**A Systematic Analysis of Protein-altering Exonic Variants in Chronic Obstructive Pulmonary Disease**

Matthew Moll1,2, Victoria E Jackson3,4,5, Bing Yu6, Megan L Grove6, Stephanie J London7, Sina A Gharib8, Traci M Bartz9,10, Colleen M Sitlani10, Josée Dupuis11, George T O’Connor12, Hanfei Xu11, Patricia A Cassano13,14, Bonnie Kaufmann Patchen13, Woo Jin Kim15, Jinkyeong Park1,16, Kun Hee Kim17, Buhm Han18, R Graham Barr19, Ani Manichaikul20, Jennifer N Nguyen20, Stephen S Rich20, Lies Lahousse21,22, Natalie Terzikhan21, Guy Brusselle21, Phuwanat Sakornsakolpat23, Jiangyuan Liu1, Christopher J Benway1, Ian P Hall24, Martin D Tobin3,25, Louise V Wain3,25, Edwin K Silverman1,26, Michael H Cho\*1,2,26, Brian D Hobbs\*1,2,26

\*Jointly supervised and co-corresponding authors

Affiliations:

1 Channing Division of Network Medicine, Department of Medicine, Brigham and Women’s Hospital, Boston, MA 02115, USA

2 Division of Pulmonary and Critical Care Medicine, Department of Medicine, Brigham and Women’s Hospital, Boston, MA 02115, USA

3 Department of Health Sciences, University of Leicester, Leicester, LE1 7RH, UK.

4 Population Health and Immunity Division, The Walter and Eliza Hall Institute of Medical Research, Parkville, Victoria, Australia.

5 Department of Medical Biology, University of Melbourne, Parkville, Victoria, Australia.

6 School of Public Health, University of Texas Health Science Center, Houston, TX 77030, USA

7 Epidemiology Branch, National Institute of Environmental Health Sciences, National Institutes of Health, Department of Health and Human Services Research, Research Triangle Park, North Carolina, USA

8 Center for Lung Biology, Division of Pulmonary & Critical Care Medicine, Dept of Medicine, University of Washington, Seattle, WA, USA

9 Department of Biostatistics, University of Washington, Seattle, WA, USA

10 Cardiovascular Health Research Unit, Department of Medicine, University of Washington, Seattle, WA, USA

11 Department of Biostatistics, Boston University School of Public Health, Boston, MA, USA

12 Pulmonary Center, Boston University School of Medicine, Division of Pulmonary, Allergy, Sleep, and Critical Care Medicine, Department of Medicine, Boston Medical Center, Boston, Massachusetts.

13 Division of Nutritional Sciences, Cornell University, Ithaca, NY 14853, USA.

14 Division of Epidemiology, Department of Population Health Sciences, Weill Cornell Medicine, NY, NY 10075 USA

15 Department of Internal Medicine, Kangwon National University, Chuncheon, South Korea.

16 Department of Internal Medicine, Dongguk University Ilsan Hospital, Goyang-Si, Gyeonggi-do, South Korea.

17 Department of Convergence Medicine and Asan Medical Center, University of Ulsan College of Medicine, Seoul, Korea.

18 Department of Medicine, Seoul National University College of Medicine, Seoul, South Korea.

19 Department of Medicine, Columbia University Medical Center, New York, NY, USA

20 Center for Public Health Genomics, University of Virginia, Charlottesville, VA, 22908, USA.

21 Department of Epidemiology, Erasmus University Medical Center, Rotterdam, 3000, CA, The Netherlands.

22 Department of Bioanalysis, FFW, Ghent University, Ghent, 9000, Belgium.

23 Department of Medicine, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok, Thailand

24 NIHR Nottingham Biomedical Research Centre, Queen’s Medical Centre, Nottingham NG7 2UH, UK

25 National Institute for Health Research Leicester Respiratory Biomedical Research Centre, Glenfield Hospital, Leicester, UK

26 Harvard Medical School, Boston, MA USA

Correspondence:

Brian D. Hobbs, MD, MMSc. [rebdh@channing.harvard.edu](mailto:rebdh@channing.harvard.edu) and

Michael H. Cho, MD, MPH. [remhc@channing.harvard.edu](mailto:remhc@channing.harvard.edu)

Channing Division of Network Medicine

181 Longwood Ave 4th floor

Boston, MA 02115

Funding:

MM is supported by T32HL007427.

BDH is supported by NIH K08HL136928 and R01 HL089856.

MHC is supported by NIH R01HL137927 and R01HL135142.

EKS is supported by NIH R01 HL137927, R01 HL147148, R01 HL133135, and P01 HL114501

M.D.T. is supported by a Wellcome Trust Investigator Award (WT202849/Z/16/Z). The research was partially supported by the NIHR Leicester Biomedical Research Centre; the views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health.

LVW holds a GSK / British Lung Foundation Chair in Respiratory Research (C17-1).

IPH holds an NIHR Senior investigator Award.

SJL is supported by the Intramural Research Program of the NIH, National Institute of Environmental Health Sciences. ZO1 ES43012.

**Competing Interests:**

EKS received grant support from GlaxoSmithKline and Bayer. MHC has received grant support from GlaxoSmithKline and Bayer, consulting fees from Genentech and AstraZeneca, and speaking fees from Illumina. MDT receives grant support from GlaxoSmithKline and Orion. LVW received grant support from GlaxoSmithKline and Orion. IPH has received grant support from Glaxo Smith Kline and Orion. The other authors declared no conflicts of interest.

**Author Contributions:**

Study Design: MM, VEJ, BDH, MHC

Acquisition, analysis, or interpretation of the data: MM, VEJ, LVW, MDT, BDH, MHC, EKS

Critical revision of the manuscript for important intellectual content: All authors

Statistical analysis: MM, BDH, MHC, HX, TMB, SAG, CMS, JD

Obtained funding: MHC, BDH, EKS

Word count

Abstract: 248

Manuscript: 4,238

**Supplementary appendix is available at:**

https://doi.org/10.6084/m9.figshare.14538222.v1

**Abstract**

Genome-wide association studies (GWASs) have identified regions associated with chronic obstructive pulmonary disease (COPD). GWASs of other diseases have shown an approximately 10-fold overrepresentation of nonsynonymous variants, despite limited exonic coverage on genotyping arrays. We hypothesized that a large-scale analysis of coding variants could discover novel genetic associations with COPD, including rare variants with large effect sizes. We performed a meta-analysis of exome arrays from 218,399 controls and 33,851 moderate-to-severe COPD cases. All exome-wide significant associations were present in regions previously identified by GWAS. We did not identify any novel rare coding variants with large effect sizes. Within GWAS regions on chromosomes 5q, 6p, and 15q, four coding variants were conditionally significant (p < 0.00015) when adjusting for lead GWAS SNPs. A common *GSDMB* splice variant (rs11078928) previously associated with decreased risk for asthma, was nominally associated with decreased risk for COPD (MAF = 0.46, p=1.8e-4). Two stop variants in *CCHCR1*, a gene involved in regulating cell proliferation, were associated with COPD (both p < 0.0001). The *SERPINA1* Z allele was associated with a random effects odds ratio of 1.43 for COPD (95% CI: 1.17-1.74), though with marked heterogeneity across studies. Overall, COPD-associated exonic variants were identified in genes involved in DNA methylation, cell-matrix interactions, cell proliferation, and cell death. In conclusion, we performed the largest exome array meta-analysis of COPD to date and identified potential functional coding variants. Future studies are needed to identify rarer variants, and further define the role of coding variants in COPD pathogenesis.

