12-19-2023

Investigating the relationship between smoking, genomics, and brain imaging

Yoonhoo Chang
Washington University in St. Louis

WASHINGTON UNIVERSITY IN ST. LOUIS
Division of Biology & Biomedical Sciences
Human and Statistical Genetics

Dissertation Examination Committee:
Laura J. Bierut, Chair
Arpana Agrawal
Andrey Anokhin
Janine Bijsterbosch
Ryan Bogdan

Investigating the Relationship Between Smoking, Genomics, and Brain Imaging

by
Yoonhoo Chang

A dissertation presented to
Washington University in St. Louis
in partial fulfillment of the
requirements for the degree
of Doctor of Philosophy

December 2023
St. Louis, Missouri

Table of Contents

List of Figures

List of Tables

Acknowledgments

This dissertation owes its existence to the invaluable support and motivation provided by the following exceptional individuals.

First of all, my thesis advisor, Dr. Laura Bierut. Thank you for being a stable source of valuable lessons not only about research but also about life. I will never forget to think, socialize, exercise, sleep, and eat well. Most importantly, I will always remember that we should all try to make the world a better place. Thank you for trusting in my potential, pushing me to see the bigger picture, and being the best mentor I could have asked for.

My thesis committee: Drs. Janine Bijsterbosch, Andrey Anokhin, Arpana Agrawal, and Ryan Bodgan. Thank you for always being there whenever I requested one-on-one meetings, and for being a reliable source of honest feedback and guidance.

Current and past members of the Bierut group. I thank Michael Bray and Jacob Borodovsky, my academic uncles, for the valuable pieces of advice and moral support. I extend my gratitude to Louis Fox, whose guidance was a guiding light in the complex realm of statistics and data science. I'd also like to express my appreciation to Sherri Fisher, who expertly guided me through the challenging manuscript submission process. Thanks to Tina Hoffman, who, with her exceptional administrative skills, played a crucial role in scheduling meetings and advancing my PhD progress. To Dr. Li-Shiun Chen, Dr. Alex Ramsey, and Giang Pham, I am grateful for the opportunity to be a part of your amazing projects.

I'd like to thank Dr. Youssef Idaghdour and the entire Idaghdour lab team, including Massar Dieng, Wael Said Abdrabou, and Manikandan Vinu, for igniting my scientific passion. If it weren't for our meeting, I might not have ventured into the world of genetics and research. I hope my time in the lab is remembered positively, perhaps as the adventurous soul with ever-changing hair colors who also dabbled in skydiving.

Thank you, Sara Holmes, Drs. Nathan Stitziel, John Rice, and Nancy Saccone, for being instrumental in my introduction to the HSG program and the St. Louis community. Your warm welcome, admission, support, and the one-on-one meetings have been truly appreciated.

Thank you to my co-authors, fellow hocus-pocus witches/younglings, Vera Thornton and Ariya Chaloemtoem. I owe my accomplishments in this lab to you, and I'm immensely grateful. Your unwavering support, teamwork, and sharing of the best memes for mental support have been truly invaluable.

To my NYUAD-to-HSG family, JooHee Choi, Jian Ryou, and Ariya Chaloemtoem:
JooHee, my mother, I am forever grateful to you for kickstarting this lineage and guiding me through the challenges of my first year in my Ph.D. program. Jian and Ariya, my daughter and granddaughter, thank you for always being there, and I will cherish the moments we've shared. I wish you all the best in your future endeavors.

I am grateful for those who shared in the Disney magic with me: Hakyung Lee, Michelle Cho, Ariya Chaloemtoem, Jian Ryou, Juhee Son, and Stacey Sohn. Your companionship during our Disney adventures, reliving the joys of childhood, fills me with deep appreciation.

My university friends, Sally Oh and Dahee Kim. The countless cups of tea and stacks of cookies we indulged in will forever hold a special place in my heart. Your unwavering presence, just a message away when I needed it most, and the warm welcome to Korea each time I return, mean the world to me. I miss you dearly, and I genuinely believe I couldn't have achieved anything without your attentive listening and your unconditional support, often delivered through adorable pet photos. I also would like to extend my deepest gratitude to Shin Won Kim, who was always one call away when I needed her wisdom.

Thank you to the high school friends who share unique memories from the Gangwon province: Na Yeon Stephanie Kim, So Yoon Lee, HaRyung Kim, Do-Hyeong Myeong, Seung Hee Chae, and Jooyeon Chae. With a tip of my academic hat, your presence and our frequent conversations were a source of encouragement to my PhD journey.

I extend my gratitude to my graduate school friends, with a special mention to those within 'The Housewarming Group.' I'm immensely thankful for our shared watch parties, cinema outings, potlucks, brunches, and the countless conversations revolving around our PhD journeys. Your collective presence and camaraderie have been a profound source of inspiration, motivating me to grow as both a scientist and a human being.

I thank my parents (Ji Min Chang and Min Jung Park), and grandparents (Ju Cheol Chang, Seung Ja Chung, Jeong Hee Lee) for your endless love and support. And my dear grandmother, I remember the very first English phrase you taught me: “Variety is the spice of life.” Ever since that moment, I've embraced this adage as my guiding principle. Though at times it leads to a touch of chaos, more often than not, it results in moments of pure joy. Words fail to convey the depth of my gratitude adequately; I am truly fortunate to be your granddaughter. I also would like to express gratitude to my aunt and uncle (Ji Hwang Chang and Jong In Choi) and my younger siblings (Ikjun Chang, Yejun Chang, Yeseo Chang).

Lastly, I extend my thanks to the feline and canine companions of my friends whose presence added a touch of serenity during the long nights of research: Miumiu, Gustav, Cenizo, Woojoo, and Bona. May your tails wag with joy and your whiskers twitch with contentment.

Yoonhoo Chang
St. Louis, MO
December 2023

ABSTRACT OF THE DISSERTATION

Investigating the relationship between smoking, genomics, and brain imaging

By
Yoonhoo Chang

Doctor of Philosophy in Biology and Biomedical Sciences
Human and Statistical Genetics
Washington University in St. Louis, 2023
Professor Laura J. Bierut, Chair

Cigarette smoking has been linked to adverse health outcomes, including several forms of cancer, respiratory and cardiovascular disease, and dementia. Despite the public health campaigns aimed at reducing tobacco use, the behavior remains prevalent, and its association with various organs is an active area of research. Recent studies showed that smoking behaviors have strong genetic contributions. Through genome-wide association studies, the genetic risk score of individuals can be created and used as a basis for more effective and precise healthcare solutions. Chapter 2 explores the creation of a polygenic risk score (PRS) for smoking cessation and its utility in two clinical trials. PRS for the later age of smoking initiation, and the combined PRS of four smoking behaviors (later age of smoking initiation, persistent smoking, cigarettes per day, and ever smoking) predicted bio-verified smoking abstinence. With this information, chapter 3 examines the utility of PRS in a more general setting (UK Biobank) with healthy individuals. The PRS for persistent smoking significantly predicted smoking cessation in the UK Biobank population. Also, individuals were divided into 7 groups of different PRS for persistent smoking, and those with the risk score of bottom 10% and top 10% had 3 years difference in median age of smoking cessation.

Chapter 4 extends the relationship between genetics and smoking behavior to an actual association with the relatively understudied organ in tobacco research: the brain. It is widely known that smoking adversely affects the lungs and heart, but studies on the relationship

between smoking and the brain are comparatively lacking. Using the smoking questionnaire data, brain imaging data, and genetic data from the UK Biobank, I found that the genetic risk for smoking is not associated with brain volume, while ever smoking was significantly negatively associated with the total brain volume. Furthermore, I also identified several regions more significantly associated with a history of daily smoking than the other regions.

Chapter 5 explores the association of smoking with the structural and functional connectivity of the brain. The UK Biobank provides diffusion MRI-derived phenotypes that measure structural connectivity within and across the regions of the brain. I examined the inter/intra-regional tracts that are more affected by smoking than others. For the functional connectivity, I used the resting-functional MRI-derived phenotypes and found that the functional connectivity within frontal lobe regions was modestly associated with ever smoking. Overall, this dissertation advances our understanding of the clinical utility of PRS for smoking behaviors, and the association of smoking behavior with the brain.

Chapter 1. Introduction

1.1 Genetics of smoking behavior

Every year, cigarette smoking contributes to more than 8 million preventable deaths worldwide, with an additional 1.3 million deaths attributed to second-hand smoke [1]. Despite increased public awareness of its health impacts, less than 10% of the smoking population successfully quit annually [2]. Understanding the pivotal role genetics play in the challenge of quitting has led researchers to delve into the genetic underpinnings of smoking behaviors.

The period following the completion of the Human Genome Project (1990-2003) marked a new era in genetic research [3]. Genome-wide association Studies (GWAS) leveraged millions of single-nucleotide polymorphisms (SNPs) to explore their association with various smoking traits [3]. This exploration included identifying SNPs linked to dichotomous traits, like ever versus never smoking, as well as quantitative traits, such as cigarettes smoked per day and lifetime number of pack years smoked (number of cigarette packs (one pack = 20 cigarettes) smoked per day times the number of years smoked). Over the years, GWAS studies unearthed genes linked to nicotine use disorder [4, 5] replicated findings on genes associated with smoking quantity [6, 7], and further established connections between nicotine use disorder and specific genes [8]. Variation in nicotinic acetylcholine receptor subunits and nicotine metabolizing genes are the strongest findings and thousands of other variants of small effect are associated. These genetic determinants of smoking were identified as significant risk factors for lung cancer [9] and chronic obstructive pulmonary disease (COPD) [10], intensifying public awareness of smoking's detrimental health effects.

More recently, meta-analyses of over 3.4 million individuals (GSCAN) consolidated genetic variants associated with four smoking behaviors (smoking initiation, age of smoking initiation, cigarettes per day, persistent smoking/smoking cessation) and drinks per week [11]. The high-powered summary statistics enable the creation of polygenic risk scores (PRSs) [12]. A PRS is made by combining the effects of many genetic signals from a GWAS into a single risk variable, which can be used to estimate an individual's genetic propensity to develop a disease or trait. Ultimately, the goal of precision medicine is to perform a genetically informed intervention that motivates people to quit smoking [13-15]. The return of the polygenic risk scores that are categorized, and interpreted is an important step toward this goal [13-15].

1.2 Neuroimaging and smoking behavior

While smoking affects various organs of the human body, its connection with the brain remains relatively underexplored. Studies have shown that cognitive decline and dementia are associated with cigarette smoking [16-18]. However, understanding the direct relationship between smoking through neuroimaging has been hindered by small sample sizes, resulting in underpowered findings [19].

Recent biobank studies, particularly utilizing data from the UK Biobank, have provided a critical opportunity to examine the impact of smoking on a larger and healthier populations. These investigations have delved into the correlation between smoking behaviors and brain-related changes [20-23], revealing a noticeable decrease in brain volume associated with smoking. Unraveling the neurobiology of substance use has been complex, but throughout the years, researchers have identified the specific brain regions involved in addiction, shaping the behavior changes in individuals during substance use [24, 25]. Still, an ongoing area of debate and research revolves around the scope of smoking's effect on the brain—whether it manifests globally or within specific regions—and whether this effect is reversible [26-29].

1.3 Scope of dissertation

In summary, the genetic studies of smoking provided many opportunities to expand precision medicine and motivate people to reduce/quit smoking. Building on genetic discovery in large scale genome wide association studies, I focused on validating the potential clinical utility of using genetic predictors in a clinical setting and a general population setting. Then I combined genetic predictors to further examine the relationship between smoking behaviors and neuroimaging measures.

Chapter 2 assesses the use of polygenic risk scores and a novel combined polygenic risk score to predict smoking cessation in two clinical trials. The summary statistics used to create the individual PRSs come from GWAS & Sequencing Consortium of Alcohol and Nicotine use (GSCAN), and the participant data is from the Genetically Informed Smoking Cessation Trial (GISC) and the Transdisciplinary Tobacco Use Research Centers (TTURC). Our findings show that the genetic risk scores of smoking behaviors predict smoking cessation in a clinical setting and potentially we can use a combined polygenic risk score to further improve smoking cessation success. This chapter has been published in the Journal *Nicotine & Tobacco Research*.

Chapter 3 assesses the use of polygenic risk scores to predict smoking cessation in a general population setting. The summary statistics used to create the PRS come from GSCAN2, and the UK Biobank provides the participant data. We have identified that the addition of genetic predictors (PRS) significantly improves the model beyond clinical predictors and found that the median age of smoking cessation increases as the genetic risk of persistent smoking increases by

3 years in the highest decile of genetic risk compared lowest decile of genetic risk. At the time of dissertation defense, this chapter was in preparation for submission.

Chapter 4 examines the association between smoking, the brain, and the genetic background of an individual in the UK Biobank dataset. Using summary statistics from the largest genome wide association study of smoking behaviors at the time, polygenic risk scores were developed for each participant in UK Biobank. Our findings in this study showed that smoking behavior is strongly associated with a decrease in global brain volume and a decrease in structural/functional connectivity. We have also identified the negative association between pack years of smoking and the brain volume and there is no evidence for recovery after smoking cessation. Additionally, we provided evidence that genetics influence smoking behavior, and smoking behavior in turn influences the brain imaging. The major findings of this chapter have been published in the journal *Biological Psychiatry: Global Open Science*.

Collectively, this Thesis shares representative work from my PhD and explores the fields of substance use, genomics, and neuroimaging.

1.4 References

1. World Health Organization. *Tobacco*; 2020 [cited 2020 8 October]. Available from: https://www.who.int/news-room/fact-sheets/detail/tobacco.

2. CDC. *Smoking & Tobacco Use*; 2020. Available from: https://www.cdc.gov/tobacco/data_statistics/fact_sheets/cessation/smokingcessation-fast-facts/index.html.

3. Collins FS, Green ED, Guttmacher AE, Guyer MS, Institute UNHGR. A vision for the future of genomics research. *Nature.* 2003;422(6934):835-847.

4. Bierut LJ, Madden PAF, Breslau N, Johnson EO, Hatsukami D, Pomerleau OF, et al. (2006): Novel genes identified in a high-density genome wide association study for nicotine dependence. *Human Molecular Genetics*. 16:24-35.

5. Bierut LJ, Stitzel JA, Wang JC, Hinrichs AL, Grucza GA, Xuei X, et al. (2008): Variants in Nicotinic Receptors and Risk for Nicotine Dependence. *American Journal of Psychiatry*. 165:1163-1171.

6. Saccone NL, Culverhouse RC, Schwantes-An TH, Cannon DS, Chen X, Cichon S, et al. (2010): Multiple independent loci at chromosome 15q25.1 affect smoking quantity: a meta-analysis and comparison with lung cancer and COPD. *PLoS Genet*. 6.

7. Berrettini W, Yuan X, Tozzi F, Song K, Francks C, Chilcoat H, et al. (2008): Alpha-5/alpha-3 nicotinic receptor subunit alleles increase risk for heavy smoking. *Mol Psychiatry*. 13:368-373.

8. Thorgeirsson TE, Geller F, Sulem P, Rafnar T, Wiste A, Magnusson KP, et al. (2008): A variant associated with nicotine dependence, lung cancer and peripheral arterial disease. *Nature*. 452:638-642.

9. Amos CI, Wu X, Broderick P, Gorlov IP, Gu J, Eisen T, et al. (2008): Genome-wide association scan of tag SNPs identifies a susceptibility locus for lung cancer at 15q25.1. *Nat Genet*. 40:616-622.

10. Pillai SG, Ge D, Zhu G, Kong X, Shianna KV, Need AC, et al. (2009): A genome-wide association study in chronic obstructive pulmonary disease (COPD): identification of two major susceptibility loci. *PLoS Genet*. 5:e1000421.

11. Saunders GR, Wang X, Chen F, Jang S-K, Liu M, Wang C, et al. (2022): Genetic diversity fuels gene discovery for tobacco and alcohol use. *Nature*. 612:720-724.

12. Choi SW, Mak TS-H, O'Reilly PF (2020): Tutorial: a guide to performing polygenic risk score analyses. *Nature Protocols*. 15:2759-2772.

13. Chen LS, Bloom AJ, Baker TB, Smith SS, Piper ME, Martinez M, et al. (2014): Pharmacotherapy effects on smoking cessation vary with nicotine metabolism gene (CYP2A6). *Addiction*. 109:128-137.

14. Ramsey AT, Bourdon JL, Bray M, Dorsey A, Zalik M, Pietka A, et al. (2021): Proof of Concept of a Personalized Genetic Risk Tool to Promote Smoking Cessation: High Acceptability and Reduced Cigarette Smoking. *Cancer Prev Res (Phila)*. 14:253-262.

15. Bray M, Chang Y, Baker TB, Jorenby D, Carney RM, Fox L, et al. (2022): The Promise of Polygenic Risk Prediction in Smoking Cessation: Evidence From Two Treatment Trials. *Nicotine Tob Res*. 24:1573-1580.

16. Zhong G, Wang Y, Zhang Y, Guo JJ, Zhao Y (2015): Smoking is associated with an increased risk of dementia: a meta-analysis of prospective cohort studies with investigation of potential effect modifiers. *PLoS One*. 10:e0118333.

17. Livingston G, Huntley J, Sommerlad A, Ames D, Ballard C, Banerjee S, et al. (2020): Dementia prevention, intervention, and care: 2020 report of the Lancet Commission. *Lancet*. 396:413-446.

18. Durazzo TC, Mattsson N, Weiner MW (2014): Smoking and increased Alzheimer's disease risk: a review of potential mechanisms. *Alzheimers Dement*. 10:S122-145.

19. Marek S, Tervo-Clemmens B, Calabro FJ, Montez DF, Kay BP, Hatoum AS, et al. (2022): Reproducible brain-wide association studies require thousands of individuals. *Nature*. 603:654-660.

20. Gray JC, Thompson M, Bachman C, Owens MM, Murphy M, Palmer R (2020): Associations of cigarette smoking with gray and white matter in the UK Biobank. *Neuropsychopharmacology*. 45:1215-1222.

21. Peng P, Li M, Liu H, Tian Y-R, Chu S-L, Van Halm-Lutterodt N, et al. (2018): Brain structure alterations in respect to tobacco consumption and nicotine dependence: a comparative voxel-based morphometry study. *Front Neuroanat*. 12:43.

22. Fritz H-C, Wittfeld K, Schmidt CO, Domin M, Grabe HJ, Hegenscheid K, et al. (2014): Current smoking and reduced gray matter volume—a voxel-based morphometry study. *Neuropsychopharmacology*. 39:2594-2600.

23. Elbejjani M, Auer R, Jacobs DR Jr., Haight T, Davatzikos C, Goff DC Jr., et al. (2019): Cigarette smoking and gray matter brain volumes in middle age adults: the CARDIA Brain MRI sub-study. *Transl Psychiatry.* 9:78.

24. Bogdan R, Hatoum AS, Johnson EC, Agrawal A (2023): The Genetically Informed Neurobiology of Addiction (GINA) model. *Nature Reviews Neuroscience*. 24:40-57.

25. Koob GF, Volkow ND (2016): Neurobiology of addiction: a neurocircuitry analysis. *Lancet Psychiatry*. 3:760-773.

26. Hatoum AS, Johnson EC, Agrawal A, Bogdan R (2021): Brain structure and problematic alcohol use: a test of plausible causation using latent causal variable analysis. *Brain Imaging and Behavior*. 15:2741-2745.

27. Logtenberg E, Overbeek MF, Pasman JA, Abdellaoui A, Luijten M, Van Holst RJ, et al. (2022): Investigating the causal nature of the relationship of subcortical brain volume with smoking and alcohol use. *The British Journal of Psychiatry*. 221:377-385.

28. Lin W, Zhu L, Lu Y (2023): Association of smoking with brain gray and white matter volume: a Mendelian randomization study. *Neurological Sciences.* 1-7.

29. Baranger DA, Demers CH, Elsayed NM, Knodt AR, Radtke SR, Desmarais A, et al. (2020): Convergent evidence for predispositional effects of brain gray matter volume on alcohol consumption. *Biological Psychiatry*. 87:645-655.

Chapter 2. The promise of polygenic risk prediction in smoking cessation: Evidence from two treatment trials

2.1 Abstract

Introduction: Tobacco use disorder is a complex behavior with a strong genetic component. Genome-wide association studies (GWAS) on smoking behaviors allow for the creation of polygenic risk scores (PRSs) to approximate genetic vulnerability. However, the utility of smoking-related PRSs in predicting smoking cessation in clinical trials remains unknown.

Aims and methods: We evaluated the association between polygenic risk scores and bioverified smoking abstinence in a meta-analysis of two randomized, placebo-controlled smoking cessation trials. PRSs of smoking behaviors were created using the GWAS and Sequencing Consortium of Alcohol and Nicotine use (GSCAN) consortium summary statistics. We evaluated the utility of using individual PRS of specific smoking behavior versus a combined genetic risk that combines PRS of all four smoking behaviors. Study participants came from the Transdisciplinary Tobacco Use Research Centers (TTURCs) Study (1091 smokers of European descent), and the Genetically Informed Smoking Cessation Trial (GISC) Study (501 smokers of European descent).

Results: PRS of later age of smoking initiation (OR [95% CI]: 1.20, [1.04-1.37], p = .0097) was significantly associated with bioverified smoking abstinence at end of treatment. In addition, the combined PRS of smoking behaviors also significantly predicted bioverified smoking abstinence (OR [95% CI] 0.71 [0.51-0.99], p = .045).

Conclusions: PRS of later age at smoking initiation may be useful in predicting smoking cessation at the end of treatment. A combined PRS may be a useful predictor for smoking abstinence by capturing the genetic propensity for multiple smoking behaviors.

Implications: There is a potential for polygenic risk scores to inform future clinical medicine, and a great need for evidence on whether these scores predict clinically meaningful outcomes. Our meta-analysis provides early evidence for potential utility of using polygenic risk scores to predict smoking cessation amongst smokers undergoing quit attempts, informing further work to optimize the use of polygenic risk scores in clinical care.

2.2 Introduction

Tobacco use disorder is the leading cause of preventable death worldwide [1]. In the United States, cigarette smoking contributes to approximately one in five deaths annually, which results in a 10-year reduced life expectancy [2]. However, successful smoking cessation can improve long-term health outcomes. For example, individuals who quit smoking before the age of 40 reduce the risk of a smoking-related death by ~90% [2, 3]. Unfortunately, cigarette smoking is highly addictive and although 68% of adult smokers desire to quit, less than 10% of adult smokers successfully quit annually [4].

The ability to successfully quit smoking is heritable, with heritability estimates from twin studies explaining up to 54% of variance in smoking cessation outcomes [5]. However, the genetic influences of smoking cessation are complex [6]. Large genome-wide association studies (GWAS) of smoking cessation and other smoking-related behaviors highlighted two well-known and documented genetic loci. The first genetic locus is *CHRNA5* on chromosomal region 15q25 and is a nicotinic receptor [7]. Within *CHRNA5*, one SNP, rs16969968, in particular drives this association [7–10]. The rs16969968 risk allele (A) is associated with a lower likelihood of smoking cessation, and the high-risk genotype (AA) is associated with a four-year earlier median age of lung cancer diagnosis [11–13]. The second genetic loci is *CYP2A6* on chromosomal region 19q13 [6, 7] and is the primary nicotine metabolizing gene [14]. *CYP2A6* is highly polymorphic, and genetic variation within *CYP2A6* is associated with changes with nicotine metabolism [15], which can affect rates of smoking cessation [16, 17].

Recent well-powered genome-wide association studies (GWAS) have identified many genetic variants associated with various smoking behaviors [7, 18, 19]. Specifically, the GWAS and Sequencing Consortium of Alcohol and Nicotine use (GSCAN) discovered over 500 genetic variants associated with four key smoking behaviors: ever smoking, later age of smoking initiation, cigarettes smoked per day, and persistent smoking (failed smoking cessation), allowing for the opportunity to develop polygenic risk scores (PRSs) [7]. A PRS is made by combining the effects of many genetic signals from a GWAS into a single risk variable, which can be used to estimate an individual's genetic propensity to develop a disease or trait. PRSs have been useful in predicting health outcomes including lung cancer, other cancers, and psychiatric disorders [20–24].

Existing research has demonstrated the utility of PRS in predicting smoking behaviors, indicating the potential use of PRSs in clinical care [25, 26]. For example, Belsky and colleagues (2013) [26] observed that individuals with high smoking-related polygenic risk scores were more likely to develop nicotine use disorder if they smoked and were less likely to quit smoking [26]. Other examples evaluate how PRS of nicotine metabolism markers can be used to predict smoking cessation [27, 28]. For example, research suggests the potential of using PRSs in predicting nicotine metabolism that could potentially inform treatment response. Although research is emerging with the use of PRSs in the prediction of smoking behaviors and health outcomes, there is still a gap of knowledge in whether PRSs of smoking behaviors derived from these large population studies predict clinical smoking cessation success among smokers making a quit attempt.

Previous research has shown that variation in individual genes such as *CHRNA5* [8, 11, 12] and *CYP2A6* [14, 15] can predict smoking cessation in both population studies and clinical trials. For example, evidence from University of Wisconsin Transdisciplinary Tobacco Use Research Center (UW-TTURC) trial suggested that *CHRNA5* genotypes may moderate the responses to nicotine replacement [12] in individuals of European ancestry. More recent evidence from the Genetically Informed Smoking Cessation Trial (GISC) trial suggested that *CHRNA5* genotypes may moderate the response to nicotine replacement and varenicline in individuals of African American ancestry [29]. However, the results of such individual gene prediction studies have not always been consistent [30–32]. Recent GWAS have identified PRSs that predict smoking behaviors in large cross-sectional population studies. To the extent that PRSs comprise multiple genetic variants which may enhance the prediction of cessation outcomes, their use should result in more accurate and consistent associations with such outcomes. We hypothesize that smokers with higher PRSs of problematic smoking behaviors (e.g. earlier onset, heavy smoking, or persistent smoking) are less likely to achieve smoking cessation in clinical trials. This hypothesis is based on the notion that characteristics, such as age of initiating regular smoking and smoking heaviness, have shown strong associations with tobacco use disorder and smoking cessation failure in previous research [33–35] Using meta-analyses of two randomized placebo-controlled smoking cessation trials, we examine the utility of PRSs of four key smoking behaviors (ever smoking, later age of smoking initiation, cigarettes per day, and persistent smoking) in predicting smoking cessation among smokers attempting to quit [2, 7].

2.3 Methods

2.3.1 Study Samples

The Transdisciplinary Tobacco Use Research Centers (TTURCs) Study

The TTURC study is a randomized, placebo-controlled smoking cessation clinical trial at the University of Wisconsin Center for Tobacco Research and Intervention focusing on the genetic association of time to relapse after quitting [13]. Each participant was randomly assigned to one of the six conditions: (1) placebo, (2) nicotine patch, (3) nicotine lozenge, (4) sustained-release bupropion, (5) nicotine patch and nicotine lozenge (combination nicotine replacement [C-NRT]), or (6) bupropion and nicotine lozenge. In addition, all participants received individual cessation counseling. For this study, 1091 smokers of European ancestry were included.

