2017

A Multi-Method Exploration of the Genetic and Environmental Risks Contributing to Tobacco Use Behaviors in Young Adulthood

Elizabeth K. Do
Virginia Commonwealth University

A Multi-Method Exploration of the Genetic and Environmental Risks Contributing to Tobacco Use Behaviors in Young Adulthood

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at Virginia Commonwealth University.

by

Elizabeth Kieuvan Do, B.A., M.P.H.
C. Kenneth and Diane Wright Center for Clinical and Translational Research;
Virginia Institute for Psychiatric and Behavioral Genetics;
Virginia Commonwealth University

Director: Hermine H. Maes, Ph.D.

Associate Professor, Departments of Psychiatry and Human and Molecular Genetics and Massey Cancer Center

Virginia Commonwealth University

Richmond, Virginia

May 2017

ACKNOWLEDGEMENTS

In addition to the funding mechanisms that have financially supported my dissertation work (R25DA026119-08, PI: Neale, Research Education in Statistical Genetics of Substance Abuse; R01DA025109-04, PI: Maes, Developmental Genetic Epidemiology of Smoking; UL1TR000058, and the CCTR Endowment Fund of Virginia Commonwealth University) and all the wonderful participants, without whom there would be no data to analyze, I would like to thank several people, without whose love and support this would not have been made possible!

First and foremost, I would like to thank my family (mom, dad, Quynh, Theresa, and Marian) for their undying support throughout this process. They were and continue to be witnesses to a multitude of my ever-changing and cycling emotions, and take all of it in stride. In addition to keeping me grounded, they encourage me to reach for the stars and to never give up. I would also like to thank my doctoral mentor/advisor/and friend, Dr. Hermine H. Maes for her continual support and mentorship. She encourages me to stay curious about the world, and to work hard to accomplish what often seems like impossible undertakings. I am deeply appreciative of all the opportunities she has encouraged me to pursue. I would also like to thank my doctoral dissertation committee (Drs. Hermine Maes, Briana Mezuk Ratliff, B. Todd Webb, Timothy York, Nathan Gillespie, and Michael Neale) for their support during the dissertation writing process – not only were they available to help me think through my research projects more critically and challenge me, but also provided guidance when I ran into roadblocks. In addition to my dissertation committee, I would like to extend thanks to my pre-doctoral qualification committee (Drs. Roxann Nay-Roberson, Nathan Gillespie, Brien Riley, Timothy York, Danielle Dick, Hermine Maes), who helped me reorganize a couple of the manuscripts included in this dissertation, which provided the basis of much of my dissertation project.

Outside of these committees, I would like to thank Drs. Briana Mezuk Ratliff and Elizabeth Prom-Wormley, for keeping the door open for me when I needed it to talk about life and science. Through many a meeting, these women scientists helped me work through conceptual problems with my research, while also providing me with advice and guidance on next steps in my career. I'd also like to give special thanks to Drs. Roseann Peterson, Alexis Edwards, and Bradley T. Webb for their assistance in helping me to understand genetic analyses, which I had never conducted previously. I am appreciative of their patience in helping me get over the steep learning curve. I am eternally grateful to the members of my cohort (Arden Moscati, Jeanne Savage, and Cassie Overstreet) for being the best officemates I could ever ask for, and for the hours of scientific musing and mental and emotional support that they've provided me. I'd also like to thank Megan Cooke, one of the first students of the PBSG program, for blazing the trail ahead for the rest of us students, as well as for all the hours she has spent with me working at coffee shops, solving problems with statistical modeling. Mackenzie Lind has also been quite helpful with day-to-day problem solving. I am lucky to be a part of a growing network of PBSG students, who I am happy to call my friends, and who have also helped me think through many of the problems addressed in this dissertation. They also make work a whole lot more fun, since we are all very supportive of one another and our goals!

Without the love and support of all these individuals and friends, who have provided me with an immense amount of support but there is not space to thank personally, I would not have been able to climb the figurative mountains required to complete this dissertation thesis.

Thank you all, from the bottom of my heart!

TABLE OF CONTENTS

*Indicates manuscripts that have already been accepted for publication. These manuscripts are included in this dissertation in full, along with its original citation.

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

5-HTTLPR..... Serotonin Transporter Gene

A ..... Additive Genetic Factors

AA..... African American

AAF..... African American Female

AAM..... African American Male

AIC..... Akaike Information Criterion

ANKK1 ..... Ankyrin Repeat and Kinase Domain Containing 1

C ..... Shared Environmental Factors

CAPA..... Child and Adolescent Psychiatric Assessment

CCC..... Common-Causal-Contingent

CPD ..... Cigarettes Per Day

DAT1 ..... Dopamine Transporter Gene

DZ..... Dizygotic

DZF..... Dizygotic Female

DZM..... Dizygotic Male

E ..... Environmental Factors

FDA ..... Food and Drug Administration

FTND ..... Fagerström Test for Nicotine Dependence

GCTA..... Genome-Wide Complex Trait Analysis

GRM ..... Genetic Relatedness Matrix

GRS..... Genotypic Risk Score

GWAS..... Genome Wide Association Study

GxE..... Gene-Environment Interaction

HBM..... Health Belief Model

LD ..... Linkage Disequilibrium

LRC ..... Likelihood Ratio Chi-Square

MAF ..... Minor Allele Frequency

MZ ..... Monozygotic

MZF ..... Monozygotic Female

MZM ..... Monozygotic Male

nAChR ..... Nicotinic Acetylcholine Receptor

OR ..... Odds Ratio

PC..... Principal Component

PGC..... Psychiatric Genomics Consortium

PRS ..... Polygenic Risk Score

QC ..... Quality Control

Q-Q..... Quantile-Quantile

REDCap..... Research Electronic Data Capture

REML..... Restricted Maximum Likelihood

RGE..... Gene-Environment Correlation

RMSEA..... Root Mean Square Error of Approximation

S4S..... Spit for Science: The VCU Student Survey

SI ..... Smoking Initiation

SLE..... Stressful Life Event

SNP ..... Single Nucleotide Polymorphism

TFT ..... Time to First Tobacco (Use After Waking)

TSA..... Transitions to Substance Abuse

US..... United States

VNTR..... Variable-Number-of-Tandem-Repeats

VTSABD ..... Virginia Twin Studies of Adolescent and Behavioral Development

WCA ..... White/Caucasian American

WCF ..... White/Caucasian Female

WCM..... White/Caucasian Male

YAFU ..... Young Adult Follow-Up

ABSTRACT

A MULTI-METHOD EXPLORATION OF THE GENETIC AND ENVIRONMENTAL RISKS CONTRIBUTING TO TOBACCO USE BEHAVIORS IN YOUNG ADULTHOOD

By Elizabeth Kieuvan Do, B.A., M.P.H.

A dissertation submitted in partial fulfillment of the requirements for the degree Doctor of Philosophy at Virginia Commonwealth University.

Virginia Commonwealth University, 2017.

Major Director: Hermine H. Maes, Ph.D.

Associate Professor, Departments of Psychiatry and Human and Molecular Genetics and Massey Cancer Center

Tobacco use remains the leading preventable cause of morbidity and mortality in both the United States and worldwide. Twin and family studies have demonstrated that both genetic and environmental factors are important contributors to tobacco use behaviors. Understanding how genes, the environment, and their interactions is critical to the development of public health interventions that focus on the reduction of tobacco-related morbidity and mortality. However, few studies have examined the transition from adolescent to young adulthood – the time when many individuals are experimenting with and developing patterns of tobacco use. This thesis seeks to provide a comprehensive set of studies looking at risk for tobacco use behaviors and nicotine dependence using samples of young adults. The first aim is to examine the joint contributions of genetic liability and environmental contexts on tobacco use in adolescence and young adulthood

using classical twin study methodologies. The second goal is to identify genetic variants and quantifying genetic risk for tobacco use in young adulthood and examining their interaction with environmental context across development. Accordingly, the thesis is divided up into the following sections: i) reviews of existing literature on genes, environment, and tobacco use; ii) twin studies of genetic and environmental influences on tobacco use behavior phenotypes; iii) prevalence, correlates, and predictors of tobacco use behaviors; iv) genetic analyses of tobacco use behaviors; v) a commentary on the emergence of alternative nicotine delivery systems and its public health impacts; and vi) plans for an internet-based educational intervention seeking to reduce tobacco use (and nicotine dependence) by providing students attending university with information on genetic and environmental risk factors for nicotine dependence.

CHAPTER ONE: GLOBAL INTRODUCTION

Elizabeth K. Do

Introduction

Tobacco use remains the leading preventable cause of morbidity and mortality in the United States (US), as well as worldwide. In 2015, an estimated 36.5 million adults in the US currently smoked cigarettes. Cigarette smoking accounts for an estimated $96 billion in direct medical costs and $97 billion in lost productivity annually[3]. Although there has been a slight overall decline in current smoking prevalence between 2005-2011, especially among adults aged 18 to 24 years (from 24.4% to 18.9%), and a decline in the prevalence of cigarette smoking[4], there has been an increase in the use of emerging tobacco products in recent years[2]. The National Adult Tobacco Survey, the Food and Drug Administration (FDA) analyzed data between 2013 and 2014 and found that young adults aged 18-24 reported the highest prevalence of use of emerging tobacco products, such as water pipes/hookah and e-cigarettes[5].

Specifically, within the United States, between 13.6% and 32.1% of adults report trying one or more other tobacco products[6,7]. A more recent study indicates that nearly half of current adolescent and young adult tobacco users engage in dual and multiple tobacco product use, many of which fall outside of current FDA regulatory authority[8]. Given the increased morbidity and mortality associated with tobacco use, disrupting the transition to regular use of tobacco products among young adults is likely to result in a number of lives saved and disease prevented at the population level[9].

There exists a widespread assumption within the public health community that tobacco use is largely fixed by age 18, since most cigarette smokers report either having first tried or experimented with smoking by age 18[2]. However, for many young adults, tobacco use patterns continue to change following age 18. For some, the period of young adulthood is marked by the transition from occasional adolescent smoking to regular or established smoking[10]. Many users who first try smoking at younger ages do not become "regular" or daily users until much later, typically between the ages of 20-21[11]. Furthermore, studies suggest that both the proportion and intensity of smoking rise substantially after the age of 17[12]. The possible escalation and continuation of tobacco use may make it more difficult for users to stop using[13]. Thus, young adulthood is a critical period in the development of tobacco use behavior that needs to be studied.

Given the high prevalence of tobacco use, the addictive nature of nicotine, and high healthcare costs attributed to tobacco-related morbidity and mortality, developing effective methods to aid individuals reduce tobacco use is critical. However, before this is possible, a better understanding of the etiology of tobacco use and addiction, which involves the complex interplay of genetic and environmental factors across development, is needed – especially among those at elevated risk of developing patterns of regular tobacco use.

To date, only a handful of studies have focused on the transition from adolescent to young adulthood – the time at which many individuals are experimenting with and developing patterns of tobacco use. As a means to contribute to the literature, this dissertation seeks to provide a comprehensive set of studies looking at risk for tobacco use behaviors and nicotine dependence using samples of young adults, keeping in mind two aims: (1)

examining the joint contributions of genetic liability and environmental contexts on tobacco use in adolescence and young adulthood using twin methodologies and (2) identifying genetic variants and quantifying genetic risk for tobacco use in young adulthood and examining its interaction with environmental context across development. In efforts to address these aims, the proposal is divided up into a few different sections: reviews of existing literature on genes, environment, and tobacco use; twin studies of genetic and environmental influences on tobacco use behavior phenotypes; prevalence, correlates, and predictors of tobacco use behaviors; molecular genetic analyses of tobacco use behaviors; a commentary on the emergence of alternative nicotine delivery systems and its public health impacts; and, plans for an internet-based educational intervention seeking to reduce tobacco use (and nicotine dependence) by providing students attending university with information on genetic and environmental risk factors for nicotine dependence.

Reviews of Existing Literature on Genes, Environment, and Tobacco Use

Chapter 2 (Narrative Review of Genes, Environment, and Tobacco Use) and chapter 3 (Genotype x Environment Interaction in Smoking Behaviors: A Systematic Review) are reflective of efforts to understand the state of the science on the influences of genes, environments, and their interactions on tobacco use behaviors, and provide a comprehensive foundation for the specific aims of this dissertation. Whereas the narrative review provides a description of gene variants and environmental factors associated with cigarette use, and a broad overview of studies examining gene-environment interaction, the systematic review takes a more technical approach by looking at variability in tobacco use behavior phenotype definitions and methodological approaches across existing

studies of gene by environment interactions. Together, this set of studies suggest the need for more studies of gene by environment interaction, with the caveat that variations in methodological approaches within the existing literature make it difficult to interpret and summarize findings across studies, and that future studies need better strategies for the harmonization and standardization of tobacco use behavior phenotypes.

Twin Studies of Genetic and Environmental Influences on Tobacco Use Behavior Phenotypes

Chapter 4 (Genetic and Environmental Influences on Smoking Behavior across Adolescence and Young Adulthood in the Virginia Twin Study of Adolescent Behavioral Development and the Transitions to Substance Abuse Follow-Up) and chapter 5 (A Twin Study of the Genetic and Environmental Relationship of Stressful Life Events and Smoking Initiation Using the Virginia Twin Studies of Adolescent Behavioral Development) were conducted to investigate the role of genetic and environmental factors contributing to two tobacco use behaviors: smoking initiation and current quantity smoked within the Virginia Twin Studies of Adolescent and Behavioral Development (VTSABD). The study described in chapter 4 was conducted to address the existing gap in the literature regarding the underlying relationship between smoking initiation and current quantity smoked by applying a common causal contingency model. The aims of this study were to determine whether: (1) the genetic and environmental factors contributing to smoking initiation and current quantity smoked are the same, (2) the magnitude of genetic and environmental factor contributions is the same across adolescence and young adulthood, and (3) if qualitative and quantitative differences in the sources of variance between males and females exist. Meanwhile, chapter 5 takes a different approach by

focusing on smoking initiation and its relation to stressful life events. Given that no genetically informed studies of the association between stressful life events and smoking initiation have been conducted, it is unclear whether shared genetic or environmental factors contribute to the covariation between stressful life events and smoking initiation, and whether the covariation structure between these two traits differ from early adolescence to young adulthood and across sex. To address this gap in the literature, the study described in chapter 5 utilizes twin modeling analyses applied to VTSABD data.

Prevalence, Correlates, and Predictors of Tobacco Use Behaviors

Chapter 6 (Prevalence and Correlates of Tobacco Use and Nicotine Delivery Systems among Young Adults in a University Setting), chapter 7 (Initial Experiences with Nicotine and its Association with Recent Use of Tobacco and Nicotine Dependence), and chapter 8 (An Exploration of Sex Differences in Responses to Items of the Fagerström Test for Nicotine Dependence) investigate the prevalence, correlates, and predictors of tobacco use behaviors, with a focus on the following tobacco use behavior phenotypes: ever tobacco use, age of initiation/onset, current use, regular use, the Fagerström Test for Nicotine Dependence (FTND), in addition to two items of the FTND – cigarettes per day and time to first tobacco use after waking. The main objectives of chapter 6 are to: determine how prevalent tobacco use and nicotine delivery systems are among students currently attending university and whether this differs by sex or race/ethnicity, and examine whether tobacco use phenotypes are correlated with other environmental factors, such as parental autonomy granting, parental involvement, and the experience of stressful life events prior to university enrollment.

Examining differences in tobacco use and the negative health outcomes resulting from tobacco use by race/ethnicity is an important understudied area of research. Although previous studies have identified biological predispositions and social determinants associated with traditional cigarette use among adult Black/African Americans not found for White/Caucasian Americans, there are few twin and family studies with a sizeable number of non-White participants to determine whether heterogeneity exists in the heritability of tobacco use behaviors[14].

Chapter 6 provides a background chapter on Spit for Science: The VCU Student Survey (S4S) – the sample that was used for all the molecular genetic analyses of tobacco use behaviors in this dissertation. Chapter 7 leverages data from S4S and the VTSABD to determine whether initial experiences with tobacco use differs by tobacco product used and are associated with recent tobacco use and meeting criteria for nicotine dependence, as well as determine whether any of these associations differ by sex. The study described in chapter 8 was conducted to test sex differences in the response to FTND items within African and White/Caucasian Americans, and to investigate the psychometric properties of FTND items across these groups. Taken together, these chapters are meant to describe phenotypic definitions for subsequent genetic analyses and to identify correlates and predictors of multiple tobacco use behaviors, while also providing information on the prevalence of emerging alternative tobacco products and nicotine delivery systems within an active and on-going study of university students.

Molecular Genetic Analyses of Tobacco Use Behaviors

Chapter 9 (Genetic Analyses of Tobacco Use Behaviors Among an Ethnically Diverse University Sample) and chapter 10 (Polygenic Risk Scores for Tobacco use Behaviors: A Comparison of Methodologies, Applied to a University Sample) utilizes genotypic data from the Spit for Science sample to address the second aim of this dissertation: to identify genetic variants and quantify genetic risk for tobacco use in young adulthood and to examine its interaction with environmental context across development. More specifically, chapter 9 examines several tobacco use behaviors (e.g. initiation, age of onset, current use, regular use, cigarettes per day, time to first tobacco use, and the FTND) among a diverse sample of young adults attending university in efforts to calculate the heritability of tobacco use behaviors within this sample, and to identify genetic variants contributing to tobacco use behaviors in young adulthood by means of conducting genome wide association analyses on each of these phenotypes. Chapter 10 expands on the work conducted in chapter 9 by describing the development polygenic risk scores using two methodological approaches, determining whether these polygenic risk scores are predictive of tobacco use behaviors, and to assess gene-by-environment interactions between these polygenic risk scores and environmental variables (parental autonomy granting, parental involvement, and stressful life events).

Commentary on the Emergence of Alternative Nicotine Delivery Systems and its Public Health Impacts

Chapter 11 (A Moving Target: The Emergence of Alternative Nicotine Delivery Systems and Public Health Impact) is a commentary regarding the growing prevalence of

alternative tobacco products and nicotine delivery systems, policy implications of tobacco product regulation, and what this might mean for future directions of research. It remains unclear whether these alternative tobacco products and nicotine delivery systems will be replacing traditional tobacco products without expanding patterns of nicotine use among adolescents and young adults – the main target of tobacco company advertising and public health harm reduction and prevention efforts. And, as information is collected regarding how much and how these products are being used, the availability and marketing of these products will continue to grow and change with user preferences. It is for this reason that it is important that more attention is given to the relationship between alternative tobacco products, nicotine delivery systems, and traditional cigarette use. More work needs to be done to better understand the increasing complexity of tobacco use among adolescents and young adults to promote effective public health planning and to ensure that the regulation of alternative tobacco products (or lack thereof) does not undermine current anti-tobacco regulatory efforts

Plans for an Internet-Based Educational Intervention Seeking to Reduce Tobacco Use (and Nicotine Dependence)

Chapter 12 (Plans for a Pilot Randomized Control Trial of Internet-Based Educational Intervention for Reduction of Tobacco Use and Nicotine Dependence) contains information on the planning of a feasibility study involving an Internet-based educational intervention examining how providing college students with information on the influence of genes and the environment on tobacco use behaviors and nicotine dependence impacts subsequent patterns of tobacco use behavior. To date, few interventions are specifically aimed at young adult smokers, even though tobacco use is common among

college students, and college is a critical time for experimentation with and development of patterns of tobacco use. The study described in this chapter applies principles of the Health Belief Model (HBM), which assumes health behavior is determined by perceptions of perceived threat, perceived susceptibility and severity, perceived benefits, perceived barriers and the strategies available to decrease its occurrence. By applying the HBM concept and constructs to our intervention, which seeks to reduce tobacco use (and risk for nicotine dependence) in young adult tobacco users, we seek to increase their perceived threat, susceptibility, and severity of nicotine dependence while decreasing their perceived barriers to reducing tobacco use among college student participants by providing knowledge of genetic and environmental risks for nicotine dependence and means to decrease barriers to reducing tobacco use.

Altogether, this set of studies contributes to the literature by providing a better comprehensive understanding of how genes, the environment, and their potential interactions influence many tobacco use behavior phenotypes by keeping in mind two aims: (1) examining the joint contributions of genetic liability and environmental contexts on tobacco use in adolescence and young adulthood using twin methodologies and (2) identifying genetic variants and quantifying genetic risk for tobacco use in young adulthood and examining its interaction with environmental context across development.

CHAPTER 2:
NARRATIVE REVIEW OF GENES, ENVIRONMENT, AND TOBACCO USE [1]

Elizabeth K. Do and Hermine H. Maes

INTRODUCTION

Tobacco use remains the leading cause of preventable death in the United States and results in nearly $170 billion in direct medical care for adults and greater than $156 billion in lost productivity due to premature death and exposure to secondhand smoke[1]. These costs emphasize the need to understand what genes and environments are involved in the establishment of cigarette use behavior[2]. Knowing what genes and environmental risk factors impact cigarette use can help to reduce its prevalence by shaping prevention and intervention efforts. However, to date, many studies on cigarette use have focused solely on genes and environments contributing independently to risk for cigarette use and its health consequences. Fewer studies have investigated the effects of gene-environment interaction (GxE), which can be conceptualized as the difference in the contribution of genetic factors, conditional on environmental exposure[3]. Since cigarette use involves both motivational and reward systems that develop through interactions between genes and the environment, studies of the joint effects of multiple genetic mutations across different environments could be useful in understanding the range of genetic susceptibility to environmental risk factors influencing cigarette use and its health consequences[4]. GxE

[1] This chapter was previously published as an original research article in the *Annals of Medicine.* To cite information from this chapter, please use the following citation: Elizabeth Do & Hermine Maes (2016): Narrative review of genes, environment, and cigarettes, Annals of Medicine, DOI:10.1080/07853890.2016.1177196

studies have been useful in determining whether genetic effects are more or less important under particular environmental conditions[5]. For example, restricting the availability of tobacco has been found to reduce the effect of genes influencing whether individuals initiate and maintain smoking behaviors. Alternatively, under environments where there are fewer restrictions, the importance of the role of genes is expected to increase since individuals are able to express the full range of phenotypes[6], inclusive of nicotine dependence and tobacco-related health conditions such as heart disease and cancer. Quitting cigarette use can effectively reduce the risk of these tobacco-related outcomes for each individual smoker, while also substantially reducing excess health-care utilization and improved labor supply on a larger scale[4]. However, to improve strategies for disease prevention and intervention efforts focused on smoking cessation, a better understanding of genetic, social environment, and individual determinants of risk contributing to cigarette use are needed. In other words, we need to be able to disentangle the etiology of cigarette use and identify the conditions under which genes, the environment, and their interaction impact cigarette use behaviors. Through this narrative review, we seek to integrate twin and molecular genetic studies of GxE in cigarette use. Specifically, this narrative review provides a brief overview of studies investigating genetic and environmental factors influencing cigarette use separately, and then summarizes gene-environment interactions in cigarette use behaviors.

Phenotypic Measures of Cigarette Use

It is important to understand how cigarette use has been measured before getting into details about how we can determine how much of cigarette use is attributed to genes, the environment, and their interactions. The most common phenotypic measures of cigarette

use include: initiation; adolescent smoking; cigarettes per day; regular smoking; nicotine dependence; and smoking cessation. Initiation is usually a self-report measure that is assessed using a yes or no question, such as “Have you ever smoked an entire cigarette?”[7]. Although adolescent smoking is often treated as binary (yes/no) variable, the way in which it is assessed differs across studies. One study may measure adolescent smoking by asking the question, “Have you ever smoked (or tried smoking)?” to which adolescents can respond either yes or no[8]. While another study may ask adolescents to choose from a nine-point scale with multiple response categories, ranging from “I have never smoked, not even one puff” to “I smoke at least once a day” and recode responses to either no (non-smoker) to yes (smoker)[9,10]. There is also some variation in how to assess cigarettes per day: some studies collect the average number of cigarettes smoked per day, while others collect the maximum number of cigarettes per day[11]. Nicotine dependence is most often assessed using the Fagerström Test for Nicotine Dependence[12]. Smoking cessation is assessed in a variety of ways, though the most common seems to be through self-reports of abstinence (e.g. 7-day point prevalence abstinence, 30-day prolonged abstinence, 6-month prolonged abstinence) or by asking about quit attempts. These different stages of cigarette use vary in their heritability, suggesting that different points along smoking trajectories may be influenced by different etiological factors[13]. Distinguishing between these phenotypes helps to provide insight into the nature of cigarette use, which may provide guidance for potential interventions and treatments[14].

Quantitative Studies of Inferred Genetic Susceptibility for Cigarette Use

Classic twin methodologies have been useful in quantifying genetic and environmental factors associated with cigarette use phenotypes. Generally, twin study methods have been used to compare the agreement in the behavior of monozygotic or identical twins that share the same genetic make-up and dizygotic or fraternal twins who share, on average, 50% of their genetic make-up. Statistical models estimate the percentage of variance in the trait explained by genes (i.e. heritability) and by common environment (i.e. experiences that render family members more alike) and unique environment (i.e. experiences that cause dissimilarity between family members)[15]. Heritability estimates differ according to phenotype and age. For the initiation of cigarette use, shared environmental factors account for a small proportion of the liability[16], relative to additive genetic factors, which account for ~60% of the variance[17]. Data from one meta-analysis showed differences in the heritability of initiation by sex, suggesting that genetic and environmental factors may contribute differently to individual differences in initiation in male and female smokers. Whereas the weighed mean heritability for females reached ~50%, the weighted mean heritability for males was ~40%[18]. Meanwhile, heritability estimates for smoking persistence range from 50% to 70%, for smoking quantity from 40% to 60%, for nicotine dependence from 60% to 80%, and for smoking cessation ~50%[19–21]. It has also been suggested that the liability to smoking initiation, regular tobacco use, and nicotine dependence are correlated. Specifically, more than 80% of the variance in liability to initiation and regular use is shared, while a smaller proportion is shared between regular use and nicotine dependence[17]. Added to this, age-dependent genetic effects have been identified, whereby the genetic liability influencing later

cigarette use behaviors is more influential when cigarette use is initiated during adolescence[22], implying a gene-environment interaction with the environment being operationalized as age.

Gene-Finding Efforts for Cigarette Use

While twin and family studies were able to establish that cigarette use phenotypes were heritable, technological advances made it possible to sequence the human genome and look for the genes underlying these twin and family heritability estimates. Gene finding methods are used to determine the locations of gene variants that differentially impact the liability to traits. In general, these gene-finding methods are statistical in nature, such that researchers infer the probability that a locus in the genomic region under investigation contributes to liability for the trait (e.g. cigarette use phenotypes) from an examination of the distribution of genetic markers within either families, as in linkage studies, or populations, as in genome-wide association studies (GWAS)[23]. Genome-wide linkage studies were first used to identify chromosomal regions that could have contained loci contributing to cigarette use phenotypes, involved with either the neurotransmission of neuromodulators or the rewarding efforts of nicotine on the mesolimbic system[24]. Candidate gene studies investigated associations between measures of cigarette use initiation, intensity, and dependence and genes involved with nicotine receptors, dopaminergic transmission, and serotonin transporters. Despite some regions showing suggestive linkage in multiple studies, results have been heterogeneous. Added to this, genes implicated in candidate gene studies have not been reliably associated with cigarette use phenotypes in larger GWAS, the effects of most candidate genes for cigarette use remain largely ambiguous. Replication of candidate gene studies remains

a problem because of small sample sizes in each individual study, differences in measures of cigarette use, and differences in genetic and environmental backgrounds[25]. GWAS simultaneously analyzes common genetic variants across the entire genome and has have been used since the early 2000s to identify genetic variants contributing to cigarette use phenotypes[26]. Gene-finding efforts have identified associations between a variety of cigarette use phenotypes and single nucleotide polymorphisms (SNPs) within neuronal nicotinic acetylcholine receptor genes (nAChRs), the initial physiological targets of nicotine in the central and peripheral nervous system[27–29], and variable-number-of-tandem-repeats (VNTR) polymorphisms located in dopaminergic genes and serotonin transporter genes[30].

Nicotinic receptor genes

Although nAChRs in CHRNA7, CHRNA9, CHRNA5, CHRNB3, and CHRNA4 were found to be significantly associated with nicotine addiction in early candidate gene studies, GWAS failed to provide support for these findings[31]. Instead, independent GWAS have provided evidence for association between common variants within the CHRNA5-CHRNA3-CHRNB4 gene cluster located on chromosome 15 and nicotine dependence. The studies identified in this review investigated the following SNPs within this cluster: rs16969968[32,33], rs680244[33], rs3743078[32], and rs1051730[34,35] which is in near-perfect linkage disequilibrium with rs16969968 in Caucasian samples. Numerous studies have demonstrated the association between functional variant rs16969968 and cigarettes per day (CPD) and nicotine dependence[27,29,36], heavy smoking[28,37] and decreased response to nicotine antagonists in vitro[34]. The same locus was associated with the risk of lung cancer and chronic obstructive pulmonary disease in several GWAS[22]. SNP rs680244

has been associated with variability in CHRNA5 mRNA levels[33]. SNP rs3743078 is a proxy for variant rs578776, which has also been associated with nicotine dependence[32]. Gene variant rs1051730 has been previously associated with smoking quantity and increased susceptibility for lung cancer and vascular disease among smokers.

Dopaminergic genes

The dopaminergic system is also believed to play an important role in nicotine dependence, since nicotine increases dopaminergic activity in the brain to induce feelings of pleasure or reward. Candidate genes include: dopamine receptors (D2 and D4), dopamine transporter gene (DAT1), ankyrin repeat and kinase domain containing 1 (ANKK1), tetratricopeptide repeat domain 12 (TTC12), and the serotonin transporter gene (5-HTTLPR). ANKK1 contains a TaqIA1 C>T polymorphism (rs1800497) that has previously been associated with reduced dopamine D2 receptor availability and binding capacities in the brain, which is believed to cause carriers of the allele to compensate for the reduced state of reward following nicotine use. It is also weakly associated with adolescent smoking initiation[9]. Dopamine receptor D_4 is a G protein-coupled receptor encoded by the DRD4 gene that is activated by the neurotransmitter dopamine. The 48-base pair variable-number-of-tandem-repeats polymorphism in exon III of the DRD4 gene ranges from 2 to 11 repeats. Previous studies have indicated that the longer the repeat, the more dampened the response to dopamine. The DAT1 transporter gene regulates re-uptake of dopamine into presynaptic terminals, terminating dopaminergic neurotransmission, and maintaining dopamine homeostasis. DAT1 contains a polymorphic 40-base pair VNTR which has been previously associated with lower risk of early smoking onset and current smoking. The gene cluster TTC12-ANKK1-DRD2 plays

a central role in modulating dopamine reward system, by mediating the reinforcing effect of all known addictive substances[34].

Serotonin transporter gene

It has been demonstrated that 5-HTTLPR plays a role in nicotine dependence via mediating rewarding effects in the dopaminergic reward system; two common variants (a 14-repeat short (S) variant having less transcriptional activity and lower serotonin uptake and a 16-repeat long (L) variant) seem to have differential effects. While the S allele has a significant effect on smoking behavior, the L allele contributes more to smoking rate[38]. It has been suggested that the differential effects are due to interactions with other polymorphisms, though results are inconclusive[39].

Meta-Analyses, Missing Heritability, and Why Studying GxE is Important

Although independent genome wide association studies have identified variants associated with cigarette use, these variants currently explain very little of the phenotypic variation because genetic effects due to common alleles are quite small and the detection of signals requires very large sample sizes. GWAS are underpowered to detect these effects. To overcome the issue of power and false-positive findings, meta-analysis statistically synthesizes information from multiple studies[40]. The largest genetic meta-analysis of cigarette use conducted by the Tobacco and Genetics Consortium included sixteen GWAS and found five significant loci. Each of the five loci was associated with only one specific smoking phenotype: nonsynonymous rs6265 on BDNF and smoking initiation (OR = 1.06, 95%CI: 1.04, 1.08, p-value = 1.8×10^{-8}); nonsynonymous rs1051730 in 15q25 on nicotinic receptor gene CHRNA3 ($\beta = 1.03$, SE = 0.053, p-value = $2.8 \times 10^{-}$

[73]), rs1329650 on 10q25 (β = 0.367, SE = 0.059, p-value = 5.7 x 10^{-10}), and rs3733829 in 9p13 of EGLN2 (β = 0.333, SE = 0.058, p-value = 1.0 x 10^{-8}) and number of cigarettes per day, and rs3025343 near DBH on chromosome 9 and smoking cessation (OR = 1.12, 95%CI: 1.08-1.18, p-value = 3.6 x 10^{-8})[11]. Still, the variance attributed to these genetic variants only explains a small proportion of phenotypic variation in cigarette use, which does not correspond to estimates of heritability calculated from twin and family studies. A portion of this "missing heritability" might be explained by gene-environment interaction[41], emphasizing the importance of studying GxE. Although reliable demonstration of GxE requires very large sample sizes, studies of GxE can be helpful in determining why heritability estimates for cigarette use phenotypes vary, and could explain why the search for susceptibility genes from GWAS have not been especially successful. Identified genetic loci from the current literature contribute only modestly to the variability in cigarette use phenotypes. Once we can identify more genes contributing to cigarette use, studies of GxE could be used to shape smoking cessation therapies and tobacco control efforts, through interventions tailored to genotypes or environmental factors contributing to tobacco use.

Social and Environmental Risk Factors for Cigarette Use

Although it is clear from the literature that genes influence cigarette use, the motivation to begin smoking is also strongly impacted by the social environment, especially during adolescence[42]. As twin and family studies have demonstrated, shared environmental factors also account for a replicable proportion of the variation in initiation specifically[16], and smoking behaviors more generally[43]. Thus, research on genetics and cigarette use should consider social and environmental factors that may modify genetic risk, especially

when we consider cigarette use as a dynamic process in which individuals can move from initiation, to intermittent use, to regular use, and/or dependence. Understanding the genetic and environmental factors that interrupt progress along this trajectory or potentiate continued use could be useful for intervening with cigarette use and promoting either prevention of initiation or cessation after continued cigarette use[44]. Below, we review epidemiological findings of key environmental covariates that may influence cigarette use and should be considered for genetic research on cigarette use.

Sociodemographic Characteristics. Sociodemographic characteristics should be considered potential environmental covariates in genetic research on cigarette use because the prevalence of smoking tends to be higher among disadvantaged groups. Additionally, disadvantaged users of cigarettes may be more likely to initiate use, less likely to be successful in quit attempts and face higher exposure to the harms of tobacco[45]. Groups that are at higher risk for smoking include the poor, semi-skilled manual occupation groups, the unemployed, poor educational achievers, and single mothers[12,46]. Smoking prevalence among these groups may be due to reduced support for quitting, low motivation to quit, stronger addiction to tobacco, targeted marketing by tobacco companies, and psychological differences regarding self-efficacy in the ability to quit[45], which could be intensified by high feelings of anxiety[47], hopelessness, lack of social, communication, and refusal skills, and low self-esteem[48], or experiencing highly stressful events in childhood[49]. These are all potential points of intervention for cessation efforts and have the potential to reduce health costs associated with cigarette use. Cigarette use also varies by sex between countries, making it difficult to determine whether males or females are more likely to smoke[50]. However, according to a review paper of 12 studies

published between 1980 and 2010 assessing smoking initiation, boys had a lower age of smoking initiation relative to girls[51]. Meanwhile, according to longitudinal studies, girls and boys have similar levels of overall substance use during early adolescence, but boys have greater increases in substance use during middle and late adolescence after initiation[52]. Studies of adult smokers have also demonstrated that women tend to smoke fewer cigarettes per day, use cigarettes with lower nicotine content, and do not inhale as deeply as men. However, it remains uncertain whether this may be due to differences in sensitivity to nicotine or differences in other social factors associated with the experience of cigarette use[53].

Family Cigarette Use. Although family influences play an important role on the development of cigarette use, most of the research done has focused the role of parents and siblings on experimentation with and the onset of cigarette use[54–56]. As evidenced by previous studies, negative family environments characterized by low connectedness or cohesion[57] high levels of parent-child conflict, inadequate parental monitoring, and family violence contribute to tobacco use[58]. Individuals with negative family environments may be less likely to comply with parental requests to abstain from smoking and their initial use may go undetected or unpunished[59]. Alternatively, an authoritative, positive parental style[60], and parental anti-smoking socialization (i.e. messages about smoking, reactions to smoking, household smoking rules), parental expectations and opinions about the choice to smoke[61–63] may help prevent early adolescents from smoking. Added to this, there is consistent evidence demonstrating that parental smoking is a risk factor for adolescent smoking[58,64]. However, one study found that the effect of parental disapproval both in smoking and nonsmoking parents was stronger and more robust than that of

parental smoking and even attenuated the effect of peer smoking, suggesting that parental disapproval makes adolescents more resistant to peer smoking[65]. However, it remains unclear to what extent this is pure environment and passive gene-environment correlation. The presence of gene-environment correlation would imply that non-smoking parents pass on their "non-smoking" genes but also create a non-smoking environment. Siblings also influence the initiation and escalation of cigarette use, such that having older siblings who smoke increases a child's risk of smoking even after adjusting for parents' smoking[66]. Risk of initiation increases substantially as the number of smokers in an adolescent's environment increases, with adolescent females more likely to smoke than adolescent males[56,67]. Besides their smoking behavior, social connectedness between siblings appears to moderate shared environmental influences on smoking frequency and any subsequent changes on smoking frequency[68]. Given that few longitudinal studies have examined how these family influences shape cigarette use following experimentation and initiation, information that could inform the development of effective cigarette use prevention programs addressing family influences remains limited[54].

Peer Cigarette Use. Peer relationships, especially those during adolescence, contribute to an individual's initiation, progression, and trajectories of cigarette use[69]. In fact, adolescent smoking is more strongly associated with peer smoking, relative to parents' smoking[56,70–73]. It has also been suggested that parental smoking does not moderate the association between friends smoking and adolescent smoking; although, parental behaviors may effect smoking progression through their impact on the selection of friends[72] and limiting increases in the number of friends who smoke[61]. It has been previously suggested that adolescents who frequently smoke in the presence of others,

use smoking as a way to achieve social belonging[74] and are more likely to smoke when their best friends smoke. However, there is debate about whether peer influence leads to smoking (e.g. socialization) or whether individuals who smoke tend to seek out other smokers (e.g. selection)[75]. Added to this, cigarette use initiation is more likely to occur in schools with higher smoking rates[76], since smoking may seem more normative and acceptable[77] and more social sources of cigarettes may exist[78]. This might explain why, despite legislation that prohibits tobacco sales to minors, adolescents are still able to acquire cigarettes through direct purchase from others or from older friends[79]. It is also unclear to what extent this is pure environment, rather than active gene-environment correlation whereby individuals are acting on their propensity to use cigarettes by seeking out friend groups that permit cigarette use. Longitudinal study designs of adolescents and their peer groups may help to determine whether gene-environment correlation is present, while disentangling whether socialization or selection has a stronger impact on trajectories of cigarette use. Findings from these longitudinal studies may be helpful in the design of interventions. For example, interventions may want to focus on cognitive factors as a means to mitigate effects of peer group influences on cigarette use through social skills or altering social norms[69].

Age of onset. Approximately 90% of adult smokers first tried cigarettes before the age of 18, and practically all began using cigarettes before the age of 26[80]. In addition to being at higher risk for nicotine dependence[81], individuals who have an earlier onset of cigarette use are at increased risk for heavy smoking[22] and worse tobacco-related health outcomes in adulthood[82]. Added to this, one study conducted on students in grades 9-12 in Canada found that a delay of one year in the age of smoking onset was associated with lower

odds of being a current smoker (adjusted OR = 0.76, 95% CI= 0.73-0.79). Increasing the age of onset also seems to increase the likelihood of successful smoking cessation, as results from another study found that the likelihood of smoking cessation was greater in smokers who had begun cigarette smoking after age 13, relative to individuals who had begun earlier[83]. These findings suggest that early prevention and intervention are needed to avoid early-onset cigarette use to reduce negative consequences associated with cigarette use, such as nicotine dependence and tobacco-related health outcomes in adulthood[82].

Public policy. Given the toll taken by cigarette use, several public policies have been implemented to control tobacco use. The choice of public policy varies considerably between and within countries, allowing for a natural experiment in the study of the effects of tobacco control on the demand for and use of cigarettes. Examples of tobacco control policies include prohibition of paid-for advertising for tobacco products, promotion of smoke-free policies, and excise taxes on tobacco products[84]. Since countries do differ greatly in the prevalence of cigarette use, potentially due to differences in cultural norms and attitudes towards cigarette use, results might not replicate across countries. In general, studies have found that smoking restrictions in public places have a negative effect on average cigarette consumption by smokers, such as smoking restrictions in restaurants, limited cigarette sale through vending machines, and smoking restrictions in shopping areas[85] and workplaces[86]. And, as summarized from one systematic review, increasing taxes on tobacco products independently reduces smoking prevalence among youth and adults, while banning smoking in public places reduces the prevalence of

smoking among the general population, and mass media campaigns reduce the initiation of smoking in youths and prevalence in adults[87].

Religion. Although religion seems to be inversely related to all measures of tobacco use (i.e. lifetime, occasional, and regular use), findings suggest that religion's primary influence on cigarette use is the negative effect it has on ever use[88]. Importance of religion and attendance in worship services are negatively associated with smoking, such that the more religious a teenager perceives him or herself to be, the less likely it is that he or she would smoke[89]. Furthermore, private religiosity is protective against initiation of regular smoking among nonsmokers as well as the initiation of experimental smoking, but only when the young person attends religious services or a religious youth group frequently. Meanwhile, public religiosity predicts the reduction and cessation of cigarette use among regular smokers[90]. It has been suggested that religiosity may discourage the use of substances through adolescents' exposure to religious doctrines discouraging the use of substances, which implies that religious individuals may be more likely to hold conservative attitudes towards substance use, such as cigarette use, and will affiliate with peers that are similar to them[91].

Evidence of Gene-Environment Interaction in Cigarette Use

From the previous sections of this manuscript, it is clear that genes and environments contribute to risk for cigarette use. However, it is important to remember that cigarette use phenotypes are complex traits arising from interactions among social-environmental, psychological, and genetic factors[92] and these interactions need to be taken into consideration when developing downstream public health interventions. Despite progress made in the prevention of and treatment for cigarette use, available treatments are

effective for only a portion of smokers. Whereas the identification of specific genetic variants is necessary in determining the underlying biological mechanism of risks for cigarette-related health outcomes, understanding how these variants interact with aspects of the environment to influence cigarette use has the potential to more effectively tailor interventions to smokers' individual risks and needs[93]. In studies of gene-environment interaction, genetic effects can be modeled either as latent variables in twin and family studies or as genuine measured genes in molecular genetic studies. When genetic effects are modeled latently, the contribution of gene effects is inferred based on observed correlations between people with different degrees of sharing across genes or the environment[5]. These correlations are used to study whether the heritability is the same in different groups. Meanwhile, molecular genetic studies focus mostly on one specific gene of interest, rather than the aggregate effect of genes influencing a trait. Despite awareness of the importance of gene-environment interactions in tobacco use, studies available on the subject are currently limited. Evidence from twin studies have predominantly focused on the importance of genetic factors influencing cigarette initiation, as it relates to family environment, school environment, neighborhood characteristics, and religion, while molecular genetic studies of social policy and the environment have investigated whether genetic influences on initiation, daily smoking, and cessation are moderated by social policy and the environment. All studies discussed in this section still await replication.

Family environment. One Finnish twin study demonstrated that at age 14, the effect of genes on cigarette use increased and common environmental effects decreased as adolescents reported less parental monitoring. Specifically, genetic factors accounted for

more than 60% of the variance at the extreme low end, but less than 15% at the extremely high end of parental monitoring. Meanwhile, common environmental effects accounted for 20% and 80% of the variance at extremely low and high ends of parental monitoring, respectively[8]. Parental monitoring seems to have an effect on genes contributing to nicotine dependence as well, as demonstrated by a significant interaction found between rs169169968 and parental monitoring (p =0.009) in the Collaborative Genetic Study of Nicotine Dependence, whereby nicotine dependence increased with the risk genotype when combined with the lowest quartile of parental monitoring[32]. This suggests that parents moderate the likelihood of an individual at genetic risk for adolescent smoking and nicotine dependence in later life, through the restrictiveness of the social environment provided by parents. Variation in rs3743078 did not contribute to this association, as no significant interaction was found between parental monitoring and rs3743078 (p=0.80). Meanwhile, whether parents smoke may have less of an effect on adolescent smoking, as interactions between measures of environmental smoking, conceptualized as paternal smoking, maternal smoking, or sibling smoking, and genetic variants of DRD2, DRD4, or DAT1 of the dopaminergic system did not significantly contribute to variation in adolescent smoking[9]. Furthermore, only one significant interaction found between maternal smoking and rs1051730 influenced occasional smoking at 14 years[34]. One study investigated the effect of smoking-specific parenting messages across: how often parents talked with their child about smoking- related issues in the past 12 months (e.g. “frequency”), how respectful parents were to children about communicating about smoking-related issues (e.g. “quality), and whether there were smoking-specific rules at home (e.g. “house rules”). The effect of these smoking-specific parenting messages seems limited, as the

Dutch study found no evidence for interaction between smoking-specific parenting in terms of frequency, quality, or house rules, and dopaminergic genes on adolescent smoking behavior[10]. Dopaminergic genetic variants DRD2, DRD4, and DAT1 were chosen for their associations with smoking from previous studies.

