

Smoking Interaction with a Polygenic Risk Score for Reduced Lung Function

Woori Kim^{1†}, Matthew Moll^{2,3†}, Dandi Qiao², Brian D. Hobbs^{2,3}, Nick Shrine⁴, Phuwanat Sakornsakolpat⁵, Martin D. Tobin^{4,6}, Frank Dudbridge⁴, Louise V. Wain^{4,6}, Christine Ladd-Acosta^{7,8}, Nilanjan Chatterjee^{9,10,11}, Edwin K. Silverman^{2,3}, Michael H. Cho^{2,3#}, Terri H. Beaty^{8#}.

Affiliations:

1 Systems Biology and Computer Science Program, Ann Romney Center for Neurological Diseases, Department of Neurology, Brigham and Women's Hospital, Boston MA, 02115

2 Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital, Boston MA, 02115

3 Division of Pulmonary and Critical Care Medicine, Department of Medicine, Brigham and Women's Hospital, Boston MA, 02115

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

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

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

7 Wendy Klag Center for Autism and Developmental Disabilities, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD 21205.

8 Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD 21205.

9 Department of Biostatistics, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD 21205.

10 Department of Oncology, School of Medicine, Johns Hopkins University, Baltimore, MD 21205.

11 Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Department of Health and Human Services, Bethesda, MD 20892.

† = These authors contributed equally

= Co-senior authors

Correspondence:

Terri H. Beaty, Ph.D.

615 N. Wolfe Street

Baltimore, MD 21205

tbeaty1@jhu.edu

and/or

Michael H. Cho, M.D., M.P.H.

181 Longwood Avenue

Boston, MA 02115

remhc@channing.harvard.edu

Funding:

MM is supported by T32HL007427.

DQ is supported by NIH K01HL129039 and the NHLBI BioData Catalyst Fellow Program.

BDH is supported by NIH K08HL136928 and U01 HL089856.

MHC is supported by NIH R01HL137927 and R01HL135142.

EKS is supported by NIH R01 HL137927, R01 HL147148, U01 HL089856, 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).

Disclosures:

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. The other authors report no disclosures.

Author Contributions:

Study Design: Matthew Moll, Woori Kim, Terri H. Beaty, Michael H. Cho

Acquisition, analysis, or interpretation of the data: Woori Kim, Matthew Moll, Brian D. Hobbs, Michael H. Cho, Edwin K. Silverman, Terri H. Beaty.

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

Statistical analysis: Matthew Moll, Woori Kim, Brian D. Hobbs, Michael H. Cho, Dandi Qiao, Edwin K. Silverman

Obtained funding: Terri H. Beaty, Edwin K. Silverman, Michael H. Cho.

Word counts:

Key Points: 100 (limit: 100)

Abstract: 350 (limit: 350)

Manuscript: 3,245 (limit: 3000)

Key Points

Question: Does cigarette smoking modify the magnitude of genetic effects on forced expiratory volume in 1 second (FEV₁)/forced vital capacity (FVC)?

Findings: FEV₁/FVC is influenced by PRS-by-smoking interactions. Smoking was detrimental across all categories of predicted genetic risk, though it was worse for those with highest predicted genetic risks. For every reported 20 pack-years of reported smoking, individuals in the top compared to the bottom decile of genetic risk showed nearly twice the reduction in FEV₁/FVC.

Meaning: Elucidating mechanisms for the interaction between smoking and genetic risk could yield greater insight into the pathogenesis of chronic obstructive pulmonary disease (COPD).

Abstract

Importance: Risk to airflow limitation and Chronic Obstructive Pulmonary Disease (COPD) is influenced by combinations of cigarette smoking and genetic susceptibility, yet it remains unclear whether gene-by-smoking interactions contribute to quantitative measures of lung function.

Objective: Determine whether smoking modifies the effect of a polygenic risk score's (PRS's) association with reduced lung function.

Design: United Kingdom (UK) Biobank prospective cohort study.

Setting: Population cohort.

Participants: UK citizens of European ancestry aged 40-69 years, with genetic and spirometry data passing quality control metrics.

Exposures: PRS, self-reported pack-years of smoking, ever- versus never-smoking status, and current- versus former-/never-smoking status.

Main Outcomes and Measures: Forced expiratory volume in 1 second (FEV₁)/forced vital capacity (FVC). We tested for interactions with models including the main effects of PRS, different smoking variables, and their cross-product term(s). We also compared the effects of pack-years of smoking on FEV₁/FVC for those in the highest versus lowest decile of predicted genetic risk for low lung function.

Results: We included 319,730 individuals (24,915 with moderate-to-severe COPD). The PRS and pack-years were significantly associated with lower FEV₁/FVC, as was the interaction term (β [interaction] = -0.0028 [95% CI: -0.0029, -0.0026]; all $p < 0.0001$). A stepwise increment in estimated effect sizes for these interaction terms was observed per 10 pack-years of smoking exposure (all $p < 0.0001$). There was evidence of significant interaction between PRS with ever/never smoking status (β [interaction] = -0.0064 [95% CI: -0.0068, -0.0060]) and current/not-current smoking (β [interaction] = -0.0091 [95% CI: -0.0097, -0.0084]). For any given level of pack-years of smoking exposure, FEV₁/FVC was significantly lower for individuals in the tenth compared to the first decile of genetic risk ($p < 0.0001$). For every 20 pack-years of smoking, those in the top compared to the bottom decile of genetic risk showed nearly a twofold reduction in FEV₁/FVC.

Conclusions and Relevance: COPD is characterized by diminished lung function, and our analyses suggest there is substantial interaction between genome-wide PRS and smoking

exposures. While smoking has negative effects on lung function across all genetic risk categories, effects of smoking are highest in those with higher predicted genetic risk.

