Distribution Agreement

In presenting this thesis as a partial fulfillment of the requirements for a degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis in whole or in part in all forms of media, now or hereafter now, including display on the World Wide Web. I understand that I may select some access restrictions as part of the online submission of this thesis. I retain all ownership rights to the copyright of the thesis. I also retain the right to use in future works (such as articles or books) all or part of this thesis.

Victoria Risner April 5, 2019

Multi-Trait Polygenic Prediction of Nicotine Dependence

by

Victoria Risner

Rohan Palmer
Adviser

Neuroscience and Behavioral Biology

Rohan Palmer

Adviser

Alicia Smith

Committee Member

David Weinshenker

Committee Member

2019

Multi-Trait Polygenic Prediction of Nicotine Dependence

By

Victoria Risner

Rohan Palmer

Adviser

An abstract of
a thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Arts with Honors

Neuroscience and Behavioral Biology

2019

Abstract

Multi-Trait Polygenic Prediction of Nicotine Dependence
By Victoria Risner

The estimated heritability of nicotine dependence (ND) ranges 40-70%, but Polygenic Risk Scores (PRS) calculated for ND have limited utility using only tobacco use Genome-Wide Association Study (GWAS) summary statistics. We examined a multi-trait PRS model using known genetic correlates of ND liability and show incremental gains in accuracy and power.

Multi-Trait Polygenic Prediction of Nicotine Dependence

By

Victoria Risner

Rohan Palmer

Adviser

A thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Arts with Honors

Neuroscience and Behavioral Biology

2019

Acknowledgements

I would like to thank Dr. Chelsie Benca-Bachman and Dr. Rohan Palmer for their continued support and advice throughout the process of completing my honors thesis. Additionally, I would like to thank the members of my honors thesis committee, Dr. David Weinshenker and Dr. Alicia Smith, for their valuable knowledge and feedback. I am also grateful to the researchers and staff that contributed to the body of literature and datasets used in this study, as well as the sharing of data that was made possible via the Database for Genotypes and Phenotypes at the National Institutes of Health. This work was funded by a grant from the National Institute on Drug Abuse (R01 DA042742).

TABLE OF CONTENTS

INTRODUCTION

Epidemiology of Tobacco/Nicotine Dependence

Despite attempts to regulate tobacco use since the early 1950's, tobacco consumption remains prevalent in the USA with approximately 15.5% (37.8 million) of Americans reporting daily use (Warren, Alberg, Kraft, & Cummings, 2014). The prevalence of tobacco use poses a significant public health concern in the United States, as roughly 500,000 Americans die from smoking or exposure to smoke each year, and 16 million Americans live with serious and costly illnesses (like cancer, cardiovascular disease, pulmonary disease, etc.) caused by tobacco use (Xu, Bishop, Kennedy, Simpson, & Pechacek, 2015). Although the predominant cause of the detrimental health effects of tobacco use lies with the toxic chemicals that comprise cigarette and other tobacco products, nicotine has been identified as the addictive component that explains why so many people use tobacco, and continue to use tobacco products despite the negative effects.

Until recently, Nicotine dependence (ND) was the prevalent phenotype of tobacco-related problems, however the 5th revision of the DSM (American Psychiatric Association, 2013) expanded the definition and renamed it Tobacco Use Disorder (TUD). Both terms (ND/TUD) reflect a maladaptive pattern of tobacco/nicotine consumption that results in both physiological and psychosocial impairments. Nicotine dependence, as defined by the DSM-IV (American Psychiatric Association, 2000) was defined by 7 symptoms (see Table 1 of DSM-IV Nicotine Dependence symptoms). Tobacco Use Disorder (TUD) further expands the definition of ND by including four additional symptoms (craving tobacco, given up/reduced activities for tobacco, tobacco use in physically hazardous situations, continued use despite recurrent social/interpersonal problems) bringing the total to 11 (American Psychiatric Association, 2013).

In addition to the DSM, severity of tobacco related problems has also been referred by heaviness of smoking indices, such as cigarettes per day (Heatherton, Kozlowski, Frecker, Rickert, & Robinson, 1989), pack years of smoking (National Cancer Institute), and the Fagerström Test for Nicotine Dependence (FTND) (Heatherton, Kozlowski, Frecker, & Fagerstrom, 1991)

Etiology of Tobacco/Nicotine Dependence

Research into the etiology of tobacco use and associated problems has been conducted using both population and family-based samples. Twin and family studies have allowed for the estimation of additive genetic effects, common environmental effects, and unique environmental effects on tobacco use behaviors (Zyphur, Zhang, Barsky, & Li, 2013). The estimated heritability (i.e., proportion of individual differences due to additive genetic effects in a population) of tobacco dependence (i.e., meeting the 3+ past-year/lifetime Diagnostic and Statistical Manual of Mental Disorders [version IV] criterion]) ranges from 40-70%, and is consistent across genders (Maes et al., 2011). Unfortunately, while twin, adoption, and family studies indicate that tobacco use and disorders run in families, for largely genetic reasons, they cannot identify which loci and genes in the human genome contribute to the heritability of this disease (Agrawal & Lynskey, 2008). Fortunately, association studies provide some clue as to the location of genomic loci that contribute to additive genetic effects, by measuring the effect of DNA variants in populations. Genome-wide Association Study (GWAS), a type of analysis that examines a wide array of genomic variants, co-localizes single nucleotide polymorphisms (SNPs), copy number variants (CNVs), repeat polymorphisms, and duplications (to name a few) with disease status/severity (Tsao & Florez, 2007). The polymorphisms employed in a GWAS are often commonly occurring (i.e., minor allele frequency (MAF) > 5%) in the population, though enhancements in array

technologies and genomic imputation have allowed for the interrogation of rare variants (MAF < 5%) (Auer & Lettre, 2015).

Early research, using candidate genes, utilized knowledge about biological mechanisms to predict which genes may have an effect on a trait (Lewis & Knight, 2012). However, it was soon discovered that constraining genetic analyses to only a few genes in mechanisms thought to be involved in a disease limited the ability to accurately model all of the suspected sources of genetic variance (Hewitt, 2008). Research to date has shown that tobacco/nicotine dependence is a phenotypically complex and polygenic trait that is influenced by thousands of gene loci that each have a small effect (Bidwell, Palmer, Brick, McGeary, & Knopik, 2016) (Yin et al., 2017).

The agnostic GWAS approach builds upon candidate gene studies by leveraging linkage disequilibrium (LD) to interrogate the entire genome. GWAS utilizes SNPs from genotyping arrays that span the entire genome by tagging LD blocks (Hewitt, 2008). As such, the entire genome of an individual does not need to be genotyped. In order to account for common variants missing from the genotyped data, imputation is often performed on individual-level genotypes in order to provide a more complete genome for analysis. As such, GWAS allows for the estimation of the effect of variants across the genome. However, prior research has shown that most GWAS lack the statistical power to find loci because of various factors including poor GWAS design, small SNP effect sizes, small sample sizes, and statistical corrections (e.g. Bonferroni), to name a few (Teo, 2008). In regards to statistical power, tens of thousands of individuals are necessary to identify a significant effect of an individual SNP, which is often very rare and in low linkage disequilibrium with nearby markers (Moonesinghe, Khoury, Liu, & Ioannidis, 2008). One proxy method for increasing sample size has been the use of meta-analyses on prior Nicotine dependence (ND) studies. These studies have confirmed associations from previous GWAS and

proposed additional SNPs contributing to ND (Evangelou & Ioannidis, 2013). More recently, Liu et al. (2019) examined genome-wide effects on smoking initiation using 1.2 million individuals and identified 378 associated variants, and 55 variants for heaviness of smoking (i.e., number of cigarettes/day) using 337,334 individuals (M. Liu et al., 2019). Among these consumption phenotypes, cigarettes per day (CPD) most closely aligns with the ND as many smokers have a difficult time curbing their consumption.

As the ability to identify loci has been steadily increasing with larger sample sizes in GWAS, many have begun to inquire how to leverage the observed effects to make prediction about risk for developing maladaptive patterns of nicotine/tobacco consumption. To date, there have been 17 GWAS of ND and three meta-analyses that have suggested loci, most notably variants involved in the coding of nicotinic acetylcholine receptor subunits, whose potential to understand risk in ND has been unrealized. Though important genes involved in the biological pathways of ND have been identified, current research has failed to take it one step further by applying these variants to predict ND in a systematic manner (Palmer, McGeary, et al., 2012). Polygenic Risk Scores (PRS) involve the use of one or more GWAS results of a certain phenotype to predict the risk of the same phenotype in a different sample. A natural expectation of any model that makes inference about causal factors in a population is that the observed effects may generalize to other sample ascertained using similar methods. Our review of the literature, to date, identified nine studies that have attempted to use a PRS to predict individual differences in smoking phenotypes ((Otto, Gizer, Bizon, Wilhelmsen, & Ehlers, 2016) (L. S. Chen et al., 2018) (Belsky et al., 2013) (Chang et al., 2019) (Stevens et al., 2017) (Allegrini et al., 2018) (Vink et al., 2014) (Musci, Uhl, Maher, & Ialongo, 2015) (Marees et al., 2018)), with just one specifically focusing on nicotine dependence or tobacco use disorder (Belsky et al.,

2013; Marees et al., 2018). Of these, two studies utilized a PRS to predict ND/TUD, but yielded inconsistent results. Marees et al. tested whether rare variants identified from a GWAS of TUD predicted CPD outcomes and found no significant contribution (Marees et al., 2018). On the contrary, Belsky et al. developed a risk score of CPD variants from a meta-analyses of three GWAS (J. Z. Liu et al., 2010; Thorgeirsson et al., 2010) (Tobacco And Genetics Consortium, 2010) and successfully predicted CPD and progression to ND (OR=1.27[1.09-1.27]); however, their risk score was not a better predictor than Family History (OR=1.53[1.29-1.80]), indicating that their risk score may not have a strong clinical applicability for predicting ND. Altogether, these recent studies highlight the promise of addition PRS, as well as their limitations (Belsky et al., 2013).

To date, there have been a very limited number of PRS calculated for smoking phenotypes. Notably, the vast majority of these studies have only employed PRS derived from GWAS of smoking outcomes. This approach has failed to capture the complex nature of ND/TUD as has been previously reported in genetic and behavioral studies (The Brainstorm Consortium et al., 2018). Specifically, risk for ND is associated with a wide variety of behaviors and traits (see Figure 1). Moreover, the evidence of genetic and phenotypic comorbidity between a wide variety of traits and ND, and the lack of PRS investigating the predictability of ND based on this comorbidity, reveals that added knowledge could be gained by expanding the set of possible sources of information for genetic risk.

Study Aims

The primary aims of this honors thesis were to identify robust PRSs for ND and to examine whether increased predictive power is gained when using multiple etiologically

associated behaviors/traits, as opposed to a model using only a singular tobacco-related PRS. We compared/contrasted a traditional PRS model based on information from a GWAS of CPD to a multi-trait PRS model that integrates GWAS information from many phenotypes that are etiologically comorbid with ND (see Figure 1). The multi-trait PRS model accounts for pleiotropy (i.e., shared genetic effects between traits that explain the phenotypic association between them) between traits. *It was hypothesized that GWAS SNP effects on ND are reproducible and explain individual differences across multiple populations (i.e. holding constant environment and gene-environment correlation and interaction effects)*. It was further hypothesized that single-trait GWAS effects are positively inflated, due to confounding with other traits that are unaccounted for in GWAS. Finally, we hypothesized that additional variance in ND will be explained using an additive model based on known pleiotropic effects.

METHODS

Identification of Traits to Predict ND

In order to supplement measures of ND, we conducted a multi-trait analysis using select phenotypes that are associated with ND and have genetic relatedness to ND. This approach has been demonstrated to increase the strength of predictive scores for genetically complex behavioral phenotypes (Turley et al., 2018). To achieve this task, we conducted a literature review to identify know comorbid indicators of ND. We identified four primary domains: personality, endophenotypes, externalizing behaviors, and internalizing behaviors. Based on these categories, we identified GWAS studies that met the following criteria: 1) The GWAS of the predictor must be conducted within the last 10 years, (to minimize the possibility of cohort effects and genotyping array differences), 2) the predictor must have both a phenotypic and a

genetic relationship to ND (Zvolensky, Taha, Bono, & Goodwin, 2015) (Ittermann et al., 2013; Xu et al., 2015) (Breslau, Kilbey, & Andreski, 1993) (Kheradmand, Ziaaddini, & Vahabi, 2011), and 3) the GWAS of the predictor must have used a sample of at least 10,000 individuals of European Ancestry. These studies were used to source discovery summary statistics. Based on these criteria, PRS were developed using six phenotypes in addition to a measure of nicotine consumption (cigarettes per day): Neuroticism, Schizophrenia, Depression, BMI, Self-Perceived Risk-Taking, and Educational Attainment (see Table 2). Additionally, we used genome-wide association data from a study on height as a control to compare the amount of effectiveness of our PRS in explaining variance in ND. A description of these phenotypes and the selected studies follows. Note that our analyses employed the entirety of the SNP summary statistics from these studies (i.e., not just those SNP effects that met genome-wide significance (i.e., $p < 5x10^{-8}$) because doing so would bias our findings and poorly reflect the polygenic nature of these phenotypes).

