
Burkart, Kristin M. and Sofer, Tamar and London, Stephanie J. and Manichaikul, Ani and Hartwig, Fernando P. and Yan, Qi and Artigas, María Soler and Avila, Lydiana and Chen, Wei and Thomas, Sonia Davis and Diaz, Alejandro A. and Hall, Ian P. and Horta, Bernardo L. and Kaplan, Robert C. and Laurie, Cathy C. and Menezes, Ana M. and Morrison, Jean V. and Oelsner, Elizabeth C. and Rastogi, Deepa and Rich, Stephen S. and Soto-Quiros, Manuel and Stilp, Adrienne M. and Tobin, Martin D. and Wain, Louise V. and Celedon, Juan C. and Barr, R. Graham (2018) A genome-wide association study in Hispanics/Latinos identifies novel signals for lung function: the Hispanic Community Health Study/Study of Latinos. American Journal of Respiratory and Critical Care Medicine . ISSN 1535-4970

**A Genome-Wide Association Study in Hispanics/Latinos Identifies Novel Signals for Lung Function.** The Hispanic Community Health Study/Study of Latinos

Kristin M Burkart, MD, MSc[1]*, Tamar Sofer, PhD[2]*, Stephanie J. London, MD, DrPH[3], Ani Manichaikul, PhD[4], Fernando P Hartwig, MSc[5], Qi Yan, PhD[6], María Soler Artigas, PhD[7], Lydiana Avila, MD[8], Wei Chen, PhD[6], Sonia Davis Thomas, DrPH[9], Alejandro A. Diaz, MD, MPH[10], Ian P Hall, Prof.[11], Bernardo L Horta, MD, PhD[5], Robert C Kaplan, PhD[12], Cathy C. Laurie, PhD[13], Ana M Menezes, MD, PhD[5], Jean V. Morrison[14], Elizabeth C Oelsner, MD, MPH[1], Deepa Rastogi, MBBS, MS[15], Stephen S Rich, PhD[4], Manuel Soto-Quiros, PhD[6],Adrienne M. Stilp, PhD[13], Martin D Tobin, Prof., PhD[7,16], Louise V Wain, PhD[7], Juan C. Celedón, MD, DrPH[6]*, R. Graham Barr, MD, DrPH[1,17]*

*These authors contributed equally to this work

**Author's affiliations:**

[1]Department of Medicine, College of Physicians and Surgeons, Columbia University, New York, NY

[2] Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Division of Sleep and Circadian Disorders, Boston, MA

Department of Biostatistics, University of Washington School of Public Health, Seattle, WA

[3] Epidemiology Branch, National Institute of Environmental Health Sciences, National Institutes of Health, Department of Health and Human Services, Research Triangle Park, NC

[4]Center for Public Health Genomics, University of Virginia School of Medicine, Charlottesville, VA

[5] Postgraduate Program in Epidemiology, Federal University of Pelotas, Pelotas, Brazil

[6]Division of Pulmonary Medicine, Allergy and Immunology, Department of Pediatrics, Children's Hospital of Pittsburgh of UPMC, University of Pittsburgh, Pittsburgh, PA

[7] Genetic Epidemiology Group, Department of Health Sciences, University of Leicester, Leicester, UK

[8]Division of Pediatric Pulmonology, Hospital Nacional de Niños, San José, Costa Rica

[9]Department of Biostatistics, University of North Carolina, Chapel Hill, NC

[10]Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA

[11]Division of Respiratory Medicine, Queen's Medical Centre, University of Nottingham, Nottingham NG7 2RD, UK

[12]Department of Epidemiology & Population Health, Albert Einstein College of Medicine, New York, NY

[13] Department of Biostatistics, University of Washington School of Public Health, Seattle, WA

[14] Department of Human Genetics and Statistics, University of Chicago, Chicago, IL

[15]Department of Pediatrics, Albert Einstein College of Medicine, New York, NY

[16]National Institute for Health Research (NIHR) Leicester Respiratory Biomedical Research Unit, Glenfield Hospital, Leicester, UK

[17]Department of Epidemiology, Mailman School of Public Health, Columbia University, New York, NY

**Correspondence:**
Kristin M Burkart, MD, MSc
Columbia University Medical Center
630 West 168th Street, PH 8 East - Room 101
New York, NY 10032
Fax: 212-305-8464
E-mail: kb2319@cumc.columbia.edu

**Word Count: 3,464**
**Abstract Word Count: 250**
**Short title:** Genome-wide study of lung function in Hispanics/Latinos

**Funding: HCHS/SOL**: We thank the participants and staff of the Hispanic Community Health Study/Study of Latinos (HCHS/SOL) for their contributions to this study. The baseline examination of HCHS/SOL was carried out as a collaborative study supported by contracts from the NHLBI to the University of North Carolina (N01-HC65233), University of Miami (N01-HC65234), Albert Einstein College of Medicine (N01-HC65235), Northwestern University (N01-HC65236), and San Diego State University (N01-HC65237). The following institutes, centers, and offices contributed to the first phase of HCHS/SOL through a transfer of funds to the NHLBI: National Institute on Minority Health and Health Disparities, National Institute on Deafness and Other Communication Disorders, National Institute of Dental and Craniofacial Research (NIDCR), National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), National Institute of Neurological Disorders and Stroke, and NIH Office of Dietary Supplements. The Genetic Analysis Center at the University of Washington was supported by NHLBI and NIDCR contracts (HHSN268201300005C AM03 and MOD03). Additional analysis support was provided by 1R01DK101855-01and13GRNT16490017. Genotyping efforts were supported by the NIH Department of Health and Human Services (HSN26220/20054C), National Center for Advancing Translational Science

Clinical Translational Science Institute (UL1TR000124), and NIDDK Diabetes Research Center (DK063491). This manuscript has been reviewed by the HCHS/SOL Publications Committee for scientific content and consistency of data interpretation with previous HCHS/SOL publications. TS is supported by NHLBI 1R35HL135818. AAD is supported by K01HL118714 and the Brigham and Women's Hospital Minority Faculty Career Development Award. SJL is supported by the Intramural Research Program of the NIH, National Institute of Environmental Health Sciences. RGB is supported by R01-HL077612 and R01-HL092081.

**COPD-Costa Rica, MESA Lung, 1982 Pelotas Birth Cohort Study, and SpiroMeta**: See online supplement

**AUTHOR CONTRIBUTION**

**Study design:** KMB, JCC, RGB

**Data collection:** FPH, IPH, BLH, RCK, LA, DR, SSR, MS-Q, AMS, AMM, MDT, RCK, SJL, JCC, RGB

**Data analysis:** TS, AM, WC, FPH, MSA, LVW

**Obtaining funding:** RCK, CCL, SSR, RGB

**Drafting Manuscript:** KMB

**Critical revision of manuscript:** KMB, TS, AM, WC, FPH, MSA, LA, AD, IPH, BLH, RCK, CCL, SJL, AMM, JVM, ECO, DR, SSR, MS-Q, AMS, MDT, LVW, QY, JCC, RGB

## AT A GLANCE COMMENTARY

### Scientific Knowledge on the Subject

Lung function and chronic obstructive pulmonary disease (COPD) are heritable traits. Prior genome-wide association studies (GWAS) of lung function have identified numerous genetic variants associated with COPD risk; however, much of the individual variance in lung function remains unexplained. Furthermore, most of these studies have been limited to participants of European ancestry. Including multi-ethnic populations in GWAS research may identify novel variants that contribute to the etiology of lung function.

### What this Study Adds to the Field

This study, the largest GWAS of lung function and COPD to exclusively include Hispanic/Latino participants, identified eight novel signals of which three replicated in independent populations. A novel locus for $FEV_1$ (rs4791658; *ZSWIM7*) replicated in a cohort of European ancestry. A locus for $FEV_1$/FVC (rs145174011; *HAL*); in a previously identified region for $FEV_1$/FVC in individuals of European ancestry and for percent emphysema in a Hispanic population remained independent in conditional analyses but did not replicate. Admixture mapping identified a novel region associated with Amerindian ancestry and $FEV_1$ which included a candidate variant (rs4133185) in the *AGMO* gene that replicated. Additionally, we identified a SNP (rs7709630) for COPD, which replicated in individuals of European ancestry. Several loci previously identified in European samples were also associated with lung function traits among Hispanic/Latino participants in HCHS/SOL at the genome-wide significance level. These results emphasize the importance of including admixed populations when performing genetic

studies to identify variants that may contribute to the genetic etiology of pulmonary function and COPD.

## ABSTRACT

***Rationale***: Lung function and chronic obstructive pulmonary disease (COPD) are heritable traits. Genome-wide association studies (GWAS) have identified numerous pulmonary function and COPD loci, primarily in cohorts of European ancestry.

***Objectives***: Perform a GWAS of COPD-phenotypes in Hispanic/Latino populations to identify loci not previously detected in European populations.

***Methods***: GWAS of lung function and COPD in Hispanic/Latino participants from a population-based cohort. We performed replication studies of novel loci in independent studies.

***Measurements and Main Results***: Among 11,822 Hispanic/Latino participants, we identified eight novel signals; three replicated in independent populations of European Ancestry. A novel locus for forced expiratory volume in one second ($FEV_1$) in *ZSWIM7* (rs4791658; $p=4.99\times10^{-9}$) replicated. A rare variant (MAF=0.002) in *HAL* (rs145174011) was associated with $FEV_1$ to forced vital capacity ($FEV_1$/FVC) ($p=9.59\times10^{-9}$) in a region previously identified for COPD-related phenotypes; it remained significant in conditional analyses but did not replicate. Admixture mapping identified a novel region, with a variant in *AGMO* (rs41331850), associated with Amerindian ancestry and $FEV_1$, which replicated. A novel locus for $FEV_1$ identified among ever smokers (rs291231; $p=1.92\times10^{-8}$) approached statistical significance for replication in admixed populations of African ancestry and a novel SNP for COPD in *PDZD2* (rs7709630; $p=1.56\times10^{-8}$) regionally replicated. Additionally, loci previously identified for lung function in European samples were associated in Hispanic/Latino participants in HCHS/SOL at the genome-wide significance level.

***Conclusions***: We identified novel signals for lung function and COPD in a Hispanic/Latino cohort. Including admixed populations when performing genetic studies may identify variants contributing to genetic etiologies of COPD.

**Word Count: 250**

**Key Words:** Hispanic/Latino, genome-wide association study, single nucleotide polymorphisms, lung function, airflow limitation, chronic obstructive pulmonary disease

## INTRODUCTION

The Hispanic/Latino population is the largest and fastest-growing minority population in the United States (U.S.), accounting for 17% of the U.S. population currently and an estimated 33% by 2060(1). The Hispanic/Latino population in the U.S. is genetically diverse, with a mixture of European, African and Amerindian genetic ancestries(2-4).

Chronic obstructive pulmonary disease (COPD), characterized by persistent airflow obstruction(5, 6), is the third-leading cause of death in the U.S.(7). The prevalence of COPD in some Hispanic groups is lower than in non-Hispanic whites and African-Americans in the U.S.(8-10). In the Hispanic Community Health Study/Study of Latinos (HCHS/SOL), the largest and best characterized cohort of Hispanics/Latinos in the US, the prevalence of COPD varied substantially by Hispanic/Latino heritage with Puerto Ricans and Cubans having a much higher prevalence of COPD than other Hispanic/Latino groups(10). This difference in HCHS/SOL was explained by smoking and asthma history; however, other investigators have noted that Hispanic ethnicity is inversely associated with COPD-phenotypes compared to whites(11, 12) and have hypothesized that genetic differences may lower risk of COPD in some Hispanic groups(11, 13).