At the end of treatment (8 weeks), bioverified smoking abstinence was verified by expired-carbon monoxide level of less than 10 ppm, documenting abstinence at post treatment.

The Genetically Informed Smoking Cessation Trial (GISC) Study

The GISC is a prospective, randomized, placebo-controlled smoking cessation trial conducted at Washington University in St Louis [29]. Each participant was randomly assigned into one of the three groups stratified by genotypes of rs16969968: nicotine patch and nicotine lozenge (C-NRT), varenicline tartrate, or placebo. All participants received cessation counseling. For this study, 501 smokers of European ancestry were included.

The primary outcome is 7-day point prevalence bioverified smoking abstinence at end of treatment (week 12). All participants self-reported smoking status for the primary endpoint, verified by an expired carbon monoxide level of less than 8 ppm.

Ethical Review Board Both studies were approved by the appropriate institutional review boards and all participants provided informed consent.

2.3.2 Genetic Data

Genotyping

TTURC participants were genotyped at the Center for Inherited Disease Research at Johns Hopkins University using the Illumina Omni2.5 microarray. Gene-Environment Association Studies (GENEVA) Coordinating Center at the University of Washington led the data cleaning process. GISC participants were genotyped using Illumina Global Microarray.

Imputation

Using PLINK software [36], standard GWAS QC was performed to TTURC and GISC datasets. Single nucleotide polymorphisms (SNPs) were aligned to the 1000 Genomes Reference (+strand, build 37) and imputed on the University of Michigan Imputation server using 1000 Genome Reference (build 37, phase 5) [9, 37]. Pre-imputed QC steps included: removing individuals with low genotyping efficiency (<95%), removing related subjects, removing subjects with discordant or inconsistency between reported and reported sex. We also removed SNPs with low genotyping efficiencies (<95%), SNPs with low minor allele frequencies (MAF) (<0.01), and SNPs with no chromosome location. Prior to genetic imputation, SNPs were aligned to the + strand of the 1,000 Genomes. Genotyped SNPs were imputed within the University of Michigan Imputation server using the 1000 Genomes build 37 phase 5 reference panel. Imputed SNPs that had an info score ≥0.9 and a minor allele frequency ≥1% were converted to hard calls. After imputation and QC, there were 47,109,470 SNPs in UW-TTURC and 47,109,465 SNPs in GISC.

Principal components (PCs) of genetic ancestry for GISC were created using EIGENTRAT [38]. Information on how PCs were generated for TTURC can be found in the database of Genotypes and Phenotypes (dbGAP) repository (https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000404.v1.p1) under accession number (phs000404.v1.p1).

Polygenic Risk Scores

We generated PRSs of four smoking behavior phenotypes (ever smoking, later age of smoking initiation, cigarettes smoked per day, and persistent smoking [failed smoking cessation]) from GSCAN GWAS summary statistics using PRSice software [39]. SNPs in each data set were pruned by PRSice (version 1.23) using p-value–informed linkage disequilibrium clumping: R2 <0.10 in a 500-kb window, collapsed to the most significant variant. For each of the four smoking-related summary statistics, we tested the predictability of PRSs that were generated from eight p-value thresholds (.5, .05, .005, .0005, 5×10^{-5}, 5×10^{-6}, 5×10^{-7}, 5×10^{-8}). In general, the inclusion of more SNPs led to more variance explained of the outcome and thus, all subsequent analyses involved PRSs generated using SNPs with p-value threshold of .5 or greater unless otherwise noted. For the GISC data set, there was a total of 136,874 SNPs for generating the later age of smoking initiation PRS, 136,689 SNPs for the cigarettes per day PRS, 135,124 SNPs for the ever smoking PRS, and 135,609 SNPs for the persistent smoking PRS. For the TTURC data set, there was a total of 183,998 SNPs for generating the later age of smoking initiation PRS, 183,342 SNPs for the cigarettes per day PRS, 181,341 SNPs for the ever smoking PRS, and 181,362 SNPs for the persistent smoking PRS. To ensure the interpretability across PRSs of different traits, all PRSs were standardized to Z-scores.

2.3.3. Combined Polygenic Risk Scores

In addition, we generated a combined PRS for risk smoking behaviors based on PRSs of four smoking behaviors (ever smoking, earlier initiation, higher cigarettes per day, persistent smoking) by taking the mean of all four z-transformed PRSs. We acknowledge the mean is an agnostic combination of these PRSs and have shown the correlation of these PRSs in the samples (Table S1).

2.3.4 Statistical Analysis

We modeled the association of each PRS to predict the outcome of bioverified smoking abstinence at end of treatment in both trials. Z-transformed PRSs were modeled as continuous variables. We examined the associations between PRSs and bioverified smoking abstinence at the end of the treatment in both trials using logistic regression models in R [40]. We included the following covariates: age, sex, PC1, and PC2. Additional covariates included cigarettes smoked per day (CPD) and treatment by FDA-approved medication (GISC: combination nicotine replacement therapy (nicotine patch, nicotine lozenge), varenicline; TTURC: combination nicotine replacement therapy, bupropion). PRSs were converted to quartiles for the ease of interpretation. In addition, meta-analysis of both trials was performed in R (metafor package) [40]. We reported both fixed and random effects models. In addition, we generated receiver operating characteristic curve (ROC) and estimated the area under the curve (AUC) for the prediction of smoking cessation using clinical predictors and genetic predictors (SAS 9.4).

2.4 Results

2.4.1 Sample Characteristics

We examined whether polygenic risk scores of smoking behaviors were associated with bioverified smoking abstinence in both smoking cessation trials, TTURC and GISC. The sample characteristics for these two trials are shown in Table 1. The samples from both studies are of European descent. The mean ages were 46.6 (GISC) and 44.5 (TTURC). The ratio of males and females was similar between the two trials (55.9% female for GISC and 58.1% female for TTURC). The mean baseline CPD was 19.1 for GISC and 21.8 for TTURC. Bioverified smoking abstinence at the end of treatment was 20.6% for GISC (N = 103) and 47.5% for TTURC (N = 518) (Table 1).

2.4.2. PRS of Specific Smoking Behaviors and Bioverified Smoking Abstinence

We evaluated the association between each PRS (ever smoking, later age of smoking initiation, cigarettes per day, and persistent smoking) and end of treatment bioverified smoking abstinence using meta-analyses from the results of the two trials. In random effect meta-analysis models (Figure 1), the PRS of later age of smoking initiation was significantly associated with bioverified smoking abstinence at end of treatment (OR [95% CI]: 1.20, [1.04–1.37], p = .0097). Specifically, smokers with PRS-later age of smoking initiation in the highest quartile compared with those with the lowest quartile were more likely to quit successfully (45.1% vs. 32.8%, Figure 1, a). The PRSs of the other risk smoking behaviors were associated with reduced abstinence but did not reach statistical significance.

2.4.3. Combined PRS of Smoking Behaviors and Bioverified Smoking Abstinence

We evaluated and meta-analyzed the association of the combined PRS of smoking behaviors and bioverified smoking abstinence. These four PRSs had low correlations (all ≤ 0.16) (Supplementary Table S1). Using an agnostic approach, we computed the combined PRS of smoking behaviors using the mean of these four PRSs. This combined PRS of smoking behavior was significantly associated with bioverified smoking abstinence (random effect model OR [95% CI]: 0.71 [0.51–0.99], p = .045) (Figure 2). Smokers with the highest quartile compared with those in the lowest quartile of the combined PRS were less likely to quit smoking (33.0% vs. 48.5%, Figure 2).

2.4.4. Prediction of Bioverified Smoking Abstinence

We evaluated whether the addition of basic clinical predictors (cigarettes per day and treatment), additional clinical predictors (baseline Fagerstrom Test for Nicotine Dependence [FTND], baseline carbon monoxide [CO], age of smoking initiation, history of depression/anxiety, and the addition of genetic predictors (PRSs of smoking behaviors) beyond use of the demographic predictors increases the prediction of bioverified smoking abstinence at end of treatment (Figure 3). In evaluating the utility of adding genetic predictors to clinical predictors, we have compared model prediction with the area under the curve (AUC). We found that adding basic clinical, additional clinical, and genetic predictors significantly increased the AUC (basic clinical predictors—0.64 to 0.69, p < .0001; additional clinical predictors—0.69 to 0.70, p = .019; genetic predictors—0.70 to 0.71, p = .034).

2.5 Discussion

Our findings provide novel evidence for the utility of PRSs derived from smoking behaviors to predict smoking cessation in clinical trials. Meta-analysis of two trials revealed associations between PRSs and successful smoking cessation at the end of treatment. Specifically, the PRS of later age of initiation predicts an increase in successful smoking cessation, and the PRS of persistent smoking and number of cigarettes smoked per day trended toward predicted successful smoking cessation but did not reach statistical significance. In addition, a combined PRS, which summarizes the PRSs of multiple aspects of smoking behaviors, could be useful in predicting smoking cessation among smokers making a quit attempt.

Most genetically informed research of smoking-related outcomes focus on single gene regions encompassing genes *CHRNA5* [8, 11, 12] and *CYP2A6* [14, 15, 17]. For example, meta-analyses show strong evidence of the associations between *CHRNA5* on chromosome 15q25 with cigarettes smoked per day, smoking cessation, and lung cancer [8]. In addition, several studies have focused on the *CYP2A6* region encompassing chromosome 19q13 [6, 14]. *CYP2A6* encodes the enzyme that is the primary metabolizer of nicotine [14] and is, therefore, a proxy for the nicotine metabolite ratio (NMR), a biomarker of nicotine metabolism. There is substantial evidence that *CYP2A6* is associated with smoking cessation. For instance, slow metabolism associated with this gene is associated with increased smoking cessation during adolescence [41]. In addition, NMR has been implicated in smoking cessation in a previous study by Lerman and colleagues (2015) [42]. These authors observed that amongst individuals taking varenicline, normal nicotine metabolizers were more likely to quit smoking than slow metabolizers [42].

Thus, there is evidence that both *CHRNA5* and *CYP2A6* provide information regarding the risk of smoking cessation success. Our findings, however, present evidence that PRSs based on many variants across the genome that are derived from smoking behaviors such as smoking initiation, age of smoking initiation, cigarettes per day, and persistent smoking hold the potential for predicting smoking cessation success. This suggests that risk for smoking cessation failure is highly heterogeneous and can also be assessed using broad polygenic approaches. Such heterogeneity is consistent with multifactorial assessments of cigarette use disorder [33] and with the evidence of heterogeneity from motivational mechanisms linked with persistent smoking [42–49].

Our evidence suggests that a combined PRS, which captures the effects of different smoking behaviors, could be a more useful predictor than a PRS derived from a single trait. Another study has combined PRSs of lung imaging phenotypes and patterns of lung growth to test the power of a combined PRS to predict chronic obstructive pulmonary disease (COPD). This study has shown that the combined PRS was significantly associated with COPD [50]. A combined PRS, or composite PRS, can capture genetic risk from multiple risk behaviors. In addition, a combined PRS may create a biologically relevant score for clinical trials and improve the relevance of genetic risk scores. There is emerging research on methods to combine individual PRSs to improve predictive power. Our study took an agnostic approach in combining individual PRSs based on the observed low correlation among the individual PRSs. Future work to improve this methodology for combining individual PRSs has the potential to revolutionize how PRS are utilized.

Our study is based on two smoking cessation datasets investigating PRSs in clinical trials. Even so, some limitations exist. First, the power of this study was limited due to a modest sample size of 1592 individuals enrolled in trials. To counteract the sample size, we have analyzed two different datasets to reveal convergent results. For future studies, more replications are needed to confirm the associations between smoking behavior PRSs and smoking cessation. Future studies with larger sample sizes would be required for evaluating the interactive effects of polygenic risk scores and specific medications. Second, we are limiting our research to smokers with European ancestry. This is largely because the GWAS and Sequencing Consortium of Alcohol and Nicotine Use (GSCAN) summary statistics is based on 1.2 million smokers of European descent. Several studies prove that the risk scores derived from European population underperform in non-European populations [7]. Third, even though 1000 genome reference panel is proven to provide valuable genomic resources that can augment the power of GWAS in groups with European ancestry [51], we acknowledge the limitation of using 1000 Genomes reference panel instead of more current reference panels (HRC or TopMed) for imputation [52, 53]. Finally, the participants in two studies received different medications for treatment, but we currently do not have enough power to evaluate how PRSs moderate response to specific medications. Furthermore, we observed variation in smoking cessation outcomes across the two trials due to heterogeneity in study design, outcome definitions, and time of the trial. Future studies involving increased number of trials and samples are needed to address this important question.

Ultimately, the genetic studies on smoking behaviors aim to enhance the prevention of smoking and aide in successful smoking cessation. The use of PRSs in clinical trials may be a helpful tool by incorporating multiple genetic signals to assess the inherent ability to quit smoking amongst

smokers. This study is motivated by available GWAS results of smoking behaviors based on large general population studies, while GWAS for refined clinical outcomes such as smoking cessation in smoking cessation trials are not available due to limited sample sizes. Future research can explore different strategies for developing PRSs, including forming them on the basis of phenotypes that critically affect smoking cessation success such as withdrawal, craving, severity, and reward sensitivity. In addition, methods such as machine learning might suggest component weightings to use in future applications. Another future direction is using weighted genetic scores to approximate a biomarker or a mechanistic pathway. In a previous study by Buchwald and colleagues (2020), the authors observed that the *CYP2A6* gene region explained up to ~36% of genetic variance of the nicotine metabolite ratio [50]. Existing research, including our prior work, showed the potential utility of PRS in predicting nicotine metabolism [17, 27]. It is probable that if a selected gene region could explain a substantial proportion of genetic variance of the nicotine metabolite ratio, then a regional PRS would be predictive of smoking cessation at the end of treatment as well. Furthermore, PRSs should permit the study of interactions between non-biological factors and treatments. For example, there is a strong association between *CHRNA5* nicotine receptor gene variant and partner smoking, which is largely an environmental factor, but also represents the genetic factors of partner selection [12]. The genetic risk factor coupled with the environmental risk factor can result in a low rate of smoking reduction; PRSs should permit the exploration of other environmental factors that modulate smoking cessation success.

This study presents evidence on the potential utility of smoking behavior PRSs in predicting smoking cessation success amongst smokers in treatment trials. Our evidence based on the meta-

analysis of two trials suggest that the PRS of later age of smoking initiation and the PRS of persistent smoking may be useful in predicting smoking cessation at the end of treatment. A combined PRS may be a useful predictor of smoking cessation by capturing the genetic propensity for multiple smoking behaviors. These findings help answer important clinical questions about the utility of polygenic risk scores in clinical smoking cessation outcomes.

2.7 Acknowledgements

2.7.1. Contributor Information

Michael Bray, Department of Psychiatry, Washington University School of Medicine, St. Louis, MO, USA. Department of Genetic Counseling, Bay Path University, Longmeadow, MA, USA.

Yoonhoo Chang, Department of Psychiatry, Washington University School of Medicine, St. Louis, MO, USA.

Timothy B Baker, Department of Medicine, School of Medicine and Public Health, Center for Tobacco Research and Intervention, University of Wisconsin, Madison, WI, USA.

Douglas Jorenby, Department of Medicine, School of Medicine and Public Health, Center for Tobacco Research and Intervention, University of Wisconsin, Madison, WI, USA.

Robert M Carney, Department of Psychiatry, Washington University School of Medicine, St. Louis, MO, USA.

Louis Fox, Department of Psychiatry, Washington University School of Medicine, St. Louis, MO, USA.

Giang Pham, Department of Psychiatry, Washington University School of Medicine, St. Louis, MO, USA.

Faith Stoneking, Department of Psychiatry, Washington University School of Medicine, St. Louis, MO, USA.

Nina Smock, Department of Psychiatry, Washington University School of Medicine, St. Louis, MO, USA. The Alvin J. Siteman Cancer Center, Washington University School of Medicine, St. Louis, MO, USA .

Christopher I Amos, Department of Medicine, Baylor College of Medicine, Institute for Clinical and Translational Research, Houston, TX, USA. Department of Biomedical Data Science, Geisel School of Medicine, Dartmouth College, Hanover, NH, USA.

Laura Bierut, Department of Psychiatry, Washington University School of Medicine, St. Louis, MO, USA. The Alvin J. Siteman Cancer Center, Washington University School of Medicine, St. Louis, MO, USA .

Li-Shiun Chen, Department of Psychiatry, Washington University School of Medicine, St. Louis, MO, USA. The Alvin J. Siteman Cancer Center, Washington University School of Medicine, St. Louis, MO, USA .

2.7.2. Funding

M.B. was supported by the National Institutes of Health Training Grant 5T32MH014677. L.-S.C. was supported by the National Institute on Drug Abuse grant R01 DA038076, Siteman Cancer

Center and NCI Cancer Center Support Grant P30 CA091842. T.B.B.'s involvement was supported in part by R01 HL109031. C.I.A. is a research scholar of the Cancer Prevention Research Institute of Texas and partially supported by RR170048. L.-S.C., L.J.B. and C.I.A. are partially supported by U19CA203654. L.J.B. was supported by National Center for Advancing Translation Sciences grant UL1TR002345 and by National Institute on Aging grant R56AG058726. The content is solely the responsibility of the authors and does not necessarily represent the official view of the National Institutes of Health.

2.7.3. Declaration of Interests

R.M.C. or a member of his family owns stock in Pfizer Inc. L.J.B. is listed as an inventor on Issued US Patent 8,080,371 "Markers for Addiction" covering the use of certain single nucleotide polymorphisms in determining the diagnosis, prognosis, and treatment of addiction, and served as a consultant for the pharmaceutical company Pfizer Inc. (New York City, NY) in 2008. All other authors declared no competing interests for this work.

2.7.4. Data Availability

Data is available from dbGaP and the NIDA Center for Genetic Studies:

https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000404.v1.p1

https://nidagenetics.org/studies/study-35-genetically-informative-smoking-cessation-trial.

2.8 References

1. World Health Organization. *Tobacco*; 2020 [cited 2020 8 October]. Available from: https://www.who.int/news-room/fact-sheets/detail/tobacco.

2. CDC. *Tobacco-Related Mortality: Centers for Disease Control and Prevention*; 2018 [updated January 17, 2018. Available from: https://www.cdc.gov/tobacco/data_statistics/fact_sheets/health_effects/tobacco_related_mortality/index.htm.

3. Jha P, Ramasundarahettige C, Landsman V, et al. 21st-Century hazards of smoking and benefits of cessation in the United States. *N Engl J Med.* 2013;368(4):341–350.

4. CDC. *Smoking & Tobacco Use*; 2020. Available from: https://www.cdc.gov/tobacco/data_statistics/fact_sheets/cessation/smokingcessation-fast-facts/index.html.

5. Xian H, Scherrer JF, Madden PA, et al. The heritability of failed smoking cessation and nicotine withdrawal in twins who smoked and attempted to quit. *Nicotine Tob Res.* 2003;5(2):245–254.

6. Hancock DB, Markunas CA, Bierut LJ, Johnson EO. Human genetics of addiction: new insights and future directions. *Curr Psychiatry Rep.* 2018;20(2):8.

7. Liu M, Jiang Y, Wedow R, et al. Association studies of up to 1.2 million individuals yield new insights into the genetic etiology of tobacco and alcohol use. Nat Genet. 2019;51(2):237–244.

8. Improgo MR, Scofield MD, Tapper AR, Gardner PD. The nicotinic acetylcholine receptor CHRNA5/A3/B4 gene cluster: Dual role in nicotine addiction and lung cancer. *Prog Neurobiol.* 2010;92(2):212–226.

9. Bray MJ, Chen LS, Fox L, et al. Dissecting the genetic overlap of smoking behaviors, lung cancer, and chronic obstructive pulmonary disease: a focus on nicotinic receptors and nicotine metabolizing enzyme. *Genet Epidemiol.* 2020;44(7):748–758.

10. Bierut LJ, Stitzel JA, Wang JC, et al. Variants in nicotinic receptors and risk for nicotine dependence. *Am J Psychiatry.* 2008;165(9):1163–1171.

11. Chen LS, Hung RJ, Baker T, et al. CHRNA5 risk variant predicts delayed smoking cessation and earlier lung cancer diagnosis—a meta-analysis. *J Natl Cancer Inst.* 2015;107(5).

12. Chen LS, Baker TB, Piper ME, et al. Interplay of genetic risk factors (CHRNA5-CHRNA3-CHRNB4) and cessation treatments in smoking cessation success. *Am J Psychiatry.* 2012;169(7):735–742.

13. Piper ME, Smith SS, Schlam TR, et al. A randomized placebo-controlled clinical trial of 5 smoking cessation pharmacotherapies. *Arch Gen Psychiatry.* 2009;66(11):1253–1262.

14. Xu C, Goodz S, Sellers EM, Tyndale RF. CYP2A6 genetic variation and potential consequences. *Adv Drug Deliv Rev.* 2002;54(10):1245–1256.

15. Tanner JA, Tyndale RF. Variation in CYP2A6 Activity and Personalized Medicine. *J Pers Med.* 2017;7(4):18.

16. Taylor AE, Fluharty ME, Bjorngaard JH, et al. Investigating the possible causal association of smoking with depression and anxiety using Mendelian randomisation meta-analysis: The CARTA consortium. *BMJ Open.* 2014;4(10):e006141.

17. Chen LS, Bloom AJ, Baker TB, et al. Pharmacotherapy effects on smoking cessation vary with nicotine metabolism gene (CYP2A6). *Addiction.* 2014;109(1):128–137.

18. Furberg H, Kim Y, Dackor J, et al. Genome-wide meta-analyses identify multiple loci associated with smoking behavior. *Nat Genet.* 2010;42(5):441.

19. Caporaso N, Gu F, Chatterjee N, et al. Genome-wide and candidate gene association study of cigarette smoking behaviors. *PLoS One.* 2009;4(2):e4653.

20. Dai J, Lv J, Zhu M, et al. Identification of risk loci and a polygenic risk score for lung cancer: a large-scale prospective cohort study in Chinese populations. *Lancet Respir Med.* 2019;7(10):881–891.

21. Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. *Nature.* 2014;511(7510):421–427.

22. Maas P, Barrdahl M, Joshi AD, et al. Breast cancer risk from modifiable and nonmodifiable risk factors among white women in the United States. *JAMA Oncol.* 2016;2(10):1295–1302.

23. Aly M, Wiklund F, Xu J, et al. Polygenic risk score improves prostate cancer risk prediction: results from the Stockholm-1 cohort study. *Eur Urol.* 2011;60(1):21–28.

24. Belsky DW, Moffitt TE, Houts R, et al. Polygenic risk, rapid childhood growth, and the development of obesity: evidence from a 4-decade longitudinal study. *Arch Pediatr Adolesc Med.* 2012;166(6):515–521.

25. Chen LS, Horton A, Bierut L. Pathways to precision medicine in smoking cessation treatments. *Neurosci Lett.* 2018;669:83–92.

26. Belsky DW, Moffitt TE, Baker TB, et al. Polygenic risk and the developmental progression to heavy, persistent smoking and nicotine dependence: evidence from a 4-decade longitudinal study. *JAMA Psychiatry* 2013;70(5):534–542.

27. Chen LS, Hartz SM, Baker TB, et al. Use of polygenic risk scores of nicotine metabolism in predicting smoking behaviors. *Pharmacogenomics.* 2018;19(18):1383–1394.

28. El-Boraie A, Taghavi T, Chenoweth MJ, et al. Evaluation of a weighted genetic risk score for the prediction of biomarkers of CYP2A6 activity. *Addict Biol.* 2020;25(1):e12741.

29. Chen LS, Baker TB, Miller JP, et al. Genetic Variant in CHRNA5 and response to varenicline and combination nicotine replacement in a randomized placebo-controlled trial. *Clin Pharmacol Ther.* 2020;108(6):1315–1325.

30. Tyndale RF, Zhu AZ, George TP, et al. Lack of associations of CHRNA5-A3-B4 genetic variants with smoking cessation treatment outcomes in caucasian smokers despite associations with baseline smoking. *PLoS One.* 2015;10(5):e0128109.

31. Sarginson JE, Killen JD, Lazzeroni LC, et al. Markers in the 15q24 nicotinic receptor subunit gene cluster (CHRNA5-A3-B4) predict severity of nicotine addiction and response to smoking cessation therapy. *Am J Med Genet B Neuropsychiatr Genet.* 2011;156(3):275–284.

32. Baker TB, Weiss RB, Bolt D, et al. Human neuronal acetylcholine receptor A5-A3-B4 haplotypes are associated with multiple nicotine dependence phenotypes. *Nicotine Tob Res.* 2009;11(7):785–796.

33. Piper ME, Bolt DM, Kim S-Y, et al. Refining the tobacco dependence phenotype using the Wisconsin Inventory of Smoking Dependence Motives. *J Abnorm Psychol.* 2008;117(4):747–761.

34. Cox LS, Wick JA, Nazir N, et al. Predictors of early versus late smoking abstinence within a 24-month disease management program. *Nicotine Tob Res.* 2011;13(3):215–220.

35. Partos TR, Borland R, Yong H-H, Hyland A, Cummings KM. The quitting rollercoaster: how recent quitting history affects future cessation outcomes (data from the International Tobacco Control 4-country cohort study). *Nicotine Tob Res.* 2013;15(9):1578–1587.

36. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet.* 2007;81(3):559–575.

37. Das S, Forer L, Schönherr S, et al. Next-generation genotype imputation service and methods. *Nat Genet.* 2016;48(10):1284–1287.

38. Price AL, Patterson NJ, Plenge RM, et al. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet.* 2006;38(8):904–909.

39. Euesden J, Lewis CM, O'Reilly PF. PRSice: polygenic risk score software. *Bioinformatics.* 2015;31(9):1466–1468.

40. R Core Team. *R: A Language and Environment for Statistical Computing.* Vienna: R Foundation for Statistical Computing; 2013.

41. Chenoweth MJ, O'Loughlin J, Sylvestre MP, Tyndale RF. CYP2A6 slow nicotine metabolism is associated with increased quitting by adolescent smokers. *Pharmacogenet Genomics.* 2013;23(4):232–235.

42. Lerman C, Schnoll RA, Hawk LW, Jr, et al. Use of the nicotine metabolite ratio as a genetically informed biomarker of response to nicotine patch or varenicline for smoking cessation: A randomised, double-blind placebo-controlled trial. *Lancet Respir Med.* 2015;3(2):131–138.

