ORIGINAL ARTICLE

Lung Development Genes and Adult Lung Function

Laura Portas[1]*, Miguel Pereira[1,2]*, Seif O. Shaheen[3], Annah B. Wyss[4], Stephanie J. London[4], Peter G. J. Burney[1], Matthew Hind[1,5‡], Charlotte H. Dean[1,6‡], and Cosetta Minelli[1‡]

[1]National Heart and Lung Institute, Imperial College London, London, United Kingdom; [2]Congenica Ltd., Wellcome Genome Campus, Cambridge, United Kingdom; [3]Institute of Population Health Sciences, Queen Mary University of London, London, United Kingdom; [4]Department of Health and Human Services, National Institute of Environmental Health Sciences, NIH, Research Triangle Park, North Carolina; [5]Department of Respiratory Medicine, Royal Brompton & Harefield NHS Foundation Trust, London, United Kingdom; and [6]MRC Harwell Institute, Oxfordshire, United Kingdom

ORCID IDs: 0000-0003-1789-1893 (L.P.); 0000-0002-7830-2717 (M.P.); 0000-0002-7273-8691 (S.O.S.); 0000-0002-9756-5205 (A.B.W.); 0000-0003-4911-5290 (S.J.L.); 0000-0001-8635-5678 (P.G.J.B.); 0000-0002-9966-9446 (M.H.); 0000-0002-8846-5472 (C.H.D.); 0000-0001-9166-3958 (C.M.).

Abstract

Rationale: Poor lung health in adult life may occur partly through suboptimal growth and development, as suggested by epidemiological evidence pointing to early life risk factors.

Objectives: To systematically investigate the effects of lung development genes on adult lung function.

Methods: Using UK Biobank data, we tested the association of 391 genes known to influence lung development with FVC and FEV_1/FVC. We split the dataset into two random subsets of 207,616 and 138,411 individuals, using the larger subset to select the most promising signals and the smaller subset for replication.

Measurements and Main Results: We identified 55 genes, of which 36 (16 for FVC, 19 for FEV_1/FVC, and one for both) had not been identified in the largest, most recent genome-wide study of lung function. Most of these 36 signals were intronic variants; expression data from blood and lung tissue showed that the majority affect the expression of the genes they lie within. Further testing of 34 of these 36 signals in the CHARGE and SpiroMeta consortia showed that 16 replicated after Bonferroni correction and another 12 replicated at nominal significance level. Of the 55 genes, 53 fell into four biological categories whose function is to regulate organ size and cell integrity (growth factors; transcriptional regulators; cell-to-cell adhesion; extracellular matrix), suggesting that these specific processes are important for adult lung health.

Conclusions: Our study demonstrates the importance of lung development genes in regulating adult lung function and influencing both restrictive and obstructive patterns. Further investigation of these developmental pathways could lead to druggable targets.

Keywords: genetic association study; UK Biobank; FVC; FEV_1/FVC; COPD

(Received in original form December 6, 2019; accepted in final form May 11, 2020)

*Co–first authors.

‡Co–senior authors.

M.P. was funded by the National Heart and Lung Institute Foundation. C.H.D. and M.H. are supported by the Royal Brompton and Harefield Hospitals Charity. M.H. is also supported by an award from Mr. and Mrs. Youssef Mansour. A.B.W. and S.J.L. are supported by the Intramural Research Program of the NIH, National Institute of Environmental Health Sciences NIH ZO1 ES43012. A.B.W. is also supported by contract no. HHSN273201600003I. Infrastructure for the CHARGE Consortium is supported by the NHLBI grant R01HL105756.

Author Contributions: L.P., M.P., M.H., C.H.D., and C.M. designed the study. L.P. and M.P. performed the statistical analyses. S.O.S., M.H., C.H.D., and C.M. wrote the manuscript. P.G.J.B. contributed to the interpretation of the data. A.B.W. and S.J.L. contributed to the replication of the findings. All authors contributed to and approved the final version of the manuscript.

Correspondence and requests for reprints should be addressed to Cosetta Minelli, M.D., Ph.D., National Heart and Lung Institute, Imperial College London, Emmanuel Kaye Building, 1B Manresa Road, London SW3 6LR, UK. E-mail: cosetta.minelli1@imperial.ac.uk.

This article has a related editorial.

This article has an online supplement, which is accessible from this issue's table of contents at www.atsjournals.org.

Am J Respir Crit Care Med Vol 202, Iss 6, pp 853–865, Sep 15, 2020

Originally Published in Press as DOI: 10.1164/rccm.201912-2338OC on May 11, 2020
Internet address: www.atsjournals.org

At a Glance Commentary

Scientific Knowledge on the Subject: Epidemiological studies on early life risk factors suggest that poor lung health in adult life may be partly due to suboptimal growth and development. Although the early environment has been implicated in the etiology of impaired lung function, there has been no systematic investigation of the role of genes known to play a vital role in lung development.

What This Study Adds to the Field: Our findings show a clear effect of lung development genes on adult lung function, influencing both restrictive and obstructive patterns. Further investigation of these developmental pathways could ultimately lead to druggable targets aimed at optimizing adult lung health and preventing chronic obstructive pulmonary disease.

Gaining a full understanding of the genetic and environmental causes of impaired lung function is important if we are to discover ways to prevent chronic obstructive pulmonary disease (COPD) and to optimize lung health. Furthermore, the public health benefits of improving lung function are far-reaching, given that poor lung function, especially a lower FVC, is a powerful predictor of increased mortality, in particular from cardiovascular disease, even in nonsmokers (1, 2).

A long-standing hypothesis states that low lung function and COPD in late adult life may occur partly through suboptimal growth and development, with failure to attain maximal lung capacity in young adulthood (3–6). There is substantial epidemiological and experimental evidence supporting the concept of the developmental origins of adult lung disease and impaired lung function (7). Epidemiological evidence includes tracking of lung function from early childhood to adulthood, which implicates environmental factors operating early in life (3); various prenatal, perinatal, and postnatal risk factors have been linked to impaired adult lung function, including maternal smoking, low birth weight, prematurity, and respiratory tract infections (4, 5).

Although the early environment has been implicated in the etiology of impaired lung function, there has been no systematic investigation of the role of genes known to play a vital role in lung development. Genetic variants affecting adult cross-sectional lung function have shown little or no effect on longitudinal lung function decline (8), and some of these variants have been identified in children as well as adults. These observations suggest that lung function at a given point in adulthood may be more influenced by genetic factors that affect the developmental trajectory of lung function rather than the rate of subsequent decline. Indeed, lung development gene variants have been identified in genome-wide association studies (GWASs) of lung function (9); some of these have been associated with infant lung function (10), and for others, there is evidence of differential expression during human fetal lung development (11). However, it is likely that other lung development gene variants genuinely associated with lung function may not have achieved the stringent genome-wide significance thresholds (typically 5×10^{-8}) required to protect against false positive findings. Taking a complementary hypothesis-driven approach, here, we investigate 391 genes known to influence lung development for association with adult lung function, in particular FVC and the ratio of FEV_1 to FVC (FEV_1/FVC), using data from the large UK Biobank (UKB) dataset.

This article was previously published in preprint form (https://doi.org/10.1101/447367).

Methods

UKB Data

The UKB is a study of 502,543 volunteer participants aged 39–70 recruited from 22 study centers across the United Kingdom, which collected data on a large number of genetic and nongenetic risk factors for chronic disease and related disease traits (12, 13). We included in our analyses 346,027 individuals of self-reported white ethnicity with available good quality genetic and lung function data, as shown in Figure E1 in the online supplement. For lung function data, we used FVC and FEV_1 "best measure," as proposed in the UK BiLEVE (Biobank Lung Exome Variant Evaluation) study (14). Table E1 provides UKB data field numbers and web links for full descriptions of all variables used in the analyses.

For the genetic data, quality control and genotype imputation were performed by UKB, as previously described (13); we used the genetic dataset made available on July 2017.

Selection of Genes Related to Lung Development

The list of genes related to lung development was prepared by two experts (C.H.D. and M.H.), as previously described (15). An initial list of genes was compiled by each expert separately based on their knowledge from both human and experimental data, including orthologs of genes known to affect lung development in a variety of model organisms. The two lists were compared, and they agreed on a common list. This list was further extended to include relevant additional genes identified based on pathway information from Kyoto Encyclopedia of Genes and Genomes (KEGG) (16) (relevant genes lying in the same pathways as those in the list) and literature data from Human Genome Epidemiology (HuGE) Navigator (17) (genes considered as associated with lung development in previous genetic association studies). In the case of large gene families, if in doubt about which genes to select, we chose those with higher gene expression in fetal lungs, using information from BioGPS (Human U133A/GNF1H Gene Atlas database) (18).

From this list of 403 genes, after excluding genes on the X chromosome, we considered 391 genes (Table E2). Within these genes, 106,384 variants were available in the UKB after the exclusion of variants with minor allele frequency of <0.01 and imputation quality (info score) of <0.5.