School environment. Two twin studies investigated the moderating effect of school-level variables on heritability of adolescent smoking behavior[94,95]. Findings from Daw et al. (2013) suggest that an individual's susceptibility to school-level patterns of smoking is conditional on the number of short alleles in 5-HTTLPR. The greater the number of short alleles, the stronger the individual's response to the school health behavioral environment[38]. No interaction effects were found between dopaminergic genes and peer smoking[9]. Institutional control, which incorporated measures of school smoking policies implemented by adults and whether teachers could smoke on school grounds, was not found to significantly interact with genetic influences on daily smoking among youth[94]. There was also no evidence for interaction between state-level smoking by adults, measured by the percentage of adults reporting regular use in the Behavioral Risk Factor Surveillance System (1992-1993), and genetic influences on regular use during adolescence. This was not the case for state-level smoking by youth, measured by the percentage of 9th to 12th graders reporting frequent smoking, which was found to be negatively associated with genetic influences on regular smoking. Within schools, the effect of genes on daily smoking decreased as the prevalence of smoking among popular students increased, suggesting that social pressures within schools moderate the heritability of daily smoking. These interactions were not found for smoking onset[94]. One study also tested whether the response to a substance use prevention/intervention

program varied based upon a set of five markers (rs16969968, rs1948, rs578776, rs588765, and rs684513) and found that there was a main effect of both the intervention (b=-0.24, p-value<0.05) and genotype at rs16969968 (b = 0.14, p-value <0.05) on high school smoking. The genotype x intervention interaction effect was also found, where those with the A/A and G/A genotypes reduced their levels of smoking to levels similar to those with G/G genotypes following the intervention phenotype (G/G vs. A/A: b = -0.67, p < 0.05; A/G vs. A/A: b = -0.61, p < 0.05; G/G vs. A/G n.s.)[96].

Neighborhood environment. Neighborhood-level factors have previously been associated with the risk of smoking initiation. To test whether genetic factors and social context influence cigarette use, one molecular study investigated the interaction between an aggregated genotypic risk score (GRS) combining the top genetic variants (i.e. all SNPs reaching a p-value threshold of $<5 \times 10^{-7}$) from a meta-analysis previously conducted on African Americans, and neighborhood-level effects on smoking behavior. Among individuals who had ever smoked cigarettes, the GRS significantly predicted the number of cigarettes smoked per day (measured by “In the past 30 days, on those days when you smoked, on average, how many cigarettes did you smoke per day?”) and accounted for ~3% of the variance. Significant interactions were observed between the GRS and number of traumatic events experienced and average neighborhood social cohesion, but not neighborhood physical disorder. The association between the GRS and cigarettes per day increased with increasing number of traumatic events and decreased with increasing levels of neighborhood social cohesion[97].

Religion. Most studies investigating the effect of religion on cigarette use have focused on the association between measures of religiosity and smoking initiation. Only one twin

study investigating the moderating effect of religion on cigarette use was identified, which investigated the interaction between self-rated religiousness, religious affiliation, and organizational religious activity and smoking initiation heritability. This study provided no evidence for interaction between religious affiliation or organizational religious activity and genetic influences of smoking initiation. It did, however, find that high levels of self-rated religiousness attenuated the additive genetic component for smoking initiation[7].

Public Policy. Given that smoking ranks highly among public health problems in the world, public policy initiatives have been implemented to decrease smoking prevalence, while also emphasizing the negative health consequences of cigarette use. Examples of legal and regulatory policies related to tobacco include prohibition of smoking in public places and workplaces, restrictions on sale and marketing of tobacco products (especially to children), and federal legislation giving government agencies the authority to regulate tobacco[98]. One study conducted in the Netherlands explored whether a change in environmental conditions – that is, smoking policies such as cigarette pack warnings about health consequences and bans on smoking advertisements – led to a change in the relative contribution of genetic factors to smoking initiation by comparing data on two cohorts of young adult twins. This study found that although the changes in policies and attitudes towards smoking led to a decrease in the prevalence of smoking, it did not change the heritability of smoking. These findings did not provide support for GxE between initiation and public policy initiatives[99]. Meanwhile, a few studies demonstrate interactions between policy initiatives and the heritability of daily and regular smoking have found evidence for GxE[95,100,101], suggesting that historical time periods can be characterized as distinct social environments that moderate the contribution of genes to

cigarette use[102]. One study using twin pairs from the National Survey of Midlife Development in the United States found that the timing of the first Surgeon General's Report coincides with an increase in the genetic influences on regular smoking, but subsequent legislation prohibiting smoking in public places reduced these influences[100]. Another study conducted using data from the National Longitudinal Study of Adolescent Health, investigated interactions between state-level measures characterizing social and institutional effects on smoking and daily smoking and smoking onset of adolescents. At the state level, the effect of genes on daily smoking were lower in states with relatively high taxes on cigarettes and greater controls on vending machine and cigarette advertising, while there was no variation in heritability estimates for smoking onset among adolescents[95]. Fletcher (2012) also found that variation in the SNP rs2304297 of nicotinic acetylcholine receptor CHRNA6 moderated the influence of tobacco taxation on multiple measures of tobacco use, such that individuals with the protective G/G polymorphism responded to taxation while others had no response. Only one study investigated GxE between policy and cessation, as a study by Boardman et al. (2011) demonstrated that the genetic influences on smoking desistance (measured using a pair-wise measure indicating the length of time in years for a twin to quit smoking after his/her sibling had quit) increased in importance following restrictive legislation on smoking behaviors during the early and mid-1970s[102].

Pharmacological treatment. Studies provide support for the role of genetic variation in response to bupropion and nicotine replacement therapy for smoking cessation. Generally, variations in genes within the dopamine and opioid pathways and in nicotine-metabolizing enzymes appear to play a role in the efficacy of nicotine-replacement

therapy, while variation in dopamine pathway genes are important for response to bupropion[93]. In the one study investigating pharmacological treatment on genetic risk for smoking cessation, genetic variants rs16969968 and rs680244 were used to categorize patients into three haplotypes: (1) low smoking risk allele at rs16969968 and low mRNA expression allele at rs680244, (2) low smoking risk and high mRNA expression, and (3) high smoking risk and high mRNA expression. These haplotypes are located CHRNA5-CHRNA3-CHRNB4 on chromosome 15 and were chosen for their consistent association with measures of smoking heaviness and nicotine dependence in other studies, and potential relation with cessation likelihood. In the smoking cessation trial, haplotype interacted with treatment in affecting success of cessation, in that active treatment was strongly associated with a lower risk of relapse in individuals with haplotype 3 (relative hazard = 0.48, p-value = $9.7x10^{-7}$) and haplotype 2 (relative hazard = 0.48, p-value = $2.7x10^{-8}$), but not haplotype 1 (relative hazard ratios = 0.83, p-value = 0.36). No significant differences were found in the effect of haplotype on abstinence/relapse between bupropion only, nicotine replacement therapy only, and combined therapies treatment groups[33]. Exposure to environmental smoking-related cues may also play an important role in promoting relapse, as individual differences in response to the sight or smell of a lit cigarette may be mediated by the DRD4 VNTR polymorphism. Participants who were homozygous or heterozygous for the seven repeat or longer allele demonstrated significantly higher craving, more arousal, less positive affect, and more attention to the smoking cues than participants for whom this polymorphism was absent[103]. The integration of genetic testing into standard clinical practice would be premature now, but

pharmacologic studies of treatments for nicotine dependence eventually may guide individualized smoking-cessation treatments.

DISCUSSION AND CONCLUSION

Both twin/family and molecular genetic studies provide preliminary evidence that gene-environment interactions have differential effects on cigarette use over the course of development. Twin and family studies demonstrate that the relative contributions of genetic and environmental factors to cigarette use changes across time from adolescence, when most smokers initiate cigarette use, through adulthood, when many smokers have established patterns of cigarette use. Familial and environmental factors contribute to whether individuals initiate cigarette use. However, as individuals move from initiation to more established patterns of use, the importance of common environmental factors decreases while the influence of genes increases. As the contribution of genes to cigarette use increases, the influence of environmental factors does not go away, but rather, environmental factors begin playing a different, but still important role – that is, as a moderator of the influence of genetic susceptibilities[104]. This implies the presence of a gene-environment interaction, such that certain environments allow for greater expression of genetic effects, possibly due to the availability of opportunities for individuals to show their genetic predispositions[6]. While twin studies of gene-environment interaction have been useful in explaining how the effect of genes may change as a function of the environment, molecular genetic studies of gene-environment interaction have been useful in parsing out the role of specific genes influencing cigarette use behaviors through the testing of the main effect of a specific gene of interest from

candidate and genome wide association studies on a given cigarette use phenotype and the testing of interactions between the specific genes of interest and the environment.

From this review of the literature on the influence of genes, environments, and their interaction on cigarette use, we find that significant GxE interactions vary across cigarette use phenotypes. Let us first consider gene-environment interactions contributing to cigarette initiation. Religion was the only environmental variable found to moderate genetic influences on initiation during adolescence[7]. More specifically, of the studies investigating gene-environment interaction contributing to initiation[7,94,95,105], only one twin study yielded a significant interaction between aggregated genetic risk and self-rated religiousness. The Timberlake et al. (2006) study was the only twin study that included this very specific environmental factor on smoking initiation, even though previous associations have been found between religion and decreased risk for smoking initiation in epidemiological studies[7]. To our knowledge, the interaction between specific genetic variants and self-rated religiousness has not been tested in molecular genetic studies and none of the genetic association studies investigating gene-environment interaction contributing to initiation yielded positive GxE results[9,10]. These findings suggest a few different things: either the contribution of genes on initiation remains consistent across different environmental contexts, the effect of GxE in twin studies is quite small, and/or current genetic association studies investigating GxE in cigarette initiation are underpowered to detect effects. Under the first scenario, the environment would have no effect on genetic influences contributing to initiation and encouraging a change in the environment (e.g. increasing self-rated religiousness) would not necessarily reduce cigarette prevalence. The second scenario suggests that the effect of gene-environment

interaction is small and certain environments provide only a minimally greater expression of genetic effects. In the context of religiousness, this might be explained by the fact that genetic influences on smoking have been found to be low or nonexistent among individuals raised with a strong religious upbringing [7]. The third scenario implies a problem with power, so investigators will need to look towards increasing their sample sizes in future studies of gene-environment interaction in cigarette initiation to detect an effect if it is there.

Gene-environment interactions contributing to other cigarette use behaviors, such as adolescent smoking, cigarettes smoked per day, nicotine dependence, and cessation have yielded significant findings as well. A couple of measures of the parental environment moderated genetic influences on adolescent smoking[8,34]. Specifically, significant interactions were found between parental monitoring and rs16969968 of CHRNA5[32] and maternal smoking during pregnancy and rs1051730 of CHRNA3[34] for smoking at age 14. Meanwhile, social pressures to smoke, prevalence of smoking among popular students, marketing and vending restrictions on the sale of cigarettes, and school-level smoking moderated the heritability of daily smoking among adolescents[38,94,95]. Only one molecular genetic study investigated and found a significant interaction between 5-HTTLPR and school tobacco use in influencing tobacco use frequency, such that the greater the number of short alleles, the stronger the individual's response to the school health behavioral environment[38]. In adulthood, the experience of traumatic events and neighborhood social cohesion interacted with aggregated genetic risk to influence the number of cigarettes smoked per day. The interaction between the experience of traumatic events and neighborhood social cohesion and aggregated genetic risk seemed

to be largely driven by a single variant (rs203652) located on the CHRNA5-CHRNA3-CHRNB4 gene cluster[97]. Parental monitoring also interacted with genetic variant, rs16919968, to determine nicotine dependence in adulthood[32], while treatment status interacted with genetic risk for smoking cessation[33].

From these studies, we can see that the environment moderates the effect of genes across different cigarette use phenotypes. However, the extent to which the environment moderates the genetic influences across different cigarette use phenotypes varies. There are a couple of reasons why it is the case that GxE is found for some cigarette use phenotypes and not others. It could be that the genes influencing initiation may be different from the genes influencing other cigarette use behaviors[95,106,107], such as adolescent smoking, daily smoking, number of cigarettes smoked per day, nicotine dependence, and smoking cessation. It is also possible that, the effect of certain environmental measures of smoking that are potentially influenced by both genes and environment [e.g. smoking status of father, sibling, friend, or best friend[34]] seems to vary among carriers of nicotinic receptor genes, but not among carriers of dopaminergic gene variants - possibly, suggesting that either: the effect of nicotinic receptor genes is larger than that of dopaminergic genes or that the effect of dopaminergic genes does not vary as a function of environmental context. Under these assumptions, we might hypothesize that cigarette use initiation may be more heavily influenced by genes predisposing individuals to addictive behaviors via effects on neurotransmitter pathways, such as genetic variants that contribute to novelty seeking[108], while daily smoking, the number of cigarettes smoked per day, and nicotine dependence may have more to do with genes

that contribute to nicotine response, such as genes influencing nicotine metabolism[109–111].

However, we are unable to conclude from some small GxE studies that the phenotypes are not genetically the same, if GWAS on the same phenotypes do not come up with the same list of associated genes. As such, we remain unable to make definitive claims regarding the nature of changing gene-environment interaction contributing to cigarette use due to the limited availability of studies investigating this phenomenon. Each study reviewed here examined the association between specific cigarette use phenotypes and a given environmental measure within either adolescents or adults. This made it difficult to determine how environmental contexts differentially influence genetic factors that contribute to cigarette use phenotypes such as initiation, daily use, nicotine dependence, and cessation and demonstrates how the use of various cigarette use phenotypes may complicate the literature and comparability of findings across studies. Added layers of complexity are found in the fact that there is a great deal of variability in heritability across each cigarette use phenotype[95] and heritability estimates may be contingent on social and institutional characteristics of the environment, such as temporal changes in genetic epidemiology of smoking, changes in smoking norms, changes in the cost of smoking, and legal limits placed upon smokers[102]. Using longitudinal data with repeated measures of different cigarette use phenotypes and environmental contexts would allow researchers to evaluate genetic contributions to the inter-individual variability of each cigarette use phenotype and assess the stability or change of individual differences in each cigarette use phenotype over time. The same longitudinal data could be used to predict cigarette use behaviors over time[112]. Future studies might also want to include a range of nicotinic

receptor, dopaminergic, and serotoninergic gene variants to parse out the effect sizes of main effects on cigarette use phenotypes and interaction effects with different environmental contexts.

In this review of the literature, only a handful of significant gene-environment interactions influencing cigarette use were identified and none were replication studies. To ensure that these findings are not false positives, replication studies using alternative samples are needed. It has been suggested elsewhere that gene-environment interaction studies will be underpowered to detect effects under the following conditions: when the estimated main effects of genes are weak, when the genetic effect is found only among individuals exposed to a particular environmental risk, and when environmental influences are not detected because risk is only conferred among individuals with genetic liability[6]. Replications of findings from studies that have identified significant gene-environment interactions influencing cigarette use would imply that the under-examined role of genetic factors in response to particular environments would be an important step in efforts to further reduce smoking rates[101].

Currently, efforts to reduce smoking rates have focused on the implementation of policies that restrict availability or use of cigarettes in public places. Anti-smoking policies directed at adolescents address onset of cigarette use, while emphasizing the role of immediate social influences and refusal skills, which have been shown to reduce initiation by 30%[113]. However, it is possible that these policies may only be effective for those who are not genetically susceptible to smoking. Furthermore, although restrictions on smoking in public places, anti-tobacco ads, and increased costs of purchasing cigarettes through excise taxes have also aided smokers in quitting, there remain concerns that policies

have focused too heavily on implementing social restrictions on cigarette use, while doing less to help genetically vulnerable smokers quit[95]. To address this gap in the literature and further reduce smoking rates, greater focus needs to be placed on determining the extent to which individual differences are due to genes, environmental factors, or their interaction. Gene-environment interaction studies may help us to better understand how prevention and intervention efforts can be tailored to genotypes under different environmental contexts at the level of family, school, neighborhood, and public policy.'

CHAPTER 3:
GENOTYPE X ENVIRONMENT INTERACTION IN SMOKING BEHAVIORS: A SYSTEMATIC REVIEW[2]

Elizabeth K. Do and Hermine H. Maes

BACKGROUND

This article presents a systematic review of the evidence for gene-environment interaction (GxE) in smoking behaviors, inclusive of smoking initiation, smoking frequency, smoking quantity, nicotine dependence, and smoking cessation. Smoking remains the most preventable cause of morbidity and mortality, yet approximately six million people die from tobacco consumption annually[1]. Twin and family studies have demonstrated that, like other complex traits, smoking behavior is influenced by both genetic and environmental risk factors[2]. Added to this, heritability estimates seem to differ according to the smoking behavior being studied. For smoking initiation, heritability estimates account for approximately 60% of the variance[3], while heritability estimates for smoking persistence ranges from 55 to 69%, smoking quantity ranges from 40 to 56%, nicotine dependence ranges from 60 to 76%, and smoking cessation is approximately 50%[4–6]. Although this suggests that smoking behavior is moderately to highly heritable, few genetic association studies have identified robust associations between specific genes and smoking behavior, aside from studies investigating genes in the *CHRNA5-CHRNA3-CHRNB4* gene cluster and nicotine dependence[7–11]. Alternatively, environmental risk factors for smoking behavior have been well documented and include: socioeconomic status[12,13], parental smoking[14,15], lack of parent-child involvement as

[2] This paper was previously published as: Genotype x Environment Interaction in Smoking Behaviors: A Systematic Review. Elizabeth K. Do; Hermine H. Maes, Nicotine & Tobacco Research (2016): doi: 10.1093/ntr/ntw153.

evidenced by low connectedness or cohesion[16], having siblings who smoke[17], and having friends who smoke[18,19].

Exposure to these environments does not necessarily guarantee the development of smoking behaviors, which brings up the question of what role individual differences in genetic vulnerability to adverse environments plays in shaping smoking behaviors. The study of gene-environment interactions, in part, addresses this question by examining whether individuals with specific genotypes are sensitive to the effects of their environment. Given that genetic factors and one's social context may jointly shape one's risk for smoking behaviors[20–23], the study of gene by environment interactions (GxE) is essential to fully understand the etiology of smoking behaviors and has become an active area of research.

To our knowledge, no systematic review of GxE studies of smoking behavior has been previously published. Thus, the aim of the current article is to identify and summarize studies that test for GxE in relation to smoking behavior among adolescents and adults systematically. We focused specifically on study characteristics related to methods and findings.

METHODS

Systematic review search strategy. A systematic review of the English language literature exploring GxE in smoking behaviors was undertaken. Studies were identified using the electronic databases of Google Scholar, PubMed, ScienceDirect, and Elsevier and included articles on twin-based and molecular genetic studies through May 2014. Search terms included combinations of "smoking", "smoking behavior", "smoking

cessation “, “genetic factors”, “environmental factors”, “twin”, “gene by environment”, “gene-environment”, “interaction”, and “moderation.” To be included in this review, the article had to measure smoking behavior as an outcome of interest and investigate the effect of some environmental factor on the heritability of a given smoking behavior. We allowed for the inclusion of both twin and molecular genetic studies. Initial searches were supplemented by reviewing the reference sections of identified studies. Through these searches, we located sixteen studies.

Extraction of references. Titles and abstracts of all references were initially assessed for relevance. For completeness, bibliographies of extracted references were manually searched for further relevant references. Where relevant references were found, their bibliographies were also manually searched.

Data extraction. All data were extracted for the following variables: (1) study name, (2) study population, (3) study design, (4) definition of environmental risk factor, (5) definition of genetic risk factor, (6) definition of smoking behavior (i.e. outcome of interest), (6) statistical parameters utilized, and (7) primary results presented. Focus was placed on measures of association (i.e. odds ratios, hazards ratios). Data were extracted into a prepared, structured Microsoft Excel database (Microsoft; Redmond, WA, USA).

RESULTS

For the purposes of describing the current state of smoking-related GxE research, we summarize the research design and study samples, measurements of outcome, environment, and genotype, and main study findings of these sixteen studies below.

Research Design and Study Samples

Research Design. The research design varied across studies. Samples were obtained from an assortment of sources including: national registries[24–26], population-based case-control[27], longitudinal studies with school based-study designs[21,23,28], longitudinal community samples[20,29–31], hospital samples[32], epidemiological studies[22,33], and randomized smoking cessation trials[34]. Of the identified studies, six utilized twin samples[21,22,24,25,35], while the remainder were molecular genetic studies[20,23,26,27,29–34]. Although many of these studies collected longitudinal data, associations between environmental exposures and outcome were not always determined prospectively. Three studies utilized cross-sectional data[27,32,34], while four studies assessed the environmental exposures repeatedly[25,26,29,30] and seven studies accounted for gene-environment correlation[20,24,28–31,35].

Sample. All sixteen studies differed in their sample size, both within and between molecular and family studies of smoking behavior. The sample size of family studies ranged from 1,310[22] to 4,120[21] individuals, while molecular genetics study samples ranged from 365[29] to 14,560[23] individuals. The identified twin and molecular genetic studies included eight unique samples from the United States [i.e. The National Health and Nutrition Health Examination III (NHANES) Phase 2 study (1991-1994)[33], the National Longitudinal Study of Adolescent Health[21,23,28,35], 1995 National Survey of Midlife Development in the United States[22], Collaborative Genetic Study of Nicotine Dependence[27], Detroit Neighborhood Health Study[20], 1987 Atherosclerosis Risk in Communities Study[34], the Smoking Cessation Trial of the University of Wisconsin Transdisciplinary Tobacco Research Center[34], and a combined sample from the University of Connecticut Health Center, Yale University School of Medicine, Medical

University of South Carolina, Mclean Hospital of Harvard Medical School, and the University of Pennsylvania School of Medicine[32]. A handful were European studies conducted in the Netherlands [i.e. Family and Health Study[29,30] and Netherlands Twin Register[25]], and Finland [i.e. Finntwin12[24] and the 1966 Northern Finland Birth Cohort[26]]. One study used data obtained from the Dunedin Multidisciplinary Health and Development Study of New Zealand[31].

Race/ethnicity. Studies also varied with respect to the amount of racial/ethnic diversity in the samples. Most studies exclusively focused on participants that were Caucasian, except for one study that investigated a sample residing in Detroit that was predominantly African American[20].

Sex. Studies seemed to be balanced with respect to sex and none of the identified studies limited their sample to only males or females.

Age. Some studies focused on adolescent samples, followed longitudinally[21,24,28–30], while others focused on adult samples, aged 18 and older[20,22,25,27,34,35]. Two studies included individuals in early adolescence to young adulthood, ranging from age 11 to 31 years[23,26]. More specific details for each of the studies identified in this systematic review can be found in the online supplementary material, under Supplementary Table 3.5: Twin Studies of Gene by Environment Interaction of Smoking Behavior and Supplementary Table 3.6: Genetic Association Studies of Gene by Environment Interaction on Smoking Behavior.

Measurement of Smoking Behavior

Outcome measures. The studies assessed smoking behavior by smoking initiation or onset[25,28,35], adolescent smoking[24,26,29,30], smoking frequency, including: number of cigarettes smoked in the past month[23,28], regular smoking[22], and cigarettes smoked per day[20]; nicotine dependence[27,31,32], and age at smoking cessation and relapse[34], as described in Table 3.1: Measures of Smoking Behavior.

Table 3.1: Measures of Smoking Behavior

Phenotype	How was it measured?
Smoking initiation	Did you ever smoke? (1) Yes (2) No/A Few Times to Try
	Have you ever smoked an entire cigarette? (1) Yes (2) No
Adolescent Smoking	Lifetime measure of adolescent smoking, measured by adolescents reporting on smoking level based on a nine-point scale: (1) I have never smoked, not even one puff (2) I smoke at least once a day
	Adolescent smoking at age 14, measured by "Have you ever smoked (or tried smoking)? (1) Yes (2) No If indicated "yes" then asked, "How many cigarettes have you smoked altogether up to now?" (1) Only one (2) About 2-10 (3) About 11-50 (4) Over 50 These variables were recoded into one variable with five categories: (1) No (2) Yes, only one (3) Yes, about 2-10 (4) Yes, about 11-50 (5) Yes, over 50
	Measure of whether adolescents smoked at age 14 and how much, which was categorized into: (1) Nonsmokers: adolescents who had never smoked or had smoked once/twice in their lives (2) Occasional smokers: adolescents smoking occasionally or about twice per week (3) Regular smokers: everyone else
Daily Smoking	Ever smoking at least once cigarette every day for 30 days
Smoking Frequency	Total number of cigarettes smoked by the respondent in the past month derived by multiplying responses to the following two questions: (1) "During the past 30 days, on how many days did you smoke cigarettes" (2) "During the past 30 days, on days you smoked, how many cigarettes did you smoke each day?"
Regular Smoking	Regular smoking was measured using two questions: (1) "Have you ever smoked cigarette regularly – that is, at least a few cigarettes every day?" and (2) If yes, "On average, about how many cigarettes did you smoke per day during the one year in your life when you smoked most heavily?" These two questions were dichotomized into: (1) Never been regular smokers: smoked less than three cigarettes per day during the time of heaviest smoking (2) Regular smokers

Cigarettes Per Day	Respondents provide a quantitative measure to the following question: “In the past 30 days, on those days when you smoked, on average, how many cigarettes did you smoke per day?”
Nicotine Dependence	Fagerström Test for Nicotine Dependence scores, dichotomized by: (1) Nicotine dependent: FTND scores ≥ 4 (2) Not nicotine dependent: FTND scores < 4
Smoking cessation	Self-reported age of smoking cessation measured by asking, “How old were you when you stopped smoking?”
Smoking relapse	Any smoking on seven consecutive days after the target quit date

Data collection methods. Data on the outcomes of interest were collected by six studies exclusively through self-reported survey response[24–26,33–35], while other studies collected data through in-home face-to-face interviews[21,30], semi-structured interviewing (i.e. Semi-Structured Assessment for Drug Dependence and Alcoholism (SSADDA))[32], family reports[31], and school reports[28]. Two studies used a combination of these methods[23,29]. Three studies indicated that data were collected through telephone interviews, though information regarding the background of the interviewer is not clear[20,22,27]. Only two of these studies investigated a biochemical indicator of smoking behavior[33,34]. However, of these two studies, only one indicated that the biochemical indicator was laboratory-based serum cotinine levels (mg/ml)[33].

Measurement of genes

Heritability estimates. In contrast to estimates of genetic variance based on measured genotypic data, biometrical genetic methods rely on the expected variance and covariance estimates of MZ and DZ twins. The expected variance and covariance estimates of MZ and DZ twins were used to estimate the latent genetic influences for the liability of smoking behaviors in different ways across twin and family studies. While two studies calculated heritability estimates from comparisons of correlations of identical and

fraternal twins[24,25], three studies estimated heritability based on extended twin pair designs that included full and half siblings[28,28,35].

Polymorphisms examined. Eight single nucleotide polymorphisms (SNPs) were investigated across two studies[23,33]. These SNPs included variants in neuronal nicotinic acetylcholine receptor genes (nAChRs; including *rs16969968, rs680244, rs3743078, rs1051730,* and *rs2304297*), variable-number-of-tandem-repeats (VNTR) polymorphisms in dopaminergic genes (*rs1800497*) and serotonin transporter genes (*5-HTTLPR*). One study focused on the polymorphic region of the promoter region of the serotonin transporter gene (SLC6A4), *5-HTTLPR*[23]. Another study divided individuals based upon their *CHRNA6* genotype (C/C, C/G/, G/G in the *rs2304297* SNP), noting that the G/G genotype has previously been related to lower likelihood of tobacco use[33]. These polymorphisms were chosen because of their potential role in the development of nicotine addiction. Specifically, nAChRs are the initial physiological targets of nicotine in the central and peripheral nervous system, while the dopamine and serotonin mediates feelings of pleasure or reward within the dopaminergic reward system[36], such that dopaminergic activity in the brain is increased by exposure to nicotine.

Data Collection Methods and Genotyping. Three studies used blood or saliva samples for genetic analysis. Limited information was provided by these studies on which specific cell lines were used and it seems to be the case that none of the studies utilized the same processing facilities, with more specific details described in Table 3.2: Genotyping Data Collection Methods.

Table 3.2: Genotyping Data Collection Methods

Study	Genotyping Data Collection Methods
Xie (2012)	Extracted DNA from immortalized cell lines directly from blood or saliva and implementing the Taqman method to genotype SNP *rs16969968* at the Yale University School of Medicine
Hiemstra et al. (2013)	Taqman analyses performed on the 7500 Fast Real-Time PCR system and scored genotypes using the algorithm and software supplied by Applied Biosystems
Ducci et al. (2011)	Did not disclose the type of samples used but explained that genome-wide genotyping was performed on DNA available at the Broad Institution Biological Sample Repository using Illumina Infinium 370cnvduo array
Meyers et al. (2013)	Isolated DNA from whole blood, or saliva if unavailable, which was then sent to the Applied Genomics Technology Facility (Wayne State University, Detroit, MI, USA) for genotyping using the humanomniexpress Beadchips (Illumina, San Diego, CA, USA)
Chen et al. (2009)	Collected blood samples for genotypic analyses with initial genotyping performed by Perlegen Sciences using custom arrays, and follow up genotyping done by Center for Inherited Disease Research (CIDR) using Illumina Golden Gate technology with focus on two SNPs
Chen et al. (2012)	Genotyping performed by the Center for Inherited Disease Research at Johns Hopkins University using Illumina Omni2.5 microarray, with data cleaning led by the GENEVA coordinating center at the University of Washington
Daw et al. (2013)	Did not indicate how samples were collected

Measurement of the environment

Types of environmental exposures assessed. The types of environmental risk factors assessed were diverse. Five studies investigated the role of family level factors, such as parental monitoring[24,27], maternal smoking during pregnancy[26], smoking-specific parenting including frequency, quality of communication, and house rules regarding smoking[30], and environmental smoking by father, mother, and siblings[29]. Three studies examined the role of school and peer level factors, including social pressure to smoke within schools[28], institutional control over smoking in schools[28], prevalence of youth smoking[23,28]and youth drinking[23], and smoking by friends and best friends[29]. Two studies investigated childhood adversity[32] and childhood maltreatment[31]. Two studies observed the role of neighborhood level factors, inclusive of social cohesion, physical disorder, lifetime trauma[20], and socioeconomic status as measured by marital status of mothers during pregnancy, socioeconomic status of cohort collected at age 31 years, and family socioeconomic status based upon occupation of father during pregnancy and at age 14

years[26]. One study investigated the role of treatment status[34] and another focused on the role of religion, as measured by religious affiliation, organizational religious activity, and self-rated religiousness[35]. Four studies assessed the role of public policy initiatives, examining the effect that cigarette restrictions, tobacco control, prevention budgets, excise tax per pack of cigarettes[21], cohort effects[25], and tobacco taxation policies[33] have on heritability estimates of smoking behavior. More detail about the measures of environmental exposures is provided in Table 3.3: Measures of Environmental Exposure.

Table 3.3: Measures of Environmental Exposure

What is the environmental exposure of interest?	How was it measured?	# of Items	First Author (Year) Population	Was GxE present? If so, with what genotype & for what smoking outcome?
FAMILY LEVEL				
Parental Monitoring	Adolescent report on the degree to which the parent: (1) Discuss with them their daily plans, (2) Know of their interests and activities, and (3) Know their whereabouts and the identity of their associates when they are not at home Measure was standardized and treated as a semi-continuous measure	3	Dick et al. (2007) Finland	Yes, heritability of adolescent smoking changed with varying levels of parental monitoring for smoking at age 14.
	Adolescent report on the degree to which the parent: (1) Expected a specific time for them to come home (2) Noticed them coming home later than expected (3) Arrived home soon after they arrived home from school and the degree to which the adolescent: (4) Told parent when they would be back (5) Left a note about where they were going (6) Checked in with parent before going out again (7) Knew how to get in touch with parent (8) Talked with parent about their plans for the coming day Low parental monitoring defined by the lowest quartile in the parent monitoring sum of scores (treated as ordinal	9	Chen et al. (2009) USA (Detroit, MI and St. Louis, MO)	Yes, interaction found between SNP (*rs16969968)* and parental monitoring influencing nicotine dependence. Nicotine dependence increased with the risk genotype when combined with lowest quartile of parental monitoring. No evidence of interaction between SNP *(rs3743078)* and parental monitoring.
Paternal Smoking	Adolescent report indicating stage of smoking, ranked on an 8-point scale ("My father/mother have never smoked" to "My father/mother smokes more than 31 cigarettes a day") Self-report indicating which stage of smoking applied to them ("I have never smoked, not even one puff" to "I smoke at least once a day") Recoded into three categories: "Never smoked", "former smoker", and "current smoker"	2	Hiemstra et al. (2014) Netherlands	No evidence for interaction between paternal smoking and *DRD2, DRD4, or DAT1.*
Maternal Smoking	Adolescent report indicating stage of smoking, ranked on an 8-point scale ("My father/mother have never smoked" to "My father/mother smokes more than 31 cigarettes a day") Self-report indicating which stage of smoking applied to them ("I have never smoked, not even one puff" to "I smoke at least once a day") Recoded into three categories: "Never smoked", "former smoker", and "current smoker"	2	Hiemstra et al. (2014) Netherlands	No evidence for interaction between maternal smoking and *DRD2, DRD4, or DAT1.*

	Maternal smoking during (2nd month of) pregnancy, classified as nonsmokers, light smokers (1-10 cigarettes per day), and heavy smokers (>10 cigarettes per day)	1	Ducci et al. (2011) Finland	Yes, significant interaction detected between maternal smoking during pregnancy and *TTC12* (rs10502172) influencing adolescent smoking.
Smoking-Specific Parenting	Frequency: average of adolescent reported scores assessing how often parents talked with child about smoking related issues in the past 12 months	8	Hiemstra et al. (2013) Netherlands	No moderating effects of the dopaminergic genes by the frequency of smoking-specific parenting were found.
	Quality: average of adolescent reported scores on quality of communication about smoking issues	6	Hiemstra et al. (2013) Netherlands	No moderating effects of the dopaminergic genes by quality of smoking-specific parenting were found.
	House Rules: average of adolescent reported scores assessing the existence of smoking-specific rules at home	5	Hiemstra et al. (2013) Netherlands	No moderating effects of the dopaminergic genes by house rules were found.
Sibling Smoking	How many of your siblings smoke, on a scale from 0 to 4 (None of my brothers/sisters smokes to four of my brothers/sisters smokes) Answers were dichotomized into "having no smoking siblings" and "having one or more smoking siblings"	2	Hiemstra et al. (2014) Netherlands	No evidence for interaction between sibling smoking status and *DRD2, DRD4, or DAT1.*
SCHOOL LEVEL				
School-Level Smoking Pattern	Pressure to smoke within a school: characterizes the popularity status of smokers and nonsmokers and indicates the extent to which the most popular students smoke by asking all students to write down the names of their five closest female friends and five closest male friends and linking these names to self-reports on smoking during the past year	2	Boardman et al. (2008) USA	Yes, attending a school where more popular students are more likely to smoke increases heritability of smoking daily.
	Each student was asked: "During the past twelve months, how often did you smoke cigarettes?" Responses ranged from 0 to 6 (never to nearly every day) School-specific mean response used for analysis	1	Daw et al. (2013) USA	Yes, adolescents smoke more cigarettes when attending schools with higher rates of tobacco use.
Institutional Control	School smoking policy: measured by summing administrator responses regarding disciplinary action of the school upon first and second incidences of smoking on school grounds, inclusive of verbal warming, minor action, in-school suspension or expulsion Assessment of whether or not teachers can smoke on school grounds	2	Boardman et al. (2008) USA	No evidence for interaction, though the heritability estimate is reduced within schools where normative pressures to avoid smoking are present (this finding is not significant, however).
Prevalence of Smoking	*Adult Smoker Prevalence:* State-level measure of the percentage of adults who reported regular smoking (i.e. those who have smoked at least 100 cigarettes and smoke currently), obtained from the Behavioral Risk Factor Surveillance System (1992-1993) *Adolescent Smoker Prevalence:* percentage of 9th to 12th graders who reported frequent smoking (i.e. those who have smoked a cigarette on at last 20 of the past 30 days), obtained from the Youth Risk Behavior Surveillance study	2	Boardman (2009) USA	No evidence for interaction between state-level smoking by adults and genetic influence on regular smoking. State-level smoking by youths negatively associated with genetic influences on regular smoking.
	Adolescent Smoker Prevalence: Proportion of students reporting that they have ever smoked a cigarette by the date of the in-school survey	1	Boardman et al. (2008) USA	Yes, attending a school where more popular students are likely to smoke increases the heritability of daily smoking.
Racial Composition	Proportion of students who are non-Hispanic and white	1	Boardman et al. (2008) USA	Yes, heritability of daily smoking is reduced within school where most students are non-Hispanic and white.
Friend's smoking	How many of your friends smoke, ranging from 1 to 5 (no one to all of them) Answers were dichotomized into "having no smoking friends" and "having smoking friends"	2	Hiemstra et al. (2014) Netherlands	No evidence for interaction between friend's smoking and *DRD2, DRD4, or DAT1.*
Best Friend's Smoking	Adolescent report indicating stage of smoking, ranked on an 8-point scale ("My best friend has never smoked" to "best friend smokes more than 31 cigarettes a day") Self-report indicating which stage of smoking applied to them ("I have never smoked, not even one puff" to "I smoke at least once a day") Recoded into three categories: "Never smoked", "former smoker", and "current smoker"	2	Hiemstra et al. (2014) Netherlands	No evidence for interaction between best friend's smoking and *DRD2, DRD4, or DAT1.*
NEIGHBORHOOD LEVEL				

Social Cohesion	Asks respondents whether they agree or disagree on a 4-point scale (strongly agree to strongly disagree) with the following statements: (1) This is a close-knit or unified community (2) People around here are willing to help their neighbors (3) People in this neighborhood generally do not get along with each other (reverse coded) (4) People in this neighborhood do not share the same values (reverse coded) (5) People in this neighborhood can be trusted Items were summed and then averaged by neighborhood (via census tracks), to create neighborhood-wide measure.	5	Meyers et al. (2013) USA (Detroit, MI)	Yes, significant interactions were found between genetic risk score and average neighborhood social cohesion influencing cigarettes smoked per day.
Physical Disorder	Items were adapted from the New York City IMPACT neighborhood evaluation scale and then factor analyses were conducted, yielding 3 factors: (1) Presence of buildings with broken windows; boarded-up windows, or boarded-up doors (2) Presence of buildings with outside damage that can only be corrected by major repairs (3) Presence of entirely vacant buildings Principal component values for each block group were calculated and averaged by neighborhood.	19	Meyers et al. (2013) USA (Detroit, MI)	No evidence for interaction between physical disorder and genetic risk score.
Socioeconomic Status	Family SES: Based on occupation of father, collected during pregnancy and at age 14 years. Classified as professionals; skilled-workers; unskilled workers and farmers. SES of Cohort: collected at age 31 years and classified as professionals skilled workers; unskilled workers farmers; and others Marital status of mothers during pregnancy dichotomized as married or unmarried (including divorced and widowed).	3	Ducci et al. (2011) Finland	No evidence for interaction between socioeconomic status and *TTC12(rs10502172).*
	Maternal SES: Proportion of student's mothers who have complete college	1	Boardman et al. (2008) USA	No, maternal smoking was not found to moderate genetic risk of smoking behavior.
Lifetime Trauma	Traumatic Events: Checklist of 19 items occurring in the individual's entire lifetime. Number of items endorsed was summed to create score from 0 to 19. Higher scores reflect greater number of traumatic events.	19	Meyers et al. (2013) USA (Detroit, MI)	Yes, significant interactions were found between genetic risk score and the number of traumatic events experienced.
Childhood Adversity	Semi-Structured Assessment for Drug Dependence and Alcoholism (SSADDA) Environmental section, which asked whether either of their parents died before they were 6 years old and whether before the age of 13 they had witnessed or experienced a violent crime, had been sexually abused, or had been physically abused.	4	Xie et al. (2012) USA	Yes, childhood adversity significantly increased ND risk in both women and men, and the effect in women was twice that than in men.
PHARMACOLOGICAL TREATMENT				
Treatment Status	These were randomly assigned, as it was part of a clinical trial. Groups included: placebo, nicotine patch, nicotine lozenge, sustained-release bupropion, nicotine patch and nicotine lozenge, or bupropion and nicotine lozenge.	1	Chen et al. (2012) USA	Yes, smokers with high-risk haplotype were three times as likely to respond to pharmacologic cessation treatment compared to smokers with the low-risk haplotype
RELIGION				
Religious Affiliation	Individuals were characterized as those who had any affiliation and those who did not (inclusive of atheists, agnostics, and those without any affiliation).	1	Timberlake et al. (2006)	No evidence for interaction.
Organizational Religious Activity	Frequency of religious attendance and participation in special activities in the past 12 months, with responses coded from 0 to 6 (never to more than once a week). Items were summed.	2	Timberlake et al. (2006)	No evidence for interaction.
Self-rated religiousness	Indicates the importance of religious faith and extent of being a religious person, with responses ranging from 0 to 3 (not important/not religious at all to more important than anything else/very religious).	2	Timberlake et al. (2006)	Yes, high levels of self-rated religiousness attenuated the additive genetic determinant of smoking initiation.
PUBLIC POLICY				
Cigarette Restrictions	Sum of two characteristics: restrictions on the location of vending machines selling cigarettes and prohibition of billboard advertising for tobacco products within 500 feet of schools	2	Boardman (2009) USA	Yes, marketing and vending machine restrictions slightly reduce genetic influences on regular smoking

Full-time staff equivalent for tobacco control	Number of staff that are dedicated to tobacco control, not including nonprofit organizations in 1994	1	Boardman (2009) USA	No evidence for interaction.
Prevention budget	Total amount spent in logged dollars on tobacco control within each state (drawn from multiple sources)	1	Boardman (2009) USA	No evidence for interaction.
Excise Tax per pack of Cigarettes	Derived from data from the Centers for Disease Control and Prevention State Tobacco Activities Tracking and Evaluation System (reflecting legislation active as of December 1, 1995)	1	Boardman (2009) USA	Yes, state-level excise tax on cigarettes reduce genetic influences on regular smoking
State-level per pack of tobacco tax rate	Rates were matched to data at the state and year-levels and are not adjusted for inflation for the four years of data.	1	Fletcher (2012)	Yes, G/G polymorphism of *CHRNA6* responded to state-level taxation while others did not.
Birth Cohort	Four birth cohorts (1920-1939, 1940-1949, 1950-1959, and 1960-1970)	1	Boardman et al. (2010) USA	For those born in the 1940s and 1950s, genetic factors do not significantly contribute to the risk of regular smoking. While for those born in the early 1930s and mid-1950s, genetic influences are the most pronounced.
	Two birth cohorts (1993-1995 and 2009-2010)	1	Vink et al. (2011) Netherlands	No, the heritability of smoking initiation did not change as a function of environmental exposure.

Classification of exposures. There was considerable variation in how studies treated exposure status in the analysis. In one study, the environment was treated as a binary variable (i.e. Exposed versus unexposed to childhood adversity)[32]. In another study, the environmental variables, which measured sociodemographic factors such as father's occupation and mother's marital status during pregnancy, were treated as categorical[26]. Most studies (62.5% or n=10) used continuous measures or scales, derived using sum scores of different sets of items[20,21,23,24,27–31,33,35], while two studies utilized proportions of smokers in contact with respondents as a measure of environmental smoking exposure[28,31]. One randomized control trial randomly assigned participants to one of six environmental exposures or treatments (i.e. Placebo, nicotine patch, nicotine lozenge, bupropion SR, nicotine patch and nicotine lozenge, or bupropion and nicotine lozenge)[34]. Classification of exposure was unclear in two studies investigating genetic influences on smoking by birth cohort[22,37].