43. Cook JW, Piper ME, Leventhal AM, et al. Anhedonia as a component of the tobacco withdrawal syndrome. *J Abnorm Psychol.* 2015;124(1):215–225.

44. Bechara A, Berridge KC, Bickel WK, et al. A neurobehavioral approach to addiction: implications for the opioid epidemic and the psychology of addiction. *Psychol Sci Public Interest.* 2019;20(2):96–127.

45. Leventhal AM, Zvolensky MJ. Anxiety, depression, and cigarette smoking: a transdiagnostic vulnerability framework to understanding emotion-smoking comorbidity. *Psychol Bull.* 2015;141(1):176–212.

46. Ditre JW, Zale EL, LaRowe LR. A reciprocal model of pain and substance use: transdiagnostic considerations, clinical implications, and future directions. Annu Rev Clin Psychol. 2019;15(1):503–528.

47. Cannon DS, Baker TB, Piper ME, et al. Associations between phenylthiocarbamide gene polymorphisms and cigarette smoking. *Nicotine Tob Res.* 2005;7(6):853–858.

48. Baker TB, Piper ME, Schlam TR, et al. Are tobacco dependence and withdrawal related amongst heavy smokers? Relevance to conceptualizations of dependence. *J Abnorm Psychol.* 2012;121(4):909–921.

49. Moll M, Sakornsakolpat P, Shrine N, et al. Chronic obstructive pulmonary disease and related phenotypes: polygenic risk scores in population-based and case-control cohorts. *Lancet Respir Med.* 2020;8(7):696–708.

50. Buchwald J, Chenoweth MJ, Palviainen T, et al. Genome-wide association meta-analysis of nicotine metabolism and cigarette consumption measures in smokers of European descent. *Mol Psychiatry.* 2021;26(6):2212–2223.

51. Belsare S, Levy-Sakin M, Mostovoy Y, et al. Evaluating the quality of the 1000 genomes project data. *BMC Genomics.* 2019;20(1):620

52. Iglesias AI, Van Der Lee SJ, Bonnemaijer PW, Höhn R, Nag A, Gharahkhani P, et al. (2017): Haplotype reference consortium panel: Practical implications of imputations with large reference panels. *Human mutation*. 38:1025-1032.

53. Kowalski MH, Qian H, Hou Z, Rosen JD, Tapia AL, Shan Y, et al. (2019): Use of> 100,000 NHLBI Trans-Omics for Precision Medicine (TOPMed) Consortium whole genome sequences improves imputation quality and detection of rare variant associations in admixed African and Hispanic/Latino populations. *PLoS genetics*. 15:e1008500.

2.9 Tables

Table 2.1. Descriptive Statistics for the samples from GISC and TTURC studies used for the analysis

	GISC	TTURC
Total N	501	1091
Age (Mean, SE)	46.6, 0.51	44.5, 0.34
Sex (n, %)		
Male	221, 44.1	457, 41.9
Female	280, 55.9	634, 58.1
Baseline CPD (Mean, SE)	19.1, 0.34	21.8, 0.28
Randomized to active pharmacotherapy (n, %)	329, 65.7	955, 87.5
History of Anxiety/Depression (n, %)	129, 25.7	226, 20.7
CO (Mean, SE)	28.9, 0.59	26.5, 0.38
FTND (Mean, SE)	4.9, 0.1	5.3, 0.07
Smoking Age of initiation (Mean, SE)	17.5, 0.19	17.3, 0.12
Smoking Abstinence at EOT (n, %)	103, 20.6	518, 47.5

SE, standard error. CPD is defined as cigarettes per day. EOT is defined as end of treatment. CO is carbon monoxide level, and FTND is Fagerstrom Test for Nicotine Dependence. TTURC2, The Transdisciplinary Tobacco Use Research Centers Study. GISC, The Genetically Informed Smoking Cessation Trial Study. Please note that TTURC is missing 2 CO, 14 FTND, 2 Smoking Age of initiation, 36 CPD values.

Table 2.2. Correlation of PRS of smoking behaviors in the study sample

(A) GISC

	PRS of ES	PRS of DAI	PRS of CPD	PRS of PS
PRS of ES				
PRS of DAI	-0.094*			
PRS of CPD	0.11*	-0.093*		
PRS of PS	0.13**	0.011	0.16***	

(B) TTURC

	PRS of ES	PRS of DAI	PRS of CPD	PRS of PS
PRS of ES				
PRS of DAI	-0.1***			
PRS of CPD	0.038	-0.067*		
PRS of PS	0.058*	-0.014	0.076*	

*: P <0.05

**: P <0.005

***: P <0.0005

ES: ever smoking

DAI: delayed age of smoking initiation

CPD: cigarettes per day

PS: persistent smoking

2.10 Figures

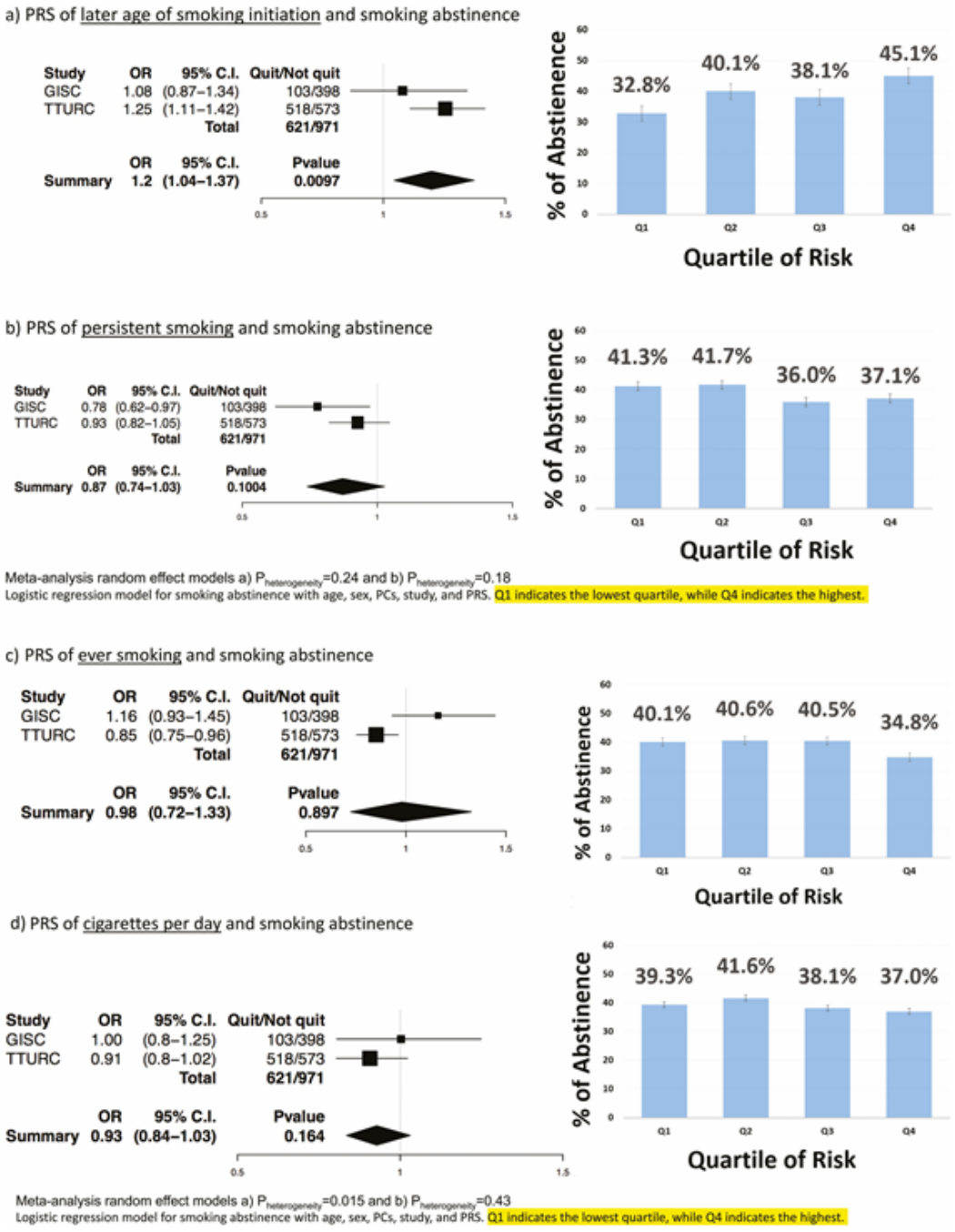
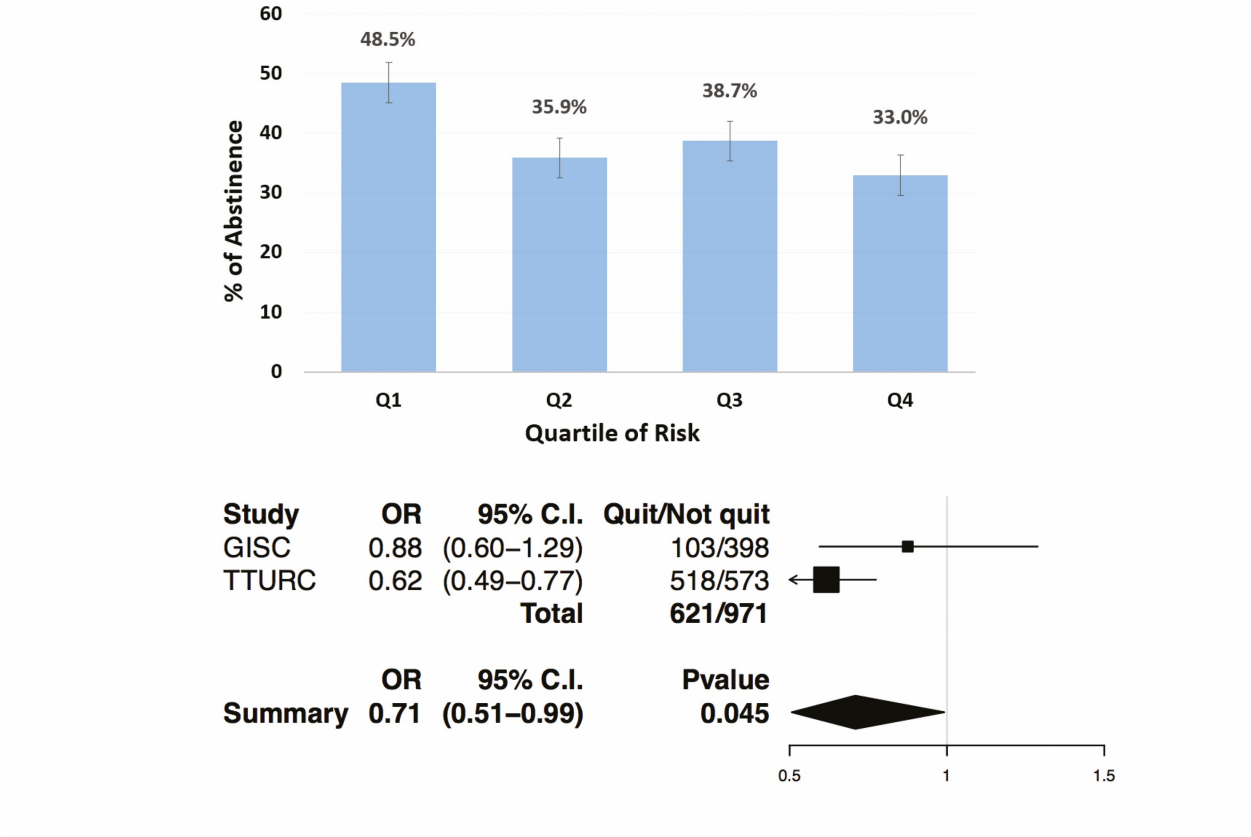


Figure 2.1. Polygenetic risk scores (PRS) of later age of smoking initiation, persistent smoking, ever smoking and cigarettes per day and bioverified end of treatment smoking abstinence: Meta-Analysis of two treatment trials



Meta-analysis random effect models. $P_{heterogeneity}$=0.013.
Logistic regression model for smoking abstinence with age, sex, PCs, study, and PRS. Q1 indicates the lowest quartile, while Q4 indicates the highest.

Figure 2.2. Combined polygenetic risk score (PRS) and bioverified end of treatment smoking abstinence: Meta-analysis of two treatment trials

Model	N=1543	R^2	AUC	Compare with the prior model	
				X2 (df)	P
1 Base model	Age, sex, and study	0.0724	0.643		
2 Add basic clinical	Age, sex, study, treatment, and CPD	0.107	0.6917	19.6 (1)	<0.0001
3 Add additional clinical	Age, sex, study, treatment, CPD, FTND, CO, history of anxiety/depression, and age of initiation	0.117	0.703	5.48 (1)	0.019
4 Add genetics	age, sex, study, CPD, treatment, FTND, CO, anxiety/depression, age of initiation, PCs, PRS of ever smoking, PRS of delayed initiation, PRS of CPD, and PRS of persistent smoking,	0.126	0.712	4.50 (1)	0.034

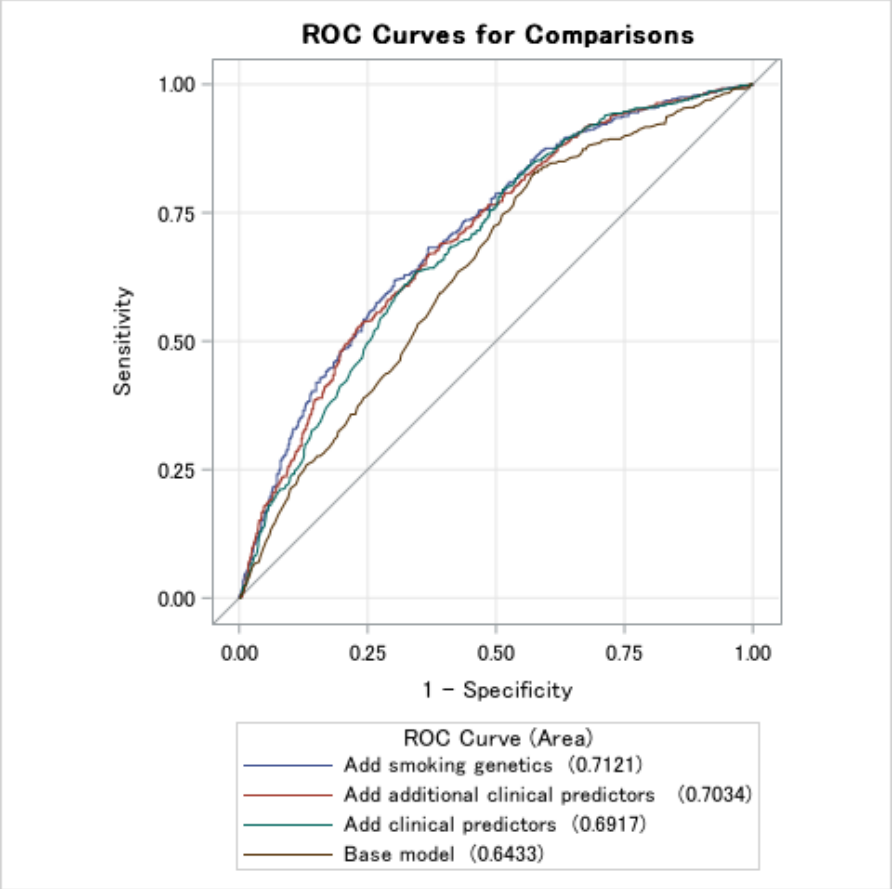


Figure 2.3. Pooled analyses: Prediction of Smoking Abstinence. CPD: cigarettes per day. FTND: baseline Fagerstrom Test for Nicotine Dependence. CO: baseline carbon monoxide PCs: Principle components. ROC curve: Receiver Operating Characteristic Curve. PRS: polygenic risk scores. AUC: area under the curve. R2 is a statistical measure that represents the proportion of the variance for included covariates in the model.

Chapter 3. Use of polygenic risk score beyond clinical factors enhances prediction of smoking cessation

3.1. Abstract

Introduction: Genetic factors play an important role in smoking behavior, with heritability estimates influencing cessation outcomes. The Genome-Wide Association Studies & Sequencing Consortium of Alcohol and Nicotine Use (GSCAN) provides valuable insights, including summary statistics for constructing Polygenic Risk Scores (PRS) for smoking cessation. This study uses the UK Biobank dataset to validate PRS in a large population, addressing the gap in population-based studies.

Methods: Analyzing data from 200,000 UK Biobank participants with a history of daily smoking, we utilized GSCAN summary statistics to create PRS for persistent smoking for each participant. We used logistic regression to examine the association between PRS and smoking cessation, adjusting for relevant covariates such as age, sex, education, smoking in household, ever depressed, and ancestral principal components. We evaluated prediction accuracy through comparing the area under the ROC curve (AUC). Survival analyses assessed quit probability and median quit age across PRS risk groups.

Results: PRS for smoking cessation exhibited a significant association with quit outcomes, with higher PRS linked to reduced success. Integration of genetic predictors significantly improved cessation prediction beyond clinical factors. Individuals in higher PRS risk groups showed lower cessation success rates, with a 3-year difference in median age of cessation between top and bottom 10% PRS groups.

Discussion: This study validates the use of PRS for smoking cessation, showcasing its potential beyond conventional predictors. The findings underscore the clinical utility of genetic information in predicting long-term smoking cessation outcomes. Despite limitations, the large-scale UK Biobank dataset provides robust evidence supporting the integration of PRS into precision medicine approaches for smoking cessation.

Conclusions: The study contributes to the translational effort in precision medicine, demonstrating the predictive capacity of PRS for smoking cessation in a healthier population. The results emphasize the potential of genetic information to enhance treatment decisions, motivating individuals to seek evidence-based interventions. This research bridges the gap between genetic discoveries and clinical application, paving the way for more effective and precise healthcare solutions in smoking cessation.

3.2. Introduction

Cigarette smoking is a significant public health problem that results in more than 8 million preventable deaths worldwide [1]. It has been linked to various adverse health outcomes, including several forms of cancer, respiratory and cardiovascular diseases, and cognitive impairment [2-4]. Despite widespread awareness of the detrimental effects of smoking, quitting remains a challenging process for many people. Only 6-10% of adults who smoke succeed in quitting smoking each year, which calls for a need for more effective smoking cessation strategies [5]. Personalized approaches to smoking cessation may improve successful quit rates.

Genetics plays a crucial role in smoking behaviors, with the heritability estimates from twin studies explaining up to 54% of variance in smoking cessation outcomes [6]. Large-scale genome-wide association studies (GWAS) were performed in recent years to unravel the genetic architecture of different smoking behaviors including smoking initiation, age of smoking initiation, number of cigarettes smoked per day, and successful smoking cessation [7-9]. GWAS & Sequencing Consortium of Alcohol and Nicotine Use (GSCAN), is the meta-analysis of such GWAS findings across studies, with accumulated samples of over 3.4 million individuals [10]. GSCAN provided comprehensive summary statistics that allowed the construction of polygenic risk scores (PRSs). PRS, the risk score made by combining the effects of many genetic signals into a single risk variable, can be used to estimate an individual's genetic propensity to develop a disease or a trait [11]. PRSs have been utilized in predicting health outcomes such as obesity, cancers, and psychiatric disorders [12-14].

However, despite the availability of extensive summary statistics, there remains a lack of studies validating the use of PRS in large population-based samples. The UK Biobank, an unprecedented resource of genetic information and detailed questionnaires from approximately 500,000 individuals, allows the investigation of genetics and smoking behavior in a vast population-based cohort [15]. The UK Biobank data can play an important role in validating the predictive capacity of PRS for smoking cessation outcomes. Moreover, this dataset offers a unique advantage for studying the distribution of PRS within the general population and addressing general population-specific questions that clinical trials' limited sample sizes could not answer, such as the predictive capability, pattern, and significance of PRS in a larger and healthier population.

The ultimate goal of precision medicine is to guide and support individuals in making lifestyle changes and treatment decisions based on their individual genetic and environmental risk profiles. Prediction of smoking cessation outcomes traditionally relied on clinical predictors, including age, sex, environmental factors, and socio-economic status [16-20]. However, PRS can potentially enhance the accuracy of prognosis prediction beyond these conventional predictors. By incorporating genetic information, PRS may improve the identification of individuals with later age of successful smoking cessation, enabling targeted interventions and support. In this study, we investigate the median age of smoking cessation as a clinically interpretable measure for smoking cessation prognosis. This study aims to contribute to this translational effort in precision medicine by validating the use of PRS to identify those at the highest risk for later smoking cessation.

3.3. Methods

3.3.1. Study Samples

UK Biobank

UK Biobank is a long-term prospective epidemiological study of 500,000 participants [21]. The study includes information on smoking behaviors and genome-wide genotype data of the participants. UK Biobank also provides extensive demographic data including age, sex, smoking status, age started smoking, age stopped smoking, cigarettes smoked per day, pack-years of smoking, smoking in the household, socioeconomic status, and mental health status.

Smoking Behaviors

Smoking phenotypes were defined using data from self-report surveys obtained during in-person assessment center visits at baseline ('instance 0', 2006-2010). Current daily smoking was defined by the answer 'Yes, on most or all days' to the question 'Do you smoke tobacco now?' (Data-field 1239). Former daily smoking was defined by the answer 'Smoked on most or all days' to the question 'In the past, how often have you smoked tobacco?' (Data-field 1249). Those with a history of occasional smoking, but not smoking daily, and those with no history of daily smoking were excluded from the analysis.

GWAS & Sequencing Consortium of Alcohol and Nicotine use

GSCAN is a meta-analysis of over 30 GWAS in over 3.4 million participants with European ancestry on substance use [10]. The study discovered genetic variants associated with a key smoking behavior: smoking cessation. The included GWAS are imputed using either 1000 Genome or Haplotype Reference Consortium (HRC) or a combination of two reference panels

with more specific panels. The released dataset presents summary statistics of smoking cessation with and without UK Biobank samples.

Ethical Review Board

The UK Biobank study was approved by the National Health Service National Research Ethics Service (11/NW/0382). All the participants provided informed consent to participate in the UK Biobank study (Our study ID: 48123).

3.3.2. Genetic Data

Polygenic Risk Scores

We retrieved genome-wide data for all participants of European ancestry from the UK Biobank genetic dataset (dataset version/number = ukb48123). We used GSCAN summary statistics generated with the UK Biobank samples removed to create a polygenic risk score (PRS) for persistent smoking [failed smoking cessation]) with genetic variants using PRSice-2 [22, 23]. The PRS results have been pruned for sites with minor allele frequency (MAF) > 0.001, imputation quality (Effective N/N) > 0.3, and an effective sample size of at least 10% of the maximum sample size. Insertions and deletions were not included in GSCAN2 summary statistics or in the calculation of PRS. PRSice-2 uses a p-value selection threshold approach. Thus, according to the different p-value thresholds (.5, .05, .005, .0005, 5×10^{-5}, 5×10^{-6}, 5×10^{-7}, 5×10^{-8}), we included SNPs with a GWAS association p-value below each threshold. To ensure the interpretability across PRSs, all PRSs were standardized to Z-scores.

3.3.3. Statistical analysis

We analyzed how each PRS was linked to predicting smoking cessation. The z-score transformed PRSs were treated as continuous factors in our analysis. We utilized logistic regression models in the R to explore the connections between PRSs and smoking cessation.

Covariates that might confound the association between smoking cessation data and genetic predisposition to smoking behaviors were included in the following analyses: age, sex, ancestral principal components (PCs), education, smoking in the household, and ever-depressed for a whole week. We generated a receiver operating characteristic curve (ROC) and estimated the area under the curve (AUC) for the prediction of smoking cessation using clinical predictors and genetic predictors (SAS 9.4). In addition, we created a survival analysis curve between quit probability and time to quit for 7 PRS risk groups (Bottom 10%, 10-20%, 20-40%, 40-60%, 60-80%, 80-90%, Top 10%) and computed median age for each group. The survival analysis and median age calculation were done for the PRS of smoking cessation (p-value threshold 0.5), and the combined PRS.

3.4. Results

3.4.1. Sample characteristics

We examined the prediction ability and distribution of polygenic risk scores of smoking cessation in UK Biobank samples. The sample characteristics are shown in Table 1. All samples are of European descent. The mean age was 57.7, and the percentage of men and women was similar (Male: 51.9, Female: 48.1). There were 48,275 participants who is currently daily smoking, and 160,643 who formerly daily smoked. The mean of age started smoking was 17.4, and the mean of age stopped smoking was 44.5. On average, ever-smoked participants have smoked 17.8 cigarettes per day and had a 23.8 pack-year history. 31,222 (14.9%) participants had one or more household members who smoked. For the related covariates, 54,880 (28.2%) reported they have a college degree; 180,975 (86.6%) participants answered "yes" to the question asking if they were ever depressed for a whole week.

3.4.2. PRS of smoking cessation and persistent smoking cessation

We evaluated the association between PRS for smoking cessation and the participant's smoking cessation in 8 different P-value thresholds. For the most stringent threshold (P < 5 x 10-8), OR was 0.95, and the P-value was 2.93 x 10-22. For the least stringent threshold (P < 0.5), OR was 0.90 and P-value was 1.03×10^{-85}. For the other 6 thresholds in between 0.5 and 5×10^{-8}, OR steadily increased from 0.95 and the P-value steadily decreased from 2.93×10^{-22} as the p-value threshold became less stringent (Figure 1).

3.4.3. Prediction of smoking cessation using clinical predictors with genetic predictors

We also evaluated the effect of the additional genetic predictors to 1) base model with age and sex, 2) model with clinical predictors added (age, sex, education, pack years, age of initiation of

smoking, smoking in household). We found that the model with genetic predictors added (age, sex, education, pack years, age of initiation, smoking in household, PC1-3, PRS for smoking cessation at 0.5 p-value threshold, and risk allele count for the SNP rs16969968) significantly increased the area under the curve (AUC) (clinical predictors—0.6229 to 0.7164, p = p < .0001; genetic predictors—0.7164 to 0.7178, p = p < .0001) as shown in Figure 2.

3.4.4. Varying probability of smoking cessation success among individuals across smoking cessation PRS risk group

We found that individuals in the higher PRS risk groups were less likely to successfully quit compared to those in the bottom 10% PRS group). For example, individuals in the top 10% PRS are less like to quit successfully compared with those in the bottom 10% (OR=0.73, 95% CI=0.69-0.77, p<0.0001). Detailed results from the regression model are included in Table S2. The median age of smoking cessation for each 7 groups PRS for smoking cessation (P < 0.5) was divided into started from 45 (Bottom 10%) and increased to 47 (Top 10%) as shown in Figure 3.

3.5. Discussion

We validate the PRS of smoking cessation and present evidence to show the potential use of PRS for smoking cessation. PRS of smoking cessation is significantly associated with an individual's

smoking cessation outcome. Second, PRS of smoking cessation can enhance the prediction of smoking cessation beyond the use of demographic and clinical predictors. Third, PRS may place individuals in different risk categories that are associated with a variable genetic underpinning for quit success that can vary by 3 years.