Introduction

Chronic obstructive pulmonary disease (COPD) is characterized by airflow obstruction, traditionally defined by a low forced expiratory volume in 1 second (FEV_1)/forced vital capacity (FVC), and cigarette smoking is the greatest environmental risk factor^{1,2}. Only a minority of smokers develop COPD^{3,4}, and genetic factors account for some of this variation in susceptibility, with ~40% of the variability in spirometric measures of pulmonary function attributed to genetic variation⁵⁻⁷. Therefore, it has long been thought that airflow obstruction may develop partially as the result of gene-by-smoking interactions.

Despite the important contribution of both smoking and genetic factors to lung function, compelling evidence for gene-by-smoking interactions has been limited. Genome-wide interaction studies have identified a handful of spirometric- and COPD-associated loci that appear to interact with smoking status,⁸⁻¹⁴ suggesting at least a portion of the variability in spirometric measures of lung function may be attributable to gene-by-smoking interactions. A major challenge in identifying gene-by-smoking interactions on lung function and risk to COPD is that individual genetic variants tend to be of small effect size and account for a low degree of phenotypic variability in lung function, diminishing the power to detect gene-by-smoking interactions.

Pooling individual GWAS variants into a single genetic risk score can account for a greater proportion of phenotypic variability^{15–20}, and should improve power to detect interactions. Genetic risk scores have been used to investigate gene-by-environment interactions in psychiatric²¹ and cardiovascular diseases²². Aschard and colleagues²³ were unable to detect individual single nucleotide polymorphism (SNP)-by-smoking interactions for FEV₁/FVC for 26 variants identified as significant in a genome-wide joint meta-analysis of SNP-by-smoking associations of pulmonary function¹⁴; however, when the authors summed these variants to create a genetic risk score, they found evidence of interaction between the genetic risk score and ever smoking status²³. By contrast, Shrine et al. performed the largest GWAS of lung function to date, developed a genetic risk score including estimated effects of 279 variants showing significant effects on lung function, and reported no evidence of interaction between this genetic risk score and ever smoking status¹⁹, though the authors did observe an interaction of the genetic risk score with ever smoking status on moderate-to-severe COPD. We constructed a polygenic risk score (PRS) based on GWASs of FEV₁ and FEV₁/FVC¹⁹ that explained more of the variability in lung function than seen with the 279-variant risk score used by Shrine et al. (~30% versus <10%)²⁰, and here we further investigate gene-by-smoking interactions.

We hypothesized that multiple measures of smoking exposure would significantly modify the effect of this genome-wide PRS on FEV₁/FVC (i.e. because it is associated with lower lung function) in the United Kingdom (UK) Biobank population-based cohort.

Methods

Study population

We included participants from the UK Biobank, a cohort of over 500,000 individuals aged 40-69 years²⁴. All participants provided written informed consent and study protocols were approved by local institutional review boards/research ethics committees. Participants were excluded if spirometry or genetic data did not meet quality control standards; further details on the impact of these inclusion and exclusion criteria are shown in Figure S1. Quality control (QC) of spirometric data has been previously described^{18,19,24}. Briefly, to determine lung function, FEV₁ and FVC were derived from the spirometry volume-time series data at the time of study enrollment, as previously reported¹⁹.

Genotyping was performed as previously described¹⁹, using Axiom UK BiLEVE array and Axiom Biobank array (Affymetrix, Santa Clara, California, USA) and imputed to the Haplotype Reference Consortium version 1.1 panel (accepting imputation accuracy $R^2 > 0.5$). We dropped variants with minor allele frequency < 0.01 and those showing deviation from Hardy-Weinberg equilibrium ($p < 1e-6$). We used only subjects of European ancestry based on a combination of self-reported ethnicity and k-means clustering of principal components of genetic ancestry, as previously reported¹⁹.

Statistical analyses

All analyses were done in R version 4.0.3 (www.r-project.org). Normality of continuous variables was assessed by visual inspection of histograms. Results are reported as mean \pm standard deviation or median [interquartile range], as appropriate. Differences in continuous

variables were assessed with Student t-tests or Wilcoxon tests, and categorical variables were compared by ANOVA or Kruskal-Wallis tests, as appropriate.

Overview of study design

The primary outcome was the FEV₁/FVC ratio, as clinical COPD is characterized by airflow obstruction (FEV₁/FVC < 0.7), and severity graded based on decrements in FEV₁ % predicted^{1,2}. We first assessed whether three measures of smoking exposure (see below) modified the effect of a PRS on quantitative measures of FEV₁/FVC. We then considered the joint effects of smoking exposures and being in the highest versus lowest decile of the PRS (i.e. highest versus lowest categories of predicted genetic risk). We examined “norms of reaction” for the relationship between pack-years of smoking and FEV₁/FVC for those in the tenth compared to the first decile of predicted genetic risk.

Smoking exposures

We examined three measures of cigarette smoking exposure: 1) pack-years of smoking, 2) ever-versus never-smoking status, and 3) current smoker versus former-/never-smoking status. All smoking information was obtained by self-report. Pack-years of smoking was examined as continuous and categorical (pack-year categories: ≤10, 10.1-20, 20.1-30, 30.1-40, 40.1-50 and >50; where the reference group is ≤10 pack-years) variables. ‘Ever-smokers’ included individuals reporting current smoking, smoking most days, or smoking occasionally. ‘Never smokers’ included those who smoked less than 100 cigarettes in their lifetime. ‘Current smokers’ included those who reported current smoking, and ‘former smokers’ included non-current smokers who smoked more than 100 or more cigarettes in their lifetime.

Polygenic risk score for lung function

A polygenic risk score (PRS) for lung function was calculated as previously described²⁰. Briefly, this PRS was based on genome-wide association results for FEV₁ and FEV₁/FVC in UK Biobank and SpiroMeta¹⁹, and was developed using a penalized regression framework accounting for linkage disequilibrium²⁵. PRSs were calculated for FEV₁ and FEV₁/FVC, and then summed into a composite PRS, which was scaled and centered. The PRS was oriented such that a higher PRS was associated with lower FEV₁ and FEV₁/FVC.