Data collection methods. The most commonly employed method for obtaining information about environmental exposures were questionnaires. The remaining studies relied on interviews, review of epidemiological data, or a combination of approaches. Like data obtained on smoking behavior outcomes, data on environmental exposures were collected by: self-reported survey response[24–26,33–35]; in-home face-to-face interviews[21,30]; semi-structured interviewing (i.e. Semi-Structured Assessment for Drug Dependence and Alcoholism (SSADDA)[32]; Semi-Structured Assessment for Nicotine Dependence (SSAND), Semi-Structured Assessment for the Genetics of Alcoholism (SSAGA); Composite International Diagnostic Interview (CIDI)[27]); telephone interviews[20,22,27]; family reports[31]; school reports[28]; or a combination of these methods[23,29].

Main study findings

There was considerable heterogeneity in the methods and analyses used across studies to test for GxE, making it difficult to summarize this research and provide a synthesis of main findings. Thus, statistical significance is emphasized over the magnitude of effects.

Main effect of genotype. Thirteen (87.5%) of the sixteen identified studies found significant main effects for either specific genes or genetic factors associated with smoking behaviors[20–24,26–31,33–35]; three did not[30,32,37]. More specifically, SNPs *rs16969968* of *CHRNA5* and *rs3743078* of *CHRNA3* were associated with nicotine dependence[27], while *5HTTLPR*[23], *CHRNA5-CHRNA3-CHRNB4*[34], and a genetic risk score[20] were associated with smoking heaviness or number of cigarettes smoked, and *DRD4* was associated with smoking onset[29]. Interestingly these main effects were not

necessarily consistent across studies, as no main effect of *rs16969968* on nicotine dependence was reported in one study[32]. No significant main effect for *DRD2* or *DAT1* on smoking onset was reported[29,30].

Main effect of environment. Eight studies found significant main effects for at least one of the environmental variables[20,21,23,24,26–29,32] and all but two of these studies[21,28] reported the effect sizes of these main effects. Two studies found the main effects to be nonsignificant[30,38]. The remaining three studies did not provide sufficient information to make this determination[22,35,37].

Gene by environment interaction effects. Of the sixteen studies identified, thirteen studies found at least one significant GxE effect ($p \leq 0.05$). Significant interactions were detected between environmental factors and SNPs within the *CHRNA5-CHRNA3-CHRNB4* gene cluster [i.e. Parental monitoring and *rs16969968* of *CHRNA5* for nicotine dependence[27] and maternal smoking during pregnancy and *rs1051730* of *CHRNA3* for smoking at age 14[26]]. Another significant interaction was detected between school-tobacco use and the serotonin promoter polymorphism, *5-HTTLPR* for tobacco use frequency[23]. However, no significant interactions between environmental factors and dopaminergic genes were identified[29,30].

Effect Size. It was not clear from some studies whether the effect for genotype or environment was larger, due in part to studies not reporting the main effect of genotype and environment[21,22,28,31]. Studies that did report this information tended to find that the effects of the environment were larger relative to genetic effects and GxE effects.

Effects by developmental period. Of the studies focused on adolescent samples, only two did not find at least some evidence for GxE[29,30]. Only one study investigating adult samples did not find at least some evidence for GxE[25].

Results demonstrate heterogeneity in both conceptual and methodological approaches to conducting tests for gene by environment interactions related to smoking behavior. This heterogeneity could be an artifact of the cross-disciplinary nature of GxE research and may reflect differences in conceptual understanding and methodological conventions adopted across academic disciplines. Given these differences in approach, it was difficult to synthesize findings, which remains a major limitation to the current literature on smoking-related GxE. This is emphasized further in a meta-analysis of 103 gene by environment interaction studies conducted in the first decade of this millennium by Duncan and Keller (2011). Results from the meta-analyses are "consistent with the existence of publication bias, low statistical power, and a high false discovery rate[39]. What this means is that thus far, gene by environment research might not have produced many reliable results[40].

Despite the differences in approach and difficulty to synthesize findings, we can come away with a few general conclusions/themes regarding the role of GxE interaction in smoking behavior. Specifically, the influence of parents and peers seem to moderate the genetic and environmental influences contributing to the initiation and maintenance of smoking behaviors, such that greater influence from parents or peers attenuates the relative importance of genetic versus environmental factors. The magnitude of this moderation is dependent upon the outcome of interest (e.g. initiation, frequency, nicotine dependence, cessation, or relapse). Thus, more attention needs to be paid to the

outcomes of interest and how results are being reported, particularly since the heritability estimates vary by the smoking behavior being measured and the age at which these outcomes are collected.

Studies of GxE interaction demonstrate that the influence of genes may change as a function of the environment and the phenotype being measured, but also suggests that restricting the availability of tobacco (whether through parental monitoring, prevalence of smoking among peers, or public policy initiatives to reduce tobacco use through restriction of use in public spaces or taxation) generally decreases the influence of genes that influence the initiation and maintenance of smoking behaviors.

Although we can come away with these conclusions, it is still the case that differences in methodological approaches are likely contributors to potential discrepancies of GxE effects across studies, and prevents the field from a deeper understanding of GxE interactions regarding smoking behavior. In efforts to guide future research and address the current challenges that exist in synthesizing findings of GxE in smoking behavior, we offer the following suggestions focused on: (1) choice of measurement for environmental variables, (2) testing and reporting of main and interaction effects, (3) testing for artifactual interaction via conducting sensitivity analyses and checking for scaling artifacts, (4) treatment of covariates, and (5) reporting gene-environment correlation (rge). Table 3.4. Assessing Validity of Gene by Environment Interaction demonstrates how each study addresses these concerns.

Table 3.4. Assessing Validity of Gene by Environment Interaction

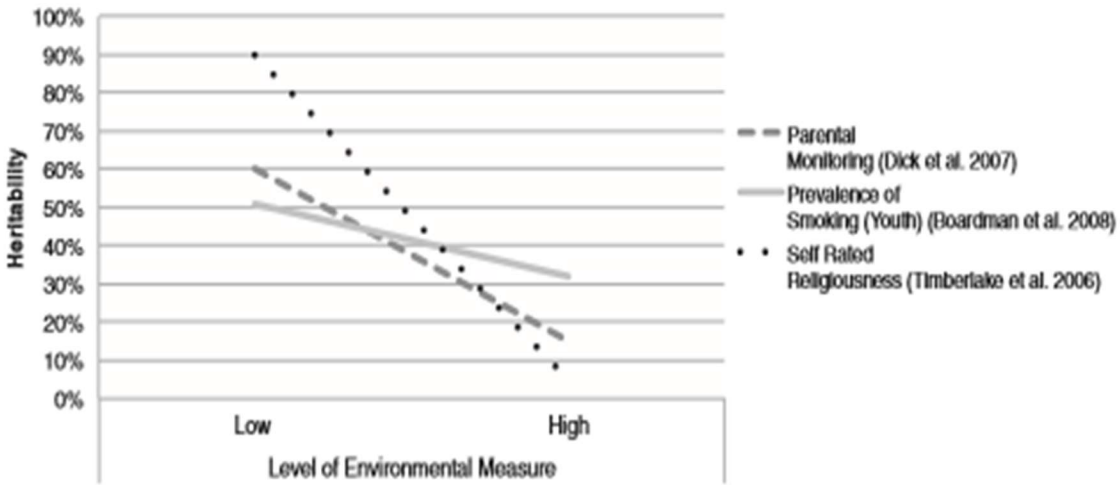
First Author (Year)	Account for rGE	Substitute genotype	Sensitivity Analyses	Variable Transformation	Effect Estimate CI	# of tests with p-values or MTC?	Reports Non-Significant Findings
Dick et al. (2007)	Yes, Analyses are adjusted for rGE.	No	No	No	Yes	No	No
Chen et al. (2009)	No	No	No	No	Yes	No	Yes
Hiemstra et al. (2013)	Yes, Pearson's correlations among the study variables were provided.	No	No	No	Yes	No	Yes
Hiemstra et al. (2014)	Yes, rGE found between maternal and paternal smoking at time 1 and DAT1.	No	No	No	Yes	Yes	Yes
Ducci et al. (2011)	No	No	No	No	No (gives SE)	Yes	Yes
Boardman et al. (2008)	Yes, Controls for passive and active rGE via maternal and peer smoking.	No	No	No	No	No	Yes
Daw et al. (2013)	No	No	Yes	No	No	No	Yes
Meyers et al. (2013)	Yes	No	Yes	No	Yes	No	Yes
Xie et al. (2012)	No	No	No	No	Yes	No	Yes
Chen et al. (2012)	No	No	No	No	Yes	No	Yes
Timberlake et al. (2006)	Yes, cross-trait Spearman correlations between siblings within a pair by zygosity were assessed in the initial testing for presence or absence of rGE.	No	No	No	No	No	Yes
Vink et al. (2011)	No	No	No	No	Yes	Yes	Yes
Boardman (2009)	No	No	No	No	No (gives SE)	No	Yes
Boardman et al. (2010)	No	No	No	No	Yes	Yes	Yes
Fletcher (2012)	No	No	No	No	Yes	No	Yes

CI = confidence interval; MTC = multiple testing correction

Choice of measurement for environmental variables

The choice of measurement for environmental variables varied across studies identified in this systematic review. Few studies from this systematic review overlapped in their measurement of environmental constructs making it difficult to compare findings across studies. Only two studies investigated the effects of parental monitoring[24,27] and two other studies investigated the effect of the prevalence of smoking among youth[21,28]. However, the findings of these studies do demonstrate a general trend: increasing the restrictiveness of an environment (e.g. increasing parental monitoring, decreasing the prevalence of smoking among youth, increasing self-rated religiousness) decreases the influence of genes on an individual's behavior, such as smoking, as seen in Figure 3.1. Moderation of the Heritability of Smoking by Environmental Measures.

Figure 3.1. Moderation of the Heritability of Smoking by Environmental Measures



Furthermore, studies investigating the same environmental constructs did not use the same means of measurement. In the study conducted by Dick et al. (2007), parental monitoring was measured using responses from three items, standardized and then

treated as a semi-continuous measure[24], while the study conducted by Chen et al. (2009) used a sum score from eight items, defined the lowest quartile, and treated the measure of parental monitoring as an ordinal variable[27]. As for the prevalence of adolescent smokers, one study utilized a state-level measure of the percentage of 9th to 12th graders reporting frequent smoking from the Youth Risk Behavior Surveillance study[21], while the other calculated the proportion of students reporting they had ever smoked a cigarette by the date of the in-school survey[28]. These observations suggest that there may be "noise" in the assessments of environments. Thus, more rigorous methods need to be undertaken to establish both reliable and valid assessments of environmental exposures across studies.

There was also wide variation in the timing of exposures assessed across studies in terms of the temporal relationship between exposure and outcome (i.e. Prospective versus cross-sectional) and the developmental period considered (i.e. Adolescence versus adulthood). A couple of studies tried to demonstrate causality through cross-lagged methodologies, but most studies measured exposures and outcomes simultaneously. To better understand timing of environmental exposures, future studies may want to incorporate more rigorous research designs, including experimental and quasi-experimental approaches that utilize the longitudinal nature of the data being used in many of these studies.

Obtaining consistent environmental variables would make it much easier to synthesize study findings. However, as can be seen from the studies included in this systematic review, comparable results found across slightly different environmental variables measuring the same construct can also provide some evidence of a moderating effect.

Additionally, it would be beneficial for future studies to focus on a broader array of proximal environments, since very little attention was paid to protective factors, such as neighborhood social cohesion[41]. However, it is important that genetic factors do not influence these proximal environments (i.e. gene-environment correlation is limited) and attention must be paid to developmental period of measurement, since certain environments are more salient at specific ages (i.e. family and peer influences during adolescence).

Testing and reporting of main and interaction effects

Few studies fully described their methods and analyses, including how tests for interaction were conducted and the nature of the association between exposure and outcome. Many studies indicated that they used moderated regression analysis or twin modeling approaches, but few provided citations or details for how estimates were calculated. Future research should report basic descriptive information that may be suggestive of GxE, such as a data table reflecting genotype by exposure by outcome. This recommendation is made based on the finding that some studies did not include univariate analyses on environmental exposures and smoking outcomes.

Future studies should explicitly note the scale (i.e. additive or multiplicative) used to detect GxE effects, as it has been previously demonstrated that the way that outcome measure is scaled and whether the GxE effect is tested on the additive or multiplicative scale influences whether a GxE effect is observed. Under the simple definition of gene-environment interactions, which suggest that either a different effect of an environmental exposure on disease risk in persons with different genotypes or a different effect of a

genotype on disease risk in persons with different environmental exposures, the presence or absence of interactions may depend on the scale of measurement (e.g. whether effects are additive or multiplicative). Where risks are measured on an additive scale, the effect of the environmental exposure differs among persons with different genotypes, but if risk factors are measured on a multiplicative scale, the effect of the exposure differs among people with different genotypes[42]. Thus, changing the scale of the outcome may create interactions that may not have previously existed or eliminate interactions that were once present. For example, binary outcomes have been shown through simulations to incorrectly detect GxE effect when none existed, thus raising concerns about the validity of results based on diagnoses[43]. Most of the identified studies of GxE in smoking behavior provided confidence intervals of effect estimates with p-values[20,22–25,27,29–34]; though, several studies provided standard error estimates with p-values instead[21,26,28]. Only a few of the studies reported the number of tests conducted[22,25,26], and all but one of the studies reported non-significant GxE.

There remains a need for more thorough reporting standards, especially as they apply to conducting tests for interaction. This should include regression coefficients for all parameters included in a regression model and explicitly noting what variables were included. Investigators should be cautious about interpreting any genotype or environmental main effect reported in previous studies, unless authors explicitly describe parameters in regression models.

Tests for artifactual interaction

Only one study tested for artifactual interactions by conducting sensitivity analyses[23], suggesting that this has not been a predominant concern. However, this remains an important task, as it brings to question the validity of findings. There are various ways to check for artifactual interaction, including: substituting genotype, conducting sensitivity analyses, and running analyses with transformed variables[43,44]. Substituting the genotype entails trying to remove a significant interaction by replacing genotypic data with a similarly distributed polymorphism that has no association with smoking behavior. Conducting sensitivity analysis entails testing whether or not using different measures that share construct validity for the behavior of interest still reveals an interaction[44]. Had GxE been observed in one of a set of measures, but not the other, then it would have suggested the occurrence of a scaling artifact[45]. The transformation entails trying to remove the interaction effect by re-running analyses with transformed variables and checking to see if the detected interaction is still significant. This can be done different ways, including monotone transformations [e.g. taking a logarithm or square root]. If the significance of the interaction is removed through transformation, an additive relationship between variables on different scales is implied. If not removable, the interaction effect could be interpreted as: a robust fan-shaped interaction not removable by transformation, a crossover effect, or a qualitative interaction[44].

Treatment of covariates

The treatment of covariates was not uniform across studies, as studies controlled for a combination of the following: age, sex, race/ethnicity or ancestry (measured by principal components), smoking behavior of parents, socioeconomic status, and education. The covariates of sex, age or developmental period, and race/ethnicity or ancestry should be

included more explicitly in future GxE research. These factors are important for understanding the etiology of smoking behavior, as well as environmental exposure patterns, and may be related to differences in genotypic frequency.

Sex. Most studies controlled for or stratified results, though the studies did not always note whether they found different GxE effects for males when compared to females.

Age. Exploration of the importance of age for GxE is necessary for several reasons. Some environmental exposures are age-specific. For example, parental monitoring decreases over time and may peak around early adolescence. Without accounting for explicitly exploring how age influences GxE effects, research may be biased. Added to this, there is now substantial evidence that genetic risks for smoking problems have age-dependent effects. Specifically, genetic risk of heavy smoking is greater in early-onset smokers (i.e. Prior to 16 years) when compared to later-onset smokers (i.e. 16 years or older). This association in early-onset smokers is consistent with the epidemiologic observation of increased vulnerability to dependence among early-onset smokers[46].

Race/ethnicity or ancestry. By not controlling for race/ethnicity, studies may lead to biased results. This is related to population stratification whereby different allele frequencies may exist among different sub-populations or ancestral groups. Ideally, research should describe methods to assess or address population stratification and control for self-reported race/ethnicity and conduct sensitivity analyses to test whether GxE effects vary by race.

Reporting of gene-environment correlation

Only about half of the identified studies accounted for gene-environment correlation (rge)[20,24,28–30,35], which refers to the phenomenon where an individual's genotype also influences his/her exposure to the environment. This implies that individuals shape their environments through heritable behaviors and that the relationship between environmental exposure and behavior may be confounded by genotype. If rge is not accounted for in studies of GxE, it is unclear whether the environment is moderating genetic effects or if genes influencing a trait are more likely to be present in a given environment[47]. Conducting tests for rge can be accomplished through simple tests of association between environmental exposures and genotype, but may be limited by genotypes measured. To account for rge, researchers can: limit studies to moderators that are uncorrelated with the outcome[43], utilize a moderator in means model to remove genetic effects shared by a trait and moderator from covariance[48], or explicitly model rge in a bivariate model. The studies accounting for rge did so in a variety of ways including model adjustment for rge[24], controlling for passive and active rge via inclusion of maternal[28,29], paternal[29], and peer smoking[28] as covariates, and providing Pearson's[30] or cross-trait Spearman correlations[35]. However, no explanations were provided for why these specific methodologies were used in the studies. We recommend that future research test and report whether rge is present.

CONCLUSION

This systematic review of the literature was conducted in attempts to understand the state of the science on GxE research of smoking behavior, focused on methodological approaches used across studies. A total of sixteen studies were identified, with thirteen finding at least some evidence to suggest a GxE effect. However, among these thirteen

studies, none of the findings seemed to overlap. The heterogeneity in results is likely related to the variation in conceptual and methodological approaches used to test for GxE. Studies varied in the populations sampled, methods used to assess environmental exposures, and means by which they tested for GxE effects. Methodological heterogeneity made it difficult to interpret and summarize findings. However, we hope that the recommendations provided will help to guide future studies towards reducing heterogeneity and capturing the joint contribution of genetic and environmental factors to smoking behavior.

Supplementary Table 3.5: Twin Studies of Gene by Environment Interaction of Human Smoking Behavior

Dick et al. (2007): FinnTwin12 (Finland)						
411 MZM 401 MZF 391 DZF 439 DZM	11.4 years (0.3 years); Follow-up at 14.1 years (0.1 years)	Adolescent smoking (at age 14)	Parental monitoring	NR	Yes	As parental monitoring increases, C and A decrease, and E increases
Boardman et al. (2008): National Longitudinal Study of Adolescent Health (United States)						
163 MZTP 240 DZTP 647 Sib Pairs 148 Half-Sib Pairs	7th to 12th grade	Smoking onset; Daily smoking	Social pressure to smoke within schools; Institutional control; Prevalence of smoking; Racial composition of school	SES (maternal education)	Yes	Social pressures within schools moderate the heritability of daily smoking, but not smoking onset. As prevalence of smoking among popular students increased, the heritability of daily smoking increased. As the proportion of non-Hispanic Whites increased, the heritability of daily smoking increased.
Boardman (2009): Wave II of the National Longitudinal Study of Adolescent Health (United States)						
248 MZTP 378 DZTP 1066 Sib Pairs 368 Half-Sib Pairs	12 to 21 years	Regular Smoking	Marketing and vending machine restrictions; Full-time staff equivalent for tobacco control; Prevention budget; Excise Tax per pack of Cigarettes; Adult Smoking Prevalence; Youth Smoking Prevalence	Sex; Age	NR	As marketing and vending machine restrictions increased, genetic influences on regular smoking increased. Genetic influences on regular smoking increased as: state-level excise tax on cigarettes increased and prevalence of state-level smoking by youths increased.
Boardman et al. (2010): 1995 National Survey of Midlife Development in the United States (United States)						
340 MZ pairs 315 SSDZ pairs	25 to 75 years	Regular smoking	National trends in cigarette consumption	NR	NR	Timing of first Surgeon General's Report coincides with increase in genetic influences on regular smoking, while subsequent legislation reduced influences. For those born in the 1940s and 1950s, genetic factors do not significantly contribute to the risk of regular smoking. While for those born in the early 1930s and mid-1950s, genetic influences are the most pronounced.
Vink et al. (2011): Two birth cohorts (1993-1995 and 2009-2010) from the Netherlands Twin Register (Netherlands)						
415 MZM 363 DZM 658 MZF 462 DZF 769 DZO	18 to 25 years	Smoking initiation	Changes in policy and smoking attitudes	NR	NR	Heritability of smoking initiation did not change as a function of environmental exposure.
Timberlake et al. (2006): Wave III of the National Longitudinal Study of Adolescent Health (United States)						
237 MZTP 315 DZTP 779 Sib Pairs 233 Half-Sib Pairs	22.4 years (1.7 years) Range: 18-27.4 years	Smoking initiation	Religious affiliation; Organizational religious activity; Self-rated religiousness	NR	Yes	As self-rated religiousness increased, genetic influences on smoking initiation increased.

Supplementary Table 3.6: Genetic Association Studies of Gene by Environment Interaction of Smoking Behavior (N=10)

Subjects	Mean Age (SD) or Age Range	Smoking Behaviors	Environmental Measure	Genotypic Measure	Covariates	rGE	GxE
Chen et al. (2009): Population-based case-control from Collaborative Genetic Study of Nicotine Dependence (United States, from Detroit, MI and St. Louis, MO)							
1,032 cases; 995 controls	25 to 44 years	Nicotine dependence (current smokers with FTND score > 4)	Parental monitoring	*CHRNA5 (rs16969968); CHRNA3 (rs3743078);* combined risk	Age; Gender	NR	Nicotine dependence increased with risk genotype (AA of *rs16969968*) when combined with lowest quartile of parental monitoring. No evidence of interaction between *rs3743078 and* parental monitoring
Ducci et al. (2011): Prospective cohort from 1966 Northern Finland Birth Cohort (Finland)							
2,476 females; 2,286 males	14 and 31 years	Smoking behavior at 14	Maternal smoking during pregnancy; SES	*CHRNA3(rs10517 30); TTC12 (rs10502172); ANKK1(rs2734849); DRD2 (rs1076563)*	Maternal smoking; Family SES; Maternal marital status	NR	Significant interaction between maternal smoking during pregnancy and *CHRNA3(rs10517 30)* for smoking at age 14.
Hiemstra et al. (2013): Longitudinal cohort/family-based design using sibling dyads from five waves of data from the Dutch "Family and Health" Study (Netherlands)							
108 boy-boy; 118 boy-girl; 106 girl-girl; 96 girl-boy	Older siblings: 14-16 years Younger siblings: 13-15 years	Smoking initiation and smoking onset	Smoking-specific parenting (i.e. frequency, quality of communication, and house rules)	*DRD2 (rs1800497); DRD4;* 40-base pair VNTR in *SLC6A3 (DAT1)* gene	Gender	Yes, Pearson's correlations among the study variables were provided.	Moderating effects of dopaminergic genes were not found.
Hiemstra et al. (2014): Longitudinal cohort/family-based design using sibling dyads from five waves of data from the Dutch "Family and Health" Study (Netherlands)							
Study 1: 465 Boys 526 Girls Study 2: 175 Boys 190 Girls	Study 1: 12.52 years (0.57 years) Study 2: 14.16 years (1.07 years)	Adolescent smoking (lifetime)	Environmental smoking, including paternal, maternal, sibling, friend's and best friend's smoking	Dopamine receptors (*DRD2* and *DRD4*) and dopamine transporter *DAT1*	Age; Gender; Ethnicity (Dutch or Other); Education level	Yes rGE found between maternal/ paternal smoking at time 1 and DAT1.	No significant interactions between environmental smoking variables and dopaminergic genes.

Daw et al. (2013): Longitudinal cohort/school-based study design from Wave I and II from the National Longitudinal Study of Adolescent Health (United States)							
14,560 individuals (53% female)	11 to 22 years (mean = 16.4 years)	Number of cigarettes smoked in the past month	School-level smoking; School-level drinking	Serotonin transporter gene (*5-HTTLPR*)	Race/ethnicity; Sex; Age; Home access to alcohol and tobacco; School penalties for drug use	NR	More short alleles are associated with stronger response to the school health behavioral environment.
Meyers et al. (2013): Longitudinal cohort from the Detroit Neighborhood Health Study (United States)							
778 individuals	18 to 95 years (mean = 52.62 years)	Cigarettes per day	Traumatic events, average neighborhood social cohesion, average neighborhood physical disorder	Genetic risk score consisting of *rs2036527, rs667282, rs3101457, rs938682, rs547843,* and *rs3813550*	Sex; Age; Ancestry	Yes, Tests of Spearman's rGE were conducted.	Significant interactions were found between genetic risk score and the number of traumatic events experienced and average neighborhood social cohesion, but not average neighborhood physical disorder
Chen et al. (2012): Prospective epidemiologic study and randomized placebo-controlled smoking cessation trial from the 1987 Atherosclerosis Risk in Communities Study (United States) and the Smoking Cessation Trial of the University of Wisconsin Transdisciplinary Tobacco Research Center							
12,771 individuals	45 to 64 years	Age at smoking cessation and relapse, following attempts to quit	Treatment status	*CHRNA5-CHRNA3-CHRNB4 (rs16969968* and *rs680244)*	Gender; Age (quartiles); CPD; Treatment	NR	Smokers with high-risk haplotypes had increased likelihood of responding to pharmacologic cessation treatment when compared to low-risk haplotype. No significant differences in haplotypic effects on abstinence or relapse between treatment groups
Belsky et al. (2013): Prospective, longitudinal study of a representative birth cohort obtained from the Dunedin Multidisciplinary Health and Development Study (New Zealand)							
1,037 males and females	11 to 38 years	Smoking initiation, conversion to daily smoking, progression to heavy smoking, nicotine dependence (e.g., Fagerström Test for Nicotine Dependence), cessation difficulties	Family history of smoking	Genetic risk score, derived from 3 recent meta-analyses of GWAS using cigarettes smoked per day (*CHRNA5-CHRNA3-CHRNB4, CYP2A6, rs7937, rs4105144)*	Adolescent developmental phenotypes mediating associations between GRS and mature phenotypes	Family history score and GRS were uncorrelated.	GRS was unrelated to smoking initiation, but individuals at higher genetic risk were more likely to convert to daily smoking as teenagers, progressed more rapidly from smoking initiation to heavy smoking, persisted longer in smoking heavily, developed nicotine dependence more frequently, more reliant on smoking to cope with stress, and more likely to fail

							cessation attempts.
Fletcher (2012): Cross-sectional data of US adults from the National Health and Nutrition Examination Survey (NHANES) (United States)							
7,200 respondents with survey and biological specimen data	42.83 years (SD: 17.09 years)	Current tobacco use [self-report (tobacco use, number of cigarettes daily) and laboratory-based serum cotinine levels (ng/ml)]	State-level per-pack tobacco tax rate	*CHRNA6 (rs2304297)*	Age, sex, race/ethnicity, income, marital status	NR	Variation in nicotinic acetylcholine receptor moderates the influence of tobacco taxation on multiple measures of tobacco use. G/G polymorphism responded to taxation while others did not.
Xie et al. (2012): Subjects recruited for linkage and association studies of the genetics of drugs and alcohol dependence at give US sites: the University of Connecticut Health Center (n=1102), Yale University School of Medicine (n=866), the Medical University of South Carolina (n=155), McLean Hospital of Harvard Medical School (n=57), and the University of Pennsylvania School of Medicine (n=26)							
1771 subjects recruited as substance dependence cases and unaffected controls	38.1 years (SD: 11.0 years)	Semi-Structured Assessment for Drug Dependence and Alcoholism (SSADA), lifetime substance dependence according to DSM-IV criteria	Childhood adversity, as assessed by SSADDA Environment section	*rs16969968*	Ancestry proportion score, sex, age	Accounted for correlated data from individuals in the same family	Childhood adversity increased ND risk in both men and women, with the effect in women being two times that in men. Significant interactive effects of childhood adversity and *rs16969968* were observed in men, but not women.

CHAPTER 4:
GENETIC AND ENVIRONMENTAL INFLUENCES ON SMOKING BEHAVIOR ACROSS ADOLESCENCE AND YOUNG ADULTHOOD IN THE VIRGINIA TWIN STUDY OF ADOLESCENT BEHAVIORAL DEVELOPMENT AND TRANSITIONS TO SUBSTANCE ABUSE FOLLOW-UP[3]

Elizabeth K. Do, Elizabeth C. Prom-Wormley, Lindon J. Eaves, Judy L. Silberg, Donna R. Miles, Hermine H. Maes

BACKGROUND

Cigarette smoking is the leading cause of preventable death in the US and has been associated with considerable economic, social, and personal costs. Annually, tobacco use costs the nation an estimated $193 billion, inclusive of lost productivity and direct health care expenditures[1]. Yet, 19% of all US adults, or approximately 43.8 million people, smoke cigarettes[2]. Of these adult smokers, 70% began smoking regularly by age 18[3].

Despite notable declines in cigarette smoking over the past 40 years, smoking behavior among adolescents remains a huge public health concern. Every day, about 3,900 children under the age of 18 try their first cigarette. Of these children, an estimated 950 will become new, regular daily smokers[4]; approximately half will die as a result of nicotine addiction and other smoking-related causes[5]. Twin studies have suggested that both genetic and environmental factors contribute to smoking behavior. However, many of

[3]This chapter was previously published as an original research article in *Twin Research and Human Genetics.* To cite information from this chapter, please use the following citation: Do EK; Prom-Wormley EC; Eaves LJ; Silberg JL; Miles DR; Maes HH, 2015. Genetic and Environmental Influences on Smoking Behavior across Adolescence and Young Adulthood in the Virginia Twin Study of Adolescent Behavioral Development and the Transitions to Substance Abuse Follow-Up. *Twin Res Hum Genet* 18(1):43-51

these twin studies investigate the influences of genes and the environment on cigarette use among adults, so less information is known regarding the genetic and environmental influences of cigarette use in adolescents.

Early twin studies investigating the genetic and environmental influences of smoking behavior of adolescents analyze various stages of smoking behavior such as initiation, progression, dependence, and addiction separately[6–8]. These studies find that the initiation of tobacco use in adolescence is primarily explained by shared environmental factors[9,10] while genetic factors contribute more to individual differences in other smoking behaviors, such as daily quantity of cigarettes smoked[6] or smoking progression, which has an estimated heritability of 0.80[9,11,12]. Furthermore, population based twin studies provide evidence that genetic influences come to play a larger role in smoking behavior by late adolescence, when the etiological structure of smoking initiation closely resembles that of adult samples[13,14].

Among adult samples, heritability estimates for smoking initiation range from 0.32 to 0.78, making it a moderately heritable trait[8,15–20]. On average, estimates are higher in women relative to men[8,21–23], suggesting that the heritability of smoking initiation may differ by gender. However, this finding has not been replicated across all studies[18].

As a consequence of analyzing smoking behavioral factors separately, we lack information on whether any overlap exists across stages[20]. Although we know from adult studies that utilize multivariate analyses that significant genetic and environmental covariance exists between initiation and dependence[8,18,24,25], it remains unclear whether the genetic and environmental factors influencing the relationship between smoking initiation and progression in adulthood are the same across adolescence into early

adulthood[6,21,26]. We also do not know if qualitative and quantitative sex differences found in adult samples exist in adolescent samples[21–27].

Thus, this study seeks to answer these questions by: examining the relationship between smoking initiation and current quantity smoked from adolescence to early adulthood, determining if qualitative and quantitative sex differences exist in this relationship, and estimating the contributions of genetic and environmental factors to smoking initiation and current quantity smoked in this younger age group.

MATERIALS & METHODS

Sample. Data were obtained from the Virginia Twin Study of Adolescent Behavioral Development (VTSABD) and its young adult follow-up, Transitions to Substance Use (TSA). The VTSABD is a multi-wave, cohort-sequential prospective study of adolescent psychopathology and its risk factors, in over 1,400 Caucasian juvenile twin pairs aged 8 to 17 years and their parents[28], greater detail about the ascertained sample have been provided elsewhere[29]. To be included the present study, individual twins had to have responded to questions regarding smoking initiation and current quantity smoked. The total sample size of this study was 2,804 twins (including 632 MZ male twins, 829 MZ female twins, 367 DZ male twins, 389 DZ female twins, and 587 DZ opposite sex twins). Data obtained for the 22 to 32-year age group (N = 1,074) was obtained from one wave of the TSA, to which all participants of earlier waves of the VTSABD were invited.

Measures. Data from each of the five waves of the VTSABD were merged and then re-categorized into age groups to ensure that there was an adequate sample size (i.e. 12-13 years, 14-15 years, and 16-17 years). However, since there was only one assessment

during age period from 22-32 years, subdividing the TSA sample by age was not warranted. Two main variables of interest were recoded across each of these age groups: one measuring whether twins had ever smoked at least one whole cigarette and another measuring the current quantity of cigarettes smoked daily. The “ever smoke” variable was binary, coded as 0 for those who had never smoked at least one whole cigarette and 1 for those who had indicated that they had ever smoked at least one whole cigarette. If respondents indicated that they had “ever smoked” in a given age group (i.e. 14-15 years), they would be given a value of 1 for “ever smoke” in that age group and every subsequent age group (i.e. 14-15 years, 16-17 years, and 22-32 years). Otherwise, if the respondents indicated that they had not “ever smoked” across all age groups, they were given a value of 0 for “ever smoke”. To measure current quantity smoked, respondents had to indicate the number of cigarettes smoked daily, in the past three months. Free responses were coded into three categories. These categories indicated: zero cigarettes smoked daily (“non-current smoker”), one to five cigarettes smoked daily (“current, light smoker”), and five or more cigarettes smoked daily (“current, heavy smoker”). Only responses where twins indicated that they had smoked before under the “ever smoke” variable were included in the quantity of cigarette use variable. Otherwise, responses for individuals who had indicated that they had never tried cigarettes were coded as missing for the quantity of cigarette use variable.

Descriptive Statistics. Prevalence estimates for smoking initiation and quantity are reported using percentages.

Genetic Analyses. All data analyses were conducted using the open-source structural equation modeling software OpenMx[30,31]. Due to inadequate sample size for smoking

quantity in 12-13-year-olds, only univariate genetic analysis on smoking initiation was conducted in this age group. Causal-common-contingent (CCC) models were fit, individually, for smoking initiation and smoking quantity across all other age groups (i.e. 14-15 years and 16-17 years in the VTSABD, and 22-32 years in the TSA).

Using the causal-common-contingent (CCC) model originally developed by Kendler and colleagues (1999) [Are you sure? I think it's Neale (2006)], smoking behavior was conceptualized as a two-stage process incorporating initiation and current quantity smoked. This model was chosen because it allows for estimating the relative magnitude of the contributions of genetic and environmental factors to smoking liability, as well as for testing the strength of the association between initiation and current quantity smoked stages for smoking via a beta pathway between the two stages[18,24,26,32,33].

The significance of an estimated beta pathway between the two stages is used to assess whether the two stages are independent or correlated processes. Specifically, if an estimated beta coefficient is found to be not significant, the liabilities for initiation and current quantity smoked are said to be independent of one another, implying that smoking initiation and current quantity smoked have separate genetic and environmental risk factors. Otherwise, if the estimated beta coefficient is significant, the liabilities for smoking initiation and current quantity smoked are said to share genetic and environmental risk factors. In this case, the beta coefficient provides an estimate of the magnitude of strength of association between smoking initiation and current quantity smoked. The greater the estimated beta coefficient, the larger the magnitude of the strength of the association between smoking initiation and current quantity smoked (i.e. beta coefficient of zero suggests that the two stages do not share genetic and environmental risk factors while a

beta coefficient of one suggests that the genetic and environmental risk factors for these two stages are identical). The estimated 95% confidence intervals around the beta coefficient give further information regarding the degree of overlap between the two stages. Again, lower limits approaching zero (or below) support independent liabilities and upper limits approaching 1 provide support for identical liabilities.

Using this model also allows for the direct estimation of additive genetic effects (a^2), shared/common environmental effects (c^2), and unique environmental effects (e^2) on both smoking initiation and current quantity smoked. However, since current quantity smoked is modeled conditionally upon smoking initiation, the genetic and environmental influences unique to current quantity smoked are estimated after those on initiation are taken into account. Thus, the proportion of variance in current quantity smoked explained by the respective influences on initiation can be calculated by multiplying them by the squared beta coefficient. The proportion of the variance in liability to current quantity smoked that is explained by genetic factors is the sum of the proportion of variance in initiation explained by genetic factors multiplied by the squared beta parameter and the proportion of variance explained by unique genetic factors contributing only to the current quantity smoked stage, with the same principle applied for environmental factors.

Nested models were fitted to test specific hypotheses about the nature of association between the two stages of smoking initiation and current quantity smoked. More explicitly, to determine whether qualitative sex differences exist in the relationship between smoking initiation and current quantity smoked, we tested the significance of the genetic and shared environmental correlations between male and female factors. A model constraining the correlation between males and females to 1, suggesting that the same

factors contribute to male and female smoking behavior, was compared to a model that freely estimated correlations between male and female factors, suggesting that different factors contribute to male and female smoking behavior. This was done separately to test whether the same genes or same environmental factors contribute to the liability of smoking initiation and current quantity smoked in males and females.

Quantitative sex differences were tested for simultaneously to answer the question of whether genetic and environmental factors explain the same proportion of the liability of smoking initiation and current quantity smoked in males and females. To test for quantitative sex differences, a model equating all parameters (i.e. genetic, shared environmental and unique environmental factors, but not thresholds) for males and females and was compared to one allowing for free estimation of parameters for males and females separately. If the model equating parameters between males and females fit the data best, it was concluded that quantitative sex differences did not exist. This process was repeated for each age group.

Following these tests for qualitative and quantitative sex differences, other alternative models were fit to the data. Specifically, nested models were created to test if there is a direct relationship between smoking initiation and current quantity smoked and whether genetic or common environmental factors could be dropped from initiation and current quantity smoked stages. Where the beta pathway could be dropped from the model without significant loss to goodness-of-fit to the data, it was determined that smoking initiation and current quantity smoked had independent liabilities. Alternatively, when dropping the beta pathway led to significant loss to goodness-of-fit, smoking initiation and current quantity smoked were said to have shared liabilities. Regardless of whether this

finding was significant, we moved on to test whether we could drop genetic or shared environmental factors from either initiation or current quantity smoked. Where genetic or environmental factors could not be dropped without significant loss to goodness-of-fit, the factor was said to contribute significantly to the smoking phenotype.

Nested models were compared using likelihood ratio chi-square (LRC) statistics, in which the degrees of freedom equal the difference between the degrees of freedom of the full and nested sub models. LRC is calculated as the difference in -2 log likelihood (-2LL of a comparison model and the -2LL of a reduced nested model[34,35]. Where the LRC comparing the two models is non-significant, the reduced model is selected as the better fitting model. Akaike Information Criterion (AIC) was also used as an index of model fit, as well as an index of parsimony[36,37].

RESULTS

Smoking prevalence. At age 12-13 years, 10.4% of the total sample had indicated that they had ever smoked. This increased to 27.4% by age 14-15 years, 46.6% by age 16-17 years, and 79.1% by age 22 to 32 years. Across all age groups, most respondents indicated that they were not current smokers (i.e. indicated that in the past 3 months, they smoke zero cigarettes daily). Although the majority (approximately 71%) of adolescents who tried smoking did not become ‘current, heavy smokers’, the proportion of ‘current, light smokers’ and ‘current, heavy smokers’ did increase consistently from the younger to the older age groups (Table 4.1).

Table 4.1: Smoking Initiation and Current Quantity Smoked Prevalence of Sample

	% Initiated smoking (indicated having ever smoked)			
	Age 12–13	Age 14–15	Age 16–17	Age 22–32
Total sample	10.4	27.4	46.6	79.1
MZ males	12.0	32.0	48.0	77.5
MZ females	8.6	22.2	35.0	76.6
DZ males	11.2	31.3	46.8	84.2
DZ females	8.0	26.8	32.6	78.5
DZ opposite sex	12.4	27.8	50.2	82.6
	% Non-current smokers (0 cigarettes smoked daily)			
	Age 12–13	Age 14–15	Age 16–17	Age 22–32
Total sample	91.2	85.1	84.3	58.0
MZ males	80.6	86.4	83.7	59.3
MZ females	96.9	88.1	88.1	62.3
DZ males	100.0	83.9	77.8	53.1
DZ females	100.0	86.0	87.2	58.0
DZ opposite sex	89.7	79.5	84.4	52.3
	% Current, light smokers (1–5 cigarettes smoked daily)			
	Age 12–13	Age 14–15	Age 16–17	Age 22–32
Total sample	8.8	8.2	5.9	13.3
MZ males	19.4	6.8	7.8	15.7
MZ females	3.1	5.9	2.4	10.4
DZ males	0.0	9.7	7.4	12.5
DZ females	0.0	8.0	2.1	12.5
DZ opposite sex	10.3	12.3	7.1	16.5
	% Current, heavy smokers (5+ cigarettes smoked daily)			
	Age 12–13	Age 14–15	Age 16–17	Age 22–32
Total sample	0.0	6.7	9.8	28.8
MZ males	0.0	6.8	8.5	25.0
MZ females	0.0	5.9	9.5	27.3
DZ males	0.0	6.5	14.8	34.4
DZ females	0.0	6.0	10.6	29.5
DZ opposite sex	0.0	8.2	8.4	31.3

Qualitative and quantitative sex differences. Genetic analyses indicated that no significant qualitative or quantitative sex differences existed in the contribution of genetic or environmental factors to liability of smoking initiation and current quantity smoked, and in the relationship between smoking initiation and current quantity smoked for any of the age groups in this sample (Table 4.2). More specifically, the same genes and environmental factors contributed to the liability of smoking initiation and current quantity smoked in males and females, and genetic and environmental contributions could be equated across sex across ages 14-15, 16-17, and 22-32. (Ages 12-13 were not included

in these analyses due to inadequate sample size and ages 22-32 were combined to ensure adequate sample size for analyses.)