Examination of genetic risk among non-European populations may reveal novel variants that yield new pathways to treatments, such as *PCSK9*(14). We are aware of only two genome-wide association studies (GWAS) of lung function or COPD-related phenotypes that analyzed Hispanics separately(13, 15). One identified a novel variant near *MAN2B1* that was associated with percent emphysema on computed tomography (CT)(16) and the other identified two loci that approached genome wide statistical

significance for COPD(13). This literature contrasts with that among persons of European ancestry, in whom GWAS have identified multiple loci in genes for lung function and COPD(17-31)

To improve our current understanding of the genetic architecture of lung function-related traits in Hispanics/Latinos, we performed a meta-analysis for $FEV_1$, $FEV_1$/FVC, airflow limitation and COPD among six Hispanic/Latino groups in the HCHS/SOL cohort(10, 32). Findings were replicated in European, Hispanic and African populations. Some of the results have been previously reported in abstract form(33).

## METHODS

### Study Sample

HCHS/SOL is a community-based cohort study of 16,415 self-identified Hispanic/Latino persons aged 18-74 years recruited from four U.S. communities. The study design, cohort recruitment(10, 32), and baseline clinical examination(34) have been previously described. Institutional Review Boards at each field center approved study protocols and written informed consent was obtained from all participants.

For genetic analyses, HCHS/SOL participants were classified in six genetic analysis groups: Central American, Cuban, Dominican, Mexican, Puerto Rican, and South American. The genetic ancestry structure of these groups, including principal components (PCs) plots for all individuals, have been previously reported(35). Henceforward we will refer to the six genetic analysis groups as Hispanic/Latino ancestry groups.

Exclusion criteria were lack of valid spirometric or genetic data, missing covariates, near exclusive Asian ancestry and participants who were not classified in the Hispanic/Latino ancestry groups. Details are provided in the online supplement.

**Phenotypic and Outcome Measures**

Spirometry was conducted in accordance with American Thoracic Society/European Respiratory Society guidelines(36) using a dry rolling sealed spirometer as previously described(10). Participants with pre-bronchodilator $FEV_1$/FVC ratios less than 70% or the lower limit of normal were selected for post-bronchodilator spirometry, the former defined by NHANES reference equations(37) since those from HCHS/SOL(38) were not available during data collection.

Airflow limitation was defined as a pre-bronchodilator $FEV_1$/FVC ratio < 70% and COPD as a post-bronchodilator $FEV_1$/FVC ratio < 70%(5). Controls had normal lung function defined as a pre-bronchodilator $FEV_1$/FVC ratio > 70% and FVC > 80 percent predicted.

**Genotyping, Quality Control and Imputation**

Consenting HCHS/SOL subjects were genotyped at Illumina on the HCHS/SOL custom 15041502 B3 array. We applied standard quality-assurance and quality-control methods(39) as previously described(35). Genome-wide imputation was carried out with the full, cosmopolitan 1000 Genomes Project phase 1 reference panel(40); associations for novel loci were confirmed with 1000 Genome phase 3 imputation panel. Details are

provided in the online supplement.

## Statistical Analyses

See the online supplement for details.

*Quantitative Lung Function Analyses*

Analyses of $FEV_1$ and $FEV_1$/FVC ratio employed linear mixed models, stratified by Hispanic/Latino ancestry group, and adjusted for age, age$^2$, sex, height, height$^2$, study center, smoking status, pack-years, sampling weights, and the first 5 PCs, as fixed effects. We used random effects for genetic relatedness (kinship), and household and community (block unit) to account for environmental correlation. Results from each Hispanic/Latino ancestry group were then meta-analyzed using the MetaCor method(41). Genome-wide significance threshold was defined as $p<5x10^{-8}$, the Bonferroni adjustment for 1 million independent tests (42). *A priori,* we planned additional meta-analyses stratified by never vs ever smoking.

*Analyses for Airflow Limitation and COPD*

We analyzed airflow limitation and COPD in pooled analyses implementing the GMMAT software(43) to fit a logistic mixed model adjusting for age, sex, study center, smoking status, pack-years, sampling weights, the first 5 PCs and Hispanic/Latino ancestry groups, as fixed effects and with random effects for kinship and block unit. Participants younger than 45 years were excluded.

*Admixture Mapping and Analyses*

Local ancestry estimates were previously inferred in the HCHS/SOL(44). A genome-wide admixture mapping scan was performed using a linear mixed model, as described above, testing European, African, and Amerindian ancestries at each available local ancestry interval (LAI). Statistical significance for admixture mapping was set at $p<5.68x10^{-5}$ based on previous simulation results for HCHS/SOL(45). After discovering a genome-wide significant LAI association (at the admixture mapping level), we identified a candidate variant and performed conditional admixture mapping analysis(46).

*Replication of Lung Function SNPs*

We pursued replication of novel loci associated with $FEV_1$ and $FEV_1$/FVC in populations with Hispanics(13, 47-51), African-American(47, 48), and European representation(17). We performed a look-up in each individual study then meta-analyzed across studies. For airflow limitation and COPD, we performed regional replication in publically available GWAS results of airflow obstruction in participants of European ancestry(23). For admixture mapping signals, we performed a look-up in the UK Biobank results for individuals of European ancestry (https://sites.google.com/broadinstitute.org/ukbbgwasresults).

*Generalization analyses of previously reported SNPs*

We looked-up previously reported SNPs, their effects, standard errors and p-value in HCHS/SOL(52).

## RESULTS

The mean age of the 11,822 HCHS/SOL participants with valid lung function and genetic data was 46±14 years. Thirty-nine percent had ever smoked cigarettes with median pack-years of 7.5. The COPD analysis included 363 and 5,253 individuals with and without COPD, respectively. Characteristics of the participants are shown in **table 1**. Manhattan and Quantile-Quantile (Q-Q) plots for $FEV_1$, $FEV_1$/FVC, airflow limitation and COPD are shown in **figures E1, E2, E3 and E4**. Genomic inflation factors ranged from 1.020–1.026 in quantitative lung function analyses and 0.988–1.011 in analyses of COPD.

*GWAS of Lung Function across all Hispanic/Latino ancestry groups*

Across all Hispanic/Latino ancestry groups, seven signals achieved genome-wide statistical significance, of which two were novel for lung function (**table 2**). The lead novel SNP, rs4791658, in the locus associated with $FEV_1$ (**figure 1a**) was an intron variant in the gene *ZSWIM7* on chromosome 17 (Beta=33.4, $p=4.99\times10^{-9}$). Its minor allele frequency (MAF) ranges from 0.39–0.49 across Hispanic/Latino ancestry groups (**figure 1b**). SNP, rs4791658 neared genome-wide significance for $FEV_1$ among ever smokers (Beta=46.8, $p=1.53\times10^{-6}$). Effect size and p-values for rs4791658 across all lung function and COPD-related traits are shown in **Table E1**.

The lead novel SNP, rs145174011, in the locus associated with $FEV_1$/FVC (**figure 1c**) was an intronic variant in the gene *HAL* on chromosome 12 (Beta=4.86, $p=9.59\times10^{-9}$). The SNP is rare with a MAF that ranges from 0.001–0.008 across Hispanic/Latino ancestry groups (**figure 1d**). The minor allele does not exhibit an outlier

effect **(figure E5).** Sensitivity analysis using rank-normalized residuals is shown in supplement. This locus is near the gene *CCDC38*, which was previously identified for $FEV_1$/FVC in individuals of European Ancestry (17) and the gene *SNRPF*, which was previously identified for percent emphysema on CT in racially/ethnically diverse participants (16). In conditional analysis adjusting for rs1036429 (*CCDC38*) and rs7957346 (*SNRPF*) in HCHS/SOL, rs145174011 (*HAL* SNP) remained associated with $FEV_1$/FVC (Beta=4.92 $p=6.4\times10^{-9}$), suggesting its association with $FEV_1$/FVC is independent of previously identified variants. The findings for rs145174011 among ever and never smokers are concordant and trended towards significance (**table E1)**. Regional association plots and forest plots of previously reported genome wide significant associations for $FEV_1$ and $FEV_1$/FVC meta-analyzed across all Hispanic/Latino ancestry groups are shown in **figure E6**.

*Stratification by Smoking Status*

In the meta-analysis across all Hispanic/Latino ancestry groups stratified by smoking status, we identified one novel locus among ever smokers for $FEV_1$ and one novel locus among never smokers for $FEV_1$/FVC (**table 3**). The lead variant in the locus associated with $FEV_1$ in ever smokers was an Indel variant ($p=1.66\times10^{-8}$) on chromosome 11. The top SNP, rs291231 (Beta=57.28, $p=1.92\times10^{-8}$), in this locus has a MAF that ranges from 0.28 – 0.42 across Hispanic/Latino ancestry groups. This SNP lies between the genes *EED* and *CCDC81* (**figure 2a and 2b**). The SNP-by-pack-years interaction trended towards significance (p-interaction=0.06) and the genotype effect became greater at higher pack-years (**figure E7**).

The top SNP (rs7228593) in the locus associated with $FEV_1$/FVC in never smokers (Beta=1.34, $p=3.47\times10^{-8}$) is located between the genes *SMIM21* and *LCOC339298* (**figure 2c**) on chromosome 18. The findings in never smokers for airflow limitation (OR=0.46, $p=2.63\times10^{-4}$) and COPD (OR=0.36, p=0.005) were directionally concordant and trended towards significance (**table E1**). Its MAF varies significantly across different Hispanic/Latino ancestry groups (**figure 2d**) with the highest in participants from the Dominican ancestry group (MAF=0.16) and lowest in Mexican ancestry group (MAF=0.02). Forest plots and regional association plot of remaining genome-wide significant associations for $FEV_1$ and $FEV_1$/FVC stratified by smoking status are shown in **figure E8**.

*GWAS of COPD*

We identified two previously unreported SNPs for COPD (**table E2 and figure E9**). SNP rs7709630 is an intron in the *PDZD2* gene on chromosome 5 ($p=1.56\times10^{-8}$; MAF=0.11) and SNP rs2286351 is a noncoding transcript variant in *CDRT15P1* on chromosome 17 ($p=1.97\times10^{-8}$, MAF=0.09). Lung function association analyses of these COPD SNPs are reported in **table E3**.

*Admixture Mapping*:

Manhattan plots from admixture mapping analyses for $FEV_1$, and $FEV_1$/FVC testing Amerindian, African, and European ancestry counts individually in LAIs are shown in **figures E10 and E11**. We identified one novel significant local ancestry association

region of Amerindian ancestry for $FEV_1$ **(figure E10a)** on chromosome 7 (Beta =-37.51, $p=2.86x10^{-6}$). There were no significant associations in the European versus others or African versus others analyses.

We identified a candidate variant (rs4133185) within this novel LAI region (Beta = -31.44, $p= 7.43x10^{-7}$; **table E4 and figure 3**) with considerable differences in ancestry-specific allele frequencies (**figure 4**). The lead SNP, rs4133185, in this candidate variant was an intron variant in the gene *AGMO*. The admixture mapping signal is less significant after conditional analyses adjusting for rs4133185 (**table E5, figures 4 and E**12).