Interaction analyses

We performed multivariable linear regressions of FEV₁/FVC on main effects of the combined PRS, smoking exposure, and cross-product interaction terms. We included covariates age, age², sex, height, genotyping array and the first 10 principal components of genetic ancestry in the linear regression model. Age was scaled and centered prior to squaring. We also performed stratified analyses amongst those in the lowest and highest deciles of the PRS, separately for never and ever smokers.

Investigation of gene-by-environment interactions has been considered to be deviation from either an additive or multiplicative model. Therefore, we additionally examined the joint effects of smoking and PRS to assess a departure of the observed joint effect from the expected effect under an additive model. We focused on comparing the 1st and 10th deciles of PRS as previously done²⁰. We created a categorical variable with mutually exclusive strata formed by the cross classification of smoking and PRS (10th vs. 1st decile). The reference category was the group

with the lowest relative smoking exposure (e.g. pack-years ≤ 10) in the first PRS decile. We then constructed multivariable linear regression models to evaluate the effects of this categorical variable on FEV₁/FVC. The expected effect for those in the highest decile with the highest smoking exposure was estimated under an additive model, and calculated by summing the estimated effect size for the lowest decile vs. the highest smoking exposure group and the highest decile vs. the lowest smoking exposure group.

Norms of Reaction

A “norm of reaction” describes the relationship between a phenotype and environmental exposure for a given genotype²⁷. We assessed norms of reaction for pack-years of smoking and FEV₁/FVC for those in the lowest and highest deciles of predicted genetic risk. We plotted pack-years of smoking versus FEV₁/FVC, stratifying by lowest and highest deciles of genetic risk. We then compared the slopes of these lines with an Analysis of Covariance (ANCOVA) using the `rstatix` R package²⁸. For the purposes of clinical interpretability, we trained multivariable linear regression models to assess the effect of 20 pack-years of smoking on FEV₁/FVC for those in the highest and lowest deciles of predicted genetic risk, adjusting for the covariates detailed as above. As sensitivity analyses, we repeated these analyses in ever smokers and in a dataset excluding all related individuals; to select unrelated individuals, we removed at least one individual from each related pair with kinship coefficient > 0.0625 , favoring the inclusion of COPD cases. We also transformed reported pack-years of smoking (ln, scaling and centering, and rank normalization) to ensure that the effects of interaction terms were not due to misspecification of the main effects of smoking.

Results

Characteristics of study participants

Characteristics of study participants are shown in Table 1. We included 319,730 participants; 24,915 participants met criteria for moderate-to-severe COPD cases (Global Initiative for Chronic Lung Disease (GOLD) spirometry grades 2-4¹), 38,713 had preserved ratio with impaired spirometry (PRISm)²⁹, and 256,102 met criteria for GOLD spirometry grades 0/1.

Interaction of a polygenic risk score with smoking

The PRS was weakly correlated with pack-years of smoking ($r = 0.041$, $p < 0.0001$; Figure S2). The relationship between PRS and FEV₁/FVC stratified by pack-years of smoking categories is illustrated in Figure 1. In multivariable analyses (Table 2), the main effects of the PRS ($\beta = -0.0304$ [95% CI: -0.0307, -0.0302]) and pack-years of smoking categories were associated with FEV₁/FVC (all $p < 0.0001$), and estimated effect sizes increased with each incremental category of smoking exposure. The same incremental trend for interaction terms between PRS and pack-years of smoking categories (all tests for interaction yielded $p < 0.0001$) were observed.

Considering pack-years of smoking as a continuous variable (Table S1), the cross-product interaction term was also associated with FEV₁/FVC (β [interaction] = -0.0028 [95% CI: -0.0029, -0.0026], $p < 0.0001$). We also performed three transformations of pack-years of smoking, and the PRS*pack-years interaction term was significant in each instance (all $p < 0.0001$, Table S2).

The relationship between the PRS and FEV₁/FVC stratified by ever- vs. never- and current- vs. former-/never-smoking statuses are shown in Figures S3A and B, respectively. Ever-smoking and the PRS*ever-smoking status interaction term were significantly associated with FEV₁/FVC (both $p < 0.0001$, Table S3). Similarly, current smoking status and the PRS*current-smoking

status interaction term were significantly associated with FEV₁/FVC (both $p < 0.0001$, Table S4). In stratified analyses, we observed similar results between PRS and smoking exposures (Table S5, Figure S4). Additionally, ever smoking status and the PRS*ever-smoking status interaction term were significantly associated with FEV₁/FVC in the lowest (β [interaction]=-0.0033 [95% CI: -0.0058,-0.00085], $p = 0.0082$) and highest (β [interaction]=-0.0095 [95% CI: -0.013,-0.0056], $p < 0.0001$) deciles of predicted genetic risk (Table S5).

The joint effects of pack-years of smoking categories and PRS are shown in Figure 2. Being in the lowest decile of predicted genetic risk and having more than 50 pack-years of smoking exposure ($\beta=-0.022$ [95% CI: -0.026, -0.018], $p < 0.0001$) had a similar estimated effect size as being in the highest decile of genetic risk and having 11-20 pack-years of smoking exposure ($\beta=-0.024$ [95% CI: -0.026,-0.023], $p < 0.0001$). We observed a greater effect size of being in the highest decile of genetic risk and having more than 50 pack-years of smoking exposure ($\beta=-0.051$ [95% CI: -0.054,-0.047]) than would be expected, confirming possible interaction between the PRS and pack-years of smoking; we observed a similar association for those in the highest genetic risk decile who were current smokers (Table S6), but a non-significant difference for ever smokers in the highest genetic risk decile (Table S7).