**Introduction**

Chronic obstructive pulmonary disease (COPD) is characterized by irreversible airflow limitation (1), and is one of the leading causes of morbidity and mortality worldwide (2). Genome-wide association studies (GWASs) have lent considerable insight into the genetic risk to COPD (3–7). Most GWAS variants are non-coding and are thought to affect COPD susceptibility through gene regulation (8). As such, identifying disease-causing variants in COPD GWAS regions remains challenging. While coding regions make up only about 1% of the genome, ~10% of GWAS signals in complex diseases are attributable to nonsynonymous variants(8). Specific rare coding variants may confer a particularly high risk for complex diseases such as COPD. For example, alpha-1 antitrypsin deficiency (AATD) is associated with a ~15-fold increased odds for emphysema (9) and is most commonly caused by homozygosity for the *SERPINA1* Z allele (rs28929474), which is a missense variant found in ~2-3% of the United States population (10). Exome sequencing of a French-Canadian family with early-onset emphysema identified a rare non-synonymous causal variant in protein tyrosine phosphatase non-receptor type 6 (*PTPN6*)(11). Germline mutations in telomerase genes have been observed in severe COPD cases(12). Thus, examining coding variants across the genome may identify important functional (protein-altering) variants associated with COPD.

Exome arrays were designed to allow genotyping of a large fraction of functional (nonsynonymous, splice, stop-gain) variants across the genome (13), and association analyses have been reported for COPD and lung function(14–16). However, several important questions regarding the utility of exome array studies in COPD remain unanswered. It is not known whether increasing sample size and power will identify novel rare coding variants that markedly increase COPD risk. There have since been several large-scale GWASs for lung function and COPD, yet GWASs have poor coverage of exonic variants and are not intended to identify rare coding variants; exome array results have not been directly compared to GWAS results, which may elucidate the functional variants being tagged by GWAS-identified SNPs. While the *SERPINA1* Z allele (rs28929474) is a known risk factor for COPD - even in heterozygous individuals (17) - the largest GWASs to date (3,4,6,7) did not identify an association of the *SERPINA1* Z allele with COPD or lung function; one reason for this result may be the smoking dependent effects of the Z allele and/or imputation inaccuracies, as the Z allele is not present on most genotyping arrays. However, the Z allele is present on exome arrays, allowing for direct assessment of the association of the Z allele with COPD risk. Further, many COPD case-control studies intentionally exclude ZZ individuals, which could introduce selection bias. A gene-by-environment interaction of cigarette smoking may be important for *SERPINA1* variants to contribute to COPD (18), which may affect the association of the Z allele on COPD in population-based cohorts (as opposed to COPD cohorts enriched for cigarette smoking).

We hypothesized that a larger exome array meta-analysis would provide increased power to detect rare and poorly imputed functional exonic variants associated with COPD and identify the most likely causal variants in previously-defined COPD GWAS regions. We also leveraged the exome array data to assess effect size heterogeneity of the Z allele across studies.

**Methods**

Study Cohorts

We included 12 cohorts in our analysis: ARIC (Atherosclerosis Risk in Communities) Study with African ancestry [Aa] and European ancestry [Ea] participants (19), CHS (Cardiovascular Health Study) including Ea and Aa participants(20), COPDGene (Genetic Epidemiology of COPD) with non-Hispanic white [NHW] and African-American [AA] participants (21), EOCOPD/ICGN (Boston Early-onset COPD Study (22,23) and International COPD Genetics Network (24)) studies, the FHS (Framingham Heart Study) (25–27), HABC (Health, Aging, and Body Composition) Study with Ea and Aa participants(28), KARE (Korean Association Resource) Study (29), MESA (Multi-Ethnic Study of Atherosclerosis) including non-Hispanic African-American, Chinese-American, Hispanic, and non-Hispanic White subpopulations (30,31), RS (Rotterdam Study) (32,33), TCGS (Transcontinental COPD Genetics Study) from Poland and South Korea (34), UK COPD Exome Chip Consortium (UKECC)(16), and UK Biobank (35). Moderate-to-severe COPD was primarily defined by pre-bronchodilator FEV1/FVC ratio < 0.7 and FEV1 < 80% predicted; post-bronchodilator measures were only performed in a minority of studies and were used when available. Individual study details, including genotyping methods, are available in the Supplementary Materials.

Statistical Analyses

Genetic association analysis was performed for case-control moderate-to-severe COPD status using an additive genetic model adjusted for age, sex, cigarette smoking pack-years, and principal components of genetic ancestry. In the family-based studies including FHS and EOCOPD/ICGN, we utilized logistic regression with generalized estimating equations to adjust for familial clustering. Quality control on summary statistics from all cohorts was performed with EasyQC (36) to assure common variant names and reference strand across cohorts and minor allele count (MAC) > 10 within each cohort. In addition to these exome array results, we also included the subset of matching variants (MAC > 4) in a case-control association analysis of UK Biobank (4); models were adjusted for age, sex, pack-years of smoking, ever smoking (when available), and principal components of genetic ancestry.

Power calculations were performed using the genome association study (GAS) calculator available at <http://csg.sph.umich.edu/abecasis/cats/gas_power_calculator/index.html> (37), based on a COPD prevalence of 0.10 (38,39). To account for relatedness amongst individuals, total effective sample size was calculated as previously described (40) and used in power calculations to approximate the number of independent individuals represented by our sample. The total effective sample size was 53,117.

We performed an inverse variance fixed effects meta-analysis of exome array results with METAL (41), and limited our analysis to putative functional (non-synonymous, stop, and splice) variants, which were annotated using wANNOVAR (42). Variants were considered for analysis if they were present in the UK Biobank and at least half of the other cohorts. Exome-wide significance was determined using Bonferroni adjustment (p < 0.05/20,536 variants < 2.4e-6). Replication of signals from previously reported exome array studies (14–16) was defined as a consistent direction of effect and exome-wide statistical significance. Plink v1.9(43) --clump was used to choose a single “index” variant from all variants with R2≥0.2 in each significantly associated genetic locus. We also performed a targeted analysis of splice and stop variants, considering p-values below a Bonferroni-adjusted threshold (p < 0.05/number-of-stop/splice-variants < 0.05/257 < 0.00019) to be nominally significant. We examined the relative association of exome-wide significant COPD variants with the spirometric parameters FEV1 and FEV1/FVC in the GWAS results from Shrine et al. (3). As genetic effects may vary with age, we examined whether age modifies the effect of exome-wide significant variants in the UK Biobank. Additionally, we evaluated the associated of alleles with age of COPD diagnosis in the COPDGene study. To examine the differential effects of associated variants based on smoking exposure, we performed stratified analyses in ever- versus never-smokers and heavy- (>20 pack-years) versus light- (≤ 20 pack-years) smokers in UK Biobank, and compared effect sizes between strata.

