Multiethnic meta-analysis identifies ancestry-specific and cross-ancestry loci for pulmonary function

Annah B. Wyss¹ et al.#

Nearly 100 loci have been identified for pulmonary function, almost exclusively in studies of European ancestry populations. We extend previous research by meta-analyzing genome-wide association studies of 1000 Genomes imputed variants in relation to pulmonary function in a multiethnic population of 90,715 individuals of European (N = 60,552), African (N = 8429), Asian (N = 9959), and Hispanic/Latino (N = 11,775) ethnicities. We identify over 50 additional loci at genome-wide significance in ancestry-specific or multiethnic meta-analyses. Using recent fine-mapping methods incorporating functional annotation, gene expression, and differences in linkage disequilibrium between ethnicities, we further shed light on potential causal variants and genes at known and newly identified loci. Several of the novel genes encode proteins with predicted or established drug targets, including KCNK2 and CDK12. Our study highlights the utility of multiethnic and integrative genomics approaches to extend existing knowledge of the genetics of lung function and clinical relevance of implicated loci.
Pulmonary function traits (PFTs), including forced expiratory volume in the first second (FEV1) and forced vital capacity (FVC), and their ratio FEV1/FVC, are important clinical measures for assessing respiratory health, diagnosing chronic obstructive pulmonary disease (COPD), and monitoring the progression and severity of various other lung conditions. Further, even when within the normal range, these parameters are related to mortality, independently of standard risk factors1–3.

In addition to lifestyle and environmental factors, such as smoking and air pollution, genetics influences pulmonary function4–6. Previous genome-wide association studies (GWAS) have identified nearly 100 loci associated with PFTs7–13. These analyses have been primarily conducted using HapMap imputed data among European ancestry populations7–12. Recently, the UK BiLEVE Study (N = 48,943) and SpiroMeta Consortium (N = 38,199) have also examined associations between 1000 Genomes imputed variants and PFTs, but only among Europeans13–15.

The present cohorts for heart and aging research in genomic epidemiology (CHARGE) meta-analysis builds upon previous studies by examining PFTs in relation to the more comprehensive 1000 Genomes panel in a larger study population (90,715 individuals from 22 studies, Table 1) comprised of multiple ancestral populations: European (60,552 individuals from 18 studies), African (8429 individuals from 7 studies), Asian (9959 individuals from 6 studies), and Hispanic/Latino (11,775 individuals from 6 ethnic background groups in 1 study). Along with look-up of our top findings in existing analyses of lung function traits and COPD, we additionally investigate correlation with gene expression in datasets from blood and lung tissue, identify known or potential drug targets for newly identified lung function associated loci, and assess the potential biological importance of our findings using recently developed methods integrating linkage disequilibrium (LD), functional annotation, gene expression, and the multiethnic nature of our data. By conducting a GWAS meta-analysis in a large multiethnic population and employing recently developed integrative genomic methods, we identify over 50 additional loci associated with pulmonary function, including some with functional or clinical relevance.

Results

Ancestry-specific meta-analyses. Each study used linear regression to model the additive effect of variants on PFTs, adjusting for age, sex, height, cigarette smoking, weight (for FVC only), and center, ancestral principal components, and a random familial effect to account for family relatedness when appropriate. Ancestry-specific fixed-effects inverse-variance weighted meta-analyses of study-specific results, with genomic control correction, were conducted in METAL (http://www.sph.umich.edu/csg/abecasis/metal/). Meta-analyses included approximately 11.1 million variants for European ancestry, 18.1 million for African ancestry, 4.2 million variants for Asian ancestry, and 13.8 million for Hispanic/Latino ethnicity (see Methods).

European ancestry meta-analyses identified 17 novel loci (defined as more than 500 kb in either direction from the top variant of a known locus as has been used in previous multiethnic GWAS16), which were significantly (defined as $P < 5.0 \times 10^{-8}$ 14,17) associated with pulmonary function: two loci for FEV1 only, 6 loci for FVC only, 7 loci for FEV1/FVC only, and two loci for both FEV1 and FVC (Table 2, Fig. 1, Supplementary Figures 1 and 2). The African ancestry meta-analysis identified eight novel loci significantly associated with pulmonary function: two loci for FEV1, one locus for FVC, and five loci for FEV1/FVC (Table 3, Supplementary Figures 1–3). Five of these loci were also significant at a stricter $P < 2.5 \times 10^{-8}$ threshold as has been suggested for populations of African ancestry17. Six of the African ancestry loci were identified based on variants with low allele frequencies (0.01–0.02) in African ancestry and which were monomorphic or nearly monomorphic (allele frequency < 0.004) in other ancestries (European, Asian, and Hispanic; Supplementary Table 1). In the Hispanic/Latino ethnicity meta-analysis, we identified one novel locus for FVC (Table 3, Supplementary Figures 1–3). Another locus was significantly associated with FEV1, but this region was recently reported by the Hispanic Community Health Study/Study of Latinos (HCHS/SOL)18. For FEV1/FVC among Hispanics/Latinos, all significant variants were in loci identified in previous studies of European ancestry populations. In the Asian ancestry meta-analysis, all variants significantly associated with PFTs were also in loci previously identified among European ancestry populations (Supplementary Figure 3). Within each ancestry, variants discovered for one PFT were also looked-up for associations with the other two PFTs (Supplementary Table 2).

Multiethnic meta-analysis. In multiethnic fixed-effects meta-analyses of 10.9 million variants, we identified 47 novel loci significantly associated with pulmonary function. Thirteen of these

---

**Table 1 Sample size and location of studies included in the CHARGE consortium 1000 Genomes and pulmonary function meta-analysis**

| Study | Country | Sample size by ancestry |
|-------|---------|-------------------------|
|       |         | European | African | Hispanic/Latino | Asian |
| AGESb | Iceland | 1620      |         |                |      |
| ALHS  | United States | 2844     |         |                |      |
| ABCb  | United States | 8878     | 1837    |                |      |
| CHSb  | United States | 3135     | 566     |                |      |
| FamHS | United States | 1679     |         |                |      |
| FHSb  | United States | 7689     |         |                |      |
| GOYAb | Denmark | 1456      |         |                |      |
| HCHS/CHARGE | United States | 11775    |         |                |      |
| HCSb  | Australia | 1822      | 943     |                |      |
| ABCb  | United States | 1472     | 943     |                |      |
| Healthy | South Korea | 2098      |         |                |      |
| Twin  | United States | 2015     |         |                |      |
| JHS   | United States | 7861     |         |                |      |
| KARE3 | South Korea | 11851     |         |                |      |
| LifeLines | Netherlands | 3787     |         |                |      |
| LLFSb | United States | 1339     | 863     |                |      |
| MESAa | United States | 5460     |         |                |      |
| NEO   | Netherlands | 1357      | 1322    |                |      |
| 1982  | Brazil | 1232      |         |                |      |
| Pelotas | Nigeria | 1135      |         |                |      |
| RSIIb | Netherlands | 2216     |         |                |      |
| Total |         | 60,552    | 8429    | 11,775         | 9959 |

---

*AGES Age Environment Susceptibility Study; ALHS Agricultural Lung Health Study (1180 asthma cases and 1664 controls); ARIC Atherosclerosis Risk in Communities Study; CARDIA coronary artery risk development in young adults; CHS Cardiovascular Health Study; FamHS Family Heart Study; FHS Framingham Heart Study; GOYA Genetics of Obesity Overweight Young Adults Study (670 obese cases and 786 controls); HCHS/SOL Hispanic Community Health Study/Study of Latinos; HCS Hunter Community Study; JHS Jackson Heart Study; KARE3 Korean Association Resource Phase 3 Study; LLFS Long Life Family Study; MESA Multi-Ethnic Study of Atherosclerosis; NEO Netherlands Epidemiology of Obesity Study; RS Rotterdam Study SbStudies included in one or more previous CHARGE papers: Hancock et al. (2010) included ARIC, CHS, FHS, RSII, and RSII; Soler Artigas et al. (2010) included AGES, ARIC, CHS, FHS, Health ABC, ABC, RSII in stage 1 and HCS, CARDIA, LifeLines, MESA, and RSII in stage 2; and Loth et al. (2014) included AGES, ARIC, CARDIA, CHS, FHS, Health ABC, HCS, MESA, RSII, and RSII in stage 1 and LifeLines and LLFS in stage 2.
Table 2 Top variants from novel loci discovered in European ancestry meta-analysis of pulmonary function in the CHARGE consortium

| N  | Value | LOC728989 | FVC rs12724426 | 1:146494027 | a | 0.21 | 31315 |
|----|-------|-----------|----------------|------------|---|-----|-------|
| N  | Value | KCNK2     | FVC rs512597   | 1:215095003 | t | 0.81 | 60507 |
| N  | Value | C1orf140  | DUSP10, FVC rs6657854 | 1:221630555 | a | 0.72 | 60508 |
| N  | Value | AFAP1     | AP3B1, FEV1 rs252746 | 5:77392117 | a | 0.78 | 60551 |
| N  | Value | LINC00340 | BA13, FEV1/FVC rs9351637 | 6:67863782 | t | 0.61 | 60528 |
| N  | Value | BAI3      | FEV1/FVC rs1404154 | 7:146651409 | t | 0.99 | 23748 |
| N  | Value | LINC00340 | BA13, FEV1/FVC rs771924 | 9:1555835 | a | 0.42 | 60507 |
| N  | Value | WNT3      | FEV1 rs1432468 | 17:43685698 | a | 0.79 | 39416 |
| N  | Value | DCC       | FEV1 rs8089865 | 18:50957922 | a | 0.59 | 60509 |
| N  | Value | KLHL22    | FVC rs2236519 | 20:45529571 | a | 0.38 | 60508 |

Additional loci also include C1orf140/DUSP10.