Discovery Samples

Height

The PRS of height was developed to be used as a control variable for predicting nicotine dependence. Though early nicotine use in adolescence has been shown to stunt growth in males (O'Loughlin, Karp, Henderson, & Gray-Donald, 2008), prior research found no phenotypic or genetic association between height and nicotine dependence in adults. GWAS summary statistics for height were based on a recent meta-analysis of genome-wide association studies (Yengo et al., 2018). Yengo et al. gathered height data for 693,529 individuals from a combination of individuals from the UK Biobank and GIANT consortium all of European Ancestry. Yengo et al.

identified 3,290 genome-wide significant SNPs (i.e., $p < 5x10^{-8}$) associated with height and found that these SNPs explained 24.6% of variance in height.

BMI

The PRS for BMI was identified based on prior associations suggesting a direct relationship with smoking. Two independent longitudinal studies have suggested that current smokers, on average, have lower BMI than non-smokers or ex-smokers (Munafo, Tilling, & Ben-Shlomo, 2009) (Sneve & Jorde, 2008). Both studies suggest that smoking cessation resulted in an increase in BMI. Similarly, two recent cross-sectional studies also suggest a relationship between obesity prevalence and smoking status. Dare et al., 2015 used individuals from the UK Biobank, to show that current smokers had the smallest likelihood of obesity (defined as a BMI ≥ 30 kg/m^2), and that ex-smokers had the highest likelihood of obesity when comparing between never, current, and ex-smokers (Dare, Mackay, & Pell, 2015). The second study found similar results as Dare et al., in Japanese adults, but only found this association in males (Watanabe et al., 2016). The results from these four studies indicates a phenotypic relationship between BMI and smoking phenotypes, suggesting that smoking serves as a metabolic mediator influencing BMI. Previous research has shown that nicotine mediates metabolism biochemically by influencing the release of appetite suppressing chemicals and increasing the amount of daily energy expenditure (Audrain-McGovern & Benowitz, 2011). A recent study published in 2019 by Wills and Hopfer sheds light on the genetic relationship between BMI and cigarette smoking. Wills and Hopfer found a positive correlation (r = 0.28-0.38) between genetic factors that influence BMI and genetic factors that influence smoking phenotypes in former and current smokers (Wills & Hopfer, 2019).

BMI summary statistics employed in this study were obtained from a meta-analysis of the UK Biobank and GIANT consortium (Yengo et al., 2018). Yengo et al. (2018) used BMI data from 681,275 individuals all of European Ancestry and found 941 genome-wide significant SNPs (p-value < $5x10^{-8}$) associated with BMI , explaining 6.0% of variance in BMI (Yengo et al., 2018).

Cigarettes Per Day

Cigarettes Per Day (CPD), a criterion in the Fagerström Test for Nicotine Dependence (FTND), is a measure of nicotine consumption quantity (Heatherton et al., 1991). Previous research has indicated that smokers who consume more frequently and in higher quantities are more likely to develop ND and less likely to quit (Breslau & Johnson, 2000) (O'Loughlin et al., 2003). The CPD Genome-Wide Association summary statistics were obtained from a meta-analysis of 16 samples in the Tobacco and Genetics (TAG) Consortium (n = 74,053) (Tobacco And Genetics Consortium, 2010). The TAG consortium assessed CPD in 38,181 individuals and identified 5 genome-wide significant (p-value < $5x10^{-8}$) SNPs (3 loci). The most significant SNP accounted for 0.5% of variance in ND (Tobacco And Genetics Consortium, 2010).

Depression

Decades of research on the phenotypic relationship between depression and nicotine dependence has revealed a consistent correlation (Breslau, Kilbey, & Andreski, 1991) (Breslau, 1995) (Tully, Iacono, & McGue, 2010). A recent literature review on cigarette smoking phenotypes and psychiatric disorders analyzed 23 research studies looking at the relationship of tobacco dependence with later depression (5 studies), and depression with later tobacco dependence (18 studies) and found that 78.3% of these studies discovered an association between the two (Fluharty, Taylor, Grabski, & Munafo, 2017). Lyons et al. conducted a twin-study

examining the relationship between depression and nicotine dependence and found a significant overlapping genetic influence (Lyons et al., 2008). On a molecular level, Li et al. (1998) found that nicotine receptor activation can increase serotonin release, a neurotransmitter thought to be involved in depression (Li, Rainnie, McCarley, & Greene, 1998).

GWAS Summary statistic data were obtained from a study conducted by Howard et al published in 2018. We used the phenotype "broad depression," which was determined by a self-reported survey question asking if individuals sought help for problems with "nerves, anxiety, tension, or depression" (Howard et al., 2018). We decided to use this phenotype of "broad depression" instead of two other measures of depression (Probable MDD and ICD-coded MDD) because Howard et al. found that "broad depression is the most tractable UK Biobank phenotype for discovering genes and gene-sets that further our understanding of the biological pathways underlying depression" (Howard et al., 2018). The sample size for their study was 322,580 individuals from the UK Biobank, which was useful in identifying 14 SNPs for Broad Depression that met genome-wide significance with a SNP heritability of 10.2% (Howard et al., 2018).

Educational Attainment

According to the Surgeon General's Report in 2014, smoking has a negative relationship with level of education: 31.5% of individuals with less than a high school diploma reported being current smokers, whereas 10.4% of individuals with at least a college degree reported being current smokers (U.S. Department of Health and Human Services, 2014). A twin study conducted by Grant et al. found differences in education to be genetically correlated with nicotine consumption (Grant et al., 2012). Interestingly, a separate twin study by McCaffery et

al. found a gene x environment interaction with educational attainment and smoking initiation, and they found that educational attainment moderated the variance in smoking initiation (McCaffery, Papandonatos, Lyons, Koenen, et al., 2008). However, they found no correlation between educational attainment and nicotine dependence (McCaffery, Papandonatos, Lyons, & Niaura, 2008). Research into the genetic relationship between educational attainment and smoking phenotypes is still developing, and not yet understood.

GWAS summary statistics for Educational Attainment were obtained from a study by Lee et al. published in Nature Genetics in 2018 (Lee et al., 2018). The total sample size with the educational attainment phenotypes, measured at ≥ 30 years of age, was 766,345 individuals meta-analyzed from 74 separate GWAS. The Lee et al. study identified 1,271 SNPs that met genome-wide significance for contributing to the level of educational attainment, explaining 11-13% of variance in Educational Attainment (Lee et al., 2018).

Neuroticism

Neuroticism is a personality trait characterized by a tendency to experience negative emotion, especially in response to stress. Neuroticism is correlated with anxiety, depression, impulsivity, and has been found to increase the risk of becoming a smoker (Zinbarg et al., 2016) (Spielberger & Jacobs, 1982). A recent study by Choi et al. found that individuals who scored high on neuroticism were more likely to have nicotine dependence (Choi, Payne, Ma, & Li, 2017). Early researchers hypothesized that neuroticism was associated with smoking because of the mood-regulating abilities of smoking (H. J. Eysenck, 1983) (Tate, Pomerleau, & Pomerleau, 1994), but recent evidence has indicated that neuroticism has a shared genetic etiology with

smoking and nicotine dependence (Criado, Gizer, Edenberg, & Ehlers, 2014) (Sallis, Davey Smith, & Munafo, 2018).

The GWAS summary statistics for Neuroticism were obtained from a study conducted by Nagel et al. in 2018 using individuals from the UK Biobank sample (Nagel, Watanabe, Stringer, Posthuma, & van der Sluis, 2018). Nagel et al. used a sum score of 12 dichotomous neuroticism items from the Eysenck Personality Questionnaire—Revised Short Form (S. B. G. Eysenck, Eysenck, & Barrett, 1985). Using the 380,506 individuals with this phenotype, Nagel et al. found 255 SNPs that met genome-wide significance with an estimated SNP heritability between 8-12% (Nagel et al., 2018).

Schizophrenia

Smoking prevalence in Schizophrenic populations is 60-90%, far greater than the general population (Sagud et al., 2009) (Lohr & Flynn, 1992). Kelly and McCreadie noted an overlap of interest on the dopamine receptor system in Schizophrenia and Nicotine Dependence research, pointing to a potential underlying shared biological etiology (Kelly & McCreadie, 2000). A recent genetic study by Hartz et al. found nicotine dependence to have a genetic correlation of 0.14 with schizophrenia (Hartz et al., 2018).

Summary statistics for Schizophrenia were selected from a study by the Schizophrenia Working Group of the Psychiatric Genetics Consortium. The sample size was 152,805 and the Schizophrenia Working group identified 128 Genome-Wide Significant SNPs associated with Schizophrenia, one of which being the 15q24 locus, which is the strongest genetic contributor to nicotine dependence (Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014).

Self-Perceived Risk Taking

Risk-taking, defined as "any consciously or non-consciously controlled behavior with a perceived uncertainty about it outcome...benefits...or costs", is significantly correlated with reward/punishment sensitivity, sensation seeking, and impulsivity (Trimpop, 1994) (Maher, Thomson, & Carlson, 2015). Previous research has found that smokers engage in more risk-taking behaviors than non-smokers (Jenks, 1992) (O'Cathall et al., 2011). A recent genetic study with the UK Biobank found a significant genetic correlation between risk-taking and smoking (r_g=0.17) (Strawbridge et al., 2018).

The current study used GWAS summary statistics for Self-Perceived Risk-Taking from a recent study conducted by Karlsson Linner et al. in 2019. Karlsson Linner et al. derived their phenotype of self-reported "general risk tolerance" from a binary question in a survey: "Would you describe yourself as someone who takes risks?" (Karlsson Linner et al., 2019). Karlsson Linner et al. used a sample of 939,908 from a meta-analysis of 17 studies and identified 124 genome-wide significant SNPs with an estimated SNP heritability of 4.6% (Karlsson Linner et al., 2019).

Target Sample Description and Quality Control

GWAS summary data from the aforementioned discovery studies were applied to an independent target sample (N=6,344 individuals of European ancestry; 45.25 % met the criteria for DSM-IV nicotine dependence [ND], 46.78% female) of pooled public use datasets that were obtained with permission from the database of Genotypes and Phenotypes (dbGaP); none of these data were included in the aforementioned studies. This sample comprises pooled data from four dbGAP studies: The Study of Addiction: Genetics and Environment (SAGE; study

accession phs000092.v1.p1), the Alcohol Dependence GWAS in European and African Americans (Yale Study; study accession phs000425.v1.p1), the Australian twin-family study of alcohol use disorder (OZ-ALC; study accession phs000181.v1.p1), and the GWAS of Heroin Dependence (Heroin GWAS study; study accession phs000277.v1.p1). Each study collected DSM-IV symptoms (coded as present or absent) for nicotine dependence (ND) by using the Semi-Structured Assessment for the Genetics of Alcoholism (SAGE study), the adapted Semi-Structured Assessment for the Genetics of Alcoholism OZ (OZ-ALC study), or the Semi Structured Assessment for Drug Dependence and Alcoholism (Yale Study, Heroin GWAS) (Bucholz et al. 1994; Hesselbrock et al. 1999; Pierucci-Lagha et al. 2005) (Brick, Keller, Knopik, McGeary, & Palmer, 2019).

The aggregation, quality control, and genomic imputation of these data has been described in detail elsewhere (Brick et al., 2019). Briefly, Plink (version 1.9) was used to exclude low frequency SNPs (MAF< 0.01), multi-allelic markers, and variants that failed the Hardy Weinberg Equilibrium test (i.e., H-W p-value < 0.001) (Purcell et al., 2007). The data were imputed after strict selection for individuals of European Ancestry using genomic principal components and multidimensional scaling. Genomic imputation was conducted using the 1000 genome EUR reference panel via the Michigan Imputation Server (Das et al., 2016).

Statistical Analyses

Development and Optimization of PRS

Effect sizes (odds-ratios or standardized regression coefficients [Betas; β]) from the selected GWASs were used to develop a PRS for each predictor. We treated the observed GWAS results for each phenotype as fixed effects. Specifically, we assumed that all SNPs (across traits)

share the same variance-covariance matrix of effect sizes, regardless of test sample. PRSs were derived using the PRSice package in R (version 3.3.3) (Jack Euesden, Cathryn Lewis, & Paul O'Reilly, 2015) (R Core Team, 2017). PRSs were derived while accounting for Linkage Disequilibrium (LD) amongst the SNPs using the clumping option (i.e., employing the β of the SNP with the highest LD in a 250 kilobase [Kb] window). PRSice models included age, gender, and alcohol dependence as covariates. Alcohol dependence (AD) was used as a covariate in an attempt to find genetic effects more specific to nicotine dependence and not AD, which is highly associated with generalized substance use (Hasin, Stinson, Ogburn, & Grant, 2007) (Stinson et al., 2005) (Palmer et al., 2009) (Palmer, Button, et al., 2012). The optimal model for each PRS was obtained by maximizing the amount of phenotypic variance explained in the target sample by varying the SNP inclusion threshold based on the significance level of the SNPs in the trait GWAS. Specifically, we predicted the likelihood that an individual met DSM-IV criterion for ND. We tested a separate logistic regression models using each trait PRS as an individual predictor, after first assessing the contribution of covariates (Model I). In Model I, we predicted ND in our target sample using the CPD PRS. In Models II through *VIII*, we examined the effects of PRS for Height, Neuroticism, Schizophrenia, Depression, BMI, Educational Attainment, and Self-perceived risk-taking, respectively (see Model Equations). The selected threshold for each trait was that which maximized the proportion of variance explained by a given PRS indicated by a particular p-value threshold. As such, thresholds and the SNPs that contributed to each PRS were allowed to vary across traits; a complete list of SNPs can be obtained by applying the thresholds presented in the results section to the publicly available summary statistics files generated by each discovery study.