To determine whether the exome signals were novel, or accounted for by previously described associations, index exonic variants from each locus were compared to prior COPD and lung function GWAS results (3,4,6,7,26) and were considered distinctly associated if outside of a 2Mb window. For SNPs within this 2Mb window, we assessed linkage disequilibrium (LD) between exonic variants and prior GWAS variants by calculating an R2 value using a reference panel of 10,000 randomly selected UK Biobank participants (4). To determine if exonic SNPs were distinct from previously described lead GWAS variants, we used results from GCTA-Conditional and joint (COJO) analyses (44) from a prior GWAS, as exome arrays do not assay genome-wide variants (4). Being previously performed, this conditional and joint (COJO) analysis was necessarily limited to variants and cohorts present in the prior GWAS (i.e., all cohorts in the current analysis except HABC and UK COPD Exome Chip Consortium). We calculated a conditionally-significant p-value threshold by performing a Bonferroni correction for the total number of functional exonic variants genotyped within 2 Mb of the index GWAS variants in which exome-wide significant variants were found (0.05/340 variants (across all regions) = 0.000147).

To examine whether exonic SNPs explained the lead signal at previously reported GWAS loci, we examined whether the exonic variant was present within the 99% credible sets from a recent COPD GWAS (4), obtained using the method of Wakefield, *et al*. (45). We also evaluated predicted functional consequences of amino acid mutations using PolyPhen 2.0 and scaled CADD (Combined Annotation Dependent Depletion) scores (46,47). Briefly, CADD scores are based on a support vector machine model predicting the relative deleteriousness of a mutation within a dataset; scaling these scores on a rank order magnitude scale allows for external comparisons. For example, a scaled CADD score of 10 means the mutation is in the top 10% of deleterious mutations, a scaled CADD score of 20 means the mutation is in the top 1% of deleterious mutation, and so forth (46,47). To gain insight into potential biological pathways affected by exonic variants, we also queried gene names at genetics.opentargets.org, which reports relevant biological pathways based on the Reactome Database of Pathways(48,49).

We performed expression quantitative trait locus (eQTL) lookups for COPD-associated exonic variants, extracting eQTL-regulated genes (eGenes) with PeQTL < 1e-8 from prior publications and publicly available data. We queried four previously published eQTL data sources, including GTEx ([www.gtexportal.org](http://www.gtexportal.org)) (50,51) analysis release V6, cis- and trans- eQTLs from Westra et al (52), lung tissue eQTLs from the Hao et al study of asthma (53), and cis- and trans-eQTLs from Vosa et al study in the eQTLGen Consortium(54). For all eQTL sources, a false discovery rate (FDR) of < 0.05 was considered a statistically significant eQTL association. Due to sparsity inherent to exome array association analyses, colocalization with eQTLs could not be performed. Therefore, for each COPD-associated exonic variant that was an eQTL for an eGene, we calculated R2 values between the COPD-associated eQTL and the eGene’s sentinel eQTL SNP using the UK Biobank as an LD reference panel, considering an R2 > 0.2 to be indicative of a shared causal variant in the eQTL and exome array analyses. Sentinel eQTL SNPs for each eGene were defined as the eQTL SNP with the lowest p-value. Protein QTL (pQTL) analyses were performed by querying SNP-regulated proteins from Sun et al. (55), and considering p-values less than a Bonferroni-corrected threshold (0.05/128,037 SNP-protein pairs = 3.9e-7) to be statistically significant.

We also examined the association of the non-synonymous *SERPINA1* Z allele (rs28929474), the most common cause of alpha-1 antitrypsin deficiency, with COPD in our study. For the Z allele, we examined the impact of including Z allele homozygotes in a study. For COPD-associated functional exonic variants and the Z allele, we constructed forest plots using the meta R package (56). To examine heterogeneity across studies, we performed meta-regression (56) of COPD-associated variant effect sizes across studies, evaluating the contribution of age, FEV1 % predicted, pack-years of smoking, and whether studies excluded ZZ homozygotes (for the Z allele) to individual variant effect sizes.

**Results**

Characteristics of cohorts

Characteristics of study participants in each cohort are shown in Table 1. In total, there were 218,399 controls and 33,851 moderate-to-severe COPD cases, which provided 99% power to detect variants with MAF of 0.01 and odds ratio of 1.3. (Table S1). The cohorts were diverse with respect to case ascertainment, sex distribution, cigarette smoking history, and ancestry. For example, COPDGene, BEOCOPD, ICGN, TCGS (Korea and Poland), and the UK COPD Exome Chip Consortium were COPD case-control studies and thus the participants were enriched for COPD cases, ever-smoking status, pack-years of smoking, and lower FEV1 % predicted compared to individuals in the population-based cohorts. TCGS-Korea had the highest percentage of males (>95%) in both case and control participants, while the lowest proportion of males was observed amongst cases in BEOCOPD (39.9%) and controls (27.6%) in CHS Aa. Amongst COPD cases, there was a predominance of males, with 14 cohorts reporting over 55% of cases to be male. With respect to genetic ancestry, there were 7,493 African (including African-Americans), 236,312 European, 7,771 East Asian, and 674 Hispanic participants. Not surprisingly, with this sample size and the different cohort characteristics, all of the ANOVA or chi-squared P-values across were significant (P < 1x10-3).

Exome chip meta-analysis

An overview of the study design is shown in Figure 1. Exome arrays containing 109,036 non-synonymous, stop, and splice variants from ICGC (n=51,458) and UK Biobank (n= 200,792) were meta-analyzed; 20,536 variants were reported in UK Biobank and ≥ 50% of the other studies. Distribution of variant allele frequencies are shown in Figure S1. Of these, 80 variants reached exome-wide Bonferroni-adjusted level of significance (p < 2.4e-6) (Figure 2, Table S2). After clumping, these 80 variants were represented by 35 lead variants. All 35 lead variants were within 2 Mb of previously reported GWAS SNPs (Table 2). Eight of these variants met criteria (see Methods) for replication of exonic signals from prior exome array and genome-wide studies (Table S3). Twenty-one exonic variants were in low LD (R2<0.2) with nearby GWAS variants.

Of the 35 exome-wide significant lead variants, we identified 4 novel conditionally significant exonic SNPs (Table 3), meaning that these SNPs were within 2Mb of COPD GWAS variants, though retained regional significance after conditioning on the lead COPD GWAS SNP using GCTA-COJO (Bonferroni p-value = 0.05/340 variants = 0.000147; see Methods). Seven of the 35 exonic variants were index variants, so conditional and joint analyses were not performed for these variants (rs721917, rs28929474, rs12373142, rs11205303, rs2571445, rs1800888, rs1334576) (4). The rs2454206 variant in *TET2* was significant after conditioning on the rs34712979 index variant, though this variant exists at a locus with two additional independent variants (rs2047409 and rs10516528)(5). We observed that the exonic rs2454206 variant association was no longer significant after conditioning on rs2047409 (p=0.24). In stepwise joint modeling considering these four *TET2* locus variants, only rs2454206 and rs34712979 were selected for the final model.

We also evaluated the 35 exome-wide significant lead variants in the 99% credible sets for the COPD GWAS loci from Sakornsakolpat et al. (4) (Table 4); 18 lead exonic variants were present in the 99% credible sets. Three variants had a posterior probability of association (PPA) > 10%, and 10 variants were in the top 20% of their respective credible sets (Figure S2), suggesting these are more likely to be causal variants. Only rs1334576 in *RREB1* had a PPA > 10%, ranked within the top 20% of its credible set, and was predicted to be damaging by PolyPhen and CADD. The remaining 17 top exonic variants (Table S4) were not present in their respective 99% credible set.

