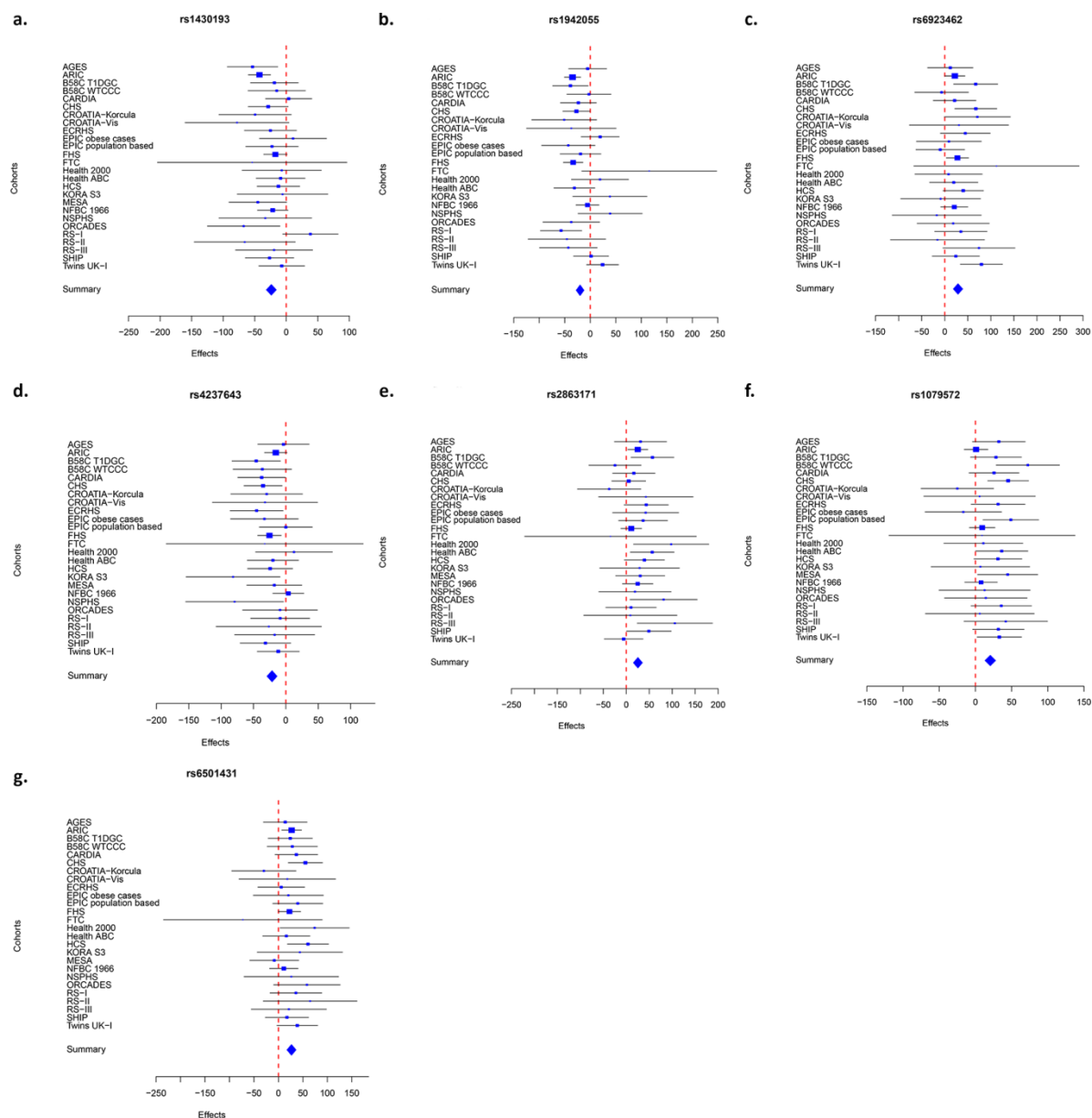
The Quantile-Quantile (QQ) plot shows $-\log_{10}(P)$ of observed genome-wide association results against expected association results for FVC. λ_{gc} before applying genomic control was 1.12. The QQ-plot for all SNPs is shown in black. The results in red show the $-\log_{10}(P)$ of observed genome-wide association results for FVC after excluding SNPs within the 26 previously reported regions¹⁻³ for FEV₁ and FEV₁/FVC; SNPs within 500kb of the lead reported SNP in the SpiroMeta-CHARGE meta-analysis of FEV₁ and FEV₁/FVC³ were excluded.

Supplementary Figure 2. **Regional Plots.**



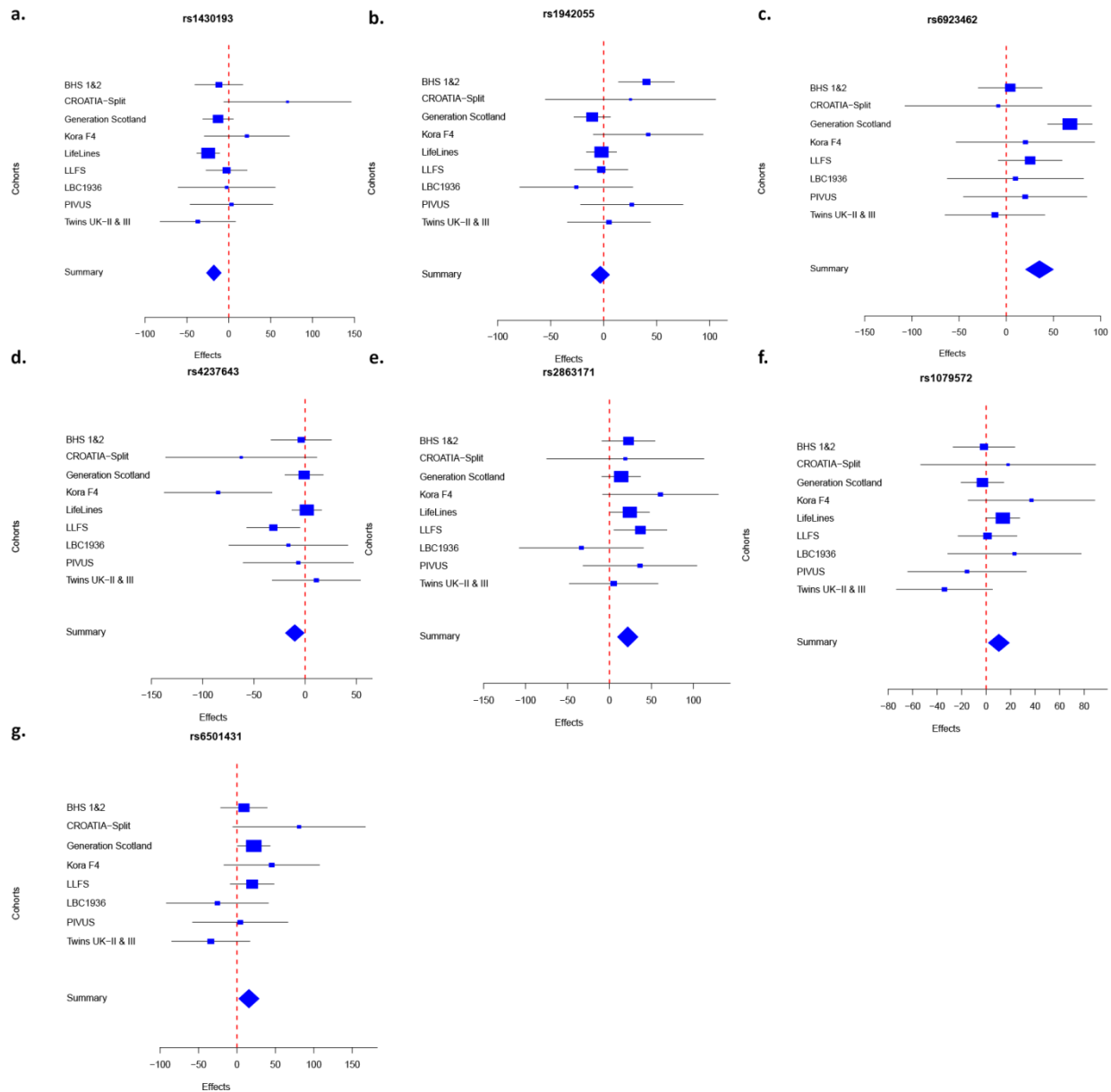
Regional association plots of seven FVC- associated loci with a P value $< 5 \times 10^{-7}$. Statistical significance of each SNP on the $-\log_{10}(P)$ scale as a function of chromosome position (NCBI Build 36) in the meta-analysis of stage 1. The sentinel SNP at each locus is shown with a purple diamond with the correlations (r^2) of surrounding SNPs to the sentinel indicated by color (red: $r^2 > 0.8$, orange: $r^2 > 0.6$, green: $r^2 > 0.4$, light blue: $r^2 > 0.2$, purple: r^2 unknown). The fine scale recombination rate is shown in blue.

Supplementary Figure 3. **Forest plots for the cohort specific effects in Stage 1.**



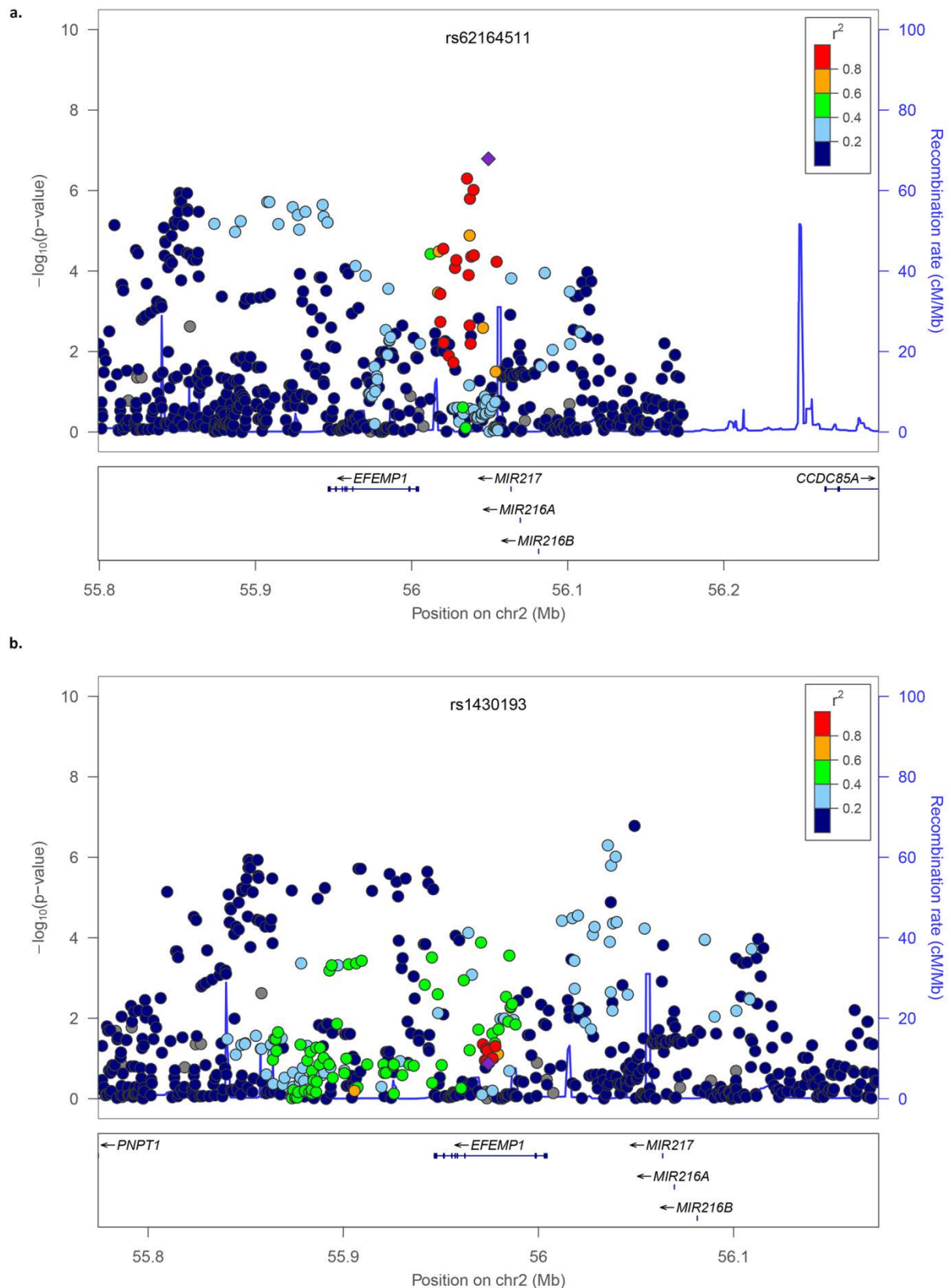
Forest plots for the seven loci associated with FVC for stage 1. Six of the SNPs included in the figure showed a genome-wide significant association ($P < 5 \times 10^{-8}$) with FVC after combining both stages. The contributing effect (β , in ml) from each study is shown by a square, with confidence intervals indicated by horizontal lines. The contributing weight of each study to the meta-analysis is indicated by the size of the square, where the size per study is relative to the total contribution of the stage. The combined meta-analysis per stage is shown at the bottom of each graph.

Supplementary Figure 4 Forest plots for the cohort specific effects in Stage 2.



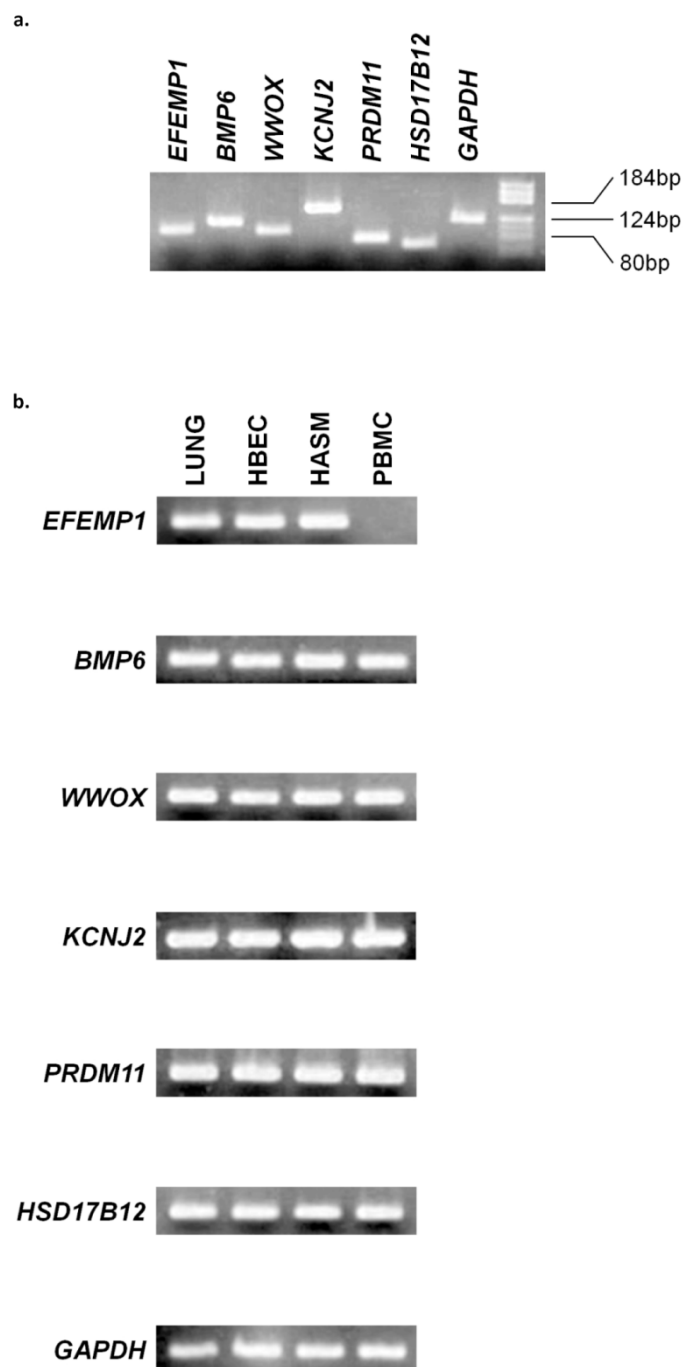
Forest plots for the seven loci associated with FVC for stage 2. Six of the SNPs included in the figure showed a genome-wide significant association ($P < 5 \times 10^{-8}$) with FVC after combining both stages. The contributing effect (β , in ml) from each study is shown by a square, with confidence intervals indicated by horizontal lines. The contributing weight of each study to the meta-analysis is indicated by the size of the square, where the size per study is relative to the total contribution of the stage. The combined meta-analysis per stage is shown at the bottom of each graph.

