Common and rare variants analysis of smoking related traits among current and former smokers of European and African ancestry

Sharon M. Lutz[1], Brittni Frederiksen[2], Ferdouse Begum[3], Michael H. Cho[4,5], Brian Hobbs[4,5], Merry-Lynn McDonald[4], Margaret M. Parker[4], Dawn L. DeMeo[4,5], Marissa Eringher[6], Kendra Young[2], Lai Jiang[3], Marilyn G. Foreman[7], Greg L. Kinney[2], Barry J. Make[8], David A. Lomas[9], Per Bakke[10], Amund Gulsvik[10], James D. Crapo[8], Edwin K. Silverman[4,5], Terri H. Beaty[3], John E. Hokanson[2], the ECLIPSE and COPDGene Investigators

Correspondence to Dr. Sharon M. Lutz, Department of Biostatistics, University of Colorado Anschutz Medical Campus, 13001 E. 17th Place, B119 Bldg. 500, W3128 Aurora, CO 80045 (email: sharon.lutz@ucdenver.edu)

1. Department of Biostatistics, Colorado School of Public Health, University of Colorado, Anschutz Medical Campus, Aurora, Colorado, USA
2. Department of Epidemiology, Colorado School of Public Health, University of Colorado Anschutz Medical Campus, Aurora, Colorado, USA
3. Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, USA
4. Channing Division of Network Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, USA
5. Division of Pulmonary and Critical Care Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, USA
6. Institute of Behavioral Genetics, University of Colorado, Boulder, USA
7. Morehouse School of Medicine, Atlanta, Georgia, USA
8. Department of Medicine, National Jewish Health, Denver, Colorado, USA
9. UCL Respiratory, University College London, London, UK
10. Department of Clinical Science, University of Bergen, Norway

ABSTRACT

Background: Cigarette smoking is a major environmental risk factor for many diseases, and the most important risk factor for chronic obstructive pulmonary disease (COPD). Genetic risk factors for smoking related traits in cohorts of COPD are largely unknown.

Methods: We performed genome-wide association analyses for smoking intensity (average cigarettes/day) across the COPDGene Non-Hispanic White (NHW) (n=6,659) and African American (AA) (n=3,260), GenKOLS (n=1,671), and ECLIPSE (n=1,942) cohorts. In addition, we performed exome array association analyses across the COPDGene Non-Hispanic White (NHW) (n=6,659) and African American (AA) (n=3,260) cohorts. We considered analyses across the entire cohort and stratified by COPD case control status.

Results: We identified genome-wide significant associations for average cigarettes/day in the 15q25 region across all cohorts (lowest $p=1.78x10^{-15}$), except in the COPDGene AA cohort alone. Previously reported associations on chromosome 19 (*RAB4* $p=1.95x10^{-6}$; *CYP2A7* $p=7.50x10^{-5}$; *CYP2B6* $4.04x10^{-4}$) had suggestive and directionally consistent associations. When we examined these signals stratified by COPD case control status, this region on chromosome 15 [*CHRNA5/3*] was nominally associated in both NHW COPD cases (Beta=0.11, $p=5.58x10^{-4}$) and controls (Beta=0.12, $p=3.86x10^{-5}$). However, on chromosome 19, there was a stronger signal with average cigarettes/day in COPD cases than controls for *RAB4* (Beta=-0.13, $p=2.49x10^{-5}$ vs Beta=-0.05, p=0.05) and there was a stronger signal with average cigarettes/day in COPD controls than cases for *CYP2B6* (Beta=-0.03, $p=1.19x10^{-3}$ vs Beta=-0.1, p=0.44). For the gene-based exome array association analysis of rare variants, there were no exome-wide significant associations. For these previously replicated signals, the most significant results were among COPDGene NHW subjects for *CYP2A7* ($p=5.2x10^{-4}$).

Conclusions: In a large genome-wide association study of both common variants and a gene-based association of rare coding variants with quantitative measures of smoking in current and former smokers, we found genome-wide significant associations at the 15q25 region with these smoking traits for common variants, but not for rare variants. These results were directionally consistent among COPD cases and controls.

BACKGROUND

The most recent Surgeon General's report "The Health Consequences of Smoking- 50 Years of Progress" notes tobacco continues to be an immense public health burden in the United States, and tobacco has killed more than 20 million people prematurely since the first Surgeon General's report on smoking in 1964. [1] The prevalence of current cigarette smoking among adults in the U.S. has declined from 42% in 1965 to 18% in 2012; however, more than 42 million Americans still smoke. The findings in this recent Surgeon General's report show the decline in the prevalence of smoking has slowed in recent years, but the burden of mortality related to smoking is expected to remain at high levels for decades to come. [1]