We also analyzed 257 stop or splice variants, of which 2 stop variants (rs3130453 and rs72856718) and 1 splice variant (rs11078928) reached Bonferroni-adjusted significance (p < 0.00019; see Methods) (Table S5). The stop variant rs3130453 (MAF=0.49) in the Coiled-Coil Alpha-Helical Rod Protein 1 (*CCHCR1*) gene was associated with an odds ratio of 1.04 [95% CI: 1.02-1.06, p=1.3e-5] for COPD. The rs72856718 stop variant (MAF=0.09), also in the *CCHCR1* gene, was associated with an odds ratio of 1.08 [95% CI: 1.04-1.13, p = 8.8e-5] for COPD. The splice variant, leading to an exon 6 deletion in *GSDMB*, had an odds ratio of 1.04 [95% CI: 1.02-1.06, p = 1.8e-4] in association with COPD. Forest plots for all exome-wide significant, stop, and splice variants are shown in Figure S3. Reactome pathways for the genes associated with conditionally significant, stop, and splice variants are shown in Table S6. Twelve exonic variants (including stop/splice variants), many of which were highly correlated with each other (Figure 3), are located within the complex HLA region (hg19; chromosome 6:28477797-33448354).

Similar associations of variants in Table 2 were observed in a prior GWAS of FEV1 and FEV1/FVC, except that three variants (rs3885951, rs11078928, rs28929474) did not reach our level of exome-wide significance for either GWAS phenotype (Table S7). We also assessed for interactions of exome-wide significant variants in Table 2 with age in UK Biobank, and found no significant interactions (all p < 0.05). Furthermore, we evaluated whether each of these variants associated with earlier age of COPD diagnosis in the COPDGene cohort, and observed that the smoking behavior-associated rs16969968 SNP in *CHRNA5* was associated with earlier age of COPD diagnosis (p = 0.001, Figure S4). Comparing ever- to never- and heavy- to light-smokers, the effect sizes are generally similar between strata (Figure S5). The exceptions include rs1422795, rs2571446, and rs3829947 in ever- vs. never-smokers, and rs1422795 in heavy- versus light-smokers.

*eQTL and pQTL analyses*

The correlation between the 35 lead exonic variants and sentinel eQTL SNPs (i.e., the eQTL SNP with the lowest p-value of all eQTL SNPs assigned to an eGene) in which the LD R2 is > 0.2 are shown in Table 5, and the full set of eQTL-regulated SNPs are shown in Table S8. Several exonic variants are associated with eQTL SNPs that regulate the same gene within lung tissue, including *FBXO38*, *ADGRG6*, *RREB1,* *C15orf40*, and *EFCAB5*. In pQTL analyses (55), 12 exonic variants were significantly associated with protein expression (Table 6).

*SERPINA1* Z allele effects

The *SERPINA1* Z allele, rs28929474 was associated with a 1.18 odds ratio for COPD in fixed effects analysis ([95% CI: 1.10-1.26], p=1.74e-6) (Table S5). In the Z allele meta-analysis, there is evidence of heterogeneity (I2=0.6) and the African Americans from the ARIC cohort exhibited an opposite direction of effect which was not statistically significant (Figure 4). Given the observed Z allele effect size heterogeneity, we performed a random effects meta-analysis, and the rs28929474 variant demonstrated association with a 1.43 odds ratio for COPD ([95% CI: 1.17-1.74], p=0.0043).

*Meta-regression*

For each exome-wide significant variant, we performed meta-regression to examine the cohort-specific effects of FEV1 % predicted and pack-years of smoking; for the Z allele (rs28929474), we also examined the effects of inclusion of Z allele homozygotes on the reported variant effect sizes (Table S9). Mean differences in FEV1 % predicted from individual cohorts did not account for the observed heterogeneity, nor did whether a study excluded Z allele homozygotes. Heterogeneity of effect sizes was at least partially attributable to mean differences in pack-years of smoking for several variants (rs12373142, rs1334576, rs16969968, rs2523989, rs3130453, rs7750641).

**Discussion**

In this study, we meta-analyzed exome array data from 33,851 moderate-to-severe COPD cases and 218,399 controls. We report 4 exonic variants on chromosomes 5q, 6p, and 15q, as well as 2 stop variants and 1 splice variant associated with COPD. We also examined the association of the *SERPINA1* Z allele (rs28929474) with COPD, and heterogeneity of effect sizes across cohorts. These results lend further insight into the potential pathogenesis of this disease and identify potential loci for laboratory-based validation.

Compared to prior studies, this exome array meta-analysis includes significantly more participants and extends prior findings by providing in-depth characterization of exonic variants. Using the criteria that exome-wide significance is reached, and the direction of effect and risk allele are the same in both the current and prior studies, 8 exonic variants from prior genome-wide and exome array studies were replicated (3,4,6,7,14–16,57). Of these, multiple lines of evidence suggest that rs1334576 in *RREB1* is likely a functional variant; *RREB1* is a zinc-finger transcription factor that binds to Ras-related elements in gene promoters, and has been implicated in cell differentiation (58–60). Smoking history was associated with the *CHRNA5* variant in meta-regression, which is not surprising given the well-established role of this locus in smoking behavior (61). This variant was also associated with early age of COPD diagnosis. Although we had adequate power to detect variants with a MAF of 0.01 with an effect size of 1.3 or greater, we did not identify any novel rare variants with large effect sizes. These data suggest that low frequency protein coding variants (down to 1%) with large effect sizes do not play a substantial role in COPD pathogenesis. However, this study was not powered to assess the impact of very rare variants (MAF < 0.01), nor the effects of low frequency and common protein coding variants on COPD subtypes.

While all variants were near previously identified COPD GWAS loci, applying a more relaxed multiple testing threshold (62,63) led to identification of four independently associated exonic variants that remained significant after conditioning on the lead nearby GWAS variant. The rs2287749 variant may be a causal variant based on conditional, credible set, and PolyPhen analyses, and is located within *ADAM19*, a metalloproteinase (64) involved in cell-matrix interactions and invadopodia formation in cancer cells (49) and previously implicated in COPD risk (4,65). Three independent variants have been previously reported at the *TET2* 4q24 locus (5). Conditional and joint analyses suggest that rs2047409 and rs34712979 account for the signal observed at this locus. *TET2* is involved in DNA demethylation and regulation of gene expression, and variants have been associated with extremes of FEV1 (5), and linked to age-associated clonal hematopoiesis and self-reported COPD and/or asthma (66). Our results further highlight the importance of *TET2* in COPD risk. Two novel variants, rs4842860 and rs17361375, in *C15orf40* were identified. In eQTL analyses, the former SNP was correlated with eQTL SNP rs6603041 in both lung and blood. Neither variant was found in their respective 99% credible sets from a COPD GWAS. This finding might indicate that these variants are not causal or could indicate that this locus was not adequately characterized by GWAS. Deeper characterization of this locus could help clarify its role in COPD pathogenesis.

