Published in final edited form as:

Genet Epidemiol. 2020 October ; 44(7): 748–758. doi:10.1002/gepi.22331.

Dissecting The Genetic Overlap Of Smoking Behaviors, Lung Cancer, And Chronic Obstructive Pulmonary Disease: A Focus On Nicotinic Receptors and Nicotine Metabolizing Enzyme

Michael J. Bray^a, Li-Shiun Chen^{a,b}, Louis Fox^a, Dana B. Hancock^c, Robert C. Culverhouse^{d,e}, Sarah M. Hartz^a, Eric O. Johnson^{c,f}, Mengzhen Liu^g, James D. McKay^h, Nancy L. Saccone^{i,e}, John E. Hokanson^j, Scott I. Vrieze^g, Rachel F. Tyndale^k, Timothy B. Baker^l, Laura J. Bierut^{a,b}

^aDepartment of Psychiatry, Washington University School of Medicine, St. Louis, MO

^bThe Alvin J. Siteman Cancer Center, Washington University School of Medicine, St. Louis, MO.

^cCenter for Omics Discovery and Epidemiology, Behavioral Health Research Division, RTI International, Research Triangle Park, NC.

^dDepartment of Medicine, Washington University School of Medicine, St. Louis, MO.

^eDivision of Biostatistics, Washington University School of Medicine, St. Louis, MO.

^fFellow Program, RTI International, Research Triangle Park, NC.

^gDepartment of Psychology, University of Minnesota, Minnesota, MN.

^hInternational Agency for Research on Cancer, World Health Organization, Lyon, France.

ⁱDepartment of Genetics, Washington University School of Medicine, St. Louis, MO.

^jDepartment of Epidemiology, University of Colorado, Denver, Aurora, CO.

^kCentre for Addiction and Mental Health and Department of Psychiatry, University of Toronto, Toronto, Ontario, Canada; Department of Pharmacology & Toxicology, University of Toronto, Toronto, Ontario, Canada.

^lCenter for Tobacco Research and Intervention, University of Wisconsin, School of Medicine and Public Health, Madison, WI.

Corresponding Author: Laura J. Bierut, PhD, MS, Washington University School of Medicine, 660 South Euclid, Campus Box 8134, St. Louis, MO 63110, laura@wustl.edu, Phone: 314-362-3492.

CONFLICTS OF INTEREST STATEMENT

Laura J. Bierut and the spouse of Nancy L. Saccone are listed as inventors on Issued U.S. Patent 8,080,371, “Markers for Addiction” covering the use of certain SNPs in determining the diagnosis, prognosis, and treatment of addiction. Rachel F Tyndale has consulted for Quinn Emanuel and Ethismos Research Inc. There are no other conflicts of interest to disclose.

DATA AVAILABILITY STATEMENT

Correspondence and requests for data accessibility of the genome-wide association study (GWAS) and Sequencing Consortium of Alcohol and Nicotine use (GSCAN) GWAS summary results should be addressed to D.J.L or S.V. from Liu et al. (2019) (Liu et al., 2019). GWAS summary statistics for the lung cancer risk study (McKay et al., 2017) are available in the database of Genotypes and Phenotypes (dbGaP) repository, https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs001273.v1.p1, under accession number [phs001273.v1.p1]. GWAS summary statistics of COPD risk study (Hobbs et al., 2017) are available in the dbGaP repository, https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000179.v5.p2, under accession number [phs000179.v5.p2].

Abstract

Smoking is a major contributor to lung cancer and chronic obstructive pulmonary disease (COPD). Two of the strongest genetic associations of smoking-related phenotypes are the chromosomal regions 15q25, encompassing the nicotinic acetylcholine receptor subunit genes *CHRNA5-CHRNA3-CHRNA4*, and 19q13.2, encompassing the nicotine metabolizing gene *CYP2A6*. In this study, we examined genetic relations between cigarettes smoked per day, smoking cessation, lung cancer, and COPD. Data consisted of genome-wide association study (GWAS) summary results. Genetic correlations were estimated using linkage disequilibrium (LD) score regression software. For each pair of outcomes, Z-score-z-score (ZZ) plots were generated. Overall, heavier smoking and decreased smoking cessation showed positive genetic associations with increased lung cancer and COPD risk. The chromosomal region 19q13.2, however, showed a different correlational pattern. For example, the effect allele-C of the sentinel SNP (rs56113850) within *CYP2A6* was associated with an increased risk of heavier smoking (z-score=19.2; $p=1.10 \times 10^{-81}$), lung cancer (z-score=8.91; $p=5.02 \times 10^{-19}$), and COPD (z-score=4.04; $p=5.40 \times 10^{-5}$). Surprisingly, this allele-C (rs56113850) was associated with increased smoking cessation (z-score=-8.17; $p=2.52 \times 10^{-26}$). This inverse relationship highlights the need for additional investigation to determine how *CYP2A6* variation could increase smoking cessation while also increasing the risk of lung cancer and COPD likely through increased cigarettes smoked per day.