The heritability of nicotine dependence has been estimated at 59% [2]. Genetic studies of smoking quantity have consistently identified the chromosome 15q25.1 region, which includes a cluster of genes coding for cholinergic nicotinic acetylcholine receptor subunits, *CHRNA5-CHRNA3-CHRNB4*. Single nucleotide polymorphisms (SNPs) on chromosome 15q25 have been associated with smoking quantity, among subjects of European ancestry, specifically non-Hispanic whites (NHW) [3–11], and African Americans (AA) [12]. This same 15q25 region has also been associated with spirometric measures of pulmonary function [13] and diseases such as COPD [14] and lung cancer [15] in which smoking may mediate the relationship between these genes and clinically relevant outcomes [16,17]. Recently, rare coding variants in *CHRNA5* were marginally associated with increased risk of nicotine dependence among European Americans (OR=12.9, p=0.01) and in the same risk direction among African Americans (OR=1.5, p=0.37) [18]. In addition to the 15q25.1 region, common variants in several other regions have been associated with smoking related traits such as chromosome 1 [*LPPR5*], chromosome 2 [*TEX41/PABPC1P2*], chromosome 6 [*DNAH8*], chromosome 7 [*PDE1C*], chromosome 8 [*CHRNB3/CHRNA6*], chromosome 9 [*DBH*], chromosome 10 [*LOC1001889*], chromosome 11 [*BDNF, NCAM1*], 19 [*RAB4B, EGLN2, CYP2A7, CYP2B6*], and chromosome 20 [*NOL4L*]. [19]

Given the observed relationship between chromosome 15q25 and COPD and the previous reports of an association between this same region and smoking, we performed a GWAS using smoking intensity

(average cigarettes/day) in the COPDGene study, a large cohort of current and former smokers enriched with COPD cases and a meta-analysis across COPDGene, the ECLIPSE and the GenKOLS studies, which were all enriched for COPD cases. In addition, we performed a gene-based association analysis of rare coding variants with average cigarettes per day among COPDGene NHW and AA. For all of these analyses, we examined the signal of these regions with average cigarettes/day stratified by COPD case control status.

RESULTS

Common Variants: Non-Hispanic Whites and African Americans in the COPDGene Study and a Meta-Analysis of COPDGene, ECLIPSE, and GenKOLS cohorts

Among all NHW COPDGene participants, multiple SNPS at the 15q25 region reached genome-wide significance ($p<5.0e-8$) for cigarettes/day as seen in Table 2. There were no genome-wide significant results for average cigarettes per day among African-Americans in the COPDGene cohort. For the meta-analysis of COPDGene, ECLIPSE, and GenKOLS cohorts, all genome-wide significant results were again in the 15q25 region. Figure 1 includes regional plots for average cigarettes per day for this chromosome 15 region from the meta-analysis, among NHW in the COPDGene study, and among AA in the COPDGene study. Note that there is a similar signal in this region for the meta-analysis and among NHW in the COPDGene study, but there is no genome-wide significant association between any SNPs in this region and average cigarettes per day among AA in the COPDGene study.

Rare Variants: Exome Genotyping in COPDGene Non-Hispanic Whites and African Americans

Using a Bonferroni correction for the number of regions tested in the exome array, we considered a significance level of $3.4x10^{-6}$ in the SKAT-O gene-based tests for association between rare variants and cigarettes/day. No genes reached this significance threshold among NHW in the COPDGene study. Among AA in COPDGene study, only one gene reached this significance threshold [*MYLIP* $p=1.41x10^{-6}$].

However, there were only 7 SNPs in this region and only one SNP rs151199797 (MAF=0.005) was responsible for this signal. When this SNP [rs151199797] was excluded from this region in the rare variant analysis, the *MYLIP* gene was no longer associated with average cigarettes per day among AA (p=0.51). This region on chromosome 6 [MYLIP] including rs151199797 was not replicated among NHW in the COPDGene study (p=0.034).

Comparison to Previously Published GWAS studies

We considered genome-wide significant results from a previously published GWA analysis of smoking related traits in the UK BiLEVE study. [19] Table 2 shows p-values of these loci from the UK BiLEVE analyses in the COPDGene sub-populations analyzed above. Previously reported associations on chromosome 19 (*RAB4* $p=1.95x10^{-6}$; *CYP2A7* $p=7.50x10^{-5}$; *CYP2B6* $4.04x10^{-4}$) had suggestive and directionally consistent associations. In Table 4 for the exome array association analysis of rare variants for these previously replicated signals, the most significant results were among COPDGene NHW subjects for *CYP2A7* ($p=5.2x10^{-4}$).

Associations Stratified by COPD Case Control Status

In table 3 when we examined these signals stratified by COPD case control status, this region on chromosome 15 [*CHRNA5/3*] was nominally associated in both NHW COPD cases ($p=5.58x10^{-4}$) and controls ($p=3.86x10^{-5}$). However, on chromosome 19, there was a stronger signal with average cigarettes/day in COPD cases than controls for *RAB4* (Beta=-0.13 vs Beta=-0.05) and there was a stronger signal with average cigarettes/day in COPD controls than cases for *CYP2B6* (Beta=-0.03 vs Beta=-1.0).

DISCUSSION

In a large genome-wide association study of current and former smokers, we have replicated previous signals showing common variants on chromosome 15q25 are strongly associated with cigarettes per day. This region contains a cluster of genes coding for nicotinic acetylcholine receptors and other genes that are associated with smoking behavior, pulmonary function, as well as diseases directly related to smoking including lung cancer and COPD. [14-18] These findings were replicated in the COPDGene cohort and then confirmed in a meta-analysis of two other European ancestry cohorts enriched in COPD (i.e.,

ECLIPSE and GenKOLS). This signal was consistent among both NHW COPD cases and controls. In addition, the exome array analysis of rare variants in the COPDGene study showed no evidence that rare variants in this region [chromosome 15q25] were associated with cigarettes per day. However, previously reported associations on chromosome 19 (*RAB4* p=1.95E-06; *CYP2A7* p=7.50E-05; *CYP2B6* 4.04E-04) had suggestive and directionally consistent associations among common variants and for rare variants among COPDGene NHW subjects for *CYP2A7* ($p=5.2x10^{-4}$).