[Norms of reaction for highest versus lowest predicted genetic risk deciles](#)

In Figure 3, we show different norms of reaction for the effects of pack-years of smoking on FEV₁/FVC among those in the highest vs. lowest deciles of predicted genetic risk. For any given level of pack-years of smoking, those in the highest PRS decile had lower FEV₁/FVC compared to those in the lowest decile of PRS. Analysis of covariance confirms the slopes of the lines (Figure 3) are significantly different ($p < 0.0001$). We observed similar results in ever-smokers

(Figure S5). For every 20 pack years of smoking, those in the first decile had a change of -0.0084 (95% CI: -0.0091, -0.0076) in FEV₁/FVC, while those in the 10th decile of predicted genetic risk had a change of -0.017 (95% CI: -0.018, -0.016), representing an approximately twofold reduction in FEV₁/FVC for each 20 pack-years of smoking for those in the highest compared to the lowest decile of predicted genetic risk.

Discussion

In this study of over 300,000 UK Biobank participants, we found three measures of smoking exposure modified the effect of a polygenic risk score (PRS) on the quantitative measure of lung function (FEV₁/FVC). As expected, smoking was detrimental to lung function across all categories of predicted genetic risk. For any given level of pack-years of smoking exposure, however, those at highest genetic risk showed lower FEV₁/FVC than those with the lowest predicted genetic risk. The effects of heavy smoking and being in the highest decile of predicted genetic risk were greater than would be expected based on the additive effects of both risk factors. These results support the idea that diminished pulmonary function (a measure of airflow obstruction) are, at least partially, due to gene-by-smoking interactions, and those in higher genetic risk categories are more susceptible to the deleterious effects of smoking.

Compared to previous studies, our study included more participants, leveraged a more powerful measure of genetic predisposition for low lung function (i.e. PRS), examined three different measures of smoking exposure (pack-years, ever smoking, current smoking), and examined ‘norms of reactions’ for those in the highest compared to the lowest deciles of predicted genetic risk. Our findings are consistent with Aschard et al.²³ who reported an interaction between ever-

smoking status and a genetic risk score for FEV₁/FVC based on 26 different variants. In a family-based study of the rs28929474 variant (Z allele) in *SERPINA1*, which leads to alpha-1 antitrypsin deficiency (AATD) and greatly increased risk of emphysema, there was a significant genotype-by-smoking interaction on FEV₁³⁰. A strong smoking interaction of rs28929474 heterozygote status with ever-smoking on lung function and COPD has been reported in UK Biobank³¹. Genome-wide interaction studies have confirmed the interaction of smoking with rs28927474 on risk to COPD¹¹, and identified gene-by-smoking interactions at several other loci⁸⁻¹⁴.

By contrast, Shrine et al.¹⁹ constructed a genetic risk score from 279 variants shown to influence lung function, but did not observe any evidence of interaction with ever-smoking status on FEV₁/FVC; further, the authors reported an interaction with moderate-to-severe COPD status in the opposite direction as expected. A prior GWAS selected ~50,000 individuals with low, average, and high FEV₁, reported no gene-smoking interactions at genome-wide significance³². The reasons for disparate findings between these studies and the current one are unclear, but may represent the effects of different loci with varying degrees of interaction effect^{4,19,20,33}. While the PRS used in the current study was derived from the Shrine et al. GWAS results¹⁹, not all variants reached genome-wide significance, and consequently, included many more variants (~2.5 M); this more refined level of predicted genetic risk may have provided the power to detect gene-by-smoking interactions.

Smoking was detrimental even to those with low predicted genetic risk, and the effects were greater for those with high predicted genetic risk. For any given level of pack-years of smoking,

those in the highest decile had lower FEV₁/FVC compared to those in the lowest decile of predicted genetic risk. These findings are in contrast to observations in cardiovascular disease, where the association between smoking and coronary heart disease was greater for those in the lowest compared to the highest tertile of a PRS²². This difference may reflect that many individuals can develop coronary disease in the absence of cigarette smoking, and that smoking is a greater risk factor for those with low polygenic risk for coronary disease. Meanwhile, airflow obstruction primarily occurs in the setting of cigarette smoking exposure. Furthermore, those with low predicted genetic risk and high smoking exposure had similar risk for low FEV₁/FVC as those with high genetic risk and low smoking exposure. Taken together, these results emphasize that abstaining from smoking is crucial to preventing obstructive lung disease regardless of an individual's predicted genetic risk, and that those in the highest risk groups might benefit from intensive smoking cessation measures with respect to the phenotypes examined in this study.

Our results suggest that the PRS includes variants that represent biological pathways by which smoking exerts deleterious effects. Some of these variants may act to confer resilience^{34,35} or susceptibility to the effects of cigarette smoke. Further investigation into the role of specific variants in susceptibility to cigarette smoke may be performed by inclusion of other “Omics” data types and leveraging network analytic techniques could help elucidate mechanisms of resilience and susceptibility. For example, the effect of occupational exposures was modified by rs9931086 in *SLC38A8* on FEV₁, and network analyses suggested inflammatory processes involving *CTLA-4*, *HDAC*, and *PPAR-alpha*, may provide mechanistic links for the observed interaction³⁶; however, this was a small study that needs replication.

Strengths of this study include use of a large volunteer cohort, utilizing the most powerful measure for genetic risk for low lung function available to date (i.e. a genome-wide PRS), and comparing individuals at extremes of predicted genetic risk. Limitations, inherent to study design, include that the UK Biobank is a single cohort observed in cross-section. Examining the effects of gene-by-smoking interactions on incident COPD should be pursued. We were not able to model the time-varying effects of smoking exposure. The PRS was partially developed using samples from UK Biobank, leading to overfitting of the PRS with respect to spirometric measures; while this issue should not affect interaction assessments, these results should ideally be replicated in future studies. However, the strong effect sizes and robustness to stratified and transformed analyses does lend confidence to our results. We included European-ancestry participants only because the PRS was derived solely from Europeans. Identification of causal variants and genetic prediction in single ancestry populations demonstrate limited portability to multi-ancestry populations^{37,38}. The 279 lung function variants from Shrine et al.¹⁹ was curated to ensure variants for smoking behavior were excluded, but the PRS used in the current study included ~2.5 million variants and was not similarly curated. Including variants that are causal for smoking behavior could bias the interaction term³⁹; however, there was a very weak correlation between this PRS and smoking exposure in UK Biobank, and previously no correlation with smoking in case-control cohorts was observed²⁰, suggesting that the PRS used in the current study largely reflects the genetics of lung function. Finally, we emphasize that smoking cessation is the main preventive intervention to airflow obstruction regardless of genetic susceptibility. Our interaction findings further support a more intensive smoking cessation is required to those with high genetic susceptibility to diminished lung function. The effectiveness of targeted genetically-informed smoking cessation interventions is unclear, though there is

evidence that knowledge of genetic risk for AATD can increase smoking cessation⁴⁰. Clinical utility of the PRS will depend on dissecting biological mechanisms of susceptibility to the harmful effects of smoking.