In addition to the fixed-effects multiethnic meta-analysis, we conducted a random-effects meta-analysis using the Han and Eskin method\(^\text{19}\) in METASOFT (http://genetics.cs.ucla.edu/meta/) as a sensitivity analysis. In instances where significant heterogeneity is present, the Han-Eskin method mitigates power loss.\(^\text{19}\) In the Han-Eskin random-effects model, 37 of the 47 loci identified in the fixed-effects model at $P < 5 \times 10^{-8}$ had a $P$ value below the same threshold (Supplementary Table 4). Among the ten loci that did not, eight loci still gave a $P < 5 \times 10^{-7}$ in the Han-Eskin random-effects model (PIK3CB2, SUZ12P1, NCO2/RARB1, CTAGE1/RBBP8, C20orf112, COMT/D1/ZNF503-AS1, EDAR, and RBMS3) while only two did not (CRADD and CCDC41) (Supplementary Table 4). In addition, there were six loci for FEV1/FVC that were genome-wide significant in the Han-Eskin random-effects model that had not quite achieved genome-wide significance in the fixed-effects model: GSTD1/GSTO2 (chr10, rs10883990), FRMD4A (chr10, rs1418884), ETFA/SCAPER (chr15, rs12440815), APP (chr21, rs2830155), A4GNT (chr3, rs9864090), UBASH3B (chr11, rs4935813) (Supplementary Table 4).

X-chromosome meta-analysis. Imputed data for X-chromosome variants were available in 12 studies (ARIC, FHS, CHS, MESA, AGES, ALHS, NEO, RS1, RS2, RS3, JHS, Pelotas; N = 43,153). Among these studies, fixed-effects inverse-variance weighted meta-analyses were conducted separately in males and females using METAL and the resulting sex-specific results were combined using a weighted sums approach. No X-chromosome variants were associated with PFTs at genome-wide significance in ancestry-specific or multiethnic meta-analyses. Although the absence of associations between X-chromosome variants and PFTs could reflect the reduced sample size, previous GWAS of pulmonary function have only identified one variant\(^\text{13}\).

Look-up replication of European and multiethnic novel loci. Our primary look-up replication was conducted in the UK BiLEVE study (N = 48,943)\(^\text{14}\). Since this study only included individuals of European ancestry, we sought replication only for the 52 novel loci (excluding the major histocompatibility complex, MHC) identified in either the European ancestry or multiethnic discovery meta-analyses. Data for the lead variant was available in the UK BiLEVE study for 51 loci, including 49 loci with a consistent direction of effect between our results and those from UK BiLEVE (Supplementary Table 5). Based on a two-sided $P < 9.6 \times 10^{-4}$ (0.05/52), 15 loci replicated for the same trait based on the lead variant from our analysis: DCBLD2/MIR548G,
Fig. 1 Manhattan plots of genome-wide association results for pulmonary function in the following CHARGE meta-analyses: a FEV₁ European ancestry; b FVC European ancestry; c FEV₁/FVC European ancestry; d FEV multiethnic; e FVC multiethnic; f FEV₁/FVC multiethnic. Novel loci indicated by magenta. Significance level (5x10⁻⁸) indicated by dashed line.
Look-up replication of African and Hispanic novel loci. Look-up replication of lead variants for novel African ancestry loci was sought in three smaller studies of African Americans: COPDGene \((N = 3219)^{21,22}\), SAPPHIRE \((N = 1707)^{23,24}\), and SAGE \((N = 1405\); predominantly children)\(^{25}\). We did not find evidence of replication for most of the African ancestry loci identified in our study (Supplementary Table 6). This could possibly reflect low power given the smaller sample sizes of the external studies combined with the low minor allele frequencies (MAF) of most (six out of eight) of the African ancestry variants. We found the strongest evidence for replication for \(R Y R 2\) \((r s 3766889)\). This variant was common \((M A F = 0.18)\) and well imputed \((r^2 > 0.90)\) in CHARGE. The effect size was similar across CHARGE \((\beta = 52.21 \text{ml}, \text{SE} = 4.12 \times 10^{-3})\) and the two adult replication studies \((\text{COPDGene} \quad \beta = 46.85 \text{ml}, \quad \text{P} = 0.03 \quad \text{and} \quad \text{SAPPHIRE} \quad \beta = 22.00 \text{ml}, \quad \text{P} = 0.32)\); meta-analysis of these adult studies resulted in a significant combined association \((\beta = 47.35 \text{ml}, \quad \text{SE} = 8.00 \text{ml}, \quad \text{P} = 3.30 \times 10^{-5})\). In SAGE, which includes mostly children and examined percent predicted values, the result was in the opposite direction and not significant. In our Hispanic ethnicity meta-analysis, the lead variant from the single novel locus \((r s 6746679, \quad \text{DKFZp686O1327/PABPC1P2})\) did not replicate in two smaller external studies of Hispanics: MESA \((N = 806; \quad \text{MESA Hispanics not included in discovery})\) and GALA II \((N = 2203; \quad \text{predominantly children})\)\(^{26}\) (Supplementary Table 6).

Overlap of newly identified loci with COPD. Pulmonary function measures are the basis for the diagnosis of COPD, an important clinical outcome; therefore, we also looked-up the 52 novel loci identified in the European ancestry or multiethnic meta-analyses in the International COPD Genetics Consortium (ICGC). This consortium recently published a meta-analysis of 1000 Genomes imputed variants and COPD primarily among individuals of European ancestry \((N = 15,256 \quad \text{cases} \quad \text{and} \quad 47,936 \quad \text{controls})\)\(^{27}\).
controls), including some of the same individuals included in the present lung function analysis. Ten lead variants representing eight novel loci were associated with COPD at \( P < 9.6 \times 10^{-14} \); RBMS3, OTUD4/SMAD1, TMEM338/ZNF462, NCOR2/SCARB1, SUZ12P1, WNT3, SOGA2, C20orf112 (Supplementary Table 7). Directions of effects were consistent between our results and the COPD findings for these variants (i.e., variants associated with increased pulmonary function values were associated
Peripheral blood samples from 5311 participants in EGCUT, (GTEx) (https://www.gtexportal.org/home/)28; (2) lung samples from 278 individuals in genotype-tissue expression eQTL and mQTL signals variants from all 60 novel loci identified in any ancestry-specific or multiethnic meta-analyses in the following datasets: (1) lung samples from 278 individuals in genotype-tissue expression (GTEx) (https://www.gtexportal.org/home/)28; (2) lung samples from 1111 participants in studies from the Lung eQTL Consortium including Laval University, the University of Groningen and the University of British Columbia29–31; (3) whole blood samples from 5257 Framingham Heart Study participants32; (4) peripheral blood samples from 5311 participants in EGCUT, InCHIANTI, Rotterdam Study, Fehmarn, HIV, SHIP-TREND and DILGOM33; and (5) peripheral blood samples from 2116 participants in four Dutch studies collectively known as BIOS34,35. We examined both whole blood and lung datasets to take advantage of the much larger size, and higher statistical power, of available blood eQTL datasets since we have previously found substantial overlap between lung and blood eQTLs for lung function GWAS loci, as well as complementary information from these two different tissues.29 The Lung eQTL Consortium study used a 10% FDR cut-off, while all other studies used a 5% FDR cutoff (see Supplementary Note 1 for further study descriptions and methods).