Smoking Burden

Among all NHW COPDGene participants, multiple SNPS at the 15q25 region reached genome-wide significance (p<5.0e-8) for cigarettes/day; however, the genetic effect size for the most significant SNP rs8192482 [*CHRNA3*] is modest (Beta=0.13). This modest effect size creates a substantial smoking burden over the lifetime of a smoker. For instance, given that a Non-Hispanic White COPDGene subject smokes for 35 years, each copy of the disease allele for rs8192482 contributes to an average increase of 47 cigarettes per year and an average increase of 1,660 cigarettes over the lifetime of this smoker. If instead the Non-Hispanic White COPDGene subject smokes for 65 years, each copy of the disease allele for rs8192482 contributes to an average increase of 3,084 cigarettes over the lifetime of this smoker.

Differences Among Non-Hispanic White and African American COPDGene Subjects

Within the COPDGene study, on average, Non-Hispanic White subjects smoked more cigarettes per day (mean=25.80, sd=11.42) compared to African American subjects (mean=21.29, sd=10.40). However, there is a larger proportion of current smokers among African American subjects (80%) compared to Non-Hispanic White Subjects (39%). For the most significant signal (rs8192482 [*CHRNA3*]) for cigarettes per day, the allele frequency for the coded T allele is 73% for Non-Hispanic White COPDGene subjects and 12% for African American subjects. There is also a significant difference in sample sizes among the Non-Hispanic White subjects (n=6,659) and African American COPDGene subjects (n=3,260). These factors may have contributed to the differing results between the 2 populations.

Potential Limitations

The COPDGene cohort was ascertained based on smoking history and is enriched for COPD cases. While this ascertainment scheme maximizes the efficiency of case-control studies from the COPDGene cohort, analyzing secondary phenotypes (e.g., quantitative measure of smoking behaviors) in any case-control study can be biased due to the ascertainment scheme. This is only an issue for SNPs associated with both the ascertainment condition and the secondary phenotype. Since our analysis focused on quantitative measures of smoking history (one of the primary ascertainment conditions), when we adjusted for GOLD stage (a measure of severity of COPD) in an additional analysis, we did see a decrease in the strength of the signal (as measured by its p-value) but markers on chromosome 15q25 still gave the most significant signal. This suggests our analysis should be robust against this source of sampling bias due to ascertainment. [20]

While COPDGene included both substantial numbers of AA and NHW subjects, the sample size for AA subjects was considerably smaller and therefore had less statistical power. In addition, cigarettes per day is self-reported and may suffer from measurement error. [17] Biomarkers for smoking history such as carbon monoxide and the nicotine metabolite ratio could provide better measurements of smoking exposure than the reported number of cigarettes per day. [21]

METHODS

COPDGene Study

The Genetic Epidemiology of COPD (COPDGene) Study is a multicenter observational study designed to identify genetic factors associated with COPD and to characterize COPD-related phenotypes. [23] This study recruited 10,192 adult smokers (current and former) who were non-Hispanic whites or African Americans ages 44 to 81 with at least 10 pack-years of smoking history. Table 1 details characteristics of the COPDGene participants included in the genome-wide association analysis. We excluded subjects with severe alpha-1 antitrypsin deficiency or genotyping failure, which resulted in 9,978 unrelated subjects. There were a total of 6,659 non-Hispanic whites (NHW) and 3,260 African Americans (AA) with

complete genotype and phenotype data available. Details of genotyping quality control and imputation have been described previously. [14]

Meta-Analysis Study Populations: ECLIPSE and GenKOLS

The Evaluation of COPD Longitudinally to Identify Predictive Surrogate Endpoints (ECLIPSE) study was a longitudinal, observational study conducted at 46 clinical centers in 12 countries. [24] For this study, we included 1,764 COPD cases and 178 current or ex-smoking controls from the ECLIPSE study; we also included 863 COPD cases and 808 controls from the GenKOLS study in Bergen, Norway. Genotyping methods for the genome-wide SNP panels and study descriptions for the GenKOLS and ECLIPSE cohorts have been described previously. [25-26]

Phenotypes of smoking related traits

Due to non-normality, we categorized average cigarettes smoked per day into 7 categories (e.g., 1 if average cigarettes smoked per day is less than or equal to 10, up to 7 if the average cigarettes per day is greater than 60, and intermediate numbers 2-6 for each increment of 10). We also considered pack-years of cigarettes smoked but due to the similarity and high correlation (r=0.92) with average number of cigarettes smoked per day, we did not present the results here. We also considered age of smoking initiation and current smoking status, but did not find any significant results for these phenotypes. As a result, we did not present these results here.