In conclusion, diminished FEV₁/FVC and airflow obstruction, which are characteristic of COPD, may be partially attributable to gene-by-smoking interactions. As expected, smoking was harmful across all genetic risk groups, but worse for those in the highest decile of predicted genetic risk. Large-scale replication and further investigations into mechanisms of interaction are needed.

References

1. Vogelmeier CF, Criner GJ, Martinez FJ, 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*. 2017;195(5):557-582.
doi:10.1164/rccm.201701-0218PP
2. Bhatt SP, Balte PP, Schwartz JE, et al. Discriminative Accuracy of FEV₁:FVC Thresholds for COPD-Related Hospitalization and Mortality. *JAMA*. 2019;321(24):2438.
doi:10.1001/jama.2019.7233
3. Rennard SI, Vestbo J. COPD: the dangerous underestimate of 15%. *Lancet (London,*

- England*). 2006;367(9518):1216-1219. doi:10.1016/S0140-6736(06)68516-4
4. Fletcher,C,Peto R, Tinker C SF. *The Natural History of Chronic Bronchitis and Emphysema: An Eight-Year Study of Early Chronic Obstructive Lung Disease in Working Men in London*. Oxford University Press, Oxford; 1976.
 5. Wilk JB, Chen T-H, Gottlieb DJ, et al. A genome-wide association study of pulmonary function measures in the Framingham Heart Study. *PLoS Genet*. 2009;5(3):e1000429. doi:10.1371/journal.pgen.1000429
 6. Palmer LJ, Knuiman MW, Divitini ML, et al. Familial aggregation and heritability of adult lung function: results from the Busselton Health Study. *Eur Respir J*. 2001;17(4):696-702. <http://www.ncbi.nlm.nih.gov/pubmed/11401066>.
 7. Zhou JJ, Cho MH, Castaldi PJ, Hersh CP, Silverman EK, Laird NM. Heritability of chronic obstructive pulmonary disease and related phenotypes in smokers. *Am J Respir Crit Care Med*. 2013;188(8):941-947. doi:10.1164/rccm.201302-0263OC
 8. de Jong K, Boezen HM, Hacken NH, Postma DS, Vonk JM. GST-omega genes interact with environmental tobacco smoke on adult level of lung function. *Respir Res*. 2013;14(1):83. doi:10.1186/1465-9921-14-83
 9. de Jong K, Vonk JM, Imboden M, et al. Genes and pathways underlying susceptibility to impaired lung function in the context of environmental tobacco smoke exposure. *Respir Res*. 2017;18(1):142. doi:10.1186/s12931-017-0625-7
 10. Kim S, Kim H, Cho N, et al. Identification of FAM13A gene associated with the ratio of FEV1 to FVC in Korean population by genome-wide association studies including gene-environment interactions. *J Hum Genet*. 2015;60(3):139-145. doi:10.1038/jhg.2014.118

11. Kim W, Prokopenko D, Sakornsakolpat P, et al. Genome-wide Gene-by-smoking Interaction Study of Chronic Obstructive Pulmonary Disease. *Am J Epidemiol*. October 2020. doi:10.1093/aje/kwaa227
12. Park B, Koo S-M, An J, et al. Genome-wide assessment of gene-by-smoking interactions in COPD. *Sci Rep*. 2018;8(1):9319. doi:10.1038/s41598-018-27463-5
13. Park B, An J, Kim W, et al. Effect of 6p21 region on lung function is modified by smoking: a genome-wide interaction study. *Sci Rep*. 2020;10(1):13075. doi:10.1038/s41598-020-70092-0
14. Hancock DB, Artigas MS, Gharib SA, et al. Genome-Wide Joint Meta-Analysis of SNP and SNP-by-Smoking Interaction Identifies Novel Loci for Pulmonary Function. Gibson G, ed. *PLoS Genet*. 2012;8(12):e1003098. doi:10.1371/journal.pgen.1003098
15. Busch R, Hobbs BD, Zhou J, 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. doi:10.1165/rcmb.2016-0331OC
16. Wain L V, Shrine N, Artigas MS, et al. Genome-wide association analyses for lung function and chronic obstructive pulmonary disease identify new loci and potential druggable targets. *Nat Genet*. 2017;49(3):416-425. doi:10.1038/ng.3787
17. Hobbs BD, de Jong K, Lamontagne M, et al. Genetic loci associated with chronic obstructive pulmonary disease overlap with loci for lung function and pulmonary fibrosis. *Nat Genet*. 2017;49(3):426-432. doi:10.1038/ng.3752
18. Sakornsakolpat P, Prokopenko D, Lamontagne M, et al. Genetic landscape of chronic obstructive pulmonary disease identifies heterogeneous cell-type and phenotype