Association of Lung Development Genes with Adult Lung Function

We first considered which of the 391 genes were associated with adult lung function in the largest, most recent GWAS by Shrine and colleagues (9); 19 of them were reported either as novel signals or as replications of findings from previous studies (14, 19–27), and their results for FVC and FEV_1/FVC in UKB ($n = 346,027$) are presented in Table E3.

To identify and replicate further associations in the remaining 372 genes, we randomly split the UKB dataset into two subsets of 60% ($n = 207,616$) and 40% ($n = 138,411$) of the total sample. Main participants' characteristics, including lung

function, for the whole study sample and for the two subsets separately are summarized in Table E4. We used the larger subset (stage 1) to select the most promising signals, taking the "best SNP" for each gene (i.e., the SNP with the lowest *P* value, if the *P* value was lower than an arbitrary screening threshold of 1×10^{-3}), and used the smaller subset (stage 2) for replication. In stage 1, we tested all 98,255 variants in the 372 genes; for each gene, we selected the best SNP for replication. In stage 2, we tested all best SNPs and considered as replicated those associations with effect in the same direction as in stage 1 and a one-sided *P* value below a Bonferroni-corrected threshold (0.05 divided by the number of SNPs sent to replication: 102 SNPs, $P < 4.9 \times 10^{-4}$, for FVC; 113 SNPs, $P < 4.4 \times 10^{-4}$, for FEV_1/FVC). The use of Bonferroni correction in stage 2, on which all our inferences are based, fully addresses the issue of multiple testing.

In both stage 1 and stage 2, we estimated the association of each variant with FVC and FEV_1/FVC using linear mixed models as implemented in BOLT-LMM (28), accounting for cryptic relatedness and the fine-scale population structure that can be found within self-reported white ethnicity. The analyses assumed an additive genetic model and were adjusted for age, age^2, sex, height, smoking status (ever vs. never), genotyping array, and assessment center. Adjustment for height ensures the genetic effects on lung function are independent of body size.

For both FVC and FEV_1/FVC, we evaluated whether our replicated SNP for a gene was in linkage disequilibrium (LD) ($r^2 > 0.1$) with the best SNP for a different gene, in which case we performed conditional analyses, mutually adjusting one for the other.

We performed the following three sets of secondary analyses on replicated SNPs: *1*) we assessed their association with spirometrically defined COPD (defined as an FEV_1/FVC below the lower limit of normal [LLN] based on the NHANES (National Health and Nutrition Examination Survey) III study equation for white ethnicity [29]), adjusting the models for the same variables as in the main analyses; *2*) we repeated the main analyses stratified by smoking status; if lung development genes are largely influencing maximal level attained through lung growth, then we might expect stronger associations in nonsmokers; in contrast, if their influence on lung function is through increasing lung repair in response to insults such as smoking, which would affect lung function decline, then we might expect stronger associations in smokers; and *3*) we repeated the main analyses stratifying participants below and above the median age of 58. If a lung development gene affects lung regeneration, we might expect a stronger effect in older people and vice versa, although the age range in the UKB (39–70 yr) limits the extent to which effect modification by age can be investigated in this dataset. To increase the statistical power of these secondary analyses, we performed them on the whole UKB sample ($N = 346{,}027$), which included 35,840 spirometrically defined COPD cases (FEV_1/FVC < LLN) (10.4%) and 211,689 ever-smokers (61.2%).

Using results in individuals of European ancestry from the CHARGE (Cohorts for Heart and Aging Research in Genomic Epidemiology) and SpiroMeta consortia, we further tested for replication signals that had not been reported by Shrine and colleagues (9) either as newly identified or as replicated from previous studies. For CHARGE, we used the results of a GWAS meta-analysis of 18 studies (26), and for SpiroMeta, we used the publicly available results of a GWAS meta-analysis of 22 studies downloaded from the GWAS catalog (www.ebi.ac.uk/gwas/publications/30804560). As in our analyses, all studies in CHARGE and SpiroMeta controlled for age, age^2, sex, height, and smoking status (with additional adjustments in CHARGE for height^2 and smoking pack-years as well as weight for FVC) as well as population stratification and, when necessary, family relatedness and center. We could not use a standard meta-analysis to combine the results from CHARGE and SpiroMeta because the latter used rank-based inverse normal transformation, and we therefore used Fisher's meta-analysis of *P* values. Replication was defined as an effect in the same direction as that in the UKB, with a one-sided *P* value below the Bonferroni-corrected threshold. In Fisher's meta-analysis, *P* values were inverted for estimates in the opposite direction.

Results

Stage 1 results for all 98,255 SNPs in the 372 genes are reported in Table E5 for both FVC and FEV_1/FVC. Taking to stage 2 the best SNP per gene, 102 SNPs with *P* values $< 1 \times 10^{-3}$ were tested for replication for FVC, and 113 were tested for replication for FEV_1/FVC; results for all SNPs tested for replication are reported in Table E6 for both traits.

In conditional analyses adjusting the effect of a replicated SNP for any other replicated SNP in LD with it, we identified three signals (two for FVC and one for FEV_1/FVC) in which the effect disappeared, and these were dropped (Table E6).

We replicated signals in 42 genes (*P* value in stage 2 below the Bonferroni-corrected thresholds of $P < 4.9 \times 10^{-4}$ for FVC or $P < 4.4 \times 10^{-4}$ for FEV_1/FVC). To assess whether these associations might be explained by neighboring genes previously associated with lung function, we repeated the analyses after adjusting for SNPs reported by Shrine and colleagues (9) that were in LD ($r^2 > 0.1$) with the 42 SNPs we had identified. After these conditional analyses, 36 signals remained as independent findings (Table E7), and all further analyses focused on them.

The results for these 36 genes are reported in Tables 1 and 2; of these, 16 were uniquely associated (replication in stage 2) with FVC, 19 were uniquely associated with FEV_1/FVC, and only one signal was associated with both traits. In the secondary analysis testing the association with spirometrically defined COPD, 14 of the 36 genes showed a statistically significant association after Bonferroni correction ($P < 1.4 \times 10^{-3}$), and a further seven showed nominal statistical significance ($P < 0.05$), with the odds ratio for COPD always in a consistent direction with the effect on FEV_1/FVC (lower or higher ratio) (Table 3). In the secondary analysis stratified by smoking, results for FVC and FEV_1/FVC were broadly similar in smokers versus never-smokers (Figures E2 and E3), with no statistically significant interactions after Bonferroni correction. The same was observed for the analysis stratified by age, with results broadly similar and no significant interactions after Bonferroni correction (Figures E4 and E5).

For the external replication in CHARGE and SpiroMeta, 34 of the 36 signals had available data for the SNP or a proxy (LD $r^2 \geq 0.8$). The sample sizes varied across SNPs (Tables 4 and 5) from 108,318 to 143,612 in the meta-analysis of CHARGE and SpiroMeta. Overall, of these

Table 1. UKB Results from Stage 1 and Stage 2 (Internal Replication) for the 17 Signals for FVC*

Gene	SNP	Chr	BP	EA	EAF	Functional Consequence†	Blood eQTL *P* Value	Lung eQTL *P* Value	Stage 1 (*n* = *207,616*)			Stage 2 (Internal Replication) (*n* = *138,411*)		
									β‡	SE	*P* Value	β‡	SE	*P* Value
ACTN3	rs57127845	11	66,318,325	C	0.82	Intron variant	7.4×10^{-6}	0.022	−12.3	2.4	2.5×10^{-7}	−14.7	2.9	2.3×10^{-7}
ACTN4	rs189809900	19	39,147,164	A	0.98	Intron variant	0.112	0.005	−25.7	7.2	3.5×10^{-4}	−35.8	8.9	2.8×10^{-5}
CLDN20	rs34268254	6	155,590,120	TA	0.39	Intron variant	NA	NA	10.2	1.9	4.5×10^{-8}	9.0	2.3	4.6×10^{-5}
GSK3B	rs6805251	3	119,560,606	T	0.38	Intron variant	3.0×10^{-97}	1.1×10^{-5}	7.4	1.9	6.8×10^{-5}	8.0	2.3	2.1×10^{-4}
HOXA1	rs45571645	7	27,135,096	G	0.98	Missense variant	2.7×10^{-5}	0.324	33.3	6.8	9.8×10^{-7}	28.7	8.3	2.9×10^{-4}
HOXB4	rs201603635	17	46,653,038	T	0.94	3′ UTR variant	NA	NA	−12.8	3.8	6.9×10^{-4}	−16.3	4.6	1.9×10^{-4}
KAT8	rs138259061	16	31,136,066	A	0.64	Intron variant	3.5×10^{-70}§	2.8×10^{-23}§	−8.7	1.9	4.1×10^{-6}	−7.7	2.3	4.5×10^{-4}
ITGB5	rs17282078	3	124,481,760	T	0.87	3′ UTR variant	3.2×10^{-6}	0.326	9.1	2.7	7.4×10^{-4}	15.7	3.3	9.0×10^{-7}
MMP24	rs7280	20	33,864,484	A	0.58	3′ UTR variant	2.2×10^{-36}	0.499	10.3	1.8	2.7×10^{-8}	9.7	2.3	9.5×10^{-6}
NCOR2	rs72451021	12	124,811,393	ACT	0.89	Intron variant	6.3×10^{-11}‖	0.429	12.8	2.9	1.2×10^{-5}	12.8	3.6	1.9×10^{-4}
NR3C1	rs72801051	5	142,685,670	A	0.84	Intron variant	1.1×10^{-8}	0.352	12.1	2.4	6.9×10^{-7}	10.7	3.0	1.6×10^{-4}
ROR2	rs12684752	9	94,682,990	T	0.94	Intron variant	1.4×10^{-9}	0.266	−13.8	3.9	3.7×10^{-4}	−17.8	4.7	8.5×10^{-5}
RUNX1	rs12483501	21	36,224,276	T	0.63	Intron variant	0.926	NA	8.9	2.0	9.3×10^{-6}	10.1	2.5	2.1×10^{-5}
SERPINC1	rs2227603	1	173,882,548	A	0.97	Intron variant	0.967	0.776	−21.6	5.4	5.8×10^{-5}	−23.1	6.6	2.5×10^{-4}
SOX9	rs796209434	17	70,122,505	CT	0.53	3′ UTR variant	NA	NA	7.3	1.8	6.2×10^{-5}	7.9	2.2	2.0×10^{-4}
WNT2B	rs351370	1	113,054,659	C	0.41	Intron variant	0.191	6.9×10^{-4}	7.0	1.8	1.2×10^{-4}	7.7	2.2	3.2×10^{-4}
WNT9A	rs35799012	1	228,133,322	C	0.83	Intron variant	0.322	0.041	−7.9	2.4	9.1×10^{-4}	−13.5	2.9	2.2×10^{-6}