Supplementary Figure 5. **Regional plots for loci in African Americans.**



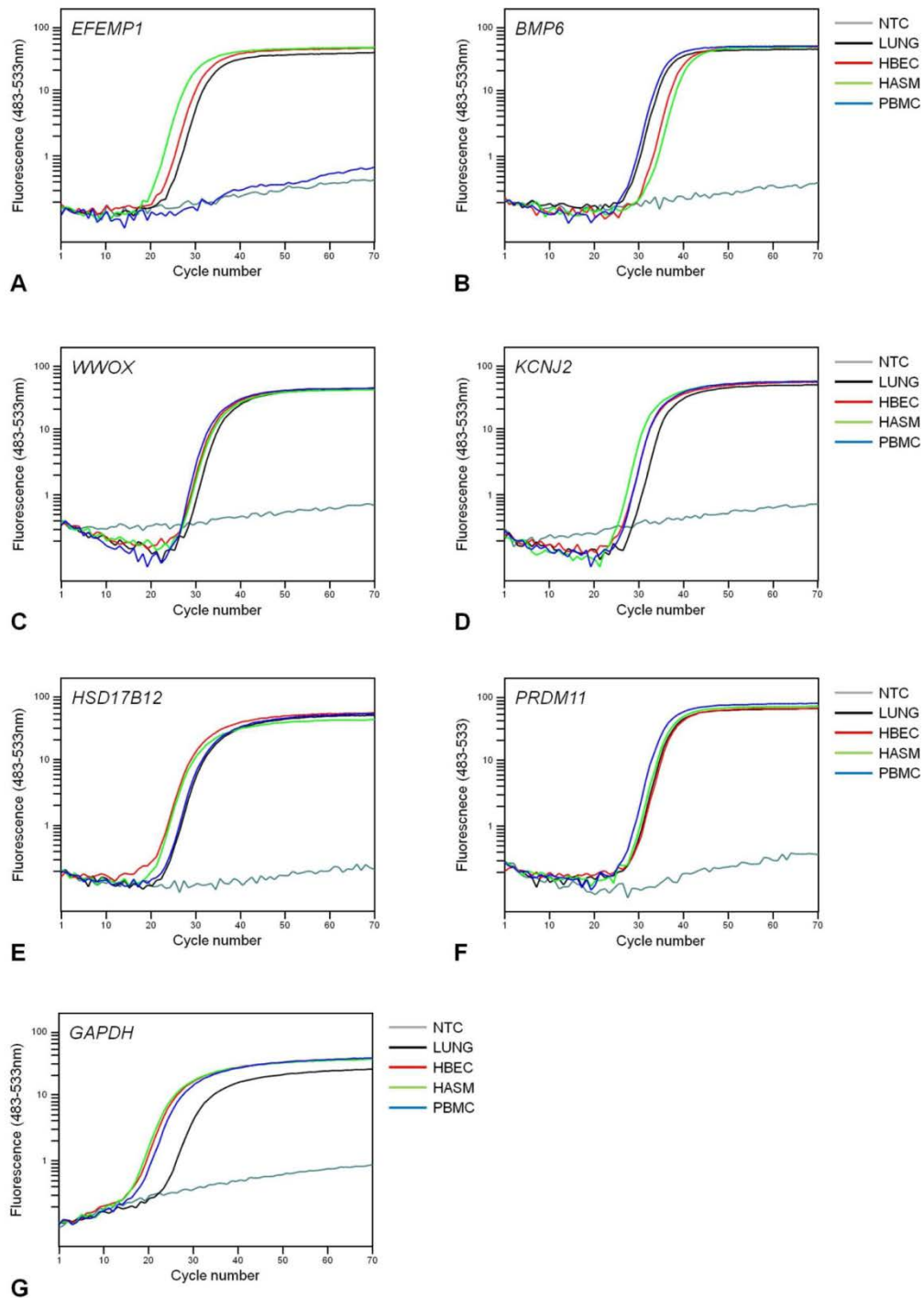
Regional plot of $-\log_{10}(P \text{ value})$ of *EFEMP1* in African American samples. The LD is based on CEU 1000G reference panel. **A.** rs62164511 is the most significant SNP in this locus in African American data. **B.** rs1430193 is the index SNP identified in European ancestry individuals. Note the low linkage disequilibrium ($r^2 = 0.16$) between these SNPs and the evidence for allelic heterogeneity at the locus.

Supplementary Figure 6. **Expression analysis of candidate genes in lung tissue and primary cells.**



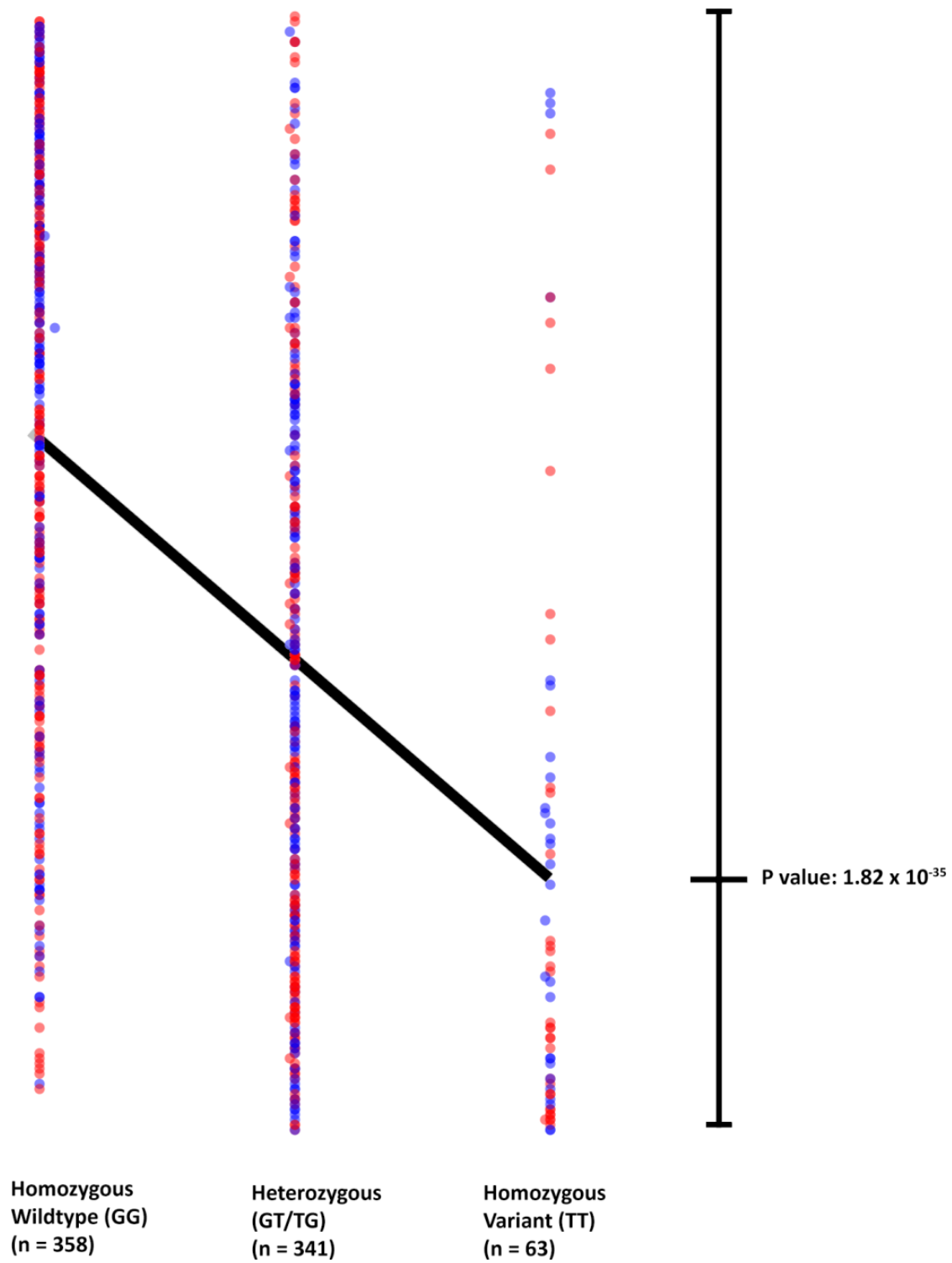
A. RT-PCR profiling of gene transcripts demonstrates expression of all seven candidate genes in total lung tissue: *EFEMP1* (86bp), *BMP6* (108bp), *WWOX* (89bp), *TMEM163* (90bp), *KCNJ2* (160bp), *PRDM11* (77bp) and *HSD17B12*(65bp). **B.** mRNA expression profiling in human primary cells: human bronchial epithelial cells (HBEC), human airway smooth muscle cells (HASM) and peripheral blood mononuclear cells (PBMC).

Supplementary Figure 7. **Real-time Reverse transcription polymerase chain reaction of candidate genes in lung tissue and primary cells.**



Taqman amplification curves of *EFEMP1* (A), *BMP6* (B), *WWOX* (C), *KCNJ2* (D), *HSD17B12* (E), *PRDM11* (F) and *GAPDH* (G) on total lung tissue, human bronchial epithelial cells (HBEC), human airway smooth muscle cells (HASM) and peripheral blood mononuclear cells (PBMC). (NTC = non template control)

Supplementary Figure 8. **Expression Quantitative Trait Loci (eQTL) Association.**



This graph shows the cis-effects for the sentinel SNP rs4237643 on probe 3420341 – *HSD17B12*, the only significant eQTL-association we were able to identify in whole blood. eQTL-analysis was performed using Illumina Whole-Genome Expression Beadchips (HumanHT-12 v4). The blue and red dots show the sex of the individual (blue for males, red for females). The black diagonal line crosses the mean gene expression level per genotype group (μ). The vertical line at the right gives the average Z-score and its P value.

Supplementary Table 1. **Baseline characteristics of the studies included in stage 1.**

	N total	N male	N female	Age range (year) at FVC measurement	Mean age, y (s.d.)	Height, cm (s.d.)	Mean FEV ₁ , l (s.d.)	Mean FVC, l (s.d.)	Mean FEV ₁ /FVC (s.d.)	N never- smokers	N ever- smokers	Genomic inflation factor (λ) FVC	
Stage 1: GWAS												Ever- smokers	Never- smokers
AGES	1,620	618	1,002	66-95	76.2 (5.6)	166 (9)	2.10 (0.68)	2.83 (0.84)	0.74 (0.11)	737	883	1.01	1.02
ARIC	9,078	4,279	4,799	44-66	54.3 (5.7)	169 (9)	2.94 (0.78)	3.99 (0.98)	0.74 (0.08)	3,620	5,458	1.03	1.01
B58C T1DGC	2,343	1,131	1,212	44-45	44.5 (0)	169 (9)	3.31 (0.78)	4.19 (0.96)	0.79 (0.08)	692	1,651	1.00	1.01
B58C WTCCC	1,372	691	681	44-45	44.5 (0)	169 (9)	2.93 (0.75)	4.18 (0.96)	0.79 (0.08)	394	978	1.01	1.00
CARDIA	1,626	768	858	17-32	25.6 (3.3)	171 (9)	3.68 (0.81)	4.70 (1.00)	0.82 (0.06)	932	694	0.97	0.99
CHS	3,140	1,226	1,914	65-95	72.3 (5.4)	160 (9)	2.12 (0.66)	3.00 (0.87)	0.71 (0.11)	1,543	1,597	1.03	1.02
CROATIA- Korcula	825	300	525	18-90	55.5 (13.5)	168 (9)	2.84 (0.81)	3.37 (0.93)	0.84 (0.09)	397	428	1.03	1.06
CROATIA-Vis	769	323	446	18-88	56.3 (15.3)	168 (10)	3.39 (1.22)	4.38 (1.43)	0.77 (0.09)	328	441	1.04	1.02
ECRHS	1,594	784	810	19-48	33.9 (7.2)	171 (9)	3.78 (0.82)	4.59 (1.03)	0.83 (0.07)	699	895	1.01	1.01
EPIC obese cases	1,104	476	628	39-76	59.1 (8.8)	166 (9)	2.35 (0.69)	2.84 (0.87)	0.82 (0.17)	489	615	1.00	1.02
EPIC population based	2,336	1,100	1,236	39-77	59.2 (9.0)	167 (9)	2.50 (0.72)	3.04 (0.90)	0.85 (0.16)	1,061	1,275	1.01	1.01
FHS	7,692	3,544	4,148	19-92	51.9	169	3.04	4.03	0.75	3,554	4,138	1.04	1.04

					(14.6)	(10)	(0.94)	(1.14)	(0.08)				
FTC	134	13	121	23–76	57.4 (19.3)	158 (6)	2.69 (0.94)	2.93 (0.61)	0.79 (0.09)	104	30	1.04	1.01
Health 2000	821	394	427	30-75	50.5 (10.9)	169 (9)	3.29 (0.90)	4.16 (1.07)	0.79 (0.07)	249	572	0.99	1.02
Health ABC	1,475	786	689	70-79	73.7 (2.8)	167 (9)	2.30 (0.70)	3.11 (0.81)	0.74 (0.08)	643	832	1.00	1.00
HCS	1,820	915	905	55-86	66.0 (7.4)	166 (9)	2.44 (0.69)	2.97 (0.83)	0.83 (0.08)	1,012	808	0.99	1.00
KORA S3	555	261	294	29–73	47.6 (9.0)	170 (9)	3.43 (0.78)	4.18 (0.99)	0.83 (0.07)	266	289	1.00	1.01
MESA	1,433	728	705	48-90	66.1 (9.7)	169 (10)	2.57 (0.76)	3.52 (0.99)	0.73 (0.09)	615	818	1.03	1.05
NFBC 1966	4,556	2,182	2,374	31–31	31.0 (0)	171 (9)	3.96 (0.79)	4.73 (0.99)	0.84 (0.06)	1,648	2,908	1.02	1.01
NSPHS	549	255	294	18-91	50.0 (19.1)	164 (10)	3.02 (0.95)	3.68 (1.12)	0.82 (0.09)	464	85	1.01	1.00
ORCADES	692	322	370	19–93	54.9 (15.3)	167 (9)	2.88 (0.84)	3.58 (0.98)	0.80 (0.09)	404	288	1.02	1.07
RS-I	1,152	502	650	72-96	79.3 (4.6)	166 (9)	2.19 (0.65)	2.90 (0.83)	0.75 (0.08)	362	790	0.99	1.00
RS-II	862	386	476	58-88	67.2 (6.3)	168 (9)	2.71 (0.78)	3.61 (1.08)	0.76 (0.09)	290	572	1.01	0.99
RS-III	1,043	464	579	46-88	56.5 (5.4)	171 (9)	3.19 (0.81)	4.06 (1.04)	0.79 (0.07)	354	689	1.00	1.00
SHIP	1,777	870	907	25–85	52.3 (13.7)	170 (9)	3.28 (0.89)	3.87 (1.03)	0.87 (0.06)	773	1,004	1.00	1.00
Twins UK-I	1,885	0	1,885	18–79	48.4 (12.2)	162 (6)	2.73 (0.56)	3.40 (0.61)	0.80 (0.08)	943	942	1.00	1.00

Sample population characteristics for each study. Characteristics are shown for studies analyzed in stage 1 (GWAS meta-analysis), stage 1 studies: AGES, Age, Gene/Environment Susceptibility; ARIC, Atherosclerosis Risk in Communities Study; B58C T1DGC, British 1958 Birth Cohort–Type 1 Diabetes Genetics Consortium; B58C WTCCC, British 1958 Birth Cohort–Wellcome Trust Case Control Consortium; CARDIA: Coronary Artery Risk Development in Young Adults; CHS, Cardiovascular Health Study; the CROATIA- Korcula study; the CROATIA-Vis study; ECRHS, the European Community Respiratory Health Survey; EPIC obese cases, European Prospective Investigation into Cancer and Nutrition, Obese Cases; EPIC population based, European Prospective Investigation into Cancer and Nutrition Cohort; FHS, Framingham Heart Study; FTC, Finnish Twin Cohort incorporating FinnTwin16 and FITSA; H2000, Finnish Health 2000 survey; Health ABC, Health, Aging, and Body Composition; HCS, Hunter Community Study; KORA S3, Cooperative Health Research in the Region of Augsburg; MESA, Multi-Ethnic Study of Atherosclerosis; NFBC1966, Northern Finland Birth Cohort of 1966; NSPHS: The Northern Swedish Population Health Study; ORCADES, Orkney Complex Disease Study; RS-I, RS-II, RS-III, Rotterdam Study; SHIP, Study of Health in Pomerania; the TwinsUK-I study.

Definition of abbreviations: N, number; cm, centimeter; s.d.; standard deviation; l, liters; FEV₁, Forced Expiratory Volume in 1 second; FVC, Forced Vital Capacity.

Supplementary Table 2. **Baseline characteristics of studies included in stage 2 and the multi-ethnic follow-up.**

Stage 2 cohorts

a. Participants of European descent										
Study	N, total	N, males (%)	N, females (%)	Age(years) Mean (SD)	Height (cm) Mean (SD)	FVC (L) Mean (SD)	Weight (kg) Mean (SD)	N, never-smokers (%)	N, ever-smokers (%)	Pack-years in Smokers Mean (SD)
BHS 1 & 2	4,094	1,795 (43.8)	2,299 (56.2)	52.9 (15.8)	168 (9)	3.84 (1.15)	74.3 (14.2)	2,124 (51.9)	1,970 (48.1)	NA*
CROATIA-Split	493	210 (42.6)	283 (57.4)	49.1 (14.6)	173 (9)	3.80 (1.06)	80.6 (16.3)	239 (48.5)	254 (51.5)	21.8 (24.0)
Generation Scotland	8,077	3,379 (41.8)	4,698 (58.2)	51.6 (13.3)	168 (10)	3.91 (1.01)	76.1 (16.3)	4,312 (53.4)	3,765 (46.6)	18.5 (20.2)
KORA F4	903	425 (47.1)	478 (52.9)	53.8 (4.5)	169 (9)	4.20 (0.97)	79.6 (16.9)	344 (38.1)	559 (61.9)	18.3 (18.6)
LBC1936	991	501 (50.6)	490 (49.4)	69.6 (0.8)	166 (9)	3.04 (0.87)	77.3 (14.5)	437 (44.1)	554 (55.9)	31.1 (27.9)
LifeLines	12,399	5,123 (41.3)	7,276 (58.7)	48.4 (11.2)	174 (9)	4.43 (1.03)	80.2 (14.9)	344 (38.1)	7,329 (59.1)	13.5 (11.5)
LLFS	3,899	1,734 (44.5)	2,165 (55.5)	68.8 (15.2)	166 (10)	3.20 (1.1)	75.3 (16.5)	2,203(56.7)	1,683(43.3)	9.1 (17.7)
PIVUS	900	483 (48.7)	417 (51.3)	70.2 (0.2)	169 (9)	3.20 (0.9)	77.3 (14.4)	438 (48,7)	462 (51,3)	11.9 (15.3)
TwinsUK-II & III	1,161	0	1,161 (100)	54.1 (17)	162 (13)	3.21 (0.6)	68.1 (6.3)	761 (65.5)	400 (34.5)	4.9 (0.6)

Multi-ethnic follow-up studies

b. African American participants from the CARE consortium										
Study	N, total	N, males (%)	N, females (%)	Age(years) Mean (SD)	Height (cm) Mean (SD)	FVC (L) Mean (SD)	Weight (kg) Mean (SD)	N, never- smokers (%)	N, ever-smokers (%)	Pack-years in Smokers Mean (SD)
ARIC	2,579	963 (37.3)	1,616 (62.7)	53.3 (5.8)	168 (9)	3.28 (0.8)	83.4 (17.3)	1,200 (46.5)	1,379 (53.5)	NA
CARDIA	696	288 (41.4)	408 (58.6)	25.9 (2.9)	169 (10)	3.90 (0.9)	74.6 (17.0)	396 (56.9)	300 (43.1)	NA
JHS	1,901	740 (38.9)	1,161 (61.1)	49.6 (11.8)	169 (9)	3.12 (0.9)	92.6 (22.4)	1,319 (69.4)	582 (30.6)	NA
MESA	894	428 (47.9)	466 (52.1)	66.3 (9.7)	169 (10)	2.93 (0.9)	85.1 (16.7)	407 (45.5)	487 (54.5)	NA
c. Chinese										
Study	N, total	N, males (%)	N, females (%)	Age(years) Mean (SD)	Height (cm) Mean (SD)	FVC (L) Mean (SD)	Weight (kg) Mean (SD)	N, never- smokers (%)	N, ever-smokers (%)	Pack-years in Smokers Mean (SD)
MESA	563	285 (50.6)	278 (49.4)	65.5 (9.6)	162 (9)	2.92 (0.82)	62.9 (11.2)	421 (74.8)	142 (25.2)	24.7 (24.8)
d. Hispanics										
Study	N, total	N, males (%)	N, females (%)	Age(years) Mean (SD)	Height (cm) Mean (SD)	FVC (L) Mean (SD)	Weight (kg) Mean (SD)	N, never- smokers (%)	N, ever-smokers (%)	Pack-years in Smokers Mean (SD)
MESA	849	398 (46.9)	451 (53.1)	64.2 (9.7)	162 (9)	3.15 (0.90)	77.2 (15.2)	456 (53.7)	393 (46.3)	20.7 (23.5)
e. Koreans										

Study	N, total	N, males (%)	N, females (%)	Age(years) Mean (SD)	Height (cm) Mean (SD)	FVC (L) Mean (SD)	Weight (kg) Mean (SD)	N, never- smokers (%)	N, ever-smokers (%)	Pack-years in Smokers Mean (SD)
Healthy Twin & KARE3 studies	8,074	3,738	4,336 (53.7)	53.4 (10.8)	161 (9)	3.55 (0.86)	62.99 (10.2)	5,064 (62.7)	3,010 (37.8)	24.6 (19.3)

* Not available

Sample population characteristics for each study. Characteristics are shown for studies analyzed in stage 2 (follow-up meta-analysis), stage 2 studies: BHS 1 & 2, Busselton Health Study; Generation Scotland; CROATIA-Split; KORA F4, Cooperative Health Research in the Region of Augsburg; LBC1936, Lothian Birth Cohort 1936; Lifelines; LLFS, Long Life Family Study; PIVUS, Prospective Investigation of the Vasculature in Uppsala Seniors; TwinsUK-II & III. Abbreviations multi-ethnic studies: CARE, Candidate gene association resource; ARIC, Atherosclerosis Risk in Communities Study; CARDIA: Coronary Artery Risk Development in Young Adults; JHS, Jackson Heart Study; MESA, Multi-Ethnic Study of Atherosclerosis; KARE3, Korea Association Resource 3. Definition of abbreviations: N, number; cm, centimeter; SD., standard deviation; L, liters; FEV₁, Forced Expiratory Volume in 1 second; FVC, Forced Vital Capacity.