*Replication*

Characteristics of the participants in the replication cohorts are shown in **table E4.** Of the two novel loci identified for lung function, only the $FEV_1$ locus on chromosome 17 successfully replicated (replication $p=3.37x10^{-5}$; adjusted threshold for statistical significance was $p<9.2x10^{-4}$) but only in the SpiroMeta consortium look-up (n=94,612)(17). The candidate variant (rs4133185), identified through admixture mapping, associated with Amerindian ancestry and $FEV_1$ replicated in UK Biobank (Beta = -8.3, p=0.004). SNP rs291239, a genotyped SNP in the novel locus on chromosome 11 identified for $FEV_1$ among ever smokers, approached statistical significance for replication in the meta-analysis across all cohorts with admixed populations of African ancestry (replication p=0.007); however, the direction of effect was discordant in the individual replication cohorts (**table E7**).

The novel SNP (rs7709630; *PDZD2*) associated with COPD regionally replicated in publically available results from the SpiroMeta-CHARGE GWAS meta-analysis of airflow obstruction among participants of European ancestry(23). The smallest p-value in the replication region was for rs409839 ($p=8.9x10^{-4}$) and the adjusted threshold for statistical significance was $p< 4.2x10^{-3}$. The other loci were not successfully replicated.

*Generalization Analyses of Previously Identified Lung Function Loci*

Using the FDR adjusted look-up threshold, several of the previously reported SNPs for $FEV_1$, $FEV_1$/FVC and COPD identified in European populations were generalized to Hispanic/Latino populations (**tables E8, E9, E10 and Figure E13**). See online supplement for details.

## DISCUSSION

In this GWAS of 11,822 Hispanic/Latino HCHS/SOL participants, we identified eight novel signals, of which three replicated in independent cohorts of European ancestry. Three novel signals were associated with lung function. A novel locus in *ZSWIM7* (rs4791658) was associated with $FEV_1$ replicated and a rare variant in *HAL* (rs145174011) was associated with $FEV_1$/FVC did not replicate; it is in a previously identified region but remained significant in conditional analyses, suggesting an independent effect. Admixture mapping identified a novel LAI region, with a candidate variant in the *AGMO* gene (rs4133185), associated with Amerindian ancestry and $FEV_1$, which replicated. In smoking stratified analyses, a novel locus (rs291231) associated with $FEV_1$ among 4,595 ever smokers approached statistical significance for replication

in cohorts with admixed populations of African ancestry and a novel SNP for COPD in *PDZD2* (rs7709630) regionally replicated in individuals of European ancestry. In addition, we confirmed loci previously identified in European samples [*KCNE2*(17) and *GPR126*(27) ] for $FEV_1$/FVC in this Hispanic/Latino sample.

Several SNPs in *ZSWIM7* were associated with $FEV_1$ at genome-wide significance and replicated in independent cohorts of European ancestry. *ZSWIM7* mutations were associated colorectal adenomatous polyposis(53) and in a GWAS of SNP-by-smoking interaction, *ZSWIM7* neared significance for systolic blood pressure(54). *ZSWIM7* is a highly conserved eukaryotic regulator of homologous recombination and plays an important role in error-free DNA repair processes for DNA double-strand breaks (DSBs)(55-57). Unrepaired DSBs trigger cell senescence, apoptosis, and pro-inflammatory responses, all of which are established mechanisms in the pathogenesis of COPD(58-60). Oxidative stress and cigarette smoke cause DSBs(61, 62) and DSBs have been implicated in the pathogenesis of COPD(62). Furthermore, COPD patients have a greater number of unrepaired DSBs which are associated with a higher expression of markers for senescence, apoptosis, and pro-inflammatory phenotypes compared to asymptomatic smokers and nonsmokers(62). In our study, the effect size of *ZSWIM7* for $FEV_1$ was of greater magnitude among ever smokers which is consistent with known effects of cigarette smoke causing DSBs. Thus, *ZSWIM7* through its regulation of homologous recombination and role in error-free DNA repair processes for DSBs is relevant to the pathogenesis of COPD.

A rare variant in *HAL* (rs145174011) was associated with $FEV_1$/FVC among all participants. The minor allele was associated with decreases in $FEV_1$/FVC of 4.9% and

6.3% (all participants and ever smokers, respectively), and with decreases in $FEV_1$ of 100 mL and 173 mL (all participants and ever smokers, respectively). It is possible that this variant was not successfully replicated due to lack of power, given that it is a rare variant in African and Hispanic populations and is not polymorphic in European populations(63).

The *HAL* SNP (rs145174011) is near two previously identified loci for pulmonary traits (*CCDC38* and *SNRPF)*. *CCDC38* has been associated with $FEV_1$/FVC in European ancestry individuals(17), and *SNRPF* has been associated with percent emphysema on CT in racially/ethnically diverse participants in the Multi-Ethnic Study of Atherosclerosis (MESA) Lung Study (16). In the MESA Lung Study, linkage disequilibrium (LD) between *SNRPF* and *CCDC38* accounted for much of the effect seen in Whites, and less so in African Americans and Chinese; however, the observed effect of *SNRF* on percent emphysema on CT in Hispanics appeared to be independent from *CCDC38*(16). In our study, the association of SNP rs145174011 (*HAL* locus) with $FEV_1$/FVC was maintained in analyses conditioning on previously reported loci. Collectively, these data suggest there are three independent loci associated with COPD-related phenotypes in this region.

Although the *HAL* locus contained numerous SNPs, they were all imputed rare variants and did not replicate; therefore, further investigation, such as whole-genome sequencing of this region and functional studies, is required to refine these loci, confirm association, and understand their biological implications for COPD.

The *HAL* gene encodes the enzyme histidine ammonia-lyase(64, 65) - the first step in histidine catabolism(66). Low levels of histidine have been associated with

increased inflammation and oxidative stress(67)(68) and decreased concentrations of histidine were associated with advanced COPD and emphysema(69). Recently, whole-exome sequencing in African-Americans identified three rare loss-of-function variants in *HAL* that were associated with increased histidine levels and were inversely related to coronary heart disease risk in African-American and European-American populations. Thus, a functional *HAL* variant may decrease histidine levels and increase susceptibility to COPD and emphysema through its effects on oxidative stress.

The region associated with Amerindian ancestry and $FEV_1$ on chromosome 7 contained a candidate variant (rs4133185) within the *AGMO* gene. This SNP, rs4133185, has large differences in ancestry-specific allele frequencies (~0.8 for Amerindian and ≤0.2 for European and African) and is associated with obstructive physiology, as measured by $FEV_1$ ($p=7.43x10^{-7}$). In the conditional analysis, rs4133185 explains part but not all of the admixture signal. SNP rs4133185 may not be the causal variant and the presence of multiple SNPs within the LAI region may explain the residual signal. A prior report showed an association of AGMO with FVC in a multi-ethnic population(70). Future whole-genome sequencing studies are needed for better fine mapping of this region.

The chromosome 11 locus (rs291231) associated with $FEV_1$ among ever smokers nominally replicated in the meta-analysis across all cohorts with admixed populations of African ancestry. However, the direction of effect was discordant in some replication cohorts. Different directions of associations may be due different patterns of LD with the causal variant between populations. The minor allele in rs291231 might be

protective in ever smokers as demonstrated by a per-allele increase in $FEV_1$ of 57.28 mL.

The chromosome 18 locus associated with $FEV_1$/FVC among never smokers contained multiple SNPs. The lead SNP (rs7228593) varies in MAF across different Hispanic/Latino ancestry groups in HCHS/SOL (0.02-0.16), is common in African populations (MAF=0.39) and is not polymorphic in European populations (MAF=0)(71) These variable allele frequencies and small sample size for replication may have contributed to lack of replication. Differing environmental exposures and LD may explain the observed genetic association of $FEV_1$/FVC among never smokers in HCHS/SOL and lack of replication in independent cohorts of different ancestry backgrounds.

In the COPD analysis, SNP rs7709630, a genotyped intron variant in *PDZD2*, was a single SNP association with COPD. This genomic region replicated for airflow obstruction in an independent population of European ancestry(23) suggesting a true association. Whereas the *CDRT15P1* locus only contained imputed SNPs and did not replicate, thus reducing the probability that this is a true association with COPD.

We hypothesize that novel loci did not replicate due to small sample size of the Hispanic/Latino replication cohorts along with differing environmental exposures, differing LD, and the complex racial admixture(2) that varies by different Hispanic/Latino ancestry groups. Furthermore, the novel locus for $FEV_1$/FVC in *HAL* and the chromosome 18 locus for $FEV_1$/FVC among never smokers have marked differences in MAF between race and ethnic populations suggesting genetic diversity between races in these loci. The highest MAF for the lead SNPs in the *HAL* and in the chromosome 18 locus were observed in African populations (MAF = 0.005 and 0.38, respectively)

compared with virtually no variability in European populations for these loci (MAF=0)(63). The lack of variance of these loci in European populations limits the ability to replicate these loci as the vast majority of GWAS of lung function have been performed in studies of European ancestry. Evidence for replication of the novel locus on chromosome 11 for $FEV_1$ in ever smokers was stronger in the meta-analysis of cohorts with admixed populations that include African ancestry compared with the meta-analysis of all cohorts (p=0.007 and p=0.04, respectively) supporting the need for additional studies on admixed populations to identify possible ethnicity- or race-specific variants that may elucidate novel pathways in the pathogenesis of COPD.

The study has several potential limitations. This population-based cohort of diverse Hispanic/Latino participants is subject to population stratification. Multiple loci demonstrated varying MAF across Hispanic/Latino groups highlighting the complex racial admixture of Hispanic/Latino populations and raises the possibility of residual population stratification. However, we adjusted our analysis for PCs and our inflation factors ranged from 0.988 to 1.026 which indicates good control of population stratification with small inflation. The COPD analysis consisted of 363 and 5,253 participants with and without COPD (respectively), which is a relatively small sample size for a GWAS. We optimized COPD phenotype by limiting the analysis to age greater than or equal to 45 and used post-bronchodilator spirometry measurements to define COPD.

In conclusion, we identified novel biologically plausible signals associated with clinically important pulmonary measures with evidence for replication in *ZSWIM7, AGMO,* and *PDZD2*; confirmed previous reports of association with $FEV_1$/FVC in

*KCNE2* and *GPR126,* and established that loci previously identified in European populations are generalizable in Hispanic/Latino populations. There is a paucity of genetic studies for lung function that include under-represented minority populations, such as Hispanic/Latino(2, 72) and African populations(14, 26, 73-75). Our findings emphasize the importance of including admixed and multi-racial populations when performing genetic studies of complex diseases and have the potential to advance our understanding of genetic risks for lung disease affecting Hispanic/Latino populations.

**ACKNOWLEDGEMENTS**

The HCHS/SOL investigators thank the National Heart, Lung, and Blood Institute and the research institutions, study investigators, and field staff for their support in creating this study. We also thank the study participants, without whom this endeavor would not have been possible. The authors thank the other investigators, staff, and participants of HCHS/SOL for their valuable contributions.