Evidence is growing on the optimal use of PRS in clinical applications. Most of the current translational work focuses on how PRS can help enhance diagnoses [12-14, 24, 25]. Here we present an example of using PRS to identify the long-term prognosis of individuals who smoke and their potential quit outcomes based on their genetic markers. Knowledge of precision risk of smoking cessation difficulty based on personal biology may motivate individuals who smoke to seek treatment, given that most individuals who smoke do not use evidence-based treatment due to common barriers such as the belief that their personalized risk for smoking is lower than average or their personal ability to quit is higher than average.

Our study is based on the UK Biobank dataset, which is large and provides ample statistical power. Even so, some limitations exist. First, we limited our research dataset to those with European ancestry. Most of the participants in both the UK Biobank and the GWAS and Sequencing Consortium of Alcohol and Nicotine Use (GSCAN) summary statistics are of European descent. Many studies have shown that the polygenic risk scores created based on the European population dataset underperform in non-European populations. Second, the UK Biobank participants may not be representative of the general population; there is evidence of a "healthy volunteer" selection bias [26]. It is commonly known that the participants in health-related studies are more health-conscious than those who do not [26, 27]. Second, we limited our research dataset to those with European ancestry. The majority of the participants in both the UK Biobank and the GWAS and Sequencing Consortium of Alcohol and Nicotine Use (GSCAN)

summary statistics are of European descent. Many studies have shown that the polygenic risk scores created based on the European population dataset underperform in non-European populations. Third, there are several ways to generate polygenic risk scores, and we used PRSice2, the P-value-based clumping and thresholding ("P+T") method [28]. Even though PRSice is based on the most commonly used method, the polygenic risk scores can be generated using different methods. Many alternative approaches assume that SNP effects are derived from combinations of different distributions, with the key parameters defining these architectures determined using Bayesian frameworks. (i.e. LDpred2, PRS-CS, SBayesR) [28, 29].

The translation of genetic discoveries to clinical interventions requires multiple steps starting from validating results, identifying clinical utility, and establishing equitable pragmatic implementation strategies in clinical settings [30]. These findings are important to start identifying the potential utility of genetic associations for individuals who smoke. While the genetic vulnerabilities for many health traits are included in consumer genomics reports (e.g., motion sickness), the knowledge of their own genetic predispositions for quit success may be highly motivational for people to make health decisions such as finding a quit date or seeking tobacco treatment.

3.6. Conclusion

To facilitate clinical translation, this study evaluated the predictive ability of PRS for smoking cessation in a large dataset of participants and demonstrated: a) its significant association with individual quit outcomes, b) its ability to improve cessation prediction beyond demographic and clinical factors, and c) its potential to categorize individuals into varying risk groups with up to a 3-year difference in genetic-based quit success. The research cycle, encompassing discovery, validation, and clinical application, stands as the pivotal step in the field of precision medicine. From identifying disease-associated genetic variants to developing robust risk assessment methods, to rigorously validating their clinical utility in clinical trials, offers tangible benefits to individuals in need of targeted interventions. Our study provides compelling evidence for the robust clinical utility of the PRS for smoking cessation. Ultimately, the purpose of our research is to provide more effective and precise healthcare solutions, and these findings help bridge the gap between the research and the application.

3.7. References

1. World Health Organization. *Tobacco*; 2020 [cited 2020 8 October]. Available from: https://www.who.int/news-room/fact-sheets/detail/tobacco.

2. US Department of Health Human Services (2010): How tobacco smoke causes disease: What it means to you. Atlanta: US Department of Health and Human Services, Centers for Disease Control and Prevention. *National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health*. 1993:18.

3. Lushniak BD, Samet JM, Pechacek TF, Norman LA, Taylor PA (2014): The health consequences of smoking—50 years of progress: a report of the Surgeon General.

4. McMaster C, Lee C (1991): Cognitive dissonance in tobacco smokers. *Addictive behaviors*. 16:349-353.

5. CDC. *Smoking & Tobacco Use*; 2020. Available from: https://www.cdc.gov/tobacco/data_statistics/fact_sheets/cessation/smokingcessation-fast-facts/index.html.

6. Xian H, Scherrer JF, Madden PA, Lyons MJ, Tsuang M, True WR, et al. (2003): The heritability of failed smoking cessation and nicotine withdrawal in twins who smoked and attempted to quit. *Nicotine & Tobacco Research*. 5:245-254.

7. Xu K, Li B, McGinnis KA, Vickers-Smith R, Dao C, Sun N, et al. (2020): Genome-wide association study of smoking trajectory and meta-analysis of smoking status in 842,000 individuals. *Nature communications*. 11:5302.

8. Saunders GR, Wang X, Chen F, Jang S-K, Liu M, Wang C, et al. (2022): Genetic diversity fuels gene discovery for tobacco and alcohol use. *Nature*. 612:720-724.

9. Siedlinski M, Cho MH, Bakke P, Gulsvik A, Lomas DA, Anderson W, et al. (2011): Genome-wide association study of smoking behaviours in patients with COPD. *Thorax*. 66:894-902.

10. Saunders GR, Wang X, Chen F, Jang S-K, Liu M, Wang C, et al. (2022): Genetic diversity fuels gene discovery for tobacco and alcohol use. *Nature*. 612:720-724.

11. Choi SW, Mak TS-H, O'Reilly PF (2020): Tutorial: a guide to performing polygenic risk score analyses. *Nature protocols*. 15:2759-2772.

12. Pantelis C, Papadimitriou GN, Papiol S, Parkhomenko E, Pato MT, Paunio T, et al. (2014): Biological insights from 108 schizophrenia-associated genetic loci. *Nature*. 511:421-427.

13. Maas P, Barrdahl M, Joshi AD, Auer PL, Gaudet MM, Milne RL, et al. (2016): Breast cancer risk from modifiable and nonmodifiable risk factors among white women in the United States. *JAMA oncology*. 2:1295-1302.

14. Belsky DW, Moffitt TE, Houts R, Bennett GG, Biddle AK, Blumenthal JA, et al. (2012): Polygenic risk, rapid childhood growth, and the development of obesity: evidence from a 4-decade longitudinal study. *Archives of pediatrics & adolescent medicine*. 166:515-521.

15. Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, et al. (2015): UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS medicine*. 12:e1001779.

16. Joffer J, Burell G, Bergström E, Stenlund H, Sjörs L, Jerdén L (2014): Predictors of smoking among Swedish adolescents. *BMC public health*. 14:1-9.

17. Caponnetto P, Polosa R (2008): Common predictors of smoking cessation in clinical practice. *Respiratory medicine*. 102:1182-1192.

18. Rodgers-Melnick SN, Zanotti K, Lee RT, Webb Hooper M (2022): Demographic and Clinical Predictors of Engaging in Tobacco Cessation Counseling at a Comprehensive Cancer Center. *JCO Oncology Practice*. 18:e721-e730.

19. Speranskaya OI (2011): P01-108 - Clinical predictors of tobacco dependence relapses. *European Psychiatry*. 26:108.

20. Gram IT, Antypas K, Wangberg SC, Løchen ML, Larbi D (2022): Factors associated with predictors of smoking cessation from a Norwegian internet-based smoking cessation intervention study. *Tob Prev Cessat*. 8:38.

21. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, et al. (2018): The UK Biobank resource with deep phenotyping and genomic data. *Nature*. 562:203-209.

22. Choi SW, O'Reilly PF (2019): PRSice-2: Polygenic Risk Score software for biobank-scale data. *Gigascience*. 8:giz082.

23. [dataset] Liu M, Jiang Y, Wedow R, Li Y, Brazel DM, Chen F, et al. (2019): Data Related to Association studies of up to 1.2 million individuals yield new insights into the genetic etiology of tobacco and alcohol use. Retrieved from the Data Repository for the University of Minnesota, https://doi.org/10.13020/3b1n-ff32.

24. Tremblay J, Haloui M, Attaoua R, Tahir R, Hishmih C, Harvey F, et al. (2021): Polygenic risk scores predict diabetes complications and their response to intensive blood pressure and glucose control. *Diabetologia*. 64:2012-2025.

25. Sun L, Pennells L, Kaptoge S, Nelson CP, Ritchie SC, Abraham G, et al. (2021): Polygenic risk scores in cardiovascular risk prediction: A cohort study and modelling analyses. *PLoS medicine*. 18:e1003498.

26. Fry A, Littlejohns TJ, Sudlow C, Doherty N, Adamska L, Sprosen T, et al. (2017): Comparison of sociodemographic and health-related characteristics of UK Biobank participants with those of the general population. *American journal of epidemiology*. 186:1026-1034.

27. Schoeler T, Speed D, Porcu E, Pirastu N, Pingault JB, Kutalik Z (2023): Participation bias in the UK Biobank distorts genetic associations and downstream analyses. *Nat Hum Behav*. 7:1216-1227.

28. Chatterjee N, Wheeler B, Sampson J, Hartge P, Chanock SJ, Park JH (2013): Projecting the performance of risk prediction based on polygenic analyses of genome-wide association studies. *Nat Genet*. 45:400-405, 405e401-403.

29. Ni G, Zeng J, Revez JA, Wang Y, Zheng Z, Ge T, et al. (2021): A Comparison of Ten Polygenic Score Methods for Psychiatric Disorders Applied Across Multiple Cohorts. *Biol Psychiatry*. 90:611-620.

30. Bray M, Chang Y, Baker TB, Jorenby D, Carney RM, Fox L, et al. (2022): The Promise of Polygenic Risk Prediction in Smoking Cessation: Evidence From Two Treatment Trials. *Nicotine Tob Res*. 24:1573-1580.

3.8. Tables

Table 3.1. Demographic and smoking variables

Total N= 194,859	**Mean ± SD**
Age*	57.9 ± 7.75
Sex (n females, %)	
Female (n, %)	94,123 (48.3)
Male (n, %)	100,736 (51.7)
College	
College of university degree (n, %)	54,880 (28.2)
Others (n, %)	139,979 (71.8)
Ever depressed for a whole week	
Yes (n, %)	169,237 (86.9)
No (n, %)	25,622 (13.1)
Smoking status	
Current daily smoking (n, %)	43,618 (22.4)
Former daily smoked (n, %)	151,241 (77.6)
Age started smoking	17.4 ± 4.18
Age stopped smoking	44.4 ± 12.8
Cigarettes per day	17.9 ± 11.3
Pack years	23.9 ± 20.5
Smoking/smokers in the household	
Yes, one or more household members (n, %)	28,804 (14.8)
No (n, %)	166,055 (85.2)

*Age at recruitment

Do not know, and prefer not to answer removed from the analysis

3.9. Figures

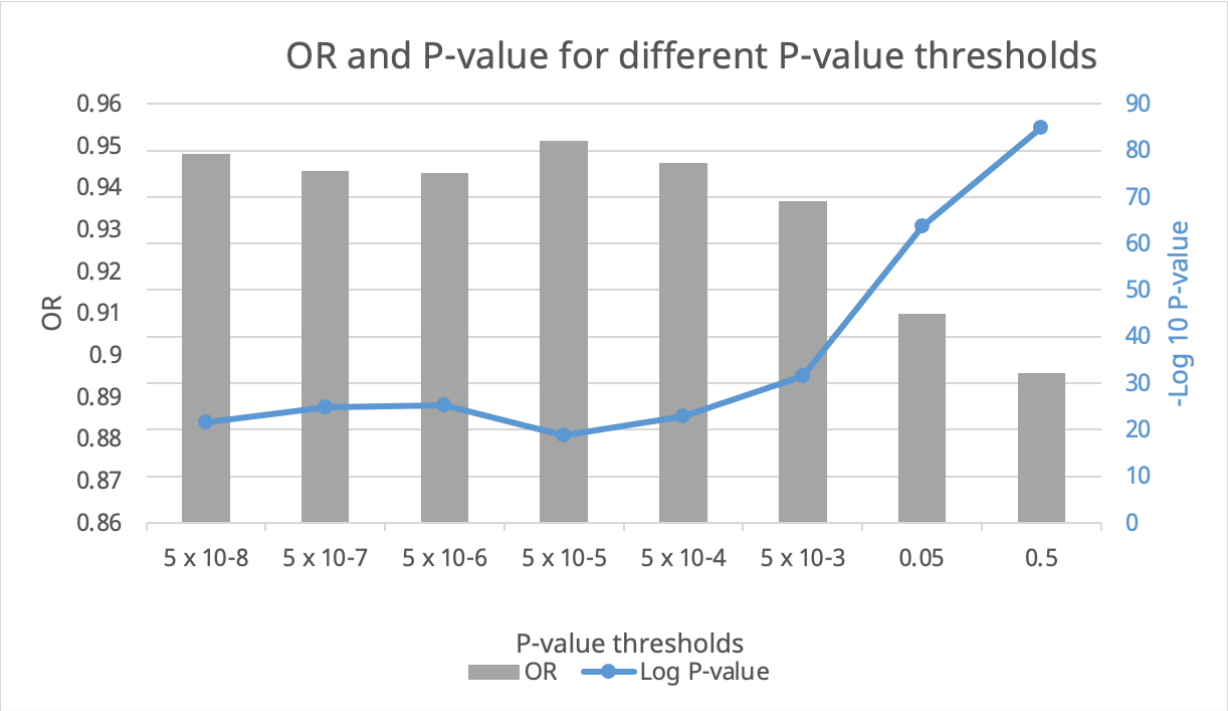
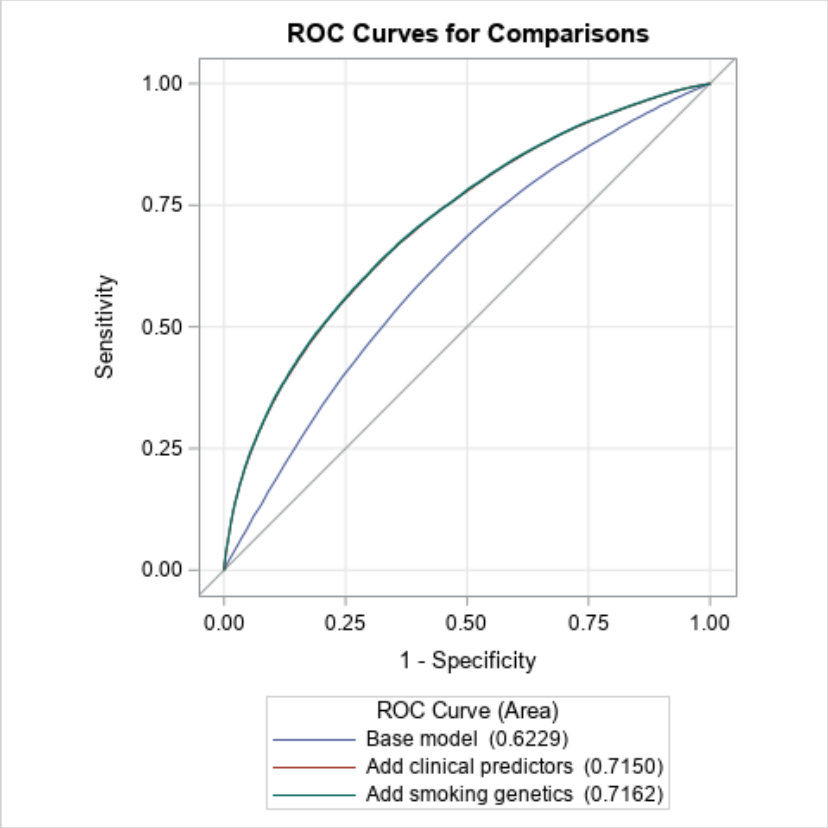


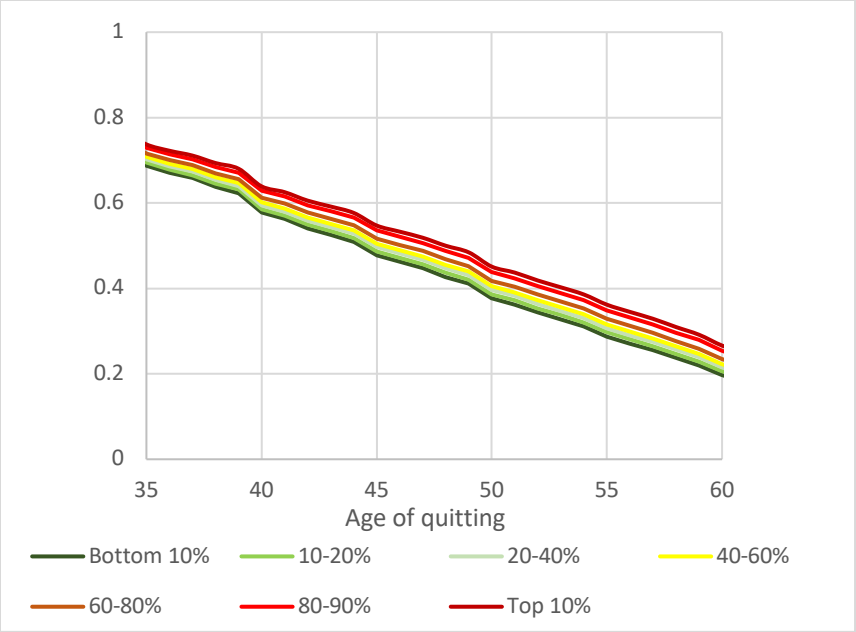
Figure 3.1. PRS of different p-value thresholds and corresponding OR and P-value



Model*	Covariates	R^2	AUC	Comparison with the prior model	
				X2	P-value
1 Base model	Age, sex	0.0322	0.6229		
2 Add clinical predictors	Age, sex, education, pack-years, age of initiation, smoker in household	0.0962	0.7150	4146.0210	<.0001
3 Add genetics	Age, sex, education, pack-years, age of initiation, smoker in household, PCs, PRS for smoking cessation (p: 0.5), rs16969968	0.0974	0.7164	50.4377	<.0001

*N=194859

Figure 3.2. ROC curve of clinical predictors with added genetic predictors



PRS for smoking cessation	Median time to quit	Quit probability	SE	CI 5%	CI 95%
Bottom 10%	44	0.509595	0.002974	0.5038	0.515456
10-20%	45	0.495998	0.002883	0.490379	0.501681
20-40%	45	0.49892	0.002103	0.494815	0.503059
40-60%	46	0.493077	0.002121	0.488937	0.497251
60-80%	46	0.501348	0.002117	0.497215	0.505515
80-90%	47	0.499644	0.002942	0.493911	0.505444
Top 10%	47	0.507241	0.002808	0.501767	0.512775

Figure 3.3. Survival analysis and median age for 7 different PRS risk groups. PRS for smoking cessation is at the p-value threshold 0.5.

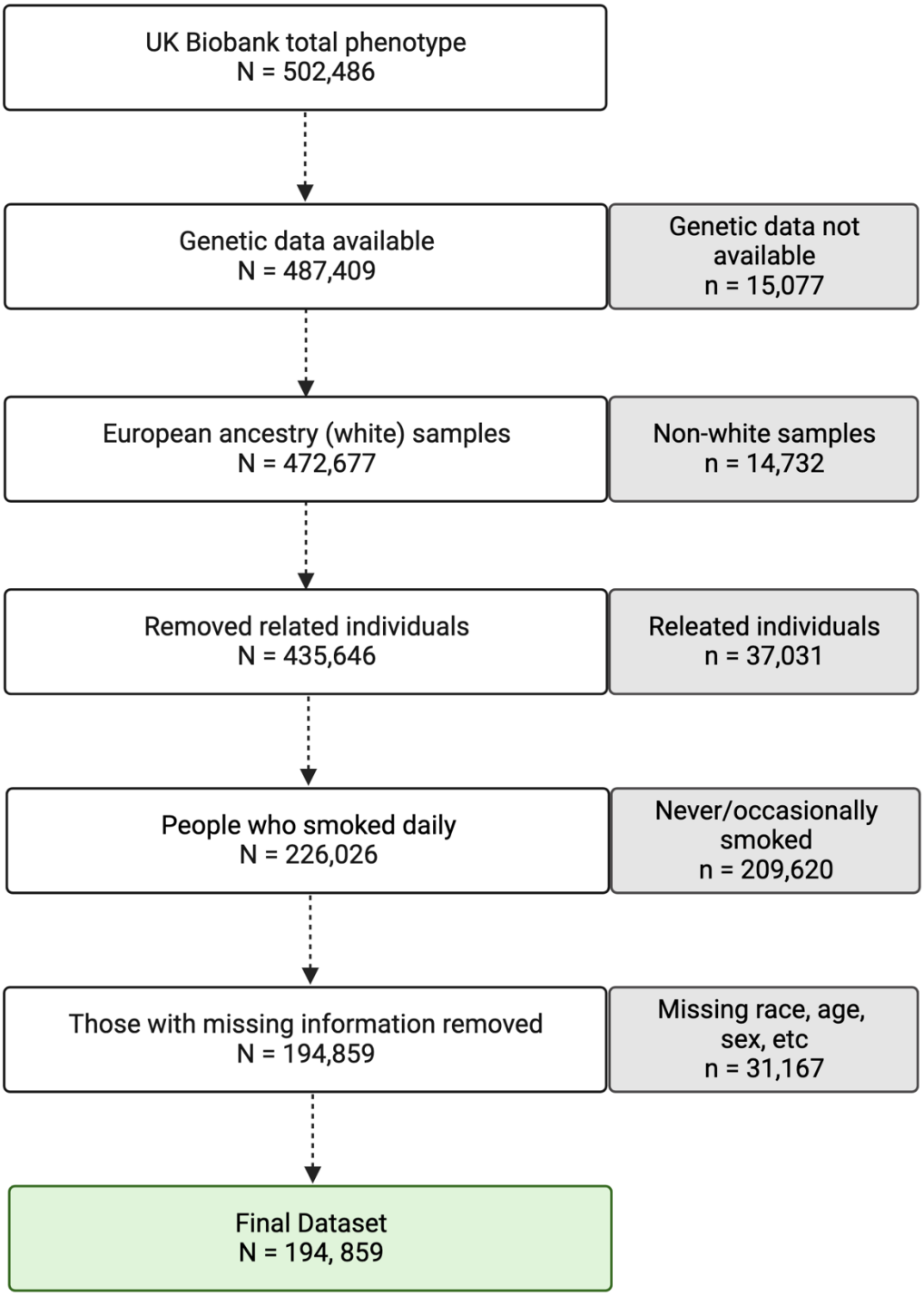


Figure 3.4. Sample processing chart

Chapter 4. Investigating the relationship between smoking behavior and global brain volume

4.1. Abstract

Introduction: Previous studies have shown that brain volume is negatively associated with cigarette smoking, but there is an ongoing debate whether smoking causes lowered brain volume or a lower brain volume is a risk factor for smoking. We address this debate through multiple methods that evaluate directionality: Bradford Hill's Criteria that is commonly used to understand a causal relationship in epidemiological studies, and mediation analysis.

Methods: In 32,094 participants of European descent from the UK Biobank dataset, we examined the relationship between a history of daily smoking and brain volumes, as well as association of genetic risk score to ever smoking with brain volume.

Results: A history of daily smoking is strongly associated with decreased brain volume, and a history of heavier smoking is associated with a greater decrease in brain volume. The strongest association was between total grey matter volume and a history of daily smoking (Effect size = -2964mm^3 p-value = 2.04 x 10^{-16}), and there was a dose response relationship with more pack years smoked associated with a greater decrease in brain volume. A polygenic risk score (PRS) for smoking initiation was strongly associated with a history of daily smoking (Effect size = 0.05, p-value = 4.20 $\times 10^{-84}$), yet only modestly associated with total grey matter volume (Effect size = - 424mm^3, p-value = 0.01). Mediation analysis indicated that a history of daily smoking is a mediator between smoking initiation PRS and total grey matter volume.

Conclusions: A history of daily smoking is strongly associated with a decreased total brain volume.

4.2. Introduction

Cigarette smoking is associated with numerous harmful health outcomes, including cardiovascular disease, respiratory disease, cancer, and diminished overall health [1-4]. The adverse effect of smoking extends into the brain, and this is shown by the association between smoking and dementia [5-7]. People who smoke are more likely to have deterioration in grey and white matter, which provides a possible explanation as to why 14% of global Alzheimer's disease cases could be attributable to cigarette smoking [8, 9].

Smoking-related behaviors are in part biologically driven. Twin studies firmly established the importance of genetic factors contributing to the onset of cigarette smoking, and smoking initiation has heritability estimates of 44% [10-12]. Recent large genome-wide association studies have identified thousands of genetic loci associated with smoking-related behaviors [13-15]. Differences in responses to nicotinic receptors, nicotine metabolism, and many other genetic factors contribute to the development of smoking behaviors. Models of addiction posit that predisposing neurodevelopmental risk factors promote the onset of cigarette smoking and other addictive behaviors [16, 17].

It is known that there are associations of smoking behavior with lower total brain volume, and grey and white matter volumes [18]. However, a significant question remains whether these associations represent predisposing features for the risk of developing cigarette smoking or are consequences of cigarette smoking. The UK Biobank presents a unique opportunity to study the association between smoking behaviors and brain features with a large sample of individuals who have completed comprehensive assessments and to shed light on whether associations with brain volumes and smoking behaviors are predisposing factors, or adverse consequences of cigarette smoking. Currently, the UK Biobank provides surveys on health behaviors and imaging

derived measures from magnetic resonance imaging (MRI) on approximately 40,000 participants. In addition, genetic data are available for UK Biobank participants. Our goal is to examine the associations between smoking behaviors, global brain volumes, and genetic variation to provide evidence for the direction of effect of the association between smoking behaviors and brain imaging measures by using traditional epidemiological methods and mediation analysis.

Bradford Hill, an eminent epidemiologist, developed criteria for establishing evidence of causality [19]. Hill's criteria of causation, originally developed to specify a causal relationship between smoking behavior and lung cancer, consists of 9 points: strength of association, consistency across sites and methods, specificity, temporality, biological gradient, plausibility, coherence, experimental evidence, and analogy (related evidence). We can use the different smoking measures (history of daily smoking, number of cigarette pack years smoked, and time since smoking cessation) available in the UK Biobank dataset to examine Hill's criteria and to build evidence as to whether observed brain differences represent predisposing factors that influence smoking behaviors or are consequences of the smoking exposure. We can study 1) the association between a history of daily smoking and global brain volumes, 2) whether there is a dose response relationship with greater cumulative exposure to smoking (measured by pack years) associated with changes in brain volumes; and 3) whether smoking cessation is associated with a reversal of changes in brain volumes, and 4) whether there are sub-regions of the brain that are more or less associated with smoking behaviors after correcting for the total brain volume changes.