Supplementary Table 3. **Cis-eQTLs** , *P* values for association with mRNA expression.

Original Gene	Original SNP or Proxy ($r^2 > 0.3$)	SNP Name	<i>P</i> Value	R^2	SNP Chr.	SNP Position	SNP Type	Probe Name	Probe Chr.	Probe Position	Allele	Direction of effect
<i>HSD17B12</i>	Proxy	rs11037676	8.42E-81	0.70	11	43827788	G/C	3420341	11	43833508	C	-
	Sentinel	rs4237643	1.82E-35	N/A	11	43604944	G/T	3420341	11	43833508	T	-
<i>PRDM11</i>	Sentinel	rs2863171	0.013	N/A	11	45207308	A/C	4810730	11	45218674	C	-
<i>KCNJ2</i>	Sentinel	rs6501431	0.029	N/A	17	66488010	T/C	5700341	17	66286834	C	+
<i>BMP6</i>	Proxy	rs11967986	0.025	N/A	6	7749203	A/C	4150731	6	7997110	C	+
	Sentinel	rs6923462	0.093	0.74	6	7746111	T/C	380328	6	7531546	C	-
<i>WWOX</i>	Proxy	rs12716848	0.061	0.71	16	76725159	G/A	6450189	16	76691257	G	+
	Sentinel	rs1079572	0.098	N/A	16	76744639	G/A	1050608	16	76870036	G	+
<i>EFEMP1</i>	Sentinel	rs1430193	0.093	N/A	2	55974357	A/T	1260040	2	56093226	T	-

This table shows the results for association with mRNA expression in whole blood from the expression Quantitative Trait Locus (eQTLs)-analysis. The lowest *P* value per gene is shown (either for a proxy or the sentinel SNP) and the results for the sentinel SNP. Abbreviations: Chr., Chromosome.

Supplementary Table 4. Results of top variants in ever-smokers and never-smokers separately.

SNP ID	Chr.	NCBI 36 Position	Nearest gene	Coded Allele	Ever-smokers			Never-smokers			P value interaction
					Beta	SE	P value	Beta	SE	P value	
rs1430193	2	55974357	<i>EFEMP1</i> (intronic)	T	-21.247	4.276	6.73E-07	-22.697	4.391	2.35E-07	0.813
rs6923462	6	7746111	<i>BMP6</i> (intronic)	T	29.94	6.214	1.45E-06	24.73	6.301	8.68E-05	0.556
rs4237643	11	43604944	<i>HSD17B12</i> (54 kb upstream)	T	-17.598	4.268	3.73E-05	-18.579	4.456	3.06E-05	0.874
rs2863171	11	45207308	<i>PRDM11</i> (3 kb downstream)	C	22.834	5.617	4.80E-05	28.044	5.712	9.11E-07	0.515
rs1079572	16	76744639	<i>WWOX</i> (intronic)	G	17.117	4.017	2.04E-05	17.626	4.193	2.62E-05	0.93
rs6501431	17	66488010	<i>KCNJ2</i> (800 kb downstream)	T	23.318	5.655	3.73E-05	22.824	5.681	5.88E-05	0.951

Results presented in this table were obtained after meta-analyzing stage 1 and stage 2 results separately in ever-smokers and never-smokers.
Abbreviations: Chr., Chromosome.

Supplementary Table 5. SNPs with significant association with FVC* from the regional lookup in African Americans and Koreans

a. Results from 6,070 African Americans, all SNPs are +/- 200 kb from rs1430193 the sentinel SNP in Europeans for the locus *EFEMP1* (Chr 2).

SNPs	Coded Allele	Freq. _{AA}	β_{AA} (ml)	S.E. _{AA}	<i>P</i> value _{AA}	Direction _{AA}	Freq. _{EUR}	β_{EUR} (ml)	S.E. _{EUR} **	<i>P</i> value _{EUR} **	Direction _{EUR}
rs62164511	a	0.902	-84.664	16.164	1.63E-07	----	NA	NA	NA	NA	NA
rs11694304	c	0.905	-82.126	16.343	5.03E-07	----	0.632	-9.882	3.804	0.009	---+-----+-----+-----+---
rs1432562	a	0.09	81.927	16.723	9.63E-07	++++	0.349	8.846	3.926	0.024	+++++-----?+- ++-++
rs934277	t	0.85	-65.024	13.365	1.14E-06	----	NA	NA	NA	NA	NA
rs1430199	a	0.85	-64.844	13.336	1.16E-06	----	0.488	-8.894	3.669	0.015	-----+---+-----+---+---+ +
rs11694790	t	0.851	-65.074	13.385	1.16E-06	----	NA	NA	NA	NA	NA
rs12614845	t	0.152	64.619	13.348	1.29E-06	++++	NA	NA	NA	NA	NA
rs17279016	t	0.092	79.209	16.498	1.58E-06	++++	0.35	9.672	3.86	0.012	+++++-----+--- ++-++
rs934276	t	0.152	63.388	13.275	1.80E-06	++++	0.513	8.845	3.667	0.016	+++++-----+---+--- ++-+-
rs11684982	a	0.152	63.331	13.264	1.80E-06	++++	NA	NA	NA	NA	NA
rs6545526	t	0.106	74.568	15.653	1.90E-06	++++	NA	NA	NA	NA	NA
rs6733730	c	0.894	-74.384	15.621	1.92E-06	----	NA	NA	NA	NA	NA
rs1367226	a	0.139	65.028	13.753	2.26E-06	++++	0.426	6.141	3.839	0.11	+-----+-----+---?+- ++-++
rs934278	a	0.894	-73.521	15.631	2.56E-06	----	0.557	-9.065	3.687	0.014	-+-----+-----+-----+---
rs1430200	t	0.157	61.212	13.085	2.90E-06	++++	0.512	8.889	3.669	0.015	+++++-----+---+---+--- ++-+-
rs9309269	a	0.815	-57.869	12.431	3.24E-06	---+	0.494	-9.522	3.718	0.01	-----+---+-----+---+--- +-+
rs9309270	t	0.156	62.21	13.38	3.33E-06	++++	NA	NA	NA	NA	NA
rs7607070	a	0.139	64.175	13.808	3.36E-06	++++	0.44	7.672	3.695	0.038	+-----+-----+---+---+---

SNPs	Coded Allele	Freq. _{AA}	$\beta_{AA}(ml)$	S.E. _{AA}	P value _{AA}	Direction _{AA}	Freq. _{EUR}	$\beta_{EUR}(ml)$	S.E. _{EUR} **	P value _{EUR} **	Direction _{EUR}
											++-++
rs11676678	a	0.85	-62.759	13.514	3.42E-06	---+	NA	NA	NA	NA	NA
rs78077363	t	0.17	-59.238	12.763	3.46E-06	----	NA	NA	NA	NA	NA
2:56125205: ATTCC	d	0.089	-79.325	17.102	3.51E-06	----	NA	NA	NA	NA	NA
rs17047277	c	0.825	58.926	12.731	3.68E-06	++++	NA	NA	NA	NA	NA
rs6545527	t	0.14	63.817	13.847	4.05E-06	++++	0.44	7.678	3.697	0.038	+-----+-----+-----+ ++-++
2:56125203: ATATT	d	0.091	-77.091	16.778	4.33E-06	----	NA	NA	NA	NA	NA
rs11125607	a	0.815	-56.703	12.349	4.39E-06	----	0.55	-7.22	3.772	0.056	-+-----++-+-----++?-+--- -
rs1346781	t	0.862	-62.635	13.808	5.73E-06	----	0.555	-7.734	3.746	0.039	-+-----+-----++?-+---
rs934274	a	0.162	59.151	13.052	5.85E-06	++++	NA	NA	NA	NA	NA
rs11693187	a	0.833	-58.377	12.889	5.91E-06	----	NA	NA	NA	NA	NA
rs1346787	t	0.861	-62.341	13.785	6.12E-06	----	0.56	-6.865	3.736	0.066	-+-----++-+-----++?-+--- -
rs75000414	t	0.141	-63.446	14.043	6.24E-06	----	NA	NA	NA	NA	NA
rs7602081	t	0.175	58.582	12.999	6.59E-06	+++-	NA	NA	NA	NA	NA
rs7609520	t	0.862	-62.957	13.985	6.74E-06	----	NA	NA	NA	NA	NA
rs934275	t	0.172	57.403	12.764	6.88E-06	++++	NA	NA	NA	NA	NA
rs4672066	a	0.149	69.992	15.593	7.17E-06	++++	0.493	-0.418	3.899	0.915	-----+-----+-----+ ---
rs4671259	a	0.208	56.028	12.488	7.24E-06	++++	NA	NA	NA	NA	NA
rs4672069	c	0.176	55.742	12.504	8.28E-06	++++	0.514	8.766	3.745	0.019	+-----+-----+-----+ ++-+-
rs6743114	a	0.184	55.137	12.437	9.28E-06	++++	0.44	7.676	3.697	0.038	+-----+-----+-----+ ++-++

SNPs	Coded Allele	Freq. _{AA}	$\beta_{AA}(ml)$	S.E. _{AA}	P value _{AA}	Direction _{AA}	Freq. _{EUR}	$\beta_{EUR}(ml)$	S.E. _{EUR} **	P value _{EUR} **	Direction _{EUR}
rs1430191	t	0.836	-56.76	12.879	1.05E-05	----	0.493	-7.522	3.663	0.04	-----+---+-----+---+---+ +
rs10496055	a	0.091	-72.862	16.558	1.08E-05	----	NA	NA	NA	NA	NA
rs75943405	a	0.095	-72.733	16.535	1.09E-05	----	NA	NA	NA	NA	NA
rs17047300	a	0.094	-72.484	16.558	1.20E-05	----	NA	NA	NA	NA	NA
rs6749890	c	0.872	-64.282	14.742	1.30E-05	----	NA	NA	NA	NA	NA
rs1961386	a	0.163	56.994	13.078	1.31E-05	++++	NA	NA	NA	NA	NA
2:56014454: C_CCT	i	0.829	-56.546	13.016	1.40E-05	---+	NA	NA	NA	NA	NA
rs74968080	a	0.904	71.102	16.499	1.64E-05	++++	NA	NA	NA	NA	NA
rs75915915	t	0.151	-57.668	13.44	1.78E-05	----	NA	NA	NA	NA	NA
2:55990766: T_TG	i	0.715	-52.216	12.173	1.79E-05	----	NA	NA	NA	NA	NA
rs11694759	t	0.164	55.595	12.964	1.80E-05	++++	0.514	8.743	3.667	0.017	+++++++---+---+---+---+ ++-+-
rs72811742	t	0.146	58.239	13.594	1.83E-05	++++	NA	NA	NA	NA	NA
rs11887339	a	0.173	54.211	12.697	1.96E-05	++++	NA	NA	NA	NA	NA
rs17047281	t	0.849	57.166	13.41	2.02E-05	++++	NA	NA	NA	NA	NA
rs10469875	c	0.185	-52.81	12.469	2.28E-05	----	NA	NA	NA	NA	NA
rs116371971	c	0.915	72.072	17.176	2.72E-05	++++	NA	NA	NA	NA	NA
rs148288741	a	0.915	71.973	17.176	2.79E-05	++++	NA	NA	NA	NA	NA
rs12617071	t	0.147	56.788	13.556	2.80E-05	++++	0.369	9.298	3.838	0.015	+++-----+-----+---+ ++-++
2:56069376: GAGA_	d	0.085	-71.822	17.177	2.90E-05	----	NA	NA	NA	NA	NA
rs1430201	a	0.085	-71.82	17.177	2.90E-05	----	NA	NA	NA	NA	NA
rs13404193	a	0.267	45.138	10.807	2.96E-05	+++-	NA	NA	NA	NA	NA
rs77784128	a	0.085	-72.392	17.341	2.98E-05	----	NA	NA	NA	NA	NA

SNPs	Coded Allele	Freq. _{AA}	$\beta_{AA}(ml)$	S.E. _{AA}	P value _{AA}	Direction _{AA}	Freq. _{EUR}	$\beta_{EUR}(ml)$	S.E. _{EUR} ^{**}	P value _{EUR} ^{**}	Direction _{EUR}
rs143157554	t	0.085	-71.691	17.179	3.00E-05	----	NA	NA	NA	NA	NA
rs143080645	t	0.915	71.528	17.172	3.11E-05	++++	NA	NA	NA	NA	NA
rs7563487	t	0.845	-55.839	13.429	3.21E-05	----	0.63	-9.549	3.782	0.012	---+-----+-----+-----+---
2:56106818: G_GTC	i	0.158	-55.319	13.357	3.45E-05	----	NA	NA	NA	NA	NA
rs11125600	c	0.17	53.775	12.995	3.50E-05	++++	NA	NA	NA	NA	NA
rs7560508	t	0.729	-44.25	10.7	3.55E-05	---+	0.414	-5.092	3.731	0.172	+-----+-----+-----+--- +-+
rs1430198	a	0.729	-44.242	10.7	3.55E-05	---+	0.414	-5.111	3.731	0.171	+-----+-----+-----+--- +-+
rs77406423	a	0.912	71.066	17.189	3.56E-05	++++	NA	NA	NA	NA	NA
rs1594308	t	0.915	70.834	17.157	3.65E-05	++++	NA	NA	NA	NA	NA
2:55988194: TC_T	d	0.79	-48.701	11.799	3.67E-05	----	NA	NA	NA	NA	NA
rs6742316	a	0.271	44.093	10.695	3.75E-05	+++-	0.586	5.152	3.731	0.167	-++++-+-+-----+-----+ ++-+-
rs2868440	t	0.899	-68.562	16.633	3.75E-05	----	0.568	-11.198	3.819	0.003	---+-----+-----+-----+---
rs11903784	t	0.085	-70.642	17.166	3.87E-05	----	NA	NA	NA	NA	NA
rs6723918	a	0.271	44.009	10.698	3.89E-05	+++-	0.586	5.223	3.731	0.162	-++++-+-+-----+-----+ ++-+-
rs1465684	a	0.144	56.033	13.649	4.04E-05	++++	0.349	8.955	3.94	0.023	+++-----+-----+-----?+- +++++
rs4233963	t	0.194	49.568	12.088	4.12E-05	+++-	0.515	8.685	3.735	0.02	++++-----+-----+-----?+- ++-+-
rs116212307	a	0.085	-70.193	17.144	4.23E-05	----	NA	NA	NA	NA	NA
rs2163714	a	0.144	55.844	13.653	4.31E-05	++++	0.349	9.66	3.862	0.012	+++-----+-----+-----+--- +++++
rs17047272	a	0.915	69.996	17.121	4.35E-05	++++	NA	NA	NA	NA	NA

b. Results from 8,074 Koreans. All SNPs are +/- 200 kb from rs6501431 (Chr 17, *KCNJ2*, 800 kb downstream)

SNPs	Coded Allele	Freq._{KOR}	β_{KOR}(ml)	S.E._{KOR}	P value_{KOR}	Direction_{KOR}	Freq._{EUR}	β_{EUR}(ml)	S.E._{EUR}**	P value_{EUR}**	Direction_{EUR}
rs12449659	t	0.862	-32.5	9.7	7.92E-04	-	0.863	-7.6	5.584	0.173	+----+---+--+---++?-- ++++
rs4793331	a	0.375	-22.5	6.9	1.22E-03	-	0.424	-3.651	3.802	0.337	-----+-----?-+- -+

* *P* value thresholds were calculated using the method of Nyholt⁴ taking into account LD between SNPs. Thresholds were 4.42×10^{-5} for African-Americans and 1.52×10^{-3} for Koreans. No SNPs were significantly associated with FVC in Hispanics or Chinese. Freq: frequency

** S.E. and *P* values presented for Europeans were obtained after applying genomic control at the meta-analysis level.

The top SNP in *EFEMP1* in European Ancestry (rs1430193) is not listed in the table because it gave a less significant P value than 4.42×10^{-5}