A significant cis-eQTL in at least one dataset was found for 34 lead variants representing 27 novel loci (Supplementary Table 8). Of these, 13 loci had significant cis-eQTLs only in datasets with blood samples and three loci only in datasets with lung samples (COMTD1/ZNF503-AS1, FAM168A, SOGA2). Eleven loci had significant cis-eQTLs in both blood and lung samples based on lead variants, with one locus having a significant cis-eQTL across all five datasets (SMAD3) and another four loci having a significant cis-eQTL in four datasets (RAB5B, CRHR1, WNT3, ZNF337). Significant trans-eQTLs in at least one dataset were found for seven lead variants representing four novel loci (TMEM38B/ZNF462, RAB5B, CRHR1, and WNT3, Supplementary Table 8).

In addition, mQTL data were available from FHS and BIOS. Significant cis-mQTLs and trans-mQTLs in at least one dataset were found for 52 lead variants (43 novel loci) and 1 lead variant (1 novel locus), respectively (Supplementary Table 8).

None of the novel variants discovered for African and Hispanic ancestry indicated significant cis-eQTLs in GTex which includes some slight multiethnic representation (12% African American and 3% other races/ethnicities). Although few ancestry-specific eQTL datasets exist, we identified two such studies with blood samples from African American participants: SAPPHIRE (N = 597) and MESA (N = 233)36. In SAPPHIRE, none of the eight African ancestry variants indicated significant cis-eQTLs (FDR < 0.05), but rs180930492 was associated with a trans-eQTL among individuals without asthma and rs11793843 and rs139215025 were associated with trans-eQTLs among individuals with asthma at FDR < 0.05 (Supplementary Table 9). In MESA, eQTL data were available for only three of the African ancestry variants (rs11748173, rs3766889, rs144296676), and none indicated significant (FDR < 0.05) cis-eQTLs (Supplementary Table 9).

Heritability and genetic correlation. We used LD score regression37 to estimate the variance explained by genetic variants investigated in our European ancestry meta-analysis, also known as single nucleotide polymorphisms (SNP) heritability. Across the genome, the SNP heritability (narrow-sense) was estimated to be 20.7% (SE 1.5%) for FEV1, 19.9% (SE 1.4%) for FVC, and 17.5% (SE 1.4%) for FEV1/FVC.

We also partitioned heritability by functional categories to investigate whether particular subsets of common variants were enriched38. We found significant enrichment in six functional categories for all three PFTs: conserved regions in mammals, DNase I hypersensitive sites (DHS), superenhancers, the histone methylation mark H3K4me1 and histone acetylation marks H3K9ac and H3K27ac (Supplementary Figure 4). Another seven categories showed enrichment for at least one PFT (Supplementary Figure 5). We observed the largest enrichment of heritability (14.5–15.3 fold) for conserved regions in mammals, which has ranked highest in previous partitioned heritability analyses for other GWAS traits (Supplementary Figure 5)38.

Since both height and smoking are important determinants of pulmonary function, and have been associated with many

---

Table 6 Top variants from novel loci discovered in multiethnic meta-analysis of FEV1/FVC* in the CHARGE consortium

| Nearest gene(abc) | Top variant | Chr:Pos | Coded allele | Allele freq | N | Betaa | SE | P value |
|------------------|-------------|---------|-------------|------------|---|--------|----|---------|
| DCAF8            | rs1591179   | 11:160206067 | t           | 0.45       | 90,624 | −0.002 | 0.0003 | 3.48E−08 |
| KCN3, NR4A2      | rs72904209  | 2:157046432  | t           | 0.88       | 90,453 | 0.003  | 0.0005 | 3.09E−08 |
| RBMS3            | rs28732417  | 3:29431565   | a           | 0.74       | 90,358 | 0.002  | 0.0004 | 1.77E−08 |
| DCBLD2, MIR548G  | rs80217917  | 3:99359368   | t           | 0.88       | 90,617 | −0.003 | 0.0005 | 2.58E−08 |
| AFAP1            | rs2852091   | 4:7846240    | t           | 0.44       | 80,715 | 0.002  | 0.0004 | 8.40E−09 |
| LINCO0340        | rs930408    | 6:22021373   | t           | 0.51       | 82,761 | −0.003 | 0.0003 | 7.45E−14 |
| FLJ35282, ELAVL2 | rs10965947  | 9:2358858    | t           | 0.39       | 90,475 | 0.002  | 0.0004 | 2.70E−09 |
| TMEM368, ZNF462  | rs2451951   | 9:10949630   | t           | 0.47       | 88,436 | 0.002  | 0.0003 | 2.36E−08 |
| JMJD1C           | rs75159994  | 10:124273671 | t           | 0.23       | 90,481 | 0.002  | 0.0004 | 1.51E−08 |
| HTRA1            | rs2293871   | 11:73280955: GA_G | d | 0.20 | 55,521 | 0.004 | 0.0006 | 2.74E−08 |
| FAM168A          | rs111793843 | 11:73280955  | d           | 0.20       | 55,521 | 0.004 | 0.0006 | 2.74E−08 |
| DDHDI, MIR588    | rs4442435   | 14:54410919  | t           | 0.54       | 80,712 | 0.002  | 0.0004 | 4.03E−08 |
| TSHZ3            | rs9636166   | 19:31829613  | a           | 0.86       | 80,714 | 0.003  | 0.0005 | 3.25E−09 |
| KHLH22, MED15    | rs4820216   | 22:20854161  | t           | 0.82       | 82,714 | −0.003 | 0.0005 | 2.61E−10 |

*Phenotype: ratio FEV1/FVC (as a proportion).

**Nearest gene: indicates gene either harboring the variant or nearest to it. HTRA1 locus also includes DMBT1. JMJD1C locus also includes EGR2, NRBF2, JMJD1C-AS1, REEP3. KHLH22/MED15 locus also includes ZNF74, SCA92.

***Loci also discovered in European ancestry meta-analyses (Table 2): RBMS3, AFAP1, LINCO0340, TSHZ3, KHLH22/MED15.

^Alleles for INDELS: I insertion, D deletion.

*Additive effect of variant on pulmonary function, adjusting for age, age2, sex, height, height2, smoking status, pack-years of smoking, weight (for FVC only), and center, ancestral principal components, and a random familial effect to account for family relatedness when appropriate.
common variants in previous GWAS, we also used LD score regression to investigate genetic overlap\(^9\) between our FEV\(_1\), FVC, and FEV\(_1\)/FVC results and publicly available GWAS results of ever smoking\(^{40}\) and height\(^{41}\). No significant genetic correlation was found between PFTs and smoking or height (Supplementary Table 10), indicating our findings are independent of those traits.

In addition, we used LD Score regression to investigate genetic overlap between each PFT and the other two PFTs, as well as with asthma. Based on the overall PFT results presented in this paper, we found significant genetic correlation between FEV\(_1\) and FVC (\(P < 0.001\)) and between FEV\(_1\) and FEV\(_1\)/FVC (\(P < 0.001\)), but not between FVC and FEV\(_1\)/FVC (\(P = 0.23\); Supplementary Table 10).

Since measures of FEV\(_1\) and FVC (independent of genetics) are involved in inflammatory and immunity related networks (Supplementary Table 13), IPA based on the multiethnic results highlighted 21 enriched networks, including similar inflammatory and immunity related networks (Supplementary Table 14).

### Identification of potential causal variants using PAINTOR.

Using a multiethnic fine-mapping analysis incorporating strength of association, variation in genetic background across major ethnic groups, and functional annotations in Probabilistic Annotation NetraTOR (PAINTOR)\(^{50}\), we examined 38 loci that contained at least five genome-wide significant variants in the European ancestry and multiethnic meta-analyses or at least one significant variant in the African ancestry or Hispanic/Latino ethnicity meta-analyses. We identified 15 variants representing 13 loci as having high posterior probabilities of causality (\(>0.8\)): 3 for FEV\(_1\), 3 for FVC, and 9 for FEV\(_1\)/FVC (Supplementary Table 15, Supplementary Figure 7). Of the 15 putative causal variants, 11 showed high posterior probabilities of causality (\(>0.8\)) before considering annotations, and 4 were identified by adding functional annotations. Nine were the top SNPs at that locus from the fixed-effects meta-analysis (loci: WNT3, PMFBP1/ZFHX3, EN1/MARCO, C2orf48/HPCAL1, CPT1C, CADPS, LOC283867/CDH5, HDC, and CDC7/TGFBR3), while 6 were not (loci: CDK2/RAB5B, BMS1P4, PMFBP1/ZFHX3, FLJ35282/ELAVL2, HDC, and COL8A1).

### Identification of independent signals using FINEMAP.

We used FINEMAP\(^{51}\) to identify variants with a high posterior probability of causality (\(>0.6\)) independent of 118 lead variants in pulmonary function loci identified in the current or previous studies\(^{14}\). We identified 37 independent variants for 23 previously identified loci and one independent variant at each of two novel loci (LINC00340 and SL2C25A1P1/BAI3; Supplementary Table 16).