- associations. *Nat Genet.* 2019;51(3):494-505. doi:10.1038/s41588-018-0342-2
19. Shrine N, Guyatt AL, Erzurumluoglu AM, et al. New genetic signals for lung function highlight pathways and chronic obstructive pulmonary disease associations across multiple ancestries. *Nat Genet.* 2019;51(3):481-493. doi:10.1038/s41588-018-0321-7
 20. Moll M, Sakornsakolpat P, Shrine N, et al. Chronic obstructive pulmonary disease and related phenotypes: polygenic risk scores in population-based and case-control cohorts. *Lancet Respir Med.* 2020;8(7):696-708. doi:10.1016/S2213-2600(20)30101-6
 21. Peyrot WJ, Van der Auwera S, Milaneschi Y, et al. Does Childhood Trauma Moderate Polygenic Risk for Depression? A Meta-analysis of 5765 Subjects From the Psychiatric Genomics Consortium. *Biol Psychiatry.* 2018;84(2):138-147.
doi:10.1016/j.biopsych.2017.09.009
 22. Hindy G, Wiberg F, Almgren P, Melander O, Orho-Melander M. Polygenic Risk Score for Coronary Heart Disease Modifies the Elevated Risk by Cigarette Smoking for Disease Incidence. *Circ Genomic Precis Med.* 2018;11(1). doi:10.1161/CIRCGEN.117.001856
 23. Aschard H, Tobin MD, Hancock DB, et al. Evidence for large-scale gene-by-smoking interaction effects on pulmonary function. *Int J Epidemiol.* January 2017:dyw318.
doi:10.1093/ije/dyw318
 24. Bycroft C, Freeman C, Petkova D, et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature.* 2018;562(7726):203-209. doi:10.1038/s41586-018-0579-z
 25. Mak TSH, Porsch RM, Choi SW, Zhou X, Sham PC. Polygenic scores via penalized regression on summary statistics. *Genet Epidemiol.* 2017;41(6):469-480.
doi:10.1002/gepi.22050

26. O'Brien RM. A caution regarding rules of thumb for variance inflation factors. *Qual Quant.* 2007;41(5):673-690. doi:10.1007/s11135-006-9018-6
27. Gupta AP, Lewontin RC. A STUDY OF REACTION NORMS IN NATURAL POPULATIONS OF DROSOPHILA PSEUDOOBSCURA. *Evolution (N Y).* 1982;36(5):934-948. doi:10.1111/j.1558-5646.1982.tb05464.x
28. Kassambara A. rstatix. <https://rpkgs.datanovia.com/rstatix/index.html>. Published 2020.
29. Wan ES, Castaldi PJ, Cho MH, et al. Epidemiology, genetics, and subtyping of preserved ratio impaired spirometry (PRISm) in COPD Gene. *Respir Res.* 2014;15:89. doi:10.1186/s12931-014-0089-y
30. Silverman EK, Chapman HA, Drazen JM, et al. Genetic epidemiology of severe, early-onset chronic obstructive pulmonary disease. Risk to relatives for airflow obstruction and chronic bronchitis. *Am J Respir Crit Care Med.* 1998;157(6 Pt 1):1770-1778. doi:10.1164/ajrccm.157.6.9706014
31. Katherine A Fawcett, Kijoung Song, Guoqing Qian, Aliko-Eleni Farmaki, Richard Packer, Catherine John, Nick Shrine, Raquel Granell, Sue Ring, Nicholas J Timpson, Laura M Yerges-Armstrong, Richard Eastell, Louise V Wain, Robert A Scott, Martin D Tobin IPH. Pleiotropic effects of heterozygosity for the SERPINA1 Z allele in the UK Biobank. *medRxiv.* 2020. doi:10.1101/2020.06.04.20115923
32. Wain L V, Shrine N, Miller S, 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.* 2015;3(10):769-781. doi:10.1016/S2213-2600(15)00283-0

33. McGeachie MJ, Yates KP, Zhou X, et al. Patterns of Growth and Decline in Lung Function in Persistent Childhood Asthma. *N Engl J Med*. 2016;374(19):1842-1852. doi:10.1056/NEJMoa1513737
34. Elbau IG, Cruceanu C, Binder EB. Genetics of Resilience: Gene-by-Environment Interaction Studies as a Tool to Dissect Mechanisms of Resilience. *Biol Psychiatry*. 2019;86(6):433-442. doi:10.1016/j.biopsych.2019.04.025
35. Tuder RM. Bringing Light to Chronic Obstructive Pulmonary Disease Pathogenesis and Resilience. *Ann Am Thorac Soc*. 2018;15(Supplement_4):S227-S233. doi:10.1513/AnnalsATS.201808-583MG
36. Liao S-Y, Lin X, Christiani DC. Gene-environment interaction effects on lung function- a genome-wide association study within the Framingham heart study. *Environ Heal*. 2013;12(1):101. doi:10.1186/1476-069X-12-101
37. Wyss AB, Sofer T, Lee MK, et al. Multiethnic meta-analysis identifies ancestry-specific and cross-ancestry loci for pulmonary function. *Nat Commun*. 2018;9(1):2976. doi:10.1038/s41467-018-05369-0
38. Duncan L, Shen H, Gelaye B, et al. Analysis of polygenic risk score usage and performance in diverse human populations. *Nat Commun*. 2019;10(1):3328. doi:10.1038/s41467-019-11112-0
39. Dudbridge F, Fletcher O. Gene-environment dependence creates spurious gene-environment interaction. *Am J Hum Genet*. 2014;95(3):301-307. doi:10.1016/j.ajhg.2014.07.014
40. Carpenter MJ, Strange C, Jones Y, et al. Does genetic testing result in behavioral health

change? Changes in smoking behavior following testing for alpha-1 antitrypsin deficiency.

Ann Behav Med. 2007;33(1):22-28. doi:10.1207/s15324796abm3301_3

Table Legends

Table 1: Characteristics of study participants. All participants were of European ancestry with high-quality genetic data and spirometry. Controls included those who met the spirometry criteria for Global Initiative for Chronic Obstructive Lung Disease (GOLD)¹ spirometry grade 0/1, and cases included GOLD 2-4 participants (i.e. those with moderate-to-severe COPD). PRISm = preserved ratio with impaired spirometry, defined as $FEV_1/FVC > 0.7$ with FEV_1 % predicted $< 80\%$ ²⁹. P-values were calculated with ANOVA or Kruskal-Wallis tests across three categories for normal and non-normal variables, respectively.