Definition of abbreviations: BP = base position (build GRCh37); Chr = chromosome; EA = effect allele; EAF = EA frequency; eQTL = expression quantitative trait loci; NA = not available; UKB = UK Biobank; UTR = untranslated region.

*FVC in milliliters.

†Functional consequence for SNPs with different consequences associated with different transcripts; we considered the most deleterious.

‡Per-allele effect estimate.

§Expression data for proxy rs9936329 ($r^2 = 0.95$).

‖Expression data for proxy rs11057583 ($r^2 = 1.0$).

Table 2. UKB Results from Stage 1 and Stage 2 (Internal Replication) for the 20 Signals for FEV_1/FVC*

Gene	SNP	Chr	BP	EA	EAF	Functional Consequence†	Blood eQTL *P* Value	Lung eQTL *P* Value	Stage 1 (*n* = 207,616)			Stage 2 (Internal Replication) (*n* = 138,411)		
									β‡	SE	*P* Value	β‡	SE	*P* Value
CSNK2B	rs3117579	6	31,633,496	G	0.80	5′ UTR variant	1.1×10^{-4}	0.025	0.24	0.02	1.2×10^{-23}	0.19	0.03	2.0×10^{-11}
CTNND1	rs665058	11	57,579,166	T	0.56	Intron variant	5.5×10^{-12}	0.460	−0.08	0.02	1.0×10^{-5}	−0.12	0.02	1.6×10^{-7}
ELN	rs2528794	7	73,480,805	G	0.88	Intron variant	NA	0.413	−0.13	0.03	7.5×10^{-6}	−0.13	0.04	1.7×10^{-4}
FARP2	rs377324224	2	242,393,182	TG	0.63	Intron variant	NA	NA	0.08	0.02	9.5×10^{-5}	0.09	0.02	1.1×10^{-4}
FGFR3	rs3135877	4	1,804,276	G	0.96	Intron variant	NA	0.479	−0.24	0.05	9.4×10^{-7}	−0.34	0.06	7.0×10^{-9}
FGFR4	rs3135911	5	176,513,896	C	0.71	5′ UTR variant	9.5×10^{-6}	0.359	0.10	0.02	8.0×10^{-7}	0.10	0.03	7.5×10^{-5}
GFI1	rs150037086	1	92,952,080	G	0.31	Intron variant	1.6×10^{-18}§	NA	0.11	0.02	1.8×10^{-7}	0.11	0.03	8.0×10^{-6}
GJE1	rs225607	6	142,455,130	C	0.54	Missense variant	NA	NA	0.10	0.02	1.1×10^{-7}	0.13	0.02	2.1×10^{-8}
KAT7	rs755736	17	47,891,904	A	0.34	Intron variant	3.5×10^{-4}	0.439	0.07	0.02	5.4×10^{-4}	0.09	0.02	1.1×10^{-4}
MAPRE1	rs853854	20	31,420,757	T	0.48	Intron variant	3.0×10^{-34}	NA	−0.08	0.02	3.2×10^{-5}	−0.08	0.02	2.1×10^{-4}
NFATC3	rs548092276	16	68,210,935	C	0.84	Intron variant	NA	NA	−0.16	0.03	3.6×10^{-9}	−0.16	0.03	7.5×10^{-7}
PDGFB	rs2267406	22	39,633,749	T	0.25	Intron variant	9.8×10^{-184}	0.396	−0.09	0.02	2.1×10^{-5}	−0.12	0.03	1.1×10^{-5}
PPARD	rs2267666	6	35,370,728	A	0.24	Intron variant	0.002	5.9×10^{-4}	−0.13	0.02	5.2×10^{-9}	−0.10	0.03	7.0×10^{-5}
RARA	rs2715554	17	38,489,170	A	0.85	Intron variant	3.6×10^{-4}	0.568	0.16	0.03	2.2×10^{-9}	0.12	0.03	1.5×10^{-4}
RUNX3	rs9438876	1	25,241,116	A	0.49	Intron variant	8.9×10^{-17}	0.323	0.11	0.02	4.1×10^{-9}	0.10	0.02	2.4×10^{-5}
SERPING1	rs11229063	11	57,369,730	G	0.73	Intron variant	7.2×10^{-253}	0.079	0.12	0.02	2.3×10^{-8}	0.12	0.03	2.5×10^{-6}
SFRP2	rs17030437	4	154,704,225	C	0.75	Intron variant	7.2×10^{-80}	0.776	0.10	0.02	1.1×10^{-5}	0.10	0.03	1.5×10^{-4}
SOX9	rs796209434	17	70,122,505	CT	0.53	3′ UTR variant	NA	NA	−0.04	0.02	2.4×10^{-2}	−0.08	0.02	4.3×10^{-4}
TCF7L1	rs4346385	2	85,504,989	A	0.29	Intron variant	0.002	0.006	−0.08	0.02	1.6×10^{-4}	−0.09	0.03	3.1×10^{-4}
WNT7A	rs73151668	3	13,920,594	G	0.85	Intron variant	0.281	1.1×10^{-5}	−0.11	0.03	3.0×10^{-5}	−0.21	0.03	9.5×10^{-11}

For definition of abbreviations, *see* Table 1.

*FEV_1/FVC expressed as a percentage.

†Functional consequence for SNPs with different consequences associated with different transcripts; we considered the most deleterious.

‡Per-allele effect estimate.

§Expression data for proxy rs4565725 ($r^2 = 0.84$).