Genotyping, Quality Control and Imputation for Common Variants

All COPDGene subjects were genotyped using the Illumina HumanOmniExpress by Illumina (San Diego, CA). Details of genotyping quality control have been previously described. [15] Imputation on the COPDGene cohorts was performed using MaCH and minimac. [27,28] Prephasing and imputation were both performed using 30 rounds and 200 states, with regions divided into 1 megabase segments with 500kb flanks. Reference panels for the NHW and AA subjects were the 1000 Genomes Phase I (v3) European (EUR) and cosmopolitan reference panels, respectively. [29] Imputed variants with an R-squared value of < 0.3 were dropped from further analysis. SNPs with MAF less than 1% were excluded.

Further details concerning genotyping, quality control, and imputation are posted on the COPDGene website (http://www.copdgene.org). All SNP locations are based on the NCBI37/hg19 assembly.

Genotyping and Quality Control for Exome Chip Variants

Genotyping with the exome Chip was done for this same COPDGene study cohort on Illumina Human exome array (v1.1 and v1.2). Standard QC protocols were followed to clean called variants on this exome array data as detailed previously. [30]

Statistical Analyses for Common Variants

GWA analyses were performed in PLINK (v1.07) stratified by race. [31] Linear regression analyses of the cigarettes per day were adjusted for age, gender, and genetic ancestry (as summarized by principal components computed within racial group) by including these covariates in the model. [32] In a secondary analysis, we stratified by COPD case-control status (GOLD spirometry stage 2-4 for cases and GOLD spirometry stage 0 for controls). For the primary analysis, we did not adjust for current smoking status due to the correlation with cigarettes per day (-0.19). For Supplemental Tables 1-3, we re-ran all analyses adjusting for current smoking status in addition to age, gender, and genetic ancestry (as summarized by principal components computed within racial group) by including these covariates in the model. The results summarized in Supplemental Tables 1-3 are very similar those shown in Table 2-4.

A fixed effects meta-analysis was performed using METAL (v 2010-08-01) [33] for cigarettes/day adjusting for the same covariates mentioned above (age, gender, and genetic ancestry as summarized by principal components) for the COPDGene, ECLIPSE and GenKOLS cohorts. Genome-wide significant associations were defined by $p<5x10^{-8}$.

Statistical Analyses for Rare Variants

Using data from the exome Array, variants with MAF<5% and minimum minor allele count of 5 were collapsed into gene sets and analyzed using SKAT-O [34] separately for NHW and AA COPDGene subjects. These analyses were restricted to putatively functional variants. The same outcomes and covariates were used for both the rare and common variant analyses with the addition of a covariate for

exome array platform (v1.1 vs v1.2) in the NHW analyses and the inclusion of exome array specific ancestry-based principal components. Among the AA and NHW in COPDGene, 14,559 and 14,677 genes were included in the analysis, respectively, which generated a Bonferroni critical value of $p= 3.4x10^{-6}$ for each cohort.

DISCLOSURES

Regarding conflicts of interest, in the past three years, Edwin K. Silverman received honoraria and consulting fees from Merck, grant support and consulting fees from GlaxoSmithKline, and honoraria from Novartis. David Lomas is a consultant and has received grant support and honoraria from GlaxoSmithKline. He chaired the Respiratory Therapy Area Board at GlaxoSmithKline (2012-15). Amund Gulsvik has participated in the advisory boards of Chesi Pharma AS, Sverige; Novartis, Norge; Takeda Nycomed, Norge, AstraZeneca, Norge and Boehringer Ingelheim, Norge. The funding sources played no role in the design of the study or the decision to submit the manuscript for publication. No other authors reported conflicts of interest.

ACKNOWLEDGEMENTS

Research reported in this publication was supported by the National Heart, Lung, And Blood Institute of the National Institutes of Health under Award Number K01HL125858. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. This work was also supported by National Heart, Lung and Blood Institute NHLBI P01 HL105339 and R01 HL089856 (E.K.S.); K08 HL097029 and R01 HL113264 (M.H.C.), and R01 HL089897 (J.D.C.), and a grant from the Alpha-1 Foundation (M.H.C.). The COPDGene study (NCT00608764) is also supported by the COPD Foundation through contributions made to an Industry Advisory Board comprised of AstraZeneca, Boehringer Ingelheim, Novartis, Pfizer, GlaxoSmithKline, Siemens, and Sunovion. The Norway GenKOLS study (Genetics of Chronic Obstructive Lung Disease, GSK code RES11080), and the ECLIPSE study (NCT00292552; GSK code SCO104960) were funded by GlaxoSmithKline. The funding sources played no role in the design of the study or the decision to submit the manuscript for publication. A full list of the COPDGene and ECLIPSE investigators are given in the supplement.

REFERENCES

1. Health consequences of smoking- 50 years of progress: A report of the surgeon general. http://www.surgeongeneral.gov/library/reports/50-years-of-progress/exec-summary.pdf

2. Li MD, Cheng R, Ma JZ, Swan GE (2003) A meta-analysis of estimated genetic and environmental effects on smoking behavior in male and female adult twins. Addiction 98: 23–31. doi:10.1046/j.1360-0443.2003.00295.x.

3. Liu JZ, Tozzi F, Waterworth DM, Pillai SG, Muglia P, et al. (2010) Meta-analysis and imputation refines the association of 15q25 with smoking quantity. Nat Genet 42: 436–440. doi:10.1038/ng.572.