Model I – VIII: Simple Regression of Nicotine Dependence

$$Y_{ND_{Symptom\ Count}} = \beta_o + \beta_1\left(PRS_{Height[control]}\right) + \varepsilon_1$$
$$Y_{ND_{Symptom\ Count}} = \beta_o + \beta_2\left(PRS_{Cigarettes\ Per\ Day}\right) + \varepsilon_i$$
$$Y_{ND_{Symptom\ Count}} = \beta_o + \beta_3\left(PRS_{Schizophrenia}\right) + \varepsilon_i$$
$$Y_{ND_{Symptom\ Count}} = \beta_o + \beta_4\left(PRS_{Depression}\right) + \varepsilon_i$$
$$Y_{ND_{Symptom\ Count}} = \beta_o + \beta_5(PRS_{Neuroticism}) + \varepsilon_i$$
$$Y_{ND_{Symptom\ Count}} = \beta_o + \beta_6(PRS_{BMI}) + \varepsilon_i$$
$$Y_{ND_{Symptom\ Count}} = \beta_o + \beta_7(PRS_{Educational\ Attainment}) + \varepsilon_1$$
$$Y_{ND_{Symptom\ Count}} = \beta_o + \beta_8\left(PRS_{Self-Perceived\ Risk\ Taking}\right) + \varepsilon_1$$

Multivariate Logistic Regression of ND on PRSs in Target Sample

Following the selection of the most informative PRS for each trait, we fitted a multivariate logistic regression model using the target sample (Model IX). We examined the joint effect of all of the predictors by including all trait PRS as predictors in the model. All models were fitted in Mplus (version 8) (Muthén & Muthén, 2017) using robust maximum likelihood estimation. Model fit was determined using the R-square test statistic. Covariates in these analyses included age, sex, and alcohol dependence status to control for differences in ND between age and sex and to control for other substance use disorders because it is known that substance use disorders are comorbid with each other.

Model IX: Multiple Regression Model of Nicotine Dependence

$$Y_{ND_{Symptom\ Count}} = \beta_o + \beta_2\left(PRS_{Cigarettes\ Per\ Day}\right) + \beta_3\left(PRS_{Schizophrenia}\right) + \beta_4\left(PRS_{Depression}\right) + \beta_5(PRS_{Neuroticism}) + \beta_6(PRS_{BMI}) + \beta_7(PRS_{Educational\ Attainment}) + \beta_8(PRS_{Self-Perceived\ Risk\ Taking}) + \varepsilon_i$$

RESULTS

Single-Trait PRS Effects on ND

Single-trait PRS results were parsed according to effects from trait PRS, age, sex, and alcohol dependence. Results from each trait PRS are shown in Figures 2 thru 9, which depict the r^2 and significance of each PRS based on various p-value selection thresholds. The covariates (age, gender, alcohol dependence) accounted for the most variance in ND (see Table 3). Compared to all other predictor PRSs, the height-PRS accounted for the least amount of variance in ND (r^2=0.0001, p=0.392) (Figure 2). The best PRS for BMI had a p-value selection threshold of 0.06 and accounted for 0.06% of variance in ND (r^2=0.0006, p=0.090) (Figure 3). The best PRS for CPD had a p-value selection threshold of 0.35 and accounted for 0.05% of variance in ND (r^2=0.0005, p=0.112) (Figure 4). The best PRS for Depression had a p-value selection threshold of 0.08 and accounted for 0.07% of variance in ND (r^2=0.0007, p=0.066) (Figure 5). The best PRS for Educational Attainment had a p-value selection threshold of 0.02 and accounted for 0.06% of variance in ND (r^2=0.0006, p=0.084) (Figure 6). The best PRS for Neuroticism had a p-value selection threshold of 0.2 and accounted for 0.07% of variance in ND (r^2=0.0007, p=0.055) (Figure 7). The best PRS for Self-Perceived Risk-Taking had a p-value selection threshold of 0.02 and accounted for 0.05% of variance in ND (r^2=0.0005, p=0.101) (Figure 9). While most PRS were non-significant (see Table 3), the PRS for Schizophrenia was statistically significant (Full (i.e., trait + covariates) r^2=0.0898, PRS (trait only) r^2=0.0013, p=0.011); the covariates (age, sex, alcohol dependence) accounted for most of the observed variance in ND. Of the total 8.90% of variance explained by the model, the covariates accounted for 8.85%, p<0.01 (see Table 3).

Correlations between Single-Trait PRS

Tables 4 shows the correlations between the optimized PRS for each trait used in our models. Correlations between PRSs ranged from moderately negative to moderately positive with several instances of non-significant associations. The most significant negative correlations were between Educational Attainment and BMI (-0.21, 95% CI: [-0.24,-0.19]) and Educational Attainment and Neuroticism (-0.19, 95% CI: [-0.21,-0.16]). Neuroticism and Depression shared the strongest positive correlation (0.38, 95% CI: [0.36, 0.41]). Schizophrenia was modestly correlated with Self-Perceived Risk Taking (0.14, 95% CI: [0.11, 0.16]), Depression (0.13, 95% CI: [0.11, 0.16]), and Neuroticism (0.10, 95% CI: [0.08, 0.12]).

Multi-Trait Logistic Regression PRS Effects on ND

The multiple logistic regression model that included the optimized PRS for each trait as a predictor did not account for any additional phenotypic variance compared to the model using only a single predictor (Δr^2=0.002) (see Table 5)). The Schizophrenia PRS was the only significant predictor of ND after controlling for the other traits and covariates, and indicated a 1% increase in risk for ND per unit increase in a person's schizophrenia standardized PRS (β= 0.029, p=0.038; Odds-ratio = 1.009, 95% Confidence Interval = [1.000, 1.018] (see Table 6)). The Schizophrenia PRS appeared to be the most robust effect compared to all other PRS predictors. The effects of the other PRS were diminished (compared to the estimates from the aforementioned model) when included in the multi-trait logistic regression, indicating that the effects seen in the other PRS were confounded with other genetic liabilities that were not accounted for in the optimization models.

DISCUSSION

Our findings reveal an association between genetic liability for Schizophrenia and Nicotine dependence, even when controlling for all other PRS effects and covariates. Moreover, we observed modest associations between the schizophrenia PRS and other genetic indicators of psychiatric liability in this sample of substance users, suggesting that these effects were not confounded with the traits examined in this study. Although we weren't able to control for all forms of substance involvement (i.e., cannabis, cocaine, opioid), by targeting the most prevalent and comorbid substance (i.e., alcohol), we aimed to target variants that contribute exclusively to ND.

Our findings agree with a similar study that found a genetic association between nicotine use and Schizophrenia (Carey et al., 2016). Our findings also parallel a study by Chen et al., which also derived a PRS of Schizophrenia from the same summary statistic data and applied it to ND (measured by FTND), which accounted for 0.07% of variance (J. Chen et al., 2016). Due to the distinctions between measurements of ND by DSM-IV and FTND, the differences in r^2 between the two studies may indicate that Schizophrenia is a weaker predictor of heaviness of nicotine consumption. Interestingly, Hartz et al. (2017) found an even stronger association between Schizophrenia PRS and ND (FTND) than our study (r^2: 0.42-5.8%); however, Hartz et al. used a target sample from COGEND, which was a subset of our sample. COGEND is comprised of individuals who have smoked at least 100 cigarettes in their lifetime, with nearly half meeting FTND criteria for ND. It is possible that the observed effects are higher in this subpopulation of our data due to the higher prevalence of smoking and ND, however, this is not consistent across samples.

These results should be interpreted in the context of several limitations. First, rather than a symptom count of ND, our target sample ND phenotype was binary (0 = ND absent, 1 = ND present). A continuous phenotype would provide better indication how people vary in their ND severity as the case-control distinction assumes that individuals with 3 criterion endorsed are equivalent to those with 7 (for example). Use of a continuous ND phenotype would provide us with more accurate approximation of variation between individuals (i.e. individual scores can range from 0-7 DSM-IV criteria met). Additionally, DSM-IV diagnosis alone is not the best measure of ND. Due to the absence of certain criteria in the DSM-IV (i.e. craving), nicotine users who might meet criteria for ND based on other measures (i.e. FTND) may not be classified as dependent according to DSM-IV criteria. Moreover, a DSM-IV ND diagnosis is merely one construct that fails to reflect all of the pharmaco- kinetic and dynamic elements of nicotine exposure. Donny and Dierker illustrated the issues with DSM-IV ND diagnosis, pointing out that 37.7% of smokers who smoke ≥ 10 cigarettes a day fail to meet DSM-IV criteria (Donny & Dierker, 2007).

Second, our method of controlling for Alcohol Dependence may have led to overcorrection of other substance use effects. By using Alcohol Dependence as a covariate, we controlled for the genetic effects, however, we also may have controlled for environmental effects of generalized substance use, which likely interact with genetic risk for addictive tendencies (i.e., gene-environment correlation). Future studies should gather summary statistic data from a GWAS of alcohol consumption or dependence, thus controlling for only the genetic effects of Alcohol Dependence. Nevertheless, our robust result of the Schizophrenia PRS reflect the results of previous studies and further implicate a shared genetic comorbidity between Schizophrenia and ND. (Maier et al., 2018)

Third, other methods of genetic prediction should also be considered, such as, weighted Multi-Trait Summary-based Best Linear Unbiased Prediction (wMT-SBLUP), with our discovery samples and compare/contrast SBLUP results with PRSice results. Rather than using the Odds Ratio/Beta estimates from the summary statistics, SBLUP, utilizes correlations and LD maps in order to adjust the estimates in order to account for more variance in the target sample. Previous research utilizing various methods of prediction has revealed that wMT-SBLUP predicts phenotypes better than single-trait PRS (Maier, 2018). Fourth, these analyses were conducted across multiple samples assuming similar effects across groups. Previous work using these data to examine alcohol dependence revealed limited effects of study effects as most of these samples are highly enriched for alcohol and tobacco dependence cases and matched controls (Brick et al., 2019). As such, additional work is needed to examine these estimates after controlling for study effects, which are suspected to be limited.

In conclusion, our research revealed that the genetic and phenotypic relationship between traits and Nicotine Dependence may be overinflated. We saw that genetic variants contributing to a trait of interest explain very small amounts of variance in nicotine dependence, despite research indicating that their relationship should be much higher. Alternatively, the methods of PRSice may result in an under-estimation of the results. In our future directions, by using an alternative method of deriving PRS (wMT-SBLUP), we can gain more insight on whether we are seeing an over-estimation from previous research or an under-estimation in our study. Additionally, researchers should be wary to indicate that a trait of interest is the single best predictor of another trait. When comparing between Single-Trait and Multi-Trait PRS results, even though most of the Single-Trait PRS do not meet the level of significance ($p < 0.05$), their individual effects were confounded and most PRS were modestly correlated. Based on the expansive literature

discussing the relationship of ND with a number of other traits, behaviors, and phenotypes, conducting a research study on the relationship between a single trait and ND would result in an inaccurate estimation of the genetic relationship between the two. ND is a complex disorder, and in order to best capture its genetic influencers for predictive purposes, one must attempt to incorporate multiple relationships in one model for the most accurate representation.

Collection of Tables

Table 1: DSM IV Nicotine Dependence Criteria	
Tolerance	Need for greatly increased amounts of the substance to achieve intoxication (or desired effect) (1a)
	Markedly diminished effect with continued use of the same amount of the substance (1b)
Withdrawal	Maladaptive behavioral change, with physiological and cognitive concomitants, that occurs when blood or tissue concentrations of a substance decline in an individual who had maintained prolonged heavy use of the substance (2a)
	Take the substance to relieve or to avoid unpleasant withdrawal symptoms (2b)
Compulsive Use	Take the substance in larger amounts or over a longer period than was originally intended (3)
Compulsive Use	There have been many unsuccessful efforts to decrease or discontinue use (4)
Compulsive Use	Spend a great deal of time obtaining the substance, using the substance, or recovering from its effects (5)
Compulsive Use	Important social, occupational, or recreational activities revolve around the substance (6)
Compulsive Use	Despite recognizing the contributing role of the substance to a psychological or physical problem, the person continues to use the substance (7)

Table 2: Summary of Traits Selected for Analyses and their GWAS results

Phenotype	**Sample**	**Sample Size**	**Paper**	**Year Published**	**# Significant SNPS**
Cigarettes Per Day	TAG Consortium	74,053	Genome-wide meta-analyses identify multiple loci associated with smoking behavior	2010	14
Neuroticism	UK Biobank	380,506	Item-level analyses reveal genetic heterogeneity in neuroticism	2018	255 (est. SNP heritability: 8-12%)
Depression	UK Biobank	322,580	Genome-wide association study of depression phenotypes in UK Biobank identifies variants in excitatory synaptic pathways	2018	14 (est. SNP heritability: 10.2%)
Schizophrenia	PGC GWAS data sets (total of 52 samples)	152,805	Biological insights from 108 schizophrenia associated genetic loci	2014	128
BMI	UK Biobank and GIANT consortium	700,000	Meta-analysis of GWAS for height and BMI in 700000 individuals of EA	2018	941 (est. SNP heritability: 22.4%)
Height (control)	UK Biobank and GIANT consortium	700,000	Meta-analysis of GWAS for height and BMI in 700000 individuals of EA	2018	3290 (est. SNP heritability: 48.3%)
Educational Attainment	Add Health, EGCUT, ELSA, FENLAND, Geisinger, GSII, NORFOLK, UKB, UKHLS, VIKING, WLS, ACPRC, AGES, ALSPAC, ASPS, BASE-II, CoLaus, COPSAC2000, CROATIA-KORCULA, deCODE, DHS, DIL, EGCUT1, EGCUT2, EGCUT3, ERF, FAMHS, FINRISK, FTC, GOYA, GRAPHIC, GS, H2000 CASES, H2000 CONTROLS, HBCS, HCD, HNRS (corexB), HNRS (oexpr), HNRS (Omni1), HRS, Hypergenes, INGI-CARL, INGI-FVG, KORA S3, KORA S4, LBC1921, LBC1936, LifeLines, MCTFR, MGS, MoBa, NBS, NESDA, NFBC66, NTF, OGP, OGP-Talana, ORCADES, PREVEND, QIMR, RS-I, RS-II, RS-III, Rush-MAP, Rush-ROS, SardiNIA, SHIP, SHIP-TREND, STR-Salty, STR- Twingene, THISEAS, TwinsUK, WTCCC58C, YFS	766,345	Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals	2018	1,115 (est. SNP heritability: 21%)
Self-Perceived Risk-Taking	Army STARRS (NSS1, NSS2, and PPDS), BASE-II, NFBC 1966, RSIII, STR 1 (TWINGENE), STR 2 (SALTY), TAG, UKB, UKHLS, VIKING, Add Health, HRS, NTR, UKB-siblings, ZURICH	939,908	Genome-wide association analyses of risk tolerance and risk behaviors in over 1 million individuals identify hundreds of loci and shared genetic influences	2019	124 (est. SNP heritability: 4.6%)