Table 4.2: Model Fit Statistics from CCC Models

Age 14–15	EP	-2LL	*df*	AIC	diffLL	diffdf	*p*
Test for quantitative differences							
1. Different ACE + beta for males and females	20	1,931.19	1,873	-1,814.81	—	—	—
2. Equated parameters	13	1,934.93	1,880	-1,825.07	3.75	7	0.81
CCC model fit comparisons							
2. Equated parameters across sex	13	1,934.93	1,880	-1,825.07	—	—	—
3. Model 2 + dropped beta	12	1,935.67	1,881	-1,826.33	0.74	1	0.39
4. Model 3 + dropped A from initiation	11	1,947.25	1,882	-1,816.75	12.32	2	0.00
5. Model 3 + dropped A from current quantity smoked	11	1,936.07	1,882	-1,827.93	1.14	2	0.57
6. Model 3 + dropped C from initiation	11	1,938.78	1,882	-1,825.22	3.84	2	0.15
7. Model 3 + dropped C from current quantity smoked	11	1,939.71	1,882	-1,824.29	4.78	2	0.09
8. Model 5 + dropped C from initiation	10	1,939.18	1,883	-1,826.82	4.24	3	0.24
9. Model 5 + dropped C from current quantity smoked	10	1,977.13	1,883	-1,788.87	42.2	3	0.00
Age 16–17							
Test for quantitative differences							
1. Different ACE + beta for males and females	20	2,096.96	1,858	-1,619.04	—	—	—
2. Equated parameters	13	2,099.69	1,865	-1,630.31	2.72	7	0.91
CCC model fit comparisons							
2. Equated parameters across sex	13	2,099.69	1,865	-1,630.31	—	—	—
3. Model 2 + dropped beta	12	2,099.75	1,866	-1,632.25	0.05	1	0.81
4. Model 3 + dropped A from initiation	11	2,112.19	1,867	-1,621.81	12.51	2	0.00
5. Model 3 + dropped A from current quantity smoked	11	2,101.05	1,867	-1,632.95	1.36	2	0.51
6. Model 3 + dropped C from initiation	11	2,103.47	1,867	-1,630.53	3.78	2	0.15
7. Model 3 + dropped C from current quantity smoked	11	2,108.69	1,867	-1,625.31	9.00	2	0.01
8. Model 3 + dropped C from initiation + A from current quantity smoked	10	2,104.77	1,868	-1,631.23	5.09	3	0.17
Age 22–32							
Test for quantitative differences							
1. Different ACE + beta for males and females	20	2,554.96	1,903	-1,251.04	—	—	—
2. Equated parameters	13	2,554.82	1,910	-1,265.18	-0.15	7	1.00
CCC model fit comparisons							
2. Equated parameters across sex	13	2,554.82	1,910	-1,265.18	—	—	—
3. Model 2 + dropped beta	12	2,559.40	1,911	-1,262.60	4.59	1	0.03
4. Model 2 + dropped A from initiation	12	2,567.20	1,911	-1,254.80	12.39	1	0.00
5. Model 2 + dropped A from current quantity smoked	12	2,558.10	1,911	-1,263.90	3.29	1	0.07
6. Model 2 + dropped C from initiation	12	2,554.82	1,911	-1,267.18	0.00	1	1.00
7. Model 2 + dropped C from current quantity smoked	12	2,555.28	1,911	-1,266.72	0.46	1	0.50
8. Model 2 + dropped A from current quantity smoked and C from initiation	11	2,558.37	1,912	-1,265.63	3.55	2	0.17
9. Model 2 + dropped A from current quantity smoked and C from current quantity smoked	11	2,574.40	1,912	-1,249.60	19.58	2	0.00
10. Model 2 + dropped C from initiation and current quantity smoked	11	2,555.28	1,912	-1,268.72	0.46	2	0.79

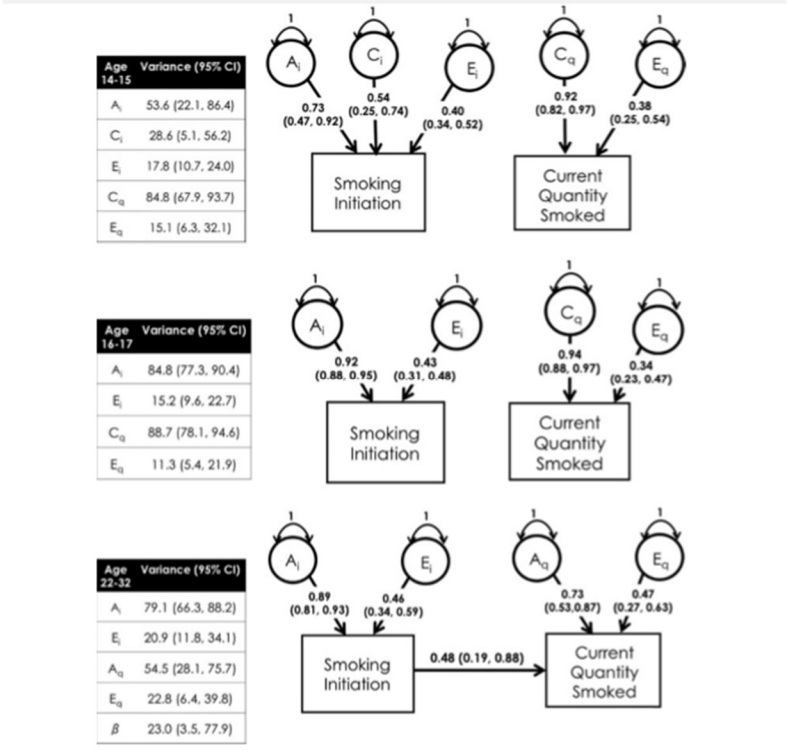
Note: EP indicates the number of estimated parameters in the model.

Relationships between smoking initiation and current quantity smoked. The relationship between smoking initiation and current quantity smoked could not be assessed for ages 12-13, due to inadequate sample size for the smoking quantity variable. Instead, univariate genetic analysis was conducted on the smoking initiation variable. The best fitting model for this age group did not include additive genetic factors, suggesting that common environmental (71.7%; 95% CI: 58.7%, 81.8%) and unique

environmental factors (28.2%; 95% CI: 18.2%, 41.2%) best explained the variance in smoking initiation at age 12 to 13 years.

Across ages 14-15 and 16-17 years, dropping the beta parameter from the CCC model did not result in significantly worse model fit. This implied that smoking initiation and smoking quantity had independent liabilities at these age groups. The best fitting models were not the same across these age groups, however. For age 14 to 15 years, the best fitting models were an ACE model for smoking initiation and a CE model for current quantity smoked, suggesting that genetic (53.5%) and environmental factors (shared: 28.6%; unique: 17.8%) contributed to smoking initiation while environmental factors contributed to current quantity smoked (shared: 84.8%; unique: 15.2%), as measured by quantity smoked. For age 16 to 17 years, the best fitting model for smoking initiation was an AE model, while a CE model still fitted the data best for current quantity smoked, suggesting that genetic (84.8%) and unique (15.2%) environmental factors contributed to smoking initiation, while environmental factors (shared: 88.7%; unique: 11.3%) contributed to current quantity smoked (Figure 4.1).

Figure 4.1: Best Fitting CCC Models and Variance Component Estimates



For ages 22-32 years, the beta parameter between the initiation and current quantity smoked stages was significant, and the best fitting model was an AE model for both initiation and current quantity smoked. This suggested that smoking initiation and smoking current quantity smoked shared liabilities to a moderate extent (β = 0.48) and was no longer independent, as with the earlier age groups. Additionally, genetic and unique environmental factors contributed to both smoking initiation and current quantity smoked, but shared environmental factors no longer exerted a signification impact on liability to smoking. Thus, of the genetic variance in liability to current quantity smoked, approximately 77.3% of the genetic variance was specific to current quantity smoked and 23.0% was shared with smoking initiation. In other words, mostly different genetic factors contributed to the liabilities of smoking initiation and current quantity smoked across

adolescence, but in young adulthood, there was some overlap between the factors influencing initiation and current quantity smoked.

DISCUSSION

No qualitative or quantitative differences were found between males and females regarding the genetic and environmental influences on individual differences in smoking initiation and current quantity smoked across adolescence into early adulthood, lending support for similar findings in other studies[6,18]. However, at age 22-32, when testing for qualitative sex differences, models constraining the genetic correlation to 1, indicating the same genes influence smoking initiation and current quantity smoked in males and females, fitted the data only slightly better than models that allowed for the free estimation of the genetic correlation. Thus, it is possible that qualitative sex differences do exist in later adulthood and that we did not have the power to detect them in the current sample. This might explain why other studies utilizing adult samples have found qualitative sex differences in the genetic and environmental influences in smoking behavior[8,16,21,27].

Unfortunately, due to sample size constraints, we were unable to determine whether genetic or environmental factors contributed more significantly during the earliest ages of adolescence (ages 12-13). However, we did find that different factors contribute to smoking initiation and current quantity smoked across mid-adolescence into early adulthood. More specifically, smoking initiation and current quantity smoked seemed to have independent liabilities until adulthood, when liabilities were shared. Genetic, shared and unique environmental factors were found to significantly contribute to smoking initiation during early adolescence (i.e. ages 14-15), but not during later adolescence (i.e. ages 16-17) or adulthood (i.e. ages 22-32 years), when genetic and unique environmental

factors significantly contribute. Shared environmental influences may be more important for 14-15 year olds relative to older age groups because they experience greater limitations on the access to and availability of cigarettes. Although 14-15 year olds and 16-17 year olds experience the same legal age restriction on the purchasing of cigarettes, 14-15 year olds might still have a harder time in gaining access to cigarettes among their peer groups if they have fewer friends who are of the legal age to buy cigarettes.

Additionally, genetic influences were not found to contribute significantly to smoking initiation until later adolescence into adulthood (beginning at age 14-15 years), much in the same way other studies suggest[9,10,13]. However, contrary to other findings, which find greater genetic influence on heavier/problem substance use, we found that genetic factors do not contribute significantly to the variance in current quantity smoked across all age groups until young adulthood (i.e. ages 22-32 years). Interestingly, it is also during this time that the liabilities of smoking initiation and quantity smoked are no longer independent of one another, but rather correlated. Again, this might be a function of access and availability to cigarettes. As access and availability of cigarettes increase, the expression of genetic predispositions towards increased smoking frequency and potential addiction may also increase, following initiation. Or, it could be the case that using a recent estimate of quantity smoked rather than an estimate from heaviest period of use is less stable and representative of adolescent youth relative to adults, and that our choice of measures for the analysis in this study could influenced the estimate of the variance components.

Limitations and Strengths. Results of this study must be interpreted in the context of a couple of limitations. First, due to low prevalence of smoking behavior among early

adolescents in this sample, the power of the current study was limited. This was apparent when we found that only univariate genetic analysis could be conducted on the smoking initiation variable among 12-13 year olds as there were too many missing values for the current quantity smoked variable and consequently, a CCC model could not be fit. It is also possible that using self-reported data underestimated the prevalence for smoking behaviors, because of social desirability bias, which could have also influenced genetic analysis. Furthermore, this study is not generalizable to all populations, as the sample included only Caucasians.

Despite these limitations, the present study does include both males and females. It is also one of only a few studies investigating the relationship between smoking initiation and current quantity smoked within an adolescent sample and adds to the literature by investigating this relationship across various age groups. Future studies could include the use of measures related to smoking progression, other than current quantity smoked to investigate their effects on the relationship between smoking initiation and current quantity smoked. It would also be interesting to see if the same relationships are found among other adolescent datasets, using different populations than the one described in the present study and if these relationships are affected by the addition of environmental covariates, such as parental monitoring or peer influences.

CHAPTER 5:
A TWIN STUDY OF THE GENETIC AND ENVIRONMENTAL RELATIONSHIP OF STRESSFUL LIFE EVENTS AND SMOKING INITIATION USING THE VIRGINIA TWIN STUDIES OF ADOLESCENT BEHAVIORAL DEVELOPMENT

Elizabeth K. Do and Hermine H. Maes

INTRODUCTION

Tobacco use is the leading cause of preventable morbidity and mortality. Despite substantial decreases in the prevalence of smoking in the last 50 years, on average, one in three adults under the age of 26 continues to use tobacco and nearly half of current users - equivalent to approximately 12.8% of the population in the United States - are nicotine dependent. The direct medical costs due to tobacco-related morbidity totals $130 billion annually[1]. Given this expenditure of public resources, a better understanding of the etiology of tobacco use is essential for the development of effective prevention efforts.

Smoking, in particular, has been identified as a mechanism for coping with stress[2,3] among young users[4] and adults[2,3]. Smoking initiation for adult users typically occurs during adolescence[5], when unhealthy and maladaptive behaviors can develop as a result of life changes and stress[6]. Although smoking initiation may occur because it is fun and pleasurable, it can also be a mechanism by which adolescents cope with stressful life events. Additionally, retrospective accounts of life events reveal that smokers report more stressful life events relative to non-smokers[4,7–9]. Added to this, stressful life events (SLEs) have been associated with smoking initiation and smoking progression[6,10–12]. Specific SLEs, such as adverse childhood experiences[13], parental divorce[14], sexual abuse, a

vulnerable family environment, and parental death during childhood are associated with lifetime history of smoking initiation[15].

Some studies investigating stress and smoking initiation suggest that these effects may be sex limited. For example, one study determined that the effect of early life stress on the risk of smoking initiation by adolescents was only significant among girls at high incidence of stressful life events. Specifically, girls who experienced high levels of stress at 7 years of age had nearly three times higher odds of tobacco use (OR = 2.94, 95% CI = 1.26, 6.83), relative to those that did not. This association was not observed for boys[16]. Another study found that the association between family related stress and smoking initiation seems to be stronger for girls[11].

However, the causes underlying these associations remain unclear and to our knowledge, no genetically informed studies of the association between stressful life events and smoking initiation have been conducted. This study seeks to address this limitation in the literature by examining a large population-based sample of Virginia twins to assess whether SLEs and smoking initiation (SI) are influenced by shared genetic and/or environmental factors. Specifically, we seek to address the following questions: (1) Is there evidence of a shared genetic and/or environmental liability for the association between SLEs and SI, (2) Does the structure of genetic and environmental influences on these traits differ by sex, and (3) Does this structure differs in early adolescence and young adulthood?

MATERIALS & METHODS

Sample. Data were obtained from the Virginia Twin Studies of Adolescent Behavioral Development, which includes the Virginia Twin Study of Adolescent Behavioral Development (VTSABD) and its two follow-up studies, which drew upon participants from the VTSABD: The Young Adult Follow-Up (YAFU) and Transitions to Substance Abuse (TSA) Follow-Up studies. The VTSABD is a cohort-longitudinal epidemiological study of the development and maintenance of childhood psychiatric disorders using a genetic twin design[16]. Adolescent male and female twins aged 8 to 16 years were ascertained through Virginia schools. In the first wave, 1412 Caucasian families participated (2775 individual twins in 1384 complete pairs). Twins under age 18 were followed every 18 months up to 4 times. All twins were targeted for a young adult assessment. 1185 pairs have been followed up in YAFU at a median age of 21 years; 399 pairs (1084 individuals) in TSA at a median age of 25 years. More details regarding the ascertainment of the sample is given elsewhere[17]. To be included in the present study, individual twins had to have provided responses to the Life Experiences Interview regarding questions on stressful life events and smoking initiation. Analytic samples included: 319 monozygotic male (MZM), 418 monozygotic female (MZF), 180 dizygotic male (DZM), 193 dizygotic female (DZF), and 293 dizygotic opposite sex (DZO) twins from the VTSABD; 263 MZM, 347 MZF, 144 DZM, 158 DZF, and 225 DZO twins from the YAFU, and 106 MZM, 201 MZF, 55 DZM, 70 DZF, and 53 DZO twins from the TSA.

Measures. The two variables of interest included smoking initiation and stressful life events. Smoking initiation was measured by the question, “Have you ever smoked, even if you were trying just one cigarette?” Responses included: no (coded as 0), yes (coded

as 1), and no answer provided (coded as missing). Stressful life events included measures regarding major family conflict, divorce/separation, and death. Responses to each of these stressful life events items was: yes (coded as 1), no (coded as 0) or missing. Since the items used to measure stressful life events differed between VTSABD, YAFU, and TSA, we summed relevant item responses and then the variables were categorized to be ordinal.

Statistical Analyses

Prior to conducting analyses, data were cleaned and recoded using SAS 9.3 (SAS Institute Inc., 2003). Data were then prepared for use as raw ordinal data and analyzed using the statistical modeling package OpenMx[18]. This approach assumes that the ordinal categories are representative of an underlying normal distribution of liability, with thresholds in liability discriminating between categories. For twin models, the liability to traits can be attributed to several latent sources of variance: additive genetic factors (A); shared environment (C), or environmental factors that lead to similarity between twin pairs; and non-shared environment (E), or environmental factors that lead to dissimilarity between twin pairs. Estimates of each of these variance components are calculated by comparing the phenotypic correlation between monozygotic twins, who share all their genes, with dizygotic twins, who share half of their genes, on average, identical by descent.

Bivariate analyses examining the association between stressful life events and smoking initiation were conducted separately for data from the VTSABD and its two follow-up studies, the YAFU and TSA. Cholesky bivariate decompositions were used to decompose

the variance of smoking initiation into genetic and environmental influences common with stressful life events and genetic and environmental influences unique to smoking initiation. Using this approach, we were also able to test for both quantitative and qualitative sex-limitation[19] with the following nested models: (1) qualitative for genetic factors and quantitative sex differences, (2) qualitative for shared environmental factors and quantitative sex differences, (3) quantitative sex differences but not qualitative sex differences, and (4) no quantitative and no qualitative sex differences, showed in the figures below.

Figure 5.1a: Quantitative and Qualitative Sex Differences for Genetic Factors: Male and Female Cholesky Paths and Female Specific A Paths

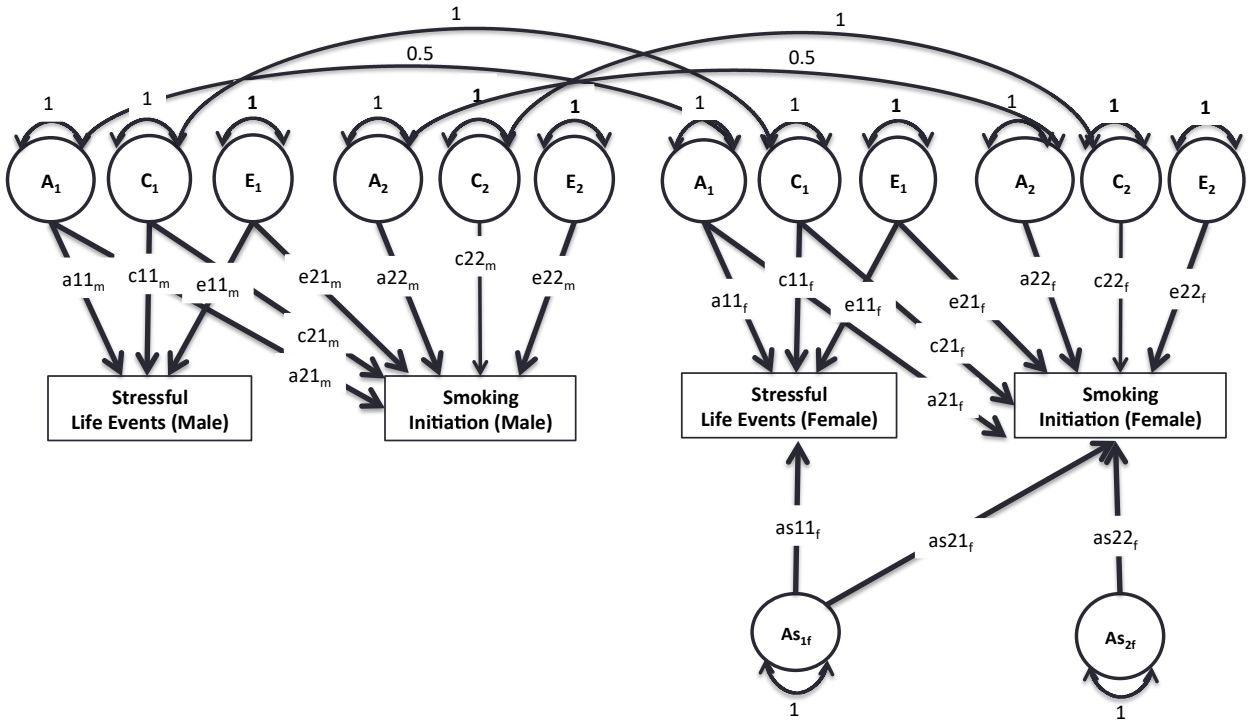


Figure 5.1b: Quantitative and Qualitative Sex Differences for Genetic Factors: Male and Female Cholesky Paths and Male Specific A Paths

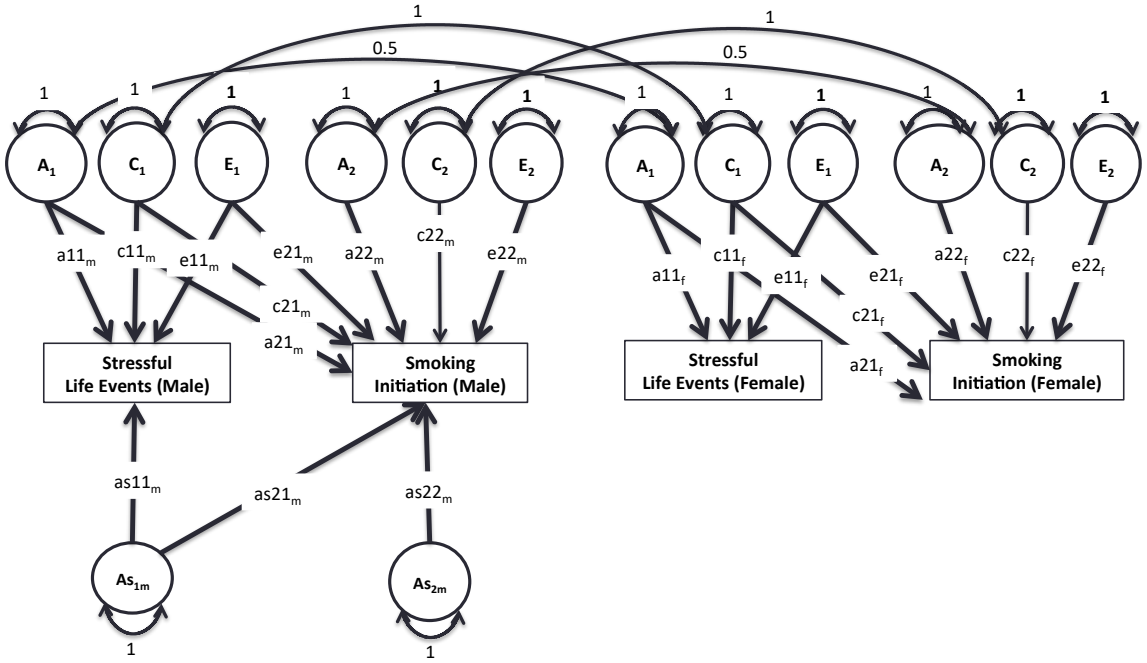


Figure 5.2a: Quantitative and Qualitative Sex Differences for Genetic Factors: Male and Female Cholesky Paths and Female Specific C Paths

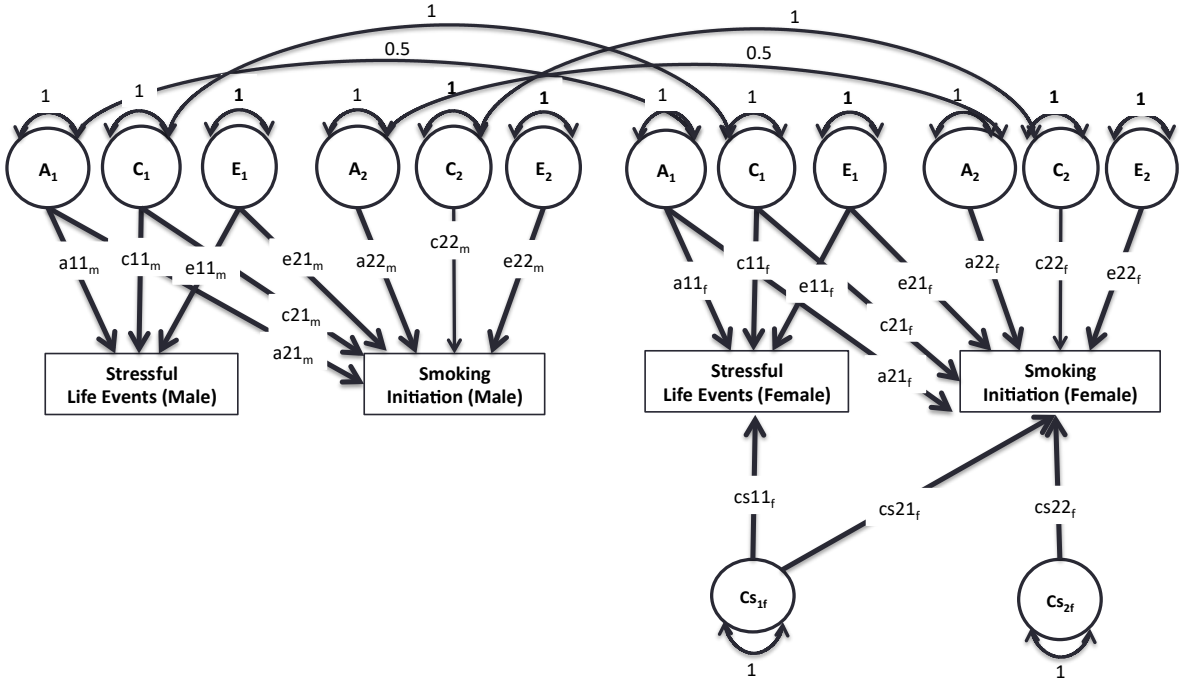


Figure 5.2b: Quantitative and Qualitative Sex Differences for Genetic Factors: Male and Female Cholesky Paths and Male Specific C Paths

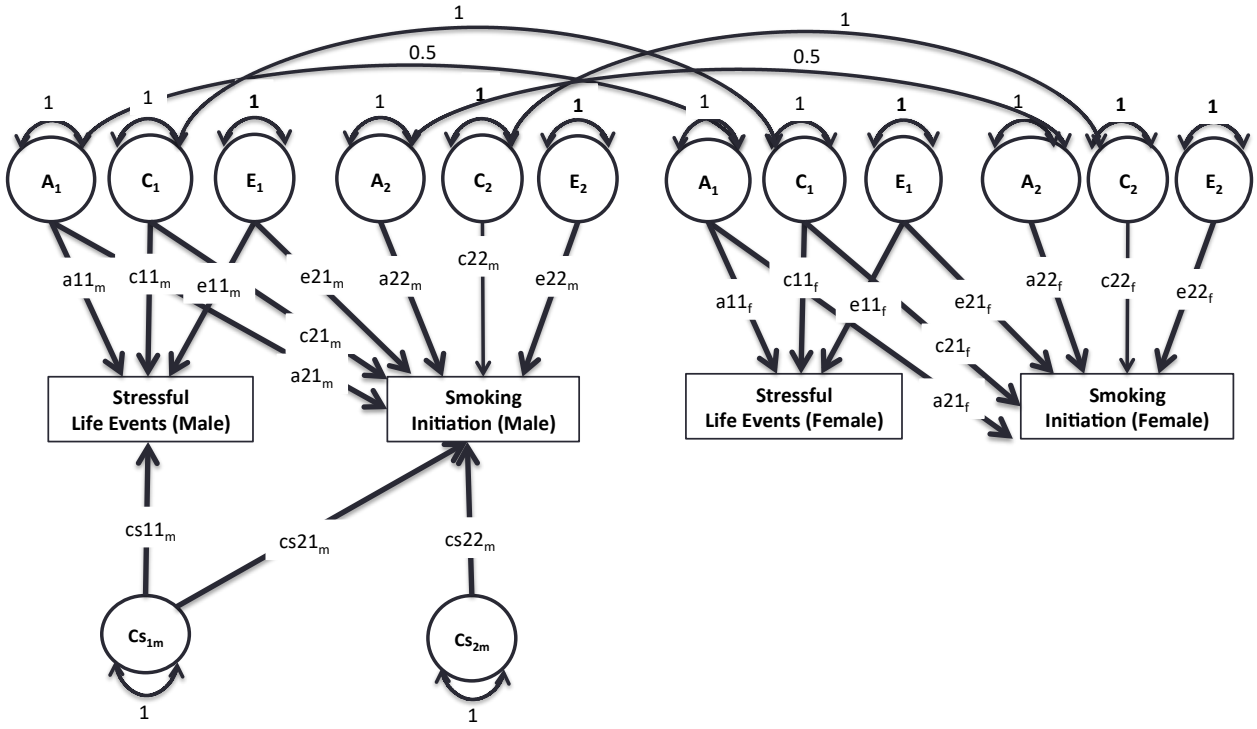


Figure 5.3: Quantitative Sex Differences for Genetic Factors: Male and Female Cholesky Paths without Male/Female Specific Paths

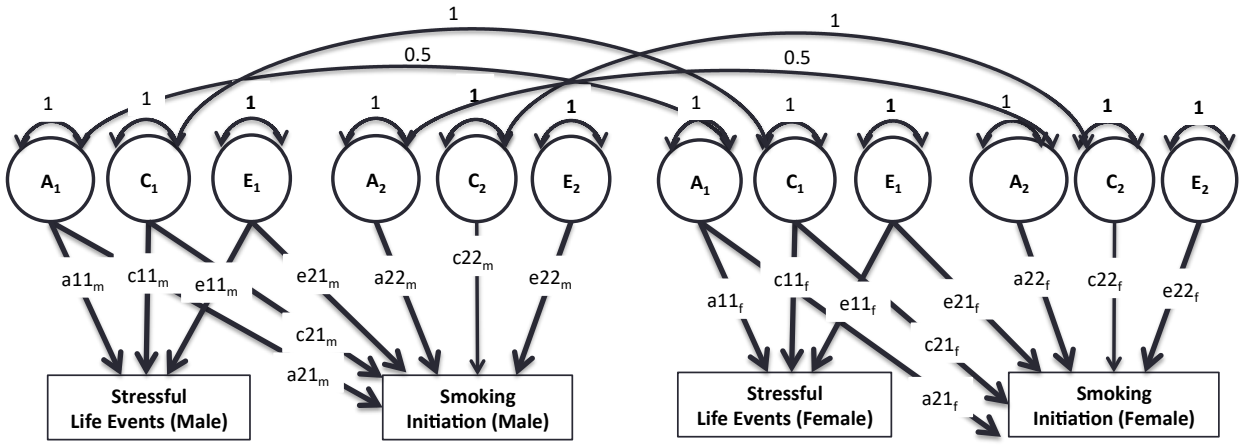
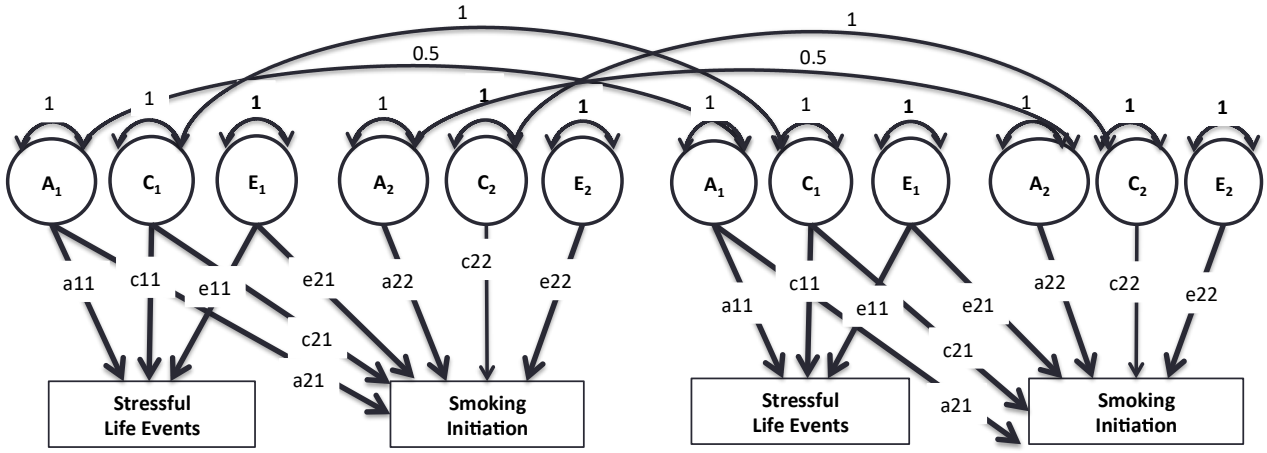


Figure 5.4: Quantitative or Qualitative Sex Differences: One Set of Parameters, Same for Males and Females



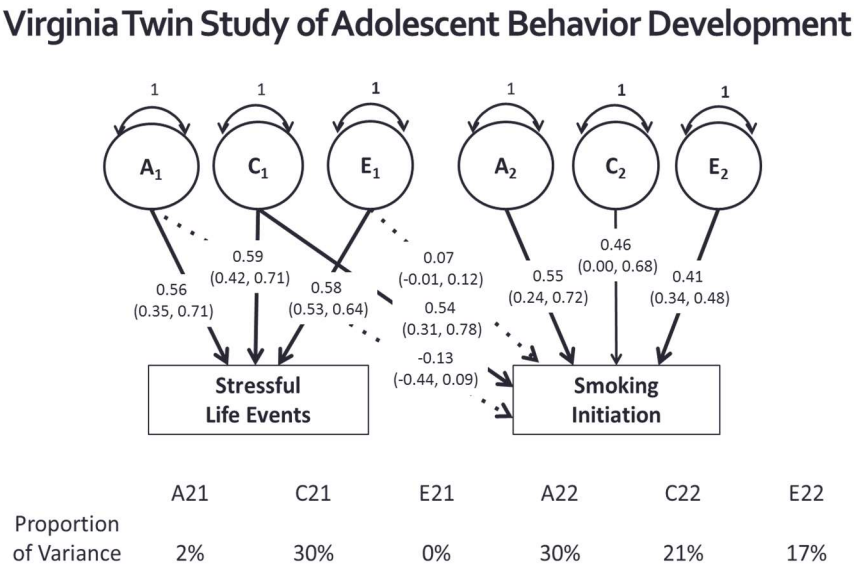
The fit of the nested models was assessed as a function of the change in the value of twice the log likelihood of the data, which is distributed as a chi-square statistic with degrees of freedom equal to the difference in the number of parameters estimated between models. A significant change in Chi Square indicates a significant deterioration in model fit. We also used Akaike's Information Criterion (AIC)[20] to select models, such that a lower AIC value indicates a better balance between the explanatory power of a model and parsimony. We did not test sub models dropping genetic and environmental paths in the Cholesky decomposition models to allow for comparability between the VTSABD, YAFU, and TSA samples.

RESULTS

After testing for quantitative and qualitative sex differences, it was determined that the homogeneity model, reflecting neither quantitative nor qualitative sex differences, fit the data best for the VTSABD, YAFU, and TSA samples (see Figure 5.4). Thus, no sex specific A or C factors significantly contributed to the covariance between stressful life events and smoking initiation, and the parameters between males and females could be

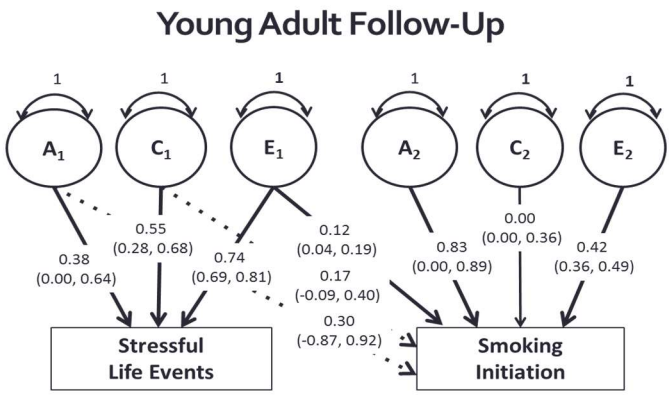
equated to be the same in each final model. In the VTSABD, common genetic (2%) and common shared environmental (30%) influences between stressful life events and smoking initiation contributed to the variance in smoking initiation, while common non-shared environmental influences did not. Genetic (30%), shared environmental (21%) and unshared environmental (17%) influences unique to smoking initiation also contributed to the variance in smoking initiation.

Figure 5.5: Model of Best Fit and Proportions of Variance (VTSABD)



Meanwhile, in the YAFU, common genetic (9%), shared environmental (3%), and non-shared environmental (1%) contributed to the variance in smoking initiation at median age of 21 years. Genetic (69%) and non-shared environmental (18%) influences unique to smoking initiation also contributed to the variance in smoking initiation.

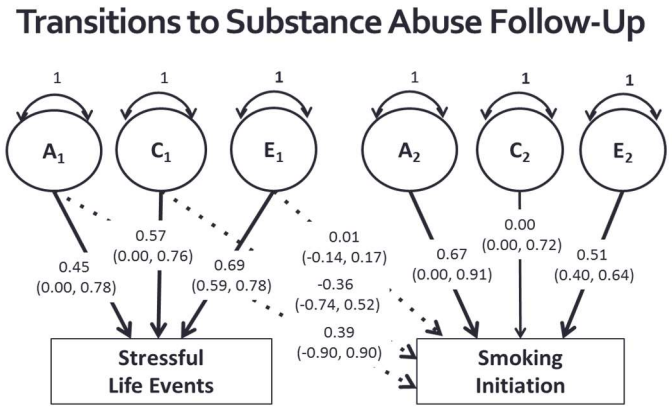
Figure 5.6: Model of Best Fit and Proportions of Variance (YAFU)



	A21	C21	E21	A22	C22	E22
Proportion of Variance	9%	3%	1%	69%	0%	18%

Finally, in the TSA, common genetic (16%) and common shared environmental (13%) influences contributed to the variance in smoking initiation at median age of 25 years. Genetic (45%) and non-shared environmental (26%) influences unique to smoking initiation contributed to the variance in smoking initiation.

Figure 5.7: Model of Best Fit and Proportions of Variance (TSA)



	A21	C21	E21	A22	C22	E22
Proportion of Variance	16%	13%	0%	45%	0%	26%

DISCUSSION

Although many studies have examined the relationship between stressful life events and tobacco use, none have investigated this relationship using a genetically informed sample. To our knowledge, this is the only study that includes stressful life events and smoking initiation. In conducting these analyses, we sought to address three major questions: (1) Is there evidence of a shared genetic and/or environmental liability for the association between SLEs and SI, (2) Does the structure of genetic and environmental influences on these traits differ by sex, and (3) Does this structure differ in early adolescence and young adulthood?

Findings from our study suggest that there is evidence of shared genetic and environmental liability for the association between SLEs and smoking initiation and that this structure differs in early adolescence and young adulthood. Moreover, the structure of genetic and environmental influences on these traits does not seem to differ by sex.

This suggests that the same genes and environments are influencing both stressful life events and smoking initiation in males and females, which differs from previous population based studies that have found stronger associations between specific stressful life events and smoking initiation among females relative to males. These findings may differ because this study aggregated the effects of multiple stressful life events and tested the association between the ordinal measure of stressful life events, rather than testing the association with each specific stressful life event.

We also find that most of the variance in smoking initiation is accounted for by genetic and environmental influences unique to smoking initiation. However, genetic and

environmental influences common to stressful life events and smoking initiation do contribute to the covariance between stressful life events and smoking initiation. While common shared environmental influences contribute more to the variance in smoking initiation during late adolescence (i.e. VTSABD), common genetic influences contribute more to the variance in smoking initiation in young adulthood (i.e. YAFU & TSA). In other words, underlying factors influencing stressful life events and smoking initiation are different for adolescent and young adult twins, such that stressful life events and smoking initiation is influenced by the same, shared environmental factors in young adolescents and by the same, genetic risk factors in young adulthood. This could suggest that once an individual is exposed to the effects of stressful life events, genetic factors come into play, and only individuals with a certain set of genes will initiate tobacco use.

Some shared environmental factors that might be involved in the association between stressful life events and smoking initiation have been identified in previous studies[4,7–9],namely the influence of peers and family. Peer influence is one of the most important determinants of smoking initiation and it is possible that the same peers experiencing stressful life events are involved with the initiation of smoking. Meanwhile, the influence of family members can be seen as both an environmental factor and influenced by genetic factors. For example, parents at high genetic risk for stressful life events may be more likely to show negative parenting behaviors, including substance use, and then transmit their genetic predisposition for substance use to their children.

Like other studies, this study is not without its limitations. Firstly, the analytic sample included only Caucasians from the Mid-Atlantic region of the United States. This limits the generalizability of the findings and future studies need to be conducted to confirm whether

the same relationships exist in other population samples. Furthermore, the measures of smoking initiation used in the current study were based upon self-report and no secondary validation measures were used. Although assurances that answers would be kept confidential were given to respondents during administration of surveys to prevent bias [and have shown to be effective in studies conducted in the United States[21]], desirability bias could have affected the way in which respondents answered questions regarding their tobacco use. To protect against this bias, future studies may consider validation of smoking behavior measures using biomarkers, such as cotinine. Similarly, different measures for stressful life events were used at each of the survey time points. To overcome this obstacle, we tried to harmonize the measures through the inclusion of related items for stressful life event measures (i.e. including measures on familial conflict, divorce/separation, and death/loss).

Despite these limitations, the study yields novel information in so much that it suggests that stressful life events and smoking initiation are associated among Caucasian adolescents and that both genetic and environmental factors play a role in the covariation between these traits. Future research should consider how these genetic and environmental factors influence the covariation between stressful life events and other tobacco use phenotypes, such as smoking progression and nicotine dependence.

CONCLUSION

This study suggests that stressful life events and smoking initiation are associated due to a common set of environmental factors during young adolescence and common set of genetic factors during young adulthood. Prevention and intervention programs for

smoking initiation should look towards targeting stressful life events, potentially through the implementation of coping strategies with stress and incorporate environmental influences of parents and peers.

CHAPTER 6:
PREVALENCE AND CORRELATES OF NICOTINE DELIVERY SYSTEMS AMONG YOUNG ADULTS IN UNIVERSITY SETTING

Elizabeth K. Do; Megan E. Cooke; Elizabeth C. Prom-Wormley; Danielle M. Dick; Kenneth S. Kendler; Hermine H. Maes

INTRODUCTION

Tobacco use is the leading preventable cause of morbidity and mortality in the United States and worldwide[1]. Among users, it can lead to several adverse outcomes, including nicotine dependence, cancer, and lung disease in later life. Each of these adverse outcomes comes with an associated cost, such as lost productivity and wages due to hospitalization or premature death. This is a key point to consider, especially when thinking about the tobacco use behaviors of young adults. Given the broader range of tobacco products available on the market, the tobacco use behaviors of young adults may be considerably higher than what has been detected by previous studies limited to the use of cigarettes[2,3]. As with cigarette smoking, the use of alternative nicotine delivery systems is associated with serious health problems, such as cardiac, pulmonary and reproductive conditions[4]. Further, their use may hinder efforts to reduce population-level tobacco use and risk for nicotine dependence. This is especially problematic, given the escalating trend for dual (e.g. using cigarettes and an alternative tobacco product) and polytobacco use (e.g. using three or more tobacco products)[5,6]. According to the Tobacco Use Supplements to the Current Population Survey, administered by the U.S. Census Bureau, alternative tobacco use occurs more often in combination with cigarette smoking than in isolation, as demonstrated by 51.3% of college students aged 18-24 years reporting concurrent use[2,7]. Concurrent use has the potential to increase exposure to

nicotine and risk for adverse health effects[8] and nicotine dependence[9]. Concurrent users also tend to ingest more nicotine daily and are less likely to stop using tobacco, relative to individuals who use only one tobacco product[7,10].

Despite public health efforts to decrease its prevalence, tobacco use remains common, especially among young adults. Young adulthood is a critical time during which individuals are exploring and/or solidifying their tobacco use patterns. Currently, the prevalence of occasional use among college students is higher compared to other adult users of tobacco[11]. Since currently available cessation services are not ideally suited for occasional users of tobacco[12,13], a better understanding the factors contributing to the patterns and nuances of tobacco use among college students could aid the design of better prevention and interventions for young adults. Most studies focus on cigarette use, which has declined from 45.5% in 1965 to 24.4% in 2005, and leveled out to about 20% within the United States. Within the past couple of years, the use of alternative tobacco products (e.g. smokeless tobacco, cigars, waterpipe/hookah, and e-cigarettes) has been increasing, particularly among adolescents and young adults.

Factors affecting tobacco use across various age groups include parental autonomy granting, parental involvement, and the experience of stressful life events. Higher levels of autonomy granting are related to lower levels of drug use[14], while associations between parental involvement and smoking suggest that young adults whose parents spend more time with them and communicate with them more frequently are less likely to use tobacco[15]. Meanwhile, studies identify tobacco use as a potential mechanism by which individuals cope with (anticipated) stress[16]. Higher levels of life stress are associated with risk of cigarette smoking initiation within adolescents[17] and retrospective accounts of life

experiences demonstrate that smokers often report more stressful life events, relative to non-smokers[18–20]. Studies of both adolescents and college students demonstrate that increased involvement with smoking is found among those experiencing negative affect, often because of stress[21]. However, it is unclear whether these factors are associated with other nicotine delivery use in young adults as well.

This study evaluates the patterns of use across nicotine delivery systems among young adults in a sample of college students attending an urban four-year university. Specifically, we want to determine: (1) age of first use for each tobacco product or nicotine delivery system, (2) the prevalence of lifetime and current use and whether this differs by sex or race/ethnicity, and (3) the degree to which use of these products are associated with environmental factors, such as parental environment as well as with experiencing stressful life events.

MATERIALS AND METHODS

Sample. Data were collected as a part of Spit for Science, an effort being led by researchers at a large, diverse, public urban university in the Mid-Atlantic region, to understand how genetic and environmental factors come together to contribute to the development of problems associated with the use of substances and emotional health. Incoming freshman (≥18 years) were invited to participate in the longitudinal cohort study, which involves multiple waves of survey data collection, including: two collection periods during the freshman year (fall and spring) and spring follow-ups conducted annually thereafter. Participants were given a $10 incentive for their participation in each of the surveys.

The study is ongoing, and the results from this study reflect the data available for cohorts one through four from Fall 2012 to Spring 2015[22]. Each cohort is followed across time, from freshman year when they are enrolled into the study. All information collected was managed using an electronic data capture tool hosted at the university called Research Electronic Data Capture (REDCap)[23]. Currently, four cohorts have enrolled in the study: 2707 enrolled in Fall 2011, 2481 enrolled in Fall 2012, 2391 enrolled in Fall 2013, and 2310 enrolled in Fall 2014 (Total N = 9,889). DNA collection is also a part of the study protocol, but not involved in the present analyses. The university Institutional Review Board (IRB) approved all Spit for Science protocols.