## References:


1. U.S. Census bureau. 2012 national population projections. 2012 [cited 2016 October 10]. Available from: http://www.census.gov/population/projections/data/national/2012.html.
2. Brehm JM, Celedón JC. Chronic obstructive pulmonary disease in hispanics. *American Journal of Respiratory and Critical Care Medicine* 2008;177:473-478.
3. Hunninghake GM, Weiss ST, Celedón JC. Asthma in hispanics. *American Journal of Respiratory and Critical Care Medicine* 2006;173:143-163.
4. Bryc K, Velez C, Karafet T, Moreno-Estrada A, Reynolds A, Auton A, Hammer M, Bustamante CD, Ostrer H. Genome-wide patterns of population structure and admixture among hispanic/latino populations. *Proceedings of the National Academy of Sciences of the United States of America* 2010;107:8954-8961.
5. Vestbo J, Hurd SS, Agustí AG, Jones PW, Vogelmeier C, Anzueto A, Barnes PJ, Fabbri LM, Martinez FJ, Nishimura M, Stockley RA, Sin DD, Rodriguez-Roisin R. Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease. *American Journal of Respiratory and Critical Care Medicine* 2013;187:347-365.
6. Vogelmeier CF, Criner GJ, Martinez FJ, Anzueto A, Barnes PJ, Bourbeau J, Celli BR, Chen R, Decramer M, Fabbri LM, Frith P, Halpin DMG, López Varela MV, Nishimura M, Roche N, Rodriguez-Roisin R, Sin DD, Singh D, Stockley R, Vestbo J, Wedzicha JA, Agusti A. Global strategy for the diagnosis, management, and prevention of chronic obstructive lung disease 2017 report: Gold executive summary. *American Journal of Respiratory and Critical Care Medicine* 2017.
7. Minino AM XJ, Kochanek KD. Deaths: Preliminary data for 2008. National vital statistics reports. Hyattsville, MD: National Center for Health Statustics; 2010.
8. Kosacz NM, Croxton TL, Kiley JP, Weinmann GG, Wheaton AG, Ford ES, Presley-Cantrell LR, Croft JB, Giles WH. Mmwr: Chronic obstructive pulmonary disease among adults - united states, 2011; 2012.
9. Díaz DA, Celli DB, Celedón DJC. Chronic obstructive pulmonary disease in hispanics: A 9-year update. *American Journal of Respiratory and Critical Care Medicine*;0:null.
10. Barr RG, Avilés-Santa L, Davis SM, Aldrich TK, Gonzalez F, Henderson AG, Kaplan RC, LaVange L, Liu K, Loredo JS, Mendes ES, Ni A, Ries A, Salathe M, Smith LJ. Pulmonary disease and age at immigration among hispanics. Results from the hispanic community health study/study of latinos. *American Journal of Respiratory and Critical Care Medicine* 2015;193:386-395.
11. Bruse S, Sood A, Petersen H, Liu Y, Leng S, Celedón JC, Gilliland F, Celli B, Belinsky SA, Tesfaigzi Y. New mexican hispanic smokers have lower odds of chronic obstructive pulmonary disease and less decline in lung function than non-hispanic whites. *American Journal of Respiratory and Critical Care Medicine* 2011;184:1254-1260.
12. Diaz AA, Come CE, Mannino DM, Pinto-Plata V, Divo MJ, Bigelow C, Celli B, Washko GR. Obstructive lung disease in mexican americans and non-hispanic whites: An analysis of diagnosis and survival in the national health and nutritional examination survey iii follow-up study. *Chest* 2014;145:282-289.
13. Chen W, Brehm JM, Manichaikul A, Cho MH, Boutaoui N, Yan Q, Burkart KM, Enright PL, Rotter JI, Petersen H, Leng S, Obeidat Me, Bossé Y, Brandsma C-A, Hao K, Rich SS, Powell R, Avila L, Soto-Quiros M, Silverman EK, Tesfaigzi Y, Barr RG, Celedón JC. A genome-wide association study of chronic obstructive pulmonary disease in hispanics. *Annals of the American Thoracic Society* 2015.
14. Burkart KM, Manichaikul A, Wilk JB, Ahmed FS, Burke GL, Enright P, Hansel NN, Haynes D, Heckbert SR, Hoffman EA, Kaufman JD, Kurai J, Loehr L, London SJ, Meng Y, O'Connor GT, Oelsner E, Petrini M, Pottinger TD, Powell CA, Redline S, Rotter JI, Smith LJ, Soler Artigas M, Tobin MD, Tsai MY, Watson K, White W, Young TR, Rich SS, Barr RG. Apom and high-density lipoprotein cholesterol are associated with lung function and per cent emphysema. *European Respiratory Journal* 2014;43:1003-1017.
15. Manichaikul A. Genome-wide study of percent emphysema on computed tomography in the general population. The multi-ethnic study of atherosclerosis lung/snp health association resource study. *American Journal of Respiratory and Critical Care Medicine* 2014;189:408-418.

16. Manichaikul A, Hoffman EA, Smolonska J, Gao W, Cho MH, Baumhauer H, Budoff M, Austin JHM, Washko GR, Carr JJ, Kaufman JD, Pottinger T, Powell CA, Wijmenga C, Zanen P, Groen HJM, Postma DS, Wanner A, Rouhani FN, Brantly ML, Powell R, Smith BM, Rabinowitz D, Raffel LJ, Hinckley Stukovsky KD, Crapo JD, Beaty TH, Hokanson JE, Silverman EK, Dupuis J, O'Connor GT, Boezen HM, Rich SS, Barr RG. Genome-wide study of percent emphysema on computed tomography in the general population. The multi-ethnic study of atherosclerosis lung/snp health association resource study. *American Journal of Respiratory and Critical Care Medicine* 2014;189:408-418.
17. Soler Artigas M, Loth DW, Wain LV, Gharib SA, Obeidat Me, Tang W, Zhai G, Zhao JH, Smith AV, Huffman JE, Albrecht E, Jackson CM, Evans DM, Cadby G, Fornage M, Manichaikul A, Lopez LM, Johnson T, Aldrich MC, Aspelund T, Barroso I, Campbell H, Cassano PA, Couper DJ, Eiriksdottir G, Franceschini N, Garcia M, Gieger C, Gislason GK, Grkovic I, Hammond CJ, Hancock DB, Harris TB, Ramasamy A, Heckbert SR, Heliovaara M, Homuth G, Hysi PG, James AL, Jankovic S, Joubert BR, Karrasch S, Klopp N, Koch B, Kritchevsky SB, Launer LJ, Liu Y, Loehr LR, Lohman K, Loos RJF, Lumley T, Al Balushi KA, Ang WQ, Barr RG, Beilby J, Blakey JD, Boban M, Boraska V, Brisman J, Britton JR, Brusselle GG, Cooper C, Curjuric I, Dahgam S, Deary IJ, Ebrahim S, Eijgelsheim M, Francks C, Gaysina D, Granell R, Gu X, Hankinson JL, Hardy R, Harris SE, Henderson J, Henry A, Hingorani AD, Hofman A, Holt PG, Hui J, Hunter ML, Imboden M, Jameson KA, Kerr SM, Kolcic I, Kronenberg F, Liu JZ, Marchini J, McKeever T, Morris AD, Olin A-C, Porteous DJ, Postma DS, Rich SS, Ring SM, Rivadeneira F, Rochat T, Sayer AA, Sayers I, Sly PD, Smith GD, Sood A, Starr JM, Uitterlinden AG, Vonk JM, Wannamethee SG, Whincup PH, Wijmenga C, Williams OD, Wong A, Mangino M, Marciante KD, McArdle WL, Meibohm B, Morrison AC, North KE, Omenaas E, Palmer LJ, Pietilainen KH, Pin I, Polasek O, Pouta A, Psaty BM, Hartikainen A-L, Rantanen T, Ripatti S, Rotter JI, Rudan I, Rudnicka AR, Schulz H, Shin S-Y, Spector TD, Surakka I, Vitart V, Volzke H, Wareham NJ, Warrington NM, Wichmann HE, Wild SH, Wilk JB, Wjst M, Wright AF, Zgaga L, Zemunik T, Pennell CE, Nyberg F, Kuh D, Holloway JW, Boezen HM, Lawlor DA, Morris RW, Probst-Hensch N, Kaprio J, Wilson JF, Hayward C, Kahonen M, Heinrich J, Musk AW, Jarvis DL, Glaser S, Jarvelin M-R, Ch Stricker BH, Elliott P, O'Connor GT, Strachan DP, London SJ, Hall IP, Gudnason V, Tobin MD. Genome-wide association and large-scale follow up identifies 16 new loci influencing lung function. *Nat Genet* 2011;43:1082-1090.
18. Soler Artigas M, Wain LV, Miller S, Kheirallah AK, Huffman JE, Ntalla I, Shrine N, Obeidat Me, Trochet H, McArdle WL, Alves AC, Hui J, Zhao JH, Joshi PK, Teumer A, Albrecht E, Imboden M, Rawal R, Lopez LM, Marten J, Enroth S, Surakka I, Polasek O, Lyytikainen L-P, Granell R, Hysi PG, Flexeder C, Mahajan A, Beilby J, Bosse Y, Brandsma C-A, Campbell H, Gieger C, Glaser S, Gonzalez JR, Grallert H, Hammond CJ, Harris SE, Hartikainen A-L, Heliovaara M, Henderson J, Hocking L, Horikoshi M, Hutri-Kahonen N, Ingelsson E, Johansson A, Kemp JP, Kolcic I, Kumar A, Lind L, Melen E, Musk AW, Navarro P, Nickle DC, Padmanabhan S, Raitakari OT, Ried JS, Ripatti S, Schulz H, Scott RA, Sin DD, Starr JM, BiLEVE UK, Vinuela A, Volzke H, Wild SH, Wright AF, Zemunik T, Jarvis DL, Spector TD, Evans DM, Lehtimaki T, Vitart V, Kahonen M, Gyllensten U, Rudan I, Deary IJ, Karrasch S, Probst-Hensch NM, Heinrich J, Stubbe B, Wilson JF, Wareham NJ, James AL, Morris AP, Jarvelin M-R, Hayward C, Sayers I, Strachan DP, Hall IP, Tobin MD. Sixteen new lung function signals identified through 1000 genomes project reference panel imputation. *Nature communications* 2015;6:8658.
19. Soler Artigas M, Wain LV, Repapi E, Obeidat Me, Sayers I, Burton PR, Johnson T, Zhao JH, Albrecht E, Dominiczak AF, Kerr SM, Smith BH, Cadby G, Hui J, Palmer LJ, Hingorani AD, Wannamethee SG, Whincup PH, Ebrahim S, Smith GD, Barroso I, Loos RJF, Wareham NJ, Cooper C, Dennison E, Shaheen SO, Liu JZ, Marchini J, Health MRCNSo, Team DRS, Dahgam S, Naluai ÅT, Olin A-C, Karrasch S, Heinrich J, Schulz H, McKeever TM, Pavord ID, Heliövaara M, Ripatti S, Surakka I, Blakey JD, Kähönen M, Britton JR, Nyberg F, Holloway JW, Lawlor DA, Morris RW, James AL, Jackson CM, Hall IP, Tobin MD, Consortium tS. Effect of five genetic variants associated with lung function on the risk of chronic obstructive lung disease, and their joint effects on lung function. *American Journal of Respiratory and Critical Care Medicine* 2011;184:786-795.
20. Artigas MS, Gharib SA, Henry A, Manichaikul A, Ramasamy A, Loth DW, Imboden M, Koch B, McArdle WL, Smith AV, Smolonska J, Sood A, Tang W, Wilk JB, Zhai G, Zhao JH, Aschard H, Burkart KM, Curjuric I, Eijgelsheim M, Elliott P, Gu X, Harris TB, Janson C, Homuth G, Hysi PG, Liu JZ, Loehr LR, Lohman K, Loos RJF, Manning AK, Marciante KD, Obeidat Me, Postma DS, Aldrich MC, Brusselle GG, Chen T-h, Eiriksdottir G, Franceschini N, Heinrich J, Rotter JI, Wijmenga C, Williams OD, Bentley AR, Hofman A, Laurie CC, Lumley T, Morrison AC, Joubert BR, Rivadeneira F, Couper DJ, Kritchevsky SB, Liu Y, Wjst M, Wain LV, Vonk JM, Uitterlinden AG, Rochat T, Rich SS, Psaty BM, O'Connor GT, North