Keywords

Genetics; Addiction; Smoking; Smoking Cessation

INTRODUCTION

Cigarette smoking is a leading cause of preventable death worldwide (World Health Organization, 2017) and is a major contributor to lung cancer (Doll, Peto, Boreham, & Sutherland, 2004; World Health Organization, 2017), chronic obstructive pulmonary disease (COPD) (Mannino & Buist, 2007), as well as many other diseases (World Health Organization, 2017). Smoking is most often initiated in adolescence and early adulthood (Edwards, Carter, Peace, & Blakely, 2013). Smoking cessation reduces the risks of developing lung cancer (Doll et al., 2004) and COPD (Mannino & Buist, 2007) later in adulthood, but because of the addictive nature of nicotine, smoking cessation is difficult.

Heritability estimates are high for nicotine use disorder (Vink, , Ming Li, and Lessov). The number of cigarettes smoked per day and ability to successfully quit smoking vary between individuals, and part of the variability of these behaviors can be explained by underlying genetic differences (Liu et al., 2019). In addition, the risk of developing lung cancer and COPD is in part genetically based (Mucci et al., 2016; Zhou et al., 2013). Heritability estimates of lung cancer (up to ~18–41%) (Hjelmberg et al., 2017; Mucci et al., 2016; Sampson et al., 2015) and COPD (up to ~38%) (Zhou et al., 2013). However, heritability estimates of lung cancer among never-smokers remains unclear (Hjelmberg et al., 2017; Sampson et al., 2015). The largest, by far, GWAS of smoking phenotypes (N up to 1.2 million individuals) identified hundreds genome-wide significant loci for smoking initiation, age of smoking initiation, cigarettes smoked per day, and smoking cessation (Liu et al.,

2019). Of these loci, two had the smallest p values, by many orders of magnitude: chromosome 15q25.1 (lead SNP $p=3.12 \times 10^{-286}$) and 19q13.2 (lead SNP $p=4.01 \times 10^{-99}$), which were associated with number of cigarettes smoked per day (Liu et al., 2019). The chromosomal region 15q25 contains the nicotinic receptor subunit gene cluster *CHRNA5-CHRNA3-CHRNA4*, and the chromosomal region 19.13.2 contains the primary nicotine metabolizing gene, *CYP2A6* (Hancock, Markunas, Bierut, & Johnson, 2018). Importantly, genome-wide association studies (GWAS) implicate these two genetic loci with multiple other smoking-related conditions (Hancock et al., 2018) including lung cancer (McKay et al., 2017) and COPD (Hobbs et al., 2017). The genetic associations of smoking-related disease such as lung cancer and COPD for chromosomal region 15q25 are likely mediated by smoking behaviors. This mediation mechanism is supported by the evidence that the chromosomal region 15q25 is not associated with lung cancer and COPD among never-smokers (Gabrielsen, Romundstad, Langhammer, Krokan, & Skorpén, 2013; Liu et al., 2019; Timofeeva et al., 2011; Truong et al., 2010).

Given the complex relation between smoking behaviors and smoking-related illnesses, it is important to dissect the relation between the underlying genetic contributions to these conditions. We examined the overall genetic correlation between smoking-related behaviors, number of cigarettes smoked per day and smoking cessation, and smoking related diseases, lung cancer and COPD, using GWAS summary statistics, and in addition, we specifically focused on the two regions (15q25 and 19q13.2) most strongly associated with smoking behaviors.

MATERIALS AND METHODS

Study Population

Smoking—GWAS summary statistics of smoking quantity (measured in cigarettes smoked per day) and smoking cessation (defined as current smoker versus former smoker) were obtained from the GWAS and Sequencing Consortium of Alcohol and Nicotine use (GSCAN) study (Liu et al., 2019). The GSCAN study performed a meta-analysis from many contributing studies (Liu et al., 2019). Number of cigarettes smoked per day differed between studies with most asking about current number of cigarettes smoked per day, others asking about maximum number of cigarettes smoked per day. Among former smokers, the typical query asked about average number of cigarettes smoked per day. A definition of a former smoker which defined the smoking cessation phenotype also varied between studies (Liu et al., 2019). Most studies allowed self identification as a former smoker with no specific time frame requirement. Other studies required a time period when a person had not smoked a cigarette. Smoking cessation motivated by medical illnesses is not recorded. Because of these varying definitions cigarettes per day and motivations of former smoker and smoking, there is variation between each contributing study and heterogeneity within the smoking behavior phenotypes.

The GSCAN study comprised individuals of European ancestry with 337,334 individuals contributing data for cigarettes smoked per day among smokers and 547,219 individuals contributing data for smoking cessation (Liu et al., 2019). SNP heritability estimates of cigarettes smoked per day and smoking cessation and genetic correlations between these

outcomes have previously been performed by Liu and colleagues (Liu et al., 2019). Liu and colleagues did not, however, perform heritability estimates of lung cancer and COPD or genetic correlations between the smoking behaviors and risk of developing lung cancer and COPD.

Lung Cancer—GWAS summary statistics for lung cancer risk were obtained from McKay and colleagues (2017) (McKay et al., 2017). This GWAS was based on the collaboration of the Transdisciplinary Research of Cancer in Lung (TRICL) of the International Lung Cancer Consortium (ILCCO). The data consisted of 26 study populations comprising 29,266 lung cancer cases and 56,450 controls of European ancestry (McKay et al., 2017). SNP heritability estimates of lung cancer from this study are at approximately 9% (McKay et al., 2017).