Measures. Demographic characteristics include race/ethnicity, sex, and age. Race/ethnicity was collected, since the prevalence of tobacco use in young adulthood is expected to differ across race/ethnicity groups. Participants were asked to select from "American Indian/Alaska Native", "Asian", "Black/African American", "Hispanic/Latino", "More than one race", "Native Hawaiian/Other Pacific Islander", "Unknown", "White", or "I choose not to answer" to answer the question: "Which one of these groups' best describes you?" The types of tobacco products and nicotine delivery systems used are also expected to differ by sex, of which participants could select: "male", "female", or "I choose not to answer." The age of respondents ranged from 18 to 35 years, with the mean age of respondents being 19.6 years.

Lifetime and recent use of cigarettes, smokeless tobacco (chaw/dip/snus), cigars (including little cigars and cigarillos), hookah, and e-cigarettes were treated as binary variables (i.e. yes = 1; no = 0).

Age of onset of tobacco use was measured using: “How old were you when you smoked a cigarette or used tobacco for the first time (including just one or two puffs)?” This variable was coded as an ordinal variable, with those who had not initiated smoking coded as zero, those who had initiated after age 18 as 1, gradually increasing to those who initiated prior to age 12 (2: 15-18 years, 3: 12-14 years, 4: <12 years). For cohort 4, measures for tobacco product-specific age of onset were available, asking “How old were you when you [smoked a cigarette/used smokeless tobacco/ smoked cigars, little cigars, cigarillos/ smoked hookah/ used an e-cigarette] for the first time?” Participants responded to this question by inputting their age in years in a free response. Responses that were missing for the age of onset of tobacco use were substituted by measures for tobacco product-specific age of onset.

Exposure to stressful life events prior to university enrollment were calculated as a sum score from five yes/no items. The items asked specifically whether before the past 12 months, the following events happened to the participant: natural disaster (flood, hurricane, tornado, earthquake, fire or explosion), physical assault (being attacked, hit, slapped, kicked, beaten up, shot, or stabbed), sexual assault (rape, attempted rape, made to perform any type of sexual act through force or threat of harm), other unwanted or uncomfortable sexual experience, transportation accident (car accident, boat accident, train wreck, plane crash). Greater values indicate higher levels of stressful life events experienced prior to university and all responses were included in analyses.

Parental involvement was measured using three items: “My parents helped me with schoolwork if there was something I didn’t understand”, “My parents knew who my friends were”, and “My parents spent time just talking with me” from the Steinberg Parental Style

Scale. Each of these items were reverse-coded and then summed. Greater values indicate higher levels of involvement. Only responses where participants answered at least 50% of the items were included in analyses.

Parental autonomy granting was measured using three items: “My parents said I should give in on arguments rather than making people angry”, “My parents told me that their ideas were correct and I should not question them”, and “My parents acted cold and unfriendly if I did something they didn’t like” from the Steinberg Parental Style Scale. Greater values indicate lower levels of autonomy granting. Only responses where participants answered at least 50% of the items were included in analyses.

Statistical Analysis. Frequencies of the relevant variables were described with percentages, while Chi-square tests and independent sample t-tests were used to assess the relationship between age of onset of tobacco use and demographic characteristics, stressful life events experienced prior to attending university, parental involvement, and parental autonomy granting. Tetrachoric and polychoric correlations were used to examine specific and overall tobacco product use by parental environment and stressful life events experienced prior to attending university. Multivariate logistic regression models were fitted to examine the associations between tobacco use (lifetime and recent use) and stressful life events experienced prior to attending university, parental involvement, and parental autonomy granting. Analyses were performed using SAS 9.3 (SAS Institute).

RESULTS

Descriptive characteristics of the sample. The total analytic sample included 9,889 individuals. Females made up 61.5% (N = 6060) of the sample. Approximately half (50.1%; N = 4881) identified as White, 19.2% (N = 1873) identified as Black/African-American, 16.6% (N = 1614) identified as Asian, 6.3% (N = 617) identified as more than one race, 6.1% (N = 594) identified as Hispanic/Latino, 0.7% (N = 67) identified as Native Hawaiian/Other Pacific Islander, 0.5% (N = 51) identified as American Indian/Alaska Native, and 0.4% (N = 39) identified as unknown race.

Age of Onset of Tobacco Use. Over half (68.4%) of students have initiated any tobacco use. Of those who had initiated, 18.4% did not report an age of onset, while more than 40% had initiated by age 18. Of those that initiated by 18, most students started using tobacco products between ages 15 to 18 years (25.6%). On average, the age of onset for smokeless tobacco was 15.82 years, followed by cigarettes at 15.96 years, cigars and hookah at 16.15 and 16.18 years, followed by e-cigarettes at 17.26 years (Table 6.1).

Table 6.1 Tobacco Use – Age of Initiation

Overall Tobacco Use	N	%	
Did Not Initiate	3091	31.26	
Initiated, but did not report age of initiation	1815	18.35	
> 18 years	1109	11.21	
15 – 18 years	2531	25.59	
12-14 years	1094	11.06	
< 12 years	249	2.52	
By Tobacco Product	**N**	**Mean**	**SD**
Cigarettes	662	15.96	2.65
Smokeless	185	15.82	3.43
Cigars	640	16.15	3.14
Hookah	918	16.18	3.81
E-Cigarettes	591	17.26	1.84

Note: Average age of tobacco use initiation was 15 years (SD = 1.2 years)

Changes in the prevalence of lifetime tobacco use. Figure 6.1 shows the prevalence in lifetime use of tobacco by nicotine delivery system (e.g. cigarettes, snus, cigars, hookah, and e-cigarettes), separated by sex and time in college by cohort. Generally, the lifetime use of tobacco products increased as the time enrolled in college increased for both males and females. On average, 30% of females and 40% of males had ever used cigarettes; less than 10% of females and 20% had ever used snus; and 50 % of males and females had ever used hookah in their lifetime, by their freshman year of college. While the percentages endorsing lifetime use of cigars, and hookah remained consistent from freshman year to senior year, the endorsement of lifetime use of cigarettes increased across all cohorts. Interestingly, the lifetime use of snus and e-cigarettes differed by cohort; however, this might be a function of availability and when this item was measured in the survey. Figure 6.2 demonstrates this point by displaying the prevalence of lifetime tobacco use by nicotine delivery system, as a function calendar time, separated by sex.

Figure 6.1: Prevalence of Lifetime Tobacco Use by Nicotine Delivery System, Sex, and Time in College – Separated by Cohort

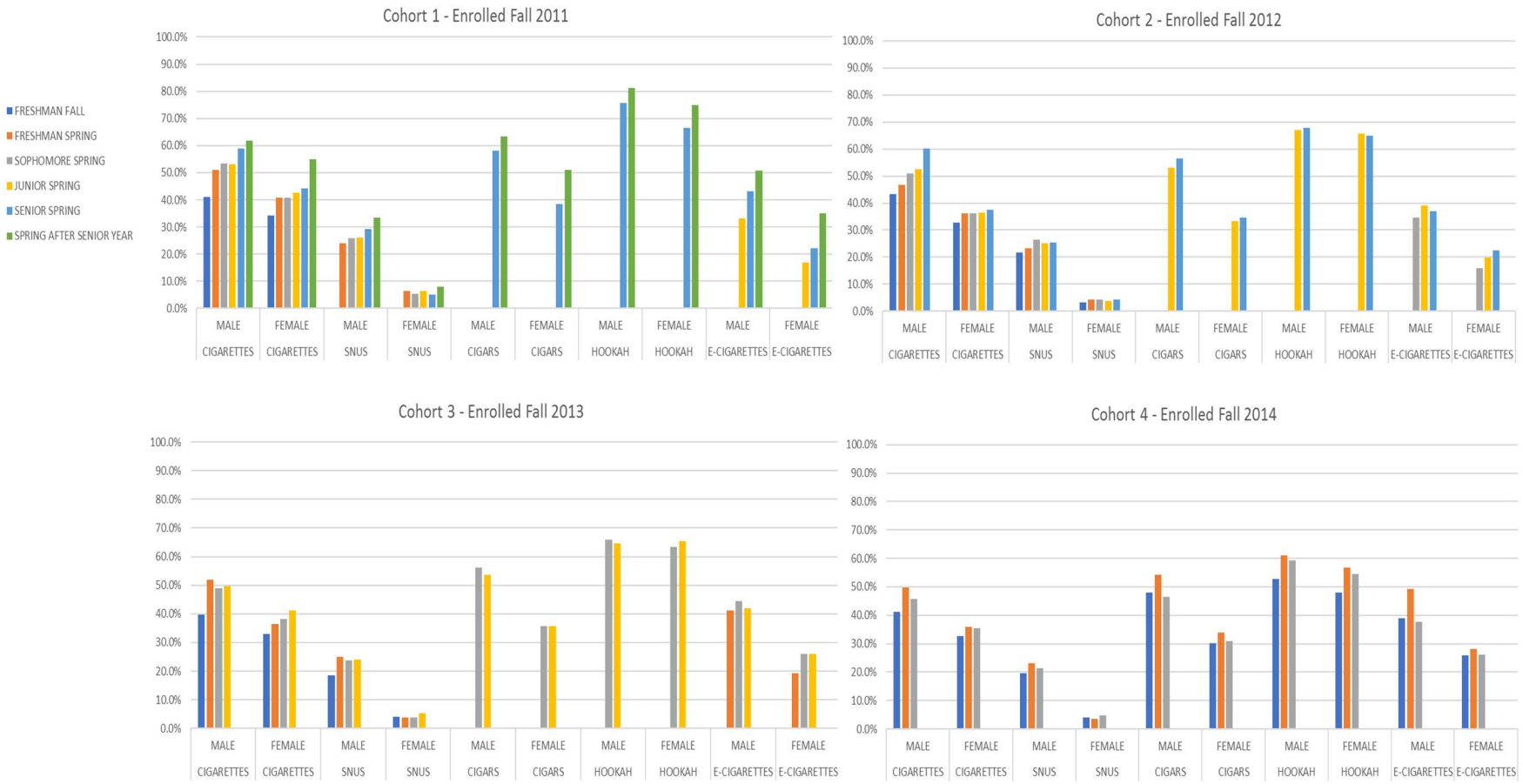
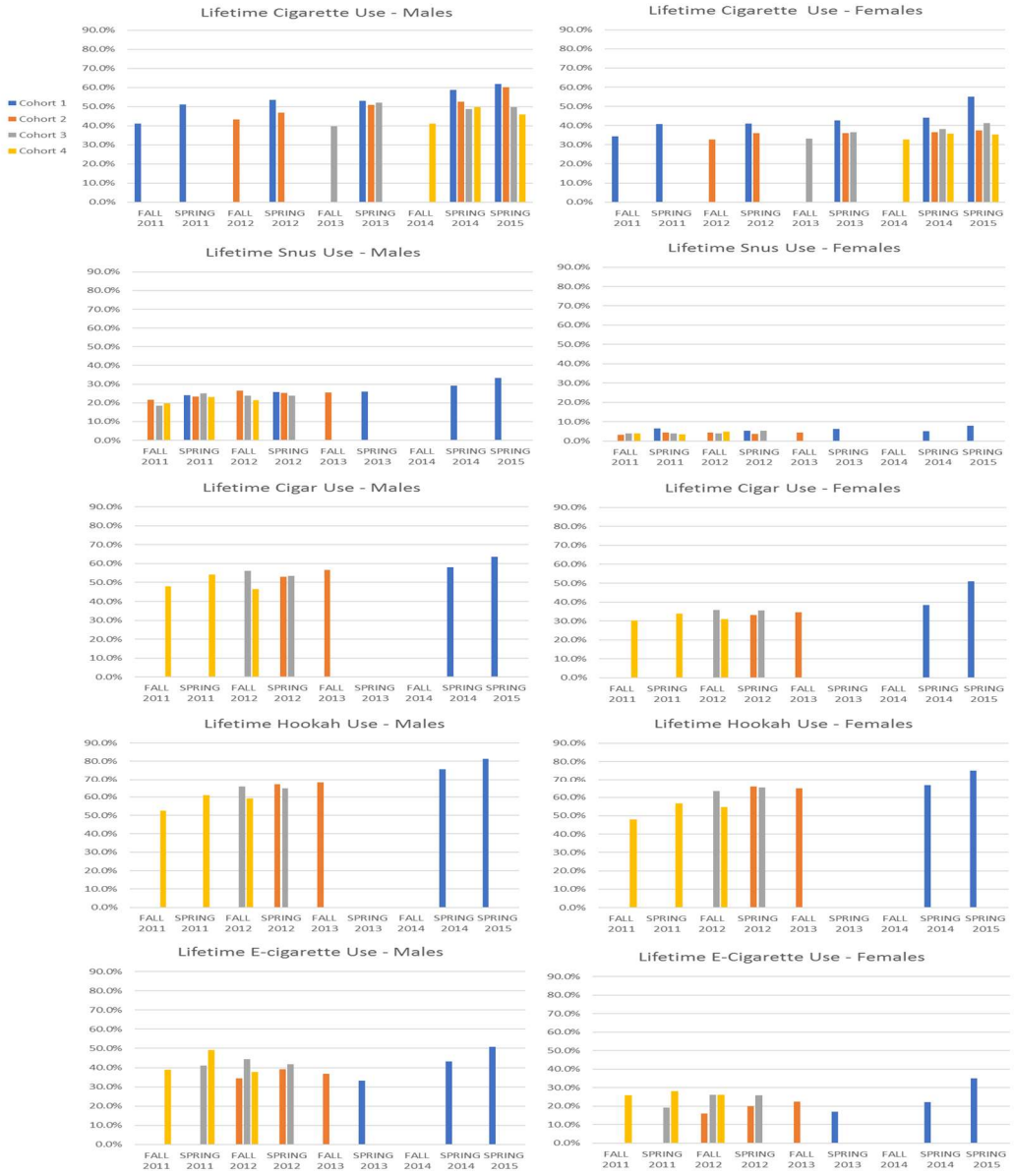


Figure 6.2: Prevalence of Lifetime Tobacco Use by Nicotine Delivery System and Time in College – Separated by Sex



Tobacco use prevalence and differences by sex and race/ethnicity. Aggregated measures of lifetime tobacco use from Table 6.2 demonstrates that the lifetime prevalence of tobacco use was highest for hookah (49.6%), followed by cigarettes (49.2%), cigars (34.9%), e-cigarettes (25.5%), and smokeless tobacco (15.1%). Overall, a higher percentage of males reported using tobacco, relative to females, across each tobacco product. The largest difference between the percentages of males vs. percentage of females using a tobacco product was found for smokeless tobacco, where 27.8% of males and 7.2% of females endorsed ever using smokeless tobacco ($\mathbf{X^2}$ = 767.46, df = 1, p-value <0.0001) (Table 6.2). The prevalence of lifetime use of tobacco products differed by race/ethnicity (Table 6.3).

The tobacco product with the highest prevalence for recent use was cigarettes (48.1%), followed by e-cigarettes (27.6%), hookah (24.3%), smokeless tobacco (18.6%), and cigars (18/3%). Looking at recent tobacco use by sex, we found that a significantly higher percentage of males used tobacco products, relative to females across all tobacco products. More than 40% of students have indicated that they used tobacco products in the past 30 days, at the time of their most recent follow-up survey (Table 6.2). The prevalence of recent use of tobacco products was similar across race/ethnicity, for hookah/waterpipe and e-cigarettes (Table 6.3).

Table 6.2: Lifetime and Recent Tobacco Use by Sex

	Male	Female	Total	$X_{df=1}$value, p-value
Lifetime Tobacco Use	**n/total n (%)**	**n/total n (%)**	**n/total n (%)**	
Overall Tobacco Use	2705/3793 (71.3)	4036/6060 (66.6)	6741/9853 (68.4)	23.99, <0.0001*
Cigarettes	2113/3793 (55.7)	2731/6060 (45.1)	4844/9853 (49.2)	105.71, <0.0001*
Smokeless	1054/3793 (27.8)	438/6060 (7.2)	1492/9853 (15.1)	767.46, <0.0001*
Cigars	1645/3793 (43.4)	1795/6060 (29.6)	3440/9853 (34.9)	194.06, <0.0001*
Hookah	1895/3793 (50.0)	2995/6060 (49.4)	4890/9853 (49.6)	0.27, 0.6033
E-Cigarettes	1199/3793 (31.6)	1312/6060 (21.7)	2511/9853 (25.5)	121.88, <0.0001**
Recent Tobacco Use	**n/total n (%)**	**n/total n (%)**	**n/total n (%)**	
Overall Tobacco Use	1880/3520 (53.4)	2164/5599 (38.7)	4044/9119 (44.4)	190.76, <0.0001*
*Cigarettes**	1167/2139 (54.6)	1195/2758 (43.3)	2362/4897 (48.2)	60.84, <0.0001*
*Smokeless**	334/1289 (25.9)	86/958 (9.0)	420/2247 (18.7)	103.70, <0.0001*
Cigars	874/3346 (26.1)	679/5148 (13.2)	1553/8494 (18.3)	226.96, <0.0001*
Hookah	993/3411 (29.1)	1165/5455 (21.4)	2158/8866 (24.3)	68.54, <0.0001*
*E-Cigarettes**	419/1203 (34.8)	301/1419 (21.2)	720/2622 (27.5)	60.61, <0.0001*

* skip pattern – if indicated no lifetime use, then skipped question

Table 6.3: Lifetime and Recent Tobacco Use by Race/Ethnicity

	1	2	3	4	5	6	7	Total	$X_{df=7}$ value, p-value
Lifetime Tobacco Use									
Any Tobacco Use *N = 9736*	3501/4881 (71.7)	1221/1873 (65.2)	930/1614 (57.6)	434/594 (73.1)	462/617 (74.9)	50/67 (74.6)	33/51 (64.7)	6659/9736 (68.4)	140.37, <0.0001*
Cigarettes *N = 9736*	2827/4881 (57.9)	610/1873 (32.6)	605/1614 (37.5)	328/594 (55.2)	332/617 (53.8)	40/76 (59.7)	21/51 (41.2)	4787/9736 (49.2)	464.88, <0.0001*
Smokeless *N = 9736*	936/4881 (19.2)	146/1873 (7.8)	181/1614 (11.2)	88/594 (14.8)	91/617 (14.8)	15/67 (22.4)	11/51 (21.6)	1475/9736 (15.2)	164.56, <0.0001*
Cigars *N = 9736*	1890/4881 (38.7)	621/1873 (33.2)621	368/1614 (22.8)	208/594 (35.0)	239/617 (38.7)	31/67 (46.3)	20/51 (39.2)	3396/9736 (34.9)	149.43, <0.0001*
Hookah *N = 9736*	2377/4881 (48.7)	970/1873 (51.8)	732/1614 (45.4)	337/594 (56.7)	329/617 (53.3)	39/67 (58.2)	24/51 (47.1)	4827/9736 (49.6)	34.46, <0.0001*
E-Cigarettes *N = 9736*	1453/4881 (29.8)	284/1873 (15.2)	347/1614 (21.5)	171/594 (28.8)	177/617 (28.7)	19/67 (28.4)	17/51 (33.3)	2478/9736 (25.4)	174.58, <0.0001*
Recent Tobacco Use									
Any Tobacco Use *N = 9008*	2216/4544 (48.8)	637/1746 (36.5)	524/1436 (36.5)	261/561 (46.5)	281/579 (48.5)	281/579 (48.5)	23/46 (50.0)	2216/4544 (48.8)	126.4, <0.0001*
*Cigarettes** *N = 4837*	1518/2847 (53.3)	211/624 (33.8)	251/619 (40.6)	144/332 (43.4)	159/331 (48.0)	19/39 (48.7)	12/22 (54.5)	2328/4837 (48.1)	101.1, <0.0001*
*Smokeless** *N = 2218*	257/1345 (19.1)	44/271 (16.2)	59/271 (21.8)	23/150 (15.3)	20/140 (14.3)	1/21 (4.8)	4/10 (40.0)	412/2218 (18.6)	14.5, 0.0426*
Cigars *N = 8389*	790/4293 (18.4)	327/1614 (20.3)	169/1305 (13.0)	96/505 (19.0)	110/537 (20.5)	16/59 (27.1)	11/43 (25.6)	1531/8389 (18.3)	42.9, <0.0001*
Hookah *N = 8759*	1045/4389 (23.8)	395/1713 (23.1)	352/1406 (25.0)	150/547 (27.4)	148/567 (26.1)	18/61 (29.5)	14/43 (32.6)	2131/8759 (24.3)	9.0, 0.2553
*E-Cigarettes** *N = 2587*	429/1491 (28.8)	64/303 (21.1)	103/387 (26.6)	52/179 (6.9)	51/185 (27.6)	4/18 (22.2)	8/16 (50.0)	713/2587 (27.6)	12.08, 0.0980

1=White/Caucasian; 2 = Black/African American; 3 = Asian; 4 = Hispanic/Latino; 5 = More than one race; 6 = Native Hawaiian/Other Pacific Islander; 7 = American Indian/Alaska Native

* skip pattern – if indicated no lifetime use, then skipped question (explains why so many missing)

More than 40% of students have used two or more tobacco products within their lifetime. Among recent tobacco users, 49.4% used only one tobacco product, 14.3% used two tobacco products, 3.3% used three tobacco products, and 0.8% used four or more tobacco products concurrently. Among those who only used one tobacco product, hookah use was the most popular (44.4%), followed by cigarettes (22.3%), and cigars (22.0%). A combination of cigarettes, hookah, cigars, and e-cigarettes was common among those who reported using two products or more concurrently.

In the total sample, 15.5% had ever smoked ≥100 cigarettes in their lifetime. Relative to females, more males reported smoking at least 100 cigarettes or more in their lifetime (21.6% of males vs. 11.7% of females; $\mathbf{X^2}$ = 202.1, df = 2, p-value <0.001). Differences in the reported prevalence of lifetime smoking of 100 or more cigarettes were also found across race/ethnicity ($\mathbf{X^2}$ = 605.7, df = 14, p-value <0.001). Compared to White/Caucasians, fewer Blacks/African Americans, Asians, and Hispanics/Latinos reported smoking ≥100 cigarettes in their lifetime.

Of the full sample, 4.9% (N = 487) met criteria for nicotine dependence, when using a cut-off score of four for the Fagerström Test for Nicotine Dependence. Males were more likely to meet criteria for nicotine dependence, relative to females (7.4% vs. 3.4% respectively; $\mathbf{X^2}$ = 83.3, df = 1, p-value <0.0001).

Lifetime tobacco use, parental involvement, and autonomy granting. Parental involvement was negatively correlated with the use of each tobacco product, while parental autonomy granting was positively correlated with the use of any tobacco product, cigarettes, smokeless tobacco, hookah, and e-cigarettes.

Table 6.4: Lifetime and Recent Tobacco Use Correlations

	Parental Involvement	Parental Autonomy Granting	Natural Disaster	Physical Assault	Sexual Assault	Other Unwanted Sexual Experiences	Transportation/ Accident	Any Stressful Life Event
Lifetime Tobacco Use								
Any Tobacco	-0.023 (-0.054, 0.007)	**0.058 (0.029, 0.088)**	-0.016 (-0.090, 0.059)	**0.197 (0.114, 0.280)**	**0.177 (0.063, 0.290)**	**0.185 (0.095, 0.275)**	**0.076 (0.001, 0.152)**	0.038 (-0.041, 0.116)
Cigarettes	**-0.049 (-0.078, -0.020)**	**0.049 (0.021, 0.078)**	-0.031 (-0.102, 0.040)	**0.284 (0.208, 0.359)**	**0.264 (0.162, 0.368)**	**0.257 (0.175, 0.339)**	**0.096 (0.024, 0.167)**	0.069 (-0.006, 0.144)
Smokeless	**-0.065 (-0.101, -0.030)**	**0.044 (0.009, 0.055)**	-0.030 (-0.115, 0.056)	**0.241 (0.152, 0.330)**	0.007 (-0.121, 0.135)	0.057 (-0.045, 0.159)	**0.101 (0.015, 0.187)**	0.062 (-0.029, 0.153)
Cigars	**-0.033 (-0.063, -0.003)**	0.029 (-0.001, 0.059)	0.044 (-0.030, 0.118)	**0.203 (0.123, 0.282)**	**0.138 (0.031,0.245)**	**0.146 (0.060, 0.232)**	0.070 (-0.005, 0.144)	**0.086 (0.008, 0.164)**
Hookah	**-0.031 (-0.060, -0.001)**	**0.058 (0.030, 0.087)**	-0.028 (-0.100, 0.042)	**0.082 (0.002, 0.161)**	**0.126 (0.020, 0.231)**	**0.126 (0.041, 0.210)**	0.010 (-0.062, 0.082)	0.00 (-0.075, 0.075)
E-Cigarettes	**-0.042 (-0.076, -0.009)**	**0.062 (0.030, 0.095)**	-0.016 (-0.093, 0.061)	**0.146 (0.063, 0.230)**	**0.251 (0.146, 0.366)**	**0.203 (0.116, 0.291)**	0.022 (-0.055, 0.100)	0.027 (-0.055, 0.108)
Recent Tobacco Use								
Any Tobacco	**-0.053 (-0.082, -0.023)**	**0.064 (0.035, 0.093)**	-0.21 (-0.095, 0.053)	**0.225 (0.145, 0.304)**	**0.168 (0.060, 0.275)**	**0.173 (0.087, 0.260)**	0.064 (-0.010, 0.139)	0.052 (-0.026, 0.131)
Cigarettes	**-0.048 (-0.088, -0.007)**	0.029 (-0.011, 0.069)	-0.026 (-0.126, 0.073)	**0.226 (0.124, 0.328)**	**0.165 (0.032, 0.299)**	**0.119 (0.007, 0.231)**	0.038 (-0.061, 0.139)	0.056 (-0.051, 0.162)
Smokeless	-0.043 (-0.113, 0.026)	-0.007 (-0.075, 0.061)	-0.132 (-0.294, 0.029)	-0.003 (-0.176, 0.169)	-0.072 (-0.321, 0.176)	-0.06 (-0.53, 0.133)	-0.098 (-0.262, 0.066)	-0.069 (-0.242, 0.134)
Cigars	**-0.061 (-0.096, -0.026)**	**0.041 (0.006, 0.076)**	0.010 (-0.080, 0.099)	**0.242 (0.150, 0.334)**	**0.148 (0.014, 0.269)**	0.048 (-0.058, 0.154)	0.054 (-0.036, 0.144)	0.076 (-0.020, 0.171)
Hookah	**-0.061 (-0.093, -0.028)**	**0.072 (0.041, 0.104)**	0.031 (-0.052, 0.113)	**0.115 (0.025, 0.205)**	0.002 (-0.121, 0.125)	0.046 (-0.053, 0.144)	0.020 (-0.063, 0.104)	-0.009 (-0.097, 0.078)
E-Cigarettes	-0.058 (0.126, 0.009)	-0.007 (-0.074, 0.059)	-0.004 (-0.150, 0.143)	0.069 (-0.087, 0.224)	-0.044 (-0.126, 0.038)	-0.063 (-0.226, 0.101)	0.047 (-0.099, 0.195)	-0.012 (-0.168, 0.145)

Boldface indicates significance at p-value ≤0.05.

Significant correlations were small in effect (Table 6.4). For example, negative correlations between parental involvement and lifetime use of tobacco products ranged from -0.065 (95%CI: -0.101, -0.030) to -0.031 (95% CI: -0.060, -0.001). Meanwhile, parental autonomy granting was statistically, positively correlated with the lifetime use of any tobacco product, cigarettes, smokeless tobacco, hookah, and e-cigarettes. These significant, positive correlations ranged from 0.044 (95% CI: 0.021, 0.078) to 0.062 (95% CI: 0.030, 0.95).

Lifetime tobacco use, physical and sexual assault. The experience of natural disaster was not significantly correlated with the use of any tobacco product (Table 6.4). However, the experience of physical assault, sexual assault, other unwanted sexual experiences or transportation/accident was significantly, positively associated with the lifetime use of tobacco products, except for the association between: sexual assault and other unwanted sexual experiences and smokeless tobacco use, and transportation/accident and lifetime cigar, hookah, and e-cigarette use. On average, these correlations ranged from 0.082 (95% CI: 0.002, 0.161) to 0.264 (95% CI: 0.162, 0.368).

Recent tobacco use, parental involvement and autonomy granting. Correlations between parental environment and recent tobacco use followed a similar pattern to those found between parental environment and lifetime tobacco use, such that significant correlations between parental involvement and recent tobacco use were negative and significant correlations between parental autonomy granting and recent tobacco use were positive. Significant negative correlations were found between parental involvement and the recent use of any tobacco product, cigarettes, cigars, and hookah, which ranged from -0.068 (95% CI: -0.082, -0.023) to -0.048 (95% CI: -0.083, -0.007). Significant positive

correlations were found between parental autonomy granting and the recent use of any tobacco product, cigars, and hookah, which ranged from 0.041 (95% CI: 0.006, 0.078) to 0.084 (95% CI: 0.036, 0.093).

Recent tobacco use, physical assault and sexual assault. The experience of physical assault, sexual assault, and other unwanted sexual experiences was correlated with the recent use of any tobacco product, and cigarette use. The experience of physical assault was also correlated with the recent use of cigars and hookah, while the experience of sexual assault was correlated with the recent use of cigars. However, the significant correlations were slightly weaker, relative to lifetime measures of tobacco use, ranging from 0.116 (95%CI: 0.025, 0.206) to 0.242 (95% CI: 0.170, 0.334).

DISCUSSION

Despite public health successes in reducing the consumption of cigarettes, the increasing popularity of alternative tobacco products poses new challenges. More than 40% of students reported tobacco use in the past 30 days at the time of their most recent follow-up survey. This is important since studies demonstrate that alternative tobacco products contribute to negative outcomes, similar to the use of cigarettes. For example, smokeless tobacco is addictive and its use has been associated with an increased risk of cancer; myocardial infarction and stroke; oral disease; and reproductive problems[24]. Hookah, or waterpipe, tobacco smoking has also been previously associated with higher odds of lung cancer, respiratory illness, low birth weight, and periodontal disease[25]. Cigar smoke contains higher concentrations of toxic and carcinogenic compounds relative to cigarettes; does not reduce the risk of nicotine addiction, and is known to cause cancers of the lung and upper aero digestive tract[26]. Additionally, individuals who use alternative

tobacco products may face the same challenges in quitting and experiencing negative tobacco-related health consequences in their futures, as those who only smoke cigarettes. Furthermore, concurrent use of tobacco products and cigarettes may make it more difficult to quit tobacco use overall[24].

It remains unclear whether alternative forms of tobacco use may serve as an initial pathway to nicotine addiction, with or without the use of cigarettes[27]. From pairwise comparisons of reported age of onset, it seems to be the case that cigarette use precedes the use of other tobacco products, such as cigars, smokeless tobacco, hookah, and e-cigarettes. However, a sizeable number of individuals reported using these substances within at least the same year. Cigar use appears to precede the use of smokeless tobacco, while smokeless tobacco use appears to precede or occur around the same time as first use of hookah or e-cigarettes. The timing of onset of the initiation of cigars, hookah, and e-cigarette use was more difficult to discern. In part, this could be attributed to the growing availability of these substances since students enrolled in S4S.

Differences by sex and race/ethnicity across lifetime and recent tobacco vary by tobacco product. In general, men had higher prevalence of tobacco use than women[28]. This finding is similar to another study that found that men used smokeless tobacco products and snus significantly more frequently than women[24]. Underlying causes for differences in tobacco product use across race/ethnicity are complex and multifactorial. It is possible that differences among certain racial/ethnic groups are related to cultural factors, such as social disapproval of smoking, particularly among women[29]. This could signal underlying cultural norms related to the use of specific tobacco products, as well as change in availability of tobacco products.

Parental environment prior to university enrollment has a small, but significant influence on tobacco use. In part, this could be attributed to the fact that as adolescents grow into young adults, protective parental influences become less important in adolescents' use of tobacco. This finding is consistent with findings from twin and epidemiological studies, which suggest that parental environment plays a smaller role in young adulthood. One study suggested that parental influences of connectedness and punishment for smoking remains important until mid-adolescence and parental monitoring continues to be important in protecting against smoking at age 16[30]. Our study suggests that parental involvement may be protective against both lifetime and recent tobacco use, while parental autonomy granting is positively associated with lifetime and recent tobacco use. This is aligned with another study by the same research group that found that a higher levels of family involvement is protective against recent smoking[31].

Experience of stressful life events prior to university might have lasting effects on tobacco use. Prior research suggests that stressful life events are associated with smoking and that stressful life events might have a differential effect on smoking among women versus men[32]. In this study, moderate correlations were found between experience of stressful life events prior to university and tobacco use, with the strongest correlations found between the experience of physical abuse and sexual abuse and tobacco use. It is possible that tobacco use functions as a coping behavior for these stressors, as previous observational studies have demonstrated that the experience of acute stressful events are associated with higher smoking prevalence. However, tobacco use is an ineffective stress-reducing strategy and perpetuates a stress response in users, which can also result in diminished self-regulation to control the urge to use. Experimental studies

demonstrate that induced stress reduces an individual's ability to resist tobacco use and increases tobacco use intensity and reward[33].

There are, however, a few limitations to note. We are unable to infer causality, despite trying to limit the exposure time for the experience of stressful life events and parental environment to prior to enrollment at university, due to the use of correlations in this study. Another potential limitation could be recall bias, since students are asked to retrospectively report their parental environment and stressful life events. Recall of information is solely dependent upon memory, which can be imperfect and potentially unreliable. Additionally, since the university has a diverse population of students, the findings from this study might not be representative of all universities of its size, or of all young adults outside of the college setting.

Despite these limitations, this study contributes to the existing literature by demonstrating that tobacco use was prevalent among college students and that tobacco use was not limited to the use of cigarettes. The use of smokeless tobacco, hookah, and e-cigarettes was also common. The use of tobacco products was correlated with other environmental factors, such as parental involvement and parental autonomy granting, physical assault, and sexual assault experienced prior to enrollment at university. Our study also demonstrated that these associations may persist over time. Given these potentially long-standing effects, it may be useful to view university enrollment as an opportunity for potential intervention or prevention of tobacco use as a coping mechanism for these stressful life events[34]. Furthermore, the endorsement of lifetime use of tobacco products and nicotine delivery systems increased across freshman to senior year, suggesting that college is a time when many individuals are trying a range of tobacco products and

nicotine delivery systems. This is important since the use of tobacco products put individuals at risk of developing nicotine dependence and lead to many negative health risks. There are few regulations associated with the marketing, sale and use of emerging tobacco products that can encourage young adults to start using tobacco products, if they have not by the time they start college[35]. Thus, further research on the prevalence, correlates, and risk factors for the use of alternative tobacco products is needed, especially as use continues to gain popularity[36].

CHAPTER 7:
INITIAL EXPERIENCES WITH NICOTINE AND ITS ASSOCIATION WITH RECENT USE OF TOBACCO AND NICOTINE DEPENDENCE

Elizabeth K. Do; Elizabeth C. Prom-Wormley; Danielle M. Dick; Kenneth S. Kendler; Hermine H. Maes

BACKGROUND

It has been suggested that initial experiences during first exposure to tobacco may be indicative of individuals' sensitivity to nicotine[1] and vulnerability to nicotine dependence[2]. However, whether innate sensitivity to nicotine dependence either enhances or inhibits the likelihood of established patterns in smoking behavior remains an area of debate[3]. Both adverse experiences, such as coughing, dizziness, and nausea, and positive experiences such as relaxation or experiencing a pleasurable rush or buzz can result from experimentation with tobacco products containing nicotine[4]. Generally, positive effects have been found to have a stronger association with smoking behavior[1,5]. Though, unpleasant reactions due to symptoms of dizziness or nausea during the initial experience of tobacco use are not necessarily protective against subsequent smoking[5–7]. Thus, a clearer understanding of the factors influencing initial experiences with tobacco use and of how these initial experiences affect subsequent use of tobacco products and the development of nicotine dependence is needed[5].

Many factors are likely to influence initial experiences with tobacco use. One such factor is the age of onset of tobacco use. One study found that pleasurable sensations upon initial exposure to tobacco use were significantly linked to the age at first cigarette, whereby the earlier the first tobacco use occurred, the higher the probability of pleasant

feelings during initial smoking. The same study reports a trend towards experiencing higher relaxation in those with an earlier age at first cigarette, and no relation between the age at first cigarette and experience of unpleasant sensations. In this study conducted by Buchmann et al. (2011), an earlier age at first cigarette is also associated with a greater likelihood of becoming a regular smoker by age 22[8].

Since previous studies have primarily focused on initial experiences with cigarette use, the present study seeks to contribute to the literature by determining: (1) the prevalence of initial experiences with tobacco, recent tobacco use, and nicotine dependence in two different samples: the Spit for Science[9] and the Virginia Twin Studies of Adolescent Behavioral Development[10], (2) whether or not initial experiences with tobacco use differ according to tobacco product used, (3) whether there is an association between initial experiences with tobacco use and recent tobacco use, or nicotine dependence, and (4) whether these associations differ by sex.

MATERIALS AND METHODS

Samples

Spit for Science. Data were collected from Spit for Science, a research study examining how genetic and environmental factors come together to contribute to the development of problems associated with the use of alcohol, the use of other substances, and emotional health. Incoming freshman, 18 years of age and older, were invited to participate in the study which included an electronic survey designed to collect broad-based information about substance use and mental health outcomes. Additionally, the study design involves multiple waves of data collection, including: two collection periods

during the freshman year occurring upon arrival on campus in the fall and again during the middle of the spring semester. Further waves of data were collected annually during the spring semester. Data collection for this study is ongoing, and results from this study reflect a subset of the total Spit for Science sample who answered questions about initial experiences with tobacco use, current tobacco use, and completed the Fagerström Test for Nicotine Dependence (FTND).

For the purposes of this study, those who did not provide information on initial experiences with tobacco were excluded from analyses. Due to small sample size, individuals who reported their self-identified race category as: Native Hawaiian or Other Pacific Islander and American Indian or Alaskan Native were excluded (n=67 & 51 respectively). Thus, data was available for 2,081 individuals. All survey information collected was managed using an electronic data capture tool hosted at Virginia Commonwealth University called Research Electronic Data Capture (REDCap). REDCap is a secure, web-based application designed to support data capture for research studies[11]. Participants were given a $10 incentive for their participation in each of the surveys. DNA collection is also a part of the study protocol, but not used in the present analysis. The VCU Institutional Review Board (IRB) approved all S4S protocols.

Virginia Twin Studies of Adolescent Behavioral Development (VTSABD). The VTSABD is a cohort-longitudinal epidemiological study that uses the genetic twin design to study the development and maintenance of child psychiatric disorders[12]. It is comprised of three studies: the Virginia Twin Study of Adolescent Behavior Development, the Young Adult Follow-Up, and the Transitions to Substance Abuse follow-up. It is the first population-based, multi-wave, cohort-sequential twin study of adolescent psychopathology and its

risk factors. Included within this study are Caucasian families of male and female monozygotic and dizygotic twins and their parents to assess the role of genes and the environment in developmental trajectories of behavior from childhood to young adulthood. The sample was ascertained through Virginia schools and assessment of the children involved semi-structured and face-to-face interviews with both twins and both parents using the Child and Adolescent Psychiatric Assessment (CAPA). Self-report questionnaires were also completed by parents, children, and teachers[13].

Measures

Demographic characteristics. Within the Spit for Science study, participants were asked to select from “American Indian/Alaska Native”, “Asian”, “Black/African American”, “Hispanic/Latino”, “More than one race”, “Native Hawaiian/Other Pacific Islander”, “Unknown”, “White”, or “I choose not to answer” to answer the question: “Which one of these groups’ best describes you?” For sex, participants could select: “male”, “female”, or “I choose not to answer”. Participants in VTSABD were predominantly White.

Age of onset of tobacco use. Within Spit for Science, age of onset was measured using the following question: “How old were you when you smoked a cigarette or used tobacco for the first time (including just one or two puffs)?” This variable was coded as an ordinal variable, with those who had not initiated smoking coded as 0, those who had initiated after age 18 as 1, gradually increasing to those who initiated prior to age 12 (2: 15-18 years, 3: 12-14 years, 4: <12 years). For cohort 4 of Spit for Science, measures for tobacco product-specific age of onset were available, asking “How old were you when you [smoked a cigarette/used smokeless tobacco/ smoked cigars, little cigars, cigarillos/

smoked hookah/ used an e-cigarette] for the first time?" Participants responded to this question by inputting their age in years in a free response. Age of onset was coded similarly for responses from the VTSABD.

Initial sensitivity to tobacco products. Initial sensitivity was measured by asking eight questions about the sensations experienced when the participant first used a specified tobacco product. Each question was worded "how much [insert sensation here] … did you feel?" with potential responses being: none, slight, moderate, or intense. The eight questions asked about feeling pleasant sensations, unpleasant sensations, nausea, relaxation, dizziness, pleasurable rush or buzz, coughing, and difficulty inhaling. These measures were the same in Spit for Science and the VTSABD.

Recent use of tobacco products. For the Spit for Science sample, recent use of tobacco products (e.g. use of tobacco products in the last 30 days) was treated as binary variables, similarly coded to lifetime use of tobacco products. Measures for recent use of cigarettes, smokeless tobacco, cigars, hookah, and e-cigarettes included the following, respectively: "How frequently did you smoke cigarettes in the past 30 days?", "How frequently did you use a smokeless tobacco product (dip/chaw/snus) in the last 30 days?", "During the last 30 days, on how many days have you smoked cigars, little cigars, or cigarillos?", "During the last 30 days, on how many days have you smoked a hookah?", and "During the last 30 days, on how many days did you use e-cigarettes?" For each of these measures, if the participant indicated 'I choose not to answer', the answer was coded as missing. If the participant indicated that he/she had not used the tobacco product, the answer was coded as 0 (no); otherwise, if the participant had indicated that he/she had used the tobacco product at least once, the answer was coded as 1 (yes). An aggregate measure

of recent use of tobacco products was also calculated by summing across each of the recent use measures of tobacco products, with 0 coded as no, and 1 or greater coded as yes. A similar approach was taken for VTSABD data, which focused on cigarette use.

Nicotine Dependence (ND). Nicotine dependence was measured using the Fagerström Test for Nicotine Dependence. A threshold of ≥4 was used to determine whether an individual met criteria for ND.

Statistical Analysis

Polychoric correlations were computed to examine the relationship between initial experiences with tobacco and recent tobacco use and nicotine dependence. Multivariate analyses relied on multiple regressions to determine whether sex was an important contributor to the association between initial experiences with tobacco use, recent tobacco use and nicotine dependence, taking into account the covariates of age of onset, race/ethnicity, first tobacco product used, and any significant interaction effects into account.

RESULTS

Descriptive Statistics

Responses regarding initial experiences with tobacco products were only available for a subset of participants from cohorts 2 (n=645), 3 (n=650), and 4 (n=786) in Spit for Science. The remaining analytic sample (n=2,081) was predominantly female (n=1388; 66.7%) and nearly half identified as White/Caucasian (n=978; 47.6%), as shown in Table 7.1. The age at first tobacco use varied by tobacco product; average age of first use was 15.6 years (SD:2.8) for cigarettes, 15.2 years (SD:4.1) for smokeless tobacco, 16.0 years

(SD:2.9) for cigars, 16.8 years (SD:1.8) for hookah, and 17.1 years (SD:1.9) for e-cigarettes. On average, most participants in the study indicate cigarettes as the product that they first used (44.0%), followed by hookah (34.3%), cigars (14.6%), e-cigarettes (3.8%), smokeless tobacco (2.5%), and other tobacco product (0.9%).

Findings from the VTSABD are similar to S4S in that most the sample was female (59.2%), White/Caucasian (98.8%), and most tobacco users had initiated use by age 18.