We can also incorporate genetic data to further establish the direction of effect of smoking behaviors and brain volume. To test the association between genetic predisposition to smoking

behavior and brain volume differences, we can use summary statistics from the GWAS and Sequencing Consortium of Alcohol and Nicotine use (GSCAN) [15], a large genetic study of smoking behaviors, to create a polygenic risk score (PRS) for ever smoking, a summary score of an individual's genetic predisposition. In UK Biobank participants, we can examine 1) the association between PRS for smoking with history of daily smoking in UK Biobank, and 2) the association between PRS for smoking with global brain volumes. Lack of a strong association between genetic predisposition to smoking and brain volume differences would add evidence that smoking is negatively related to brain volume rather than a decrease in brain volume influences smoking behavior. Finally, we can use mediation analysis as a tool to study the direction of causation and the strength of daily smoking as a mediator. Converging results from these different methodologies can provide evidence for the direction of effect of the association between smoking behaviors and imaging measures of brain volume. An overview of the study is presented in Figure 1.

4.3. Methods

4.3.1. UK Biobank participants

Our sample included the 2019 UK Biobank released data of participants with imaging data. The UK Biobank study was approved by the National Health Service National Research Ethics Service (11/NW/0382). All the participants provided informed consent to participate the UK Biobank study (Study ID: 47267, 48123).

From the imaging dataset, we removed related individuals up to third degree (n=1,123), and individuals who withdrew consent following participation. We also excluded participants with neurological conditions (n=1,122), to eliminate potential confounding effects from these conditions (18). See supplementary figure 1 for the flow chart of sample processing, and supplementary table 1 for further details of participants with neurological conditions. This study follows the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) reporting guideline for cross-sectional studies.

4.3.2. Smoking Behaviors

Smoking phenotypes were defined using data from self-report surveys obtained during in-person assessment center visits at baseline ('instance 0', 2006-2010) and at the neuroimaging visit ('instance 2', 2012-2013). A history of daily smoking (n = 8,906) was defined by a consensus of reports of former or current daily smoking on surveys at both time points (visits). Never smoking (n = 23,188) was defined by a lifetime history of never smoking or smoking fewer than 100 cigarettes on both surveys. Those with a history of occasional smoking, but not smoking daily, and those with conflicting smoking status reports on the two surveys were excluded from the analysis (n = 7,494) so that the distinction between a history of daily smoking and never smoking

would be clearer and the data are more reliable. See supplementary figure 2 for the sample size and questionnaire details for the imaging subset. See supplementary table 3 for the baseline and imaging visit comparison of reported smoking behaviors.

Smoking pack years, (number of cigarette packs (one pack = 20 cigarettes) smoked per day times the number of years smoked) was derived for those with a history of daily smoking at the imaging survey. If this value was missing, smoking pack years was taken from the baseline survey. See supplementary figure 3 for pack year distribution in categories.

Age last smoked was obtained from the imaging survey; if this value was missing, it was taken from the baseline survey. Duration of smoking cessation was derived by subtracting the age last smoked from the participants' age at the imaging assessment.

Standardized imaging confound values (age, age^2, sex, age*sex, head size, head motion rfMRI, head motion tfMRI, date, date^2, site) were curated (21). Additional covariates that might confound the association between brain measures and smoking behaviors were included in analyses: Average household income, age completed full-time education, systolic blood pressure, diastolic blood pressure, body mass index, waist-hip ratio, weekly dose of alcohol (calculated by converting drink by type into an overall sum of drinks), stress, physical activity, diabetes, cancer, vascular/heart problems and other health conditions. Additional covariates included 10 ancestral principal components (PCs). See supplementary table 4 and supplementary text for further information on the selected covariates.

Imputation of missing values for all covariates was first done using participants' reports from the baseline survey. The additional missing values were imputed using R package MICE.

Supplemental text and supplementary table 4 give further details on missing data and data wrangling.

4.3.3. Imaging Derived Measures

T1 structural imaging-derived phenotypes

Detailed information regarding the UK Biobank image acquisition parameters, preprocessing pipeline, and estimation of brain-imaging derived measures is available elsewhere (https://biobank.ctsu.ox.ac.uk/crystal/crystal/docs/brain_mri.pdf; (20)). Briefly, T1-weighted scans were acquired at 1mm isotropic resolution using a Siemens Magnetom Skyra 3T scanner. Following brain extraction and nonlinear registration to MNI space with BET and FNIRT tools, respectively, tissue-type segmentation was performed using the FAST tool (20). T1 images are also processed with Freesurfer. Cortical surface atlases for Freesurfer modelling are used to extract area, volume, and mean cortical thickness imaging-derived phenotypes (Freesurfer DKT). Freesurfer ASEG tools are used for the extraction of subcortical regions and total measures of the brain (volume of brain, volume of grey matter, volume of white matter, and volume of CSF). Variable IDs (brain measures, covariates) used in these analyses are provided in supplementary table 2.

T2 susceptibility-weighted imaging-derived phenotypes

The susceptibility-weighted MRI scan employs a 3D gradient echo acquisition with a resolution of 0.8x0.8x3mm and acquires two echo times (TE = 9.4 and 20 ms). The T2* decay times of the signals are calculated using the magnitude images obtained at two different echo times (TEs). Then, the resulting imaging-derived phenotypes are determined by taking the median T2* values of the different subcortical regions that were defined from the T1 processing.

Diffusion imaging-derived phenotypes

The diffusion MRI data is obtained using two different b-values (b=1000 and 2000 s/mm^2, b-value measures the strength of the diffusion effects) at a spatial resolution of 2mm, with a multiband acceleration factor of 3. A total of 50 distinct diffusion-encoding directions were acquired for each diffusion-weighted shell, covering 100 distinct directions over the two b-values. Diffusion tensor imaging (DTI) uses b=1000 s/mm^2 data to generate fractional anisotrophy (FA), mean diffusivity (MD) and tensor mode (OD). Neurite orientation dispersion and density imaging (NODDI) modelling uses AMICO (Accelerated Microstructure Imaging via Convex Optimization) tool to generate voxelwise microstructural parameters such as intra-cellular volume fraction (ICVF), isotropic or free water volume fraction (ISOVF), and orientation dispersion index (OD). Tractography is performed using a parametric approach to estimate fiber orientations, and a generalized ball & stick model is fit to the multi-shell data to estimate up to three crossing fiber orientations per voxel. To extract meaningful IDPs, cross-subject alignment of white matter pathways is critical. Two complementary approaches are used, including tract-based spatial statistics (TBSS) and subject-specific probabilistic diffusion tractography to identify region of interests (ROIs) for 48 and 27 tracts, respectively.

Resting-functional imaging-derived phenotypes

Resting-state functional MRI used 2.4mm spatial resolution and TR=0.735s (repetition time). Imaging-derived phenotypes were obtained using independent component analysis (ICA) performed at two different dimensionalities (25 and 100), resulting in 21 and 55 signal networks, respectively. Dual regression was applied to calculate subject-specific BOLD time series for each network. The amplitude for each network (temporal standard deviation) and functional connectivity (full or partial correlation coefficients) between network pairs were calculated.

4.3.4. Genetic dataset

We used the UK Biobank genetic dataset to retrieve genome wide data for all participants of European ancestry (dataset version/number = ukb48123). We used GSCAN summary statistics with the UK Biobank sample excluded to create a polygenic risk score (PRS) for ever smoking with variants using PRSice-2 (22, 23). The PRS results have been pruned for sites with minor allele frequency (MAF) > 0.001, imputation quality (Effective_N/N) > 0.3, and an effective sample size of at least 10% of the maximum sample size. . Insertions and deletions were not included in GSCAN summary statistics, and also not included in the calculation of PRS. PRSice-2 utilizes p-value selection threshold approach, so according to the different thresholds, only those SNPs with a GWAS association p-value below a certain threshold are included in the calculation of the PRS. We tested the PRS for ever smoking to determine whether it predicted the history of daily smoking in UK Biobank as well as total brain measures in 1) the total sample; 2) the subset of participants who never smoked; and 3) the subset of participants who reported a lifetime history of daily smoking. See figure 1 for the overview of the study including genetic dataset.

4.3.5. Statistical analysis

We performed linear regression analysis using lm package from R for each question.

Question 1) Is a history of daily smoking associated with global brain measures?

Equation: Brain volume = History of daily smoking (dichotomous variable) + covariates

The following two analyses were undertaken only in those with a history of daily smoking.

Question 2) Is there a dose-response relationship between the heaviness of smoking (defined by pack years smoked) and global brain measures?

Equation: Brain volume = Pack years (continuous variable) + covariates

Question 3) Is there evidence of positive association between brain volume and time since smoking cessation among those with history of daily smoking?

Equation: Brain volume = Time since smoking cessation (continuous variable for those who smoked daily in the past) + pack years + covariates

Question 4) Is the ever smoking PRS associated with the history of daily smoking?

Equation: History of daily smoking = Ever Smoking PRS + covariates

Question 5) Is the ever smoking initiation PRS associated with the global brain measures?

Equation: Brain volume = Ever Smoking PRS + covariates

Question 6) Are there regions of the brain more or less associated with daily smoking after correcting for the total brain volume in addition to head size?

Equation: Brain sub-region volume = History of daily smoking + total brain volume + covariates

Question 7) Is a history of daily smoking associated with structural connectivity within (diffusion skeleton measures) or across (diffusion tract-based measures) the brain?

Equation: Diffusion MRI measures = History of daily smoking + global diffusion measures (ex. average of all FA measures when calculating FA measures) + covariates

Question 8) Is a history of daily smoking associated with resting functional connectivity in the brain?

Equation: Resting-functional MRI measure groups = History of daily smoking + covariates

For questions 1 to 5, a threshold of 0.05 was set as the level of significance. For question 6 to 8, 705 sub-regions were examined, thus the threshold of significance was set at a Bonferroni correction of $0.05/705 = 7.09 \times 10^{-5}$.

4.3.6. Mediation analysis for history of daily smoking and total grey matter volume

Mediation analysis was performed using 'mediation' package in R to measure the strength of the causal mediator (daily smoking) in the relationship between polygenic risk score for smoking initiation and the outcome (total brain volume) while adjusting for various confounding variables (age, age2, sex, age*sex, head size, head motion, date, date2, site, average household income, age completed full-time education, systolic blood pressure, diastolic blood pressure, body mass index, waist-hip ratio, weekly dose of alcohol). The average causal mediated effect (ACME), or

the statistical significance of the mediator, was calculated through this package. See supplementary figure 4 for the model for the mediation analysis.

4.4 Results

4.4.1. History of daily smoking was associated with global brain measures

A history of daily smoking was associated with a decrease in total brain volume, gray matter volume, white matter volume, and increased cerebral spinal fluid (CSF) volume (Table 2). Decreased volume of grey matter had a strong association with a history of daily smoking (Effect size = -2964mm^3 p-value = 2.04 x 10^{-16}), along with decreased volume of total brain (Effect size = -3360mm^3, p-value = 2.85 x 10^{-8}). Volume of white matter was modestly associated with a history of daily smoking (Effect size = -802mm^3, P-value = 4.68 x 10^{-2}).

4.4.2. Evidence of a dose response relationship with pack years

Among participants with a history of daily smoking, there was evidence of a dose response relationship with increasing number of pack years smoked associated with a decrease in brain volume and gray matter and increased CSF volume (Table 2). Volume of grey matter had a strong association with pack years of smoking (Effect size = -84mm^3, p-value = 3.25 x 10^{-5}), as well as volume of total brain (Effect size = -129mm^3, p-value = 1.23 x 10^{-4}). A modest association was seen with volume of white matter (Effect size = -64mm^3, p-value = 0.04). There was no significant association of pack years smoked with volume of CSF.

4.4.3. Time since smoking cessation moderately associated with total grey matter volume

There was no significant association between years since smoking cessation and total brain volume, total grey matter volume, white matter volume, and CSF volume.

4.4.4. Effect of genetic predisposition to smoking on total grey matter volume among smoking population

The ever smoking PRS was strongly associated with the history of daily smoking in UK Biobank (Effect size = 0.05, p-value = 4.20×10^{-84}) corroborating that these genetic variants collectively predict this smoking behavior. There was a modest association of the ever smoking PRS with reduced total grey matter volume (Effect size = - 424mm^3, p-value = 0.01) and increased white matter volume (Effect size = 367mm^3, p-value = 0.04) in the total sample (n=30,973) (Table 3). There is no evidence of a PRS-brain volume association in the subsets including only those who have a history of daily smoking. Additionally, there was a modest evidence of PRS-white matter volume association in the subset including only those who have never smoked (Effect size = 438mm^3, p-value = 0.04).

4.4.5. Mediation analysis between total grey matter volume, smoking initiation PRS, and history of daily smoking

Because total grey matter volume was modestly associated with PRS for smoking initiation, we performed a mediation analysis between ever smoking PRS, total grey matter volume (outcome), and the history of daily smoking (mediator). The association between the PRS for ever smoking and total grey matter volume became non-significant (Effect size: 0.04, p-value = 0.21) when the mediator, a history of daily smoking, was added (total/indirect causal mediation effect size (ACME): 0.005, p-value: $< 2 \times 10^{-16}$, total direct causal mediation effect size (ADE): 0.00, p-value: 1).

4.4.6. History of daily smoking was associated with cortical volume and thickness measures

The purpose of the subregion analyses is to determine if certain regions of the brain are more or less associated with a history of daily smoking after adjusting for head size and total brain volume. The correlation between head size and total brain volume was 0.7. In these sub-region

analyses, a significance level of 7.09×10^{-5} was selected based on a Bonferroni correction for 705 tests.

Of the 186 Freesurfer DKT measures based on white matter parcellation, 33 subregions were significantly associated with a history of daily smoking and only 5 (3%) remained significantly associated after correcting for the total brain volume. Mean thickness of superior frontal cortex (both hemispheres), volume of superiorfrontal (left hemisphere), volume of rostral middle frontal cortex (left hemisphere), and volume of medial orbital frontal cortex (left hemisphere) were all negatively associated with a history of daily smoking after correcting for total brain volume (Table 4). None of the other cortical regions passed the threshold of significance based on multiple testing and demonstrated a significant association with a history of daily smoking.

4.4.7. History of daily smoking was associated with increased ventricle sizes, and decreased cerebellum and subcortical volume measures

Of the 49 Freesurfer ASEG measures, 14 (29%) were significantly associated with a history of daily smoking after correcting for the total brain volume (before correction, 22 measures were significantly associated). Volume of white matter hypointensities, choroid-plexus in both hemispheres, ventricle choroid, and 3rd ventricle in the whole brain, volume of interior lateral ventricle (left hemisphere), volume of lateral ventricle (right hemisphere), were positively associated with a history of daily smoking after correcting for total brain volume. An increase in volume of all these regions is an adverse effect. Volume of cerebellum-white-matter, volume of ventral diencephalon (all both hemispheres), amygdala (left hemisphere), volume of corpus callosum central in the whole brain, were negatively associated with a history of daily smoking after correcting for total brain volume (Table 5).

4.4.8. Susceptibility-weighted MRI measures of putamen, caudate and pallidum associated with a history of daily smoking

Of the 14 Median T2star measures, 6 (43%) passed the stringent threshold of P-value = 7.09 x 10^{-5} and remained significantly associated with a history of daily smoking after correcting for the total volume of the brain. The 6 measures are: Median T2star measures in putamen, caudate and pallium in both hemispheres of the brain. All 6 measures had region-specific negative association with a history of daily smoking (Table 6).

4.4.9. Diffusion MRI measures within and across the brain associated with a history of daily smoking

Diffusion skeleton

Of the 288 diffusion skeleton measures, 53 (18%) passed the threshold of P-value = 7.09 x 10^{-5} and remained significantly associated with a history of daily smoking after correcting for the total global diffusion measures (89, before correction). Several tracts within the regions were found to be significant. Fornix cres+stria terminalis (mean FA in both hemispheres negatively associated; OD in left hemisphere positively associated), superior corona radiata (MD in both hemisphere positively associated; ICVF in both hemispheres negatively associated), posterior corona radiata (OD in both hemisphere negatively associated, ICVF in left hemisphere negatively associated), corticospinal tract (MD and ISOVF in right hemisphere negatively associated), superior fronto-occipital fasciculus (ICVF in right hemisphere, negative), cerebral peduncle (ICVF, OD in both hemispheres positively associated), pontine crossing tract (ISOVF, negative), uncinate fasciculus (FA in both hemispheres positively associated, OD in left

hemisphere negatively associated), and posterior thalamic radiation (ICVF in left hemisphere, negative) (Supplementary Figure 8).

Diffusion tract-based

Of the 162 diffusion tract-based measures, 10 (6%) remained to be significantly associated with a history of daily smoking after correcting for the total brain volume (28, before correction). Tract middle cerebellar peduncle (Weighted-mean MD, ISOVF positively associated), tract superior thalamic radiation (MD in left hemisphere positively associated, ICVF in right hemisphere negatively associated), and anterior thalamic radiation (ISOVF in right hemisphere positively associated) were significantly associated with a history of daily smoking (Supplementary Figure 9).

4.4.10. Resting-functional MRI measures of frontal lobe areas moderately associated with a history of daily smoking

Of the 6 Resting-functional MRI groups that are dimension-reduced from the original measures, 1 group showed moderate negative association with a history of daily smoking (Effect size: -0.05, P-value: 1.6×10^{-4}). The group mainly consisted of frontal lobe and Wernicke's area (Supplementary figure 10).

4.5. Discussion

We systematically examined the relationship between a history of daily smoking and global brain volume, and the preponderance of evidence supports an adverse association of smoking with brain volume. Daily smoking is associated with a decrease in total brain volume. Using the Hill criteria as a guide to study causation, we found a strong association between a history of daily smoking and brain imaging phenotypes as reported in previous studies. Several studies using different datasets and various analytical methods have identified a strong association between a history of daily smoking and global brain volume, grey matter volume, and white matter volume [18, 25-27]. We also found a significant biological gradient, with a dose-response effect of a history of more pack years of smoking associated with greater differences in brain volume. In addition, there is evidence of biological plausibility. Daily smoking is associated with many adverse health effects across multiple organ systems and adding the brain to the list of organs adversely affected by smoking is biologically plausible. There is similar evidence of alcohol causing adverse consequences on the brain which provides analogical evidence of the harms of smoking [28, 29]. A recent study investigated the causal relationship between smoking and alcohol and subcortical brain volume variations and concluded that smoking and heavy alcohol consumption can causally reduce subcortical brain volume [30]. Additionally, another recent study performed Mendelian Randomization and found a significant association between genetic liability to ever smoking and decreased gray matter volume [31].

We used genetics as a tool to provide further evidence that a history of daily smoking may be negatively related to brain volume. Mediation analysis provides convergent evidence highlighting the plausibility of smoking associated with decreases in brain volume. We found that a polygenic risk score for ever smoking was strongly associated with history of daily

smoking in UK Biobank, but minimally associated with total grey and white matter volume. With the additional mediation analysis on the PRS for ever smoking and the total grey matter volume using history of daily smoking as a mediator, we found that the mediator effect was strong, and the association between the PRS and brain volume disappeared. Through this, we have additional support that smoking is negatively associated with the differences in brain volume.

The complexity of the relationship between smoking history and brain imaging phenotypes underscores the debate regarding causation: are brain differences predisposing to smoking behavior, or are the brain differences a consequence of smoking behaviors? There are studies suggesting that brain differences are a predisposing factor for alcohol consumption, rather than reflecting alcohol-induced atrophy [17, 32]. There is evidence that greater volume or thickness in brain regions (pars opercularis, cuneus) and lower volume in brain regions (basal forebrain, insular grey matter volume, right dorsolateral prefrontal cortex) may contribute to the development of problematic alcohol use [17, 32]. It is likely that there are also differences in brain measures that are predisposing factors for the initiation of smoking behaviors [33]. While we acknowledge that there are studies supporting the notion that regional brain differences may be a predisposing factor for alcohol consumption, we focused our investigation on the relationship between smoking behavior and global brain volume. The evidence presented in this study suggests that the changes in total brain volume, total grey matter volume, total white matter volume, and subcortical/cortical regional volumes more likely reflect adverse consequences of a history of daily smoking behavior. In addition, hippocampal volume, an important brain region effected by Alzheimer's disease, is negatively associated with a history of daily smoking. This finding is consistent with smoking, which has been identified as a

modifiable risk factor for Alzheimer's disease, accelerating the development of this illness [7]. These brain changes seem to be long-lasting and we identified no evidence of an increase in brain volume after smoking cessation.

In addition to studying the total brain measures, we examined whether sub-regions of the brain are more or less associated with daily smoking after correcting for the total brain volume. For cortical regions, we found that the thickness of superior frontal cortex is negatively associated with daily smoking, which is consistent with the evidence found in the recent studies that smoking is associated with cortical thinning [34, 35]. Additionally, we identified that the volume of superior frontal cortex, rostral middle frontal cortex, and precentral gyrus were more negatively associated with daily smoking, beyond the overall decrease total brain volume associated with a history of daily smoking. For cerebellum, the volume of cerebellum white matter in the left hemisphere was negatively associated with daily smoking, and volume of corpus callosum also showed negative association, as shown in the previous studies [36, 37]. Volume of thalamus and amygdala were more negatively associated with daily smoking, as shown from the previous studies [27, 38-40]. We found that the volume of choroid plexus, lateral ventricle, and 3rd ventricle were positively associated with daily smoking than the other regions. These areas are the essential parts or paths of the cerebrospinal fluid system [41, 42], and these findings are consistent with a compensatory increase in CSF volume as total brain volume decreases.

Finally, we examined the association of daily smoking with structural and functional connectivity. We found that increased mean diffusivity (MD), a generally a negative sign for white matter integrity, in superior thalamic radiation tract was associated with daily smoking, which is consistent with the previous findings [18]. Increased MD was also found in superior

corona radiata and cerebellar peduncle, as shown in the previous studies [18, 43]. Additionally, decreased fractional anisotropy (FA), also a negative sign for white matter integrity, in fornix was associated with daily smoking. It is not surprising that the structural connectivity of fornix, a pathway that connects several subcortical structures, was identified to be associated with smoking. With the susceptibility-weighted imaging measures, we additionally identified the strong negative association between caudate, putamen, and pallidum with daily smoking, consistent with the previous findings [44, 45]. We also found one modest association with the resting functional MRI measure group mainly associated with frontal lobe, suggesting the negative impact of smoking on functional connectivity, as well as structural connectivity.

The best way to address causation is through triangulation of data and convergent evidence including cross-sectional association, longitudinal data, and experimental paradigms UK Biobank dataset is large and provides ample statistical power, we examined cross-sectional data of brain imaging. Longitudinal data from UK Biobank neuroimaging is growing, but it remains limited at this time. Importantly, almost all participants in UK Biobank who smoked had quit smoking by the time of the first assessment, which limits longitudinal analyses of the effect of current smoking on subsequent brain imaging measures. There is also the need for prospective development data to better understand the complex interplay between behavior and brain structure. The Adolescent Brain Cognitive Development study (ABCD), the largest neuroimaging study of brain development in the U.S. will best be able to disentangle what brain measures represent predisposing factors to substance use and adverse consequences from substance use.

4.6. Conclusion

We examined the nature of the relationship between daily smoking and brain imaging phenotypes using traditional epidemiological criteria (Hill's criteria), and genetics tools (PRS and mediation analysis) in a large dataset of participants. There was a dose effect with a history of heavier smoking being associated with more severe adverse effects. We found minimal evidence that a genetic predisposition to smoking is associated with total brain volume, and this association became insignificant when a history of daily smoking was set as a mediator variable. Thus, mediation analysis further supports the effect of smoking leading to decreases in brain volume. We found that a history of smoking was strongly associated with adverse changes in total brain volumes and certain cortical, cerebellar, and subcortical regional volumes. Finally, there was no evidence of in an increase in brain volume following smoking cessation. In totality, these findings provide additional evidence that a history of daily smoking is strongly associated with long-term global adverse consequences in the brain.

4.7. Acknowledgments

4.7.1. Funding

This work was supported by National Institute on Alcohol Abuse and Alcoholism grants U10AA008401 (APA, LJB, YC) and R01AA027049 (LJB, DBH, EOJ, VT), National Institute on Drug Abuse grants K12DA041449 (LJB) and R01DA044014 (LJB, VT), and National Institute on Aging grant R56AG058726 (LJB, PI: Galama). JB is funded by the NIH (R01MH128286) and the McDonnell Center for Systems Neuroscience. APA is funded by the NIH (R01AA025646, R01DA89801). VT is additionally funded by Washington University Institute of Clinical and Translational Sciences (TL1TR002344). RB is funded by NIH (R21AA027827, R01DA054750, R01AG061162, U01DA055367). The manuscript is posted in medRxiv.

4.7.2. Disclosures

Dr. Laura J. Bierut is listed as an inventor 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. The other authors reported no biomedical financial interests or potential conflicts of interest.

4.8. References

1. Lushniak BD, Samet JM, Pechacek TF, Norman LA, Taylor PA (2014): The health consequences of smoking—50 years of progress: a report of the Surgeon General.

2. US Department of Health Human Services (2010): How tobacco smoke causes disease: What it means to you. Atlanta: US Department of Health and Human Services, Centers for Disease Control and Prevention. *National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health*. 1993:18.

3. Benowitz NL (1997): The role of nicotine in smoking-related cardiovascular disease. *Preventive Medicine*. 26:412-417.

4. Centers for Disease Control Prevention (2013): QuickStats: Number of deaths from 10 leading causes—National Vital Statistics System. *United States*.

5. Nianogo RA, Rosenwohl-Mack A, Yaffe K, Carrasco A, Hoffmann CM, Barnes DE (2022): Risk factors associated with Alzheimer disease and related dementias by sex and race and ethnicity in the US. *JAMA Neurology*. 79:584-591.

6. Norton S, Matthews FE, Barnes DE, Yaffe K, Brayne C (2014): Potential for primary prevention of Alzheimer's disease: an analysis of population-based data. *The Lancet Neurology*. 13:788-794.

7. Livingston G, Huntley J, Sommerlad A, Ames D, Ballard C, Banerjee S, et al. (2020): Dementia prevention, intervention, and care: 2020 report of the Lancet Commission. *The Lancet*. 396:413-446.

8. Barnes DE, Yaffe K (2011): The projected effect of risk factor reduction on Alzheimer's disease prevalence. *The Lancet Neurology*. 10:819-828.

9. Durazzo TC, Mattsson N, Weiner MW, Initiative AsDN (2014): Smoking and increased Alzheimer's disease risk: a review of potential mechanisms. *Alzheimer's & Dementia*. 10:S122-S145.

10. Vink JM, Beem AL, Posthuma D, Neale MC, Willemsen G, Kendler KS, et al. (2004): Linkage analysis of smoking initiation and quantity in Dutch sibling pairs. *The Pharmacogenomics Journal*. 4:274-282.