COPD—GWAS summary statistics of COPD risk were obtained from Hobbs et al. (2017) (Hobbs et al., 2017) via the International COPD Genetics Consortium. Briefly, the GWAS summary statistics obtained from the International COPD Genetics Consortium comprising of 63,192 individuals, mostly of European ancestry (15,256 COPD cases; 47,936 controls) (Hobbs et al., 2017) Do you have heritability?.

Statistical Analyses

Genetic data consisted of GWAS summary statistics from each study for the following phenotypes: number of cigarettes smoked per day, smoking cessation, lung cancer, and COPD. Consensus SNPs, variants present in all contributing GWAS, were used for each analysis resulting in a total of 4,993,689 SNPs for all phenotypes from all datasets. SNP heritability and genetic correlations were estimated using linkage disequilibrium (LD) Score Regression software (Bulik-Sullivan et al., 2015). In order to evaluate independent and specific genetic signals for each pair of phenotypes, we created Z-score-Z-score (ZZ) plots. ZZ plots were generated wherein each SNP is represented by a point with the z-scores for the two outcomes as the coordinates. The z-score of each SNP from each GWAS summary statistics were unweighted. The chromosomal 15q25.1 and 19q13.2 regions were highlighted because of their strong associations with smoking behaviors and smoking-related illnesses. These selected genomic regions included each selected gene and its 0.5 megabase (Mb) flanking regions. Genome build 37 was used to determine gene boundaries and the corresponding flanking regions: chromosomal region 15q25.1 having 2,116 SNPs encompassing *CHRNA5*, *CHRNA3*, and *CHRNA4* [boundary with 0.5 Mb - 78,357,862–79,433,587]; chromosomal region 19q13.2 having 1,992 SNPs encompassing *EGLN2*, *CYP2A6*, and *CYP2B6* [boundary with 0.5 Mb - 40,805,048–42,024,308]. Genetic correlations of the selected regions were performed using Stata/SE (College Station, Texas).

RESULTS

Strong, positive genetic correlations exist between smoking-related phenotypes

SNP genetic correlations and heritability estimates are shown in Figure 1a. Higher number of cigarettes smoked per day was correlated with decreased smoking cessation with $r = 0.43$ ($p = 1.03 \times 10^{-29}$). Correlations between the smoking phenotypes and lung cancer were

strong: number of cigarettes per day and lung cancer ($r = 0.59$; $p = 1.14 \times 10^{-10}$) and decreased smoking cessation and lung cancer ($r = 0.49$; $p = 1.52 \times 10^{-18}$). Relative to the correlations observed between smoking and lung cancer, the genetic correlations between the smoking phenotypes and COPD were more modest: number cigarettes per day and COPD ($r = 0.22$; $p = 1.70 \times 10^{-3}$) and decreased smoking cessation and COPD ($r = 0.18$; $p = 1.20 \times 10^{-3}$). There was also a modest genetic correlation between lung cancer and COPD ($r = 0.19$; $p = 0.042$).

Gene regions encompassing *CHRNA5-CHRNA3-CHRNA4* and *CYP2A6* correlate with smoking behaviors, lung cancer, and COPD

ZZ-plots reveal strong region-specific correlations (Figure 2a-f). SNPs contained in chromosomal region 15q25.1, including the nicotinic receptor subunits cluster, reveal a strong genetic correlation for all the smoking-related phenotypes (Figure 3a). The correlation of SNPs for number of cigarettes smoked per day and lung cancer was 0.98, and all the other correlations were from 0.77 to 0.94. These correlations were in the expected direction across smoking-related phenotypes with SNPs that increase the heaviness of smoking measured by number of cigarettes per day also being associated with decreased smoking cessation and increased risks of lung cancer and COPD.

The chromosomal region 19q13.2 containing the nicotine metabolizing gene *CYP2A6* showed a different pattern of correlations (Figure 3b). The SNPs comprising the *CYP2A6* gene region demonstrated positive correlations between the number of cigarettes smoked per day and increased risk of lung cancer and COPD, ranging from 0.67 to 0.81. In contrast, these SNPs demonstrated an inverse correlation with smoking cessation, i.e., SNPs associated with increased risk of heavier smoking, lung cancer, and COPD were also associated with increased smoking cessation. As an example of this correlation, the effect allele-C of the sentinel SNP (rs56113850) within *CYP2A6* was associated with greater cigarettes smoked per day (z-score = 19.2; $p = 1.10 \times 10^{-81}$), an increased risk of lung cancer (z-score = 8.91; $p = 5.02 \times 10^{-19}$) and an increased risk of COPD (z-score = 4.04; $p = 5.40 \times 10^{-5}$) (Table 1). Nevertheless, the allele-C of rs56113850 was also associated with increased smoking cessation (z-score = -8.17; $p = 2.52 \times 10^{-26}$).

Unique regions associated with lung cancer and COPD that are not associated with smoking behaviors are apparent.