Table 7.1: Prevalence of recent tobacco use and nicotine dependence

	S4S (n = 2081)		VTSABD (n = 850)	
	N	**%**	**N**	**%**
Sex				
Male	692	33.3	343	40.8
Female	1388	66.7	497	59.2
Race/Ethnicity				
White/Caucasian	978	47.6	840	98.8
Black/African American	443	21.6		
Asian	337	16.4		
Hispanic/Latino	148	7.2		
More than one race	147	7.2		
Unknown			10	1.2
Cohort				
Fall 2012	645	31.0		
Fall 2013	650	31.2		
Fall 2014	786	37.7		
Age of Initiation (Any Tobacco use)				
Did not Initiate	6	0.3	9	1.1
>18 years	289	13.9	138	16.2
15-18 years	929	44.6	396	46.6
12-14 years	325	15.6	229	26.9
<12 years	90	4.3	78	9.2
Initiated but did not report age	442	21.2		
	Mean	**SD**		
Age of Initiation (Specific Tobacco Use)				
Cigarettes (S4S n = 494)	15.97	2.39		
Smokeless, Snus, Dip, Chaw (S4S n = 133)	15.98	2.84		
Cigar (S4S n = 496)	16.34	2.56		
Hookah (S4S n = 685)	16.58	2.84		
E-cigarettes (S4S n = 449)	17.32	1.43		
Recent Tobacco Use	**N**	**%**	**N**	**%**
Any tobacco product (S4S n = 2081)	984	47.3	300	35.3
Cigarettes (S4S n = 1383)	544	39.3		
Smokeless (S4S n = 398)	398	24.6		
Cigar (S4S n = 1829)	330	18.0		
Hookah (S4S n = 1991)	422	21.2		
E-cigarette (S4S n = 1149)	270	23.5		

Fagerström Test for Nicotine Dependence (S4S n=1924; range=0-10)	0.70	1.50	2.2	2.2
	N	**%**		
Meets criteria for nicotine dependence	132	6.9	181	21.0

The highest proportion of students, who indicated that they used tobacco in the last 30 days at the most recent survey of Spit for Science, reported using cigarettes (39.3%), followed by smokeless tobacco (24.6%), e-cigarettes (23.5%), hookah (21.2%), and cigars (18.0%), as shown by Table 7.1. When comparing recent tobacco use across males and females, significant differences were found for cigarette, cigar, and e-cigarette use, such that males were more likely to endorse the recent use of all tobacco products relative to females within S4S. No significant sex differences were found for current use of cigarettes within the VTSABD. A little less than 7% of the total sample met criteria for nicotine dependence in S4S and 35.3% of the VTSABD sample. Significant differences in meeting criteria for nicotine dependence were found between males and females, such that a larger proportion of males met criteria for nicotine dependence, relative to females (10.4% vs. 5.2% in S4S; 42.5% vs. 31.6% in VTSABD), as shown in Table 7.2.

Table 7.2: Sex Differences in Endorsement of Recent Tobacco Use and Nicotine Dependence

	Males		Females		Chi-Square, df, p-value
Recent Tobacco Use (S4S)	**N**	**%**	**N**	**%**	
Any tobacco product (n = 2079)	407	58.9	576	41.5	56.0, 1, <0.0001*
Cigarettes (n = 1382)	251	47.5	293	34.3	23.9, 1, <0.0001*
Smokeless (n = 398)	70	27.2	28	19.9	2.7, 1, 0.1022
Cigar (n = 1828)	153	24.0	177	14.9	23.5, 1, <0.0001*
Hookah (n = 1990)	171	26.1	251	18.8	14.0, 1, 0.0002*
E-cigarette (n = 1148)	154	32.6	115	17.0	37.3, 1, <0.001*
Current Use (VTSABD; n=839)	128	37.4	168	33.8	1.2, 1, 0.2804
Fagerström Test for Nicotine Dependence					
FTND ≥4 (S4S; n = 1923)	66	10.4	66	5.2	18.7, 1, <0.0001*
FTND ≥4 (VTSABD; n = 497)	90	42.5	90	31.6	6.2, 1, 0.0126*

Initial experiences by age of onset

Within the S4S sample, age of onset of tobacco use was associated with stronger initial responses to tobacco across positive experiences (e.g. pleasurable buzz, pleasant sensation, relaxation), negative experiences (e.g. unpleasant sensation, dizziness, nausea), and difficulties with breathing (e.g. difficulty inhaling, coughing), even after adjusting for race/ethnicity and sex. More specifically, age of onset was associated with: relaxation (beta = 0.066, p-value = 0.040), pleasurable rush or buzz (beta = 0.107, p-value = 0.0011), unpleasant sensation (beta = 0.062, p-value = 0.0482), nausea (beta = 0.149, p-value <0.0001), dizziness (beta = 0.140, p-value <0.0001), and difficulty inhaling (beta = 0.074, p-value <0.0001).

Age of onset of tobacco use was also associated with stronger initial responses to tobacco within the VTSABD sample. Corresponding numbers were: pleasurable rush or buzz (beta = 0.098, p-value = 0.0167), unpleasant sensation (beta = 0.092, p-value = 0.0313), nausea (beta = 0.187, p-value <0.0001), dizziness (beta = 0.193, p-value <0.0001), coughing (beta = 0.159, p-value = 0.0002), and difficulty inhaling (beta = 0.16, p-value = 0.0003).

Factor structure of initial experiences with tobacco use

An exploratory factor analysis was performed on the eight items measuring initial sensitivity to tobacco products and eigenvalues showed that two factors could be selected. According to Kaiser's criteria, only factors with eigenvalues equal to or greater than 1 should be selected for factor analysis. The two-factor model using varimax rotation showed that the first factor was related to positive experiences during initial tobacco use

(e.g. pleasant sensation, relaxation, pleasurable rush or buzz), the second factor was related to negative experiences during initial tobacco use (e.g. unpleasant sensation, nausea, dizziness, coughing, and difficulties inhaling). Interestingly, the item measuring dizziness cross-loaded onto positive experiences, though not as highly as it did on negative experiences. This implies that many users rated lesser amounts of dizziness as a positive experience. This factor structure was the same across S4S and the VTSABD samples, and was used to create factor scores for use in linear regressions testing whether initial positive or negative experiences with tobacco use are predictive of recent tobacco use and/or nicotine dependence – as described later in this manuscript.

Table 7.3: Varimax Rotated Factor Patterns for Initial Experiences with Tobacco Use

	SPIT FOR SCIENCE			VTSABD		
Initial Experiences	**Factor 1**	**Factor 2**	**Communality**	**Factor 1**	**Factor 2**	**Communality**
Pleasant Sensations	**0.873**	0.065	0.766	**0.761**	-0.135	0.597
Relaxation	**0.851**	0.038	0.725	**0.646**	-0.118	0.432
Pleasurable Rush or Buzz	**0.870**	0.148	0.778	**0.821**	0.001	0.674
Unpleasant Sensation	0.038	**0.730**	0.534	-0.212	**0.639**	0.453
Dizziness	0.413	**0.600**	0.530	0.396	**0.443**	0.352
Nausea	0.189	**0.693**	0.516	0.067	**0.610**	0.376
Coughing	-0.007	**0.622**	0.387	-0.051	**0.640**	0.412
Difficulty Inhaling	0.012	**0.658**	0.434	-0.139	**0.550**	0.321
Variance Explained	2.451	2.220		1.899	1.720	

Initial experiences with tobacco by sex

Initial experiences with tobacco use differed by sex, as depicted by Table 7.4. However, significant differences in initial experiences with tobacco use by sex were not the same across samples. Whereas there were sex differences across all initial experiences, apart from difficulty inhaling in S4S, sex differences were only detected across the initial

experiences of: pleasurable rush or buzz, unpleasant sensations, and difficulty inhaling in VTSABD.

Associations between initial experiences and recent tobacco use

Significant correlations were found between initial experiences with tobacco and recent tobacco use, as depicted in Table 7.5. Initial positive experiences (pleasant sensations, relaxation, and pleasurable rush or buzz) were moderately correlated with recent use of all tobacco products. Smaller but significant correlations were observed between nausea and/or dizziness and the recent use of most tobacco products. The strongest correlations with negative initial experiences were with the recent use of smokeless tobacco products. The largest correlations were found between positive initial experiences with tobacco use and recent use of cigarettes.

Associations between initial experiences and nicotine dependence

The initial experiences of pleasant sensations, relaxation, pleasurable rush or buzz, nausea, coughing, dizziness, and difficulty inhaling were positively and significantly correlated with meeting criteria for nicotine dependence within the S4S sample. These correlations ranged in value from 0.11 (95% CI: 0.01, 0.10) to 0.38 (95% CI: 0.29, 0.46). The highest significant correlation was found between experiencing a pleasurable rush or buzz and meeting criteria for nicotine dependence, as depicted in Table 7.5. Within the VTSABD, only initial experiences of relaxation and pleasurable rush or buzz were positively and significantly correlated with meeting criteria for nicotine dependence. The correlations were lower, relative to the S4S sample, ranging from 0.15 (95% CI: 0.02, 0.27) to 0.16 (95% CI: 0.04, 0.21).

Table 7.4: Sex Differences in Initial Experiences with Tobacco Use in S4S and VTSABD

	S4S							VTSABD						
	Full		Males		Females			Full		Males		Females		
POSITIVE EXPERIENCES														
	N	**%**	**N**	**%**	**N**	**%**	**Chi-square, df, p**	**N**	**%**	**N**	**%**	**N**	**%**	**Chi-square, df, p**
Pleasant Sensations							24.9, 3, <0.0001*							6.7, 3, 0.0804
None	939	45.8	270	39.8	669	48.8		400	47.3	145	42.4	249	50.5	
Slight	543	26.5	175	25.8	367	26.8		232	27.5	96	29.1	132	26.8	
Moderate	448	21.9	187	27.6	261	19.1		160	18.9	77	22.5	83	16.8	
Severe	119	5.8	46	6.8	73	5.3		53	6.3	24	7.0	29	5.9	
Relaxation							30.6, 3, <0.0001*							7.4, 3, 0.0603
None	877	42.6	245	36.0	632	45.9		458	54.7	170	50.2	281	57.6	
Slight	557	27.1	178	26.2	378	27.5		200	23.9	85	25.1	114	23.4	
Moderate	469	22.8	199	29.3	270	19.6		146	17.4	72	21.2	72	14.8	
Severe	155	7.5	58	8.5	97	7.0		33	3.9	12	3.5	21	4.3	
Pleasurable Rush or Buzz							51.5, 3, <0.0001*							11.2, 3, 0.0106*
None	972	47.3	256	37.6	716	52.2		337	40.2	118	34.5	213	43.7	
Slight	504	24.5	170	25.0	333	24.3		232	27.7	92	26.9	137	28.1	
Moderate	387	18.8	171	25.1	216	15.7		190	22.7	94	27.5	96	19.7	
Severe	192	9.3	84	12.3	108	7.9		80	9.5	38	11.1	41	8.4	
NEGATIVE EXPERIENCES														
Dizziness							26.0, 3, <0.001*							4.0, 3, 0.2622
None	1080	52.4	316	46.2	763	55.5		283	33.8	107	31.5	172	35.3	
Slight	502	24.4	165	24.1	337	24.5		214	25.6	84	24.7	128	26.3	
Moderate	331	16.1	140	20.5	191	13.9		220	26.3	91	26.8	126	25.9	
Severe	147	7.1	63	9.2	84	6.1		120	14.3	58	17.1	61	12.5	
Unpleasant Sensations							24.4, 3, <0.0001*							10.6, 3, 0.0144*
None	1061	51.8	303	44.6	757	55.4		171	20.2	68	19.8	102	20.7	
Slight	494	24.1	184	27.1	310	22.7		202	23.9	101	29.5	99	20.1	
Moderate	369	18.0	152	22.4	217	15.9		283	33.5	108	31.5	172	34.9	
Severe	124	6.1	41	6.0	83	6.1		190	22.5	66	19.2	120	24.3	
Nausea							29.1, 3, <0.0001*							0.2, 3, 0.9728
None	1288	62.7	375	54.8	912	66.5		411	49.2	169	49.7	236	48.6	
Slight	377	18.3	153	22.4	224	16.3		179	21.4	71	20.9	108	22.2	
Moderate	277	13.5	117	17.1	160	11.7		160	19.1	66	19.4	93	19.1	
Severe	114	5.5	39	5.7	75	5.5		86	10.3	34	10.0	49	10.1	
Coughing							10.9, 3, 0.0120*							7.1, 3, 0.0685
None	761	36.9	257	37.6	503	36.5		172	20.8	74	22.0	96	20.0	
Slight	715	34.6	250	36.6	465	33.7		294	35.6	133	39.6	158	32.9	
Moderate	408	19.8	137	20.0	271	19.7		193	23.4	73	21.7	118	24.5	
Severe	180	8.7	40	5.9	140	10.2		167	20.2	56	16.7	109	22.7	
Difficulty Inhaling							3.9, 3, 0.2694							15.1, 3, 0.0017*
None	1147	55.8	385	56.5	761	55.5		209	25.2	106	31.6	101	21.0	
Slight	506	24.6	168	24.7	338	24.6		221	26.7	81	24.1	136	28.2	
Moderate	295	14.4	102	15.0	193	14.1		238	28.7	98	29.2	138	28.6	
Severe	106	5.2	26	3.8	80	5.8		160	19.3	51	15.2	107	22.2	

Table 7.5. Correlations Between Initial Experiences, Recent Tobacco Use, and Nicotine Dependence

	Pleasant Sensations	Relaxation	Pleasurable Rush or Buzz	Dizziness	Unpleasant Sensations	Nausea	Coughing	Difficulty Inhaling
SPIT FOR SCIENCE								
Any Tobacco	**0.42 (0.36, 0.47)**	**0.41 (0.36, 0.46)**	**0.47 (0.42, 0.52)**	**0.28 (0.22, 0.33)**	**0.07 (0.01, 0.13)**	**0.17 (0.11, 0.24)**	0.04 (-0.02, 0.10)	**0.13 (0.07, 0.19)**
Cigarettes	**0.31 (0.25, 0.38)**	**0.34 (0.27, 0.30)**	**0.37 (0.30, 0.43)**	**0.23 (0.16, 0.30)**	0.03 (-0.05, 0.10)	**0.11 (0.03, 0.19)**	-0.03 (-0.10, 0.04)	**0.09 (0.01, 0.16)**
Smokeless	**0.25 (0.11, 0.39)**	**0.26 (0.12, 0.40)**	**0.16 (0.02, 0.31)**	**0.32 (0.19, 0.46)**	**0.24 (0.10, 0.38)**	**0.28 (0.14, 0.42)**	**0.16 (0.02, 0.30)**	**0.27 (0.13, 0.42)**
Cigars	**0.18 (0.10, 0.25)**	**0.21 (0.13, 0.28)**	**0.15 (0.08, 0.23)**	**0.10 (0.03, 0.18)**	0.06 (-0.01, 0.14)	**0.16 (0.08, 0.23)**	**0.08 (0.01, 0.16)**	**0.11 (0.04, 0.19)**
Hookah	**0.19 (0.13, 0.26)**	**0.23 (0.17, 0.30)**	**0.21 (0.14, 0.27)**	**0.12 (0.05, 0.19)**	0.01 (-0.06, 0.08)	0.06 (-0.01, 0.14)	0.05 (-0.02, 0.12)	0.05 (-0.02, 0.13)
E-Cigarettes	**0.26 (0.18, 0.35)**	**0.28 (0.20, 0.36)**	**0.29 (0.21, 0.37)**	**0.20 (0.11, 0.29)**	**0.09 (0.00, 0.18)**	**0.15 (0.06, 0.24)**	0.06 (-0.03, 0.14)	**0.13 (0.04, 0.22)**
Nicotine Dependence	**0.26 (0.17, 0.35)**	**0.30 (0.21, 0.38)**	**0.38 (0.29, 0.46)**	**0.26 (0.17, 0.35)**	0.08 (-0.01, 0.18)	**0.24 (0.14, 0.33)**	**0.11 (0.01, 0.10)**	**0.25 (0.15, 0.34)**
VIRGINIA TWIN STUDIES OF ADOLESCENT AND BEHAVIORAL DEVELOPMENT								
Cigarettes	**0.38 (0.30, 0.47)**	**0.30 (0.20, 0.39)**	**0.35 (0.26, 0.44)**	**0.30 (0.21, 0.39)**	**-0.18 (-0.27, -0.09)**	0.06 (-0.04, 0.16)	-0.06 (-0.16, 0.03)	**-0.13 (-0.22, -0.03)**
Nicotine Dependence	0.12 (-0.01, 0.24)	**0.15 (0.02, 0.27)**	**0.16 (0.04, 0.21)**	0.10 (-0.02, 0.22)	0.00 (-0.12, 0.12)	0.05 (-0.08, 0.17)	0.10 (-0.02, 0.22)	0.03 (-0.09, 0.15)

Boldface indicates significant correlation.

Sex differences in correlations between initial experiences with tobacco and recent tobacco use and/or nicotine dependence

Within the S4S sample, significant correlations were found between positive experiences and recent use of cigarettes in both sexes; cigars and e-cigarette use in males; hookah use in females; and, nicotine dependence in both sexes, but more strongly in females. Meanwhile, significant correlations were found between negative initial experiences with tobacco use and recent use of smokeless tobacco and between difficulty breathing during initial experience with tobacco use and recent use of smokeless tobacco, in females. Significant correlations between dizziness and recent cigarette use and nicotine dependence were found in both sexes.

Significant sex differences in the correlations between initial experiences and recent tobacco use and/or nicotine dependence were also found within the VTSABD sample. Specifically, the correlation was higher among males than females for the association between the initial experiences of pleasant sensations (0.47 in males, 0.31 in females, z-statistic = 2.68, p=0.004) and dizziness (0.38 in males, 0.24 in females, z-statistic = 2.19, p=0.014) and recent cigarette use. Correlations were higher among females than males for the association between: the initial experiences of relaxation (0.04 in males, 0.23 in females, z-statistic = -2.11, p=0.017) and of coughing and current cigarette use (-0.14 in males, 0.01 in females, z-statistic = 2.11, p=0.017).

Predictors of tobacco use and nicotine dependence

Across both samples, positive initial experiences predicted both recent tobacco use and meeting criteria for nicotine dependence, after adjusting for sex and age of initiation (and race/ethnicity and first tobacco product used within the S4S sample). The direction of these effects was the same, such that higher levels of positive initial experiences predicted both recent tobacco use and meeting criteria for nicotine dependence. Having negative initial experiences with tobacco use was also predictive of meeting criteria for nicotine dependence within the S4S sample, but not in the VTSABD. Sex was also found to be a predictor of meeting criteria for nicotine dependence, such that being female reduced the likelihood of meeting criteria for nicotine dependence (as demonstrated by the negative direction of the effect). The magnitude of the sex effect was similar across samples (e.g. beta = -0.22 in S4S and -0.28 in VTSABD). Age of initiation also had a significant effect on meeting criteria for nicotine dependence, such that earlier ages of initiation were predictive of meeting criteria for nicotine dependence.

Table 7.6a: Predictors of Recent Tobacco Use and Nicotine Dependence in Spit for Science

	Recent Tobacco Use (Any)			Meets Criteria for Nicotine Dependence (FTND≥4)		
	Beta	**SE**	**p-value**	**Beta**	**SE**	**p-value**
Intercept	-2.06	85.23	0.9807	-2.81	0.29	<0.0001*
Positive Initial Experience	0.71	0.06	<0.0001*	0.57	0.11	<0.0001*
Negative Initial Experience	0.06	0.06	0.3069	0.24	0.11	0.0290*
Race – More than one race vs. White/Caucasian	0.05	0.17	0.7721	-0.48	0.43	0.2686
Race – Asian vs. White/Caucasian	0.13	0.13	0.3105	0.38	0.26	0.1427
Race – Hispanic/Latino vs. White/Caucasian	-0.22	0.17	0.1928	-0.22	0.40	0.5825
Race – Black/AA vs. White/Caucasian	0.07	0.13	0.5775	0.07	0.30	0.8085
Sex – Female vs. Male	-0.19	0.06	0.0019*	-0.22	0.11	0.0470*
Age of Initiation – Did not initiate vs <12 years	-10.05	340.9	0.9765	NA	NA	NA
Age of Initiation - >18 years vs <12 years	2.21	85.2	0.9793	-0.84	0.28	0.0026*
Age of Initiation 15-17 years vs <12 years	2.50	85.2	0.9766	-0.43	0.18	0.0156*
Age of Initiation – 12 -14 years vs <12 years	2.78	85.2	0.9740	0.42	0.19	0.0252*
First Product Used – Other vs Cigarettes	0.84	0.71	0.2388	1.30	0.74	0.0793
First Product Used – E-cigarettes vs Cigarettes	-0.12	0.28	0.6759	-0.11	0.57	0.8405
First Product Used – Hookah vs Cigarettes	-0.45	0.19	0.0167*	-0.13	0.32	0.6980
First Product Used – Cigars vs Cigarettes	-0.21	0.21	0.3075	-0.02	0.32	0.9578
First Product Used – Smokeless vs Cigarettes	0.13	0.15	0.7013	-1.23	0.65	0.0605

Table 7.6b: Predictors of Recent Tobacco Use and Nicotine Dependence in VTSABD

	Recent Tobacco Use (Any)			Meets Criteria for Nicotine Dependence (FTND≥4)		
	Beta	**SE**	**p-value**	**Beta**	**SE**	**p-value**
Intercept	-0.74	0.10	<0.0001*	-0.07	0.14	<0.0001*
Positive Initial Experience	0.74	0.10	<0.0001*	0.34	0.12	0.0044*
Negative Initial Experience	-0.17	0.10	0.0867	0.12	0.13	0.3225
Sex – Female vs. Male	-0.02	0.08	0.7897	-0.28	0.10	0.0074*
Age of Initiation - >18 years vs <12 years	-0.89	0.20	<0.0001*	-0.58	0.28	0.0416*
Age of Initiation 15-17 years vs <12 years	-0.01	0.13	0.9279	-0.20	0.17	0.2273
Age of Initiation – 12 -14 years vs <12 years	0.22	0.15	0.1339	0.31	0.18	0.0864

Discussion

The current study examined associations between initial experiences with tobacco use and recent tobacco use and nicotine dependence. We estimated the factor structure of initial experiences with tobacco products, concluding that the best fitting model was one with two factors: positive experiences (e.g. pleasant sensation, relaxation, and pleasurable rush/buzz), negative experiences (e.g. unpleasant sensation, nausea, dizziness, coughing, and difficulties inhaling). This factor structure was similar to that

found from a construct validity analysis conducted using confirmatory factor analysis by Rodriguez and Audrain-McGovern (2004), which identified two factors: pleasant (e.g. pleasant, relaxation, and rush or buzz) and unpleasant (e.g. unpleasant, coughing, difficulty inhaling, and nausea). The 'pleasant' factor overlaps with our 'positive experiences' factor and the 'unpleasant' factor overlaps with our 'negative experiences' factor[14]. Another study investigating the factor structure of early smoking experiences conducted by Baggio et al. (2013) tested associations with smoking behavior in two and three-factor models[15]. Similar to the current study and the study conducted by Rodriguez and Audrain-McGovern[14], this study identified a positive experiences factor (e.g. like the experience and felt relaxed) and negative experiences factor (e.g. did not feel very well, headache, stomach upset, heart pounding, nauseous, dizzy/lightheaded, coughed, and irritation eyes, bad taste) in the two-factor model. In the three-factor model proposed by Baggio et al. (2013) negative experiences were split into negative experiences of dizziness (e.g. did not feel very well, headache, stomach upset, heart pounding, nauseous, dizzy/lightheaded) and negative experiences of irritation (e.g. coughed, irritation eyes, bad taste)[15]. Differences in the factor structure between the current study and that conducted by Baggio et al. (2013) may be attributed to the use of different measures of initial experiences with tobacco use, limited to cigarette smoking, and the age of participants.

In the current study, we found that earlier age of onset of tobacco use was associated with stronger initial responses to tobacco across both positive and negative experiences. Given that Buchmann et al. (2011) also finds that the age of first cigarette use and pleasure experienced from cigarette use predicts smoking at age 22[8], it seems that the

age of onset of tobacco use and initial experiences with tobacco use should be considered in tobacco use interventions. In a study by Klein et al. (2013), 50.7% of study participants indicated that their experience with smoking was more negative than they expected, 30.2% indicated that the experience was about the same as they expected, and 19.1% indicated it was more positive than expected, demonstrating the variability in the immediate reaction to tobacco use. In the same study, 77.9% of participants remember the first cigarette making them feel calm and relaxed, 66.9% remember becoming dizzy, 52.1% remember coughing extensively, and 16.1% decided to continue smoking because they thought their subsequent experiences would be better than the first – suggesting that there is a small window of opportunity to intervene on the subsequent patterns of tobacco use following initial experience[16].

Like other studies, the current study found that regular use of tobacco products was related to initial experiences with tobacco use. Both negative and positive experiences seemed to have effects on regular use of tobacco products, which supports findings from previous studies[2,7,17]. Initial experiences during first tobacco use are believed to reflect the physiological and pharmacological effects of nicotine, as well an individual's sensitivity to and tolerance for nicotine. Generally, those who become regular users experience greater positive and negative reactions to nicotine compared to nonsmokers, while positive experiences may play a stronger role than unpleasant experiences in the transition to regular use[18]. This is important given the significant correlations found between initial experiences with tobacco and recent tobacco use in the current study – particularly stronger correlations found between positive experiences and recent use of cigarettes, cigars, hookah, and e-cigarettes (though the effect was not large). It has been

suggested elsewhere that social perceptions surrounding tobacco use may play a larger role in normalizing opinions regarding patterns of tobacco use, which could lead some individuals to be less resistant to experiment with tobacco[16]. Unfortunately, these findings do little to address the debate regarding whether positive reinforcement is sufficient to establish a trajectory towards nicotine dependence, or if negative reinforcement must come into play before an individual is past a 'point of improbable return'[19].

Initial sensitivity to nicotine, as measured by initial experiences with tobacco use, is only one factor associated with nicotine dependence explored in multiple studies, including the current study. The development of nicotine dependence symptoms is complex and can vary in timing of onset, level of escalation, duration, and remission of symptoms[20]. From the literature, those who progress to regular tobacco use may be sensitive to the rewarding effects of nicotine, as evidenced by the strong correlations between positive symptoms and different measures of continued use of tobacco. However, few studies have investigated how initial experiences might differ by tobacco product[18]. The current study finds that initial experiences do differ by the first tobacco product used. Another study examining the initial experiences with e-cigarette use found that few current e-cigarette users/triers had a negative first experience and that positive perceptions about first experiences were higher among current e-cigarette users when compared to former users[19]. Given the perception that e-cigarettes could be useful in cigarette cessation and the generally positive first experiences that users have with e-cigarette use, we must wonder what the downstream effects of nicotine from e-cigarettes might be on nicotine dependence. This is an especially important point to consider given the recent increase

in alternative tobacco use among youth; though, more research is needed to make more definitive conclusions.

Since data for the current study was collected using self-report measures of initial experiences with first use of tobacco products, it may be affected by recall bias – especially if recollection of initial experiences is influenced by current or continuous use. Previous studies have demonstrated that current smokers have a generally more positive recollection of their first tobacco use than former smokers[1,7,21]. Additionally, since we only had information on initial experiences for three of the four cohorts, it is unclear how missing data impacts the results. However, the findings of the current study are consistent with prior research and contribute to the literature through its assessment of multiple tobacco products. Even more importantly, our findings highlight the need for more expansive research on initial experiences with tobacco and alternative tobacco products.

Unmeasured behavioral, social, and environmental factors may play a role in shaping initial reactions to nicotine and the subsequent adoption of regular use of tobacco[22]. Social influences of peers and family may have an effect on initial experiences with tobacco use[17], as individuals with higher levels of exposure to smoking from peers and family members are more likely to report positive symptoms of initial smoking experience[22]. Exposure to smoking from peers and family members has also been found to be associated with individual reports of feeling: dizzy, relaxed, good, and high upon initial use of tobacco. Thus, initial experiences by individuals with greater exposure to smoking may be influenced by socially mediated expectancies, derived from the experience of others[23].

Alternatively, frequent or prolonged exposure to nicotine absorbed from environmental tobacco smoke may alter neurophysiology in the brain, which may be reflected by altered responses to nicotine among non-smokers[24]. Whereas a study conducted by Kozlowski & Harford (1976) found that non-smokers who were tempted to smoke were more likely than current smokers to report discomfort from smoking, possibly due to their physical reaction to cigarettes, which could discourage future use[25]. A study by Pomerleau et al. (1993) found that the initial sensitivity to noxious effects of smoking also reflects sensitivity to the reinforcing effects which may encourage future smoking[2]. Marked individual differences in the response to drugs, such as nicotine, may be attributed to both heritable contributions and unique environmental experiences[26].

Yet, few studies have investigated these genetic differences contributing to initial experiences with nicotine. One study by Sherva et al. (2008) finds an association between genetic variant *CHRNA5* and enhanced pleasurable responses to initial cigarette use among regular users[27]. Another study links a person's initial experience with smoking and their current smoking status with variation in a gene that encodes a nicotine receptor in the brain[28]. Adolescents with higher exposure to maternal smoking report lower number of unpleasant symptoms during initial smoking[29]. Genetic variation in two candidate genes has been previously associated with initial responses to nicotine. Individuals with the G-variant of OPRM1 A118G SNP are more likely to report liking of initial smoking, though findings are inconsistent[30]. Adolescents homozygous for the C-variant D4D2 Taq1A polymorphism report a lower number of unpleasant symptoms during initial smoking, indicating lower sensitivity to nicotine[31]. Those homozygous for the T-variant shows stronger perceptions of nicotine effects, indicating higher nicotine sensitivity among T-

allele carriers and may be associated with reduced feelings for reward due to reduced receptor availability[24,32].

Conclusion

We have demonstrated that current users of tobacco may be sensitive to the rewarding effects of nicotine. Additionally, initial reactions to tobacco differed by tobacco type, as well as by sex. Age of onset, sex, and positive initial experiences predicted both recent use and meeting criteria for nicotine dependence. Negative initial experiences played less of a role in meeting criteria for nicotine dependence. Thus, further research is needed to identify genetic and biological pathways influencing initial experiences with nicotine, and the social contexts that influence initial experiences with tobacco use, in efforts to potentially delay the overall age of onset for tobacco use and reduce individual's risk for nicotine dependence.

CHAPTER 8:
SEX DIFFERENCES IN FAGERSTRÖM TEST FOR NICOTINE DEPENDENCE ITEMS

Elizabeth K. Do; Danielle M. Dick; Kenneth S. Kendler; Hermine H. Maes

INTRODUCTION

The Fagerström Test for Nicotine Dependence (FTND), comprised of four dichotomous and two multi-response items[1], is the most commonly used measure of nicotine dependence. It is calculated by adding together scores for the six items, and has a range of 0 to 10. A score of four or greater is indicative of nicotine dependence. Using a sum score typically assumes a unidimensional trait. However, factor analyses of FTND items have yielded inconsistent results: while some studies indicate that the measure is comprised of only one factor, others have identified two. Single-factor model specification implies a simple linear combination of all FTND items, while two-factor model specifications propose “smoking pattern” and “morning smoking” factors[2]. Previous studies also suggest that the psychometric properties of these items may differ by sex and race/ethnicity. Psychometric studies of the FTND are necessary to gain greater insight into the structure of the test and the assessed dimensions[3]. The current study was conducted to compare sex differences in the response to FTND items within African and White/Caucasian Americans and evaluate the factor structure of FTND across these groups.

METHODS

Study Sample. The study sample was obtained through Spit for Science, a longitudinal study of college students’ behavioral and emotional health. To be included in this study,

participants had to have reported ever using tobacco in their lifetime, provide responses to the Fagerström Test for Nicotine Dependence, and self-report either “Black/African American” or “White/Caucasian’ when asked about which race/ethnicity group best describes them. Of the 9889 participants of the Spit for Science study, 6907 individuals met study criteria.

Measures. Tobacco use was measured across the following tobacco products: cigarettes, cigarillo’s/small cigars, smokeless tobacco (dip/chaw), hookah/waterpipe, and electronic cigarettes. Items of the FTND were adjusted accordingly and include four dichotomous and two multi-response items – we will use the abbreviations in square brackets: (1) How soon after you woke up did you smoke your first cigarette/use tobacco? [*wake*] (2) Did you find it difficult to refrain from smoking/using tobacco in places where it is forbidden (e.g. in church, at the library, in cinema, etc.)? [*refrain*] (3) Which cigarette/dip/chaw would you hate most to give up? [*giveup*] (4) How many cigarettes/smokeless tobacco products per day did you smoke/use? [*cpd*] (5) Did you smoke/use tobacco more frequently during the first hours after waking than during the rest of the day? [*morning*] and (6) Did you smoke/use tobacco if you were so ill that you were in bed most of the day? [*whenill*]. Responses to the *wake* question included: >60 minutes, 31-30 minutes, 6-30 minutes, and ≤5 minutes. Responses to the *refrain*, *morning* and *whenill* questions were no or yes. Responses to the *giveup* question 3 was the first one or any other. Responses to the *cpd* question included: <10 cigarettes, 11-20 cigarettes, 21-30 cigarettes, or >31 cigarettes. Since these measures were collected from the start of freshman year for all cohorts, a maximum reported FTND score was calculated across all waves for each participant. The

items from this maximum reported FTND score were included in this study, and used for statistical analyses.

Statistical Analyses. Since we were interested in determining whether males and females were statistically different in how they responded to each of the items of the FTND, chi-square tests were conducted. Scree plots, incremental variance accounted for, and interpretability were used to determine the factor structure. Exploratory factor analyses (EFA) were conducted separately by race/ethnicity and sex (e.g. African American males (AAM), African American females (AAF), White/Caucasian males (WCM), and White/Caucasian females (WCF)). Promax rotation was selected based upon findings from prior studies conducted on the FTND, suggesting the presence of correlated factors[1]. Exploratory factor analyses were performed to determine the structure of confirmatory factor analyses. Criteria for acceptable model fit included non-significant model chi-square, comparative fit index (CFI) ≥ 0.9, and a root-mean-square error of approximation (RMSEA) ≤ 0.08. Coding and chi-square tests were conducted using SAS 9.4 (SAS Institute: Cary, NC). Exploratory factor analyses were conducted using the psych and GPArotation packages, and the fa () command in R. Confirmatory factor analyses were conducted using MPlus (Muthen & Muthen).

RESULTS

Descriptive characteristics. Of the study sample of 6907 individuals (1873 African American or AA, and 4881 White/Caucasian American or WCA), 2539 (540 AA; 1999 WCA) were male and 4197 (1332 AA; 2865 WCA) were female.

Statistical differences in FTND item response by sex, separated by race/ethnicity group. Within African Americans, differences in response rates by sex were found for all items, except for three of the binary items: two related to morning smoking and one to smoking when ill. Within White/Caucasian Americans, differences in response rates by sex were found for all items, except for the cigarettes per day item.

Exploratory factor analysis. Consistent with prior studies of the factor structure of the FTND, a principal component analysis with promax rotation was conducted. The criterion for item inclusion was a factor loading of 0.30 or more[4]. Items with loadings on other factors were interpreted as belonging to the factor on which they had the highest loading. Scree plots demonstrated that a two-factor solution fit the data best across each of the four groups and accounted for 100% of the variance for each group. However, the factor loadings across groups differed.

Factor loadings (F1 and F2), communalities (H2), and uniqueness (U2) for each group are shown in the table below. Extracted communalities ranged from 0.12 to 0.77 and factor correlations ranged from 0.58 to 0.68 across all groups. Lowest communalities were observed for the *giveup* item in AA, and the *morning* item in WCA. The factor structure was similar across AA males, WCA males and females. This factor structure was such that the *wake*, *refrain*, *cpd* and *whenill* items loaded on factor 1 and the *giveup* and *morning* items loaded on factor 2, with factor loadings of 0.30 or greater. Factor 1 can be interpreted as a 'smoking pattern' factor, while factor 2 reflected a 'morning smoking' factor. Results within AA females were rather different with a separate factor being extracted for the *refrain* and *whenill* items.

Table 8.1. Factor structure of FTND Items by Race/Ethnicity Group

	AFRICAN AMERICAN FEMALES				AFRICAN AMERICAN MALES				WHITE/CAUCASIAN FEMALES				WHITE/CAUCASIAN MALES			
ITEM	**F1**	**F2**	**H2**	**U2**	**F1**	**F2**	**H2**	**U2**	**F1**	**F2**	**H2**	**U2**	**F1**	**F2**	**H2**	**U2**
WAKE	0.00	0.43	0.18	0.82	0.43	0.14	0.28	0.72	0.44	0.17	0.32	0.68	0.46	0.12	0.29	0.71
REFRAIN	0.97	-0.16	0.77	0.23	0.61	0.22	0.58	0.42	0.62	0.00	0.38	0.62	0.53	0.07	0.33	0.67
GIVE UP	0.14	0.24	0.12	0.88	-0.08	0.47	0.19	0.81	-0.01	0.87	0.75	0.25	-0.12	0.86	0.61	0.39
CPD	0.09	0.62	0.40	0.53	0.63	-0.02	0.39	0.62	0.57	0.01	0.33	0.67	0.64	-0.13	0.32	0.68
MORNING	-0.14	0.82	0.54	0.46	0.09	0.56	0.38	0.62	0.00	0.52	0.27	0.73	0.23	0.33	0.26	0.74
WHENILL	0.59	0.13	0.47	0.53	0.97	-0.22	0.74	0.26	0.74	-0.07	0.49	0.51	0.71	0.03	0.54	0.46

Values are highlighted in yellow to indicate what factors items are loading on that are higher than 0.30.

Confirmatory Factor Analysis. CFA was conducted separately for sex and race/ethnicity, with two-correlated factors (e.g. “smoking patterns” and “morning smoking”, respectively). The *wake*, *refrain*, *cpd* and *whenill* items loaded on factor 1 (“smoking patterns”) and the *giveup* and *morning* items loaded on factor 2 (“morning smoking”). The results for CFA analyses, using unstandardized estimates, are displayed in the table below.

Table 8.2. Results from Confirmatory Factor Analyses

	AAF	*AAM*	*WCF*	*WCM*
Fit indices				
Model X^2 (df)	*13.116 (8)*	*5.973 (8)*	*13.390 (8)*	*8.465 (8)*
p-value	*0.1079*	*0.6503*	*0.0991*	*0.3894*
CFI ≥ 0.9	*0.985*	*1.000*	*0.997*	*1.000*
RMSEA ≤ 0.08	*0.053*	*0.000*	*0.028*	*0.009*
F1 Loadings: Smoking Patterns				
wake	*1.000*	*1.000*	*1.000*	*1.000*
refrain	*1.356*	*1.321*	*1.017*	*1.088*
cpd	*1.265*	*1.111*	*0.982*	*0.954*
whenill	*1.388*	*1.304*	*1.104*	*1.299*
F2 Loadings: Morning Smoking				
giveup	*1.000*	*1.000*	*1.000*	*1.000*
morning	*1.416*	*1.952*	*0.686*	*1.008*
Correlation between F1 & F2	*0.361*	*0.230*	*0.600*	*0.434*
Item R^2				
wake	*0.411*	*0.496*	*0.588*	*0.480*
refrain	*0.756*	*0.865*	*0.609*	*0.568*
giveup	*0.299*	*0.239*	*0.517*	*0.613*
cpd	*0.657*	*0.611*	*0.568*	*0.437*
morning	*0.599*	*0.910*	*0.508*	*0.622*
whenill	*0.792*	*0.843*	*0.718*	*0.810*
Mean variance explained	*0.586*	*0.659*	*0.5982*	*0.588*

AAF = African American Females; AAM = African American Males; WCF = White/Caucasian Females; WCM = White/Caucasian Males

Results from these analyses indicated that each of these models fitted the data well, apparent from the non-significant chi-square model fit statistics. Based on the cut-off criteria for relatively good fit (CFI ≥ 0.9, RMSEA ≤ 0.08), the comparison of measures of approximate fit also supported conclusions based on chi-square tests of relative fit. Thus, a correlated, two-factor model fitted the data across each of these groups.

However, we also saw that the correlation between the two factors was different across each of the four groups, such that the correlations were lower within Black/African American females and males (0.361 and 0.230), when compared to White/Caucasian females and males (0.60 and 0.434, respectively).

DISCUSSION

This study employed EFA to determine the factor structure of FTND, and then used CFA to test model fit for a two correlated factors model. The results confirmed that the factor structure of FTND was not one-dimensional. Within this sample of college students, the first factor was characterized by the *wake*, *refrain*, *cpd* and *whenill* items and the second factor was characterized by the *giveup* and *morning* items. Across each group, the correlations between the factors ranged from 0.23 to 0.60, with the highest correlation found within WCA females. The correlated two-factor model suggested by the exploratory factor analyses and tested with confirmatory factor analyses was like one conducted within young smokers entering US Air Force Basic Military Training[2] and another sample of smokers enrolled in a Veteran's Affairs Medical Center Smoking Cessation Clinic[5].

Despite the consistency of our findings with these studies, the factor structure we found using our college age sample was different than what is found in studies of older adult

smokers. These studies still demonstrate two correlated factors, but items differentially load on the factors[3,6,7]. Specifically, in one study of patients admitted for pre-surgical assessments, the first factor is still characterized by morning smoking (*refrain* & *giveup* items), while the second factor assesses the degree of urgency to restore nicotine levels to a given threshold after nighttime abstinence (*cpd*, *morning* and *whenill* items), and the *wake* item permitted to load on both factors[6].

The difference in factor structure across these samples might be attributed to the age of the samples being assessed. The samples that demonstrate similar factor structures to the current study are younger. Older adults have a longer period during which they might be exposed to nicotine, and so the effects of nicotine on dependence symptoms might be different relative to younger samples who have been exposed for less time. In other words, FTND items may perform differently when used with younger, less addicted smokers relative to more mature smokers being assessed in cessation clinics.

In addition to providing another younger sample to test psychometric properties of FTND items, this study contributes to the literature by examining factor structure across sex and racial/ethnic groups. Chi-square tests indicated that differences existed across FTND items by sex. Results from the CFA conducted in this study demonstrated that a correlated, two-factor structure fit the data for African American and White/Caucasian males and females – suggesting that the factor structure of FTND items was not necessarily different by race/ethnicity, even though it did differ by sex. However, EFA suggested that the factor structure of FTND items may be different for African American females, relative to the other groups in this sample, even though a correlated, two-factor model fit the data for all groups. If the factor structure of FTND items were different for

African American females, relative to the other groups in this sample, it would mean that the constructs being measured within the FTND are biased, and scores from the FTND would interpreted differently for this group[8]. Thus, this factor structure needs to be tested in other samples, to determine its validity.

Other racial/ethnic groups were available for analyses, but not assessed in this study due to concerns over sample size. It has been suggested elsewhere that necessary sample sizes for factor analyses are dependent upon: the range of communalities, number of factors, and number of indicators. This sample had low communalities (e.g. communalities under 0.5) and two factors, including a factor which was weakly determined by two items. To achieve good recovery of population factors, it is suggested that sample sizes over 500 are needed[9]. Thus, larger samples are needed to assess whether the findings of this study generalize to other groups of young adults, who self-identify with other race/ethnicity groups not described in the current study.

Finally, other studies identifying a two-factor structure to the FTND have raised concerns that simplistic scoring of the six items of the FTND might not reflect the subtle differences between individual dependence profiles – especially if study participants score higher on one dimension and lower on the other. Using a single total score of the items – which would suggest that there is one underlying factor of nicotine dependence, as some other studies have found - produces an 'average' of dimension-specific scores. However, a single total score of the items may lack the 'sensitivity' to identify differences in dependence profiles, which could limit the potential for tailoring interventions[6]. Thus, more research needs to be conducted to determine whether using factor scores derived from

a correlated, two-factor model or the more traditional single total score of FTND derived from a one-factor model is the best approach moving forward.

CONCLUSIONS

Sex differences exist in the responses to FTND items, and their psychometric properties. Further studies are needed to determine how nicotine dependence measures perform across sex, race/ethnicity, and age and to assess whether the factor structure identified in this study and those previously conducted is generalizable to other populations.

CHAPTER 9: GENETIC ANALYSES OF TOBACCO USE BEHAVIORS AMONG AN ETHNICALLY DIVERSE UNIVERSITY SAMPLE

Elizabeth K. Do; Arden A. Moscati; Roseann E. Peterson; Bradley T. Webb; Danielle M. Dick; Kenneth S. Kendler; Hermine H. Maes

INTRODUCTION

Tobacco use encompasses a range of complex behavioral traits that are influenced, at least in part, by genes. Studies of adult twins have been useful in inferring genetic influences by comparing phenotypic similarities between monozygotic and dizygotic twins, and suggest that genetic factors account for approximately 50% of the variance in adult smoking behaviors overall. When looking at tobacco use behaviors separately, the estimates for heritability are more wide-ranging. Approximately 60% of the variance in smoking initiation is attributed to genetic factors, while genetic factors account for 55-69% of the variation for smoking persistence, 40 to 56% in smoking quantity, 60-76% in nicotine dependence, and about 50% in smoking cessation[1–3]. Generally, initiation of use is more strongly influenced by environmental factors, whereas progression to higher levels of use and dependence is more strongly influenced by genetic factors[4]. Additionally, the heritable components of tobacco use become increasingly expressed over the transition from adolescence, when many individuals initiate use, into adulthood[5].