Table 3. Results for the Association of the 36 Signals with COPD

Gene	SNP	Chr	BP	EA	EAF	OR	95% CI	*P* Value
ACTN3	rs57127845	11	66,318,325	C	0.82	1.01	0.99–1.03	0.299
ACTN4	rs189809900	19	39,147,164	A	0.98	1.03	0.97–1.10	0.326
CLDN20	rs34268254	6	155,590,120	TA	0.39	0.99	0.97–1.00	0.105
CSNK2B	rs3117579	6	31,633,496	G	0.8	0.95	0.93–0.97	**1.1×10^{-6}***
CTNND1	rs665058	11	57,579,166	T	0.56	1.03	1.01–1.04	**0.001***
ELN	rs2528794	7	73,480,805	G	0.88	1.05	1.02–1.08	**2.2×10^{-4}***
FARP2	rs377324224	2	242,393,182	TG	0.63	0.98	0.96–1.00	**0.011**
FGFR3	rs3135877	4	1,804,276	G	0.96	1.11	1.06–1.16	**1.3×10^{-6}***
FGFR4	rs3135911	5	176,513,896	C	0.71	0.96	0.94–0.97	**4.0×10^{-7}***
GFI1	rs150037086	1	92,952,080	G	0.31	0.97	0.95–0.98	**9.5×10^{-5}***
GJE1	rs225607	6	142,455,130	C	0.54	0.97	0.96–0.99	**4.1×10^{-4}***
GSK3B	rs6805251	3	119,560,606	T	0.38	0.99	0.98–1.01	0.447
HOXA1	rs45571645	7	27,135,096	G	0.98	0.98	0.92–1.04	0.743
HOXB4	rs201603635	17	46,653,038	T	0.94	1.01	0.97–1.04	0.729
ITGB5	rs17282078	3	124,481,760	T	0.87	1.00	0.98–1.03	0.759
KAT7	rs755736	17	47,891,904	A	0.34	0.98	0.96–1.00	**0.018**
KAT8	rs138259061	16	31,136,066	A	0.64	1.01	1.00–1.03	0.190
MAPRE1	rs853854	20	31,420,757	T	0.48	1.02	1.01–1.04	**0.008**
MMP24	rs7280	20	33,864,484	A	0.58	1.01	1.00–1.03	0.119
NCOR2	rs72451021	12	124,811,393	ACT	0.89	0.99	0.97–1.02	0.522
NFATC3	rs548092276	16	68,210,935	C	0.84	1.06	1.03–1.08	**1.6×10^{-6}***
NR3C1	rs72801051	5	142,685,670	A	0.84	1.04	1.02–1.07	**7.1×10^{-5}***
PDGFB	rs2267406	22	39,633,749	T	0.25	1.03	1.01–1.05	**0.006**
PPARD	rs2267666	6	35,370,728	A	0.24	1.05	1.03–1.07	**1.7×10^{-7}***
RARA	rs2715554	17	38,489,170	A	0.85	0.94	0.92–0.96	**8.6×10^{-8}***
ROR2	rs12684752	9	94,682,990	T	0.94	1.00	0.97–1.04	0.824
RUNX1	rs12483501	21	36,224,276	T	0.63	0.98	0.97–1.00	**0.044**
RUNX3	rs9438876	1	25,241,116	A	0.49	0.97	0.95–0.98	**4.1×10^{-5}***
SERPINC1	rs2227603	1	173,882,548	A	0.97	0.97	0.92–1.01	0.155
SERPING1	rs11229063	11	57,369,730	G	0.73	0.96	0.95–0.98	**2.1×10^{-5}***
SFRP2	rs17030437	4	154,704,225	C	0.75	0.97	0.96–0.99	**0.004**
SOX9	rs796209434	17	70,122,505	CT	0.53	1.01	1.00–1.03	0.079
TCF7L1	rs4346385	2	85,504,989	A	0.29	1.03	1.01–1.05	**0.003**
WNT2B	rs351370	1	113,054,659	C	0.41	1.00	0.98–1.02	0.944
WNT7A	rs73151668	3	13,920,594	G	0.85	1.06	1.03–1.08	**2.0×10^{-6}***
WNT9A	rs35799012	1	228,133,322	C	0.83	1.00	0.98–1.02	0.749

Definition of abbreviations: BP = base position (build GRCh37); Chr = chromosome; CI = confidence interval; COPD = chronic obstructive pulmonary disease; EA = effect allele; EAF = EA frequency; OR = per-allele odds ratio.
Analyses in the whole dataset: $N = 346{,}027$. In bold are results with $P < 0.05$.
*Statistically significant after Bonferroni correction ($P < 1.4 \times 10^{-3}$).

34 variants, 16 variants replicated after Bonferroni correction, and another 12 variants replicated at the nominal level of significance (Tables 4 and 5).

To help interpret our findings, we grouped all 55 genes into biological categories based on their known function, as shown in Table 6; such information was derived from the National Center for Biotechnology Information (NCBI) Gene (www.ncbi.nlm.nih.gov/gene), Ensembl (www.ensembl.org), GeneCards (www.genecards.org), and Mouse Genome Informatics (www.informatics.jax.org) databases. Of the 55 genes, 53 genes fall into only the following four categories: growth factors, transcriptional regulators, cell-to-cell adhesion and cytoskeletal, and extracellular matrix (ECM). Genes encoding growth factors, or their receptors, are the most well-represented category ($n = 19$), and within this group, Wnt-signaling genes (*CSNK2B*, *DVL2*, *GSK3B*, *ROR2*, *SFRP2*, *TCF7L1*, *WNT2B*, *WNT7A*, and *WNT9A*) are particularly prevalent. Genes encoding transcription factors are also highly represented ($n = 17$); within this category, we identified genes involved in vitamin A signaling, including the retinoic acid ligand–activated transcription factors (*RARA* and *RARB*), and glucocorticoid signaling genes, including the glucocorticoid receptor gene (*NR3C1*), *NCOR1*, and its paralogue *NCOR2* that modulate the activity of nuclear receptors, including RARs (retinoic acid receptors), PPARD, and the glucocorticoid receptor. Ten genes relate to cell-to-cell adhesion and the cytoskeleton, including three genes associated with actin microfilaments (*ACTN3*, *ACTN4*, and *TNS1*). Another seven genes relate to the ECM, including *ELN*, which encodes elastin.

Functional Annotation and Gene Expression

Using the Ensembl variant effect predictor tool (www.ensembl.org/info/genome/variation/prediction/predicted_data.html#consequences) (30), we investigated the functional consequence of the 36 novel signals; Tables 1 and 2 show that most of them are intron variants.

Table 4. Results from External Replication in CHARGE and SpiroMeta for the 17 Signals for FVC

					Internal Replication			External Replication							
					UKB Stage 2 (*N* = *138,411*)			CHARGE				SpiroMeta			Meta-Analysis CHARGE + SpiroMeta
Gene	SNP	Chr	EA	EAF	β	SE	*P* Value	*N*	β	SE	*P* Value	*n*	Direction	*P* Value	*P* Value
ACTN3	rs57127845	11	C	0.82	−14.7	2.9	2.3×10^{-7}	60,507	**−11.2**	4.2	**3.6×10^{-3}**	75,423	+	0.473[†]	**0.014**
ACTN4	rs189809900	19	A	0.98	−35.8	8.9	2.8×10^{-5}	36,112	4.7	21.3	0.414[†]	81,081	—	0.232	0.407
CLDN20	rs34268254	6	TA	0.39	9.0	2.3	4.6×10^{-5}	60,507[‡]	7.9[‡]	3.2[‡]	**6.4×10^{-3}[‡]**	75,422[‡]	+[‡]	**0.011[‡]**	**7.4×10^{-4}***
GSK3B	rs6805251	3	T	0.38	8.0	2.3	2.1×10^{-4}	60,506	**4.9**	3.2	0.064	74,551	+	0.083	**0.033**
HOXA1	rs45571645	7	G	0.98	28.7	8.3	2.9×10^{-4}	58,929	2.3	12.0	0.425	82,863	—	0.481[†]	0.559
HOXB4	rs201603635	17	T	0.94	−16.3	4.6	1.9×10^{-4}	NA	NA	NA	NA	NA	NA	NA	NA
ITGB5	rs17282078	3	T	0.87	15.7	3.3	9.0×10^{-7}	60,508	**13.5**	4.7	**2.0×10^{-3}***	75,422	+	0.365	**6.1×10^{-3}**
KAT8	rs138259061	16	A	0.64	−7.7	2.3	4.5×10^{-4}	60,508[§]	**−6.2[§]**	3.3[§]	**0.029[§]**	82,865	—	0.099	**0.020**
MMP24	rs7280	20	A	0.58	9.7	2.3	9.5×10^{-6}	60,508	**8.5**	3.3	**5.1×10^{-3}**	81,992	+	**0.014**	**7.5×10^{-4}***
NCOR2	rs72451021	12	ACT	0.89	12.8	3.6	1.9×10^{-4}	45,256	1.9	5.9	0.372	75,422[‖]	+[‖]	**0.038[‖]**	0.074
NR3C1	rs72801051	5	A	0.84	10.7	3.0	1.6×10^{-4}	60,507	**7.5**	4.4	**0.046**	75,421	+	0.111	**0.032**
ROR2	rs12684752	9	T	0.94	−17.8	4.7	8.5×10^{-5}	60,506	−2.5	6.6	0.353	75,421	—	0.398	0.415
RUNX1	rs12483501	21	T	0.63	10.1	2.5	2.1×10^{-5}	60,508	**4.0**	3.9	0.150	81,992	+	**0.026**	**0.026**
SERPINC1	rs2227603	1	A	0.97	−23.1	6.6	2.5×10^{-4}	60,508	**−20.6**	10.0	**0.020**	75,423	—	0.367	**0.044**
SOX9	rs796209434	17	CT	0.53	7.9	2.2	2.0×10^{-4}	60,506[¶]	**10.4[¶]**	3.2[¶]	**5.4×10^{-4}*[¶]**	75,422[¶]	+[¶]	**8.3×10^{-3}[¶]**	**6.0×10^{-5}***
WNT2B	rs351370	1	C	0.41	7.7	2.2	3.2×10^{-4}	60,506	**7.4**	3.4	**0.014**	74,550	+	**3.2×10^{-3}**	**4.9×10^{-4}***
WNT9A	rs35799012	1	C	0.83	−13.5	2.9	2.2×10^{-6}	60,508	**−9.4**	4.7	**0.024**	74,552	—	**2.8×10^{-3}***	**7.1×10^{-4}***

Definition of abbreviations: CHARGE = Cohorts for Heart and Aging Research in Genomic Epidemiology; Chr = chromosome; EA = effect allele; EAF = EA frequency; NA = not available; UKB = UK Biobank.