KE, Mirel DB, Meibohm B, Launer LJ, Khaw K-T, Hartikainen A-L, Hammond CJ, Glaser S, Marchini J, Kraft P, Wareham NJ, Volzke H, Stricker BHC, Spector TD, Probst-Hensch NM, Jarvis D, Jarvelin M-R, Heckbert SR, Gudnason V, Boezen HM, Barr RG, Cassano PA, Strachan DP, Fornage M, Hall IP, Dupuis J, Tobin MD, London SJ. Genome-wide joint meta-analysis of snp and snp-by-smoking interaction identifies novel loci for pulmonary function. *PLoS Genetics* 2012;8:e1003098.
21. Pillai SG. A genome-wide association study in chronic obstructive pulmonary disease (copd): Identification of two major susceptibility loci. *PLoS Genet* 2009;5:e1000421.
22. Wilk JB. A genome-wide association study of pulmonary function measures in the framingham heart study. *PLoS Genet* 2009;5:e1000429.
23. Wilk JB, Shrine NRG, Loehr LR, Zhao JH, Manichaikul A, Lopez LM, Smith AV, Heckbert SR, Smolonska J, Tang W, Loth DW, Curjuric I, Hui J, Cho MH, Latourelle JC, Henry AP, Aldrich M, Bakke P, Beaty TH, Bentley AR, Borecki IB, Brusselle GG, Burkart KM, Chen T-h, Couper D, Crapo JD, Davies G, Dupuis J, Franceschini N, Gulsvik A, Hancock DB, Harris TB, Hofman A, Imboden M, James AL, Khaw K-T, Lahousse L, Launer LJ, Litonjua A, Liu Y, Lohman KK, Lomas DA, Lumley T, Marciante KD, McArdle WL, Meibohm B, Morrison AC, Musk AW, Myers RH, North KE, Postma DS, Psaty BM, Rich SS, Rivadeneira F, Rochat T, Rotter JI, Soler Artigas M, Starr JM, Uitterlinden AG, Wareham NJ, Wijmenga C, Zanen P, Province MA, Silverman EK, Deary IJ, Palmer LJ, Cassano PA, Gudnason V, Barr RG, Loos RJF, Strachan DP, London SJ, Boezen HM, Probst-Hensch N, Gharib SA, Hall IP, O'Connor GT, Tobin MD, Stricker BH. Genome wide association studies identify chrna5/3 and htr4 in the development of airflow obstruction. *American Journal of Respiratory and Critical Care Medicine* 2012.
24. Cho MH, Boutaoui N, Klanderman BJ, Sylvia JS, Ziniti JP, Hersh CP, DeMeo DL, Hunninghake GM, Litonjua AA, Sparrow D, Lange C, Won S, Murphy JR, Beaty TH, Regan EA, Make BJ, Hokanson JE, Crapo JD, Kong X, Anderson WH, Tal-Singer R, Lomas DA, Bakke P, Gulsvik A, Pillai SG, Silverman EK. Variants in fam13a are associated with chronic obstructive pulmonary disease. *Nat Genet* 2010;42:200-202.
25. Castaldi PJ, Cho MH, Litonjua AA, Bakke P, Gulsvik A, Lomas DA, Anderson W, Beaty TH, Hokanson JE, Crapo JD, Laird N, Silverman EK, COPDGene ft, Investigators E. The association of genome-wide significant spirometric loci with chronic obstructive pulmonary disease susceptibility. *American Journal of Respiratory Cell and Molecular Biology* 2011;45:1147-1153.
26. Cho MH, McDonald M-LN, Zhou X, Mattheisen M, Castaldi PJ, Hersh CP, DeMeo DL, Sylvia JS, Ziniti J, Laird NM, Lange C, Litonjua AA, Sparrow D, Casaburi R, Barr RG, Regan EA, Make BJ, Hokanson JE, Lutz S, Murray T, Farzadegan H, Hetmanski JB, Tal-Singer R, Lomas DA, Bakke P, Gulsvik A, Crapo JD, Silverman EK, Beaty TH, on behalf of the Nett Genetics IE, Investigators CO. Risk loci for chronic obstructive pulmonary disease: A genome-wide association study and meta-analysis. *The Lancet Respiratory medicine* 2014;2:214-225.
27. Hancock DB, Eijgelsheim M, Wilk JB, Gharib SA, Loehr LR, Marciante KD, Franceschini N, van Durme YM, Chen TH, Barr RG, Schabath MB, Couper DJ, Brusselle GG, Psaty BM, van Duijn CM, Rotter JI, Uitterlinden AG, Hofman A, Punjabi NM, Rivadeneira F, Morrison AC, Enright PL, North KE, Heckbert SR, Lumley T, Stricker BH, O'Connor GT, London SJ. Meta-analyses of genome-wide association studies identify multiple loci associated with pulmonary function. *Nature Genetics* 2010;42:45-52.
28. Hancock DB, Soler Artigas M, Gharib SA, Henry A, Manichaikul A, Ramasamy A, Loth DW, Imboden M, Koch B, McArdle WL, Smith AV, Smolonska J, Sood A, Tang W, Wilk JB, Zhai G, Zhao JH, Aschard H, Burkart KM, Curjuric I, Eijgelsheim M, Elliott P, Gu X, Harris TB, Janson C, Homuth G, Hysi PG, Liu JZ, Loehr LR, Lohman K, Loos RJF, Manning AK, Marciante KD, Obeidat Me, Postma DS, Aldrich MC, Brusselle GG, Chen T-h, Eiriksdottir G, Franceschini N, Heinrich J, Rotter JI, Wijmenga C, Williams OD, Bentley AR, Hofman A, Laurie CC, Lumley T, Morrison AC, Joubert BR, Rivadeneira F, Couper DJ, Kritchevsky SB, Liu Y, Wjst M, Wain LV, Vonk JM, Uitterlinden AG, Rochat T, Rich SS, Psaty BM, O'Connor GT, North KE, Mirel DB, Meibohm B, Launer LJ, Khaw K-T, Hartikainen A-L, Hammond CJ, Gläser S, Marchini J, Kraft P, Wareham NJ, Völzke H, Stricker BHC, Spector TD, Probst-Hensch NM, Jarvis D, Jarvelin M-R, Heckbert SR, Gudnason V, Boezen HM, Barr RG, Cassano PA, Strachan DP, Fornage M, Hall IP, Dupuis J, Tobin MD, London SJ. Genome-wide joint meta-analysis of snp and snp-by-smoking interaction identifies novel loci for pulmonary function. *PLoS Genet* 2012;8:e1003098.
29. Repapi E, Sayers I, Wain LV, Burton PR, Johnson T, Obeidat M, Zhao JH, Ramasamy A, Zhai G, Vitart V, Huffman JE, Igl W, Albrecht E, Deloukas P, Henderson J, Granell R, McArdle WL, Rudnicka AR, Wellcome Trust Case Control C, Barroso I, Loos RJ, Wareham NJ, Mustelin L, Rantanen T, Surakka I, Imboden M, Wichmann HE, Grkovic I, Jankovic S, Zgaga L, Hartikainen AL, Peltonen L, Gyllensten U,

Johansson A, Zaboli G, Campbell H, Wild SH, Wilson JF, Glaser S, Homuth G, Volzke H, Mangino M, Soranzo N, Spector TD, Polasek O, Rudan I, Wright AF, Heliovaara M, Ripatti S, Pouta A, Naluai AT, Olin AC, Toren K, Cooper MN, James AL, Palmer LJ, Hingorani AD, Wannamethee SG, Whincup PH, Smith GD, Ebrahim S, McKeever TM, Pavord ID, MacLeod AK, Morris AD, Porteous DJ, Cooper C, Dennison E, Shaheen S, Karrasch S, Schnabel E, Schulz H, Grallert H, Bouatia-Naji N, Delplanque J, Froguel P, Blakey JD, Team NRS, Britton JR, Morris RW, Holloway JW, Lawlor DA, Hui J, Nyberg F, Jarvelin MR, Jackson C, Kahonen M, Kaprio J, Probst-Hensch NM, Koch B, Hayward C, Evans DM, Elliott P, Strachan DP, Hall IP, Tobin MD. Genome-wide association study identifies five loci associated with lung function. *Nature Genetics* 2010;42:36-44.

30. Wain LV, Shrine N, Artigas MS, Erzurumluoglu AM, Noyvert B, Bossini-Castillo L, Obeidat Me, Henry AP, Portelli MA, Hall RJ, Billington CK, Rimington TL, Fenech AG, John C, Blake T, Jackson VE, Allen RJ, Prins BP, Understanding Society Scientific G, Campbell A, Porteous DJ, Jarvelin M-R, Wielscher M, James AL, Hui J, Wareham NJ, Zhao JH, Wilson JF, Joshi PK, Stubbe B, Rawal R, Schulz H, Imboden M, Probst-Hensch NM, Karrasch S, Gieger C, Deary IJ, Harris SE, Marten J, Rudan I, Enroth S, Gyllensten U, Kerr SM, Polasek O, Kahonen M, Surakka I, Vitart V, Hayward C, Lehtimaki T, Raitakari OT, Evans DM, Henderson AJ, Pennell CE, Wang CA, Sly PD, Wan ES, Busch R, Hobbs BD, Litonjua AA, Sparrow DW, Gulsvik A, Bakke PS, Crapo JD, Beaty TH, Hansel NN, Mathias RA, Ruczinski I, Barnes KC, Bosse Y, Joubert P, van den Berge M, Brandsma C-A, Pare PD, Sin DD, Nickle DC, Hao K, Gottesman O, Dewey FE, Bruse SE, Carey DJ, Kirchner HL, Geisinger-Regeneron Discov EHRC, Jonsson S, Thorleifsson G, Jonsdottir I, Gislason T, Stefansson K, Schurmann C, Nadkarni G, Bottinger EP, Loos RJF, Walters RG, Chen Z, Millwood IY, Vaucher J, Kurmi OP, Li L, Hansell AL, Brightling C, Zeggini E, Cho MH, Silverman EK, Sayers I, Trynka G, Morris AP, Strachan DP, Hall IP, Tobin MD. Genome-wide association analyses for lung function and chronic obstructive pulmonary disease identify new loci and potential druggable targets. *Nat Genet* 2017;advance online publication.