ZZ-plots reveal loci that are genome-wide significant loci associated specifically with lung cancer and not cigarettes smoked per day or smoking cessation (Figure 2b,2d). In addition, there are loci associating with COPD that are not associated with cigarettes smoked per day (Figure 2c, 2e). Lastly, there are unique genetic loci associating with each lung cancer and COPD (Figure 2f).

The genetic correlation between smoking related phenotypes remains strong after censoring gene regions encompassing *CHRNA5-CHRNA3-CHRNA4* and *CYP2A6*

After censoring SNPs in the *CHRNA5-CHRNA3-CHRNA4* and the *CYP2A6* gene regions (15q25.1 and 19q13.2), the genetic correlation for each pairwise comparison remained (Figure 1b). The genetic correlation between number of cigarettes smoked per day and lung

cancer risk changed from 0.59 to 0.51 with the removal of these two key loci. The genetic correlation between lung cancer and COPD risk dropped from 0.19 to 0.12. These persistent correlations demonstrate that many other loci outside these sentinel regions on chromosome 15 and 19 contribute to the genetic relations between the smoking-related conditions.

DISCUSSION

Genome-wide, there are strong, positive genetic correlations among number of cigarettes smoked per day, decreased smoking cessation, and the risk of lung cancer and COPD. These strong correlations demonstrate that genetic variants that alter smoking behaviors are also strong contributors to the genetic risk for lung cancer and COPD. The graphs also illustrate that there are regions that are specific for each phenotype studied. Two regions stand out for the strength of their genetic associations with smoking behaviors and smoking-related diseases: chromosomal region 15q25.1, which includes genes encoding the *CHRNA5-CHRNA3-CHRNA4* nicotinic receptor subunits, and chromosomal region 19q13.2, which includes *CYP2A6* the gene encoding the enzyme that metabolizes nicotine.

The chromosomal region 15q25.1 that contains a cluster of nicotinic receptor subunit genes, *CHRNA5-CHRNA3-CHRNA4*, accounts for a strong genetic correlation between number of cigarettes smoked per day and decreased smoking cessation ($r = 0.84$). Variants in this region are also associated with an increased number of cigarettes smoked per day and increased risk of lung cancer ($r = 0.98$) and COPD ($r = 0.93$). Interestingly, variation in the 15q25 gene region is not associated with lung cancer in never-smokers (Gabielsen et al., 2013; Liu et al., 2019; Truong et al., 2010). In addition, variation in the 15q25 gene region is not associated with COPD in never-smokers (Gabielsen et al., 2013). It appears that the risk effects of this chromosomal region are mediated through smoking behaviors. For example, if a person with risk variation in the *CHRNA5-CHRNA3-CHRNA4* gene region smokes, the individual is more likely to smoke more cigarettes per day and have difficulty quitting. These behavior in turn increase the risk of lung cancer and COPD for those individuals. It is plausible that many other loci in the genome share similar effects in that genetic variation associated with an increased number of cigarettes smoked per day and decreased smoking cessation results in an increased rate of lung cancer and COPD.

Surprisingly, the *CYP2A6* gene region (19q13.2) which encompasses the primary gene that metabolizes nicotine shows a different pattern of genetic correlation. The genetic variants in *CYP2A6* associated with an increased risk of heavier smoking are also associated with increased risks of lung cancer and COPD. However, we observed that the genetic correlations between smoking cessation and the other three phenotypes were in the opposite direction of what might be expected in terms of an effect related to addictive behavior; the variants that increase the risks for smoking more cigarettes per day and for developing lung cancer or COPD were also associated with successfully quitting smoking. This unexpected finding demonstrates the complex relation between nicotine metabolism, smoking cessation, cigarettes smoked per day, lung cancer, and COPD.

The *CYP2A6* gene is a highly polymorphic enzyme that metabolizes nicotine to cotinine and then cotinine to trans-3'-hydroxycotinine (3HC) (Tanner & Tyndale, 2017). The nicotine

metabolite ratio (3HC/cotinine) denotes the efficacy of nicotine metabolism via *CYP2A6* (Tanner & Tyndale, 2017). An individual with a low nicotine metabolite ratio is a slow nicotine metabolizer, and an individual with a high nicotine metabolite ratio is considered a fast metabolizer. In a previous study by Patel and colleagues (2016), the SNP, rs56113850, the sentinel associated SNP in *CYP2A6*, explained up to 6.2% of the variation in the nicotine metabolite ratio; the C-allele was associated with increased metabolic activity (Patel et al., 2016). In GWAS of the nicotine metabolite ratio, Loukola and colleagues observed that the C-allele of rs56113850 was also associated with an increase in nicotine metabolism activity, explaining up to ~23% of the variation in the nicotine metabolite ratio in a Finnish population (Loukola et al., 2015). In the largest and more recent GWAS of the nicotine metabolite ratio, Buchwald and colleagues observed that the SNP, rs56113850, was associated with nicotine metabolite ratio among European current smokers (N = 5,185) (Buchwald et al., 2020). Extrapolating these findings with our genetic correlations, it appears that the C-allele of rs56113850 is associated with faster nicotine metabolism and is associated with increased number of cigarettes smoked per day, increased risk of lung cancer, increased risk of COPD, and increased smoking cessation.