Supplementary Table 6. **Gene Set Enrichment Analysis, resulting pathways.**

ENRICHED GENE SET	FDR	Gene _{SNP}	Gene _{Gene Set}
Development and Tissue Remodeling			
PROTEINACEOUS EXTRACELLULAR MATRIX	7,50E-04	70	98
EXTRACELLULAR MATRIX	8,00E-04	71	100
BASEMENT MEMBRANE	8,33E-04	29	37
CELL PROLIFERATION GO 0008283	1,00E-03	275	514
TISSUE MORPHOGENESIS	1,55E-03	12	14
EXTRACELLULAR MATRIX PART	2,24E-03	43	57
MORPHOGENESIS OF AN EPITHELIUM	2,61E-03	12	16
EPHA4 PATHWAY	2,64E-03	9	10
CELL FATE COMMITMENT	2,66E-03	10	13
REGULATION OF NEUROGENESIS	2,77E-03	11	14
AXONOGENESIS	2,81E-03	32	43
DEVELOPMENTAL GROWTH	3,31E-03	8	11
ERK5 PATHWAY	3,50E-03	10	17
BASAL LAMINA	5,73E-03	16	21
TRANSMEMBRANE RECEPTOR PROTEIN TYROSINE KINASE ACTIVITY	5,75E-03	30	43
EXTRACELLULAR STRUCTURE ORGANIZATION AND BIOGENESIS	5,80E-03	18	32
EMBRYONIC MORPHOGENESIS	5,81E-03	14	17
TRANSMEMBRANE RECEPTOR PROTEIN TYROSINE KINASE SIGNALING PATHWAY	5,98E-03	53	83
REGULATION OF AXONOGENESIS	6,07E-03	7	10
NEURITE DEVELOPMENT	6,76E-03	38	53
MTA3 PATHWAY	6,76E-03	7	16
ANATOMICAL STRUCTURE MORPHOGENESIS	7,50E-03	231	379
STRUCTURAL CONSTITUENT OF MUSCLE	8,43E-03	23	33
HSA04330 NOTCH SIGNALING PATHWAY	8,84E-03	25	47
HSA04514 CELL ADHESION MOLECULES	9,71E-03	84	134
Smooth Muscle			
SMOOTH MUSCLE CONTRACTION	7,50E-04	83	150
Acetylcholine			
ACETYLCHOLINE BINDING	1,00E-04	13	17
NICOTINIC ACETYLCHOLINE ACTIVATED CATION SELECTIVE CHANNEL ACTIVITY	7,14E-04	9	11
NICOTINIC ACETYLCHOLINE GATED RECEPTOR CHANNEL COMPLEX	7,14E-04	9	11
Glutamate			
GLUTAMATE RECEPTOR ACTIVITY	7,27E-04	17	20
METABOTROPIC GLUTAMATE GABA B LIKE RECEPTOR ACTIVITY	1,07E-03	9	10

Immunity			
HSA04612 ANTIGEN PROCESSING AND PRESENTATION	7,78E-04	32	83
EOSINOPHILS PATHWAY	1,62E-03	5	8
B CELL DIFFERENTIATION	3,00E-03	8	12
FOSBPATHWAY	7,47E-03	3	5
DNA and Transcription			
POSITIVE REGULATION OF NUCLEIC ACID METABOLIC PROCESS	8,00E-04	99	154
RNA BINDING	1,00E-03	114	259
RNA SPLICING FACTOR ACTIVITY TRANSESTERIFICATION MECHANISM	1,06E-03	9	19
DNA REPAIR	1,12E-03	73	125
POSITIVE REGULATION OF RNA METABOLIC PROCESS	2,09E-03	73	120
POSITIVE REGULATION OF TRANSCRIPTION DNA DEPENDENT	2,29E-03	72	118
POSITIVE REGULATION OF TRANSCRIPTION	2,58E-03	91	144
POSITIVE REGULATION OF TRANSCRIPTION FROM RNA POLYMERASE II PROMOTER	2,67E-03	39	65
RNA PROCESSING	2,76E-03	77	174
DOUBLE STRANDED RNA BINDING	3,39E-03	13	17
RNA SPLICING	3,59E-03	43	92
DNA CATABOLIC PROCESS	5,87E-03	17	23
RESPONSE TO DNA DAMAGE STIMULUS	6,02E-03	88	162
Metabolism			
HSA04940 TYPE I DIABETES MELLITUS	1,00E-04	31	45
POSITIVE REGULATION OF CELLULAR METABOLIC PROCESS	2,13E-03	135	229
POSITIVE REGULATION OF METABOLIC PROCESS	2,56E-03	137	236
HSA00750 VITAMIN B6 METABOLISM	5,05E-03	5	5
Other			
MITOCHONDRION ORGANIZATION AND BIOGENESIS	6,67E-04	26	48
ENDOPLASMIC RETICULUM LUMEN	8,46E-04	11	14
CACAM PATHWAY	1,16E-03	12	14
EXTRINSIC PATHWAY	2,31E-03	8	13
RESPONSE TO STEROID HORMONE STIMULUS	2,37E-03	6	11
LIPID RAFT	5,77E-03	19	29
PROTEIN TARGETING TO MITOCHONDRION	5,78E-03	6	11
ER TO GOLGI VESICLE MEDIATED TRANSPORT	5,85E-03	12	18
AMINE BINDING	7,43E-03	16	23
INORGANIC ANION TRANSPORT	7,49E-03	11	18
REGIONALIZATION	7,54E-03	9	15
CELL RECOGNITION	8,42E-03	13	19
PROTEIN DOMAIN SPECIFIC BINDING	9,36E-03	43	72

This table shows the resulting pathways from the Gene Set Enrichment analysis. The threshold for the False Discovery Rate is set <0.01.

Supplementary Table 7. **Fetal mRNA expression for key genes determining FVC.**

Gene	Probe ID	AveExpr	t	P value	Adjusted P value	Beta-coefficient
<i>EFEMP1</i>	201842_s_at	9.498	1.762	0.086	0.201	0.005
	201843_s_at	7.801	2.295	0.027	0.088	0.008
<i>BMP6</i>	206176_at	5.066	0.351	0.727	0.834	7.00E-04
<i>HSD17B12</i>	1554121_at	3.418	1.667	0.104	0.229	0.002
	1554122_a_at	4.531	-0.424	0.674	0.796	-7.00E-04
	217869_at	9.951	0.513	0.611	0.75	8.00E-04
	1559518_at	3.405	-0.861	0.395	0.569	-7.00E-04
<i>PRDM11</i>	220571_at	4.436	-0.949	0.348	0.525	-0.001
	229687_s_at	7.654	-6.386	1.46E-07	6.69E-06	-0.015
	229688_at	6.001	-3.681	7.00E-04	0.006	-0.008
	233067_at	4.957	-3.459	0.001	0.009	-0.006
	239763_at	6.423	2.045	0.048	0.133	0.003
<i>WWOX</i>	210695_s_at	7.181	-0.427	0.672	0.794	-0.001
	219077_s_at	7.711	0.434	0.667	0.791	0.001
	221147_x_at	5.6	-1.075	0.289	0.464	-0.002
	223747_x_at	4.447	-0.652	0.518	0.677	-0.001
	223868_s_at	4.578	-4.045	2.00E-04	0.002	-0.008
<i>KCNJ2</i>	206765_at	7.297	2.158	0.037	0.111	0.011

Fetal lung gene array data for *EFEMP1*, *BMP6*, *HSD17B12*, *PRDM11*, *WWOX*, and *KCNJ2* expression during pseudoglandular and canalicular stages of human lung development. AveExpr= average expression between all samples, t= t-statistic describing differential expression based on linear regression, P value= Unadjusted P value, Adjusted P value=Adjusted P value controlling for false discovery rate ⁴, Beta-coefficient= mean change in gene expression per day during the studied period, 7-22 weeks of gestational age).

Supplementary Table 8. ***P* values and directions for effects in previously investigated spirometry phenotypes.**

SNP	Nearest gene	Phenotype	<i>P</i> value	Direction
rs1430193	<i>EFEMP1</i> (intronic)	FVC	6.02E-12	-
		FEV ₁	1.66E-04	-
		FEV ₁ /FVC	2.63E-03	+
rs6923462	<i>BMP6</i> (intronic)	FVC	8.55E-09	+
		FEV ₁	4.13E-05	+
		FEV ₁ /FVC	0.491	-
rs4237643	<i>HSD17B12</i> (54 kb upstream)	FVC	1.25E-08	-
		FEV ₁	3.94E-05	-
		FEV ₁ /FVC	0.218	+
rs2863171	<i>PRDM11</i> (3 kb downstream)	FVC	1.15E-09	+
		FEV ₁	5.31E-06	+
		FEV ₁ /FVC	0.825	-
rs1079572	<i>WWOX</i> (intronic)	FVC	3.60E-09	+
		FEV ₁	9.43E-07	+
		FEV ₁ /FVC	0.357	+
rs6501431	<i>KCNJ2</i> (800 kb downstream)	FVC	3.16E-08	+
		FEV ₁	1.49E-05	+
		FEV ₁ /FVC	0.162	-

The top results from the FVC analysis are compared to the *P* values for these hits from the FEV₁ and FEV₁/FVC analyses. Only the directionality of effects is given here, since the β 's in FVC reflect the continuous trait in milliliters, while the β 's in FEV₁, FEV₁/FVC were ranked residuals. *P* values for association and directions of effect on FEV₁ and FEV₁/FVC were extracted from the dataset of Soler Artigas *et al*³.

Supplementary Table 9. **Genotyping and imputation information of studies included in stage 1.**

Study name	GWAS platform	Calling algorithm	Individual call rate filter applied (before imp'n)	SNP call rate filter applied before imp'n	SNP HWE filter applied (before imp'n)	SNP MAF filter applied (before imp'n)	Other filter	No of SNPs after filtering (before imp'n)	Imp'n software and version	NCBI; HapMap CEU version for imp'n	Genotype-phenotype association software and version
AGES	Illumina Hu370CNV	BeadStudio	0.97	0.90	1E-06	0.01	remove AT/GC SNPs	208340	MACH 1.0.16	36;21a	ProbABEL 0.1
ARIC	Affymetrix 6.0	Birdseed	0.95	0.95	1E-06	0.01	Discordant genotyping, sex mismatch, suspected 1 st degree relative incl individual by genotype, genetic outlier by Plink IBS or EIGENSTRAT	669450	MACH 1.0.16	36;22	ProbABEL 0.1-3
B58C T1DGC	Illumina 550K	ILLUMINUS	0.98	No	No	No	No	520010	MACH 1.0.13	35;21	ProbABEL 0.0-5b
B58C WTCCC	Affymetrix 500K	CHIAMO	0.98	No	No	No	No	490033	IMPUTE 0.2.0	35;21	SNPTEST 1.1.3

CARDIA	Affymetrix 6.0	Beagle	>= 0.98	>= 0.95	1E-04	>=0.02	PCA for excluding outliers	578,568	Beagle-3.2	36;22	ProbABEL 0.1-2
CHS	Illumina 370 CNV	BeadStudio	0.95	0.97	1E-05	heterozygote frequency >0	reproducibility errors<2	490033	BimBam 0.99	36	R
CROATIA-Korcula	Illumina HumanHap 370cnv	BeadStudio	0.98 (for SNP of call rate >=0.98,MAF >=0.02,HW E>=E-10)	0.98	1E-06	0.01	No	307728	MACH 1.0.15	36;22	GenABEL 1.4.2 , ProbABEL
CROATIA-Vis	Illumina HumanHap 300 v1	BeadStudio	0.97 (for SNP of call rate >=0.98,MAF >=0.02,HW E>=E-10)	0.98	1E-06	0.01	No	305068	MACH 1.0.15	36;22	GenABEL 1.4.2 , ProbABEL
ECRHS	Illumina Quad 610k	GenCall	None	None	None	None	No	582892	MACH 1.0	36;22	ProbABEL 0.0-9
EPIC population based	Affymetrix 500K	BRLMM	0.94	0.90	1E-06	0.01	No	397438	IMPUTE 0.3.1	35;21	SNPTEST 1.1.5
EPIC obese cases	Affymetrix 500K	BRLMM	0.94	0.90	1E-06	0.01	No	397438	IMPUTE 0.3.1	35;21	SNPTEST 1.1.5
FHS	Affy 500K + 50K Gene	Bayesian robust	0.97	0.97	1E-06	0.01	mendelian errors>100	378163	MACH 1.0.15	36;22	R version 2.9.2

	focused	linear modeling using Mahalanobis distance (BRLMM)					0; Heterozygosity > 5SD from mean				Imekin function (kinship package)
FTC	Illumina 317K	BeadStudio	0.95	0.90	1E-05	0.01	No	315987	MACH 1.0.16	36;22	PLINK 1.06
Health 2000	Illumina 610K	ILLUMINUS	0.95	0.95	1E-06	0.01	MDS-plot outliers removed (non-European ancestry)	555388	MACH 1.0	36;22	ProbABEL
Health ABC	Illumina Human 1M-Duo	BeadStudio 3.3.7	0.97	0.95	1E-06	0.01	No sex mismatch, and cryptic relatedness	914263	MACH 1.0.16.a	36;22	R version 2.9.2
HCS	Affymetrix Axiom Kaiser array	Affymetrix proprietary algorithm - BRLMM-P	0.95	0.95	<1E-06	0.01	No	549482	MACH v1.0.16	36;22	MACH2QTL v1.10, PLINK v1.07
KORA S3	Affymetrix 500K	BRLMM	0.93	No	No	No	No	490033	MACH 1.0.9	35;21	MACH2QTL 1.0.4
MESA	Affymetrix 6.0	Birdseed v2	0.95	0.95	No	No	No	872242	IMPUTE v2.1.0	36;24	ProbABEL 0.1-9c
NFBC1966	Illumina	BeadStudio	none	0.95	1E-04	0.01	No	328007	IMPUTE	35;21	SNPTEST

	HumanCNV 370-DUO								v1.0		1.1.5
NSPHS	Illumina HumanHap 300 v2	BeadStudio	0.95	0.98	1E-06	0.01	No	292220	MACH1.0.1 6	36;22	GenABEL 1.4.2, ProbABEL 0.0-6
ORCADES	Illumina HumanHap 300 v2	BeadStudio	0.98 (for SNP of call rate >=0.98,MAF >=0.02,HW E>=E-10)	0.98	1E-06	0.01	No	306207	MACH 1.0.15	36;22	GenABEL 1.4.2, ProbABEL
RS-I	Illumina HapMap 550K	Beadstudio Genecall	0.98	0.98	1E-06	0.01	excess autosomal heterozygo sity, sex mismatch or outlying identity-by- state clustering estimates	512349	MACH 1.0.15	36;22	MACH2QTL as implemente d in GRIMP
RS-II	Illumina 550K + 610 Quad	Beadstudio Genecall	0.95	0.98	1E-06	0.01	excess autosomal heterozygo sity, sex mismatch or outlying identity-by- state	466389	MACH 1.0.16	36;22	MACH2QTL as implemente d in GRIMP

							clustering estimates				
RS-III	Illumina 610 Quad	Beadstudio Genecall	0.95	0.98	1E-06	0.01	excess autosomal heterozygosity, sex mismatch or outlying identity-by-state clustering estimates	514073	MACH 1.0.16	36;22	MACH2QTL as implemented in GRIMP
SHIP	Affymetrix 6.0	Birdseed V2	0.92	No	No	No	QC callrate > 0.86 each Chip	869224	IMPUTE 0.5.0	36;22	SNPTEST 1.1.5
Twins UK-I	Illumina 317K	BeadStudio	0.95	0.95 if MAF>0.05; <0.99 if 0.01<=MAF <0.05	5.7E-07	0.01	Unexpected relatedness based on pi_hat	296293	IMPUTE 0.5.0	36;22	GenABEL 1.4.2

Stage1 study genotyping, imputation and genotype-phenotype data. Genotyping platforms, filters applied to SNPs and individuals (if any) before imputation, imputation software and genotype-phenotype association software are given. Studies: AGES, Age, Gene/Environment Susceptibility; ARIC, Atherosclerosis Risk in Communities Study; B58C-T1DGC, British 1958 Birth Cohort–Type 1 Diabetes Genetics Consortium; B58C-WTCCC, British 1958 Birth Cohort–Wellcome Trust Case Control Consortium; CARDIA: Coronary Artery Risk Development in Young Adults; CHS, Cardiovascular Health Study; the CROATIA- Korcula study; the CROATIA-Vis study; ECRHS, the European Community Respiratory Health Survey; EPIC obese cases, European Prospective Investigation into Cancer and Nutrition, Obese Cases; EPIC population based, European Prospective Investigation into Cancer and Nutrition Cohort; FHS, Framingham Heart Study; FTC, Finnish Twin Cohort incorporating FinnTwin and FITSA; H2000, Finnish Health 2000 survey; Health ABC, Health, Aging, and Body Composition; HCS, Hunter Community Study; KORA S3, Cooperative Health Research in the Region of Augsburg; MESA, Multi-Ethnic Study of Atherosclerosis; NFBC1966, Northern Finland Birth Cohort of 1966; NSPHS: The Northern Swedish Population Health Study; ORCADES, Orkney Complex

Disease Study; RS-I, RS-II, RS-III, Rotterdam Study; SHIP, Study of Health in Pomerania; the TwinsUK-I study. Abbreviations: GWAS= Genome-Wide Association Study, imp'n=imputation, HWE= Hardy Weinberg Equilibrium, MAF= minor allele frequency. MACH and IMPUTE are two software implementations that share similar underlying population genetic models³, and BIMBAM has been shown to perform similarly to MACH and IMPUTE in contrast with other imputation methods^{5,6}.

Supplementary Table 10. **Imputation quality in follow-up studies for the top results.**

SNP ID	Chr.	NCBI 36 Position	Nearest gene	Imputation quality (genotyped or imputed)								
				BHS 1 & 2	CROATIA-Split	Generation Scotland	KORA F4	LBC1936	Lifelines	LLFS	PIVUS	TwinsUK-II & III
rs1079572	16	76744639	<i>WWOX</i>	0.997 (imp)	0.999 (imp)	0.989 (imp)	0.955 (imp)	0.991 (imp)	0.836 (imp)	0.99 (imp)	0.958 (imp)	0.989 (imp)
rs1430193	2	55974357	<i>EFEMP1</i>	0.933 (imp)	0.856 (imp)	0.984 (imp)	0.873 (imp)	0.916 (imp)	0.876 (imp)	0.998 (imp)	0.939 (imp)	0.877 (imp)
rs1942055	2	135215394	<i>TMEM163</i>	0.965 (imp)	0.959 (imp)	0.976 (imp)	1 (gen)	0.952 (imp)	0.616 (imp)	0.947 (imp)	0.951 (imp)	0.96 (imp)
rs2863171	11	45207308	<i>PRDM11</i>	1 (gen)	1 (gen)	1 (gen)	0.98 (imp)	1 (imp)	0.58 (imp)	1 (gen)	1 (gen)	1 (gen)
rs4237643	11	43604944	<i>HSD17B12</i>	0.991 (imp)	0.998 (imp)	0.999 (imp)	0.991 (imp)	0.987 (imp)	0.989 (imp)	1 (imp)	0.995 (imp)	0.995 (imp)
rs6501431	17	66488010	<i>KCNJ2</i>	0.999 (imp)	0.933 (imp)	0.999 (imp)	0.925 (imp)	0.999 (imp)	0.057 (imp)*	0.999 (imp)	0.908 (imp)	0.941 (imp)
rs6923462	6	7746111	<i>BMP6</i>	0.983 (imp)	0.901 (imp)	0.952 (imp)	1 (gen)	0.986 (imp)	0.026 (imp)*	1 (imp)	0.933 (imp)	0.966 (imp)

*Due to their low imputation quality the results for these two SNPs in Lifelines did not contribute to the meta-analysis.