### Gene-based analysis of GWAS results using S-PrediXcan.

Among the novel loci identified in the current GWAS of PFTs, we identified seven variants corresponding to nine genes demonstrating genome-wide significant evidence of association with lung or whole blood tissue-specific expression (Supplementary Table 17) based on the gene-based S-PrediXcan approach\(^{52}\). Bayesian colocalization analysis\(^{53}\) indicated the following associations demonstrated at least 50% probability of shared SNPs underlying both gene expression and PFTs: ARHGEF17 and FAM168A in analysis of multiethnic GWAS for FEV\(_1\)/FVC based on GTEx whole blood models, and WNT3 in analysis of multiethnic GWAS for FVC based on GTEx lung models (Supplementary Table 18).

### Druggable targets.

To investigate whether the genes identified in our study encode proteins with predicted drug targets, we queried the ChEMBL database (https://www.ebi.ac.uk/chembl/). In addition, we incorporated an IPA to identify potential upstream targets. Genes associated with pulmonary function, but not included in the drug target analysis performed by Wain et al.\(^{14}\), were evaluated, for a total of 139 genes outside of the MHC. 110 genes representing the 60 novel loci identified in our fixed-effects ancestry-specific and multiethnic meta-analysis, 13 genes representing the 6 novel loci identified in our random-effects meta-analysis\(^{19}\), 3 genes representing an additional 3 loci near significance in the African ancestry meta-analysis (BAZZ2, NONE/PCDH10, and ADAMTS17), 9 genes representing 2 loci identified in a recent CHARGE analysis of exome variants\(^{43}\), which were also significant in our 1000 Genomes analysis (LY86/RREB1 and SEC24C), and 4 genes representing one locus identified at genome-wide significance in a separate publication from...
one of our participating studies (HCHS/SOL)\textsuperscript{18}, but also significant in our analysis (ADORa2B/ZSWIM27/TTC19/NCOR1). In the ChEMBL database, 17 of these genes encode proteins with predicted or known drug targets: NRS5A2, KCNK2, EDAR, KCNJ3, NRA4A2, BAZ2B, AAGNT, GSTO1, GSTO2, NCOR2, SMAD3, NCO1, CDK12, WNT3, PYGB, NAPN, EYA2 (Supplementary Table 19). Of these, two genes (KCNK2 and CDK12) have approved drug targets. Using IPA, four additional genes were predicted as drug targets (ADORa2B, APP, CRHR1, and MAP3K1; Supplementary Table 20) and 37 genes had drugs or chemicals as upstream regulators (Supplementary Table 21).

**Discussion**

By conducting a GWAS meta-analysis in a large multiethnic population we increased the number of known loci associated with pulmonary function by over 50%. In total, we identified 60 novel genetic regions (outside of the MHC region): 17 from European ancestry, 8 from African ancestry, 1 from Hispanic/Latino ethnicity, and 34 from a multiethnic meta-analyses.

For 32 of the 52 loci novel loci identified in our European ancestry and multiethnic meta-analyses, we found evidence for look-up replication in the UK BiLEVE study, UK Biobank study, or ICGC COPD consortium. For an additional three loci, we found support for validation using new genomic methods such as PAINTER, FINEMAP, or S-PrediXcan. Specifically, 19 novel variants replicated in look-up in a smaller independent sample of Europeans from the UK BiLEVE study\textsuperscript{14}. DCBLD2/MIR548G, SUZ12P1, CRHR1, WNT3, ZNF337, ALX1/RASSF9, MED1/CDK12, EYA2, RBM33, LINC00340, FLJ35282/ELAVL2, DDHD1/MIR5580, TSHZ3, KLHL22/MED15, FAM168A, RAB5B, JMJ1D1C, AGMO, and C20orf112. Based on a minimally adjusted publicly available analysis in a larger sample of Europeans from the UK Biobank, an additional 11 loci replicated: NRS5A2, PIK3C2B, OTUD4/SMAD1, AP3B1, CENPW/RSPO3, SMAD3, PDXDC2P, SOGA2, DCC, DNAH12, and KCNJ3/NRA4A2. Because UK BiLEVE is sampled from UK Biobank we are not able to perform a combined replication meta-analysis. Additionally, the studies adjusted for different covariates (UK BiLEVE results were adjusted for age, sex, height, pack-years and ancestral principal components while UK Biobank results were adjusted for only sex and ancestral components). Among those loci which did not directly replicate for PFTs in the UK BiLEVE or UK Biobank datasets, the lead variants in an additional two European or multiethnic loci were significantly associated in the ICGC Consortium with COPD, which was defined using PFT measures\textsuperscript{27}. TMEM38B/ZNF462 and NCO2/SCAR8B. FINEMAP and S-PrediXcan also produced evidence for causality for three European ancestry and multiethnic loci which had not replicated in UK BiLEVE, UK Biobank or ICGC: DCAF8, AFAPI, and SLC25A5/1P/BAI3.

The few additional studies with 1000 Genomes imputed variants and pulmonary function in African ancestry individuals have smaller samples sizes making replication challenging for the eight novel loci identified in our African ancestry meta-analyses. Further, lead variants for six of the eight loci were low frequency in African Ancestry (C2orf48/HPCAL1, EN1/MARCO, CAPDS, HDC, LOC283867/CDHS, and CPTIC) (MAF < 0.02), including three not well imputed (\(r^2 < 0.75\)), and monoallelic or nearly monoallelic in other ancestries (European, Asian, and Hispanic). For the two novel African ancestry variants with MAF > 0.02 and well imputed (\(r^2 > 0.90\)), we found the strongest evidence for replication for RYR2 (rs3766889). This variant had a similar effect estimate for FEV\(_1\) in CHARGE, COPDGene, and SAPPHIRE with a significant combined association across these adult studies. Although this particular variant did not quite meet genome-wide significance in the multiethnic meta-analysis for FEV\(_1\) (\(P = 6.56 \times 10^{-5}\)), another variant in this gene did for FVC (1.237929787:1.1, \(P = 4.46 \times 10^{-10}\)).

Our analysis also sheds light on additional potential causal genes at a complex locus (chromosome 17 near positions 43600000 to 44300000, hg19) previously discovered from GWAS of FEV\(_1\), which identified KANSL1 in European populations as the top finding for this region\textsuperscript{14,15}. With the exception of a single INDEL in KANSL1 in our European ancestry meta-analysis (17:44173680:T_TC, \(P = 1.03 \times 10^{-10}\)), we found CRHR1 as the strongest gene associated with FEV\(_1\) in this region. Although some variants in CRHR1 identified in our study are within 500kb of KANSL1 (e.g., rs16940672, 17:43908152, \(P = 9.06 \times 10^{-10}\)). In our multiethnic meta-analysis, several variants in CRHR1 were associated with FEV\(_1\) at smaller \(P\) values than variants in KANSL1. Definitive assessment of the causal variants at this locus, as well as other multigene GWAS loci, will likely require additional data from ongoing large-scale sequencing studies to enable detailed fine mapping.

In both our European and multiethnic meta-analyses we also noted a significant association with WNT3 on chromosome 17 near position 44800000 (hg19) which is more than 500kb from KANSL1 or CRHR1 [our definition of novel]. We found that the top variant in WNT3 for FEV\(_1\) among individuals of European ancestry (rs916888, 17:44863133, \(P = 3.76 \times 10^{-9}\)) had a high probability for causality based on PAINTOR, an analysis which integrates functional annotations along with association statistics and LD for each ethnicity\textsuperscript{50}. We also found evidence that WNT3 may be the causal gene at this locus using S-PrediXcan, a gene level association test that prioritizes potentially causal genes while filtering out LD-induced false-positives\textsuperscript{52,53}. Notably, S-PrediXcan implicated WNT3 as a likely mediating gene for FVC based on the top variant in our multiethnic meta-analyses (rs199525, 17:44847834, \(P = 7.52 \times 10^{-9}\)), which is an eQTL SNP for WNT3 in lung and other tissues. Further, the lead WNT3 variants for both FEV\(_1\) and FVC (rs916888 and rs199525) were significantly associated with COPD in a look-up of a large published meta-analysis dataset\textsuperscript{27}. In addition, other genes in the WNT signaling pathway, a fundamental development pathway, have been implicated as influencing pulmonary function\textsuperscript{54}. This pathway was also one of the significant pathways identified in our analysis. In a previous pathway analysis of asthma, SMAD3 has been shown to interact with the WNT signaling pathway\textsuperscript{55}. Finally, WNT3 also emerged as having a potential druggable target, and incorporation of pathway analysis to identify upstream regulators found an additional four drugs in clinical use for which WNT3 is a target molecule (chemotherapeutic agents doxorubicin and paclitaxel, the hormone beta-estradiol and LGK-974, a novel agent that targets a WNT-specific acyltransferase)\textsuperscript{56}. Again, further evaluation of this interesting and complex locus which contains many significant variants in LD will benefit from data being generated in ongoing large-scale sequencing studies.