Since each of tobacco use behavior constitutes steps along the trajectory towards possible nicotine dependence and problem use, it is important to try and elucidate their genetic etiology and answer such questions as: What genes are influencing tobacco use behaviors? And are the genes influencing individual tobacco use behaviors separate or overlapping? In efforts to answer these questions, many genome-wide association studies (GWAS) have been conducted to identify genes underlying susceptibility to

tobacco use behaviors (in addition to linkage and candidate gene studies). Yet, only a handful of genes, with generally small effects, have been identified. The most robust finding from the few genome wide association studies of smoking phenotypes is the association between smoking quantity and the nicotinic receptor gene cluster *CHRNA5-A3-B4*, located on chromosome 15[6].

Most genetic analyses on tobacco use behaviors have been conducted in adults of predominantly European descent, and so, less is known regarding the role of genes to tobacco behavior in adolescence and young adulthood, and among other populations. For this reason, we examine several measures of tobacco use [e.g. initiation, age of onset, current use, regular use, cigarettes per day, time to first tobacco use, and the Fagerström Test for Nicotine Dependence (FTND)] among a diverse sample of young adults attending university to: (1) calculate the heritability of tobacco use behaviors among this group, and (2) identify genetic variants contributing to tobacco use behaviors during young adulthood.

SAMPLE AND METHODS

Study population. Individuals included in this study were participants of Spit for Science ('S4S'), a longitudinal study of college students enrolled in a large, urban university in the Mid-Atlantic region of the United States. S4S is aimed at understanding how genes and environments impact substance use and mental health outcomes across time in college students[7]. To be eligible to participate in this study, incoming students had to be 18 years or older. Freshmen participants completed phenotypic assessments, covering a wide range of topics, including tobacco use. Each subsequent spring semester, students were invited to complete follow-up assessments. Students who did not participate in the first

wave of data collection (including those who turned 18 after the end of the first wave of data collection) had the opportunity to enroll in the study the following spring. Data for this study was collected and managed using REDCap electronic data capture tools[8].

Those who completed phenotypic assessments were also eligible to provide a DNA sample. A total of 9,889 university students have enrolled in S4S, including four cohorts, which matriculated from Fall 2011 to Fall 2014. Of those that enrolled into the study, 98% provided a DNA sample. Of the entire sample, 6754 indicated that they had ever used tobacco in their lifetime. Of those that indicated they had ever used tobacco in their lifetime, 5990 had provided DNA passing quality control steps, were and included in the current study, which utilizes data released in Spring 2016. The university Institutional Review Board approved study protocols and informed consent was obtained from all participants prior to participant enrollment into the study.

Tobacco Use Phenotypes. For the current study, multiple tobacco use behaviors were examined, including: ever tobacco use (e.g. indicated they had ever used tobacco in their lifetime), age of initiation, current use, regular use (e.g. indicated they had used ≥100 cigarettes in their lifetime or the equivalent), tobacco use quantity (e.g. cigarettes per day), time to first tobacco use after waking, and FTND scores. Ever, current, and regular tobacco use were treated as binary phenotypes (ever versus never). Smoking quantity, age of tobacco use initiation, and FTND were treated as continuous variables. These variables were evaluated among ever users.

Genotyping, Pre-imputation Quality Control, and Imputation. As reported and described in greater detail elsewhere[9], 6534 samples passed DNA and initial genotyping quality control (QC). Genotyping was performed at Rutgers University Cell and DNA Repository

using the Affymetrix BioBank array (653k), which contains both common GWAS framework variants (296k) for imputation, and functional variants (357k), including: rare high impact exome variants (272k), indels (18k), eQTLs (16k), and miscellaneous (51k). Off target variants were identified using SNPolisher. Variants were excluded due to: high missingness of SNPs (5%), high missingness of samples (2%), and high missingness in post-sample filtering (2%), similar to the Psychiatric Genomics Consortium (PGC)[10]. Pre-imputation QC removed 209 samples, which left 6325 samples and 560138 variants for imputation. Imputation was conducted using SHAPEIT2[11] / IMPUTE2[12] and the 1000 Genomes Phase 3 reference panel (n=2504).

Population Stratification. Given the ethnic diversity within this sample, we had to account for potential genetic heterogeneity, or when a single phenotype is caused by any one of a multiple number of alleles or non-allele mutations, caused by population stratification. Population stratification occurs when both the prevalence and allelic frequency differences exist within the population sampled for analyses. Such stratification may lead to false positive associations of genetic signals, especially when millions of markers are tested across the genome.

Genomic inflation can occur when many markers show allele frequency differences between populations and the overall distribution of test statistics is inflated. A consequence of genomic inflation can be an increase in false positives. To measure the extent of inflation due to population stratification or other potential sources of confounding, genomic control (e.g. λ, λ_{1000}) is computed. Genomic control is defined as a the median $\chi 2$ (1 degree of freedom) association statistic across SNPs divided by its theoretical median under the null distribution, with a value of 1 indicative of no stratification and a

value of >1 indicative of stratification, family structure, cryptic relatedness, differential bias, or potential other confounders. Generally, values <1.05 are considered benign; though, it is important to note that inflation in genomic control is proportional to sample size[13].

Meanwhile, methods inferring genetic ancestry have been used to correct for potential population stratification[13].The most common approach to assessing ancestry and population structure is to apply principal components analyses (PCA) to genotype data, and infer continuous axes of genetic variation. The resulting axes of variation are used to reduce the data to a small number of dimensions, or principal components, that describe as much variability as possible. These principal components are the top eigenvectors of a covariance matrix between samples and are later used in regression analyses as covariates[14]. The goal of using this approach is to maximize power for discovering etiologically relevant genetic variants, while minimizing false positive associations due to population stratification.

Ancestry Principal Components. Ancestry principal components (PCs) can be estimated either from the sample itself or from external references. Both approaches were taken in the current set of analyses. As explained by Webb et al. (2017), 1000 Genomes Project (1KGP) phase 3 variants were merged together with post quality control filtered genotypes from Spit for Science. Regions with high linkage disequilibrium were excluded, and then the common set of variants was pruned (r^2<0.1) using PLINK1.9 to yield 109,259 semi-independent variants for ancestry analyses. EIGENSOFT and smartPCA[14,15] were used to perform PCA using only the 1KGP phase 3 reference panel to determine SNP

weights for each eigenvector. The solution from the PCA was then projected onto the S4S data to generate 10 PCs[9].

Genetic Based Population Assignment. Participants were empirically assigned to 1KGP based ancestry super-populations [AFR, Africa; AMR, Americas; EAS, East Asia; EUR, Europe; and SAS, South Asia]. Using all 10 ancestry PCs, the Mahalanobis distance between the S4S sample and 1KGP population without reference population outliers (>4SD from population median, n=61) was calculated. Each subject was then assigned to the 1KGP population with the minimum Mahalanobis distance and collapsed into their respective super-population assignment. This empirically-based ancestry has several advantages to self-identified race/ethnicity including the reduction of within group variance and the ability to include “Unknown”, “More than one race”, and small groups in the analysis without an increase in genomic inflation[9].

Within Group Quality Control. To account for the diversity found in the S4S sample, filtering by Hardy-Weinberg Equilibrium (HWE), minor allele frequency (MAF), and relatedness were performed within the empirically assigned super populations. Genome-wide IBD (Π̂) was calculated using PLINK 1.9. For each group, the mean cross-sample Π̂ was calculated to find samples showing cryptic relatedness to other samples. 194 samples were excluded (>2.5 standard deviations above the mean) as outliers for average relatedness with all other samples. Clusters of probable relatives were defined using Π̂ > 0.1, Z0 >= 0.825, and Z1 < 0.175. The inclusion of Z0/Z1 is important since Π̂ > 0.1 can be due to artifacts where Z2 > 0 is extremely unlikely for cryptic relatives. Then the best performing sample for each relative cluster was retained which resulted in an additional 180 samples being excluded from the GWAS sample[9].

Within Ancestry Group PCA. As an added step to adjust for potential fine structure within each super population, within ancestry group PCA was conducted. Again, EIGENSOFT and smartPCA[14,15] were used to perform PCA for each super population found in the Spit for Science sample. Additional filtering excluded regions with high linkage disequilibrium and PLINK 1.9 was used to prune variants ($r^2 < 0.1$, MAF>0.01, HWE>$5*10^{-8}$). This yielded the following number of semi-independent variants: 71,873 (EUR); 137,042 (AFR); 84,774 (AMR); 62,046 (EAS); and 80,654 (SAS). Ten distinct PCs were generated for each super population[9].

Covariates. Within ancestry group PCs and covariates (e.g. sex and age) to include in genetic analyses were determined by stepwise linear regression for each tobacco use behavior phenotype being analyzed. Non-ancestry covariates of sex and age are kept in each model, while ancestry covariates were kept if they were retained in the best fitting model per AIC[9].

Genome-wide Complex Trait Analysis. Genome-wide Complex Trait Analysis (GCTA)[16] was used to estimate the proportion of phenotypic variance attributable to non-imputed, and directly genotyped, genetic variants [V(G)/V(P) or h^2_{SNP}]. Genetic relationship matrices, or GRMs, were derived for each ancestry group, as described in a previous study[9]. Within a mixed linear model, GCTA fits the effect of SNPs as random effects, includes the effect of sex, age, and significant within ancestry group PCs as fixed effects. The variance explained by all SNPs (e.g. the SNP-based heritability) is estimated, with heritability calculated separately by ancestry group. An ancestry group-specific minor allele frequency (MAF) cut-off of 0.01 was applied, and only unrelated individuals were

included in GRMs, resulting in the following sample sizes: N = 1339 (AFR), N = 582 (AMR), N = 557 (EAS), N = 3018 (EUR) and N=455 (SAS).

Genome Wide Association Study. Genome wide association studies (GWAS) were conducted using SNPTest[17], separately for each tobacco use behavior phenotype. Each GWAS was conducted separately by ancestry group. Association analyses were conducted under an additive model, only including markers with a minimum MAF of 0.005 and INFO of 0.5. Post-GWAS filtering was performed using ancestry specific HWE and sample size based MAFs. Rather than using a fixed MAF threshold for each group, the minimum observed minor allele count (MAC) was used, as prior research has shown a MAC of ~40 is robust for most association analyses performed in GWAS[18]. Additionally, post-filtered GWAS results were meta-analyzed using METAL[19], where sample sizes within ancestry groups were ≥400 for each tobacco use behavior phenotype. Meta-analyses using METAL implements a fixed effect model and inverse variance weighting based on sample size. Estimation of genomic inflation (λ and $\lambda_{1,000}$) for within super-population GWAS and meta-analyses was performed in R[20]. False Discovery Rate (FDR) analysis was performed using the "q-value" package (https://github/jdstorey/qvalue) using Bioconductor 3.2[21]. Genomic bins were defined for follow-up, starting with all markers with a q-value < 0.25. Initially, markers were collapsed into bins if they were within 10kb; however, post-hoc inspection showed several adjacent bins <75kb apart which were then collapsed into reported bins. The web-based plotting tool, LocusZoom[22], was used to visually display regional information regarding the strength and extent of the association signals relative to genomic position, local linkage disequilibrium, and recombination

patterns and positions of genes in the region, was used to inspect genome-wide significant SNPs.

Individual Variant Replication. Summary statistics from GWAS results of the Tobacco and Genetics (TAG) Consortium were extracted and compared to the corresponding phenotypes found in S4S. Due to differences in allele frequencies across the discovery and replication, summary statistics were not available for all markers. Replication was attempted for genome-wide significant SNPs found in TAG; however, only results based on equivalent phenotypes were examined. Nominal associations were found where p-value <0.05.

RESULTS

Descriptive statistics. Within the genetic sample, the sample was mostly female (61.2%), and had indicated that they had used any tobacco product at least once in their lifetime (67.9%). Frequencies and percentages for tobacco use behaviors within the entire sample and genetic sample only are presented in Table 9.1 below. [Though the frequencies and percentages are not shown, cohorts were equally represented.] Sex, age, and significant PCs were included as covariates in genetic analyses.

Table 9.1: Tobacco Use Behaviors of the Spit for Science Sample

Tobacco Use Phenotype	ENTIRE SAMPLE N=9889		GENETIC SAMPLE N=5990	
	N	**%**	**N**	**%**
Ever Tobacco Use	6754	68.3%	4069	67.9%
Age of Initiation				
Did not initiate	3091	31.3%	1839	39.0%
>= 18 years	1109	11.2%	806	17.1%
15-18 years	2531	25.6%	1282	27.2%
12-14 years	1094	11.1%	673	14.3%
<12 years	249	2.5%	114	2.4%
Current Use	4049	40.9%	2490	61.3%
Regular Use				
Never Used	1909	19.3%	2855	48.3%
<100 cigarettes/lifetime	3316	33.5%	2094	35.4%
>=100 cigarettes/lifetime	1529	15.5%	959	16.2%
Regular Use Among Smokers				
<100 cigarettes/lifetime	5225	52.8%	3110	76.4%
>=100 cigarettes/lifetime	1529	15.5%	959	23.6%
Cigarettes Per Day				
<10 cigarettes	3220	32.6%	1870	31.4%
11-20 cigarettes	375	3.8%	241	4.0%
21-30 cigarettes	112	1.1%	67	1.1%
>31 cigarettes	69	0.7%	42	0.7%
Time to First Tobacco Use				
>60 minutes	3230	32.7%	1915	73.9%
31-60 minutes	448	4.5%	267	10.3%
6-30 minutes	470	4.8%	296	11.4%
Within 5 minutes	200	2.0%	115	4.4%
	Mean	**Range**	**Mean**	**Range**
Fagerström Test for Nicotine Dependence	0.82	0-10	0.84	0-10

GCTA/Heritability estimates. Heritability estimates for each tobacco use behavior were calculated using GCTA[16], separately by ancestry group, as shown in Table 9.2. The heritability of ever tobacco use ranged from 0.00 to 0.28, but was only significant amongst those of East Asian (h^2 =0.13, p=0.0480) and European ancestry (h^2 =0.28, p=0.0133). Significant heritability was observed for age of initiation in those of European ancestry (h^2 =0.30, p=0.0288). None of the heritability estimates were significant for current use, regular use, and time to first tobacco use after waking across any ancestry group, while the heritability of regular use among smokers within those of European ancestry (h^2 =0.50, p=0.0041), cigarettes per day and FTND scores within the Americas ancestry group (CPD: h^2 =1.00, p=0.0212; FTND: h^2 =1.00, p=0.0211) were statistically significant at a p-

value of ≤ 0.05. However, after corrections for multiple testing, none of these results were significant at the adjusted threshold of p-value ≤ 0.00125 (or p-value ≤0.05/40 tests; where 40 tests account for eight phenotypes x five ancestry groups). Additionally, the sample sizes for the AMR, EAS, and SAS ancestry groups were smaller than suggested for GCTA analyses[23]; so those results should be interpreted with caution.

Table 9.2: Heritability Estimates for Tobacco Use Behaviors, by Ancestral Group

Tobacco Use Behavior	Ancestry Group	N	h^2_{SNP}	SE	p-value
Ever Tobacco Use	AFR	1329	0.00	0.26	0.5000
	AMR	577	0.00	0.44	0.5000
	EAS	**552**	**0.13**	**0.59**	**0.0480**
	EUR	**2980**	**0.28**	**0.13**	**0.0133**
	SAS	453	0.00	0.71	0.5000
Age of Initiation	AFR	897	0.44	0.37	0.1214
	AMR	487	0.00	0.51	0.5000
	EAS	460	0.52	0.72	0.2258
	EUR	**2523**	**0.30**	**0.16**	**0.0288**
	SAS	334	0.00	0.97	0.5000
Current Use	AFR	884	0.00	0.38	0.5000
	AMR	412	0.49	0.56	0.2046
	EAS	297	0.00	1.22	0.5000
	EUR	2172	0.00	0.18	0.5000
	SAS	284	0.99	1.11	0.1027
Regular Use	AFR	1329	0.20	0.25	0.2045
	AMR	577	0.13	0.46	0.3949
	EAS	552	0.00	0.61	0.5000
	EUR	**2980**	**0.26**	**0.13**	**0.0221**
	SAS	453	0.00	0.74	0.5000
Regular Use Among Smokers	AFR	886	0.16	0.35	0.3152
	AMR	412	0.40	0.59	0.2728
	EAS	297	0.00	1.18	0.5000
	EUR	**2175**	**0.50**	**0.18**	**0.0041**
	SAS	285	0.90	1.05	0.1911
Cigarettes Per Day	AFR	402	0.53	0.61	0.1758
	AMR	**224**	**1.00**	**1.01**	**0.0212**
	EAS	148	0.00	2.03	0.5000
	EUR	1307	0.06	0.30	0.4181
	SAS	139	0.48	1.86	0.3980
Time to First Tobacco Use After Waking	AFR	471	0.00	0.59	0.5000
	AMR	269	0.13	0.85	0.4450
	EAS	173	0.00	1.82	0.5000
	EUR	1519	0.25	0.25	0.1520
	SAS	157	0.63	1.78	0.3651
Fagerström Test for	AFR	670	0.23	0.48	0.3248
	AMR	**324**	**1.00**	**0.68**	**0.0021**
	EAS	233	1.00	1.31	0.1449

Nicotine Dependence	EUR	1773	0.31	0.22	0.0832
	SAS	217	1.00	1.40	0.1510

Primary GWAS results. Prior to applying filtering, 17,461,305 markers were available for analyses. The number of available markers following the application of filtering and meta-analyses are shown in Table 9.3, in addition to the sample size, and measures of genomic inflation (λ and λ_{1000}). To be included in the meta-analyses, sample sizes had to be ≥400 within a given ancestry group.

Table 9.3: Sample Sizes, Marker Counts, and Genomic Inflation Estimation

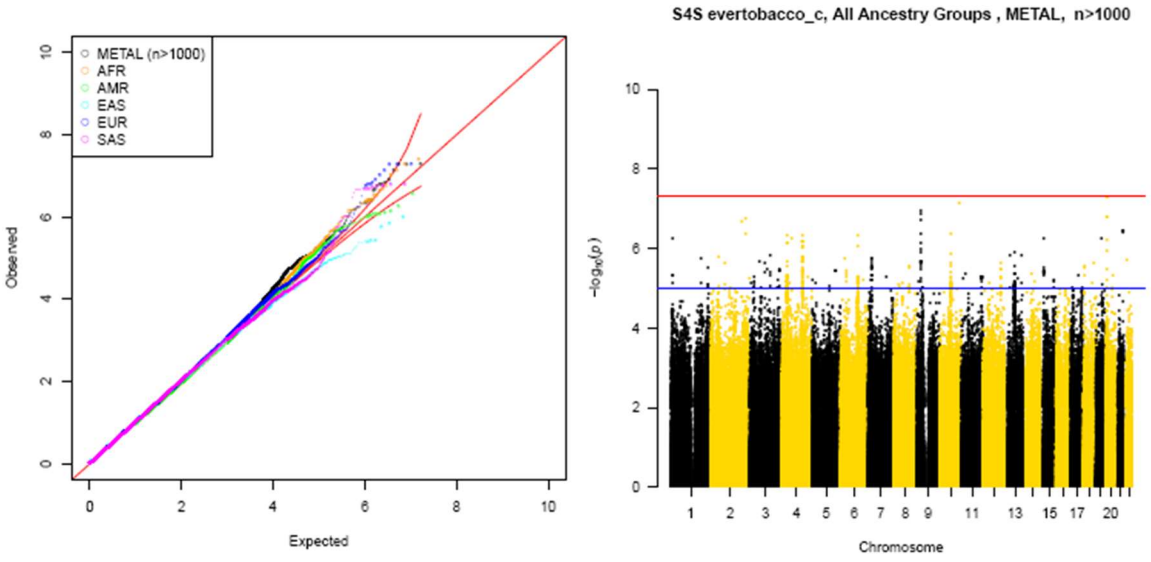
		AFR	AMR	EAS	EUR	SAS	Metal
Ever Use	n	1329	577	552	2980	453	>1000
	n Markers	14466464	10909721	6835923	10333295	7361150	16638242
	λ	1.0095	1.0220	1.0238	1.0178	1.0031	1.0034
	λ_{1000}	1.0016	1.0037	1.0040	1.0030	1.0005	1.0034
Age of Initiation	N	897	487	460	2523		>1000
	n Markers	12797220	10213567	6598175	10106533		10332814
	λ	0.9900	0.9617	1.0171	0.9970		1.0150
	λ_{1000}	0.9977	0.9912	1.0039	0.9993		1.0000
Current Use	N	884	412		2172		>1000
	n Markers	12731483	9542115		9855533		9941997
	λ	1.0057	1.0513		1.0020		1.0019
	λ_{1000}	1.0016	1.0147		1.0005		1.0000
Regular Use	N	1329	577	552	2980	453	>1000
	n Markers	14462978	8037689	6833967	10336157	7361674	16941407
	λ	0.9819	0.9738	0.9943	1.0069	0.9788	0.9934
	λ_{1000}	0.9969	0.9956	0.9990	1.0012	0.9964	1.000
Regular Use Among Smokers	N	886	412		2175		>1000
	n Markers	12717222	7339884		9856239		9941327
	λ	1.0385	1.0073		1.0206		1.0136
	λ_{1000}	1.0110	1.0021		1.0059		1.000
CPD	N	402			1307		>1000
	n Markers	6874724			6874724		6861522

	λ	0.9982			0.9999		1.0008
	λ_{1000}	0.9989			0.9999		1.0000
TFT	N	471			1519		>1000
	n Markers	4839510			4839810		4835232
	λ	0.9923			1.0047		0.9958
	λ_{1000}	0.9933			1.0016		0.9999
FTND	N	670			1773		>1000
	n Markers	7497960			7497960		7486766
	λ	1.0032			0.9904		1.0014
	λ_{1000}	1.0013			0.9960		1.0000

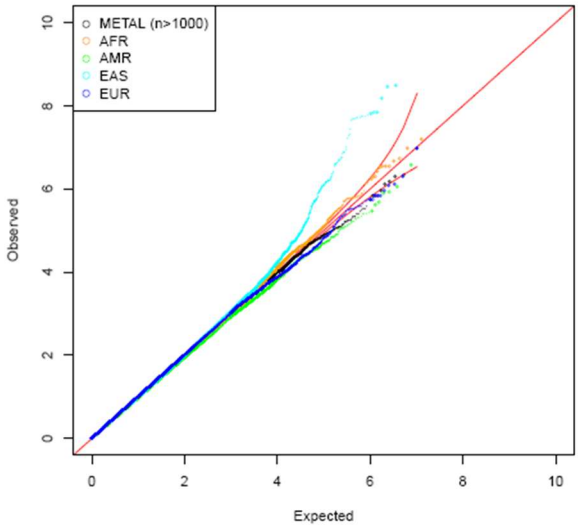
The meta-analyses showed no evidence of genomic inflation, as demonstrated by with $\lambda\ and\ \lambda_{1000}$s shown in Table 9.3. Analyses were adjusted for sex, age, and significant within ancestry group principal components (identified by step-wise linear regression). Figure 9.1, depicts the QQ- and Manhattan Plots for each of the tobacco use behaviors.

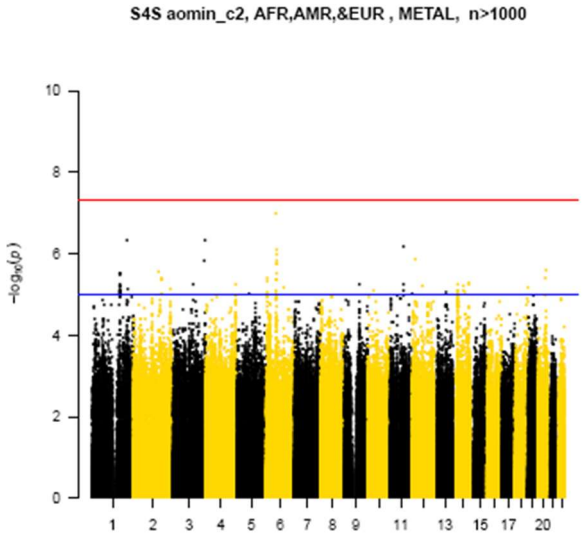
Figure 9.1: QQ-Plots and Manhattan Plots for Tobacco Use Behaviors

Ever Tobacco Use

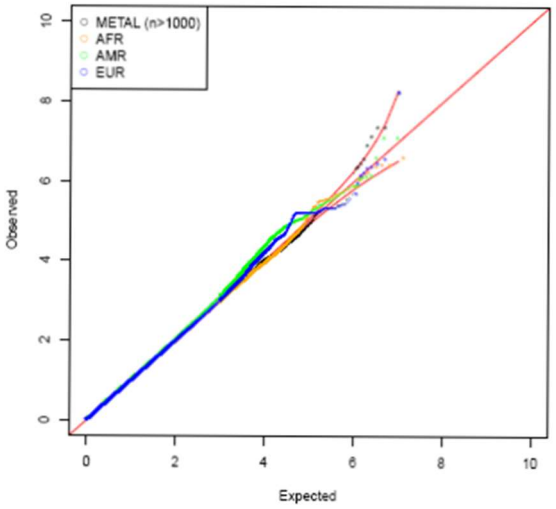


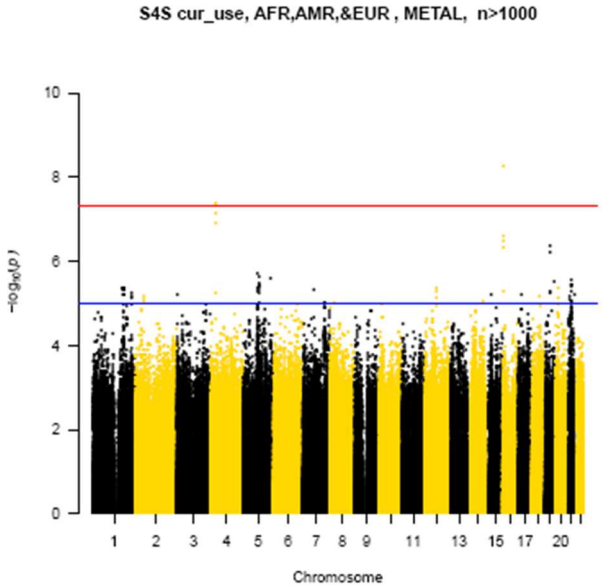
Age of Initiation



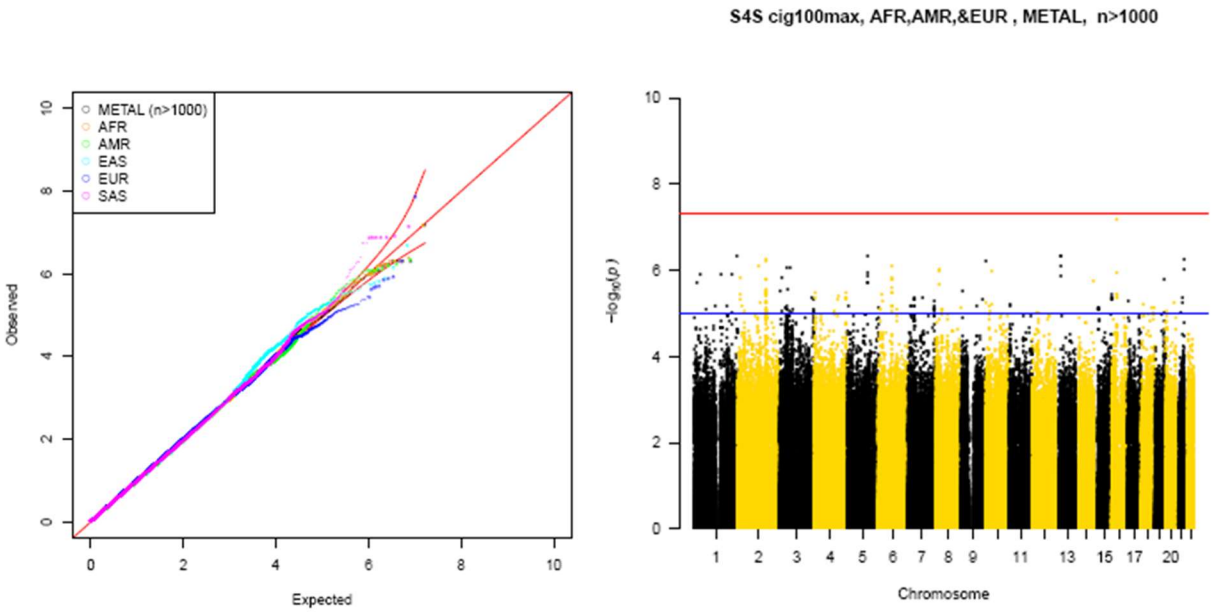


Current Use

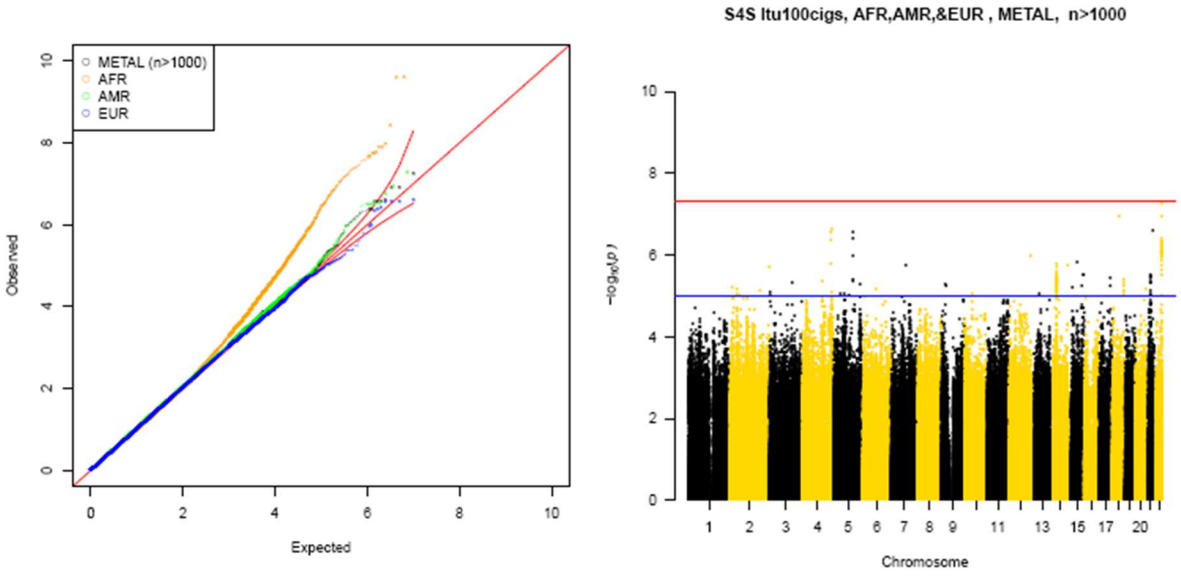




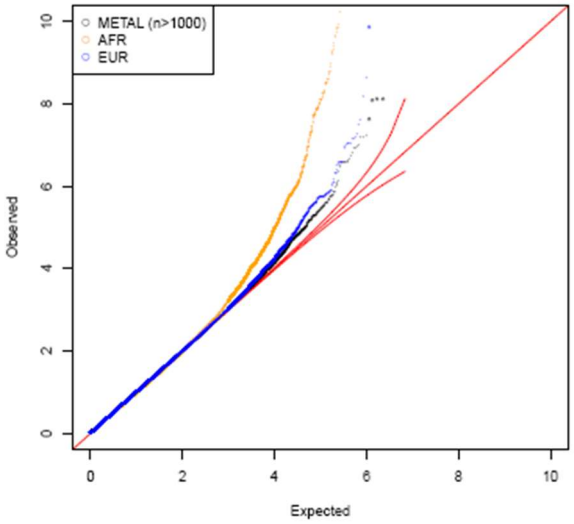
Regular Use

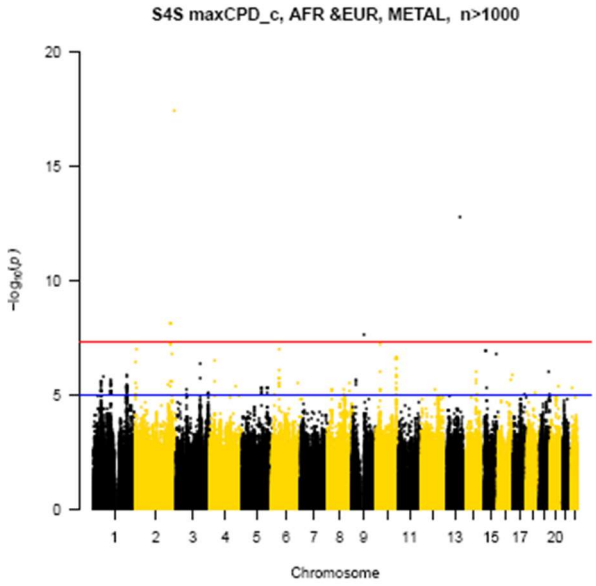


Regular Use Among Smokers

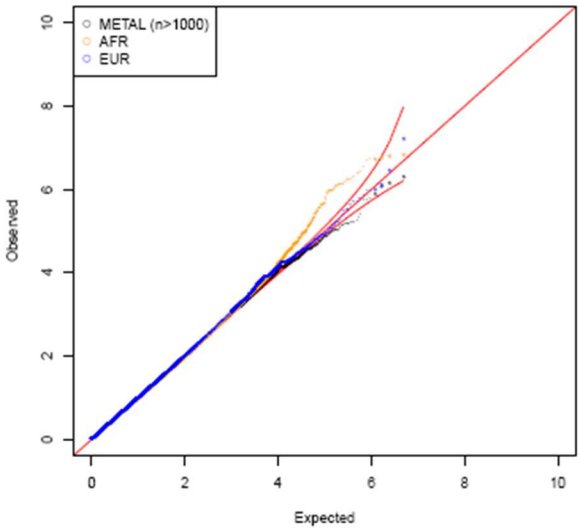


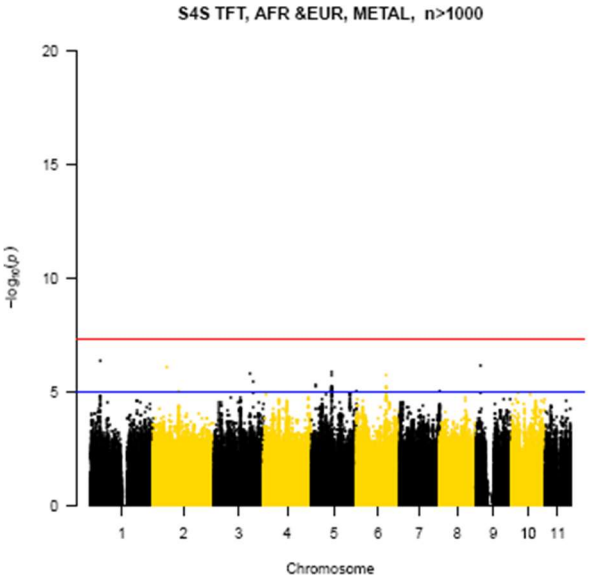
Cigarettes Per Day



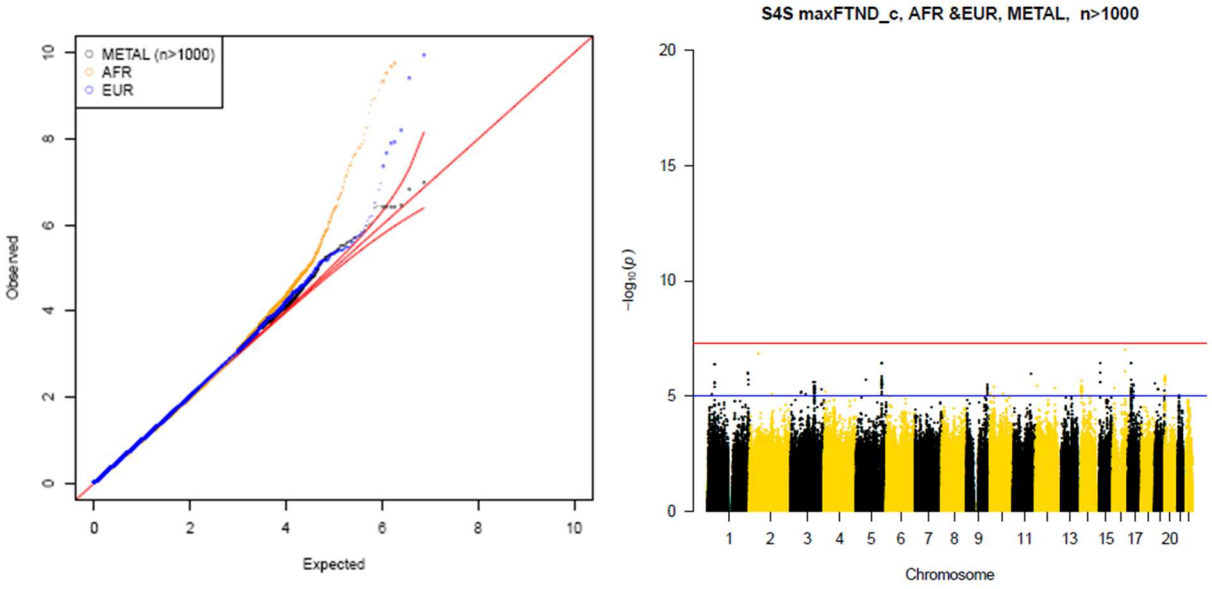


Time to First Tobacco Use After Waking





FTND



No genome-wide significant SNPs were identified for ever tobacco use, age of initiation, regular use, regular use among smokers, TFT and FTND. Significant findings were found for current use and cigarettes per day, which are described in further detail below, and shown in Table 9.4. Across the measures of current use and cigarettes per day, FDR analysis showed 38 markers with q<0.25. Each of these markers are listed in Supplementary Tables 9.7a-9.8b, with summary statistics and information on nearby genes within 75kb. These 37 markers map to 20 genomic bins, with seven genomic bins containing at least one genome-wide significant marker (p-value ≤ 5 x 10^{-8}). Regional association plots (Figures 9.2a and 9.2b) are only shown for the genome-wide significant SNPs that are found to be located within genes (e.g. rs148027841 and rs9653371).

Current Use. Three SNPs (rs148027841, rs73111343, rs73111344) were found to be associated with current use, at the significance threshold level of p-value < 5 x 10^{-8}. SNP rs148027841 is located on chromosome 16 within the protein coding gene RAB11 Family

Interacting Protein 3 or (RAB11FIP3) that has a regulatory role in the formation, targeting, and fusion of intracellular transport vesicles[24]. Figure 9.2a depicts the regional association plot for RAB11FIP3, and displays other markers that are not in high linkage disequilibrium with rs148027841, but have low p-values. rs148027841 is rare (MAF ≤ 0.02) in each of the ancestry groups. The direction of the effect was consistent across the African and European ancestry groups (e.g. negative), but in the opposite direction across the Americas ancestry group (e.g. positive). The strength of the association varied by ancestry group (e.g. -0.35 in AFR; 0.13 in AMR; and -1.40 in EUR). The other two genome-wide significant SNPs, rs73111343 and rs73111344, are common in each of the listed ancestry groups and are in high LD with one another, but are not located within any genes.

Cigarettes Per Day. Six SNPs were found to be associated with cigarettes per day (rs9653371, rs34731037, rs71427733, rs75714873, rs41319146, and rs371955890) at the genome-wide significance level. rs9653371 mapped onto a gene located on chromosome 2 (PID1), and is rare within those of European ancestry (MAF = 0.14) and is common within those of African ancestry (MAF = 0.02). Figure 9.2b depicts the regional association plot for PID1, which suggests that rs9653371 is a lone SNP, with no linkage disequilibrium with other SNPs. The direction of the effect is consistent across these two ancestry groups, but the strength of association is greater within those of European ancestry, relative to African ancestry. Although SNP rs34731037, located on chromosome 13, did not map onto any genes, it is common within those of African and European ancestry (MAF = 0.06, 0.07 respectively). The direction of the effect is consistent across these two groups, but the strength of association is greater within those

of African ancestry relative to those of European ancestry. Genome-wide significant SNPs (rs71427733, rs75714873, and rs41319146) located on chromosome 2 were also not located within any genes (or nearby any genes within 75kb) and are common within those of African and European ancestry (MAF = 0.06 for both ancestry groups). The direction of the effect is consistent across the two ancestry groups, but the strength of association is greater within the African ancestry group. rs371955890, located on chromosome 9, is not located within any genes, but is nearby to AK096159 and LOC100132352. However, the direction of the effect is not consistent across African and European ancestry groups, and the effect size is greater within the African ancestry group – even though the SNP has a higher MAF within the European ancestry group.