Analyses in the whole dataset (*N* = 346,027). For SpiroMeta, reported only effect direction because β not interpretable (use of rank-based inverse normal transformation). In bold are external replication results with $P < 0.05$. β values are per-allele effect estimates.

*External replication results significant at Bonferroni ($P < 3.1 \times 10^{-3}$).

[†]*P* value inverted in Fisher's meta-analysis to reflect the effect in opposite direction (*P* values reported for UKB stage 2, CHARGE, and SpiroMeta are all one-sided, *see text*).

[‡]Proxy: rs13220615 ($r^2 = 0.97$).

[§]Proxy: rs1978485 ($r^2 = 0.98$).

[‖]Proxy: rs11057583 ($r^2 = 1.0$).

[¶]Proxy: rs1042678 ($r^2 = 0.97$).

Table 5. Results from External Replication in CHARGE and SpiroMeta for the 20 Signals for FEV_1/FVC

					Internal Replication			External Replication							
					UKB Stage 2 (*n* = *138,411*)			CHARGE				SpiroMeta			Meta-Analysis CHARGE + SpiroMeta
Gene	SNP	Chr	EA	EAF	β	SE	*P* Value	*N*	β	SE	*P* Value	*n*	Direction	*P* Value	*P* Value
CSNK2B	rs3117579	6	G	0.80	0.19	0.03	2.0×10^{-11}	NA	NA	NA	NA	83,081	+	**5.9×10^{-5}***	—
CTNND1	rs665058	11	T	0.56	−0.12	0.02	1.6×10^{-7}	60,531	−0.05	0.04	0.132	75,639	—	0.078	0.057
ELN	rs2528794	7	G	0.88	−0.13	0.04	1.7×10^{-4}	58,707	**−0.27**	0.07	**7.0×10^{-5}***	74,767	—	**0.016**	**1.6×10^{-5}***
FARP2	rs377324224	2	TG	0.63	0.09	0.02	1.1×10^{-4}	NA	NA	NA	NA	NA	NA	NA	NA
FGFR3	rs3135877	4	G	0.96	−0.34	0.06	7.0×10^{-9}	39,004	**−0.46**	0.13	**1.6×10^{-4}***	69,559	—	0.120	**2.3×10^{-4}***
FGFR4	rs3135911	5	C	0.71	0.10	0.03	7.5×10^{-5}	51,019	**0.21**	0.05	**1.9×10^{-5}***	75,639†	+†	**1.7×10^{-3}*†**	**5.9×10^{-7}***
GFI1	rs150037086	1	G	0.31	0.11	0.03	8.0×10^{-6}	60,530‡	**0.12‡**	0.04‡	**3.9×10^{-3}‡**	75,638‡	+‡	0.054‡	**2.0×10^{-3}***
GJE1	rs225607	6	C	0.54	0.13	0.02	2.1×10^{-8}	60,531	**0.07**	0.04	0.060	75,638	+	0.114	**0.040**
KAT7	rs755736	17	A	0.34	0.09	0.02	1.1×10^{-4}	58,706	**0.04**	0.04	0.188	83,081	+	**0.018**	**0.023**
MAPRE1	rs853854	20	T	0.48	−0.08	0.02	2.1×10^{-4}	58,949	**−0.04**	0.04	0.191	83,079	—	**0.025**	**0.030**
NFATC3	rs548092276	16	C	0.84	−0.16	0.03	7.5×10^{-7}	60,531§	**−0.16§**	0.05§	**1.7×10^{-3}*§**	75,639§	−§	**4.8×10^{-3}§**	**1.0×10^{-4}***
PDGFB	rs2267406	22	T	0.25	−0.12	0.03	1.1×10^{-5}	51,669	**−0.20**	0.05	**9.1×10^{-5}***	83,079	—	0.063	**7.5×10^{-5}***
PPARD	rs2267666	6	A	0.24	−0.10	0.03	7.0×10^{-5}	60,532	**−0.22**	0.05	**2.6×10^{-6}***	83,079	—	**7.1×10^{-3}**	**3.5×10^{-7}***
RARA	rs2715554	17	A	0.85	0.12	0.03	1.5×10^{-4}	37,587	**0.04**	0.08	0.285	83,080	+	**0.017**	**0.030**
RUNX3	rs9438876	1	A	0.49	0.10	0.02	2.4×10^{-5}	58,679	**0.08**	0.05	**0.041**	82,209	+	**0.027**	**8.6×10^{-3}**
SERPING1	rs11229063	11	G	0.73	0.12	0.03	2.5×10^{-6}	60,529	**0.11**	0.05	**0.010**	75,638	+	**0.018**	**1.7×10^{-3}***
SFRP2	rs17030437	4	C	0.75	0.10	0.03	1.5×10^{-4}	60,503	**0.19**	0.05	**1.3×10^{-5}***	75,638	+	0.143	**2.6×10^{-5}***
SOX9	rs796209434	17	CT	0.53	−0.08	0.02	4.3×10^{-4}	60,529‖	**−0.08‖**	0.04‖	**0.022‖**	75,637‖	—‖	**2.1×10^{-3}*‖**	**5.1×10^{-4}***
TCF7L1	rs4346385	2	A	0.29	−0.09	0.03	3.1×10^{-4}	60,531	**−0.03**	0.04	0.276	75,638	—	**0.043**	0.065
WNT7A	rs73151668	3	G	0.85	−0.21	0.03	9.5×10^{-11}	60,530	**−0.24**	0.06	**4.6×10^{-5}***	82,210	—	**9.3×10^{-4}***	**7.7×10^{-7}***

For definition of abbreviations, *see* Table 4.

Analyses in the whole dataset ($N = 346{,}027$). In bold are external replication results with $P < 0.05$. β values are per-allele effect estimates.

*External replication results significant at Bonferroni ($P < 2.6 \times 10^{-3}$).

†Proxy: rs451643 ($r^2 = 1$).

‡Proxy: rs4565725 ($r^2 = 0.84$).

§Proxy: rs8048034 ($r^2 = 0.80$).

‖Proxy: rs1042678 ($r^2 = 0.95$).

Table 6. Gene Function and Associated Biological Categories for All the 55 Genes Identified for FVC, FEV_1/FVC, or Both*

Gene and Biological Category	Full Name	Function
Growth factors		
CSNK2B	Casein kinase 2 β	Ubiquitous protein kinase that regulates metabolic pathways, signal transduction, transcription, translation, and replication
FGFR3	Fibroblast growth factor receptor 3	Encodes a tyrosine kinase and cell surface receptor for fibroblast growth factors
FGFR4	Fibroblast growth factor receptor 4	Encodes a tyrosine kinase and cell surface receptor for fibroblast growth factors
GSK3B	Glycogen synthase kinase 3 β	Encodes a serine-threonine kinase belonging to the glycogen synthase kinase subfamily
PDGFB	Platelet-derived growth factor subunit B	Encodes a member of the protein family comprised of PDGFs
ROR2	Receptor tyrosine kinase like orphan receptor 2	Encodes a receptor protein tyrosine kinase and a type I transmembrane protein that belongs to the ROR subfamily of cell surface receptors
SFRP2	Secreted frizzled related protein 2	Encodes a member of the SFRP family that acts as soluble modulators of Wnt signaling
TCF7L1	Transcription factor 7–like 1	Encodes a member of the T-cell factor/lymphoid enhancer factor family of transcription factors
WNT2B	Wnt family member 2	Member of the WNT gene family
WNT7A	Wnt family member 7A	Member of the WNT gene family
WNT9A	Wnt family member 9A	Member of the WNT gene family
BMP4	Bone morphogenetic protein 4	Encodes a secreted ligand of the TGF-β (transforming growth factor β) superfamily of proteins
FGF10	Fibroblast growth factor 10	Encodes a member of the fibroblast growth factor family with roles in morphogenesis of epithelium, reepithelialization of wounds, hair development, and early lung organogenesis
FGF18	Fibroblast growth factor 18	Encodes a member of the fibroblast growth factor family with roles in cell growth, morphogenesis, and tissue repair and is particularly important in bone development
HHIP	Hedgehog interacting protein	Encodes a member of the HHIP family, which is a highly conserved, vertebrate-specific inhibitor of HH signaling
IGF1	Insulin-like growth factor 1	Encodes an insulin-like protein involved in mediating growth and development
KDR	Kinase insert domain receptor—vascular endothelial growth factor receptor 2	Encodes one of the two receptors of the VEGF; this receptor functions as the main mediator of VEGF-induced endothelial proliferation, survival, migration, tubular morphogenesis, and sprouting
PTCH1	Patched 1	Encodes a member of the patched family of proteins and a component of the hedgehog signaling pathway
TGFB2	Transforming growth factor β 2	Encodes a secreted ligand of the TGF-β superfamily of proteins
Transcriptional regulators		
GFI1	Growth factor independent 1 transcriptional repressor	Encodes a nuclear zinc-finger protein that functions as a transcriptional repressor
HOXA1	Homeobox A1	Encodes a DNA-binding transcription factor involved in spatial patterning in development
HOXB4	Homeobox B4	Encodes a DNA-binding transcription factor involved in spatial patterning in development
KAT7	Lysine acetyltransferase 7	Encodes a protein that is part of the multimeric HBO1 complex and possesses histone H4-specific acetyltransferase activity; this activity regulates gene transcription (e.g., VEGFR2, by influencing chromatin conformation)
KAT8	Lysine acetyltransferase 8	Encodes a member of the MYST histone acetylase protein family; the encoded protein regulates gene transcription by influencing chromatin conformation
NCOR2	Nuclear receptor corepressor 2	Encodes a protein that regulates repression of thyroid-hormone and retinoic-acid receptors
NFATC3	Nuclear factor of activated T cells 3	Encodes a member of the nuclear factors of activated T cells family of transcription factors
NR3C1	Nuclear receptor subfamily 3 group C member 1	Encodes glucocorticoid receptor
PPARD	Peroxisome proliferator-activated receptor delta	Encodes a member of the PPAR family that is believed to function as an integrator of transcriptional repression and nuclear receptor signaling
RARA	Retinoic acid receptor α	Encodes the retinoic acid receptor α that acts as a ligand-activated transcription factor
RUNX1	Runt-related transcription factor 1	Encodes for a member of the runt family of transcription factors that regulate hematopoiesis and skeletal development
RUNX3	Runt-related transcription factor 3	Encodes for a member of the runt family of transcription factors that regulate hematopoiesis and skeletal development
SOX9	SRY-box 9	The protein encoded is an HMG box DNA-binding protein
GATA6	GATA-binding protein 6	Member of the GATA family of transcription factors that regulate cellular differentiation and organogenesis during embryonic development
NCOR1	Nuclear receptor corepressor 1	Encodes a protein that regulates repression of thyroid-hormone and retinoic-acid receptors
RARB	Retinoic acid receptor β	Encodes the retinoic acid receptor β that acts as a ligand-activated transcription factor