Table 3: Summary of Results from Single-Trait PRS

Phenotype	**Number of SNPs**	**Best PRS p-value Threshold**	$\boldsymbol{\beta_{PRS}}$ **(p)**	$\boldsymbol{\beta_{age}}$ $(\boldsymbol{p})$	$\boldsymbol{\beta_{sex}}$ **(p)**	$\boldsymbol{\beta_{alcdep}}$ **(p)**	**Est. r^2 (p)**
Height	26,157	0.01	-0.012 (0.392)	-0.131 (<0.001)	-0.027 (0.051)	0.256 (<0.001)	0.082 (<0.001)
Cigarettes Per Day	61,462	0.35	0.022 (0.112)	-0.131 (<0.001)	-0.027 (0.055)	0.256 (<0.001)	0.082 (<0.001)
Depression	34,193	0.08	0.025 (0.066)	-0.130 (<0.001)	-0.026 (0.060)	0.256 (<0.001)	0.083 (<0.001)
Neuroticism	60,754	0.20	0.026 (0.055)	-0.130 (<0.001)	-0.027 (0.052)	0.256 (<0.001)	0.083 (<0.001)
Schizophrenia	34,814	0.06	**0.035 (0.011)**	-0.131 (<0.001)	-0.026 (0.065)	0.259 (<0.001)	0.083 (<0.001)
Educational Attainment	23,712	0.02	-0.007 (0.633)	-0.131 (<0.001)	-0.027 (0.051)	0.255 (<0.001)	0.082 (<0.001)
Self-Perceived Risk-Taking	14,733	0.02	0.023 (0.101)	-0.131 (<0.001)	-0.027 (0.052)	0.255 (<0.001)	0.082 (<0.001)
BMI	32,032	0.06	0.023 (0.089)	-0.131 (<0.001)	-0.027 (0.054)	0.255 (<0.001)	0.082 (<0.001

Table 4: Genetic Correlations between the best Single-Trait PRS (selected at various p-value thresholds based on R^2). *:p <0.05; **: p<0.01; ***: p<0.001; ****:p<0.0001

	BMI	Cigs/Day	Depression	Educational Attainment	Height	Neuroticism	Self-Perceived Risk-Taking	Schizophrenia
BMI								
Cigs/Day	0.04**							
Depression	0.03*	0						
Educational Attainment	-0.21****	-0.03*	-0.10****					
Height	-0.10****	0.01	-0.06****	0.15****				
Neuroticism	0	-0.01	0.38****	-0.19****	-0.07****			
Self-Perceived Risk-Taking	0.10****	0	0.06****	0	-0.03*	-0.01		
Schizophrenia	-0.04**	0.03*	0.13****	0.02	-0.07****	0.10****	0.14****	

Table 5: Adjusted R^2 estimate from Multi-Trait Logistic Regression PRS

Dependent Variable	Estimate	S.E.	Est./S.E.	Two-Tailed P-value
Nicotine Dependence	0.085	0.008	11.019	<0.001

Table 6: Summary of Results from Multi-Trait Logistic Regression PRS

NIC on	Odds Ratio	95% CI: Lower	95% CI: Upper	Two-Tailed P-Value
Polygenic Risk Scores				
Height	0.993	0.957	1.030	0.698
Cigarettes Per Day	1.001	1.000	1.002	0.131
Depression	1.048	0.939	1.169	0.404
Neuroticism	1.028	0.985	1.074	0.208
Schizophrenia	1.009	1.000	1.018	0.038
Educational Attainment	1.001	0.979	1.024	0.940
Self-Perceived Risk-Taking	1.044	0.969	1.125	0.258
BMI	1.037	0.990	1.086	0.130
Covariates				
Age	0.977	0.972	0.981	<0.001
Sex	0.910	0.821	1.009	0.073
Alcohol Dependence	2.896	2.587	3.243	<0.001

Collection of Figures

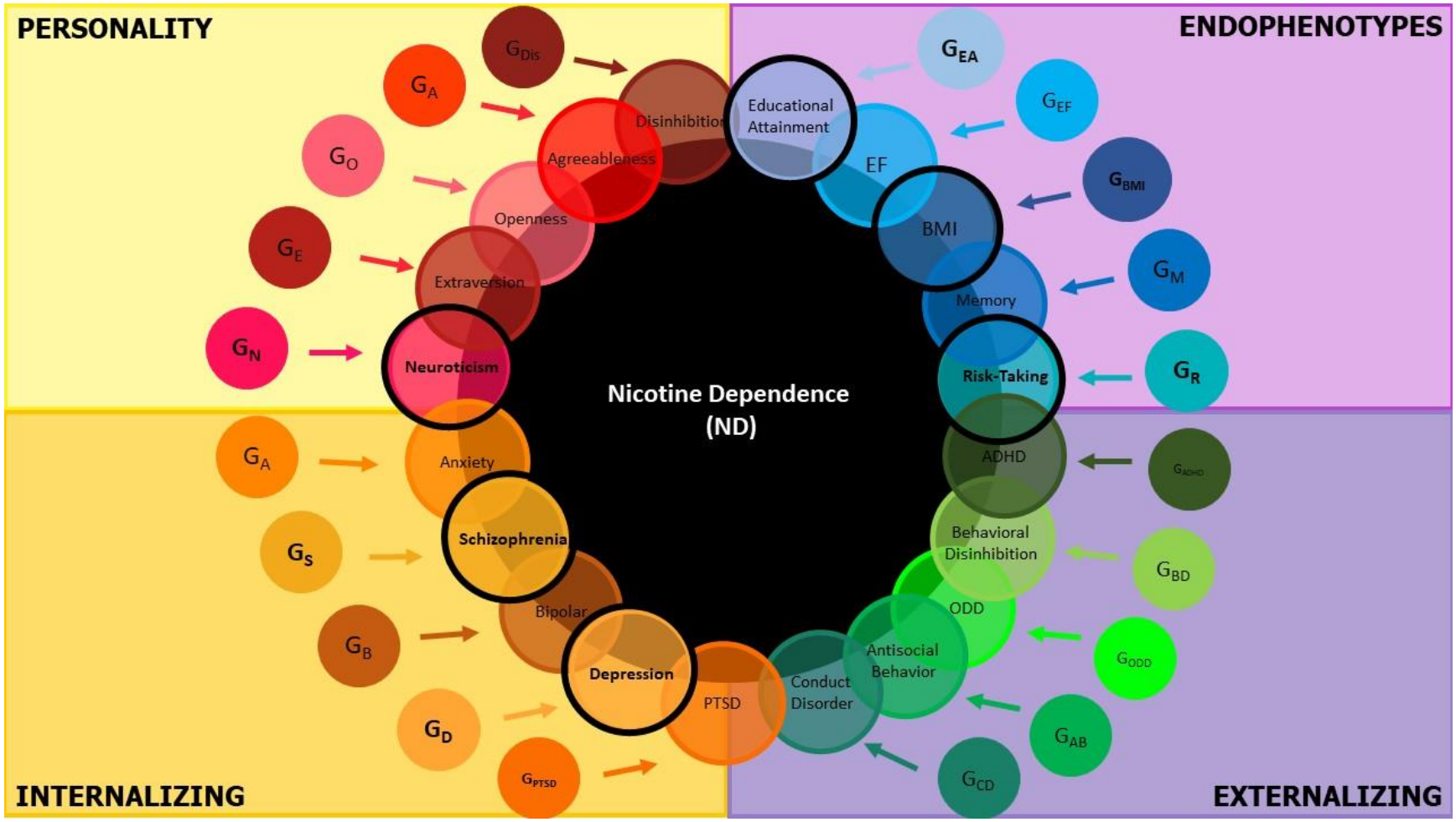

Figure 1: Model of various predictors comorbid with Tobacco Use Disorder, organized in four sections: Personality traits, Internalizing behaviors, Externalizing behaviors, and Endophenotypes

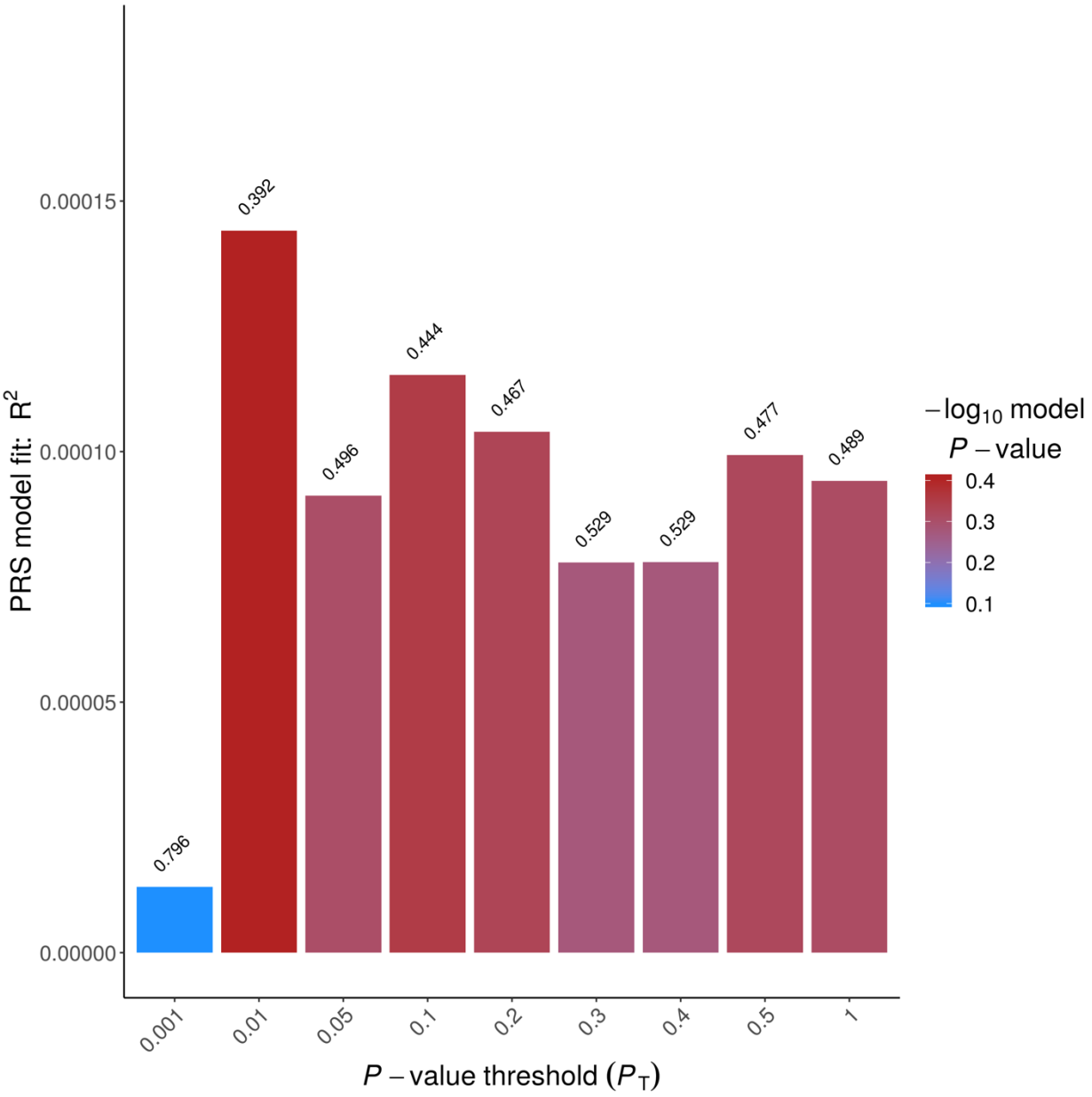

Figure 2: Results from Height PRS on Nicotine Dependence. Multiple PRS for Height were calculated at various P-value thresholds (x-axis). The best PRS, selected based on model fit (largest R^2; y-axis), was retained for multi-trait logistic regression of PRS. P-values of PRS models at each P-value threshold can be seen the top of each bar.