Smoking cessation is defined here as being a former smoker among those who have ever smoked and in this study is associated with variants that mark faster nicotine metabolism. This finding is consistent with the results in the population-based study by Loukola and colleagues, where the authors created a genetic risk score using selected SNPs within the *CYP2A6* gene region including rs56113850 (Loukola et al., 2015). In unadjusted models, the authors observed that individuals with alleles associating with fast nicotine metabolite ratio were more likely to be former smokers or more likely to have quit (odds ratio [OR] = 1.39; 95% confidence interval [CI] = 1.09–1.76; $p = 0.007$). This association, however, was no longer significant after adjustment and exclusion of potential confounders such as major depressive disorder though the effect size remained strong (OR = 1.30; 95% CI = 0.95–1.78; $p = 0.10$). Smoking and major depressive disorder have shared genetic vulnerability and are frequently comorbid (Breslau, Peterson, Schultz, Chilcoat, & Andreski, 1998; Liu et al., 2019; Taylor et al., 2014; The Brainstorm Consortium et al., 2018). However, this population-level finding of faster smoking metabolism associated with increased smoking is in contrast to what has been seen in well-characterized smoking cessation trials, where in clinical trials faster nicotine metabolism is associated with decreased smoking cessation (Chen et al., 2014; El-Boraie et al., 2019). It should be noted that using former versus current smoking status as an indication of cessation in a cross-sectional design, as used here, does not capture cessation over time in the same way as cohort studies, or clinical trials. Slower, versus faster nicotine metabolism, in a cohort study of youth found smoker metabolism associated with higher cessation rates; as adults they may not identify as former smokers (Chenoweth, O'Loughlin, Sylvestre, & Tyndale, 2013). The rate of nicotine metabolism can affect smoking cessation trajectories differently, which is not captured in a cross-sectional analyses. These findings demonstrate a complex interplay between nicotine metabolism and smoking cessation that is potentially mediated by other illnesses.

Among the four outcomes studied (cigarettes smoked per day, smoking cessation, lung cancer, and COPD), COPD had the weakest genome-wide correlations. COPD is a heterogeneous disease with different genetic associations depending on the subtype of

COPD as COPD can develop in both ever smokers and never smokers (Agusti et al., 2010; Molfino, 2004). Generating a GWAS of COPD by including only individuals who smoke would likely decrease the heterogeneity of the COPD phenotype which could then lead to stronger genetic correlations related to other phenotypes that are tightly linked with smoking (i.e., cigarettes smoked per day and smoking cessation).

Finally, we determined how much of the genetic correlations remained beyond the top two associated genomic regions. Overall, the genetic correlations persisted after censoring the *CHRNA5-CHRNA3-CHRNA4* and *CYP2A6* gene regions. This finding indicates that additional polygenic effects that contribute to heaviness of smoking and to smoking cessation also contribute to the risks of lung cancer and COPD.

Several genetic loci that underlie smoking behaviors have strong neurobiological indications for their influence on gene regulation specifically in addiction-relevant brain tissues (Hancock et al., 2018). The GSCAN study that generated the GWAS summary statistics of cigarettes smoked per day and smoking cessation examined tissue-specific gene expression of each of their summary statistics of smoking behaviors (Liu et al., 2019). Liu and colleagues observed that the tissue-specific gene expression of associated loci was enriched in multiple central nervous system tissues for each of these smoking-related outcomes (Liu et al., 2019). Lung tissue, however, did not share the same gene expression enrichment as tissues of the nervous system (Liu et al., 2019). It is most likely that gene expression in the central nervous system alters behavior activity leading to smoking-behaviors (smoking more heavily and decreased likelihood of smoking cessation). The act of smoking leads to damaging of the lungs which in turn leads to an increased risk of developing lung cancer and COPD.

There are limitations to this study. We note that our heritability estimates of smoking behaviors, lung cancer, and COPD, are lower than previous heritability estimates of these behaviors and diseases (Mucci et al., 2016; Vink, Willemsen, & Boomsma, 2005; Zhou et al., 2013). These lower heritability estimates illustrate the “missing heritability” that is seen with GWAS studies. GWAS studies capture only part of the genetic variation across the genome. Low frequency and rare variants are not fully captured in GWAS and including variants would likely increase the heritability. Nonetheless, even with the decreased heritability, the finding of the genetic correlation between smoking behaviors and lung cancer and COPD holds. In addition, the lung cancer and COPD summary statistics that we used for this study included both ever-smokers and never-smokers, but the never smokers were a small subgroup (less than 10% of the sample) (Hobbs et al., 2017; McKay et al., 2017). It is possible that performing correlational analyses using GWAS summary statistics between cigarettes smoked per day, smoking cessation, lung cancer, and COPD from analyses that include only ever-smokers could lead to higher global and loci-specific (*CHRNA5-CHRNA3-CHRNA4* and *CYP2A6* gene regions) genetic correlations. However, we do not expect that the genetic correlational trends of smoking cessation to change (being positively correlated with the other smoking-related phenotypes within the *CHRNA5-CHRNA3-CHRNA4* gene region and inversely correlated within the *CYP2A6* gene region).