31. Hobbs BD, de Jong K, Lamontagne M, Bosse Y, Shrine N, Artigas MS, Wain LV, Hall IP, Jackson VE, Wyss AB, London SJ, North KE, Franceschini N, Strachan DP, Beaty TH, Hokanson JE, Crapo JD, Castaldi PJ, Chase RP, Bartz TM, Heckbert SR, Psaty BM, Gharib SA, Zanen P, Lammers JW, Oudkerk M, Groen HJ, Locantore N, Tal-Singer R, Rennard SI, Vestbo J, Timens W, Pare PD, Latourelle JC, Dupuis J, O'Connor GT, Wilk JB, Kim WJ, Lee MK, Oh Y-M, Vonk JM, de Koning HJ, Leng S, Belinsky SA, Tesfaigzi Y, Manichaikul A, Wang X-Q, Rich SS, Barr RG, Sparrow D, Litonjua AA, Bakke P, Gulsvik A, Lahousse L, Brusselle GG, Stricker BH, Uitterlinden AG, Ampleford EJ, Bleecker ER, Woodruff PG, Meyers DA, Qiao D, Lomas DA, Yim J-J, Kim DK, Hawrylkiewicz I, Sliwinski P, Hardin M, Fingerlin TE, Schwartz DA, Postma DS, MacNee W, Tobin MD, Silverman EK, Boezen HM, Cho MH, Investigators CO, Investigators E, LifeLines I, Group SR, International CGNI, Investigators UKB, International CGC. Genetic loci associated with chronic obstructive pulmonary disease overlap with loci for lung function and pulmonary fibrosis. *Nat Genet* 2017;advance online publication.

32. LaVange LM, Kalsbeek WD, Sorlie PD, Avilés-Santa LM, Kaplan RC, Barnhart J, Liu K, Giachello A, Lee DJ, Ryan J, Criqui MH, Elder JP. Sample design and cohort selection in the hispanic community health study/study of latinos. *Ann Epidemiol*;20:642-649.

33. Kristin MB, Adrienne MS, Tamar S, Stephanie L, Sonia D, Juan CC, Barr RG. Genome-wide association study (gwas) of lung function among hispanic/latino individuals of diverse backgrounds. The hispanic community health study/study of latinos (hchs/sol). A21 genetic regulation of chronic airway inflammation: American Thoracic Society; 2015. p. A1069-A1069.

34. Sorlie PD, Avilés-Santa LM, Wassertheil-Smoller S, Kaplan RC, Daviglus ML, Giachello AL, Schneiderman N, Raij L, Talavera G, Allison M, LaVange L, Chambless LE, Heiss G. Design and implementation of the hispanic community health study/study of latinos. *Ann Epidemiol* 2010;20:629-641.

35. Conomos Matthew P, Laurie Cecelia A, Stilp Adrienne M, Gogarten Stephanie M, McHugh Caitlin P, Nelson Sarah C, Sofer T, Fernández-Rhodes L, Justice Anne E, Graff M, Young Kristin L, Seyerle Amanda A, Avery Christy L, Taylor Kent D, Rotter Jerome I, Talavera Gregory A, Daviglus Martha L, Wassertheil-Smoller S, Schneiderman N, Heiss G, Kaplan Robert C, Franceschini N, Reiner Alex P, Shaffer John R, Barr R G, Kerr Kathleen F, Browning Sharon R, Browning Brian L, Weir Bruce S, Avilés-Santa M L, Papanicolaou George J, Lumley T, Szpiro Adam A, North Kari E, Rice K, Thornton Timothy A, Laurie Cathy C. Genetic diversity and association studies in us hispanic/latino populations: Applications in the hispanic community health study/study of latinos. *American Journal of Human Genetics* 2016;98:165-184.

36. Pellegrino R. Interpretative strategies for lung function tests. *Eur Respir J* 2005;26:948-968.

37. Hankinson JL, Odencrantz JR, Fedan KB. Spirometric reference values from a sample of the general u.S. Population.[see comment]. *American Journal of Respiratory & Critical Care Medicine* 1999;159:179-187.
38. LaVange PL, Davis PSM, Hankinson DJ, Enright DP, Wilson MR, Barr DRG, Aldrich DTK, Kalhan DR, Lemus DH, Ni DA, Smith DLJ, Talavera DGA. Spirometry reference equations from the hispanic community health study/study of latinos (hchs/sol). *American Journal of Respiratory and Critical Care Medicine*;0:null.
39. Laurie CC, Doheny KF, Mirel DB, Pugh EW, Bierut LJ, Bhangale T, Boehm F, Caporaso NE, Cornelis MC, Edenberg HJ, Gabriel SB, Harris EL, Hu FB, Jacobs KB, Kraft P, Landi MT, Lumley T, Manolio TA, McHugh C, Painter I, Paschall J, Rice JP, Rice KM, Zheng X, Weir BS, for the GI. Quality control and quality assurance in genotypic data for genome-wide association studies. *Genetic Epidemiology* 2010;34:591-602.
40. The Genomes Project C. An integrated map of genetic variation from 1,092 human genomes. *Nature* 2012;491:56-65.
41. Sofer T, Shaffer JR, Graff M, Qi Q, Stilp AM, Gogarten SM, North KE, Isasi CR, Laurie CC, Szpiro AA. Meta-analysis of genome-wide association studies with correlated individuals: Application to the hispanic community health study/study of latinos (hchs/sol). *Genetic Epidemiology* 2016;40:492-501.
42. Pe'er I, Yelensky R, Altshuler D, Daly MJ. Estimation of the multiple testing burden for genomewide association studies of nearly all common variants. *Genet Epidemiol* 2008;32:381-385.
43. Chen H, Wang C, Conomos Matthew P, Stilp Adrienne M, Li Z, Sofer T, Szpiro Adam A, Chen W, Brehm John M, Celedón Juan C, Redline S, Papanicolaou George J, Thornton Timothy A, Laurie Cathy C, Rice K, Lin X. Control for population structure and relatedness for binary traits in genetic association studies via logistic mixed models. *The American Journal of Human Genetics* 2016;98:653-666.
44. Browning SR, Grinde K, Plantinga A, Gogarten SM, Stilp AM, Kaplan RC, Avilés-Santa ML, Browning BL, Laurie CC. Local ancestry inference in a large us-based hispanic/latino study: Hispanic community health study/study of latinos (hchs/sol). *G3: Genes|Genomes|Genetics* 2016;6:1525-1534.
45. Sofer T, Baier LJ, Browning SR, Thornton TA, Talavera GA, Wassertheil-Smoller S, Daviglus ML, Hanson R, Kobes S, Cooper RS, Cai J, Levy D, Reiner AP, Franceschini N. Admixture mapping in the hispanic community health study/study of latinos reveals regions of genetic associations with blood pressure traits. *PLoS One* 2017;12:e0188400.
46. Zhang QS, Browning BL, Browning SR. Asafe: Ancestry-specific allele frequency estimation. *Bioinformatics* 2016.
47. Bild DE, Bluemke DA, Burke GL, Detrano R, Diez Roux AV, Folsom AR, Greenland P, JacobsJr DR, Kronmal R, Liu K, Nelson JC, O'Leary D, Saad MF, Shea S, Szklo M, Tracy RP. Multi-ethnic study of atherosclerosis: Objectives and design. *American Journal of Epidemiology* 2002;156:871-881.
48. Kaufman JD, Adar SD, Allen RW, Barr RG, Budoff MJ, Burke GL, Casillas AM, Cohen MA, Curl CL, Daviglus ML, Diez Roux AV, Jacobs DR, Jr., Kronmal RA, Larson TV, Liu SL-J, Lumley T, Navas-Acien A, O'Leary DH, Rotter JI, Sampson PD, Sheppard L, Siscovick DS, Stein JH, Szpiro AA, Tracy RP. Prospective study of particulate air pollution exposures, subclinical atherosclerosis, and clinical cardiovascular disease: The multi-ethnic study of atherosclerosis and air pollution (mesa air). *American Journal of Epidemiology* 2012;176:825-837.
49. Victora CG, Barros FC. Cohort profile: The 1982 pelotas (brazil) birth cohort study. *International Journal of Epidemiology* 2006;35:237-242.
50. Horta BL, Gigante DP, Gonçalves H, dos Santos Motta J, Loret de Mola C, Oliveira IO, Barros FC, Victora CG. Cohort profile update: The 1982 pelotas (brazil) birth cohort study. *International Journal of Epidemiology* 2015;44:441-441e.
51. Chen W, Brehm JM, Boutaoui N, Soto-Quiros M, Avila L, Celli BR, Bruse S, Tesfaigzi Y, Celedón JC. Native american ancestry, lung function, and copd in costa ricans. *CHEST Journal* 2014;145:704-710.
52. Sofer T, Heller R, Bogomolov M, Avery CL, Graff M, North KE, Reiner AP, Thornton TA, Rice K, Benjamini Y, Laurie CC, Kerr KF. A powerful statistical framework for generalization testing in gwas, with application to the hchs/sol. *Genetic Epidemiology* 2017;41:251-258.
53. Spier I, Kerick M, Drichel D, Horpaopan S, Altmüller J, Laner A, Holzapfel S, Peters S, Adam R, Zhao B, Becker T, Lifton RP, Holinski-Feder E, Perner S, Thiele H, Nöthen MM, Hoffmann P,