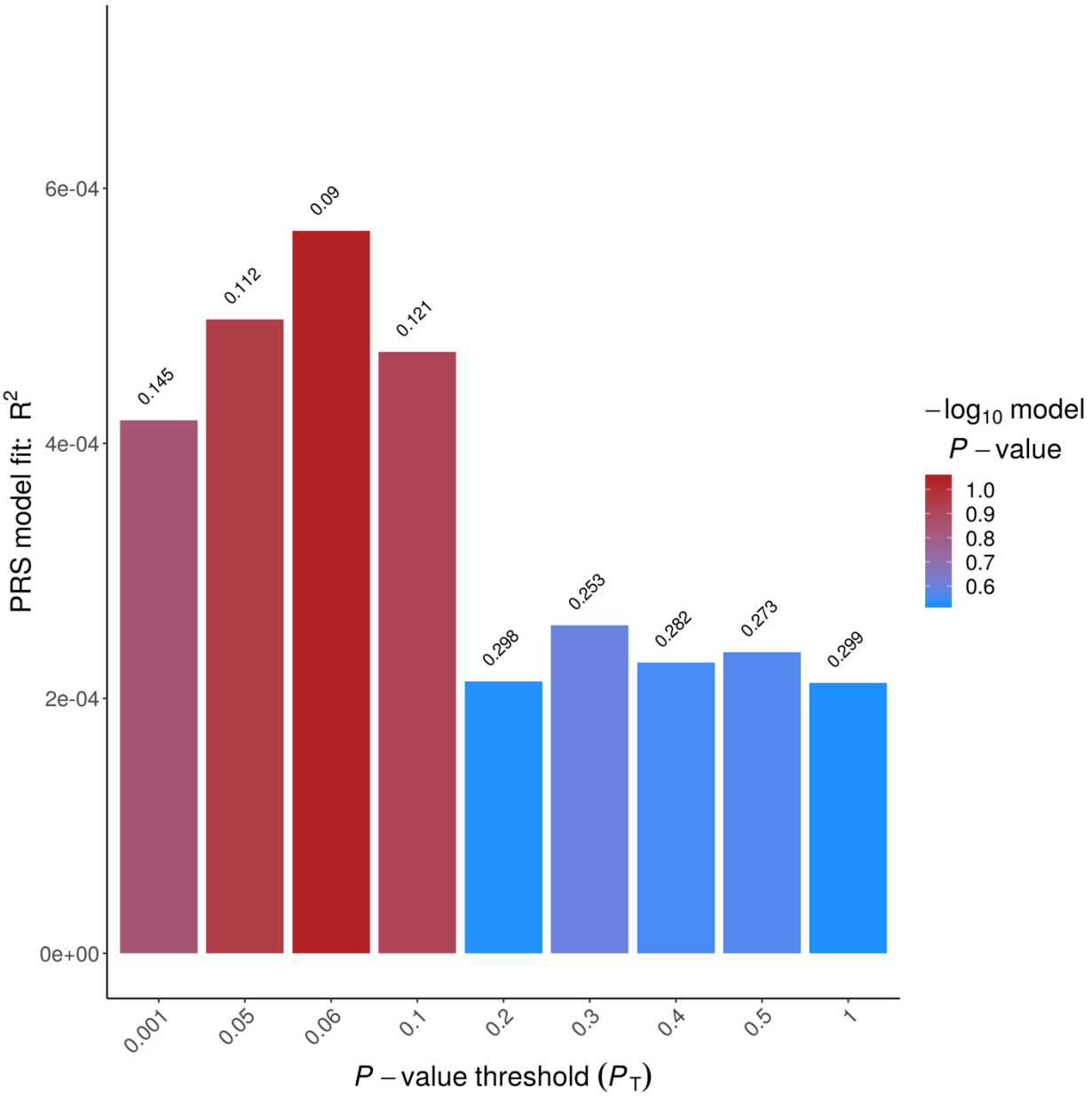

Figure 3: Results from BMI PRS on Nicotine Dependence. Multiple PRS for BMI were calculated at various P-value thresholds (x-axis). The best PRS, selected based on model fit (largest R^2; y-axis), was retained for multi-trait logistic regression of PRS. P-values of PRS models at each P-value threshold can be seen the top of each bar.

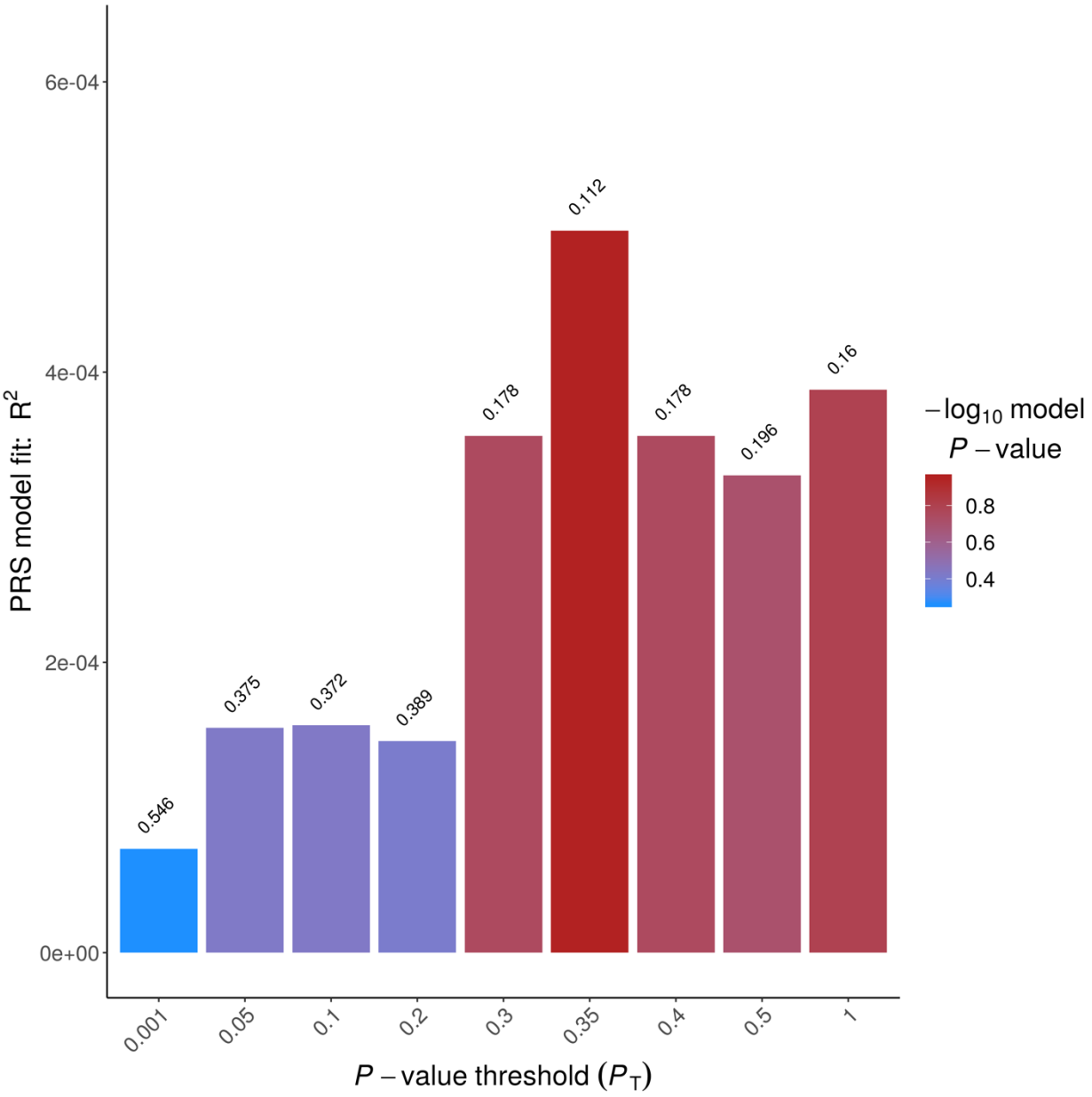

Figure 4: Results from Cigarettes Per Day PRS on Nicotine Dependence. Multiple PRS for Cigarettes Per Day were calculated at various P-value thresholds (x-axis). The best PRS, selected based on model fit (largest R^2; y-axis), was retained for multi-trait logistic regression of PRS. P-values of PRS models at each P-value threshold can be seen the top of each bar.

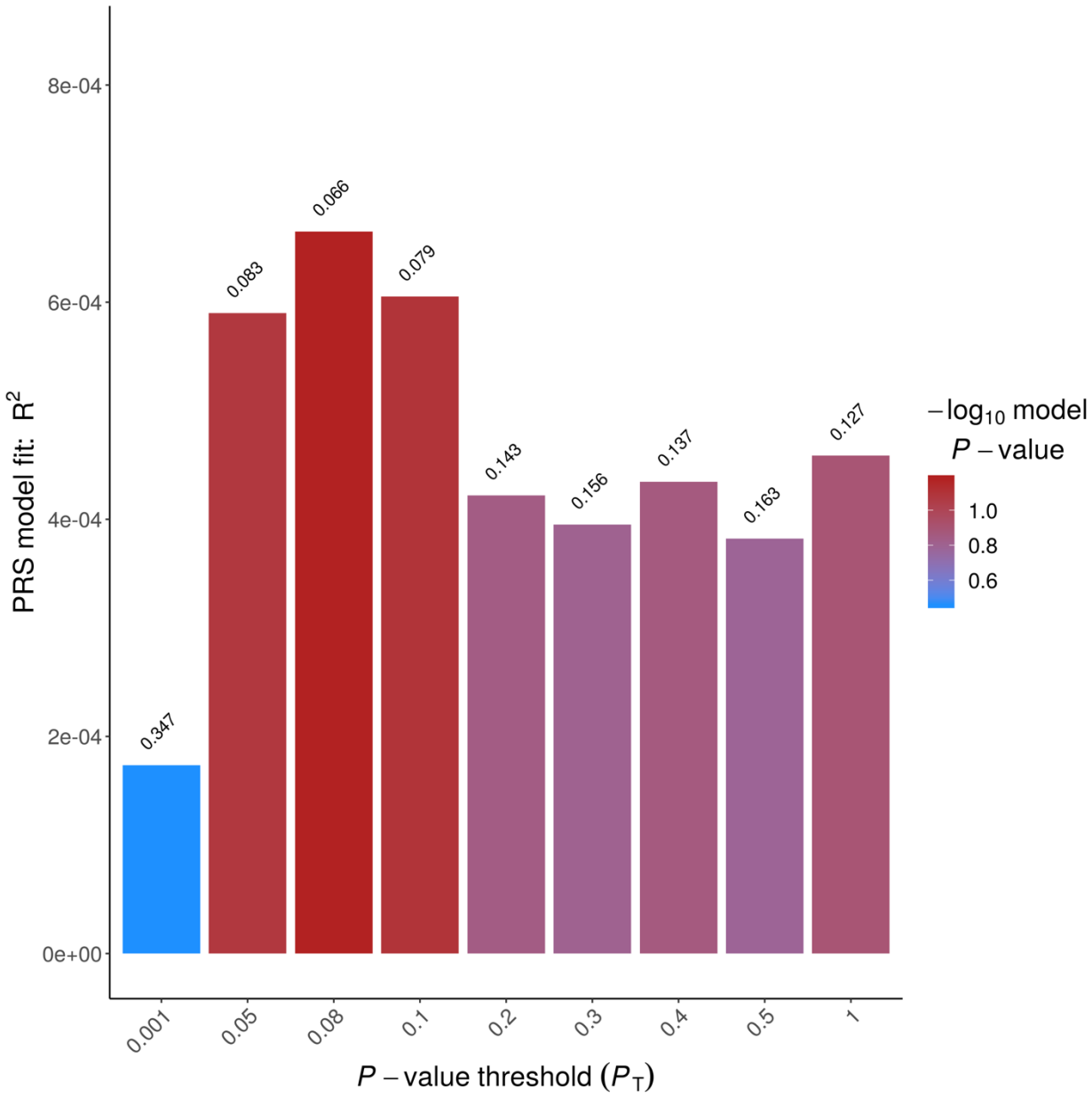

Figure 5: Results from Depression PRS on Nicotine Dependence. Multiple PRS for Depression were calculated at various P-value thresholds (x-axis). The best PRS, selected based on model fit (largest R^2; y-axis), was retained for multi-trait logistic regression of PRS. P-values of PRS models at each P-value threshold can be seen the top of each bar.