In summary, this study further characterizes the genetic relations between number of cigarettes smoked per day, smoking cessation and risk of lung cancer and COPD. In general, we see a positive genetic correlation between increased heaviness of smoking, decreased smoking cessation, and increased risk of lung cancer and COPD. In addition, it is clear that there are genetic associations specific to lung cancer and COPD that are not smoking related. However, there is a novel inverse genetic relationship between increased smoking cessation and increased risks for heavier smoking and smoking-related diseases seen with the *CYP2A6* gene region. This inverse relationship highlights the need to better understand how nicotine metabolism alters the risk of heavier smoking measured by cigarettes smoked per day, lung cancer, and COPD while also contributing to increased smoking cessation at the population level.

ACKNOWLEDGEMENTS

The study was funded by the National Institutes of Health (NIH) Training Grant 5T32MH014677 to Michael J. Bray. This study was also funded by the NIH and National Institute of Drug Abuse Grants P30CA091842 and R01DA036583 to Laura J. Bierut and R01DA042090 to Dana B. Hancock.

Research reported in this publication was supported by the Washington University Institute of Clinical and Translational Sciences grant UL1TR002345 from the National Center for Advancing Translational Sciences (NCATS) of the National Institutes of Health (NIH). The content is solely the responsibility of the authors and does not necessarily represent the official view of the NIH. Rachel F. Tyndale acknowledges the support of a Canada Research Chair in Pharmacogenomics and CIHR (FDN-154294).

The Atherosclerosis Risk in Communities study has been funded in whole or in part with Federal funds from the National Heart, Lung, and Blood Institute, National Institute of Health, Department of Health and Human Services, under contract numbers HHSN268201700001I, HHSN268201700002I, HHSN268201700003I, HHSN268201700004I, and HHSN268201700005I. The authors thank the staff and participants of the ARIC study for their important contributions. Funding for GENEVA was provided by National Human Genome Research Institute grant U01HG004402 (E. Boerwinkle).

International Agency for Research on Cancer Disclaimer: Where authors are identified as personnel of the International Agency for Research on Cancer / World Health Organization, the authors alone are responsible for the views expressed in this article and they do not necessarily represent the decisions, policy or views of the International Agency for Research on Cancer / World Health Organization.

REFERENCES

- Agusti A, Calverley PM, Celli B, Coxson HO, Edwards LD, Lomas DA, ... Evaluation of, C. L. t. I. P. S. E. i. (2010). Characterisation of COPD heterogeneity in the ECLIPSE cohort. *Respir Res*, 11, 122. doi:10.1186/1465-9921-11-122 [PubMed: 20831787]
- Breslau N, Peterson EL, Schultz LR, Chilcoat HD, & Andreski P. (1998). Major depression and stages of smoking. A longitudinal investigation. *Arch Gen Psychiatry*, 55(2), 161–166. doi:10.1001/archpsyc.55.2.161 [PubMed: 9477930]
- Buchwald J, Chenoweth MJ, Palviainen T, Zhu G, Benner C, Gordon S, ... Tyndale RF (2020). Genome-wide association meta-analysis of nicotine metabolism and cigarette consumption measures in smokers of European descent. *Mol Psychiatry*. doi:10.1038/s41380-020-0702-z
- Bulik-Sullivan BK, Loh PR, Finucane HK, Ripke S, Yang J, Schizophrenia Working Group of the Psychiatric Genomics, C., ... Neale BM (2015). LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet*, 47(3), 291–295. doi:10.1038/ng.3211 [PubMed: 25642630]
- Chen LS, Bloom AJ, Baker TB, Smith SS, Piper ME, Martinez M, ... Bierut L. (2014). Pharmacotherapy effects on smoking cessation vary with nicotine metabolism gene (*CYP2A6*). *Addiction*, 109(1), 128–137. doi:10.1111/add.12353 [PubMed: 24033696]