Table 2: Multivariable linear regression of the form $FEV_1/FVC \sim PRS + \text{Pack years of smoking} + PRS * \text{pack-years of smoking} + \text{covariates}$. Covariates include age, age², sex, height, genotyping array, principal components of genetic ancestry. Pack years of smoking is included as a categorical variable with ≤ 10 pack-years as the reference group. PRS = polygenic risk score.

Figure Legends

Figure 1: The relationship between PRS and FEV_1/FVC by pack years of smoking.

Figure 2: Joint effects of pack years of smoking category and PRS decile. PRS = polygenic risk score.

Figure 3: Norms of reaction for those with high versus low predicted genetic risk for COPD. The slopes of these lines were significantly different in analysis of covariance ($p < 0.0001$).

Table 1: Characteristics of study participants. All participants were of European ancestry with high-quality genetic data and spirometry. Controls included those who met the spirometry

criteria for Global Initiative for Chronic Obstructive Lung Disease (GOLD)¹ spirometry grade 0/1, and cases included GOLD 2-4 participants (i.e. those with moderate-to-severe COPD).

PRISm = preserved ratio with impaired spirometry, defined as $FEV_1/FVC > 0.7$ with FEV_1 % predicted $< 80\%$ ²⁹. P-values were calculated with ANOVA or Kruskal-Wallis tests across three categories for normal and non-normal variables, respectively.

| | <i>Overall</i> |
|--------------------------------------|--------------------|
| n | 319730 |
| Age (mean (SD)) | 56.45 (8.02) |
| Sex (No. (%) male) | 141864 (44.4) |
| Pack years of smoking (median [IQR]) | 0.00 [0.00, 11.00] |
| Former/Never smoker (No. (%)) | 287445 (89.9) |
| Current smoker (No. (%)) | 32242 (10.1) |
| Ever smoker (No (%)) | 146679 (45.9) |
| FEV1 % predicted (mean (SD)) | 92.17 (16.01) |
| FEV1/FVC (mean (SD)) | 0.76 (0.06) |
| Decile of polygenic risk score (%) | |
| Lowest | 31973 (10.0) |
| Highest | 31973 (10.0) |

Table 2: Multivariable linear regression of the form $FEV_1/FVC \sim PRS + \text{Pack years of smoking} + PRS * \text{pack-years of smoking} + \text{covariates}$. Covariates include age, age², sex, height, genotyping array, principal components of genetic ancestry. Pack years of smoking is included as a categorical variable with ≤ 10 pack-years as the reference group. PRS = polygenic risk score.

| <i>Variable</i> | <i>Beta (95% CI)</i> | <i>P</i> |
|---------------------------------|---------------------------|-----------|
| PRS | -0.03 (-0.031--0.03) | < 1e-256 |
| 11-20 pack years | -0.011 (-0.012--0.01) | 1.70E-225 |
| 21-30 pack years | -0.014 (-0.015--0.014) | < 1e-256 |
| 31-40 pack years | -0.026 (-0.027--0.025) | < 1e-256 |
| 41-50 pack years | -0.034 (-0.035--0.033) | < 1e-256 |
| >50 pack years | -0.04 (-0.041--0.038) | < 1e-256 |
| PRS X 11-20 pack years category | -0.0038 (-0.0046--0.0031) | 4.80E-24 |
| PRS X 21-30 pack years category | -0.0068 (-0.0077--0.006) | 1.60E-56 |
| PRS X 31-40 pack years category | -0.013 (-0.014--0.012) | 5.50E-133 |
| PRS X 41-50 pack years category | -0.015 (-0.017--0.014) | 1.90E-101 |
| PRS X >50 pack years category | -0.017 (-0.019--0.016) | 3.60E-134 |

Figure 1: The relationship between PRS and FEV1/FVC by pack years of smoking.