Timmermann B, Schweiger MR, Aretz S. Exome sequencing identifies potential novel candidate genes in patients with unexplained colorectal adenomatous polyposis. *Familial Cancer* 2016;15:281-288.
54. Sung YJ, de las Fuentes L, Schwander KL, Simino J, Rao DC. Gene–smoking interactions identify several novel blood pressure loci in the framingham heart study. *American Journal of Hypertension* 2015;28:343-354.
55. Martín V, Chahwan C, Gao H, Blais V, Wohlschlegel J, Yates JR, McGowan CH, Russell P. Sws1 is a conserved regulator of homologous recombination in eukaryotic cells. *The EMBO Journal* 2006;25:2564-2574.
56. Godin SK, Meslin C, Kabbinavar F, Bratton-Palmer DS, Hornack C, Mihalevic MJ, Yoshida K, Sullivan M, Clark NL, Bernstein KA. Evolutionary and functional analysis of the invariant swim domain in the conserved shu2/sws1 protein family from saccharomyces cerevisiae to homo sapiens. *Genetics* 2015;199:1023-1033.
57. Lou Z, Chen J. Cellular senescence and DNA repair. *Experimental Cell Research* 2006;312:2641-2646.
58. Tuder RM, Petrache I. Pathogenesis of chronic obstructive pulmonary disease. *The Journal of Clinical Investigation*;122:2749-2755.
59. MacNee W, Tuder RM. New paradigms in the pathogenesis of chronic obstructive pulmonary disease i. *Proceedings of the American Thoracic Society* 2009;6:527-531.
60. Neofytou E, Tzortzaki EG, Chatziantoniou A, Siafakas NM. DNA damage due to oxidative stress in chronic obstructive pulmonary disease (copd). *International Journal of Molecular Sciences* 2012;13:16853-16864.
61. Caramori G, Adcock IM, Casolari P, Ito K, Jazrawi E, Tsaprouni L, Villetti G, Civelli M, Carnini C, Chung KF, Barnes PJ, Papi A. Unbalanced oxidant-induced DNA damage and repair in copd: A link towards lung cancer. *Thorax* 2011.
62. Aoshiba K, Zhou F, Tsuji T, Nagai A. DNA damage as a molecular link in the pathogenesis of copd in smokers. *European Respiratory Journal* 2012;39:1368.
63. National center for biotechnology information entrez dbsnp: Rs192375903; build 38. [cited 2016 11/1]. Available from: https://www.ncbi.nlm.nih.gov/projects/SNP/snp_ref.cgi?rs=192375903.
64. Suchi M, Sano H, Mizuno H, Wada Y. Molecular cloning and structural characterization of the human histidase gene (hal). *Genomics* 1995;29:98-104.
65. Yu B, Li AH, Muzny D, Veeraraghavan N, de Vries PS, Bis JC, Musani SK, Alexander D, Morrison AC, Franco OH, Uitterlinden A, Hofman A, Dehghan A, Wilson JG, Psaty BM, Gibbs R, Wei P, Boerwinkle E. Association of rare loss-of-function alleles in hal, serum histidine levels and incident coronary heart disease. *Circulation Cardiovascular genetics* 2015;8:351-355.
66. Taylor RG, Lambert MA, Sexsmith E, Sadler SJ, Ray PN, Mahuran DJ, McInnes RR. Cloning and expression of rat histidase. Homology to two bacterial histidases and four phenylalanine ammonia-lyases. *Journal of Biological Chemistry* 1990;265:18192-18199.
67. Niu Y-C, Feng R-N, Hou Y, Li K, Kang Z, Wang J, Sun C-H, Li Y. Histidine and arginine are associated with inflammation and oxidative stress in obese women. *British Journal of Nutrition* 2012;108:57-61.
68. Watanabe M, Suliman ME, Qureshi AR, Garcia-Lopez E, Barany P, Heimburger O, Stenvinkel P, Lindholm B. Consequences of low plasma histidine in chronic kidney disease patients: Associations with inflammation, oxidative stress, and mortality. *American Journal of Clinical Nutrition* 2008;87:1860-1866.
69. Ubhi BK, Cheng KK, Dong J, Janowitz T, Jodrell D, Tal-Singer R, MacNee W, Lomas DA, Riley JH, Griffin JL, Connor SC. Targeted metabolomics identifies perturbations in amino acid metabolism that sub-classify patients with copd. *Molecular Biosystems* 2012;8:3125-3133.
70. Wyss AB, Sofer T, Lee MK, Terzikhan N, Nguyen JN, Lahousse L, Latourelle JC, Smith AV, Bartz TM, Feitosa MF, Gao W, Ahluwalia TS, Tang W, Oldmeadow C, Duan Q, de Jong K, Wojczynski MK, Wang X-Q, Noordam R, Hartwig FP, Jackson VE, Wang T, Obeidat M, en, Hobbs BD, Huan T, Kichaev G, Jin J, Graff M, Harris TB, Kalhan R, Heckbert SR, Paternoster L, Burkart KM, Liu Y, Holliday EG, Wilson JG, Vonk JM, Sanders J, Barr RG, de Mutsert R, Baptista Menezes AM, Adams HHH, van den Berge M, Joehanes R, Launer LJ, Morrison AC, Sitlani CM, Celedón JC, Kritchevsky SB, Scott RJ, Christensen K, Rotter JI, Bonten TN, Wehrmeister FC, Bossé Y, Franceschini N, Brody JA, Kaplan RC, Lohman K, McEvoy M, Province MA, Rosendaal FR, Taylor KD, Nickle DC, Gudnason V, North KE, Fornage M, Psaty BM, Myers RH, Connor G, Hansen T, Laurie CC, Cassano P, Sung J, Kim WJ, Attia JR, Lange L, Boezen HM, Thyagarajan B, Rich SS, Mook-Kanamori DO, Horta BL, Uitterlinden AG, Sin

DD, Im HK, Cho MH, Brusselle GG, Gharib SA, Dupuis J, Manichaikul A, London SJ. Multiethnic meta-analysis identifies new loci for pulmonary function. *bioRxiv* 2017.
71. National center for biotechnology information entrez dbsnp: Rs7228593 ; build 38. [cited 2016 November]. Available from: https://www.ncbi.nlm.nih.gov/projects/SNP/snp_ref.cgi?rs=7228593.
72. Pouladi N, Bime C, Garcia JGN, Lussier YA. Complex genetics of pulmonary diseases: Lessons from genome-wide association studies and next-generation sequencing. *Translational Research* 2016;168:22-39.
73. Loth DW, Artigas MS, Gharib SA, Wain LV, Franceschini N, Koch B, Pottinger T, Smith AV, Duan Q, Oldmeadow C, Lee MK, Strachan DP, James AL, Huffman JE, Vitart V, Ramasamy A, Wareham NJ, Kaprio J, Wang X-Q, Trochet H, Kähönen M, Flexeder C, Albrecht E, Lopez LM, de Jong K, Thyagarajan B, Alves AC, Enroth S, Omenaas E, Joshi PK, Fall T, Viñuela A, Launer LJ, Loehr LR, Fornage M, Li G, Wilk JB, Tang W, Manichaikul A, Lahousse L, Harris TB, North KE, Rudnicka AR, Hui J, Gu X, Lumley T, Wright AF, Hastie ND, Campbell S, Kumar R, Pin I, Scott RA, Pietiläinen KH, Surakka I, Liu Y, Holliday EG, Schulz H, Heinrich J, Davies G, Vonk JM, Wojczynski M, Pouta A, Johansson Å, Wild SH, Ingelsson E, Rivadeneira F, Völzke H, Hysi PG, Eiriksdottir G, Morrison AC, Rotter JI, Gao W, Postma DS, White WB, Rich SS, Hofman A, Aspelund T, Couper D, Smith LJ, Psaty BM, Lohman K, Burchard EG, Uitterlinden AG, Garcia M, Joubert BR, McArdle WL, Musk AB, Hansel N, Heckbert SR, Zgaga L, van Meurs JBJ, Navarro P, Rudan I, Oh Y-M, Redline S, Jarvis D, Zhao JH, Rantanen T, O'Connor GT, Ripatti S, Scott RJ, Karrasch S, Grallert H, Gaddis NC, Starr JM, Wijmenga C, Minster RL, Lederer DJ, Pekkanen J, Gyllensten U, Campbell H, Morris AP, Gläser S, Hammond CJ, Burkart KM, Beilby J, Kritchevsky SB, Gudnason V, Hancock DB, Williams OD, Polasek O, Zemunik T, Kolcic I, Petrini MF, Wjst M, Kim WJ, Porteous DJ, Scotland G, Smith BH, Viljanen A, Heliövaara M, Attia JR, Sayers I, Hampel R, Gieger C, Deary IJ, Boezen HM, Newman A, Jarvelin M-R, Wilson JF, Lind L, Stricker BH, Teumer A, Spector TD, Melén E, Peters MJ, Lange LA, Barr RG, Bracke KR, Verhamme FM, Sung J, Hiemstra PS, Cassano PA, Sood A, Hayward C, Dupuis J, Hall IP, Brusselle GG, Tobin MD, London SJ. Genome-wide association analysis identifies six new loci associated with forced vital capacity. *Nature Genetics* 2014;46:669-677.
74. Burkart KM, Wilk JB, Enright P, Ahmed FS, Hansel NN, Haynes D, Heckbert SR, Loehr LR, London SJ, Lyon H, O'Conner GT, Oelsner E, Petrini MF, Redline S, Smilth L, White WB, Rich SS, Barr RG. Candidate-wide association study of lung function among european-and african-americans in 8 population-based cohorts. The candidate-gene association resource (care). *European Respiratory Journal* 2010
Supplement 54:1566.
75. Wain LV, Shrine N, Miller S, Jackson VE, Ntalla I, Artigas MS, Billington CK, Kheirallah AK, Allen R, Cook JP, Probert K, Obeidat Me, Bossé Y, Hao K, Postma DS, Paré PD, Ramasamy A, Mägi R, Mihailov E, Reinmaa E, Melén E, O'Connell J, Frangou E, Delaneau O, Freeman C, Petkova D, McCarthy M, Sayers I, Deloukas P, Hubbard R, Pavord I, Hansell AL, Thomson NC, Zeggini E, Morris AP, Marchini J, Strachan DP, Tobin MD, Hall IP. Novel insights into the genetics of smoking behaviour, lung function, and chronic obstructive pulmonary disease (uk bileve): A genetic association study in uk biobank. *The Lancet Respiratory Medicine*;3:769-781.

**Figure 1** Regional Association Plots and Forest Plots of genome wide significant loci associated with $FEV_1$ and $FEV_1$/FVC meta-analyzed across all Hispanic/Latino ancestry groups.

For each locus, we provide regional association plots with correlations between the reference SNP (the SNP with the lowest p-value) and other SNPs in the region. The reference SNPs a purple diamond (genotyped SNP) or upside down triangle (imputed SNP). Other SNPs in the region are depicted as circles (genotyped SNPs) and "X"s (imputed SNPs) . The correlations ($r^2$) are calculated from the group of interest and are indicated by the colors shown on the plot. For each locus, we also provide a forest plot comparing the SNP-trait association testing results across the Hispanic/Latino ancestry groups. (a) Regional association plot for the $FEV_1$ locus (lead SNP rs4791658) on chromosome 17. (b) Forrest Plot for the $FEV_1$ locus (lead SNP rs4791658) on chromosome 17. (c) Regional association plot for the $FEV_1$/FVC locus (lead SNP rs145174011) on chromosome 12. (d) Forrest Plot for the $FEV_1$/FVC locus (lead SNP rs192375903) on chromosome 12.

**Figure 2** Regional Association Plots and Forest Plots of genome wide significant loci associated with $FEV_1$ and $FEV_1$/FVC stratified by smoking status and meta-analyzed across all Hispanic/Latino Ancestry groups.

For each locus, we provide regional association plots with correlations between the reference SNP (the SNP with the lowest p-value) and other SNPs in the region. The reference SNPs a purple diamond (genotyped SNP) or upside down triangle (imputed SNP). Other SNPs in the region are depicted as circles (genotyped SNPs) and "X"s (imputed SNPs). The correlations ($r^2$) are calculated from the group of interest and are indicated by the colors shown on the plot. For each locus, we also provide a forest plot comparing the SNP-trait association testing results across the Hispanic/Latino ancestry in smoking stratified group of interest. (a) Regional association plot for the $FEV_1$/FVC locus among never smokers (lead SNP rs7228593) on chromosome 18. (b) Forrest Plot for the $FEV_1$/FVC locus among never smokers (lead SNP rs7228593) on chromosome 18. (c) Regional association plot for the $FEV_1$ locus among ever smokers (lead SNP rs291231) on chromosome 11. (d) Forrest Plot for the $FEV_1$ locus among ever smokers (lead SNP rs291231) on chromosome 11.

**Figure 3** Regional Association Plot and Forest Plot of candidate variant identified in local ancestry interval region associated with Amerindian ancestry and $FEV_1$ meta-analyzed across all Hispanic/Latino ancestry groups.