- Chenoweth MJ, O'Loughlin J, Sylvestre MP, & Tyndale RF (2013). CYP2A6 slow nicotine metabolism is associated with increased quitting by adolescent smokers. *Pharmacogenet Genomics*, 23(4), 232–235. doi:10.1097/FPC.0b013e32835f834d [PubMed: 23462429]
- Doll R, Peto R, Boreham J, & Sutherland I. (2004). Mortality in relation to smoking: 50 years' observations on male British doctors. *BMJ*, 328(7455), 1519. doi:10.1136/bmj.38142.554479.AE [PubMed: 15213107]
- Edwards R, Carter K, Peace J, & Blakely T. (2013). An examination of smoking initiation rates by age: results from a large longitudinal study in New Zealand. *Aust N Z J Public Health*, 37(6), 516–519. doi:10.1111/1753-6405.12105 [PubMed: 24892149]
- El-Borai A, Taghavi T, Chenoweth MJ, Fukunaga K, Mushiroda T, Kubo M, ... Tyndale RF (2019). Evaluation of a weighted genetic risk score for the prediction of biomarkers of CYP2A6 activity. *Addict Biol*. doi:10.1111/adb.12741
- Gabrielsen ME, Romundstad P, Langhammer A, Krokan HE, & Skorpen F. (2013). Association between a 15q25 gene variant, nicotine-related habits, lung cancer and COPD among 56,307 individuals from the HUNT study in Norway. *Eur J Hum Genet*, 21(11), 1293–1299. doi:10.1038/ejhg.2013.26 [PubMed: 23443019]
- Hancock DB, Markunas CA, Bierut LJ, & Johnson EO (2018). Human Genetics of Addiction: New Insights and Future Directions. *Curr Psychiatry Rep*, 20(2), 8. doi:10.1007/s11920-018-0873-3 [PubMed: 29504045]
- Hjelmberg J, Korhonen T, Holst K, Skytthe A, Pukkala E, Kutschke J, ... Nordic Twin Study of Cancer, c. (2017). Lung cancer, genetic predisposition and smoking: the Nordic Twin Study of Cancer. *Thorax*, 72(11), 1021–1027. doi:10.1136/thoraxjnl-2015-207921 [PubMed: 29054888]
- Hobbs BD, de Jong K, Lamontagne M, Bosse Y, Shrine N, Artigas MS, ... International, C. G. C. (2017). Genetic loci associated with chronic obstructive pulmonary disease overlap with loci for lung function and pulmonary fibrosis. *Nat Genet*, 49(3), 426–432. doi:10.1038/ng.3752 [PubMed: 28166215]
- Liu M, Jiang Y, Wedow R, Li Y, Brazel DM, Chen F, ... Vrieze S. (2019). Association studies of up to 1.2 million individuals yield new insights into the genetic etiology of tobacco and alcohol use. *Nat Genet*, 51(2), 237–244. doi:10.1038/s41588-018-0307-5 [PubMed: 30643251]
- Loukola A, Buchwald J, Gupta R, Palviainen T, Hallfors J, Tikkanen E, ... Kaprio J. (2015). A Genome-Wide Association Study of a Biomarker of Nicotine Metabolism. *PLoS Genet*, 11(9), e1005498. doi:10.1371/journal.pgen.1005498
- Mannino DM, & Buist AS (2007). Global burden of COPD: risk factors, prevalence, and future trends. *Lancet*, 370(9589), 765–773. doi:10.1016/S0140-6736(07)61380-4 [PubMed: 17765526]
- McKay JD, Hung RJ, Han Y, Zong X, Carreras-Torres R, Christiani DC, ... Amos CI (2017). Large-scale association analysis identifies new lung cancer susceptibility loci and heterogeneity in genetic susceptibility across histological subtypes. *Nat Genet*, 49(7), 1126–1132. doi:10.1038/ng.3892 [PubMed: 28604730]
- Molfini NA (2004). Genetics of COPD. *Chest*, 125(5), 1929–1940. doi:10.1378/chest.125.5.1929 [PubMed: 15136409]
- Mucci LA, Hjelmberg JB, Harris JR, Czene K, Havelick DJ, Scheike T, ... Nordic Twin Study of Cancer, C. (2016). Familial Risk and Heritability of Cancer Among Twins in Nordic Countries. *JAMA*, 315(1), 68–76. doi:10.1001/jama.2015.17703 [PubMed: 26746459]
- Patel YM, Park SL, Han Y, Wilkens LR, Bickeboller H, Rosenberger A, ... Le Marchand L. (2016). Novel Association of Genetic Markers Affecting CYP2A6 Activity and Lung Cancer Risk. *Cancer Res*, 76(19), 5768–5776. doi:10.1158/0008-5472.CAN-16-0446 [PubMed: 27488534]
- Sampson JN, Wheeler WA, Yeager M, Panagiotou O, Wang Z, Berndt SI, ... Chatterjee N. (2015). Analysis of Heritability and Shared Heritability Based on Genome-Wide Association Studies for Thirteen Cancer Types. *J Natl Cancer Inst*, 107(12), djv279. doi:10.1093/jnci/djv279
- Tanner JA, & Tyndale RF (2017). Variation in CYP2A6 Activity and Personalized Medicine. *J Pers Med*, 7(4). doi:10.3390/jpm7040018
- Taylor AE, Fluharty ME, Bjorngaard JH, Gabrielsen ME, Skorpen F, Marioni RE, ... Munafo MR (2014). Investigating the possible causal association of smoking with depression and anxiety using

Mendelian randomisation meta-analysis: the CARTA consortium. *BMJ Open*, 4(10), e006141. doi:10.1136/bmjopen-2014-006141

The Brainstorm Consortium, Anttila V, Bulik-Sullivan B, Finucane HK, Walters RK, Bras J, ... Murray R. (2018). Analysis of shared heritability in common disorders of the brain. *Science*, 360(6395). doi:10.1126/science.aap8757

Timofeeva MN, McKay JD, Smith GD, Johansson M, Byrnes GB, Chabrier A, ... Brennan P. (2011). Genetic polymorphisms in 15q25 and 19q13 loci, cotinine levels, and risk of lung cancer in EPIC. *Cancer Epidemiol Biomarkers Prev*, 20(10), 2250–2261. doi:10.1158/1055-9965.EPI-11-0496 [PubMed: 21862624]

Truong T, Hung RJ, Amos CI, Wu X, Bickeboller H, Rosenberger A, ... Spitz MR (2010). Replication of lung cancer susceptibility loci at chromosomes 15q25, 5p15, and 6p21: a pooled analysis from the International Lung Cancer Consortium. *J Natl Cancer Inst*, 102(13), 959–971. doi:10.1093/jnci/djq178 [PubMed: 20548021]

Vink JM, Willemsen G, & Boomsma DI (2005). Heritability of smoking initiation and nicotine dependence. *Behav Genet*, 35(4), 397–406. doi:10.1007/s10519-004-1327-8 [PubMed: 15971021]

World Health Organization. (2017). WHO report on the global tobacco epidemic, 2017: monitoring tobacco use and prevention policies.