Some genes identified in our study play key roles in inflammation, immunity, and pulmonary biology. For example, MARCO (macrophage receptor with collagenous structure) has been shown in murine models to be required for lung defense against pneumonia and inhaled particles\textsuperscript{57,58}. SMAD3 is part of the SMAD family of proteins which are signal transducers and transcriptional modulators that mediate multiple signaling pathways. SMAD3 is activated by transforming growth factor beta (TGF-B) which plays a key role in airway remodeling. SMAD3

---

**ARTICLE**

**NATURE COMMUNICATIONS** | (2018) 9:2976 | DOI: 10.1038/s41467-018-05369-0 | www.nature.com/naturecommunications

---

**AUTHOR CONTRIBUTIONS**

This project was conceived and planned by S.M., J.L., A.K., A.F., J.T., and W.C.S. The fine-mapping and association analysis were carried out by D.T.C., J.L., and M.K. by using code written by D.T.C.. The post-processing of the dataset was carried out by J.L., W.C.S., and A.K. The lead GWAS meta-analysis was carried out by S.M., J.L., A.K., A.F., and W.C.S. The multiethnic meta-analysis was carried out by J.L., A.K., A.F., M.K., and W.C.S. The analysis of the replication meta-analysis was carried out by J.L., M.K., and W.C.S. The writing of the manuscript was done by S.M., J.L., A.K., A.F., and W.C.S. The figures and tables were prepared by J.L., M.K., and W.C.S. The authors listed above are to be considered for authorship and are listed in alphabetical order. Additional technical contributors are gratefully acknowledged in the Acknowledgments section.
has a predicted drug target and SNPs in SMAD3 are significantly associated with asthma in GWAS42,59.

Other genes identified in our study that are targeted by approved drugs include CDK12 and KCNK2. CDK12 drug targets include AT-7519, Ronicitlib, AZD-5438, and PHA-793887. Ronicitlib has been used in clinical trials including lung cancer patients60. KCNK2 (potassium channel subfamily K member 2) is targeted by five inhalational anesthetic agents. These agents have antiinflammatory effects both systemically61 and in the lungs62 and meta-analysis of clinical studies shows protection against pulmonary complications after cardiac surgery63. A recent trial suggested that one of these inhalation agents, sevoflurane, offers promise for reducing epithelial injury and improving outcomes in patients with acute respiratory distress syndrome64.

In addition to querying commonly used genome databases for functional annotation of variants, we sought to narrow down causal variants in implicated loci using recently developed methods that incorporate LD, functional data and/or the multi-ethnic analysis done in this paper. In particular, PAINTOR is a useful tool to identify potential causal variants in our novel loci as it leverages LD across ancestral groups along with association statistics and functional annotations50. PAINTOR identified 15 putative causal variants from 13 loci, including seven loci uniquely identified in the multiethnic meta-analyses such as PMFBP1/ZFHX3 and COLRA1 (part of the DCBLD2 loci). Several of the putative causal variants from PAINTOR were the top SNPs from the fixed-effects meta-analysis (e.g., rs916888 WNT3). Similarly, FINEMAP has been shown to be an accurate and efficient tool for investigating whether lead SNPs for a given loci are driven by independent variants in the same region, especially when annotation information is not available51. Among previous and novel loci identified in individuals of European ancestry, we identified 37 independent variants for 23 previously identified loci and two lead variants for two novel loci (rs1928168 LINC00340 and rs9351637 SLC2A5S1P1/BAI3) with a high probability of causality. Finally, we ran S-PrediXcan a gene level association test that prioritizes potentially causal genes52. Seven of our novel loci contained putative causal genes based on S-PrediXcan for lung or whole blood tissues, including NRFB2 (part of the JMJD1C locus) and WNT3. S-PrediXcan also highlighted the region around chromosome 11 position 73280000 (hg19), noting strong evidence for both FAM168A and ARHGEF17 which was further supported by the colocalization analysis. Interestingly, DEPICT also prioritized ARHGEF17, a member of the guanine nucleotide exchange factor (GEF) family of genes which can mediate actin polymerization and contractile sensitization in airway smooth muscle65,66.

Rather than conducting a standard gene-based analysis, we performed a newer integrative method, DEPICT, that incorporates cell and tissue-specific functional data into a pathway analysis to prioritize genes within implicated loci49. In addition to identifying potential causal variants, this approach revealed a number of fundamental development processes, including pathways related to lung development, growth regulation, and organ morphogenesis. The WNT signaling pathway was also highlighted along with processes relevant to the pathogenesis of COPD including extracellular matrix structure and collagen networks. Tissue/cell type enrichment results highlighted smooth muscle which is highly relevant for lung function. DEPICT excludes the MHC due to extended LD in this region, which likely explains the relative paucity of inflammation-related pathways identified compared to previous pathway analyses in GWAS of PFTs29,54. Indeed, standard IPA analysis of our data including the MHC region, found that 33 of 84 genes (39%) in the 3 (out of 16) enriched networks involved in immune or inflammatory processes are in the MHC. The predominance of fundamental pathways related to lung growth, differentiation and structure is consistent with recent work57 that has rekindled interest in the observation made 40 years ago68 that individuals can cross the threshold for diagnosis of COPD either by rapid decline in adulthood or by starting from a lower baseline of maximal pulmonary function attained during growth. Within this context, understanding the genetic (and environmental) factors that influence the variability in maximal lung function attained during the first three decades of life is essential to reducing the public health burden of COPD69.

In summary, our study extends existing knowledge of the genetic landscape of PFTs by utilizing the more comprehensive 1000 Genomes imputed variants, increasing the sample size, including multiple ancestries and ethnicities, and employing newly developed computational applications to interrogate implicated loci. We discovered 60 novel loci associated with pulmonary function and found evidence of replication in UK BiLEVE, UK Biobank, or ICGC for 32 novel loci and validation for another 3 loci. We also identified evidence of seven variants in these loci were missense mutations and had possible deleterious or regulatory effects, and many had significant eQTLs. Further, using new genomic methods that incorporate LD, functional data and the multiethnic structure of our data, we shed light on potential causal genes and variants in implicated loci. Finally, several of the newly identified genes linked to lung function are druggable targets, highlighting the clinical relevance of our integrative genomics approach.

### Methods

#### Studies

Member and affiliate studies from The CHARGE consortium with pulmonary function and 1000 Genomes imputed genetic data were invited to participate in the present meta-analysis. Participating studies included: AGES, ALHS, ARIC, CARDIA, CHS, FamHS, FHS, GOYA, HCHS/SOL, HCS, Health ABC, Healthy Twin, HJS, KARE3, LifeLines, LLPS, MESA, NEO, 1982 PELOTA, RSI, RII, RIII. Characteristics of these studies are provided in Supplementary Table2 and descriptions of study designs are provided in the Supplementary Note 3. Informed consent was obtained from participants in each study. Although our meta-analysis included studies of asthma (ALHS) and obesity (GOYA and NEO), exclusion of these studies did not materially change results (Supplementary Note 2). Further, previous meta-analyses of GWAS of pulmonary function have demonstrated high correlation between results when including or excluding asthma and COPD cases8.

#### Pulmonary function.

Spirometry measures of pulmonary function (FEV1, FVC, and the ratio FEV1/FVC) were collected by trained staff in each study according to American Thoracic Society or European Respiratory Society guidelines. See cohort descriptions in Supplementary Note 1 for more details.

#### Variants

Studies used various genotyping platforms, including Affymetrix Human Array 6.0, Illumina Human Omni Chip 2.5, and others, as described in cohort descriptions in the Supplementary Note 1. Using MACH, MINIMAC, or IMPUTE2, studies then used genotyped data to impute variants based on the 1000 Genomes Integrated phase 1 reference panel. One study (Hunter Community) imputed to the 1000 Genomes European phase 1 reference panel: sensitivity analyses excluding this study from the European ancestry meta-analysis showed no material differences (see Supplementary Note 2). The two Asian studies (Healthy Twin and KARE3) imputed to the 1000 Genomes Asian phase 1 reference panel.