Examining only the most deleterious (stop and splice) variants, we identified three nominally-significant associations. We found two stop variants in the *CCHCR1* gene. *CCHCR1* has a role in regulating cell proliferation and differentiation, both of which are important in the pathogenesis of emphysema. The splice variant, rs11078928, encodes a polymorphism at a splice acceptor site in *gasdermin-B* (*GSDMB)* on chromosome 17q21 (67). GSDMB is important in pyroptosis, a type of programmed cell death which releases inflammatory mediators. Pyroptosis is activated by caspase-mediated cleavage of the inhibitory C-terminus of gasdermin-B, releasing the functional N-terminus (68–70); The rs11078928 variant leads to a deletion of exon 6 in the N-terminus, rendering gasdermin-B unable to activate pyroptosis (67). The minor allele (C) has been associated with lower asthma risk (67), and was associated with lower COPD risk in our study. This finding is consistent with the notion that subpopulations of individuals have features of both asthma and COPD; indeed, childhood asthma is associated with lower lung function and increased risk for COPD in adulthood (71,72). Thus, *GSDMB* may contribute to the pathobiology of asthma-COPD overlap.

Prior investigations into the Z allele association with COPD have been conflicting. Many prior lung function and COPD genetic association studies (GWASs and exome-wide) have not reported associations with the *SERPINA1* Z allele (3–6,14–16). Yet, the Z allele has also been associated with severe COPD (73) and lung function (74) at genome-wide significance in cohorts enriched for COPD and heavy smoking, potentially with a gene-by-smoking interaction (75). We evaluated the effect of a directly genotyped (rather than imputed) Z allele and applied a random effects meta-analysis. The Z allele was associated with a 1.17-1.74 odds ratio for COPD.

We observed an asymmetric distribution of effect sizes and standard errors for the Z allele across cohorts, suggesting that there may be cohort-specific selection bias with regards to inclusion of individuals with the Z allele. The combination of the exclusion of homozygous Z allele individuals (PiZZ) from many COPD case-control studies and the inclusion of a large number of individuals with little to no smoking history in population-based cohorts likely diminished the power to detect an overall effect for Z allele. To explore this issue, we used meta-regression to assess the impact of intentional exclusion of ZZ individuals, lung function severity, and cigarette smoke exposure on Z allele effect size heterogeneity. None of these factors clearly explained the observed Z allele effect size heterogeneity across studies.

This study has several strengths and limitations. The primary strengths of this study are the large sample size, and the direct assessment of protein coding variant associations with COPD in the context of GWAS findings. We were not well-powered to detect variants with MAF < 0.01, so we are unable to assess the impact of very rare protein coding variants on COPD risk. Larger studies are needed to replicate our findings and assess the impact of very rare variants. Exome array data provide sparse coverage of variants across the genome, making colocalization analyses between COPD exonic variant associations and eQTLs or pQTLs impossible. However, we attempted to address this limitation by identifying eQTLs and pQTLs in highest LD with COPD exonic variants. Finally, we employed a limited set of bioinformatic prediction tools to identify functional variants, but the accuracy of such tools for predicting biologically important changes in protein structure and function is not clear (76). Laboratory-based validation is critical to understanding the causal influence of the exome array variants reported here.

In conclusion, we performed the largest exome array meta-analysis of moderate-to-severe COPD to date. We were unable to identify any protein-altering coding variants at exome-wide significance in regions not previously identified by GWAS. However, at previously-described GWAS loci, we report multiple coding variants associated with COPD, including 4 conditionally significant nonsynonymous variants, two stop variants, and a splice variant. These variants exist in genes important in cell-matrix interactions, cell proliferation, DNA demethylation, regulation of proteases, and regulation of cell death. We further identify heterogeneity of effects of the *SERPINA1* Z allele across cohorts. Future studies will be needed to replicate and validate these identified exonic variants, identify rarer variants, and further describe the role of coding variants in COPD pathogenesis.

References

1. Vogelmeier CF, Criner GJ, Martinez FJ, Anzueto A, Barnes PJ, Bourbeau J, et al. Global Strategy for the Diagnosis, Management, and Prevention of Chronic Obstructive Lung Disease 2017 Report. GOLD Executive Summary. Am J Respir Crit Care Med [Internet]. 2017;195(5):557–82. Available from: http://www.ncbi.nlm.nih.gov/pubmed/28128970

2. Soriano JB, Lamprecht B, Ramírez AS, Martinez-Camblor P, Kaiser B, Alfageme I, et al. Mortality prediction in chronic obstructive pulmonary disease comparing the GOLD 2007 and 2011 staging systems: a pooled analysis of individual patient data. Lancet Respir Med [Internet]. 2015 Jun;3(6):443–50. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25995071

3. Shrine N, Guyatt AL, Erzurumluoglu AM, Jackson VE, Hobbs BD, Melbourne CA, et al. New genetic signals for lung function highlight pathways and chronic obstructive pulmonary disease associations across multiple ancestries. Nat Genet [Internet]. 2019;51(3):481–93. Available from: http://www.ncbi.nlm.nih.gov/pubmed/30804560

4. Sakornsakolpat P, Prokopenko D, Lamontagne M, Reeve NF, Guyatt AL, Jackson VE, et al. Genetic landscape of chronic obstructive pulmonary disease identifies heterogeneous cell-type and phenotype associations. Nat Genet [Internet]. 2019;51(3):494–505. Available from: http://www.ncbi.nlm.nih.gov/pubmed/30804561

5. Wain L V, Shrine N, Miller S, Jackson VE, Ntalla I, Soler Artigas M, et al. Novel insights into the genetics of smoking behaviour, lung function, and chronic obstructive pulmonary disease (UK BiLEVE): a genetic association study in UK Biobank. Lancet Respir Med [Internet]. 2015 Oct;3(10):769–81. Available from: http://www.ncbi.nlm.nih.gov/pubmed/26423011

6. Wain L V, Shrine N, Artigas MS, Erzurumluoglu AM, Noyvert B, Bossini-Castillo L, et al. Genome-wide association analyses for lung function and chronic obstructive pulmonary disease identify new loci and potential druggable targets. Nat Genet [Internet]. 2017 Mar;49(3):416–25. Available from: http://www.ncbi.nlm.nih.gov/pubmed/28166213

7. Hobbs BD, de Jong K, Lamontagne M, Bossé Y, Shrine N, Artigas MS, et al. Genetic loci associated with chronic obstructive pulmonary disease overlap with loci for lung function and pulmonary fibrosis. Nat Genet [Internet]. 2017 Mar;49(3):426–32. Available from: http://www.ncbi.nlm.nih.gov/pubmed/28166215

8. Hindorff LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS, et al. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. Proc Natl Acad Sci U S A [Internet]. 2009 Jun 9;106(23):9362–7. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19474294

9. Eriksson S. Studies in alpha 1-antitrypsin deficiency. Acta Med Scand Suppl [Internet]. 1965;432:1–85. Available from: http://www.ncbi.nlm.nih.gov/pubmed/4160491

10. Hobbs BD, Silverman EK CM. Genetics and epidemiology. In: Strnad P, Brantly ML BR, editor. α1-Antitrypsin Deficiency (ERS Monograph). Sheffield, European Respiratory Society; 2019. p. 27–38.