(Continued)

Table 6. (*Continued*)

Gene and Biological Category	Full Name	Function
RUNX2	Runt-related transcription factor 2	Encodes for a member of the runt family of transcription factors that regulate hematopoiesis and skeletal development
Cell-to-cell adhesion and cytoskeleton		
ACTN3	Actinin α 3 (gene/pseudogene)	Involved in crosslinking actin filaments, part of the cytoskeleton
ACTN4	Actinin α 4	Actin-binding protein, part of the cytoskeleton
CLDN20	Claudin 20	Encodes a tight junction protein; important for cell polarity and regulating movement of molecules via the paracellular route.
CTNND1	Catenin delta 1	Armadillo protein family, which function in adhesion between cells and signal transduction
FARP2	FERM, ARH/RhoGEF, and pleckstrin domain protein 2	ρ guanidine exchange factor
GJE1	Gap junction protein epsilon 1	Gap junction protein; Gap junctions are specialized intercellular connections that enable cell-to-cell communication
MAPRE1	Microtubule associated protein RP/EB family member 1	Encodes a protein that localizes to microtubules, a dynamic network of filaments that form part of the cytoskeleton
DSP	Desmoplakin	Encodes a protein component of functional desmosomes
PARD3	Par-3 family cell polarity regulator	Encodes a member of the PARD protein family that regulates cell polarity and cell-to-cell integrity
TNS1	Tensin 1	Encodes for a protein that localizes to focal adhesions and crosslinks actin filaments
Extracellular matrix		
ELN	Elastin	Encodes a protein that is one of the two components of elastic fibers
ITGB5	Integrin subunit β 5	Encodes the integrin β subunit 5 protein
MMP24	Matrix metallopeptidase 24	Encodes a member of the peptidase M10 family of MMPs
SERPINC1	Serpin family C member 1	Encodes a plasma protease inhibitor and a member of the serpin superfamily
SERPING1	Serpin family G member 1	Encodes a highly glycosylated plasma protein involved in the regulation of the complement cascade
ITGAV	Integrin subunit α V	Encodes a member of the integrin α chain family
MMP15	Matrix metallopeptidase 15	Encodes a member of the peptidase M10 family and membrane-type subfamily of MMPs
Oxidative stress and endothelial dysfunction		
AGER	Advanced glycosylation end-product (AGE) specific receptor	Multiligand receptor; role in chronic vascular injury
Immune response and surfactant regulation		
SFTPD	Surfactant protein D	The protein encoded is part of the innate immune response and has a role in surfactant regulation

*Bold formatting indicates the 36 novel genes.

We also assessed whether the 36 signals affect the expression of the genes they lie within. For gene expression in the blood, we used cis–expression quantitative trait loci (eQTL) data from the eQTLGen Consortium (www.eqtlgen.org/cis-eqtls.html) (31), which includes 37 datasets with a total of 31,684 individuals; for the 36 SNPs, the actual sample size varied from 8,269 to 31,684. For gene expression in lung tissue, we used data from the Genotype-Tissue Expression Portal (GTEx) (www.gtexportal.org/home/eqtls/tissue?tissueName=Lung), which includes lung tissue samples from 383 individuals, with actual sample sizes varying from 12 to 286 for our 36 SNPs. Tables 1 and 2 report the effects of the 36 signals on the expression of the gene they lie within. Of 27 SNPs with available data, 22 showed eQTL evidence in the blood, and 10 showed it in the lung tissue. For four signals (*WNT2B, WNT7A, WNT9A*, and *ACTN4*), we found evidence in the lung tissue, but not in the blood, despite the very small sample size of lung eQTL data.

Discussion

Our study demonstrates the role of lung development genes in regulating adult lung function and provides further support for the developmental origins of both restrictive and obstructive impairment of adult lung function and spirometrically defined COPD. Overall, we identified 55 lung development–related genes associated with adult lung function; of these, 36 had not been reported in the largest and most recent GWAS of lung function (9), showing the value of our hypothesis-driven approach in complementing agnostic GWASs. Only 6 of the 36 signals could not be replicated in external populations from the CHARGE and SpiroMeta consortia; for three of them, this is not surprising, given the low allele frequency and, therefore, low power to detect realistic effect sizes despite the large replication sample size.

To further assess the novelty of the 36 genes, we searched the literature for any evidence of association with lung function and related outcomes, using PhenoScanner (32) and HuGE Navigator (17) and checking references of relevant papers. We found previous evidence for just four of the 36 genes. An intergenic variant annotated to *NCOR2* (*NCOR2/SCARB1* locus) was

previously associated with adult FEV_1 (26) but did not replicate in the study by Shrine and colleagues (9). *NCOR2* was also associated with FVC in young adults but could only be replicated in children (15); the same study identified, but did not replicate, *KAT8*. *SOX9* was associated with adult FEV_1 in a study that included SNP by smoking interaction (33). *NR3C1* was previously identified in a GWAS of spirometrically defined COPD (34); recently, an intergenic variant annotated to *NR3C1* (*NR3C1/ARHGAP26* locus) was also associated with FEV_1/FVC in a methodological study incorporating functional genomics data to increase power in the GWAS (35). Interestingly, two additional genes were previously associated with asthma-related phenotypes, *RUNX1* with pediatric asthma (36) and IgE concentrations (37) in two candidate-gene studies, and *ITGB5* with airway hyperresponsiveness in individuals with asthma in a GWAS (38).

Among all 55 genes, the large majority show an association with either FVC or FEV_1/FVC, but not both, which is not surprising, given that these parameters identify distinct patterns of lung function impairment. In population-based epidemiological studies, a low FVC is a marker of restriction, indicating small lung volumes, and is a strong predictor of all-cause mortality, even in the absence of chronic lung disease (1). Similarly, a low FEV_1/FVC is an epidemiological marker of COPD, which is projected to become the third leading cause of death worldwide by 2020 (39). Knowledge of whether a lung development gene affects restriction, obstruction, or both links the development of lung structure with function and points to underlying mechanistic pathways that will inform future experimental follow-up studies.

Biological Interpretation

Our finding that 53 of the 55 genes identified in this study fall into four biological categories that regulate organ size and cell integrity indicates the particular importance of these processes for adult lung health. Growth factors, the best-represented gene category, are diffusible signaling proteins that exert a variety of biological responses important for organ generation, including proliferation, morphogenesis, and angiogenesis. They are also important for maintaining homeostasis in adulthood. Abnormal production of growth factors can lead to lung diseases; for example, perturbed angiogenic growth factors can lead to bronchopulmonary dysplasia (40), and overactive TGF-β signaling can lead to idiopathic pulmonary fibrosis (41). Within this group, Wnt-signaling genes are highly represented; in addition to being critically required for all stages of lung generation, the Wnt-signaling pathway has an important role in maintaining lung health by stimulating repair after injury (42, 43).