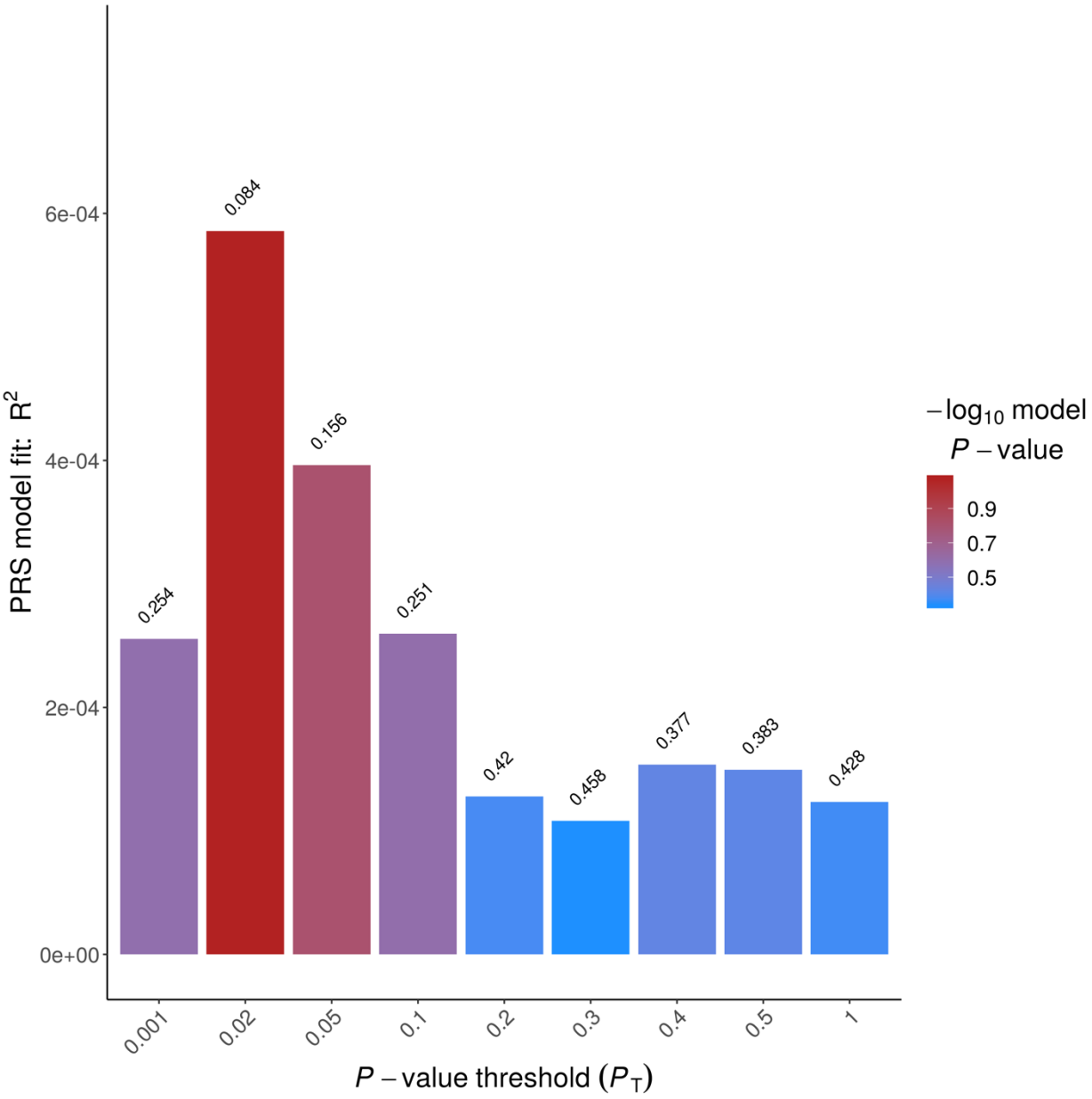

Figure 6: Results from Educational Attainment PRS on Nicotine Dependence. Multiple PRS for Educational Attainment were calculated at various P-value thresholds (x-axis). The best PRS, selected based on model fit (largest R^2; y-axis), was retained for multi-trait logistic regression of PRS. P-values of PRS models at each P-value threshold can be seen the top of each bar.

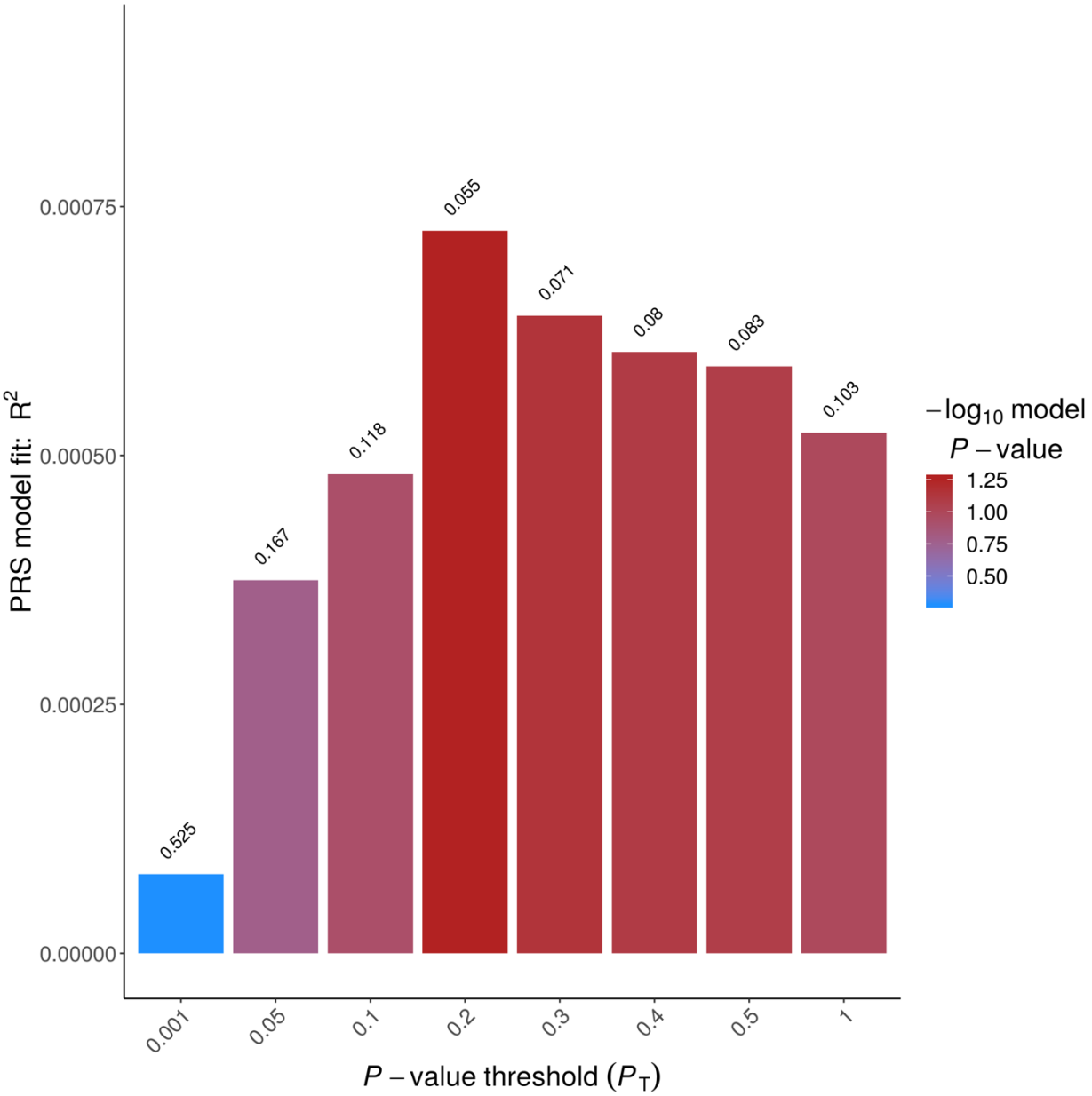

Figure 7: Results from Neuroticism PRS on Nicotine Dependence. Multiple PRS for Neuroticism were calculated at various P-value thresholds (x-axis). The best PRS, selected based on model fit (largest R^2; y-axis), was retained for multi-trait logistic regression of PRS. P-values of PRS models at each P-value threshold can be seen the top of each bar.

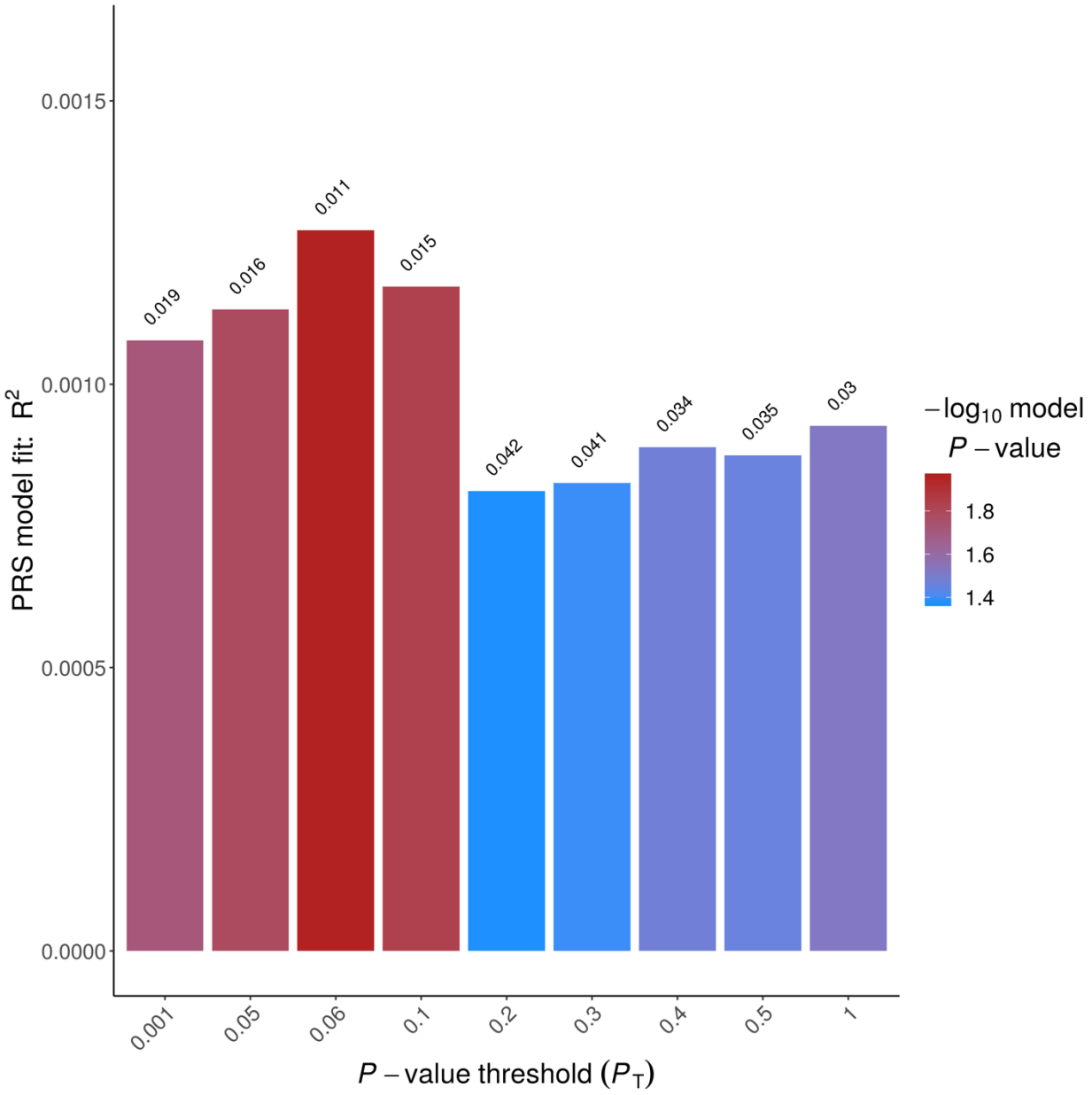

Figure 8: Results from Schizophrenia PRS on Nicotine Dependence. Multiple PRS for Schizophrenia were calculated at various P-value thresholds (x-axis). The best PRS, selected based on model fit (largest R^2; y-axis), was retained for multi-trait logistic regression of PRS. P-values of PRS models at each P-value threshold can be seen the top of each bar.

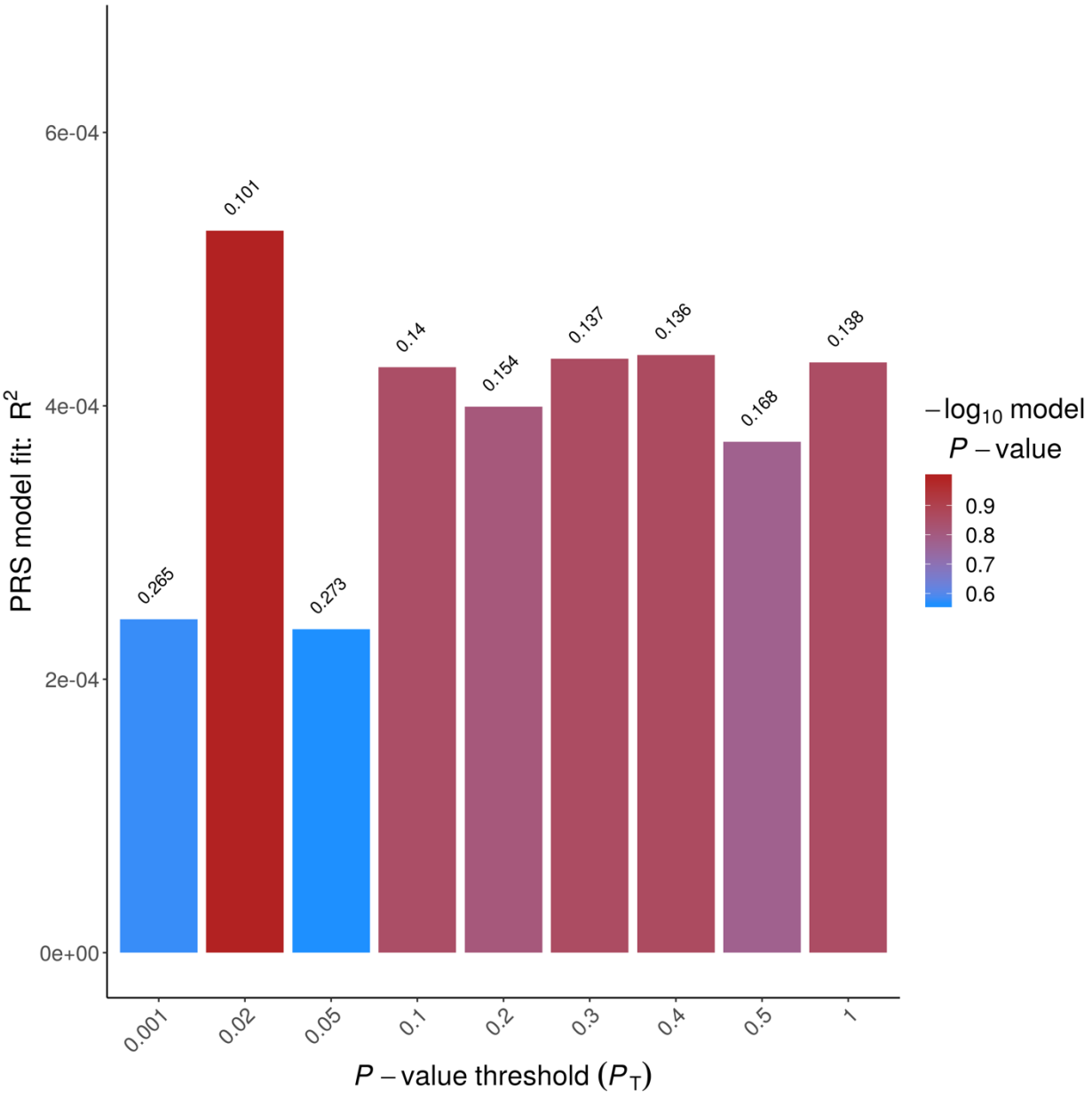

Figure 8: Results from Self-Perceived Risk-Taking PRS on Nicotine Dependence. Multiple PRS for Self-Perceived Risk-Taking were calculated at various P-value thresholds (x-axis). The best PRS, selected based on model fit (largest R^2; y-axis), was retained for multi-trait logistic regression of PRS. P-values of PRS models at each P-value threshold can be seen the top of each bar.