#### Statistical analysis.

Within each study, linear regression was used to model the additive effect of variants on PFTs. FEV1 and FVC were modeled as milliliters and FEV1/FVC as a proportion. Studies were asked to adjust analyses for age, age squared, sex, height, height squared, smoking status (never, former, and current), pack-years of smoking, center (if multicenter study), and ancestral principal components, including a random familial effect to account for family relatedness when appropriate69.

Models of FVC were additionally adjusted for weight. Analyses were conducted using ProphAbel, PLINK, FAST, or the R kinship package as described in the cohort descriptions of the Supplementary Note 1.

#### Ancestry-specific and multiethnic fixed-effects meta-analyses using inverse variance weighting of study-specific results with genomic control correction were conducted in Meta-Analysis Helper (METAL, http://www.sph.umich.edu/csg/ were used for the four ancestry-specific fixed-effects meta-analysis results were conducted using the Han-Eakin model19 in METASOFT (http://genetics.cs.ucd.edu/metasoft/). Only variants
with p-values for association <0.05 or P values for heterogeneity <0.1 from fixed-effects models were included in the random-effects models.

Variants with imputation quality scores (r2) > 0.3 and/or a minor allele count (MAC) less than 20 were excluded from each study prior to meta-analysis. Following meta-analysis, we also excluded variants with less than one-third the total sample size and with the sample size of the largest study for a given meta-analysis to achieve the following minimal sample sizes: 20,184 for European ancestry; 2810 for African ancestry; 4435 for Hispanic/Latino ethnicity; and 30,238 for Multiethnic.

Significance was defined as P < 5 × 10−8. Novel variants were defined as being more than ±500 kb from the top variant of a loci identified in a previous GWAS. To create the previous GWAS, we used association statistics (Z-scores) and LD information for each locus for each ancestry. We included our ancestry-specific meta-analysis results and used the American, European, and East Asian individuals from 1000 Genomes to calculate LD3. For PAINTOR we focused on 22 novel loci identified in our European ancestry and multiethnic fixed-effects meta-analyses which had at least five genome-wide significant variants as well as all nine African or Hispanic loci which had at least one genome-wide significant variant. In addition, we included six loci which overlapped with the UK BiLEVE 1000 Genomes paper34 and one locus with the CHARGE exome paper35, since we ran PAINTOR prior to those publications. To reduce computational burden, we limited flanking regions to ±20 kilobase (kb) from the top SNPs and included variants with absolute value of Z-score greater than 1.96.

We used 269 publicly available annotations relevant to “lungs”, “bronch”, or “pulmo” from the following: hypersensitivity sites2, superenhancers4, Fantom5 enhancer and transcription start site regions4, immune cell enhancers46, and methylation and acetylation marks3. We ran each phenotype separately to prioritize annotations based on likelihood-ratio statistics7,8,9. We included minimally correlated top annotations (less than five for each phenotype) to identify causal variants.

Pathway analysis using DEPICT and IPA. For gene prioritization and identification of enriched pathways and tissue/cell types, we used IPATH49 with associated results for FEV1, FVC and FEV1/FVC. We used 269 publicly available annotations relevant to human genome, ranks variants, and generates raw and scaled posterior probabilities of >0.8. For gene prioritization and identification of enriched pathways and tissue/cell types, we used IPATH49 with associated results for FEV1, FVC and FEV1/FVC. We used 269 publicly available annotations relevant to human genome, ranks variants, and generates raw and scaled posterior probabilities of >0.8.

Functional annotation. To find functional elements in novel genome-wide significant signals, we annotated SNPs using various databases. We used Ensembl VEP44 (Accessed 17 Jan 2017) and obtained mapped genes, transcripts, consequence of variants on protein sequence, SIFT45,46 scores, and PolyPhen-247 scores. We checked if there were deleterious variants using CADD v1.347, which integrates various functional annotation databases.

Pathway analysis using DEPICT and IPA. For gene prioritization and identification of enriched pathways and tissue/cell types, we used IPATH49 with associated results for FEV1, FVC and FEV1/FVC. We used 269 publicly available annotations relevant to human genome, ranks variants, and generates raw and scaled posterior probabilities of >0.8. For gene prioritization and identification of enriched pathways and tissue/cell types, we used IPATH49 with associated results for FEV1, FVC and FEV1/FVC. We used 269 publicly available annotations relevant to human genome, ranks variants, and generates raw and scaled posterior probabilities of >0.8.

FENEMAP. We used FENEMAP50 to identify signals independent of lead variants for pulmonary function loci identified in the current or previous studies15. The Rotterdam Study (N = 6291), one of the largest cohort studies included in our multi-ethnic analysis, was used as a reference population. SNPs with MAF of <1% were excluded, leaving 118 SNPs for analysis. Ten SNPs for FEV1 and FVC and 20 SNPs for FEV1/FVC were further excluded because the LD matrix of the reference file from the Rotterdam Study did not represent the correlation matrix of the total study population. We allowed up to 10 causal SNPs per variant in FENEMAP analyses. To reduce the chance of false positive findings, we also conducted sensitivity analyses allowing up to 15 causal SNPs for loci with more than four SNPs with posterior probabilities of >0.8.

S-PrediXcan. S-PrediXcan is a summary statistics based approach for gene-based analyses51 that was derived as an extension of the PrediXcan method for integration of GWAS and reference transcriptome data20. We used the S-PrediXcan approach for genes potential causally associated with a B3 association result and publicly available GWAS of ever smoking52 and height53 was also investigated using LD score regression49.

FENEMAP. We used FENEMAP50 to identify signals independent of lead variants for pulmonary function loci identified in the current or previous studies15. The Rotterdam Study (N = 6291), one of the largest cohort studies included in our multi-ethnic analysis, was used as a reference population. SNPs with MAF of <1% were excluded, leaving 118 SNPs for analysis. Ten SNPs for FEV1 and FVC and 20 SNPs for FEV1/FVC were further excluded because the LD matrix of the reference file from the Rotterdam Study did not represent the correlation matrix of the total study population. We allowed up to 10 causal SNPs per variant in FENEMAP analyses. To reduce the chance of false positive findings, we also conducted sensitivity analyses allowing up to 15 causal SNPs for loci with more than four SNPs with posterior probabilities of >0.8.

S-PrediXcan analysis was performed using the following publicly available tissue-specific expression models (http://prediclibd.org) from the GTEx project:55-57: (1) GTEx Lung (278 samples) and (2) GTEx whole blood (338 samples). Approximately, 85% of participants in GTEx are white, 12% African American, and 3% of other races/ethnicities. Gene-based S-PrediXcan results were filtered on the following: (1) proportion of SNPs used (~n SNPs available in GWAS summary data); (2) nominal P values in prediction model 0.05 and (2) nominal P value for each gene-based test divided by the number of genes passing specified filters in each analysis to test whether genetically regulated gene expression was associated with the trait of interest. The genome-wide S-PrediXcan results were then merged with novel loci from the current GWAS study by identifying all matches in which the novel locus SNP was within 500kb of the start of the gene.

We further incorporated a Bayesian colocalization approach58 to interpret the extent to which S-PrediXcan results may have been induced by LD within the region of interest. The Bayesian colocalization procedure was run using the following priors: p1 = 1e-4; prior probability SNP associated to trait 1, p2 = 1e-4; prior probability SNP associated to trait 2, p12 = 1e-5; prior probability SNP associated to both traits. The procedure generated posterior probabilities that were used to identify gene loci with a significant association with neither trait, (H1) associated with PFT phenotype but not gene expression, (H2) associated with gene expression but not PFT phenotype, (H3) associated with
both traits, due to two independent SNPs, and (H4) associated with both traits, due to one shared SNP.

**Drug targets.** We searched annotated gene lists against the ChEMBL database (v22.1, updated on November 15, 2016) to identify genes as targets of approved drugs or drugs in development. In addition, we used the Ingenuity Pathway Analysis (IPA, www.ingenuity.com, content of 2017-06-22) to identify drug targets and upstream regulators of the gene lists. We reported the upstream regulators in the following categories, biologic drug, chemical—endogenous mammalian, chemical—kinase inhibitor, chemical—other, chemical drug, chemical reagent, and chemical toxicant.

**Data availability.** The complete meta-analysis results have been deposited in the database of Genotypes and Phenotypes (dbGaP) under the CHARGE accession number phs000390. GWAS data for most UK studies are already available dbGaP. For all other studies, please send requests to the study PI or Stephanie London (london2@niehs.nih.gov) who will forward them to the relevant party. Requests for METAL code can be directed to Stephanie London.