For each locus, we provide regional association plots with correlations between the reference SNP (the SNP with the lowest p-value) and other SNPs in the region. The reference SNPs a purple diamond (genotyped SNP) or upside-down triangle (imputed SNP). Other SNPs in the region are depicted as circles (genotyped SNPs) and "X"s (imputed SNPs) . The correlations ($r^2$) are calculated from the group of interest and are indicated by the colors shown on the plot. We also provide a forest plot comparing the SNP-trait association testing results across the Hispanic/Latino ancestry groups. (a) Regional association plot for the $FEV_1$ locus (lead SNP rs4133185) on chromosome 7. (b) Forrest Plot for the $FEV_1$ locus (lead SNP rs4133185) on chromosome 7.

**Figure 4** Amerindian Admixture Mapping Region on Chromosome 7 for $FEV_1$

The left panel provides the admixture mapping results as two lines with the blue line representing results from the primary analysis, the green line representing results from the conditional analysis and the association results in the same region are represented as circles. The genome-wide significant threshold for admixture mapping in HCHS/SOL data set is the horizontal grey dashed line. The blue and green lines and circles are given as –log(p-value,10) against genomic positions. The red-filled triangle corresponds to the SNP used in the conditional analysis (rs4133185). The right panel provides the ancestry-specific effect allele frequencies for the SNP used in the conditional analysis (rs4133185), as estimated by ASAFE applied on HCHS/SOL data set(46).

**Table 1: Characteristics of Participants in HCHS/SOL Study**

| | Central American | Cuban | Dominican | Mexican | Puerto Rican | South American | Combined |
|---|---|---|---|---|---|---|---|
| Number of participants | 1,315 | 2,138 | 1,063 | 4,438 | 2,000 | 868 | 11,822 |
| Age (years) | 44 ± 13 | 49 ± 13 | 45 ± 14 | 44 ± 14 | 48 ± 14 | 46 ± 13 | 46 ± 14 |
| Male Sex (%) | 40 | 47 | 35 | 39 | 43 | 40 | 41 |
| Height (cm) | 160 ± 9 | 164 ± 9 | 162 ± 9 | 161 ± 9 | 163 ± 9 | 161 ± 9 | 162 ± 9 |
| BMI (kg/m$^2$) | 30 ± 6 | 29 ± 6 | 29 ± 6 | 30 ± 6 | 31 ± 7 | 29 ± 5 | 30 ± 6 |
| Smoking (never) % | 68 | 52 | 77 | 65 | 48 | 66 | 61 |
| Smoking (former) % | 19 | 19 | 13 | 19 | 20 | 20 | 19 |
| Smoking (current) % | 13 | 29 | 10 | 16 | 32 | 14 | 20 |
| Pack years of smoking* | 4.5 (1.7 - 14) | 18.5 (6 - 36) | 8.2 (3.4 – 18.3) | 4.0 (1.3 – 10.5) | 11.0 (4 – 25.5) | 4.9 (1.5 - 16) | 7.5 (2.4 - 21) |
| Pre-bronchodilator $FEV_1$ (liters) | 2.88 ± 0.74 | 2.80 ± 0.81 | 2.70±0.75 | 2.97 ± 0.78 | 2.70 ± 0.83 | 2.95 ± 0.79 | 2.86 ± 0.80 |
| Pre-bronchodilator $FEV_1$/FVC (%) | 81.5 ± 5.9 | 78.7 ± 8.3 | 81.8 ± 7.2 | 80.8 ± 6.1 | 79.3 ± 7.9 | 80.1 ± 6.5 | 80.3 ± 7 |
| COPD n (%)† | 24 (1.8) | 124 (5.8) | 24 (2.3) | 85 (1.9) | 86 (4.3) | 20 (2.3) | 363 (3.1) |
| Airflow Limitation n (%)‡ | 60 (4.6) | 260 (12.2) | 62 (5.8) | 207 (4.7) | 216 (10.8) | 49 (5.7) | 854 (7.2) |

Data presented as mean ± standard deviation for continuous measures and percentage for binary measures

*Definition of abbreviations*: $FEV_1$ = forced expiratory volume in one second, $FEV_1$/FVC = ratio of forced expiratory volume in one second over forced vital capacity %

* Pack years of smoking presented as median ± Inter Quartile Range (IQR) in former or current smokers

† Chronic Obstructive Pulmonary Disease defined as post-bronchodilator $FEV_1$/FVC ratio < 0.70

‡ Defined as pre-bronchodilator $FEV_1$/FVC ratio < 0.70

**Table 2: Genome wide significant loci associated with $FEV_1$ and $FEV_1/FVC$**

| SNP rsID | Trait | Position Build 37 Chr: basepair | Gene /nearest Gene(s) (function) | Effect allele | EAF | Genotyped or Imputed SNP | oevar* | N | Beta | SE | P Value |
|---|---|---|---|---|---|---|---|---|---|---|---|
| rs76656601 | $FEV_1/FVC$ | 6: 168428548 | *KIF25* (intron) | G | 0.996 | Imputed | 0.7 | 11822 | 4.70 | 0.77 | $1.31 \times 10^{-9}$ |
| **rs4791658** | **$FEV_1$** | **17:15884792** | ***ZSWIM7*† (intron)** | **G** | **0.541** | **Imputed‡** | **0.996** | **11822** | **33.35** | **5.70** | **$4.99 \times 10^{-9}$** |
| **rs145174011** | **$FEV_1/FVC$** | **12: 96378700** | ***HAL*† (Intron)** | **T** | **0.997** | **Imputed** | **0.98** | **11822** | **4.85** | **0.85** | **$9.59 \times 10^{-9}$** |
| rs28593428 | $FEV_1/FVC$ | 21: 35632170 | *KCNE2* | C | 0.902 | Imputed | 0.996 | 11821 | 0.80 | 0.14 | $1.45 \times 10^{-8}$ |
| rs262113 | $FEV_1/FVC$ | 6: 142824950 | *GPR126* | T | 0.812 | Imputed† | 0.999 | 11822 | - 0.57 | 0.10 | $2.83 \times 10^{-8}$ |
| rs74444778 | $FEV_1/FVC$ | 7: 42553708 | *LOC105375250* (intron) | G | 0.999 | Imputed | 0.914 | 11822 | 6.17 | 1.12 | $3.61 \times 10^{-8}$ |
| rs115745680 | $FEV_1$ | 5: 154407079 | *GEMIN5 KIF4B* | A | 0.997 | Imputed | 0.981 | 11822 | 313.24 | 57.32 | $4.63 \times 10^{-8}$ |

*Definition of abbreviations*: SNP= single nucleotide polymorphism, $FEV_1/FVC$= ratio of forced expiratory volume in one second over forced vital capacity,

Chr = chromosome, nearest gene = the nearest genes on each side of the intergenic marker, EAF = effect allele frequency, oevar = observed / expected variance, N = sample size used in testing the given variant, Beta ($FEV_1$) = per-allele change in $FEV_1$ (mL), Beta ($FEV_1/FVC$) = per-allele change in $FEV_1/FVC$ %, SE = standard error

*Gene abbreviations*: *KIF25* = (kinesin family member 25), *ZSWIM7* = (zinc finger, SWIM-type containing 7), *HAL*= (histidine ammonia-lyase*), GPR126*= (G protein-coupled receptor 126), *KCNE2* = (potassium channel, voltage gated subfamily E regulatory beta subunit 2), *GEMNIN5*= gem nuclear organelle associated protein 5, *KIF4B*= kinesin family member 4B

Analyses were performed using data imputed to 1000 Genome phase 1 data and stratified by Hispanic/Latino ancestry group and were adjusted for age, age$^2$, sex, height, height$^2$, smoking status, pack-years, sampling weights, first 5 principal components, Hispanic/Latino ancestry groups, kinship and block unit. Within subgroup estimates were then meta-analyzed. Previously unreported associations are in bold. Retrospectively, association analyses using genetic data imputed to 1000 Genome phase 3 imputation panel were performed in novel loci.

* oevar is provided by MACH software as a measure of imputation quality;

† Results for analyses using genetic data imputed to 1000 Genome phase 3 imputation panel

‡ Genotyped SNP in locus

**Table 3: Loci associated with $FEV_1$ and $FEV_1$/FVC ratio at genome-wide significance in analyses stratified by smoking status**

**Ever Smokers**

| SNP rsID | Trait | Position Build 37 Chr: basepair | Nearest Gene(s) | Effect allele | EAF | Genotyped or imputed SNP | oevar* | N | Beta | SE | P Value |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Indel (no rs#)** | **$FEV_1$** | **11: 86006803** | ***EED CCDC81*** | **GACA** | | **Imputed†** | **1.017** | | **56.95** | **10.09** | **$1.66 \times 10^{-8}$** |
| **rs291231** | **$FEV_1$** | **11: 86007090** | ***EED CCDC81*** | **G** | **0.341** | **Imputed†** | **0.999** | **4595** | **57.28** | **10.19** | **$1.92 \times 10^{-8}$** |
| rs9974878 | $FEV_1$/FVC | 21: 35642446 | *KCNE2* | G | 0.887 | Imputed† | 0.994 | 4594 | 1.28 | 0.23 | $3.58 \times 10^{-8}$ |

**Never Smokers**

| SNP rsID | Trait | | Nearest Gene(s) | Effect allele | EAF | Genotyped or imputed SNP | oevar* | N | Beta | | P Value |
|---|---|---|---|---|---|---|---|---|---|---|---|
| rs116726860 | $FEV_1$/FVC | 2: 155486042 | *KCNJ3 GALNT13* | T | 0.992 | Imputed | 0.897 | 7227 | 3.49 | 0.60 | $6.04 \times 10^{-9}$ |
| **rs7228593** | **$FEV_1$/FVC** | **18: 73575704** | ***SMIM21 LOC339298*** | **G** | **0.948** | **Imputed†** | **0.997** | **7227** | **1.34** | **0.24** | **$3.47 \times 10^{-8}$** |

*Definition of abbreviations*: SNP= single nucleotide polymorphism, $FEV_1$/FVC= ratio of forced expiratory volume in one second over forced vital capacity,

Chr = chromosome, nearest gene = the nearest genes on each side of the intergenic marker, EAF = effect allele frequency, oevar = observed / expected variance, N = sample size used in testing the given variant, Beta ($FEV_1$) = per-allele change in $FEV_1$ (mL), Beta ($FEV_1$/FVC) = per-allele change in $FEV_1$/FVC %, SE = standard error, Indel = Indel mutation

*Gene abbreviations*: *KCNJ3* = potassium voltage-gated channel subfamily J member 3*, GALNT13* = polypeptide N-acetylgalactosaminyltransferase 13, *SMIM21* = small integral membrane protein 21, *EED* = genes embryonic ectoderm development, *KCNE2* = (potassium channel, voltage gated subfamily E regulatory beta subunit 2), *MROH2A* = *(*maestro heat-like repeat family member 2A)

Analyses were stratified by Hispanic/Latino ancestry group and smoking status. Model adjusted for age, age$^2$, sex, height, height$^2$, smoking status, pack-years, sampling weights, first 5 principal components, Hispanic/Latino ancestry groups, kinship and block unit. Within subgroup estimates were then meta-analyzed.  Previously unreported associations are in bold.

* oevar is provided by MACH software as a measure of imputation quality; it is a measure of imputation quality

† Genotyped SNP in locus