Zhou JJ, Cho MH, Castaldi PJ, Hersh CP, Silverman EK, & Laird NM (2013). Heritability of chronic obstructive pulmonary disease and related phenotypes in smokers. *Am J Respir Crit Care Med*, 188(8), 941–947. doi:10.1164/rccm.201302-0263OC [PubMed: 23972146]

DATA CITATION:

[Lung cancer GWAS summary statistics]McKay JD, Hung RJ, Han Y, Zong X, Carreras-Torres R, Christiani DC, ... Amos CI; 2017; Lung cancer GWAS summary statistics; dbGaP repository: https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs001273.v1.p1;phs001273.v1.p1

[COPD GWAS summary statistics]Hobbs BD, de Jong K, Lamontagne M, Bosse Y, Shrine N, Artigas MS, ... International, C. G. C; 2017; COPD GWAS summary statistics; dbGaP repository: https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000179.v5.p2;phs000179.v5.p2

Figure 1. There are strong genome-wide genetic correlations among the smoking-related phenotypes.

a. Genome-wide genetic correlations.

b. Genome-wide genetic correlations censoring the *CHRNA5-CHRNA3-CHRNA4* and *CYP2A6* gene regions.

The top half of the matrix represents the genetic correlation of the heritability of each pairwise comparison of all outcomes in color format. Blue represents a positive correlation. Red represents a negative correlation. The size of the colored squares as well as the depth of

the color represent the strength of the genetic correlation. The bottom half of the matrix represents the numerical value of each respective genetic correlation. The diagonal represents the genetic heritability of each outcome, the asterisks represent the corresponding p-value of each respective genetic correlation (* $p < 0.05$; ** $p < 5.00 \times 10^{-5}$; *** $p < 5.00 \times 10^{-10}$).

Figure 2. ZZ-plots reveal strong correlations of chromosomal regions 15q25 and 19q13.2.

- a.** Comparing cigarettes smoked per day (y-axis) and smoking cessation (x-axis).
- b.** Comparing cigarettes smoked per day (y-axis) and lung cancer (x-axis).
- c.** Comparing cigarettes smoked per day (y-axis) and COPD (x-axis).
- d.** Comparing smoking cessation (y-axis) and lung cancer (x-axis).
- e.** Comparing smoking cessation (y-axis) and COPD (x-axis).
- f.** Comparing lung cancer (y-axis) and COPD (x-axis).

Each dot represents a SNP. The solid red line represents the genome-wide significance value ($p\text{-value} = 5.00 \times 10^{-8}$) converted to a z-score format. The chromosomal region 15q25.1 is color-coded gold. The chromosomal region 19q13.2 is color-coded purple. All other SNPs that were not in chromosomal regions 15.q25.1 and 19q13.2 were color-coded black.

Figure 3. Genome-specific comparisons reveal strong correlations on chromosomal regions 15q25 and 19q13.2.
a. The correlation of the SNPs in the *CHRNA5-CHRNA3-CHRNA4* gene region.
b. The correlation of the SNPs in the *CYP2A6* gene region.
The top half of the matrix represents the genetic correlation of each pairwise comparison of all outcomes. Blue represents a positive correlation. Red represents a negative correlation. The size of the colored squares as well as the depth of the color represents the strength of the genetic correlation. The bottom half of the matrix represents the numerical value of each

respective genetic correlation. The diagonal is the correlation of SNPs within the same outcome.

Table 1.

The leading SNP, rs56113850, in the chromosomal region 19q13.2 has pleiotropic effects.

Chromosomal Region 15q25.1 - rs16969968 – A (Effect Allele)				
Smoking-related Phenotype	Beta	Standard Error	Z-score	P-value
Lung Cancer	0.26	1.20×10^{-2}	21.36	3.32×10^{-101}
COPD	0.19	2.03×10^{-2}	9.58	1.19×10^{-21}
Cigarettes Smoked per Day	0.09	2.93×10^{-3}	30.24	2.32×10^{-200}
Smoking Cessation	0.01	2.56×10^{-3}	5.70	9.37×10^{-9}
Chromosomal Region 19q13.2 - rs56113850– C (Effect Allele)				
Smoking-related Phenotype	Beta	Standard Error	Z-score	P-value
Lung Cancer	0.12	1.38×10^{-2}	8.91	5.02×10^{-19}
COPD	0.11	2.69×10^{-2}	4.04	5.40×10^{-5}
Cigarettes Smoked per Day	0.06	2.91×10^{-3}	19.24	1.10×10^{-81}
Smoking Cessation	−0.02	2.52×10^{-3}	−8.17	2.52×10^{-26}