11. Bossé Y, Lamontagne M, Gaudreault N, Racine C, Levesque M-H, Smith BM, et al. Early-onset emphysema in a large French-Canadian family: a genetic investigation. Lancet Respir Med [Internet]. 2019;7(5):427–36. Available from: http://www.ncbi.nlm.nih.gov/pubmed/31000475

12. Stanley SE, Chen JJL, Podlevsky JD, Alder JK, Hansel NN, Mathias RA, et al. Telomerase mutations in smokers with severe emphysema. J Clin Invest [Internet]. 2015 Feb;125(2):563–70. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25562321

13. Abecasis Lab at the University of Michigan Center for Statistical Genetics. Exome chip design [Internet]. [cited 2019 Aug 8]. Available from: http://genome.sph.umich.edu/wiki/https://genome.sph.umich.edu/wiki/Exome\_Chip\_Design

14. Hobbs BD, Parker MM, Chen H, Lao T, Hardin M, Qiao D, et al. Exome array analysis identifies a common Variant in IL27 associated with chronic obstructive pulmonary disease. Am J Respir Crit Care Med. 2016;194(1):48–57.

15. Jackson VE, Latourelle JC, Wain L V, Smith A V, Grove ML, Bartz TM, et al. Meta-analysis of exome array data identifies six novel genetic loci for lung function. Wellcome open Res [Internet]. 2018;3:4. Available from: http://www.ncbi.nlm.nih.gov/pubmed/30175238

16. Jackson VE, Ntalla I, Sayers I, Morris R, Whincup P, Casas JP, et al. Exome-wide analysis of rare coding variation identifies novel associations with COPD and airflow limitation in MOCS3, IFIT3 and SERPINA12. Thorax. 2016;71(6):501–9.

17. Foreman MG, Wilson C, DeMeo DL, Hersh CP, Beaty TH, Cho MH, et al. Alpha-1 Antitrypsin PiMZ Genotype Is Associated with Chronic Obstructive Pulmonary Disease in Two Racial Groups. Ann Am Thorac Soc [Internet]. 2017 Aug;14(8):1280–7. Available from: http://www.ncbi.nlm.nih.gov/pubmed/28380308

18. Castaldi PJ, DeMeo DL, Kent DM, Campbell EJ, Barker AF, Brantly ML, et al. Development of predictive models for airflow obstruction in alpha-1-antitrypsin deficiency. Am J Epidemiol [Internet]. 2009 Oct 15;170(8):1005–13. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19726494

19. The Atherosclerosis Risk in Communities (ARIC) Study: design and objectives. The ARIC investigators. Am J Epidemiol [Internet]. 1989 Apr;129(4):687–702. Available from: http://www.ncbi.nlm.nih.gov/pubmed/2646917

20. Fried LP, Borhani NO, Enright P, Furberg CD, Gardin JM, Kronmal RA, et al. The Cardiovascular Health Study: design and rationale. Ann Epidemiol [Internet]. 1991 Feb;1(3):263–76. Available from: http://www.ncbi.nlm.nih.gov/pubmed/1669507

21. Regan EA, Hokanson JE, Murphy JR, Make B, Lynch DA, Beaty TH, et al. Genetic Epidemiology of COPD (COPDGene) Study Design. COPD J Chronic Obstr Pulm Dis [Internet]. 2011 Feb 9;7(1):32–43. Available from: http://www.tandfonline.com/doi/full/10.3109/15412550903499522

22. Silverman EK, Weiss ST, Drazen JM, Chapman HA, Carey V, Campbell EJ, et al. Gender-related differences in severe, early-onset chronic obstructive pulmonary disease. Am J Respir Crit Care Med [Internet]. 2000 Dec;162(6):2152–8. Available from: http://www.ncbi.nlm.nih.gov/pubmed/11112130

23. Patel BD, Coxson HO, Pillai SG, Agustí AGN, Calverley PMA, Donner CF, et al. Airway wall thickening and emphysema show independent familial aggregation in chronic obstructive pulmonary disease. Am J Respir Crit Care Med [Internet]. 2008 Sep 1;178(5):500–5. Available from: http://www.ncbi.nlm.nih.gov/pubmed/18565956

24. Zhu G, Warren L, Aponte J, Gulsvik A, Bakke P, Anderson WH, et al. The SERPINE2 gene is associated with chronic obstructive pulmonary disease in two large populations. Am J Respir Crit Care Med. 2007 Jul;176(2):167–73.

25. Splansky GL, Corey D, Yang Q, Atwood LD, Cupples LA, Benjamin EJ, et al. The Third Generation Cohort of the National Heart, Lung, and Blood Institute’s Framingham Heart Study: design, recruitment, and initial examination. Am J Epidemiol [Internet]. 2007 Jun 1;165(11):1328–35. Available from: http://www.ncbi.nlm.nih.gov/pubmed/17372189

26. Wilk JB, Chen T-H, Gottlieb DJ, Walter RE, Nagle MW, Brandler BJ, et al. A genome-wide association study of pulmonary function measures in the Framingham Heart Study. PLoS Genet [Internet]. 2009 Mar;5(3):e1000429. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19300500

27. Kannel WB, Feinleib M, McNamara PM, Garrison RJ, Castelli WP. An investigation of coronary heart disease in families. The Framingham offspring study. Am J Epidemiol [Internet]. 1979 Sep;110(3):281–90. Available from: http://www.ncbi.nlm.nih.gov/pubmed/474565

28. Simonsick EM, Newman AB, Nevitt MC, Kritchevsky SB, Ferrucci L, Guralnik JM, et al. Measuring higher level physical function in well-functioning older adults: expanding familiar approaches in the Health ABC study. J Gerontol A Biol Sci Med Sci [Internet]. 2001 Oct;56(10):M644-9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/11584038

29. Hong CB, Kim YJ, Moon S, Shin Y-A, Cho YS, Lee J-Y. KAREBrowser: SNP database of Korea Association REsource Project. BMB Rep [Internet]. 2012 Jan;45(1):47–50. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22281013

30. Hankinson JL, Kawut SM, Shahar E, Smith LJ, Stukovsky KH, Barr RG. Performance of American Thoracic Society-recommended spirometry reference values in a multiethnic sample of adults: the multi-ethnic study of atherosclerosis (MESA) lung study. Chest [Internet]. 2010 Jan;137(1):138–45. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19741060

31. Bild DE, Bluemke DA, Burke GL, Detrano R, Diez Roux A V, Folsom AR, et al. Multi-Ethnic Study of Atherosclerosis: objectives and design. Am J Epidemiol [Internet]. 2002 Nov 1;156(9):871–81. Available from: http://www.ncbi.nlm.nih.gov/pubmed/12397006

32. Hofman A, Darwish Murad S, van Duijn CM, Franco OH, Goedegebure A, Ikram MA, et al. The Rotterdam Study: 2014 objectives and design update. Eur J Epidemiol [Internet]. 2013 Nov;28(11):889–926. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24258680

33. Ikram MA, van der Lugt A, Niessen WJ, Koudstaal PJ, Krestin GP, Hofman A, et al. The Rotterdam Scan Study: design update 2016 and main findings. Eur J Epidemiol [Internet]. 2015 Dec;30(12):1299–315. Available from: http://www.ncbi.nlm.nih.gov/pubmed/26650042

34. Zhou X, Baron RM, Hardin M, Cho MH, Zielinski J, Hawrylkiewicz I, et al. Identification of a chronic obstructive pulmonary disease genetic determinant that regulates HHIP. Hum Mol Genet [Internet]. 2012 Mar 15;21(6):1325–35. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22140090

35. Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. PLoS Med [Internet]. 2015 Mar;12(3):e1001779. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25826379

36. Winkler TW, Day FR, Croteau-Chonka DC, Wood AR, Locke AE, Mägi R, et al. Quality control and conduct of genome-wide association meta-analyses. Nat Protoc [Internet]. 2014 May;9(5):1192–212. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24762786

37. Skol AD, Scott LJ, Abecasis GR, Boehnke M. Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. Nat Genet [Internet]. 2006 Feb;38(2):209–13. Available from: http://www.ncbi.nlm.nih.gov/pubmed/16415888

38. Buist AS, McBurnie MA, Vollmer WM, Gillespie S, Burney P, Mannino DM, et al. International variation in the prevalence of COPD (the BOLD Study): a population-based prevalence study. Lancet (London, England) [Internet]. 2007 Sep 1;370(9589):741–50. Available from: http://www.ncbi.nlm.nih.gov/pubmed/17765523

39. Pauwels RA, Rabe KF. Burden and clinical features of chronic obstructive pulmonary disease (COPD). Lancet (London, England) [Internet]. 364(9434):613–20. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15313363

40. Belle G van. Statistical Rules of Thumb. New York: John Wiley and Sons; 45–46 p.

41. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics [Internet]. 2010 Sep 1;26(17):2190–1. Available from: http://www.ncbi.nlm.nih.gov/pubmed/20616382

42. Chang X, Wang K. wANNOVAR: annotating genetic variants for personal genomes via the web. J Med Genet [Internet]. 2012 Jul;49(7):433–6. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22717648

43. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience [Internet]. 2015;4:7. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25722852

44. Yang J, Ferreira T, Morris AP, Medland SE, Genetic Investigation of ANthropometric Traits (GIANT) Consortium, DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) Consortium, et al. Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. Nat Genet [Internet]. 2012 Mar 18;44(4):369–75, S1-3. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22426310

45. Wakefield J. Bayes factors for genome-wide association studies: comparison with P-values. Genet Epidemiol [Internet]. 2009 Jan;33(1):79–86. Available from: http://www.ncbi.nlm.nih.gov/pubmed/18642345

46. Kircher M, Witten DM, Jain P, O’Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. Nat Genet [Internet]. 2014 Mar;46(3):310–5. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24487276

47. Rentzsch P, Witten D, Cooper GM, Shendure J, Kircher M. CADD: predicting the deleteriousness of variants throughout the human genome. Nucleic Acids Res [Internet]. 2019 Jan 8;47(D1):D886–94. Available from: http://www.ncbi.nlm.nih.gov/pubmed/30371827

48. Wu G, Haw R. Functional Interaction Network Construction and Analysis for Disease Discovery. Methods Mol Biol [Internet]. 2017;1558:235–53. Available from: http://www.ncbi.nlm.nih.gov/pubmed/28150241

49. Fabregat A, Sidiropoulos K, Viteri G, Forner O, Marin-Garcia P, Arnau V, et al. Reactome pathway analysis: a high-performance in-memory approach. BMC Bioinformatics [Internet]. 2017 Mar 2;18(1):142. Available from: http://www.ncbi.nlm.nih.gov/pubmed/28249561

50. GTEx Consortium. The Genotype-Tissue Expression (GTEx) project. Nat Genet [Internet]. 2013 Jun;45(6):580–5. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23715323

51. GTEx Consortium. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. Science [Internet]. 2015 May 8;348(6235):648–60. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25954001

52. Westra H-J, Peters MJ, Esko T, Yaghootkar H, Schurmann C, Kettunen J, et al. Systematic identification of trans eQTLs as putative drivers of known disease associations. Nat Genet [Internet]. 2013 Oct;45(10):1238–43. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24013639

53. Hao K, Bossé Y, Nickle DC, Paré PD, Postma DS, Laviolette M, et al. Lung eQTLs to help reveal the molecular underpinnings of asthma. PLoS Genet [Internet]. 2012;8(11):e1003029. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23209423

54. Urmo Vosa, Annique Claringbould, , Harm-Jan Westra, Marc Jan Bonder, Patrick Deelen, Biao Zeng, Holger Kirsten, Ashis Saha, Roman Kreuzhuber SK. Unraveling the polygenic architecture of complex traits using blood eQTL metaanalysis. bioRxiv. 2018;

55. Sun BB, Maranville JC, Peters JE, Stacey D, Staley JR, Blackshaw J, et al. Genomic atlas of the human plasma proteome. Nature [Internet]. 2018;558(7708):73–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/29875488

56. G S. meta: An R package for meta-analysis. R News. 2007;7:40–5.

57. Busch R, Hobbs BD, Zhou J, Castaldi PJ, McGeachie MJ, Hardin ME, et al. Genetic Association and Risk Scores in a Chronic Obstructive Pulmonary Disease Meta-analysis of 16,707 Subjects. Am J Respir Cell Mol Biol [Internet]. 2017;57(1):35–46. Available from: http://www.ncbi.nlm.nih.gov/pubmed/28170284

58. Zhang L, Zhao J, Edenberg HJ. A human Raf-responsive zinc-finger protein that binds to divergent sequences. Nucleic Acids Res [Internet]. 1999 Jul 15;27(14):2947–56. Available from: http://www.ncbi.nlm.nih.gov/pubmed/10390538

59. Mukhopadhyay NK, Cinar B, Mukhopadhyay L, Lutchman M, Ferdinand AS, Kim J, et al. The zinc finger protein ras-responsive element binding protein-1 is a coregulator of the androgen receptor: implications for the role of the Ras pathway in enhancing androgenic signaling in prostate cancer. Mol Endocrinol [Internet]. 2007 Sep;21(9):2056–70. Available from: http://www.ncbi.nlm.nih.gov/pubmed/17550981

60. Thiagalingam A, De Bustros A, Borges M, Jasti R, Compton D, Diamond L, et al. RREB-1, a novel zinc finger protein, is involved in the differentiation response to Ras in human medullary thyroid carcinomas. Mol Cell Biol [Internet]. 1996 Oct;16(10):5335–45. Available from: http://www.ncbi.nlm.nih.gov/pubmed/8816445

61. Hartz SM, Short SE, Saccone NL, Culverhouse R, Chen L, Schwantes-An T-H, et al. Increased genetic vulnerability to smoking at CHRNA5 in early-onset smokers. Arch Gen Psychiatry [Internet]. 2012 Aug;69(8):854–60. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22868939

62. Glubb DM, Maranian MJ, Michailidou K, Pooley KA, Meyer KB, Kar S, et al. Fine-scale mapping of the 5q11.2 breast cancer locus reveals at least three independent risk variants regulating MAP3K1. Am J Hum Genet [Internet]. 2015 Jan 8;96(1):5–20. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25529635

63. Gallagher MD, Chen-Plotkin AS. The Post-GWAS Era: From Association to Function. Am J Hum Genet [Internet]. 2018;102(5):717–30. Available from: http://www.ncbi.nlm.nih.gov/pubmed/29727686