Received: 11 October 2017 Accepted: 21 June 2018

Published online: 30 July 2018

**References**

1. Hole, D. J. et al. Impaired lung function and mortality risk in men and women: findings from the Renfrew and Paisley prospective population study. *Br. Med. J.* 313, 711–715 (1996), discussion 715–716.
2. Young, R. P., Hopkins, R. & Eaton, T. E. Forced expiratory volume in one second: not just a lung function test but a marker of premature death from all causes. *Eur. Respir. J.* 30, 616–622 (2007).
3. Schunemann, H. J., Dorn, J., Grant, B. J., Winkelstein, W. Jr & Trevisan, M. Pulmonary function is a long-term predictor of mortality in the general population: 29-year follow-up of the Buffalo Health Study. *Chest* 118, 650–664 (2000).
4. Wilk, J. B. et al. Evidence for major genes influencing pulmonary function in the NHLBI family heart study. *Genet. Epidemiol.* 19, 81–94 (2000).
5. Givelber, R. J. et al. Segregation analysis of pulmonary function among families in the Framingham Study. *Am. J. Respir. Crit. Care Med.* 157, 1445–1451 (1998).
6. Chen, Y., Horne, S. L., Rennie, D. C. & Dosman, J. A. Segregation analysis of two lung function indices in a random sample of young families: the Humboldt Family Study. *Genet. Epidemiol.* 13, 35–47 (1996).
7. Repapi, E. et al. Genome-wide association study identifies five loci associated with lung function. *Nat. Genet.* 42, 36–44 (2010).
8. 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).
9. 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).
10. Westra, H. J. et al. Systematic identification of trans-eQTLs as putative drivers of known disease associations. *Nat. Genet.* 45, 1238–1243 (2013).
11. Bender, M. J. et al. Disease variants alter transcription factor levels and methylation of their binding sites. *Nat. Genet.* 49, 131–138 (2017).
12. Zhanakova, D. V. et al. Identification of context-dependent expression quantitative trait loci aids interpretation of genomic association studies. *Genome Biol.* 18, 16 (2017).
13. Westra, H. J. et al. Systematic identification of trans-eQTLs as putative drivers of known disease associations. *Nat. Genet.* 45, 1238–1243 (2013).
14. Bowman, N. et al. Lung eQTLs with high expression levels are associated with chronic lung disease. *Nat. Genet.* 47, 291–295 (2015).
15. Finucane, H. K. et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet.* 47, 1228–1235 (2015).
16. Bulik-Sullivan, B. K. et al. LD Score regression distinguishes confounding from polygenicity of genomic association studies. *Nat. Genet.* 47, 1005–1012 (2015).
17. Bulik-Sullivan, B. K. et al. LD Score regression distinguishes confounding from polygenicity of genomic association studies. *Nat. Genet.* 47, 1005–1012 (2015).
18. Wood, A. R. et al. Defining the role of common variation in the genomic and biological architecture of adult human height. *Nat. Genet.* 46, 1113–1118 (2014).
19. Moffatt, M. F. et al. A large-scale, consortium-based genome-wide association study of asthma. *N. Engl. J. Med.* 363, 1211–1221 (2010).
20. Jackson, V. E. et al. Meta-analysis of exome array data identifies six novel genetic loci for lung function (version 2; referees: 2 approved). *Wellcome Open Res.* 3 https://doi.org/10.12688/wellcomeopenres.12583.2 (2018).
21. McLaren, W. et al. The Ensembl variant effect predictor. *Genome Biol.* 17, 122 (2016).
22. Kumar, P., Henikoff, S. & Ng, P. C. Predicting the effects of coding nonsynonymous variants on protein function using the SIFT algorithm. *Nat. Protoc.* 4, 1073–1082 (2009).
23. Adzhubei, I. A. et al. A method and server for predicting damaging missense mutations. *Nat. Methods* 7, 248–249 (2010).
24. Kircher, M. et al. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat. Genet.* 46, 310–315 (2014).
25. Boyle, A. P. et al. Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res.* 22, 1790–1797 (2012).
26. Pers, T. H. et al. Biological interpretation of genome-wide association studies using predicted gene functions. *Nat. Commun.* 6, 5890 (2015).
50. Kichaev, G. & Pasaniuc, B. Leveraging functional-annotation data in transethnic fine-mapping studies. Am. J. Hum. Genet. 97, 260–271 (2015).

51. Bentzen, M., et al. FINEMAP: efficient variable selection using summary data from genome-wide association studies. Bioinformatics 32, 1493–1501 (2016).

52. Barbeira, A. E. et al. Exploring the phenotypic consequences of tissue-specific gene expression variation inferred from GWAS summary statistics. Nat. Commun. 9, 1825 (2018).

53. Visscher, P. M., et al. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. PLoS Genet. 10, e1004383 (2014).

54. Gharib, S. A. et al. Integrative pathway genomics of lung function and airflow obstruction. Hum. Mol. Genet. 24, 6836–6848 (2015).

55. Barreto-Luís, A. et al. A pathway-based association study reveals variants from Wnt signaling genes contributing to asthma susceptibility. Clin. Exp. Allergy 47, 618–626 (2017).

56. Liu, J. et al. Targeting Wnt-driven cancer through the inhibition of Porcupine by LGK974. Proc. Natl Acad. Sci. USA 110, 20224–20229 (2013).

57. Arredouani, M. et al. The scavenger receptor MARCO is required for lung defense against pneumococcal pneumonia and inhaled particles. J. Exp. Med. 209, 267–272 (2008).

58. Dorrington, M. G. et al. MARCO is required for TRL2- and Nod2-mediated responses to Streptococcus pneumoniae and clearance of pneumococcal colonization in the murine nasopharynx. J. Immunol. 190, 250–258 (2013).

59. Torgerson, D. G. et al. Meta-analysis of genome-wide association studies of asthma in ethnically diverse North American populations. Nat. Genet. 43, 887–892 (2011).

60. Bahlada, R. et al. Phase I dose-escalation studies of roniciclib, a pan-cyclin-dependent kinase inhibitor, in advanced malignancies. Br. J. Cancer 116, 1503–1512 (2017).

61. Yu, Q. B., Li, H. M., Li, L. L., Wang, S. Y. & Wu, Y. B. Sevoflurane downregulates interleukin-6 and interleukin-8 levels in patients after cardiopulmonary bypass surgery: a meta-analysis. Genet. Mol. Res. 14, 19016–19027 (2015).

62. Cho, E. J., Yoon, J. H., Hong, S. J., Lee, S. H. & Sim, S. B. The effects of sevoflurane on systemic and pulmonary inflammatory responses after cardiopulmonary bypass. J. Cardiothorac. Vasc. Anesth. 40, 639–645 (2009).

63. Uhlig, C. et al. Effects of volatile anesthetics on mortality and postoperative pulmonary and other complications in patients undergoing surgery: a systematic review and meta-analysis. Anesthesiology 124, 1230–1245 (2016).

64. Jabadon, M. et al. Sevoflurane for sedation in acute respiratory distress syndrome: A Randomized Controlled Pilot Study. Am. J. Respir. Crit. Care Med. 195, 792–800 (2017).

65. Hirshman, C. A. & Emala, C. W. Actin reorganization in airway smooth muscle cells involves Gq and Gi-2 activation of Rho. Am. J. Physiol. 277, L653–L661 (1999).

66. Billington, C. K. & Penn, R. B. Signaling and regulation of G protein-coupled receptors in airway smooth muscle. Respir. Rev. 4, 2 (2003).

67. Lange, P. et al. Lung function trajectories leading to chronic obstructive pulmonary disease. N. Engl. J. Med. 373, 111–122 (2015).

68. Fletcher, C. & Peto, R. The natural history of chronic airflow obstruction. Br. Med. J. 1, 1645–1648 (1977).

69. Martinez, F. D. Early-life origins of chronic obstructive pulmonary disease. N. Engl. J. Med. 375, 871–878 (2016).

70. Hankinson, L. L., Odencrantz, J. R. & Fedan, K. B.Spirometric reference values for the present study. J.C.C. received research materials from Pharmavite (vitamin D and placebo capsules), GSK (Flevent) and Merck (AstraZeneca) to provide medications free of cost to participants in NIH-funded studies. D.C.N. is an employee of Merck Research Laboratories. D.D.S. has received research funding from Boehringer Ingleheim, Merck, and AstraZeneca to cover the costs of this work. B.M.P. serves on the DSMB of a clinical trial funded by the manufacturer (Zoll LifeCor) and on the Steering Committee of the Yale Open Data Access Project funded by Johnson & Johnson. M.H.C. has received grant support from GlaxoSmithKline. The authors declare no other competing interests.