Genes encoding transcription factors are also well represented; these regulate the expression of multiple genes by binding to specific DNA sequences to activate or repress gene transcription. During development, transcriptional regulators control growth in a highly ordered spatiotemporal manner (44), the disruption of which can affect organ size, architecture, and function. Within this category, we identified genes involved in vitamin A and glucocorticoid signaling. Vitamin A signaling has an important role not only in lung development but also in adult lung structural homeostasis, with abnormal vitamin A signaling associated with histological emphysema, driven possibly via aberrant endothelial cell repair in patients with COPD (45, 46). Interestingly, we also identified transcription factors, such as the homeobox genes *HOXA1* and *HOXB4*, which themselves are transcriptional targets of other genes that we identified, including *RARA*, *RARB*, *WNT2B*, *WNT7B*, and *WNT9A*.

Some of the genes identified relate to cell-to-cell adhesion and the cytoskeleton. Cell-to-cell adhesion is important to maintain tissue integrity; its breakdown and subsequent loss of epithelial barrier function is also frequently a component of lung disease (47). Three of the genes identified (*ACTN3*, *ACTN4*, and *TNS1*) act on the actin cytoskeleton, a network of intracellular fibers that are integral to both cell-to-cell and cell-to-ECM interactions and that are required to maintain cell integrity and movement (48).

Finally, we identified genes related to the ECM, which in the lungs, provides not only a scaffold to support cells but also a source of biological signals and mechanical strength to maintain cell integrity and health through a bioactive environment interacting with surrounding cells (48, 49). Pathological changes to the ECM are a recognized hallmark of lung diseases, including asthma, COPD, and idiopathic pulmonary fibrosis, and current regenerative medicine strategies are exploring the efficacy of targeting the ECM as a possible avenue for the treatment of lung diseases (49). Included in this category is the gene encoding elastin (*ELN*); elastin is a major component of the ECM that not only links alveoli to the conducting airways but also is a key determinant of the elastic recoil in the lung. We speculate that the association of *ELN* with FEV_1/FVC and COPD (FEV_1/FVC $<$ LLN) in our data might reflect an effect of this gene on elastic recoil and the risk of emphysema.

Strengths and Limitations

Despite the high heritability of lung function, genetic variants identified by agnostic GWASs still explain only a small proportion of its variability in the population (9). By using a hypothesis-driven approach, we have identified a substantial number of additional variants associated with lung function, especially polymorphisms with relatively low allele frequencies, which may not have reached strict genome-wide significance thresholds in previous GWASs. Although this suggests that focusing the analyses on many genes related to a pathophysiological process believed to affect the outcome is a promising approach, a practical issue is how to select the genes to be investigated. Our list of about 400 genes was previously prepared following a thorough process based on experts' knowledge from animal and human studies, integrated with data from bioinformatic tools (15). However, we acknowledge that there is a degree of subjectivity involved in this method.

Epidemiological studies have linked the early life environment to adult lung function and COPD, and it is assumed that these associations are mediated through impacts on lung growth and development. By demonstrating clear associations of multiple lung development genes with adult lung function, we have provided more direct evidence that lung development plays a crucial role in adult lung health. Furthermore, in contrast to observational studies implicating the early environment, our genetic findings are unlikely to be affected by classical environmental and lifestyle confounders, and this strengthens causal inference. That said, given the cross-sectional nature of our study and the age of UKB participants, measured lung function will reflect a combination of the maximal level attained through growth and subsequent decline. We therefore cannot determine whether the implicated lung development genes are only influencing the former or whether they may also be influencing repair and, hence, combating

insults such as smoking, which can cause accelerated decline later in life. The broadly similar results in smokers and nonsmokers do not favor one explanation over the other.

An obstructive pattern, indicated by a low FEV_1/FVC ratio, can be caused by respiratory conditions other than COPD, including bronchiectasis, bronchiolitis, and cystic fibrosis, but these are uncommon in the general population. Asthma is more common, however, and can also result in a low prebronchodilator FEV_1/FVC, which cannot exclude the presence of reversible obstruction. As post-bronchodilator lung function was not measured in the UKB, we performed sensitivity analyses excluding individuals with a self-reported doctor diagnosis of asthma, and these confirmed the results of the main analyses (data not shown).

Future Research

Further detailed investigation of our findings is required to identify the underlying causal variants and possible pathogenetic mechanisms. For some of the identified genes, there is experimental evidence of an ongoing role in adult lung homeostasis and repair through alveolar maintenance and regeneration after injury later in life, with potential implications for understanding the rapid decline of lung function and identifying future pharmacological targets. Longitudinal cohorts offer an opportunity to examine associations with lung function trajectories across the life course. If lung development genes are acting primarily on growth and development, we might expect to see stronger associations in children and young adults before lung function decline has commenced. Conversely, if they are acting primarily on repair, stronger effects might be seen on decline in older individuals. Extending the investigation of lung development genes to incorporate cross-sectional data on children, adolescents, and young adults from different studies would also help disentangle effects on lung growth from those on lung regeneration. However, such investigation would require very large sample sizes to ensure adequate power to detect signals with relatively small effects and/or low allele frequencies such as those that we have identified.

We have taken a conservative approach that only considered one best SNP per gene, but a gene may contain multiple independent signals. Similar to GWAS findings, the majority of our novel 36 signals are intronic variants, which might exert their effect by modifying the expression of other genes; however, most of them do affect the expression of the genes they lie within. These signals could be further investigated in relevant human cell lines or animal models, for example by using gene editing to delete a small region that includes the SNP identified, as recently done by Parker and colleagues (50).

Finally, further research is needed to clarify whether the identified genes act on lung function independently or through gene–gene or gene–environment interactions. For example, *NCOR2* might affect lung function through its effects on vitamin A metabolism via the RAR or alternatively through interaction with genes encoding non–nuclear receptor transcription factors like Foxp1, which are also important for lung development (51). Another example is a possible gene-to-environment interaction between lung development genes involved in vitamin A metabolism and vitamin A intake on lung function; for example, the beneficial effect of prenatal vitamin A supplementation on offspring lung function (52) may be modified by vitamin A–related genes.

In conclusion, our findings show a clear effect of lung development genes on adult lung function, influencing both restrictive and obstructive patterns. Furthermore, they demonstrate how genetic knowledge of relevant biological processes can be used to help identify novel genetic associations for complex traits. Further investigation of these developmental pathways could ultimately lead to druggable targets, with the aim of optimizing adult lung health and preventing COPD. ■

Author disclosures are available with the text of this article at www.atsjournals.org.

Acknowledgment: This research was conducted using the UK Biobank resource (application number 19136). The authors thank the participants, field workers, and data managers in the UK Biobank for all their time and efforts.

References

1. Burney PG, Hooper R. Forced vital capacity, airway obstruction and survival in a general population sample from the USA. *Thorax* 2011;66: 49–54.
2. Gupta RP, Strachan DP. Ventilatory function as a predictor of mortality in lifelong non-smokers: evidence from large British cohort studies. *BMJ Open* 2017;7:e015381.
3. Martinez FD. Early-life origins of chronic obstructive pulmonary disease. *N Engl J Med* 2016;375:871–878.
4. Melén E, Guerra S. Recent advances in understanding lung function development. *F1000 Res* 2017;6:726.
5. Stocks J, Hislop A, Sonnappa S. Early lung development: lifelong effect on respiratory health and disease. *Lancet Respir Med* 2013;1: 728–742.
6. Lange P, Celli B, Agustí A, Boje Jensen G, Divo M, Faner R, *et al*. Lung-function trajectories leading to chronic obstructive pulmonary disease. *N Engl J Med* 2015;373:111–122.
7. Krauss-Etschmann S, Bush A, Bellusci S, Brusselle GG, Dahlén SE, Dehmel S, *et al*. Of flies, mice and men: a systematic approach to understanding the early life origins of chronic lung disease. *Thorax* 2013;68:380–384.
8. John C, Soler Artigas M, Hui J, Nielsen SF, Rafaels N, Paré PD, *et al*. Genetic variants affecting cross-sectional lung function in adults show little or no effect on longitudinal lung function decline. *Thorax* 2017;72: 400–408.
9. Shrine N, Guyatt AL, Erzurumluoglu AM, Jackson VE, Hobbs BD, Melbourne CA, *et al*.; Understanding Society Scientific Group. New genetic signals for lung function highlight pathways and chronic obstructive pulmonary disease associations across multiple ancestries. *Nat Genet* 2019;51:481–493.
10. Collins SA, Lucas JS, Inskip HM, Godfrey KM, Roberts G, Holloway JW; Southampton Women's Survey Study Group. HHIP, HDAC4, NCR3 and RARB polymorphisms affect fetal, childhood and adult lung function. *Eur Respir J* 2013;41:756–757.
11. Miller S, Melén E, Merid SK, Hall IP, Sayers I. Genes associated with polymorphic variants predicting lung function are differentially expressed during human lung development. *Respir Res* 2016; 17:95.
12. Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, *et al*. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med* 2015;12:e1001779.
13. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, *et al*. The UK Biobank resource with deep phenotyping and genomic data. *Nature* 2018;562:203–209.