64. Wei P, Zhao YG, Zhuang L, Ruben S, Sang QX. Expression and enzymatic activity of human disintegrin and metalloproteinase ADAM19/meltrin beta. Biochem Biophys Res Commun [Internet]. 2001 Jan 26;280(3):744–55. Available from: http://www.ncbi.nlm.nih.gov/pubmed/11162584

65. Pérez-Rubio G, Silva-Zolezzi I, Fernández-López JC, Camarena Á, Velázquez-Uncal M, Morales-Mandujano F, et al. Genetic Variants in IL6R and ADAM19 are Associated with COPD Severity in a Mexican Mestizo Population. COPD [Internet]. 2016;13(5):610–5. Available from: http://www.ncbi.nlm.nih.gov/pubmed/27078193

66. Buscarlet M, Provost S, Zada YF, Barhdadi A, Bourgoin V, Lépine G, et al. DNMT3A and TET2 dominate clonal hematopoiesis and demonstrate benign phenotypes and different genetic predispositions. Blood [Internet]. 2017;130(6):753–62. Available from: http://www.ncbi.nlm.nih.gov/pubmed/28655780

67. Panganiban RA, Sun M, Dahlin A, Park H-R, Kan M, Himes BE, et al. A functional splice variant associated with decreased asthma risk abolishes the ability of gasdermin B to induce epithelial cell pyroptosis. J Allergy Clin Immunol [Internet]. 2018;142(5):1469-1478.e2. Available from: http://www.ncbi.nlm.nih.gov/pubmed/29330013

68. Kayagaki N, Stowe IB, Lee BL, O’Rourke K, Anderson K, Warming S, et al. Caspase-11 cleaves gasdermin D for non-canonical inflammasome signalling. Nature [Internet]. 2015 Oct 29;526(7575):666–71. Available from: http://www.ncbi.nlm.nih.gov/pubmed/26375259

69. Ding J, Wang K, Liu W, She Y, Sun Q, Shi J, et al. Pore-forming activity and structural autoinhibition of the gasdermin family. Nature [Internet]. 2016;535(7610):111–6. Available from: http://www.ncbi.nlm.nih.gov/pubmed/27281216

70. Shi J, Zhao Y, Wang K, Shi X, Wang Y, Huang H, et al. Cleavage of GSDMD by inflammatory caspases determines pyroptotic cell death. Nature [Internet]. 2015 Oct 29;526(7575):660–5. Available from: http://www.ncbi.nlm.nih.gov/pubmed/26375003

71. Hayden LP, Cho MH, Raby BA, Beaty TH, Silverman EK, Hersh CP, et al. Childhood asthma is associated with COPD and known asthma variants in COPDGene: a genome-wide association study. Respir Res [Internet]. 2018 Oct 29;19(1):209. Available from: http://www.ncbi.nlm.nih.gov/pubmed/30373671

72. McGeachie MJ, Yates KP, Zhou X, Guo F, Sternberg AL, Van Natta ML, et al. Patterns of Growth and Decline in Lung Function in Persistent Childhood Asthma. N Engl J Med [Internet]. 2016 May 12;374(19):1842–52. Available from: http://www.ncbi.nlm.nih.gov/pubmed/27168434

73. Busch R, Hobbs BD, Zhou J, Castaldi PJ, McGeachie MJ, Hardin ME, et al. Genetic association and risk scores in a chronic obstructive pulmonary disease meta-analysis of 16,707 subjects. Am J Respir Cell Mol Biol. 2017;57(1):35–46.

74. Li X, Ortega VE, Ampleford EJ, Graham Barr R, Christenson SA, Cooper CB, et al. Genome-wide association study of lung function and clinical implication in heavy smokers. BMC Med Genet [Internet]. 2018;19(1):134. Available from: http://www.ncbi.nlm.nih.gov/pubmed/30068317

75. Silverman EK, Province MA, Campbell EJ, Pierce JA, Rao DC. Family study of alpha 1-antitrypsin deficiency: effects of cigarette smoking, measured genotype, and their interaction on pulmonary function and biochemical traits. Genet Epidemiol [Internet]. 1992;9(5):317–31. Available from: http://www.ncbi.nlm.nih.gov/pubmed/1427021

76. Miosge LA, Field MA, Sontani Y, Cho V, Johnson S, Palkova A, et al. Comparison of predicted and actual consequences of missense mutations. Proc Natl Acad Sci U S A [Internet]. 2015 Sep 15;112(37):E5189-98. Available from: http://www.ncbi.nlm.nih.gov/pubmed/26269570

**Table legends**

Table 1: Characteristics of cohorts in meta-analysis. In total, there were 218,399 controls and 33,851 moderate-to-severe COPD cases.

Table 2: Thirty-five lead exonic SNPs and nearby sentinel GWAS SNPs (within 2Mb). "\*" = variant location within HLA region (hg19; chromosome 6: 28477797-33448354). "\*\*" = R2 < 0.2. Exonic variants were clumped prior to comparing to GWAS SNPs based on an R2>0.2. References to GWAS variants are for the most recent publication.

Table 3: Conditionally significant exome array SNPs (p<0.000147) within 2 Mb of GWAS SNPs identified in UK Biobank(4). Exonic variant effects were adjusted for the index SNPs indicated in the table. Chromosome positions based on build hg19. Note that the rs2454206 variant in *TET2* was significant after conditioning on the rs34712979 index variant, though this variant exists at a locus with two additional independent variants (rs2047409 and rs10516528)(5). We observed that the exonic rs2454206 variant was no longer significant after conditioning on rs2047409 (p=0.24). Note that HLA imputation was not performed, so HLA region variants were not included in conditional and joint analyses.

Table 4: Exome array variants identified in 99% credible sets derived from UK Biobank(4) using the method by Wakefield et al.(45) and wANNOVAR functional annotations. PolyPhen and CADD were used to predict consequences of mutation. PPA = Posterior Probability of Association. CADD scores are based on a support vector machine model predicting the relative deleteriousness of a mutation within a dataset; scaling these scores on a rank order magnitude scale allows for external comparisons. For example, a scaled CADD score of 10 means the mutation is in the top 10% of deleterious mutations, a scaled CADD score of 20 means the mutation is in the top 1% of deleterious mutation, and so forth(46,47). Chromosome positions based on hg19. PPA = posterior probability of association. Percentile indicates the ranking of the exonic variant within the credible set of the GWAS index SNP.

Table 5: Exon SNPs and sentinel eQTL SNPs with R2 > 0.2 and p-value (eQTL) < 1e-8. "\*" indicates HLA region.

Table 6: Exonic genes associated with pQTL-regulated proteins at Bonferroni-corrected significance level (p < 3.9e-7). "\*" indicates variant is in HLA region.

**Figure legends**

Figure 1: Overview of study design.

Figure 2: Manhattan plot of exome array variants. The horizontal red line indicates an exome-wide significance level of 2.4e-6. The exonic variants reaching exome-wide significance are annotated.

Figure 3: LD matrix (R2) of exome array variants within the HLA region (hg19; 6:28477797-33448354).

Figure 4: Forest plot of the *SERPINA1* Z allele (rs28929474). The Z allele was associated with a 1.18 odds ratio for COPD in fixed effects analysis ([95% CI: 1.10-1.26], p=1.74e-6), and 1.43 odds ratio for COPD in random effects analysis ([95% CI: 1.17-1.74), p=0.0043]) (I2 = 0.60). See Methods for cohort abbreviations.