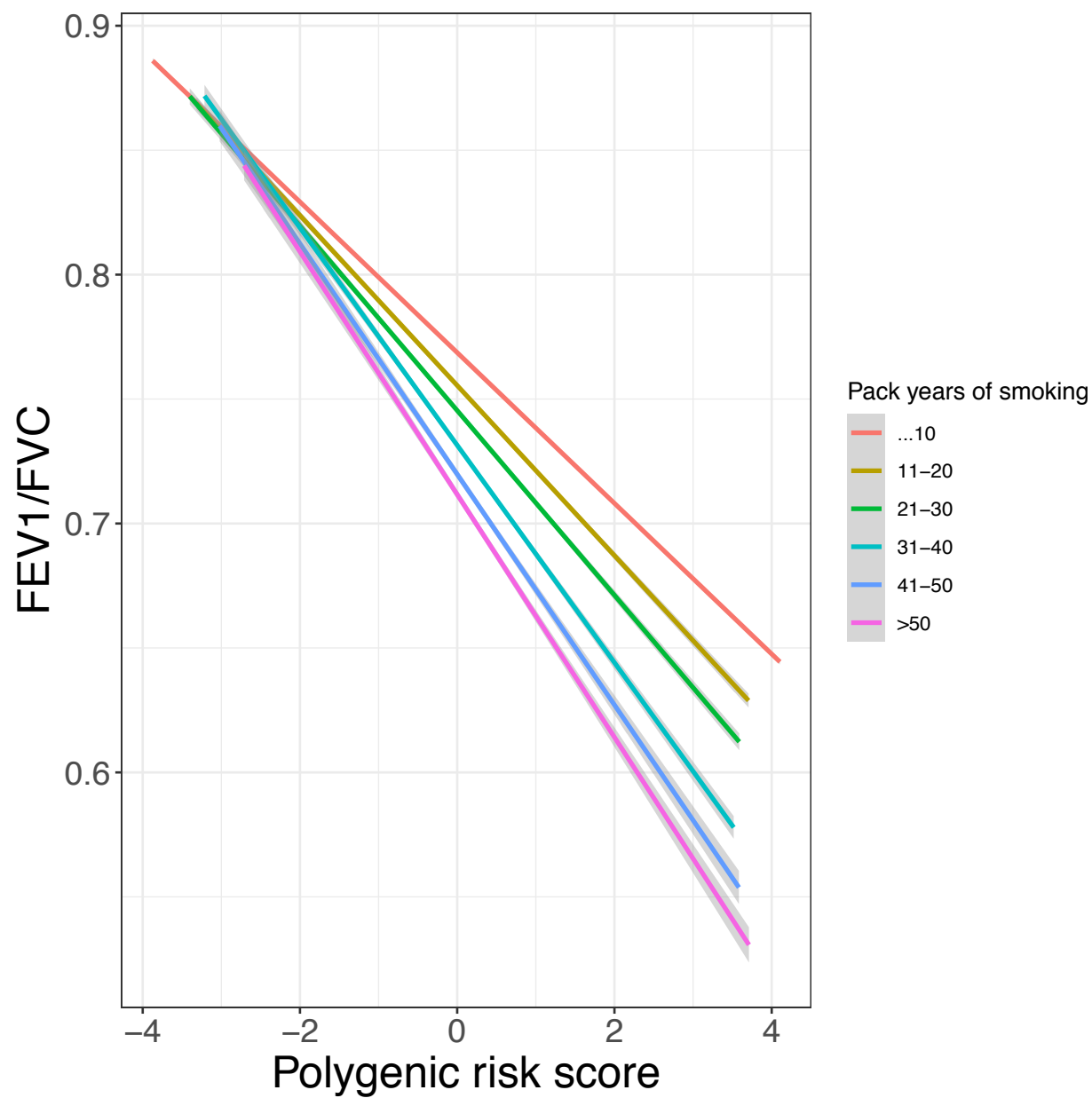


Figure 2: Joint effects of pack years of smoking category and PRS decile. PRS = polygenic risk score.

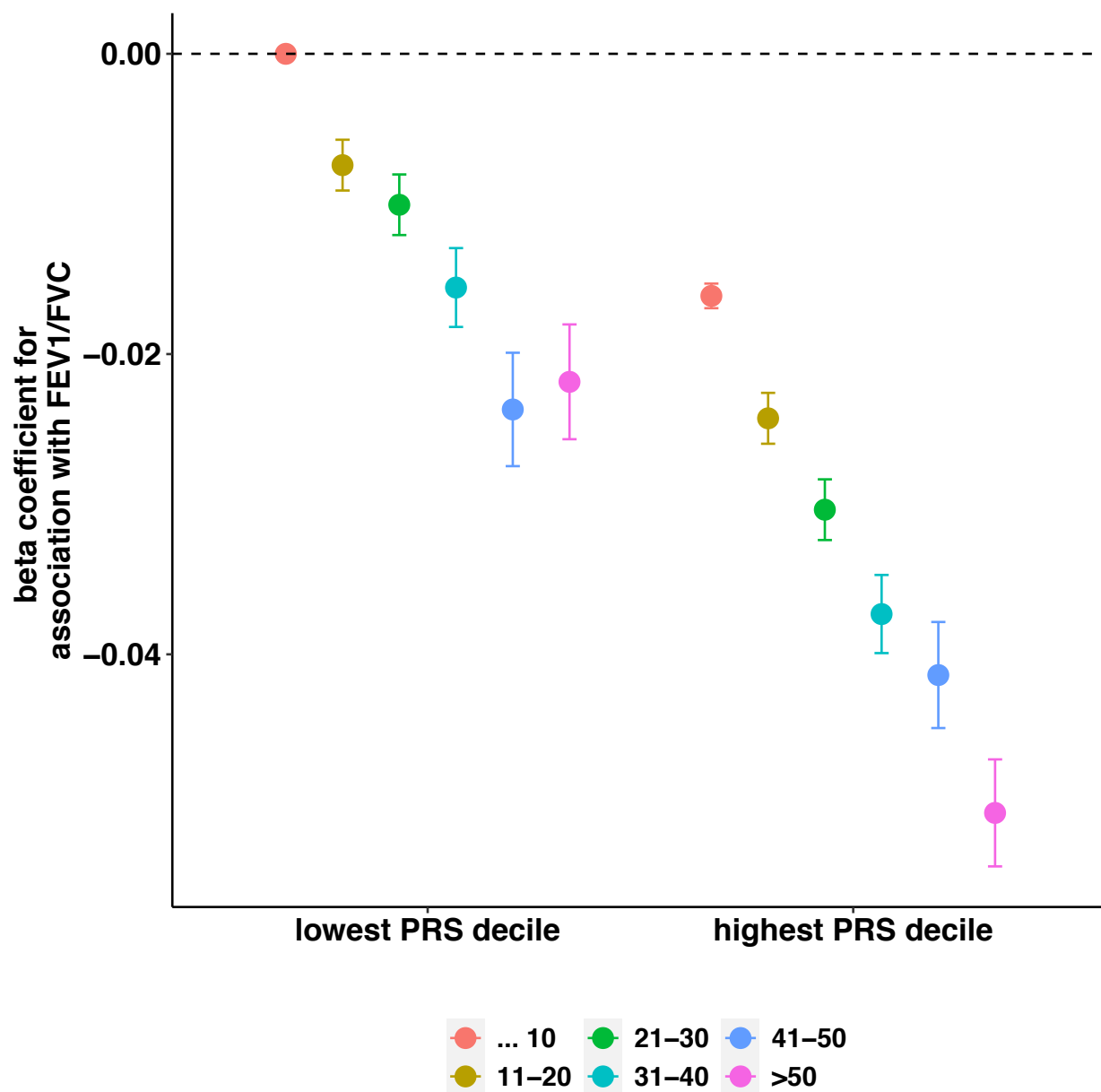


Figure 3: Norms of reaction for those with high versus low predicted genetic risk for COPD. The slopes of these lines were significantly different in analysis of covariance ($p < 0.0001$).