Shown are imputation quality metrics and information on whether each SNP was genotyped (gen) or imputed (imp) for the seven SNPs taken forward for follow-up in the 9 studies that form stage 2. Abbreviations: Chr., Chromosome.

Supplementary Table 11. **Estimated number of independent tests and corresponding *P* value thresholds.**

Replication set	Number of SNPs tested	Number of independent tests	<i>P</i> value threshold
African Americans	7,470	1,132.1	4.42E-05
Korean	72	26	1.52E-03
Chinese	7,436	1,133	4.41E-05
Hispanics	7,473	1,133.1	4.41E-05

This table shows the number of SNPs tested in the regional analyses in other ancestries. The effective number of independent loci being tested in each replication population was calculated using the technique proposed by Li and Ji ⁷, as implemented in “matSpD” ⁸ (<http://gump.qimr.edu.au/general/daleN/matSpD/>), based on the linkage disequilibrium structures of the 1000 Genomes all ancestries sample. The *P* value threshold is defined by 0.05/ (number of independent tests).

Supplementary Table 12. **Gene Expression Assays for mRNA expression profiling.**

Gene symbol	Assay ID	Gene name	Amplicon size
<i>EFEMP1</i>	Hs00244575_m1	EGF containing fibulin-like extracellular matrix protein 1	86
<i>BMP6</i>	Hs01099594_m1	Bone morphogenetic protein 6	108
<i>WWOX</i>	Hs03044790_m1	WW domain containing oxidoreductase	89
<i>KCNJ2</i>	Hs01876357_s1	Potassium inwardly-rectifying channel, subfamily J, member 2	160
<i>PRDM11</i>	Hs01075851_m1	PR domain containing 11	77
<i>HSD17B12</i>	Hs00275054_m1	Hydroxysteroid (17-beta) dehydrogenase 12	65
<i>GAPDH</i>	Hs99999905_m1	Glyceraldehyde-3-phosphate dehydrogenase	124

Assay identifiers for the genes associated with FVC which we tested in lung tissue, human bronchial epithelial cells, human airway smooth muscle, peripheral mononuclear blood cells.

Supplementary note

Gene expression profiling

The mRNA expression profiles of the candidate genes *EFEMP1*, *BMP6*, *WWOX*, *TMEM163*, *KCNJ2*, *PRDM11* and *HSD17B12* and the housekeeping gene *GAPDH* were determined in human lung tissue and primary cell samples using RT-PCR. Lung resection specimens were obtained from patients diagnosed with solitary pulmonary tumors at Ghent University Hospital (Ghent, Belgium). Lung tissue at maximum distance from the pulmonary lesions and without signs of retro-obstructive pneumonia or tumor invasion was collected by a pathologist. Primary human bronchial epithelial cells (HBEC) and human airway smooth muscle cells (HASM) 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). Peripheral blood mononuclear cells (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 RNA template using the Transcriptor Universal cDNA Master kit (Roche) following manufacturer's instructions. Expression of the candidate genes and the housekeeping gene *GAPDH* was analyzed using TaqMan Gene Expression Assays (Applied Biosystems, Foster City, CA, USA; Assay ID numbers are given in supplementary table 3). Real-time-PCR reactions were set up using diluted cDNA and the LightCycler480 Probes Master (Roche) with identical amplification conditions for each of the genes. Amplifications were performed on a LightCycler480 detection system (Roche) with the following cycling conditions: 10 minutes at 95°C, 70 cycles of 10 seconds at 95°C and 15 seconds at 60°C. Amplification was followed in real-time by gene specific TaqMan probes and final PCR products were visualized by gel electrophoresis.

Expression Quantitative Trait Loci (eQTLs)

For Rotterdam Study III (RS-III), whole-blood cells were collected in PAXgene-tubes (Becton Dickinson). Total RNA was isolated using PAXgene Blood RNA kits (Qiagen), and to ensure a constant high quality of the RNA preparations, all RNA samples were analysed using the Labchip GX (Calliper) according to the manufacturer's instructions. Samples with an RNA Quality Score ≥ 7 were amplified and labelled (AmbionTotalPrep RNA), and hybridized to the Illumina Whole-Genome Expression Beadchips (HumanHT-12 v4). Processing of the samples was performed at the Genetic Laboratory of Internal Medicine, Erasmus University Medical Center Rotterdam. The RS-III expression dataset is available at GEO (Gene Expression Omnibus) public repository under the accession GSE 33828. For normalization, raw intensity data generated with the expression arrays were exported from Illumina's GenomeStudio V 2010.1 Gene Expression Module to the R environment and quantile normalized and log₂-transformed, as well as probe-centered, and sample-standardized.

Joint analysis for variants in the KCNJ2 locus

The method developed by Peter Visscher and colleagues⁹ to approximate conditional and joint association analyses, accounting for linkage disequilibrium between SNPs was undertaken using the GCTA software (<http://gump.qimr.edu.au/gcta/>). This method uses summary level results, as well as individual level data for a subset of individuals sampled from the same discovery population to estimate linkage disequilibrium correlations. The results of the current analysis for FVC were used for two variants (rs11654749 and rs6501431) and the 1958 Birth Cohort T1DGC (N=2,593) was used as the reference population to estimate linkage disequilibrium correlations. Quality control checks were run on this dataset prior to the analysis, including the removal of individuals with outlying ancestries.

Funding and acknowledgements per study

SpiroMeta consortium

Funding & Acknowledgements:

The research undertaken by MDT, MSA and LVW was part-funded 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. MDT holds a Medical Research Council Senior Clinical Fellowship (G0902313). 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 (EMDA), part-financed by the European Regional Development Fund and delivered by Medilink East Midlands. IPH holds a Medical Research Council programme grant (G1000861).

CHARGE consortium

Funding & Acknowledgements: The CHARGE Pulmonary Working Group would like to acknowledge funding from the NHLBI (HL105756) and the CHARGE consortium for its organizational support. This work was supported in part by the Intramural Research Program of the National Institutes of Health (NIH), National Institute of Environmental Health Sciences (NIEHS).

Cohorts included in stage 1

AGES

Funding & Acknowledgements: This study has been funded by NIH contract N01-AG-1-2100, the NIA Intramural Research Program, Hjartavernd (the Icelandic Heart Association), and the Althingi (the Icelandic Parliament). The study is approved by the Icelandic National Bioethics Committee, VSN: 00-063. The researchers are indebted to the participants for their willingness to participate in the study.

ARIC

Funding & Acknowledgements: The Atherosclerosis Risk in Communities Study is carried out as a collaborativestudy supported by National Heart, Lung, and Blood Institute contracts(HHSN268201100005C, HHSN268201100006C, HHSN268201100007C, HHSN268201100008C, HHSN268201100009C, HHSN268201100010C, HHSN268201100011C, and HHSN268201100012C), R01HL087641, R01HL59367 andR01HL086694; National Human Genome Research Institute contract U01HG004402;and National Institutes of Health contract HHSN268200625226C. The authors thank the staff and participants of the ARIC study for their important contributions. Infrastructure was partly supported by Grant Number UL1RR025005, a component of the National Institutes of Health and NIH Roadmap for Medical Research.

B58C T1DGC & WTCC

Funding & Acknowledgements: We acknowledge 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://www2.le.ac.uk/projects/birthcohort>). 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 (NIDDK), National Institute of Allergy and Infectious Diseases (NIAID), National Human Genome Research Institute (NHGRI), National Institute of Child Health and Human Development (NICHD), and Juvenile Diabetes Research Foundation International (JDRF) 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).

CARDIA

Funding & Acknowledgements: The Coronary Artery Risk Development in Young Adults (CARDIA) study is funded by contracts N01-HC-95095, N01-HC-48047, N01-HC-48048, N01-HC-48049, N01-HC-48050, N01-HC-45134, N01-HC-05187, N01-HC-45205 and N01-HC-45204 from the National Heart, Lung and Blood Institute to the CARDIA investigators. A full list of principal CARDIA investigators and institutions can be found at http://www.cardia.dopm.uab.edu/o_pain.htm. Genotyping of the CARDIA participants and statistical data analysis was supported by grants U01-HG-004729 from the National Human Genome Research Institute and R01-HL-084099 from the National Heart, Lung and Blood Institute to MF and K23HL094531-01 from the National Heart, Lung and Blood Institute and M01 RR00997 from the National Center for Research Resources to AS. The authors thank the investigators and staff of the GENEVA coordinating center and genotyping center, as well as the staff and participants of the CARDIA study for their important contributions.

CHS

Funding & Acknowledgements: This CHS research was supported by NHLBI contracts HHSN268201200036C, HHSN268200800007C, N01HC55222, N01HC85079, N01HC85080, N01HC85081, N01HC85082, N01HC85083, N01HC85086, and HHSN268200960009C; and NHLBI grants HL080295, HL087652, HL105756, and HL085251 with additional contribution from the National Institute of Neurological Disorders and Stroke (NINDS). Additional support was provided through AG023629 from the National Institute on Aging (NIA). A full list of CHS investigators and institutions can be found at <http://chs-nhlbi.org/>. The provision of genotyping data was supported in part by the National Center for Advancing Translational Sciences, CTSI grant UL1TR000124, and the National Institute of Diabetes and Digestive and Kidney Disease Diabetes Research Center (DRC) grant DK063491 to the Southern California Diabetes Endocrinology Research Center.

CROATIA-Korcula

Funding & Acknowledgements: The CROATIA studies were supported through the grants from the Medical Research Council UK and Ministry of Science, Education and Sport of the Republic of Croatia (number 108-1080315-0302) and the European Union framework program 6 EUROSPAN project (contract no. LSHG-CT-2006-018947). The CROATIA cohorts would like to acknowledge the invaluable contributions of the recruitment teams in Vis, Korcula and Split (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 Vis, Korcula and Split. SNP genotyping was performed at the Wellcome Trust Clinical Research Facility in Edinburgh for CROATIA-Vis, by Helmholtz Zentrum München, GmbH, Neuherberg, Germany for CROATIA-Korcula and by AROS Applied Biotechnology, Aarhus, Denmark for CROATIA-Split.

CROATIA-Vis

Funding & Acknowledgements: The CROATIA studies were supported through the grants from the Medical Research Council UK and Ministry of Science, Education and Sport of the Republic of Croatia (number 108-1080315-0302) and the European Union framework program 6 EUROSPAN project (contract no. LSHG-CT-2006-018947). The CROATIA cohorts would like to acknowledge the invaluable contributions of the recruitment teams in Vis, Korcula and Split (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 Vis, Korcula and Split. SNP genotyping was performed at the Wellcome Trust Clinical Research Facility in Edinburgh for CROATIA-Vis, by Helmholtz Zentrum München, GmbH, Neuherberg, Germany for CROATIA-Korcula and by AROS Applied Biotechnology, Aarhus, Denmark for CROATIA-Split.

ECRHS

Funding & Acknowledgements: Acknowledgement

The co-ordination of ECRHS II (the phase of the study that collected blood samples suitable for genotyping) was supported by the European Commission, as part of their Quality of Life programme. *Genotyping was conducted within then the EU-funded GABRIEL project.* The following bodies funded the local studies in ECRHS II: **Albacete:** Fondo de Investigaciones Santarias (FIS) (grant code: 97/0035-01, 99/0034-01 and 99/0034-02), Hospital Universitario de Albacete, Consejería de Sanidad; **Barcelona:** SEPAR, Public Health Service (grant code: R01 HL62633-01), Fondo de Investigaciones Santarias (FIS) (grant code: 97/0035-01, 99/0034-01 and 99/0034-02) CIRIT (grant code: 1999SGR 00241) Red Respira ISCII; CIBER Epidemiología y Salud Pública (CIBERESP), Spain **Basel:** Swiss National Science Foundation, Swiss Federal Office for Education & Science, Swiss National Accident Insurance Fund (SUVA), USC NIEHS Center grant 5P30 ES07048; **Bergen:** Norwegian Research Council, Norwegian Asthma & Allergy Association (NAAF), Glaxo Wellcome AS, Norway Research Fund; **Erfurt:** GSF-National Research Centre for Environment & Health, Deutsche Forschungsgemeinschaft (DFG) (grant code FR 1526/1-1); **Galdakao:** Basque Health Dept; **Grenoble:** Programme Hospitalier de Recherche Clinique-DRC de Grenoble 2000 no. 2610, Ministry of Health, Direction de la Recherche Clinique, CHU de Grenoble, Ministère de l'Emploi et de la Solidarite, Direction Generale de la Sante, Comité des Maladies Respiratoires de l'Isere; **Hamburg:** GSF-National Research Centre for Environment & Health, Deutsche Forschungsgemeinschaft (DFG) (grant code MA 711/4-1); **Ipswich and Norwich:** Asthma UK (formerly known as National Asthma Campaign); **Huelva:** Fondo de Investigaciones Santarias (FIS) (grant code: 97/0035-01, 99/0034-01 and 99/0034-02); **Oviedo:** Fondo de Investigaciones Santarias (FIS) (grant code: 97/0035-01, 99/0034-01 and 99/0034-02) ; **Paris:** Ministère de l'Emploi et de la Solidarite, Direction Generale de la Sante, UCB-Pharma (France), Aventis (France), Glaxo France, Programme Hospitalier de Recherche Clinique-DRC de Grenoble 2000 no. 2610, Ministry of Health, Direction de la Recherche Clinique, CHU de Grenoble; **Tartu:** Estonian Science Foundation; **Umeå:** Swedish Heart Lung Foundation, Swedish Foundation for Health Care Sciences & Allergy Research, Swedish Asthma & Allergy Foundation, Swedish Cancer & Allergy Foundation; **Uppsala:** Swedish Heart Lung Foundation, Swedish Foundation for Health Care Sciences & Allergy Research, Swedish Asthma & Allergy Foundation, Swedish Cancer & Allergy Foundation.

EPIC

Funding & Acknowledgements: The EPIC Norfolk Study is funded by Cancer Research United Kingdom and the Medical Research Council. EPIC would like to thank Manjinder S. Sandhu (Department of Public Health and Primary Care, Institute of Public Health, University of Cambridge).

FHS

Funding & Acknowledgements: Framingham Heart Study (FHS) research was conducted in part using data and resources of the NHLBI and Boston University School of Medicine. The analyses reflect intellectual input and resource development from the FHS investigators participating in the SNP Health Association Resource (SHARe) project. This work was partially supported by NHLBI (contract no. N01-HC-25195) and its contract with Affymetrix for genotyping services (contract no. N02-HL-6-4278). A portion of this research utilized the Linux Cluster for Genetic Analysis (LinGA-II) funded by the Robert Dawson Evans Endowment of the Department of Medicine at Boston University School of Medicine and Boston Medical Center. J.B.W. received salary support from the Flight Attendant Medical Research Institute (FAMRI).

FTC

Funding & Acknowledgements: K. H. P. acknowledges Novo Nordisk Foundation, Finnish Diabetes Foundation, Finnish Foundation for Cardiovascular Research, and Finnish Academy (266286). T. R. acknowledges Academy of Finland (75507). J. K. acknowledges The European Community's Seventh Framework Programme (FP7/2007-2013), ENGAGE Consortium (HEALTH-F4-2007- 201413), The European Community's Fifth Framework Programme, GenomEUtwin (QLG2-CT-2002-01254) and

Academy of Finland, Center of Excellence in Complex Disease Genetics (213506 and 129680). A.V. acknowledges Academy of Finland (251723).

Health ABC

Funding & Acknowledgements: S.B.K. acknowledges NIH/NIA (RC1 AG035835) for analysis time funding. Y.L. and S.B.K. acknowledge NIH/NIA (R01 AG032098) for GWAS funding. P.A.C. acknowledges NIH/NHLBI (R01 HL071022). This research was supported by NIA contracts N01AG62101, N01AG62103, and N01AG62106; NIA grant R01-AG028050, and NINR grant R01-NR012459 and was supported in part by the Intramural Research Program of the NIH, National Institute on Aging. The genome-wide association study was funded by NIA grant 1R01AG032098-01A1 and genotyping services were provided by the Center for Inherited Disease Research (CIDR). CIDR is fully funded through a federal contract from the National Institutes of Health to The Johns Hopkins University, contract number HHSN268200782096C.

Health 2000

Funding & Acknowledgements: This study was financially supported by the Medical Research Fund of the Tampere University Hospital.

Hunter Community Study

Acknowledgements: The authors would like to thank the men and women participating in the HCS as well as The University of Newcastle, Vincent Fairfax Family Foundation and The HunterMedical Research Institute.

KORA S3

Funding & Acknowledgements: The KORA study 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 (BMBF) and by the State of Bavaria. 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 (BMBF, FKZ 01GI0882).

MESA-Lung

Funding & Acknowledgements: MESA and MESA SHARe acknowledges NIH/NHLBI (N01-HC-95159 through N01-HC-95169 and RR-024156). MESA SHARe acknowledges NIH/NHLBI (N02-HL-6-4278). MESA Family acknowledges NIH (R01HL071051, R01HL071205, R01HL071250, R01HL071251, R01HL071252, R01HL071258, R01HL071259). MESA Lung acknowledges NIH/NHLBI (R01 HL077612). MESA Lung SHARe acknowledges NIH/NHLBI (RC1 HL100543).

NFBC1966

Funding & Acknowledgements:

We thank the late Professor Paula Rantakallio (launch of NFBC1966), and Ms Outi Tornwall and Ms MinttuJussila (DNA biobanking). The authors would like to acknowledge the contribution of the late Academician of Science Leena Peltonen.

NFBC1966 received financial support from the Academy of Finland (project grants 104781, 120315, 129269, 1114194, Center of Excellence in Complex Disease Genetics and SALVE), University Hospital Oulu, Biocenter, University of Oulu, Finland (75617), the European Commission (EURO-BLCS, Framework 5 award QLGI-CT-2000-01643), NHLBI grant 5R01HL087679-02 through the STAMPEED program (1RL1MH083268-01), NIH/NIMH (5R01MH63706-02), ENGAGE project and grant agreement HEALTH-F4-2007-201413, the Medical Research Council, UK (G0500539, G0600705, PrevMetSyn/SALVE) and the Wellcome Trust (project grant GR069224), UK. A.R. acknowledges

European Commission (through project GABRIEL (contract number 018996 under the Integrated Program LSH-2004-1.2.5-1)) and Department of Health, UK.

NSPHS

Funding & Acknowledgements:

Swedish Medical Research Council K2007-66X-20270-01-3 and Swedish Society for Medical Research. The computations were performed on resources provided by SNIC through Uppsala Multidisciplinary Center for Advanced Computational Science (UPPMAX) under Project b2011203.