4. Thorgeirsson TE, Gudbjartsson DF, Surakka I, Vink JM, Amin N, et al. (2010) Sequence variants at CHRNB3-CHRNA6 and CYP2A6 affect smoking behavior. Nat Genet 42: 448–453. doi:10.1038/ng.573.

5. Genome-wide meta-analyses identify multiple loci associated with smoking behavior (2010). Nat Genet 42: 441–447. doi:10.1038/ng.571.

6. Saccone NL, Culverhouse RC, Schwantes-An T-H, Cannon DS, Chen X, et al. (2010) Multiple independent loci at chromosome 15q25.1 affect smoking quantity: a meta-analysis and comparison with lung cancer and COPD. PLoS Genet 6. Available: http://www.ncbi.nlm.nih.gov/pubmed/20700436. Accessed 5 September 2012.

7. Wang JC, Grucza R, Cruchaga C, Hinrichs AL, Bertelsen S, et al. (2008) Genetic variation in the CHRNA5 gene affects mRNA levels and is associated with risk for alcohol dependence. Mol Psychiatry 14: 501–510.

8. Bierut LJ, Fox L, Horton WJ, Breslau N, Budde J, et al. (2008) Variants in nicotinic receptors and risk for nicotine dependence. The American journal of psychiatry 165: 1163–1171. doi:10.1176/appi.ajp.2008.07111711.

9. Wang JC, Cruchaga C, Saccone NL, Bertelsen S, Liu P, et al. (2009) Risk for nicotine dependence and lung cancer is conferred by mRNA expression levels and amino acid change in CHRNA5. Human Molecular Genetics 18: 3125–3135. doi:10.1093/hmg/ddp231.

10. Wang JC, Bierut LJ, Goate AM (2009) Variants Weakly Correlated with CHRNA5 D398N Polymorphism Should be Considered in Transcriptional Deregulation at the 15q25 Locus Associated with Lung Cancer Risk. Clinical Cancer Research 15: 5599–5599. doi:10.1158/1078-0432.CCR-09-1108.

11. Falvella FS, Galvan A, Frullanti E, Spinola M, Calabrò E, et al. (2009) Transcription Deregulation at the 15q25 Locus in Association with Lung Adenocarcinoma Risk. Clinical Cancer Research 15: 1837–1842. doi:10.1158/1078-0432.CCR-08-2107.

12. David SP, Hamidovic A, Chen GK, Bergen AW, Wessel J, et al. (2012) Genome-wide meta-analyses of smoking behaviors in African Americans. Transl Psychiatry 2: e119.

13. Lutz SM, Cho MH, Young K, Hersh CP, Castaldi P, McDonald ML, Regan E, Mattheisen M, DeMeo DL, Parker M, Foreman M, Make BJ, Jensen RL, Casaburi R, Lomas DA, Bhatt SP, Bakke P, Gulsvik A, Crapo JD, Beaty TH, Laird N, Lange C, Hokanson JE, Silverman EK, ECLIPSE Investigators, and COPDGene Investigators. (2015) A Genome-Wide Association Study Identifies Risk Loci for Spirometric Measures among Smokers of European and African Ancestry. BMC Genetics

14. Cho MH, McDonald MN, 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 ICGN, ECLIPSE, and COPDGene Investigators. (2014) Risk Loci for Chronic Obstructive Pulmonary Disease: A Genome-Wide Association Study and Meta-Analysis. Lancet Respir Med. 2:214-225.

15. Chen LS, Hung RJ, Baker T, Horton A, Culverhouse R, Saccone N, Cheng I, Deng B, Han Y, Hansen NM, Horsman J, Kim C, Lutz S, Rosenberger A, Aben KK, Andrew AS, Breslau N, Chang SC, Dieffenbach AK, Dienemann H, Frederiksen B, Han J, Hatsukami DK, Johnson EO, Pande M, Wrensch MR, McLaughlin J, Skaug V, van der Heijden HF, Wampfler J, Wenzlaff A, Woll P, Zienolddiny S, Bickeböller H, Brenner H, Duell EJ, Haugen A, Heinrich J, Hokanson JE, Hunter DJ, Kiemeney LA,Lazarus P, Le Marchand L, Liu G, Mayordomo J, Risch A, Schwartz AG, Teare D, Wu X, Wiencke JK, Yang P, Zhang ZF, Spitz MR, Kraft P, Amos CI, Bierut LJ (2015) CHRNA5 Risk Variant Predicts Delayed Smoking Cessation and Earlier Lung Cancer Diagnosis - A Meta-analysis. JNCI

16. Lutz SM, Hokanson JE. (2014) Genetic Influences of Smoking and Clinical Disease: Understanding Behavioral and Biological Pathways with Mediation Analyses. Annals of Thoracic Medicine. 11(7):1082-1083.

17. Lutz SM, Hokanson JE. (2015) Mediation Analysis in Genome-Wide Association Studies: Current Perspectives. Open Journal Bioinformatics.