Bibliography

Agrawal, A., & Lynskey, M. T. (2008). Are there genetic influences on addiction: evidence from family, adoption and twin studies. *Addiction, 103*(7), 1069-1081. doi:10.1111/j.1360-0443.2008.02213.x

Allegrini, A. G., Verweij, K. J. H., Abdellaoui, A., Treur, J. L., Hottenga, J. J., Willemsen, G., . . . Vink, J. M. (2018). Genetic vulnerability for smoking and cannabis use: associations with e-cigarette and water pipe use. *Nicotine Tob Res*. doi:10.1093/ntr/nty150

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders (4th ed., text rev.).*

American Psychiatric Association. (2013). Diagnostic and statistical manual of mental disorders (5th ed.). Washington, DC: Author.

Audrain-McGovern, J., & Benowitz, N. L. (2011). Cigarette smoking, nicotine, and body weight. *Clin Pharmacol Ther, 90*(1), 164-168. doi:10.1038/clpt.2011.105

Auer, P. L., & Lettre, G. (2015). Rare variant association studies: considerations, challenges and opportunities. *Genome Medicine, 7*. doi:ARTN 16
10.1186/s13073-015-0138-2

Belsky, D. W., Moffitt, T. E., Baker, T. B., Biddle, A. K., Evans, J. P., Harrington, H., . . . Caspi, A. (2013). Polygenic risk and the developmental progression to heavy, persistent smoking and nicotine dependence: evidence from a 4-decade longitudinal study. *JAMA Psychiatry, 70*(5), 534-542. doi:10.1001/jamapsychiatry.2013.736

Bidwell, L. C., Palmer, R. H. C., Brick, L., McGeary, J. E., & Knopik, V. S. (2016). Genome-wide single nucleotide polymorphism heritability of nicotine dependence as a multidimensional phenotype. *Psychological Medicine, 46*(10), 2059-2069. doi:10.1017/S0033291716000453

Breslau, N. (1995). Psychiatric comorbidity of smoking and nicotine dependence. *Behav Genet, 25*(2), 95-101.

Breslau, N., & Johnson, E. O. (2000). Predicting smoking cessation and major depression in nicotine-dependent smokers. *Am J Public Health, 90*(7), 1122-1127.

Breslau, N., Kilbey, M., & Andreski, P. (1991). Nicotine dependence, major depression, and anxiety in young adults. *Arch Gen Psychiatry, 48*(12), 1069-1074.

Breslau, N., Kilbey, M. M., & Andreski, P. (1993). Nicotine Dependence and Major Depression - New Evidence from a Prospective Investigation. *Archives of General Psychiatry, 50*(1), 31-35.

Brick, L. A., Keller, M. C., Knopik, V. S., McGeary, J. E., & Palmer, R. H. C. (2019). Shared additive genetic variation for alcohol dependence among subjects of African and European ancestry. *Addict Biol, 24*(1), 132-144. doi:10.1111/adb.12578

Carey, C. E., Agrawal, A., Bucholz, K. K., Hartz, S. M., Lynskey, M. T., Nelson, E. C., . . . Bogdan, R. (2016). Associations between Polygenic Risk for Psychiatric Disorders and Substance Involvement. *Front Genet, 7*, 149. doi:10.3389/fgene.2016.00149

Chang, L. H., Couvy-Duchesne, B., Liu, M., Medland, S. E., Verhulst, B., Benotsch, E. G., . . . Consortium, G. (2019). Association between polygenic risk for tobacco or alcohol consumption and liability to licit and illicit substance use in young Australian adults. *Drug Alcohol Depend, 197*, 271-279. doi:10.1016/j.drugalcdep.2019.01.015

Chen, J., Bacanu, S. A., Yu, H., Zhao, Z., Jia, P., Kendler, K. S., . . . group, F. m.-a. (2016). Genetic Relationship between Schizophrenia and Nicotine Dependence. *Sci Rep, 6*, 25671. doi:10.1038/srep25671

Chen, L. S., Hartz, S. M., Baker, T. B., Ma, Y., N, L. S., & Bierut, L. J. (2018). Use of polygenic risk scores of nicotine metabolism in predicting smoking behaviors. *Pharmacogenomics, 19*(18), 1383-1394. doi:10.2217/pgs-2018-0081

Choi, J. S., Payne, T. J., Ma, J. Z., & Li, M. D. (2017). Relationship between Personality Traits and Nicotine Dependence in Male and Female Smokers of African-American and European-American Samples. *Front Psychiatry, 8*, 122. doi:10.3389/fpsyt.2017.00122

Criado, J. R., Gizer, I. R., Edenberg, H. J., & Ehlers, C. L. (2014). CHRNA5 and CHRNA3 variants and level of neuroticism in young adult Mexican American men and women. *Twin Res Hum Genet, 17*(2), 80-88. doi:10.1017/thg.2014.11

Dare, S., Mackay, D. F., & Pell, J. P. (2015). Relationship between smoking and obesity: a cross-sectional study of 499,504 middle-aged adults in the UK general population. *PLoS One, 10*(4), e0123579. doi:10.1371/journal.pone.0123579

Das, S., Forer, L., Schönherr, S., Sidore, C., Locke, A. E., Kwong, A., . . . Fuchsberger, C. (2016). Next-generation genotype imputation service and methods. *Nature Genetics, 48*, 1284-1287.

Donny, E. C., & Dierker, L. C. (2007). The absence of DSM-IV nicotine dependence in moderate-to-heavy daily smokers. *Drug Alcohol Depend, 89*(1), 93-96. doi:10.1016/j.drugalcdep.2006.11.019

Evangelou, E., & Ioannidis, J. P. A. (2013). Meta-analysis methods for genome-wide association studies and beyond. *Nature Reviews Genetics, 14*(6), 379-389. doi:10.1038/nrg3472

Eysenck, H. J. (1983). A note on 'Smoking, personality and reasons for smoking. *Psychol Med, 13*(2), 447-448.

Eysenck, S. B. G., Eysenck, H. J., & Barrett, P. (1985). A revised version of the psychoticism scale. *Personality and Individual Differences, 6*(1), 21-29. doi:10.1016/0191-8869(85)90026-1

Fluharty, M., Taylor, A. E., Grabski, M., & Munafo, M. R. (2017). The Association of Cigarette Smoking With Depression and Anxiety: A Systematic Review. *Nicotine Tob Res, 19*(1), 3-13. doi:10.1093/ntr/ntw140

Grant, J. D., Scherrer, J. F., Lynskey, M. T., Agrawal, A., Duncan, A. E., Haber, J. R., . . . Bucholz, K. K. (2012). Associations of alcohol, nicotine, cannabis, and drug use/dependence with educational attainment: evidence from cotwin-control analyses. *Alcohol Clin Exp Res, 36*(8), 1412-1420. doi:10.1111/j.1530-0277.2012.01752.x

Hartz, S. M., Horton, A. C., Hancock, D. B., Baker, T. B., Caporaso, N. E., Chen, L. S., . . . Bierut, L. J. (2018). Genetic correlation between smoking behaviors and schizophrenia. *Schizophr Res, 194*, 86-90. doi:10.1016/j.schres.2017.02.022

Hartz, S. M., Horton, A. C., Oehlert, M., Carey, C. E., Agrawal, A., Bogdan, R., . . . Bierut, L. J. (2017). Association Between Substance Use Disorder and Polygenic Liability to Schizophrenia. *Biol Psychiatry, 82*(10), 709-715. doi:10.1016/j.biopsych.2017.04.020

Hasin, D. S., Stinson, F. S., Ogburn, E., & Grant, B. F. (2007). Prevalence, correlates, disability, and comorbidity of DSM-IV alcohol abuse and dependence in the United States: results from the National Epidemiologic Survey on Alcohol and Related Conditions. *Arch Gen Psychiatry, 64*(7), 830-842. doi:10.1001/archpsyc.64.7.830

Heatherton, T. F., Kozlowski, L. T., Frecker, R. C., & Fagerstrom, K. O. (1991). The Fagerstrom Test for Nicotine Dependence: a revision of the Fagerstrom Tolerance Questionnaire. *Br J Addict, 86*(9), 1119-1127.

Heatherton, T. F., Kozlowski, L. T., Frecker, R. C., Rickert, W., & Robinson, J. (1989). Measuring the heaviness of smoking: using self-reported time to the first cigarette of the day and number of cigarettes smoked per day. *Br J Addict, 84*(7), 791-799.

Hewitt, J. K. (2008). Statistical genetics: Gene mapping through linkage and association. *Behavior Genetics, 38*(3), 337-338. doi:10.1007/s10519-008-9195-2

Howard, D. M., Adams, M. J., Shirali, M., Clarke, T. K., Marioni, R. E., Davies, G., . . . McIntosh, A. M. (2018). Genome-wide association study of depression phenotypes in UK Biobank identifies variants in excitatory synaptic pathways. *Nat Commun, 9*(1), 1470. doi:10.1038/s41467-018-03819-3

Ittermann, T., Thamm, M., Schipf, S., John, U., Rettig, R., & Volzke, H. (2013). Relationship of smoking and/or passive exposure to tobacco smoke on the association between serum thyrotropin and body mass index in large groups of adolescents and children. *Thyroid, 23*(3), 262-268. doi:10.1089/thy.2012.0110

Jack Euesden, Cathryn Lewis, & Paul O'Reilly. (2015). PRSice: Polygenic Risk Score software,. *Bioinformatics, 31*(9), 1466-1468.

Jenks, R. J. (1992). Attitudes, perceptions, and risk-taking behaviors of smokers, ex-smokers, and nonsmokers. *J Soc Psychol, 132*(5), 569-575. doi:10.1080/00224545.1992.9713895

Karlsson Linner, R., Biroli, P., Kong, E., Meddens, S. F. W., Wedow, R., Fontana, M. A., . . . Beauchamp, J. P. (2019). Genome-wide association analyses of risk tolerance and risky behaviors in over 1 million individuals identify hundreds of loci and shared genetic influences. *Nat Genet, 51*(2), 245-257. doi:10.1038/s41588-018-0309-3

Kelly, C., & McCreadie, R. (2000). Cigarette smoking and schizophrenia *Advances in Psychiatric Treatment, 6*, 327-331.

Kheradmand, A., Ziaaddini, H., & Vahabi, M. (2011). Prevalence of Cigarette Smoking in Schizophrenic Patients Compared to Other Hospital Admitted Psychiatric Patients.

Lee, J. J., Wedow, R., Okbay, A., Kong, E., Maghzian, O., Zacher, M., . . . Cesarini, D. (2018). Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nat Genet, 50*(8), 1112-1121. doi:10.1038/s41588-018-0147-3

Lewis, C. M., & Knight, J. (2012). Introduction to genetic association studies. *Cold Spring Harb Protoc, 2012*(3), 297-306. doi:10.1101/pdb.top068163

Li, X., Rainnie, D. G., McCarley, R. W., & Greene, R. W. (1998). Presynaptic nicotinic receptors facilitate monoaminergic transmission. *J Neurosci, 18*(5), 1904-1912.

Liu, J. Z., Tozzi, F., Waterworth, D. M., Pillai, S. G., Muglia, P., Middleton, L., . . . Marchini, J. (2010). Meta-analysis and imputation refines the association of 15q25 with smoking quantity. *Nat Genet, 42*(5), 436-440. doi:10.1038/ng.572

Liu, M., Jiang, Y., Wedow, R., Li, Y., Brazel, D. M., 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*. doi:10.1038/s41588-018-0307-5

Lohr, J. B., & Flynn, K. (1992). Smoking and schizophrenia. *Schizophr Res, 8*(2), 93-102.