ORCADES

Funding & Acknowledgements: ORCADES was supported by the Chief Scientist Office of the Scottish Government, the Royal Society, the MRC Human Genetics Unit, Arthritis Research UK and the European Union framework program 6 EUROSPAN project (contract no. LSHG-CT-2006-018947). ORCADES would like to acknowledge the invaluable contributions of Lorraine Anderson and the research nurses in Orkney, the administrative team in Edinburgh and the people of Orkney. DNA extractions were performed at the Wellcome Trust Clinical Research Facility in Edinburgh.

Rotterdam Study I, II & III

Funding & Acknowledgements: The generation and management of GWAS genotype data for the Rotterdam Study is supported by the Netherlands Organization of Scientific Research NWO Investments (nr. 175.010.2005.011, 911-03-012). This study is funded by the Research Institute for Diseases in the Elderly (014-93-015; RIDE2), the Netherlands Genomics Initiative (NGI)/Netherlands Organisation for Scientific Research (NWO) project nr. 050-060-810. We thank Pascal Arp, Mila Jhamai, Marijn Verkerk, Lizbeth Herrera and Marjolein Peters for their help in creating the GWAS database, and Karol Estrada and Maksim V. Struchalin for their support in creation and analysis of imputed data. The Rotterdam Study is funded by Erasmus Medical Center and Erasmus University, Rotterdam, Netherlands Organization for the Health Research and Development (ZonMw), the Research Institute for Diseases in the Elderly (RIDE), the Ministry of Education, Culture and Science, the Ministry for Health, Welfare and Sports, the European Commission (DG XII), and the Municipality of Rotterdam. The authors are grateful to the study participants, the staff from the Rotterdam Study and the participating general practitioners and pharmacists. We would like to thank Karol Estrada, Dr. Fernando Rivadeneira, Dr. Tobias A. Knoch, Anis Abuseiris, Luc V. de Zeeuw, and Rob de Graaf (Erasmus MC Rotterdam, The Netherlands), for their help in creating GRIMP, and BigGRID, MediGRID, and Services@MediGRID/D-Grid, (funded by the German Bundesministerium fuer Forschung und Technology; grants 01 AK 803 A-H, 01 IG 07015 G) for access to their grid computing resources.

SHIP

Funding & Acknowledgements: 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 no.01ZZ9603, 01ZZ0103, 01ZZ0403, 03IS2061A, BMBF 01GI0883). 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.

Twins UK-I

Funding & Acknowledgements: The study was funded by the Wellcome Trust; European Community's Seventh Framework Programme (FP7/2007-2013). The study also receives support

from the National Institute for Health Research (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. SNP Genotyping was performed by The Wellcome Trust Sanger Institute and National Eye Institute via NIH/CIDR.

Cohorts included in Stage 2

BHS 1& 2

Funding & Acknowledgements: 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 Busselton Health Study 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.

CROATIA-Split

Funding & Acknowledgements: The CROATIA studies were supported through the grants from the Medical Research Council UK and Ministry of Science, Education and Sport of the Republic of Croatia (number 108-1080315-0302) and the European Union framework program 6 EUROSPAN project (contract no. LSHG-CT-2006-018947). The CROATIA cohorts would like to acknowledge the invaluable contributions of the recruitment teams in Vis, Korcula and Split (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 Vis, Korcula and Split. SNP genotyping was performed at the Wellcome Trust Clinical Research Facility in Edinburgh for CROATIA-Vis, by Helmholtz Zentrum München, GmbH, Neuherberg, Germany for CROATIA-Korcula and by AROS Applied Biotechnology, Aarhus, Denmark for CROATIA-Split.

Generation Scotland

Authors and affiliations:

David J Portous,

Medical Genetics Section, Centre for Genomics and Experimental Medicine, University of Edinburgh, Edinburgh UK.

MRC Institute of Genetics and Molecular Medicine, Western General, Edinburgh, UK

Blair H Smith

Medical Research Institute, University of Dundee, Dundee, UK

Holly Trochet

MRC Human Genetics Unit, University of Edinburgh, Edinburgh, UK.

Institute of Genetics and Molecular Medicine, Western General Hospital, Edinburgh, UK.

Acknowledgements: We would like to 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. SNP genotyping was performed at the Wellcome Trust Clinical Research Facility in Edinburgh. GS:SFHS is funded by the Scottish Executive Health Department, Chief Scientist Office, grant number CZD/16/6. SNP genotyping was funded by the Medical Research Council UK.

KORA F4

Funding & Acknowledgements:

The KORA study 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 (BMBF) and by the State of Bavaria. 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 (BMBF, FKZ 01GI0882).

LBC1936

Funding & Acknowledgements: 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.

Lifelines

Funding & Acknowledgements: The Lifelines Cohort Study is supported by the Netherlands Organization of Scientific Research NOW (1750102007006), the Economic Structure Enhancing Fund (FES) of the Dutch government, the Ministry of Economic Affairs, the Ministry of Education, Culture and Science the Ministry for Health, Welfare and Sports, the Northern Netherlands Collaboration of Provinces (SNN), the Province of Groningen, the city of Groningen, the University Medical Center Groningen, the University of Groningen and the Dutch Kidney Foundation.

LLFS

Funding & Acknowledgements: The Long Life Family Study (LLFS) was supported by the National Institute on Aging (NIA) grants U01-AG023712, U01-AG023744, U01-AG023746, U01-AG023749, U01-AG023755, U19-AG023122, K24-AG025727, R01-AG032319, the Glenn Medical Research Foundation, and the National Heart Lung Blood Institute (NHLBI, R21-HL114237). The investigators would like to thank the LLFS participants and staff for their valuable contributions. We are also grateful to Heidi Dubrouillet for her efforts in project management.

PIVUS

Funding & Acknowledgements: A.P.M. acknowledges the Wellcome Trust (WT098017, WT064890, 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.

Twins UK-II & III

Funding & Acknowledgements: The study was funded by the Wellcome Trust; European Community's Seventh Framework Programme (FP7/2007-2013). The study also receives support from the National Institute for Health Research (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. SNP Genotyping was performed by The Wellcome Trust Sanger Institute and National Eye Institute via NIH/CIDR.

Study included for multi-ethnic lookup:

CARe

Acknowledgements:

The authors wish to acknowledge the support of the National Heart, Lung, and Blood Institute and the contributions of the research institutions, study investigators, field staff and study participants in creating this resource for biomedical research. The following parent studies have contributed parent study data, ancillary study data, and DNA samples through the Broad Institute (N01-HC-65226) to create this genotype/phenotype data base for wide dissemination to the biomedical research community: ARIC, CARDIA, CFS, JHS, and MESA, MESA Family, MESA Air Pollution and MESA Lung studies. A full list of participating MESA Investigators and institutions can be found at <http://www.mesa-nhlbi.org>. A full list of principal CHS investigators and institutions can be found at <http://www.chs-nhlbi.org/pi.htm>. This manuscript has been reviewed by the MESA Investigators for scientific content and consistency of data interpretation with previous MESA publications and significant comments have been incorporated prior to submission for publication. The ARIC Study is carried out as a collaborative study supported by National Heart, Lung, and Blood Institute. The authors thank the staff and participants of the ARIC study for their important contributions. Infrastructure was partly supported by Grant Number UL1RR025005, a component of the National Institutes of Health and NIH Roadmap for Medical Research. CHS research was supported by NHLBI contracts, with additional contribution from NNDS and the NIA.

Funding: NIH/NHLBI grants RC1-HL100543, R01-HL077612, and R01-HL093081, in addition to: **ARIC:** HHSN268201100005C, HHSN268201100006C, HHSN268201100007C, HHSN268201100008C, HHSN268201100009C, HHSN268201100010C, HHSN268201100011C, and HHSN268201100012C), R01HL087641, R01HL59367 and R01HL086694; NIH/NHGRI contract U01HG004402 and National Institutes of Health contract HHSN268200625226C; N01-HC-55015, N01-HC-55016, N01-HC-55021, N01-HC-55019, N01-HC-55020, N01-HC-55017, N01-HC-55018; **Broad Institute:** N01-HC-65226; **CARDIA:** N01-HC-48047, N01-HC-48048, N01-HC-48049, N01-HC-48050, N01-HC-95095, N01-HC-45204, N01-HC-45205, N01-HC-05187, N01-HC-45134, N01-HC-95100; **CFS:** R01 HL46380-01-16; **JHS:** N01-HC-95170, N01-HC-95171, N01-HC-95172; **MESA:** N01-HC-95159, N01-HC-95160, N01-HC-95161, N01-HC-95162, N01-HC-95163, N01-HC-95164, N01-HC-95165, N01-HC-95166, N01-HC-95167, N01-HC-95168, N01-HC-95169, R01-HL093081 and RR-024156. R01-HL-071051, R01-HL-071205, R01-HL-071250, R01-HL-071251, R01-HL-071252, R01-HL-071258, R01-HL-071259, EPA grant RD831697;

KARE3 and Healthy Twin Study

Acknowledgements: This work was funded by the Consortium for Large Scale Genome Wide Association Study III (2011E7300400), which was supported by the genotyping data (the Korean Genome Analysis Project, 4845-301) and the phenotype data (the Korean Genome Epidemiology Study, 4851-302) and a grant from the Centers for Disease Control and Prevention of Korea (budget 2012-E71011-00, 2011-E71011-00, and 2010-E71010-00) and also supported by the National Project for Personalized Genomic Medicine (A111218-11-GM02 and A111218-12-GM10).

Individual study descriptions

Cohorts included in stage 1

This section describes study-specific characteristics that are not presented in the tables. All participants provided written informed consent and studies were approved by local Research Ethics Committees and/or Institutional Review boards.

AGES: The Reykjavik Study cohort originally comprised a random sample of 30,795 men and women born in 1907-1935 and living in Reykjavik in 1967. A total of 19,381 attended, resulting in 71% recruitment rate. The study sample was divided into six groups by birth year and birth date within month. One group was designated for longitudinal follow-up and was examined in all stages. One group was designated a control group and was not included in examinations until 1991. Other groups were invited to participate in specific stages of the study. Between 2002 and 2006, the **AGES**-Reykjavik study re-examined 5,764 survivors of the original cohort who had participated before in the Reykjavik Study. Pulmonary function was obtained by spirometry (Vitalograph). Testing was conducted with the participant in a standing position and a disposable mouthpiece. Participants were shown how to perform the maneuver by the technician before testing. A successful test session was defined as at least three acceptable maneuvers.

Atherosclerosis Risk in Communities (**ARIC**), is a population based study of risk factors for atherosclerosis and its sequelae¹⁰ in adults from four U.S. field centers aged 45-64 at recruitment in 1987-1989. ARIC spirometry measurements were made with a Collins Survey II water-seal spirometer (Collins Medical, Inc.) and Pulmo-Screen II software (PDS Healthcare Products, Inc.). Genotyping was done using the AffymetrixGeneChip SNP Array 6.0 and imputation was performed using MACH. Quality control steps for genotyping data included exclusions for call rate <95%, minor allele frequency <1%, HWE $P < 10^{-5}$, no chromosomal location, suspected first-degree relative of an included individual based on genotype data, or more than 8 standard deviations for any of the first ten principal components. The current analysis includes 9,078 Caucasian subjects with genotyping data, pulmonary function measures and complete covariate information.

Details of the **British 1958 Birth Cohort** biomedical follow-up have been previously reported¹¹ and a full technical report is available online (<http://www.b58cgene.sgul.ac.uk/report.php>). Spirometry at age 44-45 years was done in the standing position without nose clips, using a Vitalograph handheld spirometer as previously described¹². In the analysis, all readings with a best-test variation greater than 10% were excluded.

Cross-sectional analyses of data from year 0 examination of the **Coronary Artery Risk Development in Young Adults** (CARDIA) cohort was performed. During 1985-1986, CARDIA randomly recruited 5,115 black and white men and women, aged 18 to 30 years, from the general population at Birmingham, Alabama; Chicago, Illinois; and Minneapolis, Minnesota; and from the membership of the Oakland Kaiser-Permanente Health Plan in Oakland, California. Detailed methods, instruments, and quality control procedures are described at the CARDIA website (http://www.cardia.dopm.uab.edu/ex_mt.htm) and in other published reports^{13,14}. Spirometric pulmonary function testing were performed using the Collins survey 8-liter water-sealed spirometer and the Eagle II microprocessor (Warren E. Collins, Inc., Braintree, MA) in a sitting position with noseclips, as per the 1979 American Thoracic Society criteria¹⁵. Specifically, each subject performed a minimum of three trials with expirations recorded to the FVC plateau, which occurs after six seconds of expiration in adult males and was maintained for at least one second before terminating the forced expiratory maneuver. If, at the end of the three trials, there were at least three acceptable tracings, and with the maximum FVC and FEV₁ reproduced to within 5% or 100 mL, whichever is greater, no more trials were performed.

The **CHS** is a population-based cohort study of risk factors for CHD and stroke in adults > 65 years conducted across four U.S. field centers¹⁶. CHS spirometry measurements were made with a Collins Survey I water-seal spirometer (Collins Medical, Inc.) and software from S&M Instruments. CHS genotyped 3,980 participants free of cardiovascular disease at baseline with available DNA and consent to genetic testing. After exclusions for call rate < 95%, sex mismatch or discordance with prior genotyping, 3,291 self-identified white participants remained. Of these, 3,140 had pulmonary function measures and complete covariate information.

The **CROATIA-Korcula** study is a family-based, cross-sectional study in the isolated island of Korcula that included about 1,000 examinees aged 18-95. Spirometry was performed in the sitting upright position without nose clips, using a Jaeger Toennisen spirometer. Three readings were taken for each examinee, at least 15 seconds apart¹⁷. Population stratification was partially taken into account after adjusting the measures for the 3 first principal components drawn from the population genomic kinship matrix computed using the *ibs* function of the GenABEL¹⁸. Additional correction of the standard error of effect estimates was undertaken by genomic control¹⁹. Effect estimates were obtained using the *palinear* function of the ProbABEL package.

The **CROATIA-Vis** study is a family-based, cross-sectional study in the isolated island of Vis that included about 1,000 examinees aged 18-93. Spirometry was performed in the sitting upright position without nose clips, using a Jaeger Toennisen spirometer. Three readings were taken for each examinee, at least 15 seconds apart¹⁷. Population stratification was partially taken into account after adjusting the measures for the 3 first principal components drawn from the population genomic kinship matrix computed using the *ibs* function of the GenABEL¹⁸. Additional correction of the standard error of effect estimates was undertaken by genomic control¹⁹. Effect estimates were obtained using the *palinear* function of the ProbABEL package.

Details of the methods of **ECRHS I** and **ECRHS II**, a multicenter international cohort study, have been published elsewhere^{20,21}. Participants within the ECRHS were eligible for inclusion in this analysis if they were identified by random sampling of those who fulfilled the following criteria 1) lived in centres that took part in genome-wide genotyping initiative under the auspices of GABRIEL²² AND 2) were initially selected to take part in the ECRHS clinical measurements as part of the random sample (ie not specifically selected for inclusion because of any pre-existing disease). Participants were included in this analyses if they also provided a technically satisfactory forced expiratory manoeuvre, compliant with ATS spirometry criteria, at the time of the first survey (aged 20-48). Most centers used the BIOMEDIN water-sealed spirometer for lung function measures and all centers conducted manoeuvres in the sitting position with nose clips on. Further details are available in²³.

The **EPIC Norfolk** GWA cohort includes 2,566 participants randomly selected from the EPIC-Norfolk Study, a population-based cohort study of 25,663 men and women of European descent aged 39-79 years recruited in Norfolk, UK between 1993 and 1997¹⁰⁹. Respiratory function was assessed by spirometry^{24,25}. Forced expiratory volume in 1 second (FEV₁) and forced vital capacity (FVC) were measured twice using a portable spirometer (Micro Medical, Rochester, United Kingdom). The higher values of the two readings for FEV₁ and FVC were used for the analyses.

The **FHS** is a longitudinal community-based family study that originated in 1948 with the recruitment of adults from the town of Framingham, MA. The offspring of the original cohort were recruited to participate in 1971, and the third generation (Gen3) was recruited starting in 2002. Spirometry has been measured on all three generations of the participating families as part of the clinical examinations. For the original and offspring cohorts, spirometry was performed using a 6-L water-filled Collins survey spirometer connected to an Eagle II microprocessor (Collins Medical, Braintree, MA) or in later offspring examinations to a personal computer running software developed by S&M Instruments, Doylestown, PA. For the Gen3 cohort, spirometry was performed using the CPL System

(Collins Medical, Braintree, MA) with a dry rolling-seal spirometer. All of these systems provided an automatic correction for body temperature and pressure, saturated, and provided quality assurance information. Each Framingham participant's latest examination with acceptable pulmonary function data was used in this analysis. Eligible examinations included original cohort examinations 19, 17, 16 and 13; offspring cohort examinations 7, 6, 5, and 3; and Gen3 cohort examination 1. A total of 9,274 subjects were genotyped and exclusions were based on a call rate < 97%, heterozygosity more than 5 standard deviations from the mean, or excessive non-inheritance. A total of 8,481 participants passed the genotyping quality control criteria. Of these, 7,914 participants with complete spirometry and covariate data were used in the final analysis data set.

The subjects for the **Finnish Twin Cohort (FTC)** came from two clinical sub studies of twins in Finland: 1. The FinnTwin16 study is a nationwide longitudinal cohort study of health behaviors in twins and their families²⁶. The participants were 34 MZ twin pairs, who were a subset of 347 twins recruited from the FinnTwin16 study of five consecutive birth cohorts (1975–1979) of twins (n = 4929 individuals), and who participated in a clinical study of body composition and cardiorespiratory fitness. The study protocol included a bicycle ergometer exercise test with gas exchange analysis. A mass flow sensor of the gas exchange device (Sensormedics Co) was used to measure forced expiratory volume in one second and forced vital capacity before exercise. 2. The Finnish Twin Study on Ageing (FITSA). Participants were recruited from the older Finnish Twin Cohort for a clinical study of functional limitations in older women. Clinical assessment was conducted in 2000–2001 at the University of Jyväskylä. The final sample consisted of 103 monozygotic (MZ) and 114 dizygotic (DZ) twin pairs. Lung function was measured in the standing position using an electronic spirometer (Medikro 202, Kuopio, Finland). The subject was asked to inhale maximally and to exhale as fast as possible into a mouthpiece connected to a flow transducer and a flow-volume curve was created. At least two tests were performed and the best result taken for the analyses. Spirometer was calibrated daily with a three-litre pump and was accurate to within 1%. Both sub studies were mainly genotyped as part of the GenomEUtwin project of female MZ pairs. To avoid taking into account the statistical dependence of two related individuals in a family, we selected one twin at random for inclusion in the analysis if both of the MZ twins in a pair had the phenotype.