14. Wain LV, Shrine N, Miller S, Jackson VE, Ntalla I, Soler Artigas M, *et al*.; UK Brain Expression Consortium (UKBEC); OxGSK Consortium. Novel insights into the genetics of smoking behaviour, lung function, and chronic obstructive pulmonary disease (UK BiLEVE): a genetic association study in UK Biobank. *Lancet Respir Med* 2015;3:769–781.
15. Minelli C, Dean CH, Hind M, Alves AC, Amaral AF, Siroux V, *et al*.; SpiroMeta consortium; CHARGE consortium. Association of forced vital capacity with the developmental gene NCOR2. *PLoS One* 2016; 11:e0147388.
16. Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M. KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res* 2012;40:D109–D114.
17. Yu W, Gwinn M, Clyne M, Yesupriya A, Khoury MJ. A navigator for human genome epidemiology. *Nat Genet* 2008;40:124–125.
18. Wu C, Orozco C, Boyer J, Leglise M, Goodale J, Batalov S, *et al*. BioGPS: an extensible and customizable portal for querying and organizing gene annotation resources. *Genome Biol* 2009;10: R130.
19. Repapi E, Sayers I, Wain LV, Burton PR, Johnson T, Obeidat M, *et al*.; Wellcome Trust Case Control Consortium; NSHD Respiratory Study Team. Genome-wide association study identifies five loci associated with lung function. *Nat Genet* 2010;42:36–44.
20. Wilk JB, Chen TH, Gottlieb DJ, Walter RE, Nagle MW, Brandler BJ, *et al*. A genome-wide association study of pulmonary function measures in the Framingham Heart Study. *PLoS Genet* 2009;5:e1000429.
21. Hancock DB, Eijgelsheim M, Wilk JB, Gharib SA, Loehr LR, Marciante KD, *et al*. Meta-analyses of genome-wide association studies identify multiple loci associated with pulmonary function. *Nat Genet* 2010;42: 45–52.
22. Soler Artigas M, Loth DW, Wain LV, Gharib SA, Obeidat M, Tang W, *et al*. Genome-wide association and large-scale follow up identifies 16 new loci influencing lung function. *Nat Genet* 2011;43:1082–1090.
23. Cho MH, McDonald ML, Zhou X, Mattheisen M, Castaldi PJ, Hersh CP, *et al*.; NETT Genetics Investigators; ICGN Investigators; ECLIPSE Investigators; COPDGene Investigators. Risk loci for chronic obstructive pulmonary disease: a genome-wide association study and meta-analysis. *Lancet Respir Med* 2014;2:214–225.
24. Hobbs BD, de Jong K, Lamontagne M, Bossé Y, Shrine N, Artigas MS, *et al*.; COPDGene Investigators; ECLIPSE Investigators; LifeLines Investigators; SPIROMICS Research Group; International COPD Genetics Network Investigators; UK BiLEVE Investigators; International COPD Genetics Consortium. Genetic loci associated with chronic obstructive pulmonary disease overlap with loci for lung function and pulmonary fibrosis. *Nat Genet* 2017;49:426–432.
25. Jackson VE, Latourelle JC, Wain LV, Smith AV, Grove ML, Bartz TM, *et al*.; Understanding Society Scientific Group. Meta-analysis of exome array data identifies six novel genetic loci for lung function. *Wellcome Open Res* 2018;3:4.
26. Wyss AB, Sofer T, Lee MK, Terzikhan N, Nguyen JN, Lahousse L, *et al*. Multiethnic meta-analysis identifies ancestry-specific and cross-ancestry loci for pulmonary function. *Nat Commun* 2018;9:2976.
27. Soler Artigas M, Wain LV, Miller S, Kheirallah AK, Huffman JE, Ntalla I, *et al*.; UK BiLEVE. Sixteen new lung function signals identified through 1000 Genomes Project reference panel imputation. *Nat Commun* 2015;6:8658.
28. Loh PR, Tucker G, Bulik-Sullivan BK, Vilhjálmsson BJ, Finucane HK, Salem RM, *et al*. Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nat Genet* 2015;47:284–290.
29. Hankinson JL, Odencrantz JR, Fedan KB. Spirometric reference values from a sample of the general U.S. population. *Am J Respir Crit Care Med* 1999;159:179–187.
30. McLaren W, Pritchard B, Rios D, Chen Y, Flicek P, Cunningham F. Deriving the consequences of genomic variants with the ensembl API and SNP effect predictor. *Bioinformatics* 2010;26:2069–2070.
31. Võsa U, Claringbould A, Westra H-J, Bonder MJ, Deelen P, Zeng B, *et al*. Unraveling the polygenic architecture of complex traits using blood eQTL metaanalysis [preprint]. *bioRxiv*; 2018 [accessed 2019 Jul 9]. Available from: www.biorxiv.org/content/10.1101/447367v1.
32. Staley JR, Blackshaw J, Kamat MA, Ellis S, Surendran P, Sun BB, *et al*. PhenoScanner: a database of human genotype-phenotype associations. *Bioinformatics* 2016;32:3207–3209.
33. Hancock DB, Soler Artigas M, Gharib SA, Henry A, Manichaikul A, Ramasamy A, *et al*. Genome-wide joint meta-analysis of SNP and SNP-by-smoking interaction identifies novel loci for pulmonary function. *PLoS Genet* 2012;8:e1003098.
34. Schwabe K, Vacca G, Dück R, Gillissen A. Glucocorticoid receptor gene polymorphisms and potential association to chronic obstructive pulmonary disease susceptibility and severity. *Eur J Med Res* 2009;14:210–215.
35. Kichaev G, Bhatia G, Loh PR, Gazal S, Burch K, Freund MK, *et al*. Leveraging polygenic functional enrichment to improve GWAS power. *Am J Hum Genet* 2019;104:65–75.
36. Haley KJ, Lasky-Su J, Manoli SE, Smith LA, Shahsafaei A, Weiss ST, *et al*. RUNX transcription factors: association with pediatric asthma and modulated by maternal smoking. *Am J Physiol Lung Cell Mol Physiol* 2011;301:L693–L701.
37. Chae SC, Park BL, Park CS, Ryu HJ, Yang YS, Lee SO, *et al*. Putative association of RUNX1 polymorphisms with IgE levels in a Korean population. *Exp Mol Med* 2006;38:583–588.
38. Himes BE, Qiu W, Klanderman B, Ziniti J, Senter-Sylvia J, Szefler SJ, *et al*. ITGB5 and AGFG1 variants are associated with severity of airway responsiveness. *BMC Med Genet* 2013; 14:86.
39. GBD 2016 Causes of Death Collaborators. Global, regional, and national age-sex specific mortality for 264 causes of death, 1980-2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet* 2017;390:1151–1210.
40. Thébaud B. Angiogenesis in lung development, injury and repair: implications for chronic lung disease of prematurity. *Neonatology* 2007;91:291–297.
41. Saito A, Horie M, Nagase T. TGF-β signaling in lung health and disease. *Int J Mol Sci* 2018;19:2460.
42. Frank DB, Peng T, Zepp JA, Snitow M, Vincent TL, Penkala IJ, *et al*. Emergence of a wave of Wnt signaling that regulates lung alveologenesis by controlling epithelial self-renewal and differentiation. *Cell Rep* 2016;17:2312–2325.
43. Logan CY, Nusse R. The Wnt signaling pathway in development and disease. *Annu Rev Cell Dev Biol* 2004;20:781–810.
44. Costa RH, Kalinichenko VV, Lim L. Transcription factors in mouse lung development and function. *Am J Physiol Lung Cell Mol Physiol* 2001; 280:L823–L838.
45. Ng-Blichfeldt JP, Alçada J, Montero MA, Dean CH, Griesenbach U, Griffiths MJ, *et al*. Deficient retinoid-driven angiogenesis may contribute to failure of adult human lung regeneration in emphysema. *Thorax* 2017;72:510–521.
46. Massaro D, Massaro GD. Lung development, lung function, and retinoids. *N Engl J Med* 2010;362:1829–1831.
47. Gon Y, Hashimoto S. Role of airway epithelial barrier dysfunction in pathogenesis of asthma. *Allergol Int* 2018;67:12–17.
48. Yu W, Datta A, Leroy P, O'Brien LE, Mak G, Jou TS, *et al*. Beta1-integrin orients epithelial polarity via Rac1 and laminin. *Mol Biol Cell* 2005;16: 433–445.
49. Burgess JK, Mauad T, Tjin G, Karlsson JC, Westergren-Thorsson G. The extracellular matrix: the under-recognized element in lung disease? *J Pathol* 2016;240:397–409.
50. Parker MM, Hao Y, Guo F, Pham B, Chase R, Platig J, *et al*. Identification of an emphysema-associated genetic variant near *TGFB2* with regulatory effects in lung fibroblasts. *eLife* 2019;8:e42720.
51. Mottis A, Mouchiroud L, Auwerx J. Emerging roles of the corepressors NCoR1 and SMRT in homeostasis. *Genes Dev* 2013;27:819–835.
52. Checkley W, West KP Jr, Wise RA, Baldwin MR, Wu L, LeClerq SC, *et al*. Maternal vitamin A supplementation and lung function in offspring. *N Engl J Med* 2010;362:1784–1794.