Lyons, M., Hitsman, B., Xian, H., Panizzon, M. S., Jerskey, B. A., Santangelo, S., . . . Tsuang, M. T. (2008). A twin study of smoking, nicotine dependence, and major depression in men. *Nicotine Tob Res, 10*(1), 97-108. doi:10.1080/14622200701705332

Maes, H. H., Neale, M. C., Chen, X. N., Chen, J. C., Prescott, C. A., & Kendler, K. S. (2011). A Twin Association Study of Nicotine Dependence with Markers in the CHRNA3 and CHRNA5 Genes. *Behavior Genetics, 41*(5), 680-690. doi:10.1007/s10519-011-9476-z

Maher, A. M., Thomson, C. J., & Carlson, S. R. (2015). Risk-taking and impulsive personality traits in proficient downhill sports enthusiasts. *Personality and Individual Differences, 79*, 20-24. doi:10.1016/j.paid,2015.01.041

Maier, R. M., Zhu, Z., Lee, S. H., Trazaskowski, M., Ruderder, D. M., Stahl, E. A., . . . Robinson, M. R. (2018). Improving genetic prediction by leveraging genetic correlaitons among human diseases and traits *Nature Communications, 9*(1), 989. doi:10.1038/s41467-017-02769-6

Marees, A. T., Hammerschlag, A. R., Bastarache, L., de Kluiver, H., Vorspan, F., van den Brink, W., . . . Derks, E. M. (2018). Exploring the role of low-frequency and rare exonic variants in alcohol and tobacco use. *Drug Alcohol Depend, 188*, 94-101. doi:10.1016/j.drugalcdep.2018.03.026

McCaffery, J. M., Papandonatos, G. D., Lyons, M. J., Koenen, K. C., Tsuang, M. T., & Niaura, R. (2008). Educational attainment, smoking initiation and lifetime nicotine dependence among male Vietnam-era twins. *Psychol Med, 38*(9), 1287-1297. doi:10.1017/S0033291707001882

McCaffery, J. M., Papandonatos, G. D., Lyons, M. J., & Niaura, R. (2008). Educational attainment and the heritability of self-reported hypertension among male Vietnam-era twins. *Psychosom Med, 70*(7), 781-786. doi:10.1097/PSY.0b013e3181817be6

Moonesinghe, R., Khoury, M. J., Liu, T., & Ioannidis, J. P. A. (2008). Required sample size and nonreplicability thresholds for heterogeneous genetic associations. *Proceedings of the National Academy of Sciences of the United States of America, 105*(2), 617-622. doi:10.1073/pnas.0705554105

Munafo, M. R., Tilling, K., & Ben-Shlomo, Y. (2009). Smoking status and body mass index: a longitudinal study. *Nicotine Tob Res, 11*(6), 765-771. doi:10.1093/ntr/ntp062

Musci, R. J., Uhl, G., Maher, B., & Ialongo, N. S. (2015). Testing gene x environment moderation of tobacco and marijuana use trajectories in adolescence and young adulthood. *J Consult Clin Psychol, 83*(5), 866-874. doi:10.1037/a0039537

Muthén, L. K., & Muthén, B. O. (2017). *Mplus User's Guide* (Eighth ed.). Los Angeles, California: Muthén & Muthén.

Nagel, M., Watanabe, K., Stringer, S., Posthuma, D., & van der Sluis, S. (2018). Item-level analyses reveal genetic heterogeneity in neuroticism. *Nat Commun, 9*(1), 905. doi:10.1038/s41467-018-03242-8

National Cancer Institute. NCI Dictionary of Cancer Terms. *pack year.* Retrieved from https://www.cancer.gov/publications/dictionaries/cancer-terms/def/pack-year

O'Loughlin, J., DiFranza, J., Tyndale, R. F., Meshefedjian, G., McMillan-Davey, E., Clarke, P. B., . . . Paradis, G. (2003). Nicotine-dependence symptoms are associated with smoking frequency in adolescents. *Am J Prev Med, 25*(3), 219-225.

O'Loughlin, J., Karp, I., Henderson, M., & Gray-Donald, K. (2008). Does cigarette use influence adiposity or height in adolescence? *Ann Epidemiol, 18*(5), 395-402. doi:10.1016/j.annepidem.2007.12.010

O'Cathall, S. M., O'Connell, O. J., Long, N., Morgan, M., Eustace, J. A., Plant, B. J., & Hourihane, J. O. (2011). Association of cigarette smoking with drug use and risk taking behaviour in Irish teenagers. *Addict Behav, 36*(5), 547-550.

Otto, J. M., Gizer, I. R., Bizon, C., Wilhelmsen, K. C., & Ehlers, C. L. (2016). Polygenic risk scores for cigarettes smoked per day do not generalize to a Native American population. *Drug Alcohol Depend, 167*, 95-102. doi:10.1016/j.drugalcdep.2016.07.029

Palmer, Button, T. M., Rhee, S. H., Corley, R. P., Young, S. E., Stallings, M. C., . . . Hewitt, J. K. (2012). Genetic etiology of the common liability to drug dependence: evidence of common and specific mechanisms for DSM-IV dependence symptoms. *Drug Alcohol Depend, 123 Suppl 1*, S24-32. doi:10.1016/j.drugalcdep.2011.12.015

Palmer, McGeary, J. E., Francazio, S., Raphael, B. J., Lander, A. D., Heath, A. C., & Knopik, V. S. (2012). The genetics of alcohol dependence: advancing towards systems-based approaches. *Drug Alcohol Depend, 125*(3), 179-191. doi:10.1016/j.drugalcdep.2012.07.005

Palmer, Young, S. E., Hopfer, C. J., Corley, R. P., Stallings, M. C., Crowley, T. J., & Hewitt, J. K. (2009). Developmental epidemiology of drug use and abuse in adolescence and young adulthood: Evidence of generalized risk. *Drug Alcohol Depend, 102*(1-3), 78-87. doi:10.1016/j.drugalcdep.2009.01.012

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., . . . Sham, P. C. (2007). PLINK: a toolset for whole-genome association and population-based linkage analysis. *American Journal of Human Genetics, 81*.

R Core Team. (2017). *R: A language and environment for statistical computing*. Vienna, Austria.

Sagud, M., Mihaljevic-Peles, A., Muck-Seler, D., Pivac, N., Vuksan-Cusa, B., Brataljenovic, T., & Jakovljevic, M. (2009). Smoking and schizophrenia. *Psychiatr Danub, 21*(3), 371-375.

Sallis, H. M., Davey Smith, G., & Munafo, M. R. (2018). Cigarette smoking and personality: interrogating causality using Mendelian randomisation. *Psychol Med*, 1-9. doi:10.1017/S0033291718003069

Schizophrenia Working Group of the Psychiatric Genomics Consortium. (2014). Biological insights from 108 schizophrenia-associated genetic loci. *Nature, 511*(7510), 421-427. doi:10.1038/nature13595

Sneve, M., & Jorde, R. (2008). Cross-sectional study on the relationship between body mass index and smoking, and longitudinal changes in body mass index in relation to change in smoking status: the Tromso Study. *Scandanavian Journal of Public Health, 36*(4), 397-407. doi:10.1177/1403494807088453

Spielberger, C. D., & Jacobs, G. A. (1982). Personality and smoking behavior. *J Pers Assess, 46*(4), 396-403. doi:10.1207/s15327752jpa4604_11

Stevens, V. L., Jacobs, E. J., Gapstur, S. M., Carter, B. D., Gaudet, M. M., Westmaas, J. L., & Patel, A. V. (2017). Evaluation of a Novel Difficulty of Smoking Cessation Phenotype Based on Number of Quit Attempts. *Nicotine Tob Res, 19*(4), 435-441. doi:10.1093/ntr/ntw234

Stinson, F. S., Grant, B. F., Dawson, D. A., Ruan, W. J., Huang, B., & Saha, T. (2005). Comorbidity between DSM-IV alcohol and specific drug use disorders in the United States: results from the National Epidemiologic Survey on Alcohol and Related Conditions. *Drug Alcohol Depend, 80*(1), 105-116. doi:10.1016/j.drugalcdep.2005.03.009

Strawbridge, R. J., Ward, J., Cullen, B., Tunbridge, E. M., Hartz, S., Bierut, L., . . . Smith, D. J. (2018). Genome-wide analysis of self-reported risk-taking behaviour and cross-disorder genetic correlations in the UK Biobank cohort. *Translational Psychiatry, 8*(39). doi:10.1038/s41398-017-0079-1

Tate, J. C., Pomerleau, C. S., & Pomerleau, O. F. (1994). Pharmacological and non-pharmacological smoking motives: a replication and extension. *Addiction, 89*(3), 321-330.

Teo, Y. Y. (2008). Common statistical issues in genome-wide association studies: a review on power, data quality control, genotype calling and population structure. *Current Opinion in Lipidology, 19*(2), 133-143. doi:DOI 10.1097/MOL.0b013e3282f5dd77

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

Thorgeirsson, T. E., Gudbjartsson, D. F., Surakka, I., Vink, J. M., Amin, N., Geller, F., . . . Stefansson, K. (2010). Sequence variants at CHRNB3-CHRNA6 and CYP2A6 affect smoking behavior. *Nat Genet, 42*(5), 448-453. doi:10.1038/ng.573

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

Trimpop, R. M. (1994). What Is Risk Taking Behavior. In *The Psychology of Risk Taking Behavior* (Vol. 107, pp. 1-14): Advances in Psychology.

Tsao, H., & Florez, J. C. (2007). Introduction to genetic association studies. *Journal of Investigative Dermatology, 127*(10), 2283-2287. doi:10.1038/sj.jid.5701054

Tully, E. C., Iacono, W. G., & McGue, M. (2010). Changes in genetic and environmental influences on the development of nicotine dependence and major depressive disorder from middle adolescence to early adulthood. *Dev Psychopathol, 22*(4), 831-848. doi:10.1017/S0954579410000490

Turley, P., Walters, R. K., Maghzian, O., Okbay, A., Lee, J. J., Fontana, M. A., . . . Co, S. S. G. A. (2018). Multi-trait analysis of genome-wide association summary statistics using MTAG. *Nature Genetics, 50*(2), 229-+. doi:10.1038/s41588-017-0009-4

U.S. Department of Health and Human Services. (2014). *The Health Consequences of Smoking—50 Years of Progress: A Report of the Surgeon General*. Atlanta, Georgia.

Vink, J. M., Hottenga, J. J., de Geus, E. J., Willemsen, G., Neale, M. C., Furberg, H., & Boomsma, D. I. (2014). Polygenic risk scores for smoking: predictors for alcohol and cannabis use? *Addiction, 109*(7), 1141-1151. doi:10.1111/add.12491

Warren, G. W., Alberg, A. J., Kraft, A. S., & Cummings, K. M. (2014). The 2014 Surgeon General's report: "The health consequences of smoking--50 years of progress": a paradigm shift in cancer care. *Cancer, 120*(13), 1914-1916. doi:10.1002/cncr.28695

Watanabe, T., Tsujino, I., Konno, S., Ito, Y. M., Takashina, C., Sato, T., . . . Nishimura, M. (2016). Association between Smoking Status and Obesity in a Nationwide Survey of Japanese Adults. *PLoS One, 11*(3), e0148926. doi:10.1371/journal.pone.0148926

Wills, A. G., & Hopfer, C. (2019). Phenotypic and genetic relationship between BMI and cigarette smoking in a sample of UK adults. *Addict Behav, 89*, 98-103. doi:10.1016/j.addbeh.2018.09.025

Xu, X., Bishop, E. E., Kennedy, S. M., Simpson, S. A., & Pechacek, T. F. (2015). Annual healthcare spending attributable to cigarette smoking: an update. *Am J Prev Med, 48*(3), 326-333. doi:10.1016/j.amepre.2014.10.012

Yengo, L., Sidorenko, J., Kemper, K. E., Zheng, Z., Wood, A. R., Weedon, M. N., . . . Consortium, G. (2018). Meta-analysis of genome-wide association studies for height and body mass index in approximately 700000 individuals of European ancestry. *Hum Mol Genet, 27*(20), 3641-3649. doi:10.1093/hmg/ddy271

Yin, X., Bizon, C., Tilson, J., Lin, Y., Gizer, I. R., Ehlers, C. L., & Wilhelmsen, K. C. (2017). Genome-wide meta-analysis identifies a novel susceptibility signal at CACNA2D3 for

nicotine dependence. *Am J Med Genet B Neuropsychiatr Genet*. doi:10.1002/ajmg.b.32540

Zinbarg, R. E., Mineka, S., Bobova, L., Craske, M. G., Vrshek-Schallhorn, S., Griffith, J. W., . . . Anand, D. (2016). Testing a Hierarchial Model of Neuroticism and Its Cognitive Facets: Latent Structure and Prospective Prediction of First Onsets of Anxiety and Unipolar Mood Disorders During 3 Years in Late Adolescence. *SAGE Journal, 4*(5), 805-824. doi:10.1177/2167702615618162

Zvolensky, M. J., Taha, F., Bono, A., & Goodwin, R. D. (2015). Big five personality factors and cigarette smoking: a 10-year study among US adults. *J Psychiatr Res, 63*, 91-96. doi:10.1016/j.jpsychires.2015.02.008

Zyphur, M. J., Zhang, Z., Barsky, A. P., & Li, W. D. (2013). An ACE in the hole: Twin family models for applied behavioral genetics research. *Leadership Quarterly, 24*(4), 572-594. doi:10.1016/j.leaqua.2013.04.001