11. Vink JM, Willemsen G, Boomsma DI (2005): Heritability of smoking initiation and nicotine dependence. *Behavior Genetics*. 35:397-406.

12. Xian H, Scherrer JF, Madden PA, Lyons MJ, Tsuang M, True WR, et al. (2003): The heritability of failed smoking cessation and nicotine withdrawal in twins who smoked and attempted to quit. *Nicotine & Tobacco Research*. 5:245-254.

13. Xu K, Li B, McGinnis KA, Vickers-Smith R, Dao C, Sun N, et al. (2020): Genome-wide association study of smoking trajectory and meta-analysis of smoking status in 842,000 individuals. *Nature Communications*. 11:5302.

14. Saunders GR, Wang X, Chen F, Jang S-K, Liu M, Wang C, et al. (2022): Genetic diversity fuels gene discovery for tobacco and alcohol use. *Nature*. 612:720-724.

15. Liu M, Jiang Y, Wedow R, Li Y, Brazel DM, Chen F, et al. (2019): Association studies of up to 1.2 million individuals yield new insights into the genetic etiology of tobacco and alcohol use. *Nature Genetics*. 51:237-244.

16. Bogdan R, Hatoum AS, Johnson EC, Agrawal A (2023): The Genetically Informed Neurobiology of Addiction (GINA) model. *Nature Reviews Neuroscience*. 24:40-57.

17. Hatoum AS, Johnson EC, Agrawal A, Bogdan R (2021): Brain structure and problematic alcohol use: a test of plausible causation using latent causal variable analysis. *Brain Imaging and Behavior*. 15:2741-2745.

18. Gray JC, Thompson M, Bachman C, Owens MM, Murphy M, Palmer R (2020): Associations of cigarette smoking with gray and white matter in the UK Biobank. *Neuropsychopharmacology*. 45:1215-1222.

19. Hill AB (2015): The environment and disease: association or causation? *Journal of the Royal Society of Medicine*. 108:32-37.

20. Miller KL, Alfaro-Almagro F, Bangerter NK, Thomas DL, Yacoub E, Xu J, et al. (2016): Multimodal population brain imaging in the UK Biobank prospective epidemiological study. *Nature Neuroscience*. 19:1523-1536.

21. Alfaro-Almagro F, McCarthy P, Afyouni S, Andersson JLR, Bastiani M, Miller KL, et al. (2021): Confound modelling in UK Biobank brain imaging. *NeuroImage*. 224:117002.

22. Choi SW, O'Reilly PF (2019): PRSice-2: Polygenic Risk Score software for biobank-scale data. *Gigascience*. 8:giz082.

23. [dataset] Liu M, Jiang Y, Wedow R, Li Y, Brazel DM, Chen F, et al. (2019): Data Related to Association studies of up to 1.2 million individuals yield new insights into the genetic etiology of tobacco and alcohol use. Retrieved from the Data Repository for the University of Minnesota, https://doi.org/10.13020/3b1n-ff32.

24. Hemani G, Zheng J, Elsworth B, Wade K, Haberland V, Baird D, et al. (2018): The MR-Base platform supports systematic causal inference across the human phenome. *Elife.* 7:e34408.

25. Peng P, Li M, Liu H, Tian Y-R, Chu S-L, Van Halm-Lutterodt N, et al. (2018): Brain structure alterations in respect to tobacco consumption and nicotine dependence: a comparative voxel-based morphometry study. *Front Neuroanat*. 12:43.

26. Fritz H-C, Wittfeld K, Schmidt CO, Domin M, Grabe HJ, Hegenscheid K, et al. (2014): Current smoking and reduced gray matter volume—a voxel-based morphometry study. *Neuropsychopharmacology*. 39:2594-2600.

27. Elbejjani M, Auer R, Jacobs DR Jr., Haight T, Davatzikos C, Goff DC Jr., et al. (2019): Cigarette smoking and gray matter brain volumes in middle age adults: the CARDIA Brain MRI sub-study. *Transl Psychiatry.* 9:78.

28. Topiwala A, Wang C, Ebmeier KP, Burgess S, Bell S, Levey DF, et al. (2022): Associations between moderate alcohol consumption, brain iron, and cognition in UK Biobank participants: observational and mendelian randomization analyses. *PLoS medicine.* 19:e1004039.

29. Daviet R, Aydogan G, Jagannathan K, Spilka N, Koellinger PD, Kranzler HR, et al. (2022): Associations between alcohol consumption and gray and white matter volumes in the UK Biobank. *Nature Communications.* 13:1175.

30. Logtenberg E, Overbeek MF, Pasman JA, Abdellaoui A, Luijten M, Van Holst RJ, et al. (2022): Investigating the causal nature of the relationship of subcortical brain volume with smoking and alcohol use. *The British Journal of Psychiatry*. 221:377-385.

31. Lin W, Zhu L, Lu Y (2023): Association of smoking with brain gray and white matter volume: a Mendelian randomization study. *Neurological Sciences.* 1-7.

32. Baranger DA, Demers CH, Elsayed NM, Knodt AR, Radtke SR, Desmarais A, et al. (2020): Convergent evidence for predispositional effects of brain gray matter volume on alcohol consumption. *Biological Psychiatry*. 87:645-655.

33. Zou R, Boer OD, Felix JF, Muetzel RL, Franken IH, Cecil CA, et al. (2022): Association of maternal tobacco use during pregnancy with preadolescent brain morphology among offspring. *JAMA network open.* 5:e2224701-e2224701.

34. Akkermans SEA, van Rooij D, Rommelse N, Hartman CA, Hoekstra PJ, Franke B, et al. (2017): Effect of tobacco smoking on frontal cortical thickness development: A longitudinal study in a mixed cohort of ADHD-affected and -unaffected youth. *Eur Neuropsychopharmacol.* 27:1022-1031.

35. Karama S, Ducharme S, Corley J, Chouinard-Decorte F, Starr JM, Wardlaw JM, et al. (2015): Cigarette smoking and thinning of the brain's cortex. *Molecular Psychiatry.* 20:778-785.

36. Tewari A, Hasan M, Sahai A, Sharma PK, Agarwal AK (2011): Nicotine mediated microcystic oedema in white matter of cerebellum: possible relationship to postural imbalance. *Ann Neurosci.* 18:14-16.

37. Paus T, Nawazkhan I, Leonard G, Perron M, Pike GB, Pitiot A, et al. (2008): Corpus callosum in adolescent offspring exposed prenatally to maternal cigarette smoking. *Neuroimage.* 40:435-441.

38. Lor CS, Haugg A, Zhang M, Schneider L, Herdener M, Quednow BB, et al. (2023): Thalamic volume and functional connectivity are associated with nicotine dependence severity and craving. *Addiction Biology.* 28:e13261.

39. Akkermans SEA, van Rooij D, Rommelse N, Hartman CA, Hoekstra PJ, Franke B, et al. (2017): Effect of tobacco smoking on frontal cortical thickness development: A longitudinal study in a mixed cohort of ADHD-affected and -unaffected youth. *Eur Neuropsychopharmacol.* 27:1022-1031.

40. Karama S, Ducharme S, Corley J, Chouinard-Decorte F, Starr JM, Wardlaw JM, et al. (2015): Cigarette smoking and thinning of the brain's cortex. *Molecular Psychiatry.* 20:778-785.

41. Li H, Liu Y, Xing L, Yang X, Xu J, Ren Q, et al. (2020): Association of cigarette smoking with sleep disturbance and neurotransmitters in cerebrospinal fluid. *Nature and Science of Sleep.* 12: 801-808.Brinker T, Stopa E, Morrison J, Klinge P (2014): A new look at cerebrospinal fluid circulation. *Fluids and Barriers of the CNS.* 11:1-16.

42. Brinker T, Stopa E, Morrison J, Klinge P (2014): A new look at cerebrospinal fluid circulation. *Fluids and Barriers of the CNS.* 11:1-16.

43. Hudkins M, O'Neill J, Tobias MC, Bartzokis G, London ED (2012): Cigarette smoking and white matter microstructure. *Psychopharmacology*. 221:285-295.

44. Salokangas RK, Vilkman H, Ilonen T, Taiminen T, Bergman Jn, Haaparanta M, et al. (2000): High levels of dopamine activity in the basal ganglia of cigarette smokers. *American Journal of Psychiatry*. 157:632-634.

45. Durazzo TC, Meyerhoff DJ, Yoder KK, Murray DE (2017): Cigarette smoking is associated with amplified age-related volume loss in subcortical brain regions. *Drug and alcohol dependence*. 177:228-236.

4.9. Tables

Table 4.1. Demographic, smoking and health related variables (Total N= 32,094)

	Daily smoked (N = 8,906)	**Never smoked (N = 23,188)**
	Mean ± SD	**Mean ± SD**
Age	65.14 ± 7.53	63.21 ± 7.65
Sex (n females, %)		
Female (n, %)	3,967 (44.5)	13,049 (56.3)
Male (n, %)	4,939 (55.5)	10,139 (43.7)
Income (£)		
Less than 18,000 (n, %)	1,283 (14.4)	2,692 (11.6)
18,000 to 30,999 (n, %)	2,699 (30.3)	6,057 (26.1)
31,000 to 51,999 (n, %)	2,646 (29.7)	7,039 (30.4)
52,000 to 100,000 (n, %)	1,790 (20.1)	5,579 (24.1)
Greater than 100,000 (n, %)	488 (5.5)	1,821 (7.9)
Age completed full time education	19.20 ±3.53	20.10 ±3.33
Diabetes		
Yes (n, %)	353 (4.0)	468 (2.0)
Cancer		
Yes (n, %)	552 (6.2)	1,182 (5.1)
Vascular/heart problems		
Heart attack, angina (n, %)	2,292 (25.7)	4,353 (18.8)
Other health conditions*		
Yes (n, %)	1,595 (17.9)	3,266 (14.1)
Stress, illness, bereavement		
Illness, injury, bereavement, stress (n, %)	3,819 (42.9)	9,451 (40.8)
None of the above (n, %)	5,087 (57.1)	13,737 (59.2)
Body Mass Index	27.29 ±4.28	26.25 ±4.20
Waist/hip ratio	0.89±0.09	0.86±0.09
Systolic blood pressure (mmHg)	141.38 ±20.06	139.65±19.64
Diastolic blood pressure (mmHg)	79.38±10.59	79.31±10.68
Weekly drinks of alcohol	12.66 ± 10.86	8.21±7.73
Non-vigorous physical activity**	3.44±2.30	3.43±2.26
Vigorous physical activity**	1.81±1.86	1.86±1.81

*Answer to the question: Has a doctor ever told you that you have had any other serious medical conditions or disabilities?

**Number of days/week of non-vigorous or vigorous physical activity 10+ minutes

Table 4.2. Effect size and p-value for total brain measures with the smoking phenotypes

History of daily smoking (N=32,094)

Brain measures	Effect size	SE	t	P-value
Volume of brain	-3,360.95	605.39	-5.55	2.85×10^{-8}
Volume of grey matter	-2,964.18	360.42	-8.22	2.04×10^{-16}
Volume of white matter	-801.74	403.24	-1.99	4.68×10^{-2}
Volume of cerebrospinal fluid	4.93	3.03	1.63	0.10

Pack years of smoking (N=8,622)

Brain measures	Effect size	SE	t	P-value
Volume of brain	-128.75	33.52	-3.84	1.23×10^{-4}
Volume of grey matter	-83.87	20.17	-4.16	3.25×10^{-5}
Volume of white matter	-63.61	22.28	-2.85	4.32×10^{-3}
Volume of cerebrospinal fluid	0.29	0.17	1.68	0.09

Time since smoking cessation (N=8,111)

Brain measures	Effect size	SE	t	P-value
Volume of brain	-5.01	57.08	-0.09	0.93
Volume of grey matter	9.86	34.39	0.29	0.77
Volume of white matter	15.64	38.04	0.41	0.68
Volume of cerebrospinal fluid	-0.55	0.29	-1.89	0.06

Covariates: Weekly alcohol use, diastolic and systolic blood pressure, Body Mass Index, waist-hip ratio, income, age completed full time education, diabetes, vascular/heart problems, other health conditions/disabilities, physical activity, stress, and imaging confounds (age, age2, sex, age*sex, head size, head motion rfMRI, head motion tfMRI, date, date2, site)

Effect sizes are in mm^3

Table 4.3. Effect size and p-value for total brain measures associated with the smoking initiation Polygenic Risk Score (PRS)

Smoking Initiation PRS (Total population, N=30,973*)

Brain measures	Effect size	SE	t	P-value
Volume of brain	-36.61	277.32	-0.13	0.89
Volume of grey matter	-424.48	165.33	-2.57	0.01
Volume of white matter	366.99	184.92	1.98	0.04
Volume of cerebrospinal fluid	-1.23	1.39	-0.89	0.38

Smoking Initiation PRS (Never smoked population, N=22,298)

Brain measures	Effect size	SE	t	P-value
Volume of brain	157.46	323.90	0.49	0.63
Volume of grey matter	-315.65	192.13	-1.64	0.10
Volume of white matter	437.84	216.52	2.02	0.04
Volume of cerebrospinal fluid	-0.34	1.61	-0.21	0.83

Smoking Initiation PRS (Daily smoked population, N=8,675)

Brain measures	Effect size	SE	t	P-value
Volume of brain	-124.13	531.72	-0.23	0.82
Volume of grey matter	-327.46	320.51	-1.02	0.31
Volume of white matter	262.91	353.07	0.74	0.46
Volume of cerebrospinal fluid	-4.16	2.72	-1.53	0.13

Other PRS thresholds (0.4, 0.3, 0.2, 0.1, 0.05, 0.01, $1x10^{-3}$, $1x10^{-4}$, $1x10^{-5}$, $1x10^{-6}$, $1x10^{-7}$, $5x10^{-8}$) are included in supplementary table 6

*Sample size is after filtering for robust genetic information

Effect sizes are in mm^3

Table 4.4. Effect size and p-value for total brain measures associated with the Freesurfer DKT measures

Brain measures	Hemisphere	Effect size	SE	t	P-value
Mean thickness of superiorfrontal	Left	-5.26×10^{-3}	1.01×10^{-3}	-5.23	1.67×10^{-7}
	Right	-4.07×10^{-3}	9.33×10^{-4}	-4.36	1.29×10^{-5}
Volume of superiorfrontal	Left	-122.14	25.17	-4.85	1.22×10^{-6}
Volume of rostralmiddlefrontal	Left	-116.07	28.53	-4.07	4.75×10^{-5}
Volume of medialorbitofrontal	Left	-27.66	6.27	-4.41	1.02×10^{-5}

Covariates: Weekly alcohol use, diastolic and systolic blood pressure, Body Mass Index, waist-hip ratio, income, age completed full time education, diabetes, vascular/heart problems, other health conditions/disabilities, physical activity, stress, and imaging confounds (age, age2, sex, age*sex, head size, head motion rfMRI, head motion tfMRI, date, date2, site)

Effect sizes are in mm^3

Table 4.5. Effect size and p-value for total brain measures associated with the Freesurfer ASEG measures

Brain measures	Hemisphere	Effect size	SE	t	P-value
Volume of choroid-plexus	Left	26.45	2.66	9.95	2.68×10^{-23}
	Right	31.02	2.54	12.20	3.71×10^{-34}
Volume of Inf-Lat-Vent (Inferior-lateral-ventricle)	Left	13.86	3.40	4.07	4.63×10^{-5}
Volume of 3rd-Ventricle	Whole	26.32	5.76	4.57	4.86×10^{-6}
Volume of Lateral-Ventricle	Right	304.82	75.44	4.04	5.34×10^{-5}
Volume of VentricleChoroid	Whole	703.66	164.11	4.29	1.81×10^{-5}
Volume of WM-hypointensities	Whole	152.49	33.18	4.60	4.32×10^{-6}
Volume of Cerebellum-White-Matter	Left	-128.20	22.38	-5.73	1.03×10^{-8}
	Right	-120.73	24.60	-4.91	9.23×10^{-7}
Volume of VentralDC (Ventral diencephalon)	Left	-15.02	3.49	-4.31	1.65×10^{-5}
	Right	-16.76	3.37	-4.98	6.41×10^{-7}
Volume of Amygdala	Left	-11.30	2.29	-4.93	8.43×10^{-7}
Volume of CC-Central (Corpus callosum-central)	Whole	-6.56	1.45	-4.54	5.74×10^{-6}

Covariates: Weekly alcohol use, diastolic and systolic blood pressure, Body Mass Index, waist-hip ratio, income, age completed full time education, diabetes, vascular/heart problems, other health conditions/disabilities, physical activity, stress, and imaging confounds (age, age2, sex, age*sex, head size, head motion rfMRI, head motion tfMRI, date, date2, site

Effect sizes are in mm^3

Table 4.6. Effect size and p-value for total brain measures associated with the Median T2 star measures (susceptibility-weighted IDPs)

Brain measures	Hemisphere	Effect size	SE	t	P-value
Median T2 star in putamen	Left	-0.92	0.06	-14.85	1.08×10^{-49}
	Right	-0.88	0.06	-14.56	7.27×10^{-48}
Median T2 star in caudate	Left	-0.66	0.06	-11.72	1.25×10^{-31}
	Right	-0.58	0.06	-10.36	4.17×10^{-25}
Median T2 star in pallidum	Left	-0.49	0.05	-9.98	1.99×10^{-23}
	Right	-0.49	0.05	-10.18	2.61×10^{-24}

Covariates: Weekly alcohol use, diastolic and systolic blood pressure, Body Mass Index, waist-hip ratio, income, age completed full time education, diabetes, vascular/heart problems, other health conditions/disabilities, physical activity, stress, and imaging confounds (age, age2, sex, age*sex, head size, head motion rfMRI, head motion tfMRI, date, date2, site

Effect sizes are in mm^3

4.10. Figures

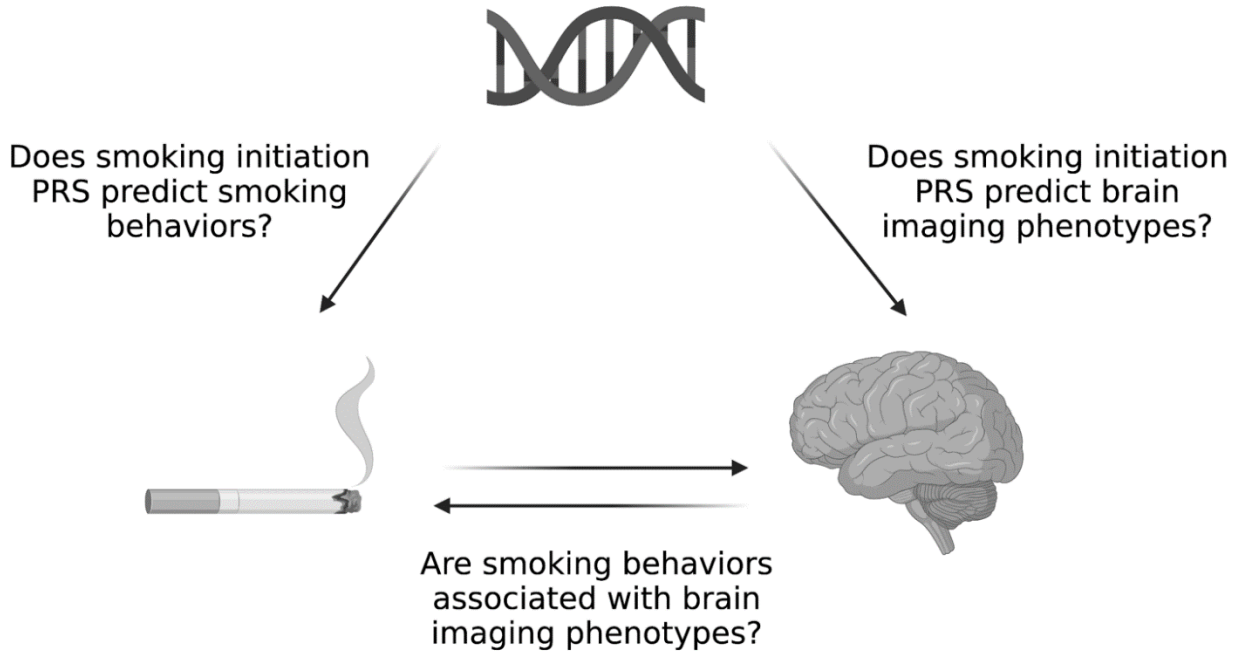


Figure 4.1: Overview of the study. We examined:

1) the predictive ability of the smoking initiation PRS for smoking for a history of daily smoking

2) the association between the smoking initiation PRS for smoking initiation and brain measures.

2) the association between smoking behaviors and brain measures

(Created with BioRender.com)

4.11. Supplementary Text

Covariates were selected to account for potentially confounding variables (1-13). Covariates include weekly alcohol use, diastolic and systolic blood pressure, body mass index (BMI), waist-hip ratio, income, age completed full time education, socioeconomic status (SES), stress, physical activity, diabetes, cancer, vascular/heart problems, other health conditions, and imaging confounds. Imaging confounds were age, age^2, sex, age*sex, head size, head motion rfMRI, head motion tfMRI, date, date^2, site.

Imaging covariates were processed according to UK Biobank-recommended scripts from Alfaro Almagro 2021 (9). UK Biobank imaging data were collected at three different sites. Every imaging covariate excluding sex was split into three sites to account for the potential confounding effect of the imaging site. Then the covariates were normalized using the median and median absolute deviation * 1.48 (one SD). The variable names were converted to site#_variable (ex. site1_age).

For non-imaging covariates, we first acquired the answers from the questionnaire completed during imaging visit (the participants were given the same touchscreen questionnaire as the baseline visit). If the answer was missing for the imaging visit, then we used the answers from the baseline visit to "backfill" the missing answers. Percent missing in supplementary table 4 indicates the missing data right after backfilling, and before imputation using MICE (10). Waist-hip ratio was acquired from waist circumference and hip circumference. Also, the only two education-related variables were age completed full time education and education qualifications. Age completed full time education was originally missing 19% of the answers after backfilling, but we used education qualifications to additionally fill in the missing data. Education qualification is a categorical variable which indicate the degree, professional qualifications, or

tests such as GCSE and A levels. We found the average age of completing such qualifications and added this age into age completed full time education variable. After doing this, the missing percentage decreased to 0.39.

Then we performed MICE to ensure that we had no missing data in our covariates. We did an approach with seed = 103, and 5 iterations. Sex was skipped since it did not have a missing value, but was used as a predictor. For all the continuous variables, we used norm method, for categorical income variable we used polyreg, and for binomial variables we used logreg. After MICE, the missing percentage for our non-imaging covariates was 0. (see UKB_sample_processing.R script in https://github.com/yoonhoochang/UKB_Global_Smoking for code details)

4.11.1. Supplemental Text References

1. Cox SR, Ritchie SJ, Tucker-Drob EM, Liewald DC, Hagenaars SP, Davies G, et al. (2016): Ageing and brain white matter structure in 3,513 UK Biobank participants. *Nature Communications*. 7:13629.

2. Yaple ZA, Yu R (2020): Functional and structural brain correlates of socioeconomic status. *Cerebral Cortex*. 30:181-196.

3. Hiscock R, Bauld L, Amos A, Fidler JA, Munafò M (2012): Socioeconomic status and smoking: a review. *Annals of the New York Academy of Sciences*. 1248:107-123.

4. Beard E, West R, Michie S, Brown J (2017): Association between smoking and alcohol-related behaviours: a time–series analysis of population trends in England. *Addiction*. 112:1832-1841.

5. Xiao P, Dai Z, Zhong J, Zhu Y, Shi H, Pan P (2015): Regional gray matter deficits in alcohol dependence: A meta-analysis of voxel-based morphometry studies. *Drug and alcohol dependence*. 153:22-28.

6. Zahr NM, Pfefferbaum A (2017): Alcohol's effects on the brain: neuroimaging results in humans and animal models. *Alcohol research: current reviews*.

7. Cox SR, Lyall DM, Ritchie SJ, Bastin ME, Harris MA, Buchanan CR, et al. (2019): Associations between vascular risk factors and brain MRI indices in UK Biobank. *European heart journal*. 40:2290-2300.

8. Dare S, Mackay DF, Pell JP (2015): Relationship between smoking and obesity: a cross-sectional study of 499,504 middle-aged adults in the UK general population. *PloS one*. 10:e0123579.

9. Jahnel T, Ferguson SG, Shiffman S, Schüz B (2019): Daily stress as link between disadvantage and smoking: an ecological momentary assessment study. *BMC Public Health.* 19:1284.

10. Heydari G, Hosseini M, Yousefifard M, Asady H, Baikpour M, Barat A (2015): Smoking and physical activity in healthy adults: A cross-sectional study in Tehran. *Tanaffos.* 14:238-245.

11. Campagna D, Alamo A, Di Pino A, Russo C, Calogero AE, Purrello F, et al. (2019): Smoking and diabetes: dangerous liaisons and confusing relationships. *Diabetol Metab Syndr.* 11:85.

12. Siemiatycki J, Krewski D, Franco E, Kaiserman M (1995): Associations between cigarette smoking and each of 21 types of cancer: a multi-site case-control study. *Int J Epidemiol.* 24:504-514.

13. Wang W, Zhao T, Geng K, Yuan G, Chen Y, Xu Y (2021): Smoking and the pathophysiology of peripheral artery disease. *Front Cardiovasc Med.* 8:704106.

14. Alfaro-Almagro F, McCarthy P, Afyouni S, Andersson JLR, Bastiani M, Miller KL, et al. (2021): Confound modelling in UK Biobank brain imaging. *NeuroImage*. 224:117002.

15. Zhang Z (2016): Multiple imputation with multivariate imputation by chained equation (MICE) package. *Annals of translational medicine*. 4:2.

4.12. Supplemental Tables

Supplementary table 4.1. Neurological condition diagnosis codes and number of participants removed for those conditions (N = 1122)

Neurological disease/trauma/conditions	Diagnosis code	Sample N	Percent
Stroke or ischaemic stroke	1081	318	0.78
Transient ischemic attack	1082	220	0.54
Epilepsy	1264	163	0.40
Meningitis	1247	112	0.27
Multiple sclerosis	1261	107	0.26
Parkinsons	1262	71	0.17
Head injury	1266	42	0.10
Encephalitis	1246	18	0.04
Brain hemorrhage	1491	16	0.04
Subarachnoid hemorrhage	1086	15	0.04
Guillan-Barre syndrome	1256	14	0.03
Meningioma	1659	13	0.03
Dementia	1263	13	0.03
Ischaemic stroke	1583	12	0.03
Subdural hematoma	1083	10	0.02
Spina bifida	1524	7	0.02
Cerebral aneurysm	1425	6	0.01
Neurological disease / trauma	1240	4	0.01
Motor neuron disease	1259	4	0.01
Other demyelinating disease	1397	4	0.01
Brain / intracranial abscess	1245	3	0.01
Chronic degenerative neurological	1258	1	2.50×10^{-3}
Cerebral palsy	1433	1	2.50×10^{-3}

Diagnosis code from UK Biobank data-field 20002 (baseline visit, primary and additional diagnoses). There are multiple diagnosis columns for this data-field. Some participants have more than one neurological condition.