18. Olfson E, Saccone NL, Johnson EO, Chen L-S, Culverhouse R, Doheny K , Foltz SM, Fox L, Gogarten SM, Hartz S, Hetrick K, Laurie CC, Marosy B, Amin N, Arnett D, Barr RG, Bartz TM, Bertelsen S, Borecki IB, Brown MR, Chasman DI, van Duijn CM, Feitosa MF, Fox ER, Franceschini N, Franco OH, Grove ML, Guo X, Hofman A, Kardia SLR, Morrison AC, Musani SK, Psaty BM, Rao DC, Reiner AP, Rice K, Ridker PM, Rose LM, Schick UM, Schwander K, Uitterlinden AG, Vojinovic D, Wang J-C, Ware EB, Wilson G, Yao J, Zhao W, Breslau N, Hatsukami D, Stitzel JA, Rice J, Goate A, Bierut LJ. (2016) Rare, low frequency and common coding variants in CHRNA5 and their contribution to nicotine dependence in European and African Americans. Molecular Psychiatry (2016) 21, 601–607.

19. Wain LV, Shrine N, Miller S, et al. (2015) Novel insights into the genetics of smoking behaviour, lung function, and chronic obstructive pulmonary disease (UK BiLEVE): a genetic association study in UK Biobank. Lancet Respir.

20. Lutz SM, Hokanson JE, Lange C. (2014) An alternative hypothesis testing strategy for secondary phenotype data in case-control genetic association studies. Frontiers in Genetics.

21. Bloom AJ, Hartz SM, Baker TB, Chen LS, Piper ME, Fox L, Martinez M, Hatsukami D, Johnson EO, Laurie CC, et al. (2014) Beyond cigarettes-per- day: a genome-wide association study of the biomarker carbon monoxide. Ann Am Thorac Soc. 11:1003–1010.

22. Wang J, Spitz MR, Amos CI, Wu X, Wetter DW, Cinciripini PM, Shete S. (2012) Method for evaluating multiple mediators: mediating effects of smoking and COPD on the association between the CHRNA5-A3 variant and lung cancer risk. PLoS ONE. 7: e47705.

23. Regan EA, Hokanson JE, Murphy JR, et al. (2010) Genetic epidemiology of COPD (COPDGene) study design. COPD. 7:32-43.

24. Vestbo J, Anderson W, Coxson HO, et al. (2008) Evaluation of COPD Longitudinally to Identify Predictive Surrogate End-points (ECLIPSE). Eur Respir J. 31:869-873.

25. Pillai SG, Ge D, Zhu G, et al. ICGN Investigators. (2009) A genome-wide association study in chronic obstructive pulmonary disease (COPD): Identification of two major susceptibility loci. PLoS Genet. 5(3):e1000421.

26. Cho MH, Boutaoui N, Klanderman BJ, et al. (2010) Variants in FAM13A are associated with chronic obstructive pulmonary disease. Nature Genetics. 42(3):200-201.

27. Howie B, Fuchsberger C, Stephens M, et al. (2012) Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. Nature Genetics. 44:955–9.

28. Li Y, Willer CJ, Ding J, et al. (2010) MaCH: sing sequence and genotype data to estimate haplotypes and unobserved genotypes. Genetic Epidemiology. 834:816–834.

29. Abecasis GR, Auton A, Brooks LD, et al. (2010) An integrated map of genetic variation from 1,092 human genomes. Nature. 491:56–65.

30. Hobbs BD, Parker MM, Chen H, et al. (2016) Exome Array Analysis Identifies a Common Variant in IL27 Associated with Chronic Obstructive Pulmonary Disease. Am J Respir Crit Care Med. 194(1):48-57. doi: 10.1164/rccm.201510-2053OC.

31. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, et al. (2007) PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. The American Journal of Human Genetics 81: 559–575. doi:10.1086/519795.

32. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, et al. (2006) Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet 38: 904–909. doi:10.1038/ng1847.

33. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics. 2010;26:2190–1.

34. Lee S, Emond MJ, Bamshad MJ, Barnes KC, et al. (2012) Optimal Unified Approach for Rare-Variant Association Testing with Application to Small-Sample Case-Control Whole-Exome Sequencing Studies. Am J Hum Genet. 91(2): 224–237.

TABLES

Table 1. Characteristics of COPDGene, ECLIPSE, and GenKOLS subjects included in genome-wide association analysis. For continuous variables, the mean is given first followed by the standard deviation.

	COPDGene NHW	COPDGene AA	ECLIPSE	GenKOLS
Sample size (#COPD cases)	6,658 (2,812)	3,280 (821)	1,942 (1,764)	1,671 (863)
Gender (%male)	47.73%	44.79%	66.17%	58.71%
Age (years)	62.0 (8.9)	54.6 (7.2)	63.1 (7.6)	60.7 (11.1)
Pack-years	47.2 (26.0)	38.3 (21.6)	48.6 (27.7)	26.0 (17.4)
Average cigarettes per day	25.8 (11.4)	21.3 (10.4)	25.3 (12.7)	14.7 (7.6)
Smoking initiation age	16.9 (4.3)	17.0 (5.3)	16.9 (4.4)	18.5 (5.0)
Current Smoking (%)	38.9%	80.0%	34.8%	43.8%

Table 2 Genome-wide significant results for average cigarettes per day and comparision of previously replicated regions from the UK BiLEVE study [19]. The "Coded Allele" column shows alleles for COPDGene NHW cohort, COPDGene AA cohort, the meta-analysis of COPDGene NHW, COPDGene AA, ECLIPSE and GenKOLS. Correcting for the 12 regions listed below, all p-values less than 4.17e-3 (0.05/12) are in bold.