The DNA archive established from the **Health 2000** Survey Cohort was used. Details of this study population and phenotyping procedures have been previously reported²⁷. Spirometry was done in the standing position without nose clips, using a Vitalograph 2150 spirometer. In the analysis, the maximum permissible difference between the two highest FEV₁ and FVC values was 10%.

The Health Aging and Body Composition (**Health ABC**) study is a prospective observational cohort of well-functioning individuals aged 70–79 years. The Health ABC study recruited 3,075 community-dwelling African and European Americans, men and women, at two field centers at the University of Pittsburgh, Pennsylvania and the University of Tennessee, Memphis. Spirometry was performed with a horizontal dry rolling seal spirometer (SensorMedics Corporation, Yorba Linda, CA) connected to a computer. Pulmonary function testing followed ATS guidelines for the standardization of spirometry, and is described in detail elsewhere¹²⁴. Health ABC genotyped 1,794 self-described white participants at baseline with available DNA and consent to genetics testing; of these 1,661 passed quality control benchmarks (call rate > 97%, no sex mismatch, and cryptic relatedness), and 1,472 had pulmonary function measurements and complete data on covariates.

The **Hunter Community Study** is a population-based prospective cohort study of community-dwelling men and women aged 55–85 years of age who reside in Newcastle, New South Wales (NSW), Australia. The cohort comprises 3253 participants that were randomly selected from the NSW State electoral roll between 2004 and 2007. Spirometry was performed using electronic spirometers (Micro Medical SpiroUSB, Cardinal Health, Kent, UK) with Spida 5 software (Carefusion Ltd, Kent, UK) according to American Thoracic Society (ATS) guidelines, i.e. 3 acceptable traces with no more

than 8 attempts. Key lung function measures were FEV₁ and FVC. The spirometer was calibrated daily using a 3-Litre syringe.

The **KORA S3** study (Cooperative Health Research in the Region of Augsburg) is also known as the third MONICA survey, which was performed 1994-95 in Augsburg, Germany. The objective and protocol of this survey have been published earlier²⁸. The third MONICA survey consisted of a random sample of all registered residents of the city of Augsburg aged 25–74 years. For all participants younger than 60, who did not smoke or use inhalers one hour before the test, FEV₁ and FVC were determined by spirometry in 1997-98. All spirometric tests were performed strictly adhering to the ECRHS protocol²⁰. Tests were accounted valid if at least two technically satisfactory manoeuvres could be obtained throughout a maximum of nine trials. FVC was defined as the maximum value within all valid manoeuvres.

The Long Life Family Study (**LLFS**) is a family-based cohort study that enrolled 4559 long-lived probands and their siblings (n=1445), their offspring (n=2329) and spousal controls (n=785) recruited from 3 U.S. field centers (Boston University Medical Center in Boston MA, Columbia College of Physicians and Surgeons in New York City NY, and the University of Pittsburgh in Pittsburgh PA) and the University of Southern Denmark. Participants were chosen from the Center for Medicare and Medicaid Services lists of Medicare enrollees who were ≥ 89 years old during study recruitment and lived in zip codes near (within 3 hours driving distance) one of the three U.S field centers. The University of Southern Denmark used the Danish National Register of Persons to identify individuals who were ≥ 90 years during the study recruitment period²⁹. Only families with a Family Longevity Selection Score (FLOSS) of 7 or higher³⁰ who had the proband, at least one living sibling, and one of their living offspring (minimum family size of 3) who could give informed consent and were willing to participate in the interview and examination including donating a blood sample were eligible to participate in this study. This strategy led to the enrollment of families with the greatest potential utility for phenotypic and genetic studies of exceptional survival in families. The interviews and examinations were conducted in the home setting with portable equipment by centrally trained and certified research assistants using a common standardized protocol. Lung function was measured with a portable spirometer (EasyOne™, NDD Medical Technologies, Andover, MA) using American Thoracic Society guidelines. After excluding participants with non-European Ancestry (n=6), participants with poor quality spirometry readings (n=295), pulmonary fibrosis (n=11) and those with lung volume reducing surgery (n=14) and those with missing genotypes (n=344), 3889 participants were included in the present analysis.

LLFS Genotyping: The Illumina Human Omni chip 2.5 v1, was used to genotype all the LLFS participants at the Center for Inherited Disease Research (CIDR). A threshold call rate of > 98% per marker, identified 2,134,578 SNPs. Principal components (PCs), for controlling for population structure, were produced with EIGENSTRAT³¹ on 1,522 LLFS unrelated individuals using 116,867 tag SNPs, where in advance any SNPs with MAF < 5%, HWE $p < 1e-6$, and with missing genotypes were excluded. Imputations were performed based on the cosmopolitan phased haplotypes of 1000 Human Genome (1000HG, version 2010-11 data freeze, 2012-03-04 haplotypes) using MACH and MINIMACH^{6,32} and a total of 38.05 million SNPs were imputed.

Details of the **Multi-Ethnic Study of Atherosclerosis** (MESA) and MESA Lung Study have been previously reported^{33,34} and a full technical report of MESA procedures is available online (www.mesa-nhlbi.org). Spirometry was conducted in 2004-2006 in accordance with the 2005 ATS/ERS recommended guidelines¹²⁷ using a dry-rolling-sealed spirometer with automated quality control software (Occupational Marketing, Inc. Houston, TX). All spirometry exams were centrally reviewed by at least one (JLH) of the authors³⁵.

The **Northern Finland Birth Cohort** (NFBC) study programme was initiated in the 1960s. The cohort of women and their offspring was established in the provinces of Oulu and Lapland and had an expected date of birth in 1966 comprising 12231 children (NFBC1966, 15). The NFBC1966 had spirometry and other measurements done at the age of 31 years. In NFBC1966, we used a Vitalograph P-model spirometer (Vitalograph Ltd., Buckingham, UK), with a volumetric accuracy of $\pm 2\%$ or ± 50 mL whichever was greater. The spirometer was calibrated regularly using a 1-Litre precision syringe. The spirometric manoeuvre was performed three times but was repeated if the coefficient of variation between two maximal readings was $>4\%$.

NSPHS

The **Northern Sweden Population Health Study** (NSPHS) represents a cross-sectional study conducted in the community of Karesuando in the subarctic region of the County of Norrbotten, Sweden, in 2006. Spirometry was performed in sitting position without noseclips using a MicroMedicalSpida 5 spirometer (<http://www.medisave.co.uk>). Three consecutive 28 lung function measurements per participant were done and the maximum value per measured lung function parameter was used for further analysis. Relatedness was taken into account by applying the "polygenic" linear mixed effects model. Genome-wide association analysis was performed using a score test, a family-based association test¹⁷ which uses the residuals and the variance-covariance matrix from the polygenic model and the SNP fixed effect coded under an additive model.

The **Orkney Complex Disease Study** (ORCADES) is an ongoing family-based, cross-sectional study in the isolated Scottish archipelago of Orkney. Spirometry was performed in the sitting position without nose clips, using a Spida handheld spirometer. Measurements were repeated once and the better reading was used for analysis. Population stratification was partially taken into account after adjusting the measures for the 3 first principal components drawn from the population genomic kinship matrix computed using the *ibs* function of the GenABEL¹⁸. Additional correction of the standard error of effect estimates was undertaken by genomic control¹⁹. Effect estimates were obtained using the *palinear* function of the ProbABEL package.

The **Rotterdam Study** is a prospective population-based cohort study founded in 1990 in a suburb of Rotterdam, the Netherlands. The first cohort (RS-I) consists of 7,983 participants, aged 55 years and over. The second cohort (RS-II) was recruited in 2000 with the same inclusion criteria. The third cohort (RS-III) consists of 3,932 participants, aged 45 years and over and was recruited in 2006. Performing of spirometry was introduced in 2004. Spirometry was performed by trained paramedical personnel using a SpiroPro® portable spirometer (Erich Jaeger, Hoechberg, Germany), according to American Thoracic Society (ATS)/European Respiratory Society (ERS) guidelines. FEV₁, FVC and FEV₁/FVC ratio were measured. A total of 6,240 subjects were genotyped in RS I, a total of 2,516 in RS II and a total of 2,420 subjects were genotyped in RS III. Exclusions were based on a call rate $< 98\%$, Hardy-Weinberg *p*-value $< 10^{-6}$ and MAF $< 0.01\%$. A total of 5,974 for RS I, 2,157 for RS II and 2,082 for RS III passed genotyping quality control. Since spirometry was introduced in only 2004, full data with regard to spirometry and covariate data was available in 1,224 participants, 852 for RS II and 1,247 for RS III and were used in the final analysis dataset. Further details can be found elsewhere³⁶.

The **Study of Health in West Pomerania** (SHIP) is a cross-sectional, population based survey in a region in the Northeast of Germany. Study details are given elsewhere^{37,38}. The examinations were conducted using a bodyplethysmograph equipped with a pneumotachograph (VIASYS Healthcare, JAEGER, Hoechberg, Germany) which meets the American Thoracic Society (ATS) criteria³⁹. The volume signal of the equipment was calibrated with a 3.0 litre syringe connected to the pneumotachograph in accordance with the manufacturer's recommendations and at least once on each day's testing. Barometric pressure, temperature and relative humidity were registered every morning. Calibration of reference gas and volume was examined under ATS-conditions (Ambient

Temperature Pressure) and the integrated volumes were BTPS (Body Temperature Pressure Saturated) corrected^{39,40}. Lung function variables were measured continuously throughout the baseline breathing and the forced manoeuvres using a VIASYS HEALTHCARE system (MasterScreen Body/Diff.). Spirometry flow volume loops were conducted in accordance with ATS recommendations⁴⁰ in a sitting position and with wearing noseclips. The participants performed at least three forced expiratory lung function manoeuvres in order to obtain a minimum of two acceptable and reproducible values⁴¹. Immediate on-screen error codes indicating the major acceptability (including start, duration and end of test) and reproducibility criteria supported the attempt for standardised procedures. The procedure was continuously monitored by a physician. The best results for forced vital capacity (FVC), forced expiratory volume in one second (FEV₁), peak expiratory flow (PEF) and expiratory flow at 75%, 50%, 25% of FVC (MEF 75, MEF 50, MEF 25) were taken. The ratio of FEV₁ to FVC was calculated from the largest FEV₁ and FVC.

The **Twins UK** cohort consisted of a group of twins ascertained to study the heritability and genetics of age-related diseases (www.twinsUK.ac.uk). These unselected twins were recruited from the general population through national media campaigns in the UK and shown to be comparable to age-matched population singletons in terms of disease-related and lifestyle characteristics⁴². Spirometry (Vitalograph model 2150, Buckingham, England) was conducted at the clinical center during a visit. Twins were instructed before the test and forced vital capacity (FVC) manoeuvres were performed in a standing position, without the use of nose clips. Three manoeuvres were performed and the maximum obtained values for FEV₁ were obtained. Because of the relatedness in the TwinsUK cohort, we utilized the GenABEL software package¹⁸ which incorporates a pair-wise kinship matrix calculated using genotyping data in the polygenic model to correct for relatedness and hidden population stratification. The score test implemented in the software was used to test the association between a given SNP and the lung function phenotypes. Genotyping was done in two different platforms. These were Illumina's Human Hap 300k Duo for part of the UK Twin Cohort, Illumina Human Hap610W Quad for the rest of the UK Twin Cohort. We excluded SNPs that had a low call rate (<95%), Hardy-Weinberg p values < 10⁻⁴ and minor allele frequencies < 1%. We also removed subjects if the sample call rate was less than 95%, autosomal heterozygosity was outside the expected range, genotype concordance was over 97% with another sample and the sample was of lesser call rate, non-Caucasian ancestry either self-identified or identified by PCA cluster analysis by comparison to the three HapMap phase 2 reference populations (CEU, YRI, CHB+JPT), or unexplained relatedness (estimated proportion of allele shared identical by descent >0.05). The overall genotyping efficiency of the GWA was 98.7 %. Imputation of genotypes was carried out using the software IMPUTE⁴³.

Cohorts included in stage 2

The **Busselton Health Study** (BHS) is a longitudinal survey of the town of Busselton in the south-western region of Western Australia that began in 1966. In 1994/1995 a cross-sectional community follow-up study was undertaken where blood was taken for DNA extraction. A sample of 1,168 European-ancestry individuals were genotyped using the Illumina 610-Quad BeadChip (BHS1), and subsequent genotyping was carried out on an independent group of 3,038 European-ancestry individuals (BHS2). Spirometric measures of forced expired volume in one second (FEV₁) and forced vital capacity (FVC) were assessed as described previously^{44,45}.

The **CROATIA-Split** study, is an ongoing population-based, cross-sectional study in the Dalmatian City of Split for which 499 of the examinees, aged 18-95, have genotype data available. Genotyping was performed using the Illumina 370CNV array. Genotypes were determined using the GenomeStudio clustering algorithm. Individual samples with a call rate of <97% and SNPs with a call rate of <98% were excluded from analyses. Spirometry was performed in the sitting upright position without nose clips, using a Jaeger Toennisen spirometer. Three readings were taken for each examinee, at least 15 seconds apart. Population stratification was partially taken into account after adjusting the measures

for the 3 first principal components drawn from the population genomic kinship matrix computed using the `ibs` function of the GenABEL package. Additional correction of the standard error of effect estimates due to related individuals was undertaken using the `mmscore` function in GenABEL. Effect estimates were obtained using the `palinear` function of the ProbABEL package.

The Generation Scotland: Scottish Family Health Study is a collaboration between the Scottish Universities and the NHS, funded by the Chief Scientist Office of the Scottish Government. GS:SFHS is a family-based genetic epidemiology cohort with DNA, other biological samples (serum, urine and cryopreserved whole blood) and socio-demographic and clinical data from ~24,000 volunteers, aged 18-98 years, in ~7,000 family groups. Participants were recruited across Scotland, with some family members from further afield, from 2006 - 2011. Most (87%) participants were born in Scotland and 96% in the UK or Ireland. The cohort profile has been published³⁶. GS:SFHS operates under appropriate ethical approvals, and all participants gave written informed consent. Spirometry was performed without nose clips using the Ndd Easy One Spirometer (Model 2001). Spirometry was performed a maximum of 4 times until satisfactory readings were obtained, all readings with a best test variation of greater than 10% were excluded. GS:SFHS was genotyped using the Illumina Omni Express + Exome Array Chip. QC was performed separately for the OMNI Express data (Individual Call Rate 97%, SNP Call Rate 98%, MAF 0.01, HWE 1e-6) and the Exome Array data (Individual Call Rate 99%, SNP Call Rate 98%, MAF 0.0001, HWE 1e-6) then combined prior to imputation. Prephasing was performed using Shapelt v2 and imputation using IMPUTE2 and the "ALL (Phase 1 integrated release v3, April 2012)" reference panel. GWAS analysis of the imputed dosages was undertaken using the `mmscore` function in ProbABEL which accounts for relatedness.

The **KORA F4** study (Cooperative Health Research in the Region of Augsburg) is a follow-up study to the KORA-Survey 2000 (S4, 10/1999 – 7/2001)²⁸. Lung function tests were performed in a subsample of the KORA F4 cohort corresponding to a random population sample of subjects born between 1946 and 1965 (age range 41 – 63 years, n=1,321). Spirometry was performed in line with the ATS/ERS recommendations⁴⁶ using a pneumotachograph-type spirometer (Masterscreen PC, CardinalHealth, Würzburg, Germany) before and after inhalation of 200 µg salbutamol. The spirometer was calibrated daily using a calibration pump (Cardinal Health, Würzburg, Germany), and additionally, an internal control (examiner) was used to ensure constant instrumental conditions. The participants performed at least three forced expiratory lung function manoeuvres in order to obtain a minimum of two acceptable and reproducible values. The final values of FVC and FEV₁ were determined based on the best maneuver performed before bronchodilation as defined by the highest sum of FVC and FEV₁. For participants who did not manage to perform at least two acceptable and reproducible values out of nine trials no lung function measurements were recorded. Genotyping was performed in 1,814 randomly selected KORA F4 participants using the Affymetrix Human SNP Array 6.0. Hybridisation of genomic DNA was done in accordance with the manufacturer's standard recommendations. Genotypes were determined using the Birdseed2 clustering algorithm. For quality control purposes we applied a positive control and a negative control DNA every 96 samples. On a chip level only subjects with overall genotyping efficiencies of at least 93% were included. In addition the called gender had to agree with the gender in the KORA study database. Imputation of genotypes was performed with Impute v0.4.2 based on HapMap II CEU reference panels. For 903 individuals genotype information and spirometry was available. The statistical analyses were performed in R (www.r-project.org).

LifeLines is an observational follow-up study in a large representative sample of the population of the northern provinces of the Netherlands covering three generations⁴⁷. Firstly, a random sample of persons aged between 25 and 50 years are invited to participate. Subsequently their family members if present are invited to also take part (parents, partner, parents in law, children), resulting in a three-generation study. The core of the LifeLines project consists of dedicated data collection and biological sample storage, including genetic samples ("biobank"). All participants receive a number

of questionnaires and a basic medical examination and are followed for many years with extensive standardized measurements. Spirometry was performed following ATS guidelines using a Welch Allyn Version 1.6.0.489, PC-based SpiroPerfect with Ca Workstation software. Genotyping was performed on 3078 unrelated individuals with IlluminaCytoSNP-12 arrays⁴⁸.

Long Life Family Study (LLFS) is a family-based cohort study that enrolled 4559 long-lived probands and their siblings (n=1,445), their offspring (n=2,329) and spousal controls (n=785) recruited from 3 U.S. field centers (Boston University Medical Center in Boston MA, Columbia College of Physicians and Surgeons in New York City NY, and the University of Pittsburgh in Pittsburgh PA) and the University of Southern Denmark. Participants were chosen from the Center for Medicare and Medicaid Services lists of Medicare enrollees who were ≥ 89 years old during study recruitment and lived in zip codes near (within 3 hours driving distance) one of the three U.S. field centers. The University of Southern Denmark used the Danish National Register of Persons to identify individuals who were ≥ 90 years during the study recruitment period²⁹. Only families with a Family Longevity Selection Score (FLOSS) of 7 or higher³⁰ who had the proband, at least one living sibling, and one of their living offspring (minimum family size of 3) who could give informed consent and were willing to participate in the interview and examination including donating a blood sample were eligible to participate in this study. This strategy led to the enrollment of families with the greatest potential utility for phenotypic and genetic studies of exceptional survival in families. The interviews and examinations were conducted in the home setting with portable equipment by centrally trained and certified research assistants using a common standardized protocol. Lung function was measured with a portable spirometer (EasyOne™, NDD Medical Technologies, Andover, MA) using American Thoracic Society guidelines. After excluding participants with non-European Ancestry (n=6), participants with poor quality spirometry readings (n=295), pulmonary fibrosis (n=11) and those with lung volume reducing surgery (n=14) and those with missing genotypes (n=344), 3,889 participants were included in the present analysis. **LLFS Genotyping:** The Illumina Human Omni chip 2.5 v1, was used to genotype all the LLFS participants at the Center for Inherited Disease Research (CIDR). A threshold call rate of $> 98\%$ per marker, identified 83,774 SNPs. Principal components (PCs), for controlling for population structure, were produced with EIGENSTRAT³¹ on 1,522 LLFS unrelated individuals using 116,867 tag SNPs, where in advance any SNPs with $MAF < 5\%$, $HWE\ p < 1e-6$, and with missing genotypes were excluded. Imputations were performed based on the cosmopolitan phased haplotypes of 1000 Human Genome (1000HG, version 2010-11 data freeze, 2012-03-04 haplotypes) using MACH and MINIMACH^{6,32} and a total of 38.05 million SNPs were imputed.