Supplementary table 4.2. Variables and corresponding UK Biobank data-field ID

Variable	Data-field ID
Age	21003
Age completed full-time education	845, 6138
Body Mass Index	21001
Current tobacco smoking	1239
Date	53
Diabetes, cancer, other health conditions	2443, 2453, 2473
Diastolic blood pressure	4079
Head size	25000
Hip circumference	49
Imaging site	54
Income	738
Neurological conditions	20002
Past tobacco smoking	1249
Physical Activity	884, 904
rfMRI motion	25741
Sex	31
Stress, illness, bereavement	6145
Systolic blood pressure	4080
tfMRI motion	25742
Vascular/heart problems	6150
Waist circumference	48
Weekly dose of alcohol	1558, 4407, 4418, 4429, 4451, 4462
Volume of Brain	26514
Volume of Gray Matter	26518
Volume of White Matter	26553, 26584
Volume of CSF	26527

Supplementary table 4.3. Smoking history at baseline vs. imaging visit (starting from N=39,588)

Baseline/ Imaging	Daily current	Daily former	More than 100 Cigs	Less than 100 Cigs	Never smoked
Daily current	630	784	89	3	0
Daily former	109	7383	1027	44	69
More than 100	39	671	3081	585	229
Less than 100	0	10	263	4150	2263
Never smoked	2	18	56	1115	15660

Bold letters are baseline visits.

Green shades indicate those with a consistent history of daily smoking.

Red shades indicate those with a consistent history of never smoking at both baseline and imaging visit.

Supplementary table 4.4. Missing data and covariates

Covariates	N Missing (%)	Processing Notes
Body Mass Index	33 (0.10)	Imaging visit answers, backfilled with baseline visit answers
Diastolic blood pressure	479 (1.49)	
Systolic blood pressure	479 (1.49)	
Waist circumference	1 (3.12 x 10^{-3})	
Hip circumference	1 (3.12 x 10^{-3})	
Income	2556 (9.03)	
Stress, illness, bereavement	65 (0.20)	
Diabetes	38 (0.12)	
Cancer	53 (0.17)	
Vascular/heart problems	30 (0.09)	
Other diagnosis	400 (1.25)	
Non-vigorous physical activity	770 (2.40)	
Vigorous physical activity	533 (1.66)	
Age completed full-time education	126 (0.39)	Variable created from age completed full time education (Field ID 845) and educational qualification (Field ID 6138)
Weekly dose of alcohol	4124 (12.85)	Variable created from dose of different types of alcohol (Field ID 1558, 4407, 4418, 4429, 4451, 4462)
Age	0	Imaging visit answers, not backfilled
Sex	0	

Total N = 32,094 for all covariates

Supplementary table 4.5. Demographic, smoking and health related variables compared between daily smoked, occasionally smoked, and never smoked population

	Daily smoked (N = 8,906)	**Smoked occasionally (N = 4,907)**	**Never smoked (N = 23,188)**
	Mean ± SD	**Mean ± SD**	**Mean ± SD**
Age	65.14 ± 7.53	63.74 ± 7.67	63.21 ± 7.65
Sex (n females, %)			
Female (n, %)	3,967 (44.5)	2,407 (49.1)	13,049 (56.3)
Male (n, %)	4,939 (55.5)	2,500 (50.9)	10,139 (43.7)
Income (£)			
Less than 18,000 (n, %)	1,283 (14.4)	575 (11.7)	2,692 (11.6)
18,000 to 30,999 (n, %)	2,699 (30.3)	1,249 (25.5)	6,057 (26.1)
31,000 to 51,999 (n, %)	2,646 (29.7)	1,508 (30.7)	7,039 (30.4)
52,000 to 100,000 (n, %)	1,790 (20.1)	1,157 (23.8)	5,579 (24.1)
Greater than 100,000 (n, %)	488 (5.5)	418 (8.5)	1,821 (7.9)
Age completed full time education	19.20 ±3.53	19.81 ±3.46	20.10 ±3.33
Diabetes			
Yes (n, %)	353 (4.0)	111 (2.3)	468 (2.0)
Cancer			
Yes (n, %)	552 (6.2)	282 (5.7)	1,182 (5.1)
Vascular/heart problems			
Heart attack, angina (n, %)	2,292 (25.7)	1,015 (20.7)	4,353 (18.8)
Other health conditions*			
Yes (n, %)	1,595 (17.9)	715 (14.6)	3,266 (14.1)
Stress, illness, bereavement			
Illness, injury, bereavement, stress (n, %)	3,819 (42.9)	2,022 (41.2)	9,451 (40.8)
None of the above (n, %)	5,087 (57.1)	2,885 (58.8)	13,737 (59.2)
Body Mass Index	27.29 ± 4.28	26.51 ± 3.99	26.25 ± 4.20
Waist/hip ratio	0.89±0.09	0.87±0.08	0.86±0.09
Systolic blood pressure (mmHg)	141.38 ±20.06	139.77±19.67	139.65±19.64
Diastolic blood pressure (mmHg)	79.38±10.59	79.29±10.58	79.31±10.68
Weekly drinks of alcohol	12.66 ± 10.86	11.28 ± 9.02	8.21±7.73
Non-vigorous physical activity**	3.44±2.30	3.69±2.24	3.43±2.26
Vigorous physical activity**	1.81±1.86	2.05±1.89	1.86±1.81

*Answer to the question: Has a doctor ever told you that you have had any other serious medical conditions or disabilities?

**Number of days/week of non-vigorous or vigorous physical activity 10+ minutes

Supplementary table 4.6. Demographic, smoking and health related variables compared between total UK Biobank sample and our study dataset (N = 32,094)

	UKB total* (N = 502,366)	**Imaging subset (N = 32,094)**
	Mean ± SD	**Mean ± SD**
Age	56.53 ± 8.10	54.81 ± 7.51
Sex (n females, %)		
Female (n, %)	273,298 (54.4)	17,016 (53.0)
Male (n, %)	229,068 (45.6)	15,078 (47.0)
Income (£)		
Less than 18,000 (n, %)	97,176 (22.9)	3,975 (12.4)
18,000 to 30,999 (n, %)	108,140 (25.4)	8,756 (27.3)
31,000 to 51,999 (n, %)	110,746 (26.0)	9,685 (30.2)
52,000 to 100,000 (n, %)	86,243 (20.3)	7,369 (22.9)
Greater than 100,000 (n, %)	22,923 (5.4)	2,309 (7.2)
Age completed full time education	16.73 ±2.34	19.85 ±3.41
Diabetes		
Yes (n, %)	26,394 (5.3)	821 (2.6)
Cancer		
Yes (n, %)	38,607 (7.7)	1,734 (5.4)
Vascular/heart problems		
Heart attack, angina (n, %)	171,175 (32.7)	6,645 (20.7)
Other health conditions**		
Yes (n, %)	99,980 (19.9)	4,861 (15.1)
Stress, illness, bereavement		
Illness, injury, bereavement, stress (n, %)	295,001 (52.0)	13,270 (41.3)
None of the above (n, %)	272,234 (48.0)	18,824 (58.7)
Body Mass Index	27.43 ±4.80	26.54 ±4.25
Waist/hip ratio	0.88±0.10	0.87±0.09
Systolic blood pressure (mmHg)	138.18 ±19.40	140.13±19.77
Diastolic blood pressure (mmHg)	81.78±10.54	79.33±10.65
Weekly drinks of alcohol	9.46 ± 10.24	9.44±8.94
Non-vigorous physical activity***	3.68±2.32	3.43±2.27
Vigorous physical activity***	1.83±1.96	1.85±1.83

*Data mainly acquired from UK Biobank data descriptions (data-field)

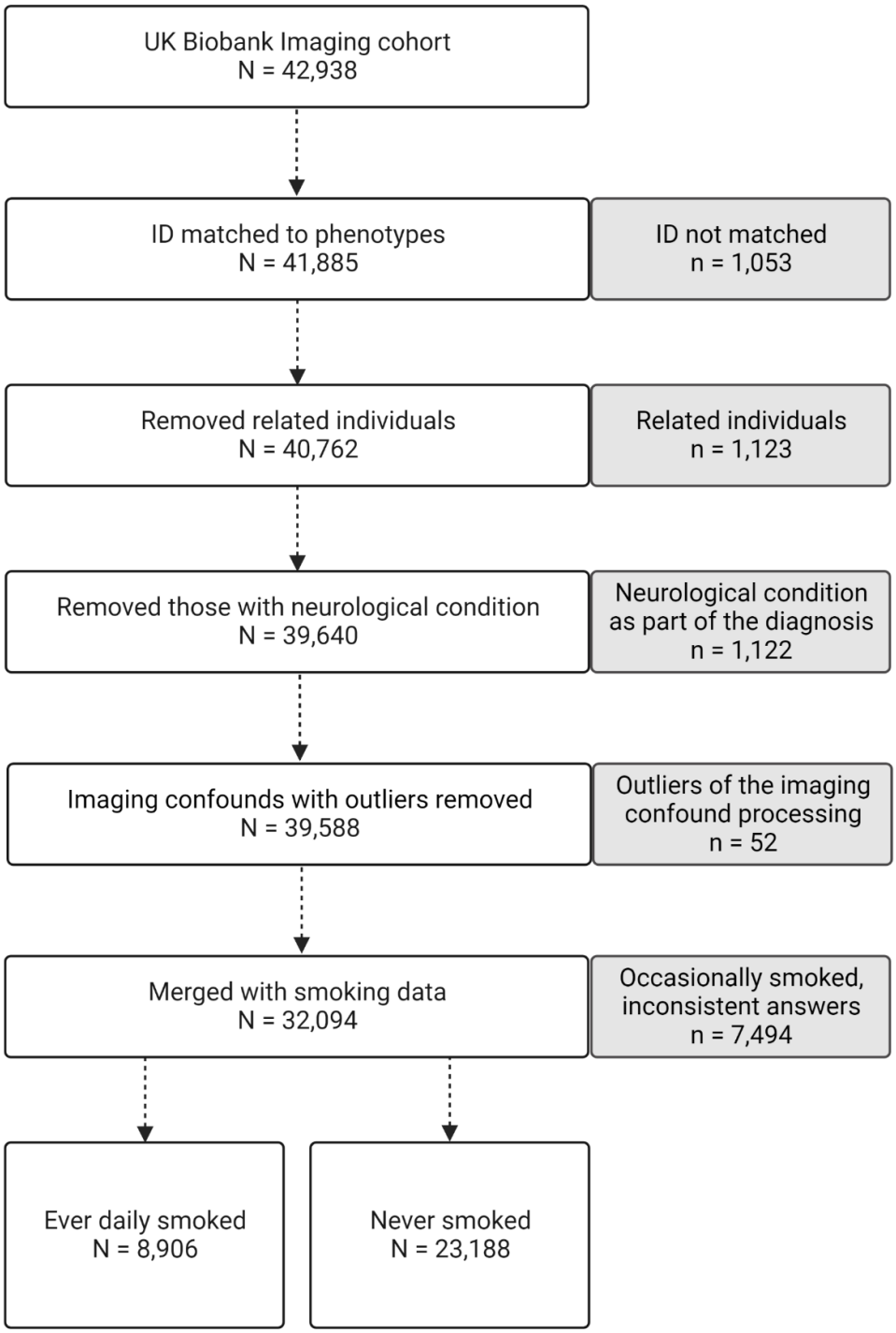
**Answer to the question: Has a doctor ever told you that you have had any other serious medical conditions or disabilities?

***Number of days/week of non-vigorous or vigorous physical activity 10+ minutes

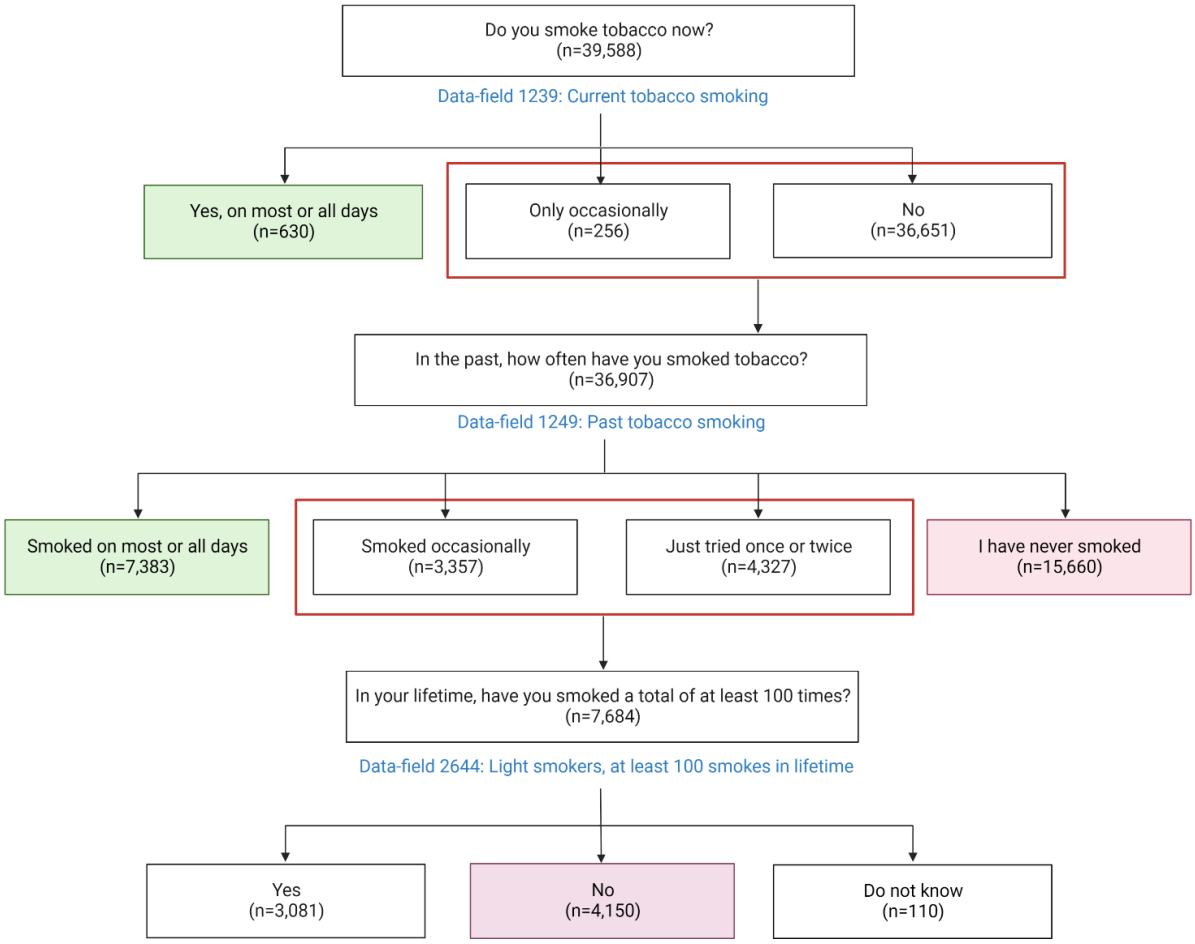
We excluded “do not know” and “prefer not to answer” from the sample count. Some participants also didn’t give consent to answer certain questions.

Age is from the first recruitment (2006-2010), not the imaging recruitment (2014+). Baseline answers are all from the first recruitment (2006-2010).

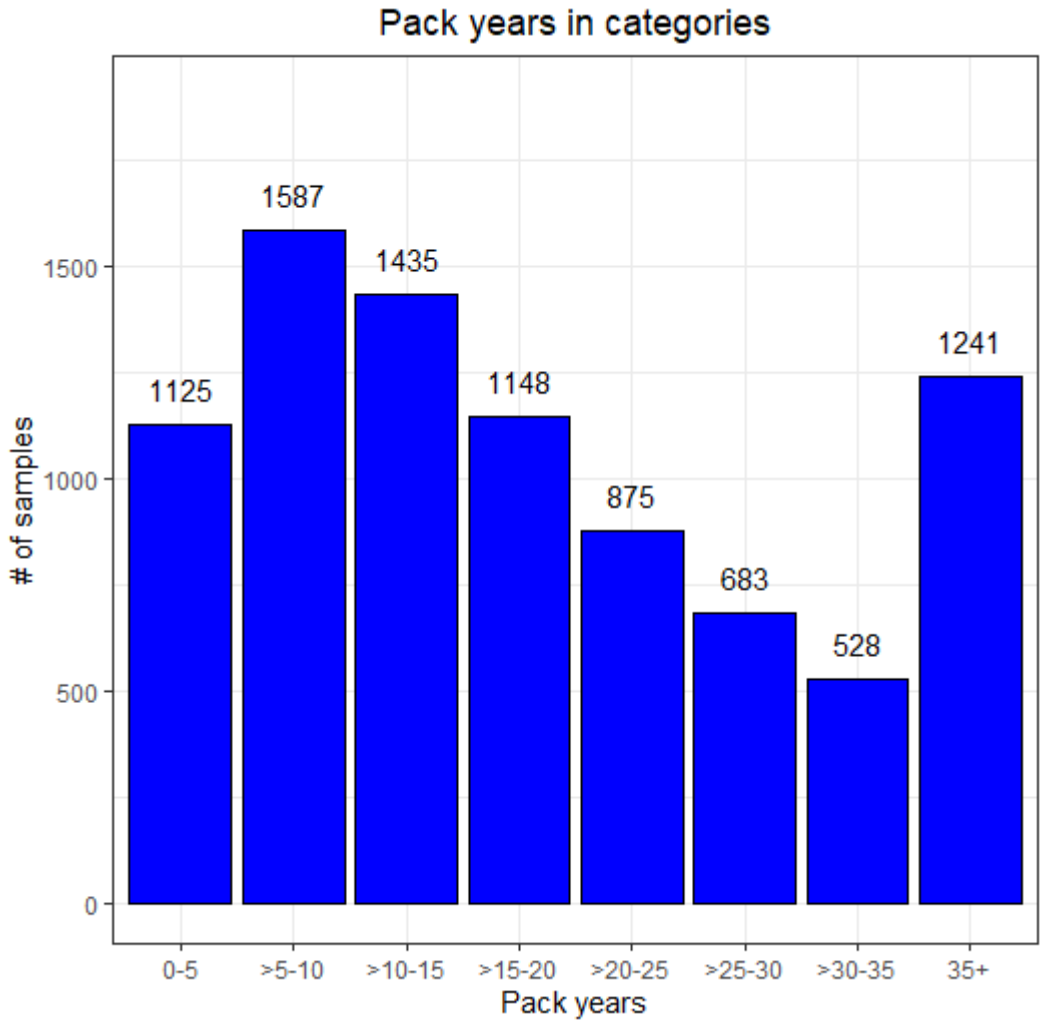
4.13. Supplemental Figures



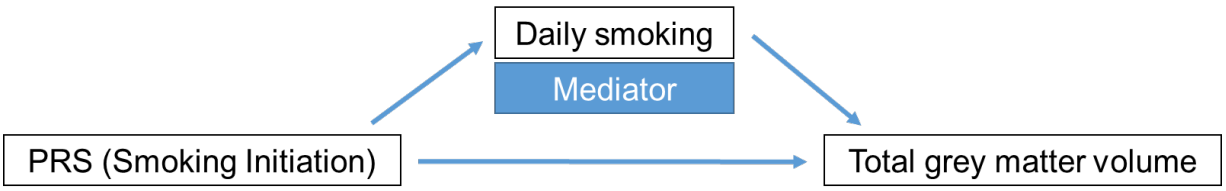
Supplementary figure 4.1. Consort chart of sample processing. Relatedness was from UK Biobank kinship file (ukb48123_kinship.txt provided from UK Biobank), which provides all pairs related up to third degree. We detected all the related pairs in our dataset and broke the pairs by removing one participant from each pair.



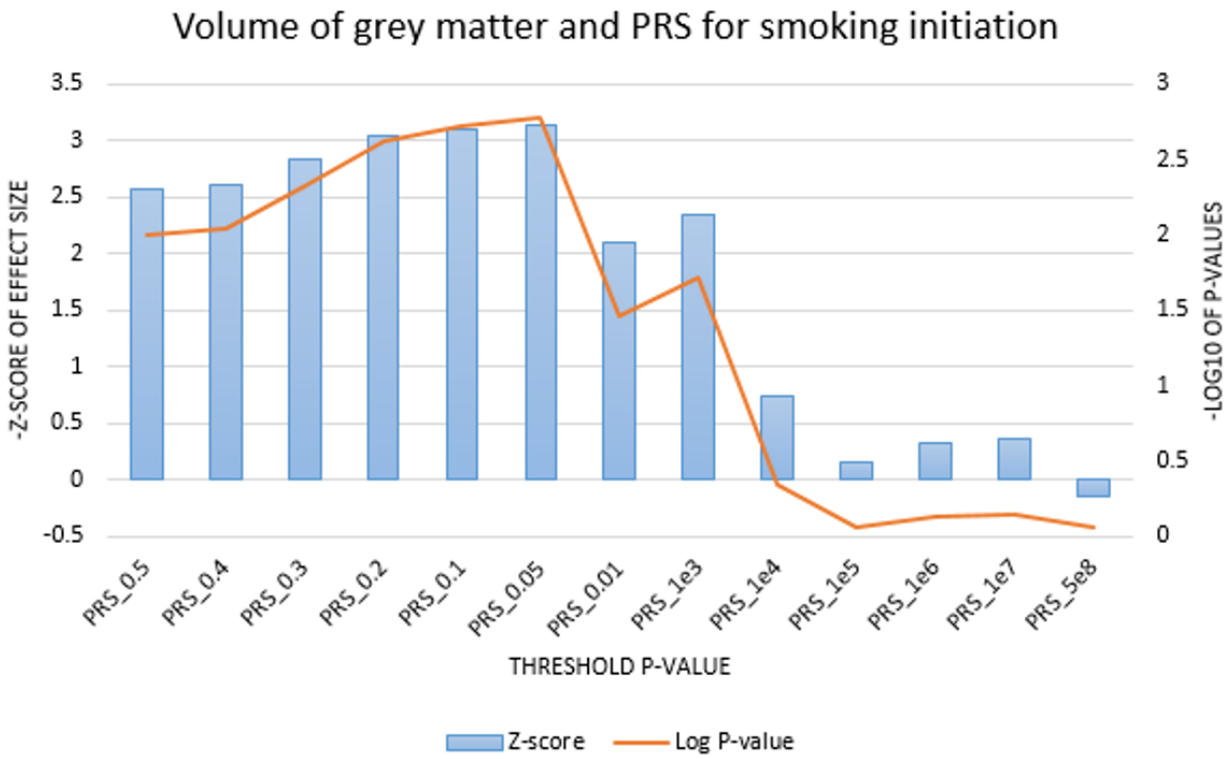
Supplementary figure 4.2. Smoking status extracted from the final subset of touchscreen questionnaire. Ever daily smoked is defined by green (Current daily smoking and Former daily smoking), and never smoked is defined by red (Never previously smoked and smoked less than 100 cigarettes in lifetime). *Note that we excluded the "Prefer not to answer" from the chart*



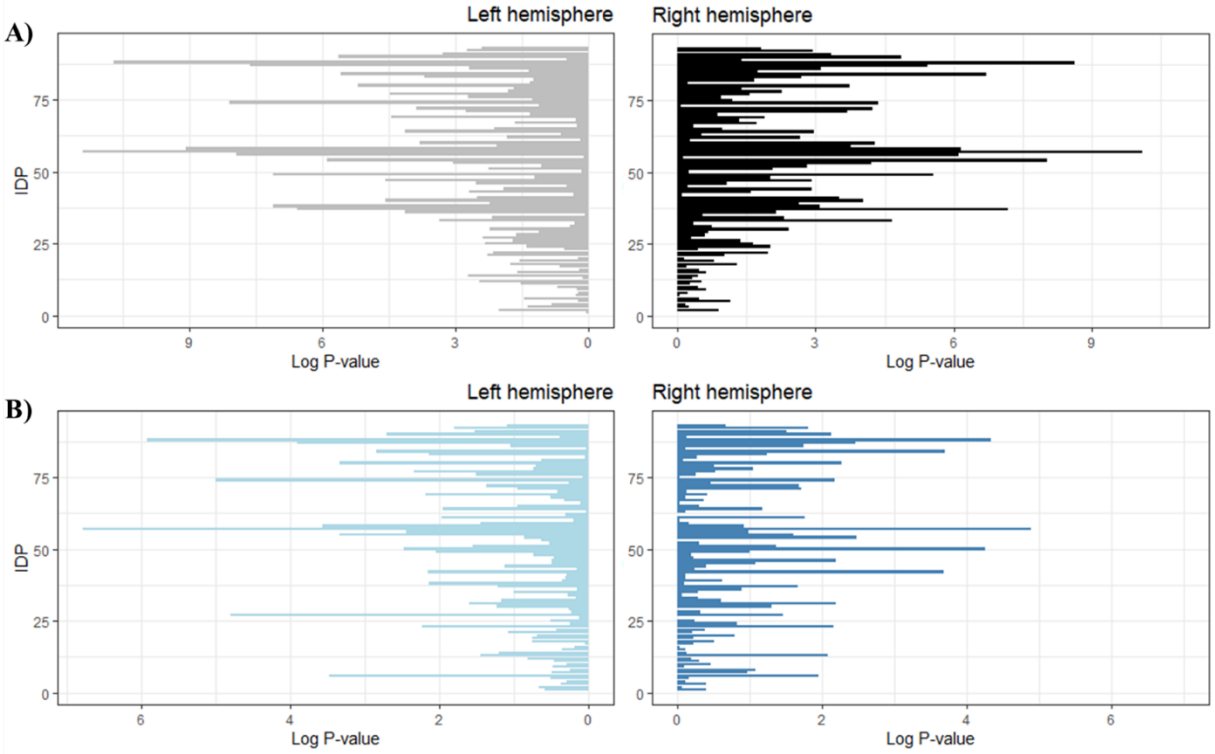
Supplementary figure 4.3. Pack year distribution in categories



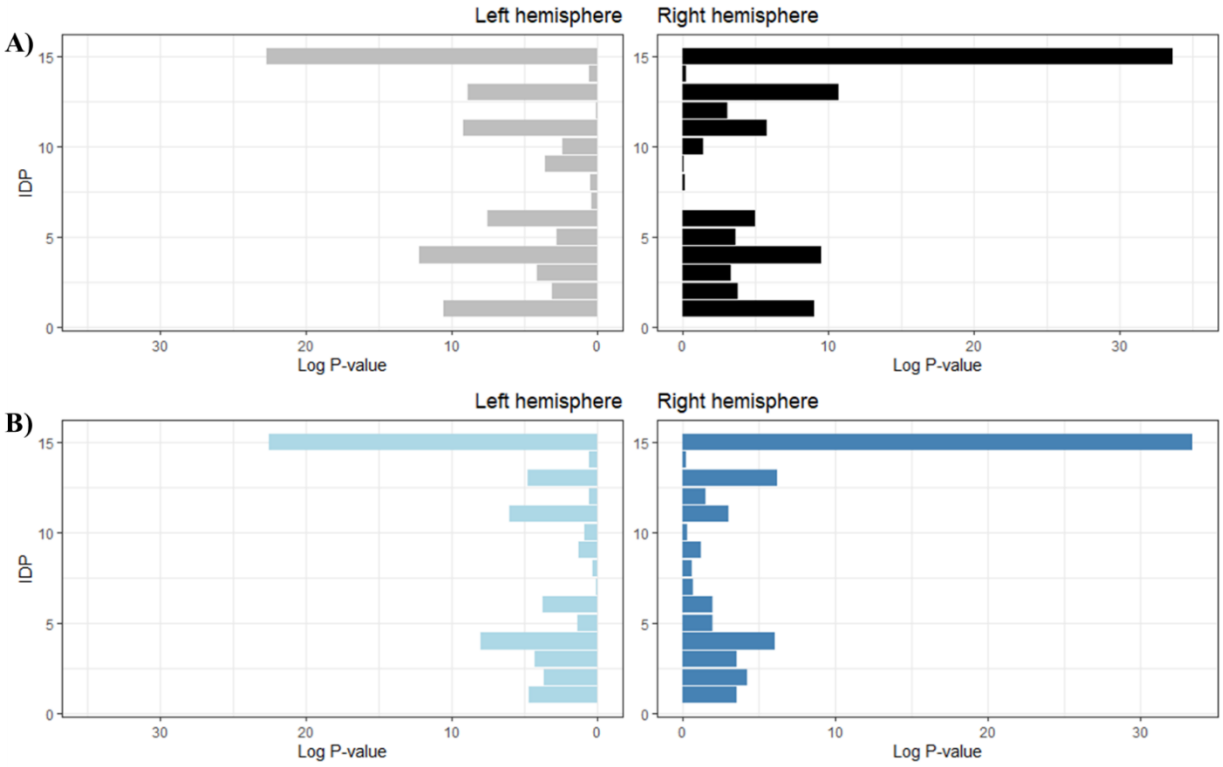
Supplementary figure 4.4. Model for Mediation analysis. Polygenic risk score (PRS) for smoking initiation is strongly associated with total grey matter volume through mediator (Daily smoking). Any statistical significance of the direct association between PRS and total grey matter volume disappears when the mediator is added to the model.