SNP	Position (bp)	CHR	Nearest Gene	Coded Allele	COPDGene NHW (n=6,658)			COPDGene AA (n=3,260)			Meta-Analysis of COPDGene NHW, COPDGene AA, ECLIPSE and GenKOLS (n=13,551)		
					Allele Freq	Beta	P	Allele Freq	Beta	P	Allele Freq	Beta	P
rs8192482	78886198	15	*CHRNA3*	T/T/T	0.73	0.13	**5.37E-11**	0.12	0.05	0.41	0.37	0.12	**1.78E-15**
rs11633958	78866445	15	*CHRNA5*	T/T/T	0.73	0.13	**9.41E-11**	0.13	0.04	0.46	0.37	0.12	**2.18E-15**
rs72738786	78828086	15	*AGPHD1*	T/T/T	0.74	0.13	**3.41E-11**	0.28	0.02	0.61	0.37	0.12	**5.78E-15**
rs2869548	78922638	15	*CHRNB4*	A/A/A	0.76	0.12	**2.17E-10**	0.13	0.05	0.31	0.39	0.12	**2.82E-14**
rs11858836	78783277	15	*IREB2*	A/A/A	0.72	0.11	**1.42E-08**	0.13	0.01	0.83	0.36	0.10	**1.26E-10**
rs61784651	99445471	1	*LPPR5*	T/T/T	0.31	-0.03	0.27	0.12	-0.08	0.15	0.16	0.02	0.32
rs10193706	146316319	2	*TEX41/ PABPC1P*	A/C/A	0.93	0.01	0.46	0.38	0.01	0.96	0.51	0.01	0.4
rs10807199	38901867	6	*DNAH8*	T/T/T	0.93	0.01	0.86	0.3	0.02	0.65	0.47	-0.01	0.56
rs215605	32336965	7	*PDE1C*	G/T/T	0.76	0.02	0.21	0.56	0.01	0.83	0.37	-0.01	0.43
rs13280604	42559586	8	*CHRNB3*	G/A/A	0.44	0.04	0.09	0.55	-0.06	0.03	0.78	-0.02	0.24
rs3025343	136478355	9	*DBH*	A/A/A	0.24	-0.02	0.52	0.05	0.07	0.40	0.12	0.001	0.99
rs1329650	93348120	10	*LOC100188947*	T/T/T	0.55	-0.01	0.58	0.21	0.02	0.63	0.56	-0.03	0.12
rs6265	27679916	11	*BDNF*	T/T/T	0.37	-0.02	0.48	0.09	-0.02	0.75	0.82	-0.02	0.39
rs4466874	112861434	11	*NCAM1*	C/C/T	0.82	-0.06	**9.34E-04**	0.82	0.05	0.06	0.58	0.03	0.02
rs7937	41302706	19	*RAB4B*	C/T/T	0.85	-0.09	**1.95E-06**	0.61	0.03	0.34	0.42	0.07	**1.05E-05**
rs3733829	41310571	19	*EGLN2*	G/G/A	0.74	0.06	**3.86E-03**	0.16	0.03	0.49	0.47	-0.04	0.01
rs12461383	41370338	19	*CYP2A7*	C/G/C	0.93	-0.09	**2.47E-04**	0.45	0.04	0.26	0.47	-0.09	**7.50E-05**
rs7260329	41521638	19	*CYP2B6*	A/A/A	0.62	-0.07	**4.04E-04**	0.3	-0.04	0.23	0.68	-0.05	**8.20E-04**
rs4911243	31162568	20	*NOL4L*	A/G/A	0.67	0.04	0.03	0.34	-0.03	0.46	0.35	0.02	0.18

Table 3. Below are the SNPs from Table 2 with the analysis stratified by COPD case control status among COPDGene NHW and AA. Correcting for the 12 regions listed below, all p-values less than 4.17e-3 (0.05/12) are in bold.