The **LBC1936** consists of 1,091 relatively healthy individuals assessed on cognitive and medical traits at about 70 years of age. They were all born in 1936 and most took part in the Scottish Mental Survey of 1947. At baseline the sample of 548 men and 543 women had a mean age 69.6 years (SD = 0.8). They were all Caucasian, community-dwelling, and almost all lived in the Lothian region (Edinburgh city and surrounding area) of Scotland. A full description of participant recruitment and testing can be found elsewhere⁴⁹. Genotyping was performed at the Wellcome Trust Clinical Research Facility, Edinburgh. Quality control measures were applied and 1005 participants remained. Lung function assessing peak expiratory flow rate, forced expiratory volume in 1 s, and forced vital capacity (each the best of three), using a Micro Medical Spirometer was assessed, sitting down without nose clips, at age 70 years. The accuracy of the spirometer is $\pm 3\%$ (to ATS recommendations Standardisation of Spirometry 1994 update for flows and volumes).

The Prospective Investigation of the Vasculature in Uppsala Seniors (**PIVUS**)⁵⁰ is a population-based study of the cardiovascular health in the elderly. The main purpose of PIVUS was to investigate the role of endothelial function in cardiovascular risk. Mailed invitations were sent to subjects who lived in Uppsala, Sweden, within 2 months after their 70th birthday. The subjects were randomly selected from the community register. A total of 1,016 men and women participated in the baseline

investigation (participation rate, 50.1%). Spirometry was performed in 901 subjects at baseline in accordance with American Thoracic Society recommendations (α spirometer; Vitalograph Ltd; Buckingham, UK). The best value from three recordings was used. The Ethics Committee of the University of Uppsala approved the study, and the participants gave their informed consent. Genotyping of all samples was undertaken using the Illumina Omni Express and CardioMetabochip. Genotypes were called using GENCALL. A total of 738,879 SNPs passed quality control (thresholds: call rate < 0.95 , and call rate < 0.99 for $MAF < 5\%$; HWE p -value $< 10^{-6}$). SNPs with $MAF < 1\%$ were removed from the imputation scaffold. Imputation was performed using IMPUTE up to CEU haplotypes from Phase II HapMap reference panel.

The **Twins UK** cohorts **II** and **III** consisted of a group of twins ascertained to study the heritability and genetics of age-related diseases (www.twinsUK.ac.uk). These unselected twins were recruited from the general population through national media campaigns in the UK and shown to be comparable to age-matched population singletons in terms of disease-related and lifestyle characteristics⁴². Spirometry (Vitalograph model 2150, Buckingham, England) was conducted at the clinical centre during a visit. Twins were instructed before the test and forced vital capacity (FVC) manoeuvres were performed in a standing position, without the use of nose clips. Three manoeuvres were performed and the maximum obtained values for FEV1 were obtained. Because of the relatedness in the TwinsUK cohort, we utilized the GenABEL software package¹⁸ which incorporates a pair-wise kinship matrix calculated using genotyping data in the polygenic model to correct for relatedness and hidden population stratification. The score test implemented in the software was used to test the association between a given SNP and the lung function phenotypes. Genotyping was done in two different platforms. These were Illumina's Human Hap 300k Duo for part of the UK Twin Cohort, Illumina Human Hap610W Quad for the rest of the UK Twin Cohort. We excluded SNPs that had a low call rate ($< 95\%$), Hardy-Weinberg p values $< 10^{-4}$ and minor allele frequencies $< 1\%$. We also removed subjects if the sample call rate was less than 95%, autosomal heterozygosity was outside the expected range, genotype concordance was over 97% with another sample and the sample was of lesser call rate, non-Caucasian ancestry either self-identified or identified by PCA cluster analysis by comparison to the three HapMap phase 2 reference populations (CEU, YRI, CHB+JPT), or unexplained relatedness (estimated proportion of allele shared identical by descent > 0.05). The overall genotyping efficiency of the GWA was 98.7 %. Imputation of genotypes was carried out using the software IMPUTE⁴³ for both Twins UK-II and III.

CARe Study Description

Study Samples

The Candidate-gene Association Resource (CARe) consortium has been previously described⁵¹. This study includes 5 of the NHLBI CARe cohorts that measured spirometry and genotyped participants using the Affymetrix Genome-Wide Human SNP Array 6.0 platform in African-American participants: Atherosclerosis Risk in Communities (ARIC), Coronary Artery Risk Development in young Adults (CARDIA), Cleveland Family Study (CFS), Cardiovascular Health Study (CHS), Jackson Heart Study (JHS) and the subset of MESA with spirometry. These cohorts have been previously described^{10,13,33,34,52-54}.

Phenotypic Measures

Spirometry: Pre-bronchodilator spirometry was performed by trained and certified spirometry technicians in accordance with the American Thoracic Society guidelines. Spirometry methods and equipment were highly standardized and in some cases identical across cohorts.

CARe Genotyping:

The Affymetrix Genome-Wide Human SNP Array 6.0 Chip, was used to genotype all the CARe participants at the Broad Institute of MIT and Harvard. A threshold call rate of > 95% per marker, identified 768,780 SNPs. Principal components (PCs), for controlling for population structure, were produced with EIGENSTRAT³¹ on 5,951 CARe unrelated individuals, where in advance any SNPs with MAF < 1% and SNPs with genotyping success rate < 95%.

Korean Studies

The Healthy Twin Study, Korea

The Healthy Twin Study (HT) is a prospective multi-center cohort study based on a large nation-wide family database. The HT has recruited 3,690 individuals, with 815 pairs of adult like-sex twins over age 30 and their first-degree families as of 2013 February. Being a twin was the only criteria for participation, and the participants were not selected by any specific health status. The protocols and measurements are described in detail in previous report⁵⁵⁻⁵⁷. Extended questionnaire and health examination are provided at recruitment and follow-up study is ongoing in every 3 year, and the 3rd wave survey will be finished in 2014. Spirometry was performed by trained paramedical personnel using a Vmax22® (Sensormedics, Yorba Linda, California, USA), according to American Thoracic Society (ATS)/European Respiratory Society (ERS) guidelines. FEV1, FEV5, FVC, FEV1/FVC ratio, FEV1/FVC ratio, FEF 50%, FEF 25-75%, FEF 200-1200 ml, PEF, and FET 100% were measured. Genotyping of the HT was performed using two platforms. First 1,861 individuals were analyzed by Affymetrix Genome Wide Human SNP Array 6.0 (Affymetrix, Inc., Santa Clara, CA, USA). Genotyped markers with a Hardy-Weinberg Equilibrium ($p < 0.001$), low call rate ($< 95\%$), and low minor allele frequency ($MAF < 0.01$) were excluded. In addition to the conventional quality control, Mendelian and non-Mendelian errors mimicking double-recombinations were detected by within-family comparisons using Merlin. This error-checking step further detected and deleted 63,777 erroneous markers, leaving 516,452 SNPs. Second, 1,047 individuals were analyzed using Illumina HumanBeadCore chip with around 300K SNPs, and the same quality control method and criteria is applied (in process).

The KARE3 (Korea Association Resource 3)

The Korea Association Resource (KARE) project was initiated in 2007 to undertake a large-scale GWA analysis among the 10,038 participants (aged between 40 and 69) of the Ansung ($n = 5,018$), a rural

area-based, and Ansan (n = 5,020) an industrialized mid-size city-based cohorts. These cohorts, established as part of the Korean Genome Epidemiology Study (KoGES) in 2001 measured epidemiologic and clinical data. Detailed protocols were described in previous report⁵⁸. Lung function measures were taken by a skilled technician using a portable spirometer (Vmax-2130, Sensor Medics, Yorba Linda, CA, USA) according to standardized protocols of the American Thoracic Society (ATS). FEV1, FVC and FEV1/FVC ratio were measured. For 8,842 individuals, genotyping using the Affymetrix Genome-wide Human SNP array 5.0 (Affymetrix Inc., Santa Clara, California, USA), with 500,568 SNP markers. Genotype call rates <5% (n=45,343), deviations from the Hardy-Weinberg equilibrium (HWE) test $p < 0.0001$ (n=35,410), or a minor allele frequency (MAF) lower than 5% (n=147,570) were excluded leaving 312,381 (Ansung) and 313,984 (Ansan) SNP markers.

KARE3 and the Twin family study were imputed using IMPUTE2. Both HapMap3 phase2 (JPT+CHB;<http://hapmap.ncbi.nlm.nih.gov/>) and Korean HapMap (<http://www.khapmap.org/>) panel data were combined to serve as the reference. For the HT, family-based imputation was further performed to check possible errors using Beagle. Imputation quality score ($0 \leq r^2 \leq 1$) was checked, and the r^2 cutoff for post-imputation SNP filtering was 0.5. The average quality score was 0.99 for the imputed markers.

References

1. Repapi, E. et al. Genome-wide association study identifies five loci associated with lung function. *Nat Genet* **42**, 36-44 (2010).
2. 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).
3. Soler Artigas, M. et al. Genome-wide association and large-scale follow up identifies 16 new loci influencing lung function. *Nat Genet* **43**, 1082-90 (2011).
4. Melen, E. et al. Expression analysis of asthma candidate genes during human and murine lung development. *Respir Res* **12**, 86 (2011).
5. 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).
6. Li, Y., Willer, C., Sanna, S. & Abecasis, G. Genotype imputation. *Annu Rev Genomics Hum Genet* **10**, 387-406 (2009).
7. Li, J. & Ji, L. Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix. *Heredity (Edinb)* **95**, 221-7 (2005).
8. Nyholt, D.R. A simple correction for multiple testing for single-nucleotide polymorphisms in linkage disequilibrium with each other. *Am J Hum Genet* **74**, 765-9 (2004).
9. Yang, J. et al. Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat Genet* **44**, 369-75, S1-3 (2012).
10. McGarvey, L.P. et al. Ascertainment of cause-specific mortality in COPD: operations of the TORCH Clinical Endpoint Committee. *Thorax* **62**, 411-5 (2007).
11. Strachan, D.P. et al. Lifecourse influences on health among British adults: effects of region of residence in childhood and adulthood. *Int J Epidemiol* **36**, 522-31 (2007).
12. Marossy, A.E., Strachan, D.P., Rudnicka, A.R. & Anderson, H.R. Childhood chest illness and the rate of decline of adult lung function between ages 35 and 45 years. *Am J Respir Crit Care Med* **175**, 355-9 (2007).
13. Friedman, G.D. et al. CARDIA: study design, recruitment, and some characteristics of the examined subjects. *J Clin Epidemiol* **41**, 1105-16 (1988).
14. Hughes, G.H. et al. Recruitment in the Coronary Artery Disease Risk Development in Young Adults (Cardia) Study. *Control Clin Trials* **8**, 68S-73S (1987).
15. ATS statement--Snowbird workshop on standardization of spirometry. *Am Rev Respir Dis* **119**, 831-8 (1979).
16. Fried, L.P. et al. The Cardiovascular Health Study: design and rationale. *Ann Epidemiol* **1**, 263-76 (1991).
17. Buist, A.S. Standardization of spirometry. *Am Rev Respir Dis* **136**, 1073-4 (1987).
18. Aulchenko, Y.S., Ripke, S., Isaacs, A. & van Duijn, C.M. GenABEL: an R library for genome-wide association analysis. *Bioinformatics* **23**, 1294-6 (2007).
19. Amin, N., van Duijn, C.M. & Aulchenko, Y.S. A genomic background based method for association analysis in related individuals. *PLoS One* **2**, e1274 (2007).
20. Burney, P.G., Luczynska, C., Chinn, S. & Jarvis, D. The European Community Respiratory Health Survey. *Eur Respir J* **7**, 954-60 (1994).
21. European Community Respiratory Health Survey, I.I.S.C. The European Community Respiratory Health Survey II. *Eur Respir J* **20**, 1071-9 (2002).
22. Moffatt, M.F. et al. A large-scale, consortium-based genomewide association study of asthma. *N Engl J Med* **363**, 1211-21 (2010).
23. Chinn, S. et al. Smoking cessation, lung function, and weight gain: a follow-up study. *Lancet* **365**, 1629-35; discussion 1600-1 (2005).
24. Canoy, D. et al. Abdominal obesity and respiratory function in men and women in the EPIC-Norfolk Study, United Kingdom. *Am J Epidemiol* **159**, 1140-9 (2004).
25. Cox, B., Huppert, F. & Whichelow, M. *The Health and Lifestyle Survey: seven years on*, (Dartmouth Publishing Company, Aldershot, United Kingdom, 1993).

26. Kaprio, J., Pulkkinen, L. & Rose, R.J. Genetic and environmental factors in health-related behaviors: studies on Finnish twins and twin families. *Twin Res* **5**, 366-71 (2002).
27. Heistaro, S. Methodology report. Health 2000 survey. in *Publications of National Public Health Institute* (2000).
28. Wichmann, H.E., Gieger, C., Illig, T. & Group, M.K.S. KORA-gen--resource for population genetics, controls and a broad spectrum of disease phenotypes. *Gesundheitswesen* **67 Suppl 1**, S26-30 (2005).
29. Pedersen, C.B., Gotzsche, H., Moller, J.O. & Mortensen, P.B. The Danish Civil Registration System. A cohort of eight million persons. *Dan Med Bull* **53**, 441-9 (2006).
30. Sebastiani, P. et al. A family longevity selection score: ranking sibships by their longevity, size, and availability for study. *Am J Epidemiol* **170**, 1555-62 (2009).
31. Price, A.L. et al. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* **38**, 904-9 (2006).
32. 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**, 816-34 (2010).
33. Bild, D.E. et al. Multi-ethnic study of atherosclerosis: objectives and design. *Am J Epidemiol* **156**, 871-81 (2002).
34. Rodriguez, J. et al. The association of pipe and cigar use with cotinine levels, lung function, and airflow obstruction: a cross-sectional study. *Ann Intern Med* **152**, 201-10 (2010).
35. Hankinson, J.L. et al. Performance of American Thoracic Society-recommended spirometry reference values in a multiethnic sample of adults: the multi-ethnic study of atherosclerosis (MESA) lung study. *Chest* **137**, 138-45 (2010).
36. Smith, B.H. et al. Cohort Profile: Generation Scotland: Scottish Family Health Study (GS:SFHS). The study, its participants and their potential for genetic research on health and illness. *Int J Epidemiol* **42**, 689-700 (2013).
37. John, U. et al. Study of Health In Pomerania (SHIP): a health examination survey in an east German region: objectives and design. *Soz Präventivmed* **46**, 186-94 (2001).
38. Volzke, H. et al. Reference intervals of serum thyroid function tests in a previously iodine-deficient area. *Thyroid* **15**, 279-85 (2005).
39. Nelson, S.B., Gardner, R.M., Crapo, R.O. & Jensen, R.L. Performance evaluation of contemporary spirometers. *Chest* **97**, 288-97 (1990).
40. Standardization of spirometry--1987 update. Official statement of American Thoracic Society. *Respir Care* **32**, 1039-60 (1987).
41. Quanjer, P.H. et al. Lung volumes and forced ventilatory flows. Report Working Party Standardization of Lung Function Tests, European Community for Steel and Coal. Official Statement of the European Respiratory Society. *Eur Respir J Suppl* **16**, 5-40 (1993).
42. Andrew, T. et al. Are twins and singletons comparable? A study of disease-related and lifestyle characteristics in adult women. *Twin Res* **4**, 464-77 (2001).
43. 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**, 906-13 (2007).
44. James, A.L. et al. Decline in lung function in the Busselton Health Study: the effects of asthma and cigarette smoking. *Am J Respir Crit Care Med* **171**, 109-14 (2005).
45. Ryan, G. et al. Decline in lung function and mortality: the Busselton Health Study. *J Epidemiol Community Health* **53**, 230-4 (1999).
46. Miller, M.R. et al. Standardisation of spirometry. *Eur Respir J* **26**, 319-38 (2005).
47. Stolk, R.P. et al. Universal risk factors for multifactorial diseases: LifeLines: a three-generation population-based study. *Eur J Epidemiol* **23**, 67-74 (2008).
48. Dolmans, G.H. et al. Wnt signaling and Dupuytren's disease. *N Engl J Med* **365**, 307-17 (2011).
49. Deary, I.J. et al. The Lothian Birth Cohort 1936: a study to examine influences on cognitive ageing from age 11 to age 70 and beyond. *BMC Geriatr* **7**, 28 (2007).
50. Lind, L., Fors, N., Hall, J., Marttala, K. & Stenborg, A. A comparison of three different methods to evaluate endothelium-dependent vasodilation in the elderly: the Prospective Investigation

- of the Vasculature in Uppsala Seniors (PIVUS) study. *Arterioscler Thromb Vasc Biol* **25**, 2368-75 (2005).
51. Musunuru, K. et al. Candidate gene association resource (CARE): design, methods, and proof of concept. *Circ Cardiovasc Genet* **3**, 267-75 (2010).
 52. Buxbaum, S.G., Elston, R.C., Tishler, P.V. & Redline, S. Genetics of the apnea hypopnea index in Caucasians and African Americans: I. Segregation analysis. *Genet Epidemiol* **22**, 243-53 (2002).
 53. Taylor, H.A., Jr. The Jackson Heart Study: an overview. *Ethn Dis* **15**, S6-1-3 (2005).
 54. Carpenter, M.A. et al. Laboratory, reading center, and coordinating center data management methods in the Jackson Heart Study. *Am J Med Sci* **328**, 131-44 (2004).
 55. Sung, J. et al. The Korean Twin Registry--methods, current stage, and interim results. *Twin Res* **5**, 394-400 (2002).
 56. Sung, J. et al. Healthy Twin: a twin-family study of Korea--protocols and current status. *Twin Res Hum Genet* **9**, 844-8 (2006).
 57. Gombojav, B. et al. The Healthy Twin Study, Korea updates: resources for omics and genome epidemiology studies. *Twin Res Hum Genet* **16**, 241-5 (2013).
 58. 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**, 527-34 (2009).