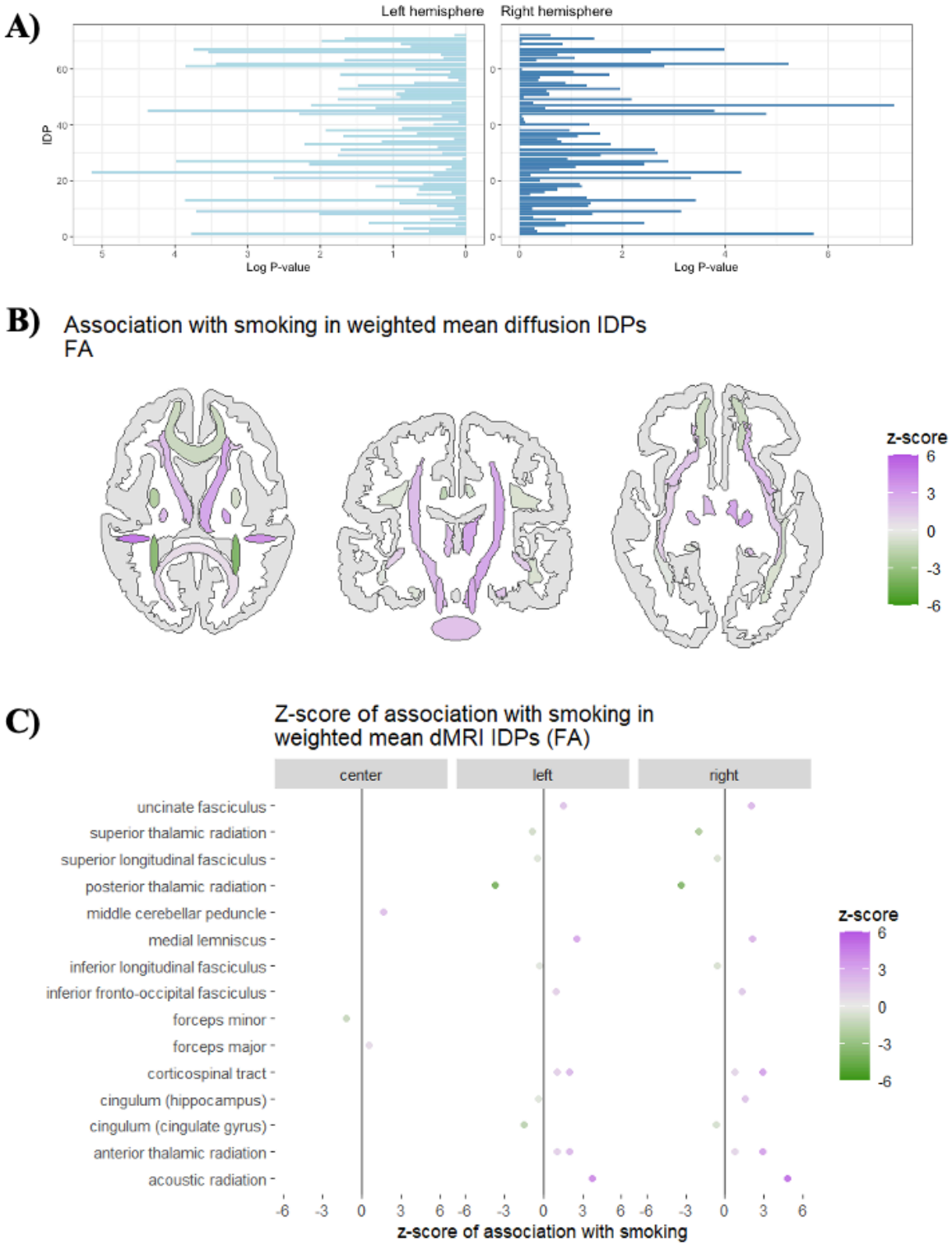
Supplementary figure 4.5. Different PRS thresholds with volume of grey matter (effect size and p-value). Z-score is the –z-score of the effect size. Log P-value is –log10 of P-value.



Supplementary figure 4.6. Log P-value of Freesurfer DKT measures in left and right hemisphere. We only included IDPs with left and right hemisphere values. A) Unadjusted DKT measures (mm3), B) Adjusted DKT measures (normalized)



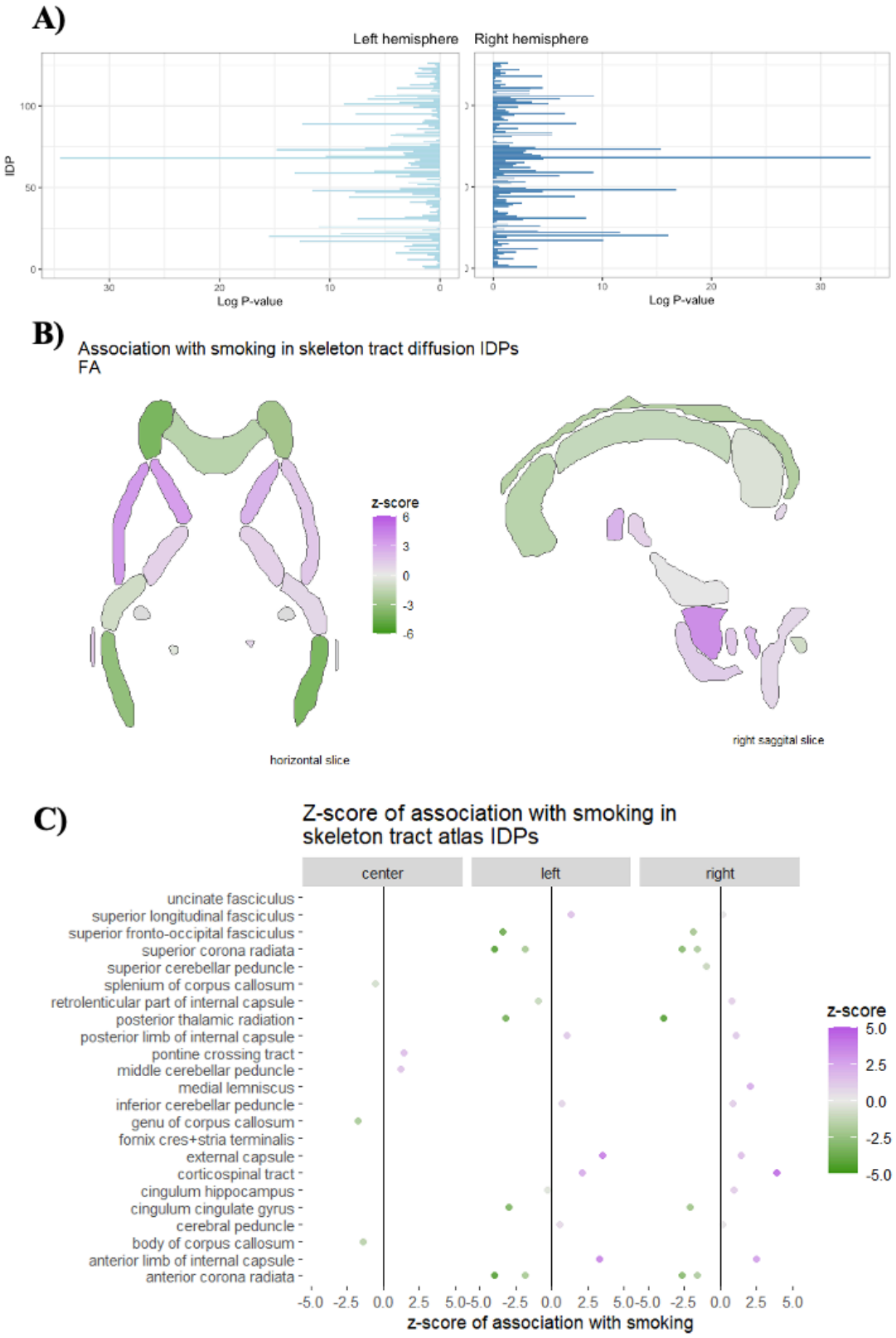
Supplementary figure 4.7. Log P-value of Freesurfer ASEG measures in left and right hemisphere. We only included IDPs with left and right hemisphere values. A) Unadjusted DKT measures (mm3), B) Adjusted DKT measures (normalized).



Supplementary Figure 4.8. Diffusion skeleton measures associated with daily smoking.

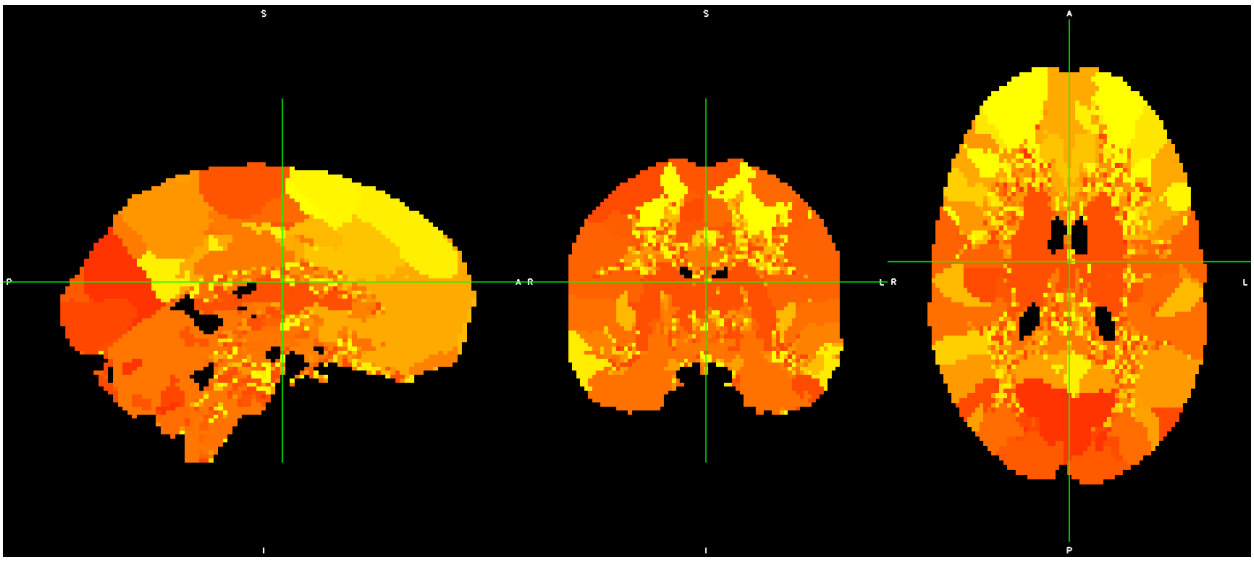
A) Log P-value of diffusion skeleton measures in left and right hemisphere. We only included IDPs with left and right hemisphere values. Adjusted for total measures. B) Z-scores of the 288

diffusion skeleton FA measures. Purple is positive association with daily smoking, while green is negative association. C) Names of the diffusion skeleton measures and their corresponding z-scores.

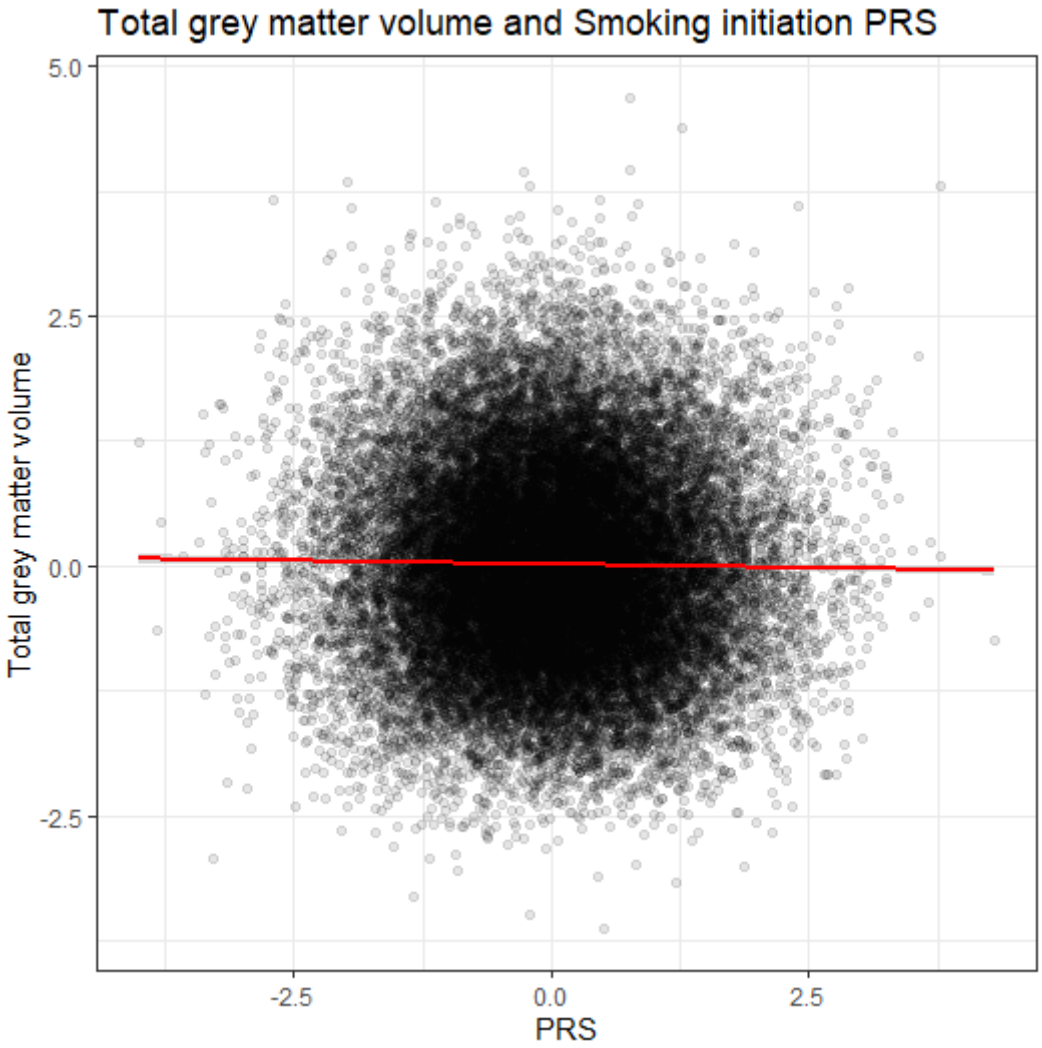


Supplementary Figure 4.9. Diffusion tract measures associated with daily smoking.

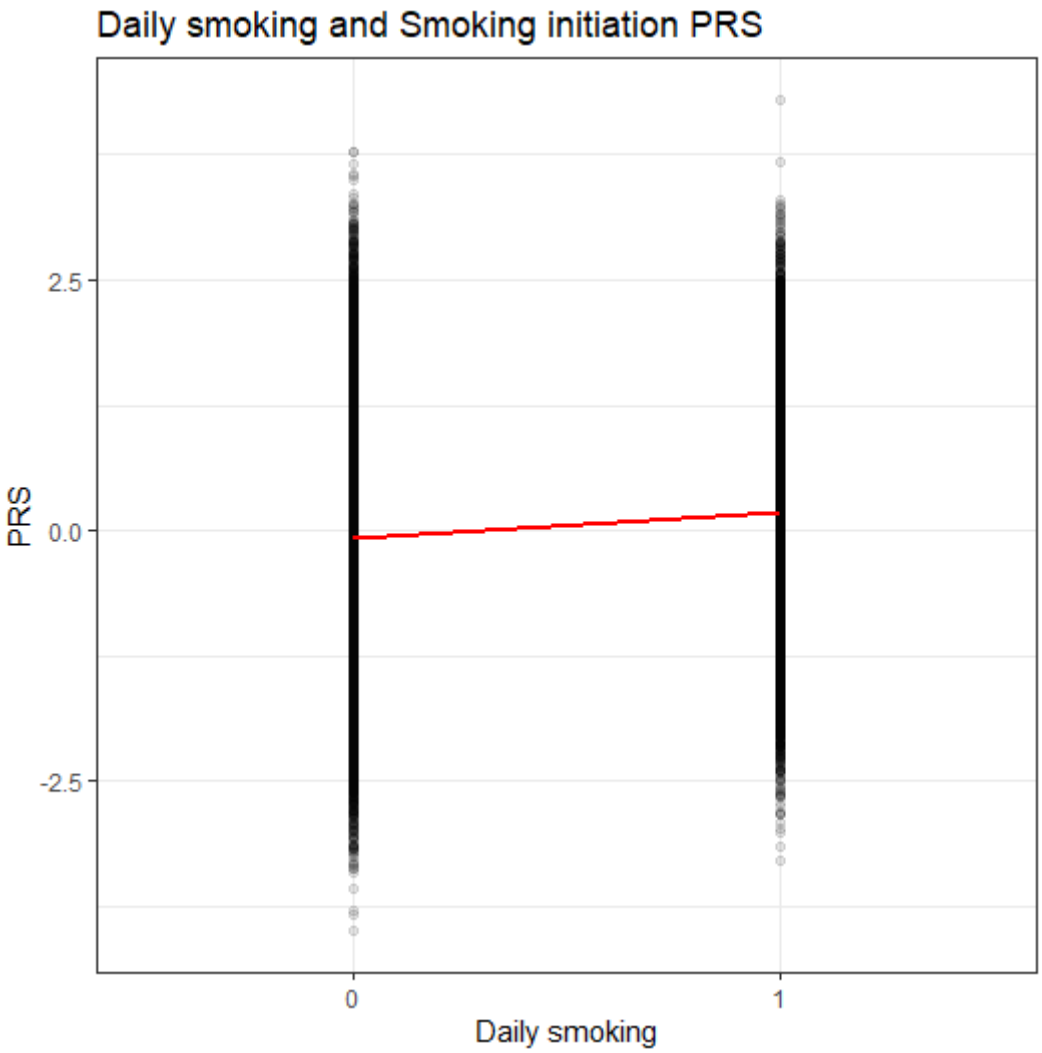
A) Log P-value of diffusion tract measures in left and right hemisphere. We only included IDPs with left and right hemisphere values. Adjusted for total measures. B) Z-scores of the 162 diffusion skeleton FA measures. Purple is positive association with daily smoking, while green is negative association. Right: horizontal slice of the brain, Left: Right saggital slice of the brain. C) Names of the diffusion skeleton measures and their corresponding z-scores.



Supplementary figure 4.10. Resting-functional MRI measure modestly associated with daily smoking. Yellow indicates strong, and red indicates weak association.



Supplementary figure 4.11. Total grey matter volume and Smoking initiation PRS. Scatter plot for PRS with grey matter volume. The fitted line is from linear regression results. Each dot represents an individual. PRS and total grey matter volume are normalized.



Supplementary figure 4.12. Daily smoking and smoking initiation PRS. Scatterplot for PRS and smoking. Since ever daily smoked has two groups (ever and never daily smoked), the line is the difference between the means of two groups. PRS is normalized.

Chapter 5. Conclusion

5.1 Summary of the Dissertation

Overall, this dissertation illustrates the potential utility of polygenic risk scores in aiding smoking cessation in individuals and shows the brain changes associated with smoking behavior. First, I illustrate in Chapter 2 that polygenic risk scores, especially PRS for the later age of smoking initiation, and a combination of PRS for multiple smoking behaviors, are associated with smoking cessation outcomes in two clinical trials. In Chapter 3, I then conducted a study of PRS for smoking cessation in a larger general population sample. By testing the association between the PRS and the age of smoking cessation, I identify that those with a higher genetic risk for persistent smoking had a later age of smoking cessation compared to those with a lower genetic risk. This finding has clinical implications because we can identify those who will more likely smoke for a longer period of time and thus may benefit for more intensive smoking cessation interventions. In Chapter 4, I showed the significant association of smoking behavior in the brain structure and function, while also proving that the genetic risk for smoking is not associated with brain volume. These findings support the adverse effects of cigarette smoking on the brain, though I recognize that there are likely brain structure and function changes that predispose one to smoke and smoke more heavily. These findings add insight into the importance of targeted treatment for smoking cessation aided by genetic risk scores, to prevent the possible negative impact on the individual's brain.

5.2 Future Directions

5.2.1. UK Biobank longitudinal dataset

The direction of effect between the brain and the smoking behavior is best addressed with the aid of a longitudinal dataset. Currently, the UK Biobank provides the repeat scan for a few thousand participants in the imaging cohort, and this sample size is continuously growing. For future studies, it will be worthwhile to identify the individuals who had daily smoked during the primary brain scan but stopped 5 years later, for the repeat scan. Examining the brain-imaging measures of these individuals will more definitively show the changes in the brain caused by the smoking behavior. Within the timeline of this dissertation, we could only identify less than a hundred such individuals who changed their smoking behavior, and so we were unable to undertake this analysis.

5.2.2. The Adolescent Brain Cognitive Development (ABCD) study

The prospective development data will also provide insight into the complex interplay between the brain and behavior. The Adolescent Brain Cognitive Development study (ABCD) is the largest neuroimaging study of brain development in the world, and the study will help determine the direction of effect between daily smoking and brain structure and function [1]. Future studies using this developmental dataset may identify that certain parts of the brain are negatively impacted by daily smoking, and other parts may be predisposing factors to smoking.

5.2.3. Ancestry-weighted PRS

The ongoing issue the genetic study faces is the lack of representation of diverse populations in genetic datasets and tools. The polygenic risk score effectively predicts the corresponding behavior in those with European ancestry but underperforms in other populations [2]. It is

important to start developing tools that are effective across all populations, and one of those is ancestry weighted PRS. This PRS, weighted using ancestral principal components (PCs) can be an effective prediction tool that can be used in precision medicine for the general population with diverse genetic ancestry. Future studies may utilize the robustly developed tools [3] to create ancestry PRS and validate its use in clinical settings.

5.2.4. Brain aging and targeted treatments

Many studies identified the association between substance use and dementia [4]. Recently, there was also an increase in the study of brain aging and substance use [5, 6]. It is known that using substances such as tobacco or alcohol prematurely age an individual's brain, thus amplifying the risk of brain conditions connected to brain volume loss such as Alzheimer's Disease. Developing a brain-aging model to identify the extent to which the brain ages as a result of behaviors such as smoking and other modifiable risk factors may have important public health implications.

Public Health Implications

Tobacco use has substantial public health implications in the US and worldwide. The development of prevention strategies and treatment plans that consider both genetic and environmental factors, as well as the public health message that the brain structure and function are negatively associated with daily smoking, may lead to more effective approaches for smoking control and cessation, ultimately reducing the burden of smoking-related disease and improving overall public health outcomes.

5.3. References

1. Casey BJ, Cannonier T, Conley MI, Cohen AO, Barch DM, Heitzeg MM, et al. (2018): The adolescent brain cognitive development (ABCD) study: imaging acquisition across 21 sites. *Developmental cognitive neuroscience*. 32:43-54.

2. Breedon JR, Marshall CR, Giovannoni G, van Heel DA, Dobson R, Jacobs BM (2023): Polygenic risk score prediction of multiple sclerosis in individuals of South Asian ancestry. *Brain Commun*. 5:fcad041.

3. Patel AP, Wang M, Ruan Y, Koyama S, Clarke SL, Yang X, et al. (2023): A multi-ancestry polygenic risk score improves risk prediction for coronary artery disease. *Nature Medicine*. 29:1793-1803.

4. Zhong G, Wang Y, Zhang Y, Guo JJ, Zhao Y (2015): Smoking is associated with an increased risk of dementia: a meta-analysis of prospective cohort studies with investigation of potential effect modifiers. *PLoS One*. 10:e0118333.

5. Daviet R, Aydogan G, Jagannathan K, Spilka N, Koellinger PD, Kranzler HR, et al. (2022): Associations between alcohol consumption and gray and white matter volumes in the UK Biobank. *Nature Communications*. 13:1175.

6. Ho YS, Yang X, Yeung SC, Chiu K, Lau CF, Tsang AW, et al. (2012): Cigarette smoking accelerated brain aging and induced pre-Alzheimer-like neuropathology in rats. *PLoS One*. 7:e36752.