71. Yang, J. et al. Genomic inflation factors under polygenic inheritance. Eur. J. Hum. Genet. 19, 807–812 (2011).

72. Genomes Project, C. et al. An integrated map of genetic variation from 1092 human genomes. Nature 491, 56–65 (2012).

73. Maurano, M. T. et al. Systematic localization of common disease-associated variation in regulatory DNA. Science 337, 1190–1195 (2012).

74. Hnisz, D. et al. Superenhancers in the control of cell identity and disease. Cell 155, 934–947 (2013).

75. Forrester, A. R. et al. A promoter-level mammalian expression atlas. Nature 507, 462–470 (2014).

76. Farh, K. K. H. et al. Genetic and epigenetic fine mapping of causal autoimmune disease variants. Nature 518, 337–343 (2015).

77. Dunham, I. et al. An integrated encyclopedia of DNA elements in the human genome. Nature 489, 57–74 (2012).

78. Kichaev, G. et al. Integrating functional data to prioritize causal variants in statistical fine-mapping studies. PLoS Genet. 10, e1004722 (2014).

79. Pickrell, J. K. Joint analysis of functional genomic data and genome-wide association studies of 18 human traits. Am. J. Hum. Genet. 94, 559–573 (2014).

80. Gamazon, E. R. et al. A gene-based association method for mapping traits using reference transcriptome data. Nat. Genet. 47, 1091–1098 (2015).
Annah B. Wyss1, Tamar Sofer2,3, Mi Kyeong Lee1, Natalie Terzikhan4,5, Jennifer N. Nguyen6, Lies Lahousse5,7, Jeanne C. Latourellle8, Albert Vernon Smith9,10,11, Traci M. Bartz12,13, Mary F. Feitosa14, Wei Gao15, Tarunveer S. Ahluwalia16,17, Wenbo Tang18, Christopher Oldmadow19, Qing Duan20, Kim de Jong21, Mary K. Wojczynski24, Xin-Qun Wang22, Raymond Noordam23, Fernando Pires Hartwig24,25, Victoria E. Jackson26, Tianyuan Wang27, Ma’en Obeidat28, Brian D. Hobbs29,30, Tianxiao Huan31, Hongsheng Gui32, Margaret M. Parker29, Donglei Hu33, Lauren S. Mogill34, Gleb Kichaev35, Jianping Jin36, Mariaelisa Graff37, Tamara B. Harris38, Ravi Kalhan39, Susan R. Heckbert40, Lavinia Paternoster25, Kristin M. Burkart41, Yongmei Liu42, Elizabeth G. Holliday39, James G. Wilson43, Judith M. Vonk21, Jason L. Sanders44, R. Graham Barr45,46, Renée de Mutsert47, Ana Maria Baptista Menezes34, Hieab H.H. Adams5,48, Maarten van den Berg49, Roby Joehanes50, Albert M. Levin51, Jennifer Liberto52, Lenore J. Launer38, Alanna C. Morrison52, Colleen M. Sitlani12, Juan C. Celedón53, Stephen B. Kretzkes54, Rodney J. Scott19,55, Kaare Christensen56, Jerome I. Rotter57, Tobias N. Bonten58,59, Fernando César Wehrmeister24, Yohan Bossé60, Shujie Xiao32, Sam Oh33, Nora Franceschini37, Jennifer A. Brody12, Robert C. Kaplan61, Kurt Lohman42, Mark McEvoy19, Michael A. Province14, Frits R. Rosendaal47, Kent D. Taylor57, David C. Nickle62, L. Keoki Williams32,63, Esteban G. Burchard33,34, Heathier E. Wheeler13, Don D. Sin28,65, Vilmundur Gudnason9,10, Kari E. North37, Myriam Fornage66, Patricia A. Cassano18,71, Joohon Sung72,73, John R. Attia19, Leslie Lange74, H. Marike Boezen21, Bharat Thyagarajan75, Stephen S. Rich6, Dennis O. Mook-Kanamori47,58, Bernardo Lessa Horta24, André G. Uitterlinden76, Hae Kyung Im77, Michael H. Cho29,30, Guy G. Brusselle4,5,78, Sina A. Gharib79, José Dupuis15,69, Ani Manichaikul6 & Stephanie J. London1

1Epidemiology Branch National Institute of Environmental Health Sciences, National Institutes of Health, US Department of Health and Human Services, Research Triangle Park, NC 27709, USA. 2Division of Sleep and Circadian Disorders, Brigham and Women’s Hospital, Boston, MA 02115, USA. 3Department of Medicine, Harvard Medical School, Boston, MA 02115, USA. 4Department of Respiratory Medicine, Ghent University Hospital, Vlagstraat 15, B-9000 Gent, Belgium. 5Department of Epidemiology, Erasmus University Medical Center, Rotterdam 3000 CA, The Netherlands. 6Center for Public Health Genomics, University of Virginia, Charlottesville, VA 22908, USA. 7Department of Bioanalysis, FFW, Ghent University, Ghent 9000, Belgium. 8Department of Neurology, Boston University School of Medicine, Boston, MA 02118, USA. 9Icelandic Heart Association, Kópavogur 201, Iceland. 10Faculty of Medicine, University of Iceland, Reykjavik 101, Iceland. 11Department of Biostatistics, University of Michigan, Ann Arbor, MI 48109, USA. 12Cardiovascular Health Research Unit, Department of Medicine, University of Washington, Seattle, WA 98101, USA. 13Division of Biostatistics, University of Washington, Seattle, WA 98195, USA. 14Division of Statistical Genomics, Department of Genetics, Washington University School of Medicine, St Louis, MO 63110, USA. 15Department of Biostatistics, Boston University School of Public Health, Boston, MA 02118, USA. 16The Novo Nordisk Foundation Center for Basic Metabolic Research, Metabolic Genetics Section, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen 2200, Denmark. 17Steno Diabetes Center Copenhagen, Gentofte 2820, Denmark. 18Department of Nutritional Sciences, Cornell University, Ithaca, NY 14853, USA. 19Hunter Medical Research Institute and Faculty of Health, University of Newcastle, Callaghan, NSW 2305, Australia. 20Department of Genetics, University of North Carolina, Chapel Hill, NC 27599, USA. 21Department of Epidemiology, University of Groningen, University Medical Center Groningen, 9713 EZ Groningen, Netherlands. 22Division of Biostatistics, Department of Public Health Sciences, University of Virginia, Charlottesville, VA 22908, USA. 23Department of Internal Medicine, Section Gerontology and Geriatrics, Leiden University Medical Center, Leiden 2300 RC, The Netherlands. 24Postgraduate Program in Epidemiology, Federal University of Pelotas, 96020-220 Pelotas, Brazil. 25Medical Research Council Integrative Epidemiology Unit, School of Social and Community Medicine, University of Bristol, Bristol BS8 2BN, UK. 26Department of Health Sciences, University of Leicester, Leicester LE1 7RH, UK. 27Integrative Bioinformatics Support Group National Institute of Environmental Health Sciences, National Institutes of Health, US Department of Health and Human Services, Research Triangle Park, NC 27709, USA. 28The University of British Columbia Center for Heart Lung Innovation, St Paul’s Hospital, Vancouver, BC V6Z 1Y6, Canada. 29Channing Division of Network Medicine, Brigham and Women’s Hospital, Boston, MA 02115, USA. 30Division of Pulmonary and Critical Care Medicine, Brigham and Women’s Hospital, Boston, MA 02115, USA. 31The Population Sciences Branch, Division of Intramural Research, National Heart, Lung, and Blood Institute, Bethesda, MD 20892, USA. 32Center for Health Policy and Health Services Research, Henry Ford Health System, Detroit, MI 48202, USA. 33School of Medicine, University of California San Francisco, San Francisco, CA 94143, USA. 34Department of Biology, Loyola University Chicago, Chicago, IL 60660, USA. 35University of California Los Angeles, Los Angeles, CA 90095, USA. 36Westat, Durham, NC 27703, USA. 37Department of Epidemiology, University of North Carolina, Chapel Hill, NC 27599, USA. 38Department of Health and Human Services, Laboratory of Epidemiology and Population Sciences, National Institute on Aging, National Institutes of Health, Bethesda, MD 20892, USA. 39Division of Pulmonary and Critical Care Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL 60611, USA. 40Department of Epidemiology, Cardiovascular Health Research Unit, University of Washington, Seattle, WA 98101, USA. 41Division of Pulmonary, Allergy and Critical Care Medicine, Department of Medicine, College of Physicians and Surgeons, Columbia University, New York, NY 10032, USA. 42Wake Forest School of Medicine, Winston-Salem, NC 27101, USA. 43Department of Physiology and Biophysics, University of Mississippi Medical Center, Jackson, MS 39216, USA. 44Department of Medicine, Massachusetts General Hospital, Boston, MA 02114, USA.