SNP	CHR	Nearest Gene	Coded Allele	COPDGene NHW COPD Cases (n=2,819)			COPDGene NHW COPD Controls (n=2,534)			COPDGene AA COPD Cases (n=821)			COPDGene AA COPD Controls (n=1,749)		
				Allele Freq	Beta	P	Allele Freq	Beta	P	Allele Freq	Beta	P	Allele Freq	Beta	P
rs8192482	15	*CHRNA3*	T/T/T/T	0.78	0.11	**5.58E-4**	0.68	0.12	**3.86E-5**	0.12	0.08	0.43	0.11	0.05	0.53
rs11633958	15	*CHRNA5*	T/T/T/T	0.78	0.11	**5.06E-4**	0.68	0.12	**5.34E-5**	0.13	0.13	0.21	0.12	0.01	0.91
rs72738786	15	*AGPHD1*	T/T/T/T	0.79	0.11	**3.33E-4**	0.69	0.12	**5.81E-5**	0.28	0.07	0.35	0.27	0.01	0.82
rs2869548	15	*CHRNB4*	A/A/A/A	0.81	0.10	**7.94E-4**	0.71	0.11	**2.81E-4**	0.13	0.07	0.49	0.12	0.07	0.35
rs11858836	15	*IREB2*	A/A/A/A	0.76	0.08	**7.17E-4**	0.67	0.11	**2.36E-4**	0.14	0.09	0.38	0.13	-0.02	0.77
rs61784651	1	*LPPR5*	T/T/T/T	0.3	-0.01	0.90	0.32	-0.05	0.24	0.11	-0.16	0.17	0.12	-0.06	0.45
rs10193706	2	*TEX41/ PABPC1P2*	A/A/C/C	0.91	0.01	0.83	0.95	0.04	0.12	0.39	-0.02	0.8	0.36	0.01	0.80
rs10807199	6	*DNAH8*	T/T/T/T	0.96	0.02	0.43	0.9	-0.04	0.14	0.3	-0.01	0.87	0.3	0.02	0.64
rs215605	7	*PDE1C*	G/G/T/T	0.77	0.02	0.57	0.75	0.02	0.56	0.55	0.03	0.62	0.56	0.05	0.24
rs13280604	8	*CHRNB3*	G/G/A/A	0.42	0.05	0.15	0.45	0.02	0.53	0.54	-0.08	0.18	0.56	-0.1	0.01
rs3025343	9	*DBH*	A/A/A/A	0.24	-0.02	0.64	0.24	0.01	0.82	0.06	0.06	0.72	0.04	0.14	0.25
rs1329650	10	*LOC100188947*	T/T/T/T	0.56	-0.01	0.79	0.55	-0.03	0.39	0.22	0.01	0.98	0.2	0.02	0.67
rs6265	11	*BDNF*	T/T/T/T	0.37	-0.02	0.64	0.38	-0.01	0.76	0.08	-0.08	0.55	0.09	-0.05	0.53
rs4466874	11	*NCAM1*	C/C/C/C	0.82	-0.07	0.02	0.8	-0.07	0.02	0.78	0.11	0.03	0.84	0.04	0.30
rs7937	19	*RAB4B*	C/C/T/T	0.8	-0.13	**2.49E-05**	0.89	-0.05	0.05	0.63	0.01	0.92	0.61	0.03	0.48
rs3733829	19	*EGLN2*	G/G/G/G	0.77	0.08	0.02	0.71	0.06	0.04	0.18	0.03	0.77	0.15	0.08	0.24
rs12461383	19	*CYP2A7*	C/C/G/G	0.91	-0.1	9.63E-03	0.96	-0.06	0.09	0.47	0.07	0.26	0.46	0.02	0.59
rs7260329	19	*CYP2B6*	A/A/A/A	0.6	-0.03	0.44	0.65	-0.1	**1.19E-03**	0.31	-0.07	0.35	0.31	-0.03	0.47
rs4911243	20	*NOL4L*	A/A/G/G	0.68	0.04	0.24	0.67	0.05	0.12	0.32	0.02	0.75	0.35	-0.08	0.08

Table 4. Using the genes from Table 2, below are p-values for the associations of the rare variants in each gene with average cigarettes per day. Correcting for the 12 regions listed below, all p-values less than 4.17e-3 (0.05/12) are in bold.

CHR	Nearest Gene	COPDGene NHW			COPDGene AA		
		All (n=6,658)	Cases (n=2,819)	Controls (n=2,534)	All (n=3,260)	Cases (n=821)	Controls (n=1,749)
15	*CHRNA3*	0.17	0.24	0.41	0.52	0.52	0.49
15	*CHRNA5*	0.06	0.15	1.00	0.30	0.33	0.45
15	*AGPHD1*	0.27	0.01	0.28	0.02	0.43	0.05
15	*CHRNB4*	0.87	0.82	0.84	0.87	0.19	0.50
15	*IREB2*	0.27	0.77	0.82	0.02	NA	0.15
1	*LPPR5*	0.83	0.76	0.81	0.19	0.88	0.11
2	*TEX41/ PABPC1P2*	0.56	0.85	0.22	0.64	1.00	0.33
6	*DNAH8*	0.87	0.35	0.87	0.38	0.42	0.51
7	*PDE1C*	1.00	0.05	0.14	0.60	0.32	0.74
8	*CHRNB3*	0.64	1.00	0.65	0.86	NA	1.00
9	*DBH*	0.52	0.64	0.50	0.70	1.00	0.45
10	*LOC100188947*	0.87	0.79	0.71	0.11	0.51	0.69
11	*BDNF*	1.00	0.76	0.46	0.26	0.48	0.56
11	*NCAM1*	0.76	0.30	0.82	0.03	0.27	0.28
19	*RAB4B-EGLN2*	0.44	0.69	0.57	0.04	0.15	0.27
19	*CYP2A7*	**5.2E-4**	0.05	0.08	0.47	0.51	0.97
19	*CYP2B6*	0.56	0.45	0.03	0.48	0.51	0.21
20	*NOL4L*	0.32	0.55	0.89	0.84	NA	NA

FIGURES

Figure 1: Below are the regional plots for average cigarettes per day for chromosome 15 from the meta-analysis, among NHW in the COPDGene study, and among AA in the COPDGene study. Note that there is a similar signal in the region for the meta-analysis and among NHW in the COPDGene study, but there is no association between any SNPs in this region and average cigarettes per day among AA in the COPDGene study.