Table 9.4: Genome-wide significant SNPs for Current Use and Cigarettes Per Day

Phenotype	Ancestry Group	N	Missing Proportion	INFO	MAF	HWE	P-value	Beta	SE
	rs148027841 (chromosome 16)								
Current Use	AFR	884	1.7E-06	0.69	0.01	1.00	0.58	-0.35	0.62
	AMR	412	NA	0.79	0.01	1.00	0.85	0.13	0.71
	EUR	2172	1.8E-06	0.75	0.02	1.00	5.5E-09	-1.40	0.24
	rs73111344 (chromosome 4)								
Current Use	AFR	884	5.1E-06	0.87	0.10	0.85	3.2E-06	0.84	0.19
	AMR	412	4.8E-06	0.92	0.06	0.63	1.0E-04	1.30	0.35
	EUR	2172	3.2E-06	0.91	0.08	1.00	5.5E-09	0.33	0.13
	rs73111343 (chromosome 4)								
Current Use	AFR	884	5.1E-06	0.87	0.10	0.85	3.2E-06	0.84	0.19
	AMR	412	6.1E-06	0.92	0.06	0.63	1.0E-04	1.30	0.35
	EUR	2172	3.2E-06	0.91	0.08	1.00	5.5E-09	0.33	0.13
	rs9653371 (chromosome 2)								
Cigarettes Per Day	AFR	402	1.1E-05	0.84	0.14	0.84	0.76	0.04	0.11
	EUR	1307	1.1E-06	0.51	0.02	1.00	2.0E-22	3.26	0.20
	rs34731037 (chromosome 13)								
Cigarettes Per Day	AFR	402	1.9E-05	0.53	0.06	1.00	1.9E-27	2.52	0.17
	EUR	1307	1.1E-05	0.67	0.07	0.52	0.02	0.27	0.10
	rs71427733 (chromosome 2)								
Cigarettes Per Day	AFR	402	1.2E-06	0.96	0.06	1.00	4.6E-05	0.64	0.16
	EUR	1307	1.2E-06	0.96	0.06	0.62	1.4E-05	0.36	0.08
	rs75714873 (chromosome 2)								
Cigarettes Per Day	AFR	402	1.2E-06	0.96	0.06	1.00	4.6E-05	0.64	0.16
	EUR	1307	1.2E-06	0.96	0.06	0.62	1.4E-05	0.36	0.08
	rs41319146 (chromosome 2)								
Cigarettes Per Day	AFR	402	3.4E-16	0.95	0.06	1.00	4.7E-05	0.64	0.16
	EUR	1307	7.7E-07	0.96	0.06	0.62	1.4E-05	0.36	0.08
	rs371955890 (chromosome 9)								
Cigarettes Per Day	AFR	402	2.0E-05	0.50	0.05	0.61	6.4E-37	3.00	0.12
	EUR	1307	2.8E-05	0.60	0.11	0.27	0.52	-0.05	0.08

Figure 9.2a: Locus Zoom plot for rs148027841 on chromosome 16, associated with current use

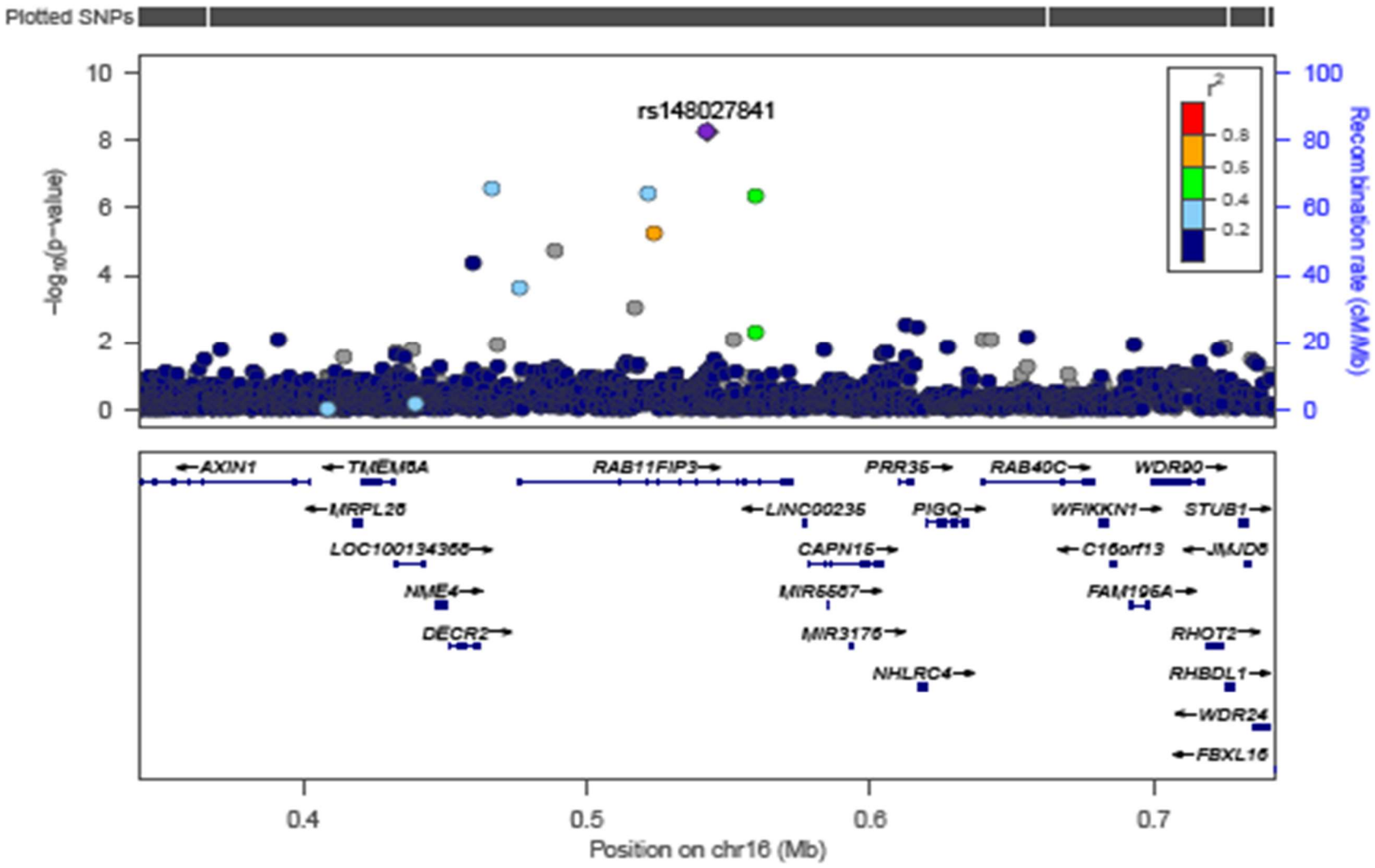


Figure 9.2b: Locus Zoom Plot for rs9653371 on chromosome 2, associated with cigarettes per day

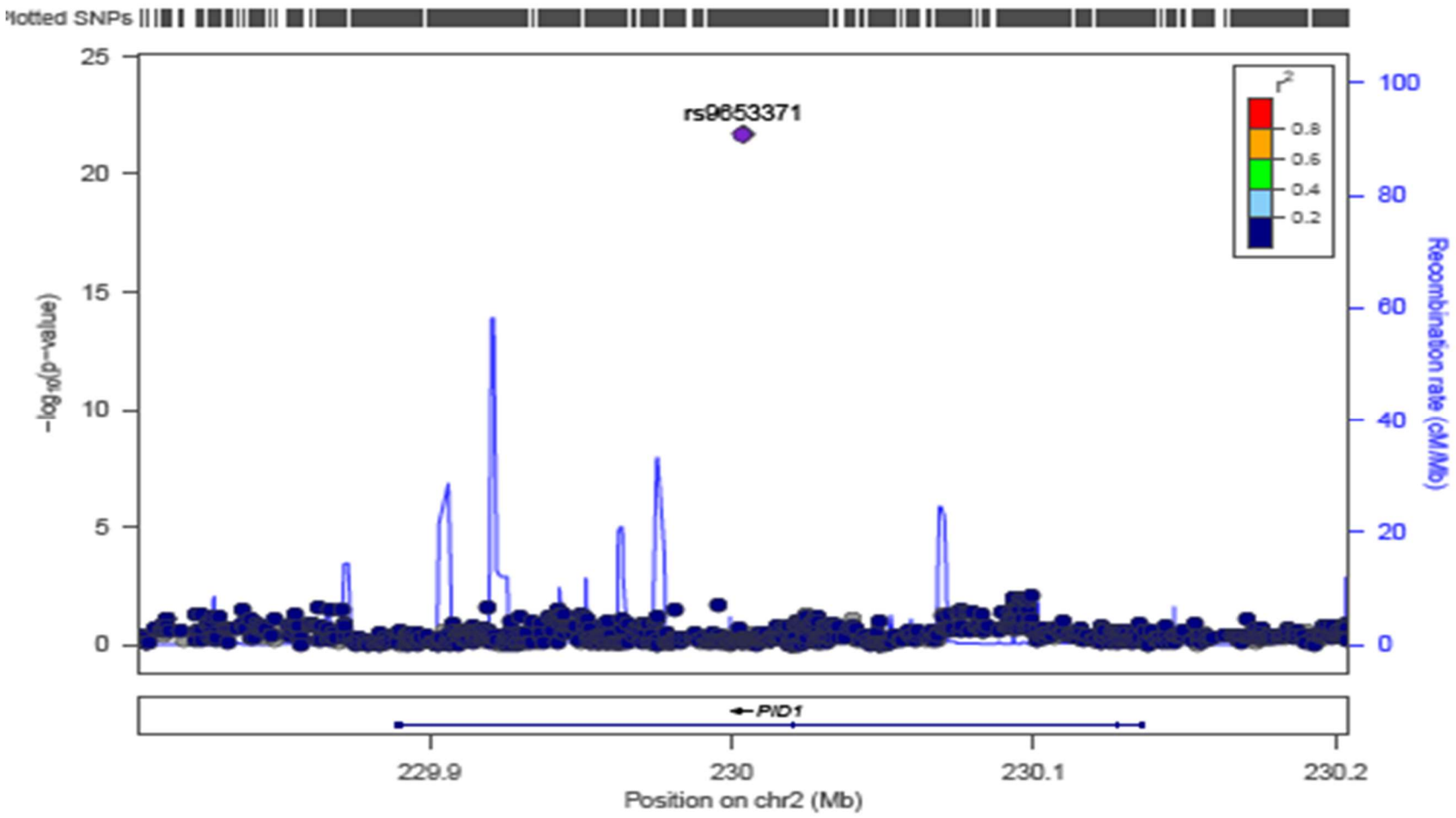


Table 9.5: Individual Variant Replication Summary

Phenotype	SNP	CHR	BP	Alleles	Tobacco and Genetics Consortium (EUR)						Spit for Science (EUR)					
					N	Coded AF	INFO	OR	SE	P	N	MAF	INFO	Beta	SE	P
CPD	rs1051730	15	76681394	G/A	38,181	0.65	1.00	-1.02	0.09	8.0E-33	1,307	0.33	1.00	0.04	0.04	0.27
	rs16969968	15	76669980	G/A	38,181	0.65	1.00	-1.02	0.09	4.5E-33	1,307	0.33	1.00	0.05	0.04	0.16
	rs1329650	10	93338100	T/G	38,181	0.28	1.00	-0.43	0.09	2.3E-06	1,307	0.27	1.00	-0.07	0.04	0.11
	rs1028936	10	93339777	C/A	37,284	0.18	1.00	-0.55	0.12	1.6E-06	1,307	0.17	0.95	-0.04	0.05	0.49
	rs3733829	19	46002411	G/A	38,181	0.36	1.00	0.35	0.09	7.7E-05	1,307	0.35	1.00	-0.06	0.04	0.11
Regular Use	rs6265	11	27636492	T/C	74,035	0.21	1.00	-0.06	0.01	1.7E-05	2,980	0.19	1.00	-0.001	0.03	0.97
	rs1013442	11	27535522	T/A	74,035	0.26	1.00	-0.06	0.01	3.4E-05	2,980	0.24	0.99	0.02	0.03	0.57
	rs4923457	11	27605156	T/A	74,035	0.23	1.00	-0.06	0.01	2.1E-05	2,980	0.21	0.99	0.03	0.03	0.27
	rs4923460	11	27613365	T/G	74,035	0.23	1.00	-0.06	0.01	2.2E-05	2,980	0.21	0.99	0.03	0.03	0.27
	rs4074134	11	27603861	T/C	74,035	0.23	1.00	-0.06	0.01	1.9E-05	2,980	0.21	1.00	0.03	0.03	0.27
	rs1304100	11	27528179	G/A	74,035	0.26	1.00	-0.06	0.01	4.9E-05	2,980	0.24	0.99	0.02	0.03	0.53
	rs6484320	11	27659764	T/A	74,035	0.24	1.00	-0.06	0.01	2.0E-05	2,980	0.22	0.97	-0.03	0.03	0.39
	rs879048	11	27595510	C/A	74,035	0.23	1.00	-0.06	0.01	2.3E-05	2,980	0.21	0.98	-0.03	0.03	0.27

Identification of Novel Genetic Variants and Replication of TAG SNPs in S4S. Despite being able to identify nine genome-wide significant SNPs contributing to either current use or cigarettes per day in the S4S sample, none have been associated with tobacco use behaviors in any previous studies. In attempts to replicate findings for individual variants from the TAG Consortium using the S4S dataset, only results based on equivalent phenotypes were examined (e.g. markers associated with cigarettes per day and regular use (e.g. which is like the current study's measure of ever use), as shown in Table 9.5. None of the individual variants from the TAG Consortium demonstrated nominal associations (p-value < 0.05) in S4S.

DISCUSSION

Study findings support previous research on complex, polygenic nature of tobacco use behaviors, and suggest the need for additional research investigating the role of genes contributing to tobacco use behaviors. More specifically, by using a population-based study of university students, we found that there are common SNPs contributing to tobacco use behaviors. Estimates of SNP-based heritability using GCTA indicate that tobacco use behaviors are moderately heritable, but only within those of the European ancestry group. Although these estimates are generally lower than those estimated from twin studies, they were non-zero and significant those of European ancestry (at least before applying multiple testing corrections), indicating that the existing sample sizes should be adequate for calculations of SNP-based heritability. The expected range heritabilities for smoking phenotypes are between 0.4 and 0.7 - at least according to twin and family studies of individuals of European ancestry. However, prior research suggests that SNP-based heritabilities using GCTA are typically 50% of that of found within twin

studies. For this reason, Table 9.6 shows a range of heritabilities from 0.2 to 0.7. From power calculations using the GCTA-GREML Power Calculator, there was sufficient (e.g. >80% power) to detect SNP-based heritabilities between 0.5 and 0.7, with sample sizes of >1,500 individuals (Supplementary Table 9.6).

Meanwhile, ancestry groups with smaller sample sizes often yielded nonsignificant heritability estimates close to zero. This could mean that either the sample sizes for these groups are too small and we do not have enough power to calculate heritability using GCTA, or that the phenotypic traits being measured are weakly (or not at all) affected by genetic variation within this sample. Since published evidence suggests these tobacco use behaviors are similarly heritable across different populations[25], it could be that larger sample sizes are needed to estimate heritability across South Asian, American, and East Asian ancestry groups. What this means is that to get the statistical power needed to determine associations between the aggregate effect of genetic variants and tobacco use behaviors for the remaining ancestry groups, larger samples may be needed.

Like other studies, the current study is not without limitations. Firstly, all the tobacco use behaviors used within this study (and all the studies included within this dissertation thesis) is based upon self-report data. Thus, the findings from this (and subsequently, all studies included in this dissertation thesis) are subject to potential reporting bias and does not offer the possibility for any external verification (e.g. outsider reporting, or verification using biomarkers, such as cotinine levels). To mitigate potential reporting bias, we examined tobacco use variables across each wave of individual data and recoded variables to maintain consistency.

Additionally, the current study was not able to replicate findings from the larger meta-analyses of tobacco use behavior. Failures to replicate significant GWAS hits from the TAG Consortium could be due to many reasons, including: variability in phenotype definitions across independent samples, inadequate sample size, false positive results, and population-specific effects[19], as described in fuller detail on the next page.

Variability in phenotype definitions can result from differences in measurement protocols across studies. The definitions used in the current study are different than definitions used in previous studies. For example, within the studies included within larger consortiums, “ever use” was defined as having used at least 100 cigarettes in a lifetime, which matches more closely with our “regular use” variable. Meanwhile, “ever use” within the context of the current study is defined as having used any tobacco product during a lifetime. This means that individuals within previous studies will have had more exposure to nicotine, relative to individuals within the current study which could influence association findings.

Another potential contributor to the failure to replicate SNP associations (even when comparing similar phenotypes of “regular use” in our study to “ever use” of previous studies) is inadequate sample size. Studies finding significant SNP associations with tobacco use phenotypes report larger sample sizes than that found in this study. For example, the ENGAGE Consortium includes more than 30,000 genotyped participants for each of the tobacco-related phenotypes[20]. Sample size is important because it directly impacts power, or the chance of discovering effects. Generally, low-powered studies produce more false negatives relative to higher-powered studies and have reduced probabilities of observing effects that pass the required threshold of claiming discovery (e.g. reaching statistical significance threshold). Additionally, even when true effects are

found within low powered studies, it is likely that the estimate of the magnitude of the effect is exaggerated[21]. Potentially, what this means for the current study is that because the study was underpowered (as evidenced by a post-hoc power calculation demonstrating that we have <80% power to replicate findings from the larger TAG Consortium), the sample sizes were not large enough to find any associations across certain tobacco use behaviors (e.g. ever tobacco use, age of initiation, regular use, and regular use among smokers) and/or the effect of the significant associations found for current use, time to first tobacco use after waking, cigarettes per day, and the Fagerström Test for Nicotine Dependence may be exaggerated. Alternatively, in the case where effect sizes are already small, false positives are likely. To correct for any false positives driven by population stratification, we conducted genetic analyses separately by ancestry and included principal components as covariates in the models. Furthermore, we calculated lambda inflation factors and implemented a false discovery rate correction.

Despite these precautions, it is also possible that our failure to replicate GWAS hits from previous studies is due to population-specific effects, or the possibility that detected effects are stronger within a specific sample, rather than the general population. Given that most tobacco-related GWAS have been conducted within older adults of European ancestry and focused on cigarette use, it is possible that the findings from previously conducted genetic studies of tobacco use behaviors are limited to that subset. This study diverges from these studies in that it includes younger individuals of varied ancestral backgrounds and has a broader definition of tobacco use. Participants were included in analyses if they had indicated they had ever used any of the following tobacco products: cigarettes, snus, cigars, hookah and e-cigarettes. The reason why these measures were

aggregated into the broader umbrella of tobacco, was that the sample sizes across each tobacco product, except for cigarettes, were not adequate to conduct separate analyses. Perhaps future studies, with larger samples of individuals who use alternative tobacco products and electronic nicotine delivery systems, will be able to identify genetic variants influencing the use of different tobacco products, or assess whether this is primarily driven by the availability of specific products. Although it is possible that the associations found within this study might not be generalizable to other genetic studies of tobacco use behaviors, a couple of associations not previously reported were found between tobacco use behaviors and markers localized within or near genes of possible biological interest. However, the robustness of these findings is limited, and require further investigation.

Finally, given that sample is representative of university students, it is relatively young and participant's behaviors may change with time. This means is that the estimated heritability, accounted for by genetic variants may change. Previous studies suggest that genetic factors may become more influential as participants mature, so we would expect the contribution of genes to increase over time, if we were to follow these individuals as they grow older. But again, further research is required.

Supplementary Table 9.6: Power Calculation Summary

TOBACCO USE BEHAVIOR	Ancestry Group	N	$h^2 = 0.70$	$h^2 = 0.60$	$h^2 = 0.50$	$h^2 = 0.40$	$h^2 = 0.30$	$h^2 = 0.20$
EVER USE	ALL	5891	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**0.961**
	AFR	1329	**0.837**	0.713	0.556	0.390	0.243	0.134
	AMR	577	0.248	0.195	0.149	0.113	0.085	0.065
	EAS	552	0.231	0.182	0.141	0.108	0.082	0.064
	EUR	2980	**1.000**	**1.000**	**0.997**	**0.965**	**0.807**	0.470
	SAS	453	0.171	0.138	0.111	0.088	0.071	0.060
AGE OF INITIATION	ALL	4367	**1.000**	**1.000**	**1.000**	**1.000**	**0.986**	0.789
	AFR	897	0.510	0.398	0.294	0.206	0.136	0.088
	AMR	487	0.190	0.152	0.120	0.095	0.075	0.061
	EAS	460	0.175	0.141	0.113	0.090	0.072	0.060
	EUR	2523	**1.000**	**0.998**	**0.979**	**0.891**	0.668	0.358
CURRENT USE	ALL	3468	**1.000**	**1.000**	**1.000**	**0.992**	**0.908**	0.592
	AFR	884	0.358	0.389	0.287	0.358	0.134	0.087
	AMR	412	0.149	0.122	0.100	0.082	0.068	0.058
	EUR	2172	**0.998**	**0.985**	**0.930**	0.785	0.540	0.279
REGULAR USE	ALL	5891	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**0.961**
	AFR	1329	**0.837**	0.713	0.556	0.390	0.243	0.134
	AMR	577	0.248	0.195	0.149	0.113	0.085	0.065
	EAS	552	0.231	0.182	0.141	0.108	0.082	0.064
	EUR	2980	**1.000**	**1.000**	**0.997**	**0.965**	**0.807**	0.470
	SAS	453	0.171	0.138	0.111	0.088	0.071	0.060
REGULAR USE AMONG SMOKERS	ALL	3473	**1.000**	**1.000**	**1.000**	**0.993**	**0.909**	0.594
	AFR	886	0.501	0.390	0.288	0.202	0.134	0.087
	AMR	412	0.149	0.122	0.100	0.082	0.068	0.058
	EUR	2175	**0.998**	**0.985**	**0.930**	0.786	0.541	0.280
CPD	ALL	1709	**0.966**	**0.990**	0.771	0.580	0.368	0.191
	AFR	402	0.145	0.119	0.097	0.080	0.067	0.057
	EUR	1307	**0.825**	0.698	0.543	0.380	0.236	0.131
TFT	ALL	1990	**0.993**	**0.965**	**0.882**	0.711	0.471	0.242
	AFR	471	0.181	0.145	0.116	0.092	0.073	0.060
	EUR	1519	**0.920**	**0.822**	0.671	0.485	0.302	0.161
FTND	ALL	2443	**1.000**	**0.996**	**0.972**	**0.871**	0.640	0.339
	AFR	670	0.317	0.246	0.185	0.136	0.097	0.071
	EUR	1773	**0.975**	**0.920**	**0.801**	0.611	0.391	0.202

Boldface indicates power ≥80%

Supplementary Table 9.7a: SNPs Contributing to Current Use, Following FDR Correction

SNP	CHR	A1	A2	AFRICAN ANCESTRY				AMERICAN ANCESTRY				EUROPEAN ANCESTRY				WEIGHT	ZSCORE	P	DIRECTION	Q_1K
				INFO	MAF	HWE	P	INFO	MAF	HWE	P	INFO	MAF	HWE	P					
rs148027841	16	a	g	0.695	0.009	1.000	0.5754	0.795	0.014	1.000	0.852	0.746	0.024	1.000	0.000	2172	5.832	5.48E-09*	??+	0.054
rs73111343	4	c	g	0.872	0.103	0.855	3.17E-06	0.925	0.059	0.628	0.000	0.909	0.083	0.777	0.023	3468	5.473	4.41E-08*	+++	0.146
rs73111344	4	a	g	0.872	0.103	0.855	3.17E-06	0.925	0.059	0.628	0.000	0.909	0.083	0.777	0.023	3468	5.474	4.41E-08*	+++	0.146
rs10489015	4	t	c	0.887	0.100	0.849	3.68E-06	0.952	0.057	0.623	0.000	0.920	0.082	0.670	0.025	3468	5.379	7.48E-08	+++	0.186
rs73111362	4	a	t	0.884	0.100	0.851	4.01E-06	0.950	0.057	0.623	0.000	0.917	0.082	0.667	0.032	3468	-5.290	1.22E-07	---	0.243

CHR = chromosome; A1 = allele 1; A2 = allele 2; INFO = imputation quality; MAF – minor allele frequency; HWE = Hardy Weinberg Equilibrium; P = p-value; Q_1k = q-value

Supplementary Table 9.7b: Genomic Bins for SNPs Contributing to Current Use, Following FDR Correction

CHR	START BP	END BP	NSNP	P-VALUE	Q-VALUE	N	WITHIN GENES	NEARBY GENES (75KB)
16	542536	542536	1	5.48E-09	0.054365312	2172	RAB11FIP3	C16orf11, LINC00235,MIR3176,MIR5587,NHLRC4,PIGQ,RAB11FIP3,SOLH
4	27350593	27350679	2	4.41E-08	0.146066436	3468	None	None
4	27368605	27373502	2	7.48E-08	0.185710913	3468	None	None

CHR = chromosome; Start BP = starting base pair; End BP = ending base pair; NSNP = number of SNPs within genomic bin

Supplementary Table 9.8a: SNPs Contributing to Cigarettes Per Day, Following FDR Correction

			AFRICAN ANCESTRY				EUROPEAN ANCESTRY										
SNP	CHR	BP	INFO	MAF	HWE	P	INFO	MAF	HWE	P	A1	A2	Weight	Zscore	P	Direction	q_1k
rs9653371	2	230003866	0.840	0.142	0.838	7.58E-01	0.506	0.016	1.000	2.00E-22	a	c	1709	-8.669	4.36E-18	--	3.0E-11
rs34731037	13	98575958	0.537	0.056	1.000	1.85E-27	0.668	0.070	0.516	1.68E-02	t	c	1709	7.357	1.89E-13	++	6.4E-07
rs71427733	2	205109090	0.957	0.058	1.000	4.64E-05	0.962	0.061	0.620	1.39E-05	t	c	1709	5.776	7.67E-09	++	1.1E-02
rs75714873	2	205109081	0.957	0.058	1.000	4.64E-05	0.962	0.061	0.620	1.39E-05	a	g	1709	-5.775	7.68E-09	--	1.1E-02
rs41319146	2	205108675	0.953	0.058	1.000	4.69E-05	0.964	0.060	0.620	1.43E-05	t	c	1709	-5.768	8.01E-09	--	1.1E-02
rs371955890	9	68695692	0.502	0.050	0.613	6.37E-37	0.601	0.113	0.266	5.18E-01	c	g	1709	-5.592	2.25E-08	-+	2.5E-02

rs12264038	10	23293025	0.865	0.185	0.505	4.80E-05	0.963	0.064	0.008	8.08E-05	t	c	1709	-5.419	5.99E-08	--	5.0E-02
rs71427734	2	205114662	1.000	0.057	1.000	1.46E-03	1.000	0.059	1.000	1.03E-05	a	g	1709	-5.401	6.61E-08	--	5.0E-02
rs16840927	2	205113668	0.992	0.057	1.000	1.48E-03	0.994	0.059	0.618	1.02E-05	a	g	1709	-5.400	6.66E-08	--	5.0E-02
rs2479719	6	41913081	0.930	0.106	0.786	3.62E-02	0.949	0.016	1.000	7.90E-07	c	g	1709	-5.334	9.60E-08	--	6.2E-02
rs185526451	2	7039120	0.907	0.081	1.000	1.76E-02	0.887	0.023	0.489	1.99E-06	t	g	1709	-5.309	1.10E-07	--	6.2E-02
15:28703003	15	28703003	0.525	0.055	1.000	2.18E-29	0.580	0.029	1.000	8.52E-01	a	g	1709	-5.295	1.19E-07	-+	6.2E-02
rs143847239	15	28703004	0.525	0.055	1.000	2.18E-29	0.580	0.029	1.000	8.52E-01	a	c	1709	-5.295	1.19E-07	-+	6.2E-02
rs7562403	2	213947476	0.891	0.210	0.880	4.59E-01	0.836	0.028	1.000	2.40E-08	t	c	1709	-5.239	1.62E-07	--	7.6E-02
rs16944923	15	91383766	0.542	0.055	1.000	1.76E-24	0.595	0.116	0.684	7.59E-01	c	g	1709	-5.221	1.78E-07	--	7.6E-02
rs10886927	10	123132045	0.985	0.186	0.506	4.83E-02	0.988	0.383	0.725	1.32E-06	t	c	1709	-5.187	2.14E-07	--	7.6E-02
rs10430703	10	123132206	0.981	0.186	0.506	4.88E-02	0.982	0.383	0.682	1.43E-06	t	c	1709	-5.172	2.32E-07	--	7.6E-02
rs11199933	10	123130356	0.999	0.189	0.623	4.51E-02	0.999	0.387	0.726	1.80E-06	c	g	1709	5.147	2.64E-07	++	7.6E-02
rs1896402	10	123130715	0.999	0.189	0.623	4.51E-02	0.999	0.387	0.726	1.80E-06	t	c	1709	-5.147	2.65E-07	--	7.6E-02
rs12220114	10	123130495	1.000	0.189	0.517	4.50E-02	1.000	0.387	0.727	1.81E-06	a	g	1709	-5.147	2.65E-07	--	7.6E-02
rs55823562	10	123130834	0.999	0.189	0.623	4.51E-02	0.998	0.387	0.726	1.81E-06	caata	c	1709	5.146	2.66E-07	++	7.6E-02
rs35059288	10	123131428	0.998	0.189	0.517	4.52E-02	0.998	0.387	0.726	1.81E-06	t	ta	1709	-5.146	2.67E-07	--	7.6E-02
rs10886926	10	123131594	0.996	0.189	0.623	4.52E-02	0.998	0.387	0.683	1.83E-06	t	c	1709	-5.144	2.69E-07	--	7.6E-02
rs10886925	10	123131571	0.996	0.189	0.623	4.52E-02	0.998	0.387	0.683	1.83E-06	a	g	1709	5.144	2.70E-07	++	7.6E-02
rs13128868	4	31329383	0.827	0.188	0.411	1.59E-04	0.937	0.392	0.072	1.72E-04	t	g	1709	5.117	3.10E-07	++	8.4E-02
rs10186370	2	1716434	0.935	0.256	0.357	3.47E-02	0.985	0.027	0.062	3.25E-06	a	g	1709	5.094	3.50E-07	++	9.1E-02
rs10713378	3	141829796	0.830	0.051	1.000	2.20E-03	0.954	0.151	0.161	4.70E-05	g	ga	1709	5.044	4.56E-07	++	1.1E-01
rs10430704	10	123132309	0.958	0.173	0.479	8.55E-02	0.972	0.378	0.480	2.48E-06	a	g	1709	4.952	7.33E-07	++	1.8E-01
rs113116955	6	41914195	0.949	0.060	0.145	6.62E-02	0.948	0.016	1.000	3.70E-06	t	c	1709	-4.938	7.90E-07	--	1.8E-01
rs5809760	14	76944781	0.820	0.304	0.639	3.52E-03	0.935	0.397	0.386	6.50E-05	ctt	c	1709	4.908	9.20E-07	++	2.1E-01
rs867188	19	51631165	1.000	0.114	0.456	2.06E-02	1.000	0.303	0.294	1.59E-05	a	c	1709	-4.898	9.69E-07	--	2.1E-01
rs12259839	10	23290952	0.835	0.113	0.134	1.58E-02	0.961	0.061	0.012	2.19E-05	t	c	1709	-4.882	1.05E-06	--	2.2E-01

Supplementary Table 9.8b: Genomic Bins SNPs Contributing to Cigarettes Per Day, Following FDR Correction

CHR	STARTBP	ENDBP	NSNP	P-VALUE	Q-VALUE	N	WITHIN GENES	NEARBY GENES (75KB)
2	230003866	230003866	1	4.36E-18	2.96E-11	1709	PID1	None
13	98575958	98575958	1	1.89E-13	6.40E-07	1709	None	IPO5
2	205108675	205114662	5	7.67E-09	0.010864774	1709	None	None
9	68695692	68695692	1	2.25E-08	0.025414477	1709	None	AK096159,LOC100132352
10	23290952	23293025	2	5.99E-08	0.050233272	1709	ARMC3	None
6	41913081	41914195	2	9.60E-08	0.062015813	1709	CCND3	BYSL, CCND3, MED20, USP49
2	7039120	7039120	1	1.10E-07	0.062015813	1709	None	CMPK2, RNF144A, RNF144A-AS1, RSAD2
15	28703003	28703004	2	1.19E-07	0.062015813	1709	None	DQ578199, DQ578700, DQ588687, DQ599733, GOLGA8F, GOLGA8G, JB175342, MIR4509-1
2	213947476	213947476	1	1.62E-07	0.076203844	1709	IKZF2	None
15	91383766	91383766	1	1.78E-07	0.076203844	1709	None	BLM, FES, FURIN, MAN2A2
10	123130356	123132309	10	2.14E-07	0.076203844	1709	None	None
4	31329383	31329383	1	3.10E-07	0.084203692	1709	None	None
2	1716434	1716434	1	3.50E-07	0.091274956	1709	PXDN	None
3	141829796	141829796	1	4.56E-07	0.114536686	1709	TFDP2	GK5, TFDP2
14	76944781	76944781	1	9.20E-07	0.208066183	1709	ESRRB	None
19	51631165	51631165	1	9.69E-07	0.212190467	1709	SIGLEC9	BC045766, CTU1, KLK13, KLK14, SIGLEC17P, SIGLEC7, SIGLEC9

CHAPTER 10:

POLYGENIC RISK SCORES FOR TOBACCO USE BEHAVIORS: ARE THEY PREDICTIVE WITHIN A UNIVERSITY SAMPLE?

Elizabeth K. Do, Jeanne E. Savage, Roseann E. Peterson, Bradley T. Webb, Danielle M. Dick, Kenneth S. Kendler, Hermine H. Maes

BACKGROUND

Previously conducted genome wide association studies identified several regions and candidate genes related to smoking behavior[1–3]. Recently, three large consortia (Oxford-GlaxoSmithKline, Tobacco and Genetics Consortium, and ENGAGE consortium) combined their summary statistics into a single meta-analysis for: smoking initiation, quantity, and cessation. This effort yielded a genome-wide significant association between the number of cigarettes per day and a cluster of nicotinic acetylcholine receptor genes on chromosome 15[4–6]. Additional genes contributing to tobacco use behaviors have been identified, such as the neuronal nicotinic acetylcholine receptor subunit beta-3 (*CHRNB3*) and alpha-6 (*CHRNA6)*[7], cytochrome P450, family 2, subfamily A, polypeptide 6 (*CYP2A6)*[6,7], and *LOC10018894*[5]. Each of these genes, which include genetic variants that contribute individually small effects, account for a very small proportion of the variance within smoking quantity (<2%).

Polygenic risk scores (PRS) are used[8] to summarize the genetic effects among a group of genetic variants that do not individually achieve genome-wide significance (p-value ≤ 5×10^{-8}) in large-scale genome wide association studies (GWAS). The effects of single nucleotide polymorphisms, or SNPs, are pooled together to represent a measured set of variants underlying a trait from GWAS summary statistics[9]. Traditionally, this is done by

abstracting the GWAS results from an initial discovery sample, ranking markers by their evidence of association (e.g. p-value), and then analyzing an independent target sample by constructing a PRS from the weighted sum of associated alleles using weights from the discovery sample within each subject. An association analysis is then conducted between a given trait and the constructed PRS. Where there is a statistically significant association (p-value ≤ 0.05) between the PRS and a given trait, the genetic effects found within the discovery sample are thought to contribute to the trait within the target sample[8]. Since the development of this method in 2009, which was first successfully implemented in a GWAS of schizophrenia[10], many other approaches have been developed to account for linkage disequilibrium[11] and improve resolution and determine the best-fit PRS[12].

PRS have the potential to be useful in increasing understanding and drawing inferences about genetic architectures both within and across many complex traits[11]. Though, within existing studies of complex traits, the variance explained by PRS seldom exceeds 2-3%. Additionally, while some studies report positive findings, others have been unable to find evidence for common genetic risk variation contributing to selected traits. Since the accuracy of the prediction score increases with the size of the discovery sample[9], it is possible that either: previously conducted studies finding null results did not have an adequate sample size within the discovery sample, there is a lack of genetic contribution to the phenotype of interest, or the genetic structure of the discovery and target samples are different[11]. This is an important point given the current lack of gene identification studies in populations of diverse ancestry, genetic architecture of tobacco use behaviors not being well described within ancestral groups outside of European ancestry, and

evidence that genetic determinants have important implications for addiction within many populations across the globe[13].

Despite the large epidemiological literature focused on social determinants of tobacco use and growing literature on the genetic epidemiology of tobacco use behaviors demonstrating the role that genes and the environment play in the development of tobacco-related phenotypes, few studies have examined how genetic variants interact with aspects of the environment to produce tobacco use behaviors. As described in a recent narrative review of the literature on genes, the environment, and their interaction on cigarette use, twin and family studies demonstrate that as individuals move from initiation to more established patterns of use, the importance of environmental factors decreases while the influence of genes increase. As this occurs, environmental factors begin to moderate the influence of genetic susceptibilities, implying gene-environment interaction. To date, these studies have included gene-environment interaction between either aggregate or individual genetic variants and their interaction with: religiosity, parental environment, traumatic events, and neighborhood factors[14].

To build upon this existing literature, we investigate two other aspects of parenting – autonomy granting and involvement – and the experience of stressful life events prior to university enrollment. In addition to being available for analyses in the current study, the parental environment and experience of stressful life events seem to be salient factors contributing to the progression and trajectory of tobacco use behaviors in young adulthood. Additionally, these analyses expand upon those conducted in previous chapters of this dissertation; namely, chapter 5 which describes a twin study of the

association between the experience of stressful life events and smoking initiation, and chapter 6 which investigates the prevalence and correlates of tobacco use behaviors.

To summarize, the current study seeks to: (1) generate polygenic risk scores from summary statistics from the TAG Consortium, (2) determine whether these polygenic risk scores are predictive of tobacco use behaviors within a college-aged sample, and (3) assess gene-by-environment interactions between polygenic risk scores, parental environment, and stressful life events prior to university enrollment, across individuals of European and African ancestry. We hypothesize that the polygenic risk scores will be predictive of tobacco use behaviors, and that associations will be higher among those in the European ancestry group, relative to the African ancestry group.

METHODS

Discovery Sample. Data from the Tobacco and Genetics (TAG) Consortium GWAS meta-analyses for smoking behavior, using genotype and smoking data from existing GWAS of other traits, was used as the discovery sample. This sample is comprised of 74,035 individuals from sixteen different studies conducted in the United States and Europe. Associations between approximately 2.5 million imputed markers and four smoking phenotypes were tested: ever versus never regular smokers, age at onset of smoking, cigarettes per day, and smoking cessation. For the purposes for the current analyses, one dimension of smoking behavior was included: ever versus never regular smokers. For ever versus never regular smokers, regular smokers were defined as those who reported having smoked ≥100 cigarettes during their lifetime and never regular smokers were defined as those who reported having smoked between 0 and 99 cigarettes during their lifetime (n=69,409). Each study conducted uniform cross-sectional analyses

using an additive genetic model for each tobacco use phenotype. Linear regression was used for quantitative traits, while logistic regression was used for dichotomous traits. Although original analyses were run separately for males and females, the TAG Consortium did not detect significant interactions by sex and data was analyzed together. Age was not included as a covariate in analyses conducted by the TAG Consortium, though case-control studies included case/control status as a covariate.

Target Sample. The target sample consisted of participants from Spit for Science, a longitudinal study of college students enrolled in a large, public, urban university in the Mid-Atlantic region, as described in previous studies[15]. This study collects population-based longitudinal data across five waves of survey data that were collected from 2011 to 2016. The phenotypes of interest within the target sample are as follows:

1. Ever Tobacco Use: lifetime measure of using any tobacco product
2. Age of Initiation: time at which tobacco product use started
3. Current Use: recent use measure of tobacco use within the past 30 days
4. Regular Use: endorses smoking at least 100 cigarettes in a lifetime
5. Cigarettes Per Day: maximum number of cigarettes smoked per day
6. Time to First Tobacco Use after Waking: amount of time taken between waking and using tobacco
7. Fagerström Test for Nicotine Dependence: standard measure of physical dependence on nicotine

Details regarding genotyping, pre-imputation quality control, imputation, population stratification, within ancestry group principal components analyses, and the inclusion of covariates can be found in the previous chapter (Chapter 9).

Polygenic risk scores. For the present study, polygenic risk scores for tobacco use behaviors were identified based on the large meta-analysis of the Tobacco and Genetics (TAG) Consortium. The risk alleles from the TAG Consortium measure of "ever vs. never regular use" were used as the discovery sample, to calculate polygenic risk scores in the Spit for Science sample. PRS were calculated using the methodology described by Purcell et al. (2009)[10].

This approach generates scores for individuals based on an allelic scoring system involving single nucleotide polymorphisms. The steps involved include performing quality control on both samples, filtering on call rate of ≥ 0.9, MAF > 0.01, and the removal of strand ambiguous SNPs (which occurs when you are unable to differentiate forward vs. backward strands, without information on allele frequency) and mismatched alleles. Prior to conducting quality control steps, 17,461,305 SNPs were contained within Spit for Science and 2,455,593 SNPs within the TAG Consortium data. Following quality control filtering and the removal of ambiguous SNPs, there were 7,653,789 and 2,455,593 SNPs retained in S4S and TAG, respectively. Across both samples, there were 1,802,970 common SNPs, after removing 1,924 mismatched SNPs. Then, LD clumping was performed on the remaining list of common SNPs, within the European and African ancestry groups. The --clump command considers LD when there are multiple significant association p-values within the same region. When performed, clumps are formed around central 'index variants' which must have p-values no larger than 0.0001 by default. For these analyses, the r^2 threshold was set to 0.1, the clump kb radius was set to 1000. Within the European ancestry group, 9,248 clumps were formed from 81,525 index variants and within the African ancestry group, 11,165 clumps were formed from 81,525

index variants. After pruning, 9,142 and 11,107 common variants were retained for the European and African ancestry groups. Scores were then created using the --score procedure with multiple p-value thresholds, and standardized to have a mean of zero and a standard deviation of one.

Statistical Analyses. Data analysis processes included: calculation of LD statistics, matching independent SNPs from the discovery and target samples, and calculation of ever regular use-based polygenic risks cores to conduct association analyses and predict tobacco use behaviors. Associations between each threshold were tested using regression analyses in R, adjusted by sex, age, and significant within ancestry group principal components from GWAS analyses conducted in the previous chapter. Where polygenic risk scores were predictive of tobacco use behaviors, regression analyses including the main effects of the PRS and the environment (e.g. parental involvement, parental autonomy granting, experience of stressful life events) and an interaction (e.g. PRS x environment), were tested to determine potential gene-by-environment interactions, using a significance threshold of p-value ≤ 0.05.

RESULTS

Table 9.1 (previous chapter) depicts the distribution of tobacco use variables for the S4S target sample, which only included individuals who provided genetic data passing quality control (n = 5,950). Information regarding the summary of environmental measures in S4S is shown in Table 10.1 on the next page. The information provided in this table is limited to post-quality control samples with genotypes, within individuals of European and African ancestries.

Table 10.1: Summary of Environmental Measures in S4S

	EUR		AFR	
Environmental Measures	**N / Total**	**%**	**N / Total**	**%**
Any Stressful Life Event	372 / 500	74.4	86 / 303	28.3
Accident	223 / 499	46.7	120 / 302	39.7
Physical Assault	125 / 500	25.0	65 / 301	21.6
Sexual Assault	36 / 498	7.2	18 / 299	6.0
Other Sexual Assault	89 / 496	17.9	41 / 298	13.8
Natural Disaster	281 / 500	56.2	170 / 302	56.3
	Mean	**Range**	**Mean**	**Range**
Parental Autonomy Granting (unstandardized)	6.49	1-12	6.62	1-12
Parental Involvement (unstandardized)	9.83	1-12	9.47	1-12

Polygenic risk scores (PRS) reflect a combined effect of selected SNPs, based upon five different p-value thresholds (p-values = $5x10^{-8}$, $5x10^{-6}$, $5x10^{-4}$, $5x10^{-2}$, and $5x10^{2}$). These selected SNPs were used to define large sets of risk alleles in the discovery sample, which were then used to generate a PRS for individuals in the independent target sample, as shown in Table 10.2. Within this table, we have indicated the total number of SNPs from the discovery (TAG Consortium) and target (S4S) samples, as well as the total number of common SNPs found across TAG and S4S, for the African (AFR) and European (EUR) ancestry groups found in S4S. For each of the p-value thresholds shown below, we also show the number of SNPs contributing to the estimated PRS. Generally, as the p-value threshold decreases, so does the number of SNPs contributing to the PRS.

Table 10.2: P-value Thresholds and Number of SNPs

	n SNPs Ever Use	
All SNPs TAG	2,457,119	
All SNPs S4S	17,461,305	
	AFR	**EUR**
Total Common SNPS: TAG and S4S	235,306	130,353
P-value threshold = $5x10^{2}$	66,282	120,080
P-value threshold = $5x10^{-2}$	7,212	13,051
P-value threshold = $5x10^{-4}$	105	194
P-value threshold = $5x10^{-6}$	37	9
P-value threshold = $5x10^{-8}$	28	6

Prior to running regression analyses, we ran correlations between each polygenic risk score set and the tobacco use behaviors measured in the target sample. Correlations between PRS and tobacco use behaviors were wide ranging for both the European ancestry group and for the African ancestry group, as shown in Table 10.3. Within the European ancestry group, negative correlations were found between each PRS with: ever tobacco use (range: -0.0224 to -0.0451) and age of initiation (range: -0.0255 to -0.0016). Alternatively, positive correlations were found between each PRS with: ever tobacco use (range: 0.0350 to 0.1014), age of initiation (range: 0.0397 to 0.0881), regular use (range: 0.0238 to 0.0685), and cigarettes per day (range: 0.1235 to 0.1494) within the African ancestry group. Where correlations are negative, it is suggested that polygenic risk scores are associated with lower levels of a given tobacco use behavior phenotype (e.g. polygenic risk scores were associated with never using tobacco products in a lifetime and older age of initiation within those of European ancestry). Meanwhile, positive correlations indicate that polygenic risk scores are associated with higher levels of tobacco use behavior phenotypes (e.g. polygenic risk scores were associated with having used tobacco products in a lifetime, younger age of initiation, having smoked 100 cigarettes in a lifetime, and the use of more cigarettes per day within those of African ancestry).

Table 10.3: Correlations between PLINK-based Polygenic Risk Scores and Tobacco Use

EUROPEAN ANCESTRY GROUP: TAG EVER VS. NEVER REGULAR USE PRS								
P-value Threshold	**EU**	**AO**	**CU**	**RU**	**RUS**	**CPD**	**TFT**	**FTND**
5x10^{2}	-0.0451	-0.0016	0.0343	0.0111	0.0241	-0.0698	0.0167	0.0072
5x10^{-2}	-0.0224	-0.0023	0.0331	-0.0095	0.0042	-0.0159	0.0289	0.0188
5x10^{-4}	-0.0528	-0.0255	-0.0112	-0.0265	-0.0037	0.0219	0.0444	0.0192
5x10^{-6}	-0.0374	-0.0102	-0.0133	-0.0224	-0.0154	0.0241	0.0607	0.0308
5x10^{-8}	-0.0388	-0.0141	-0.0227	-0.0247	-0.0188	0.0299	0.0677	0.0395
AFRICAN ANCESTRY GROUP: TAG EVER VS. NEVER REGULAR USE PRS								
P-value Threshold	**EU**	**AO**	**CU**	**RU**	**RUS**	**CPD**	**TFT**	**FTND**
5x10^{2}	0.1014	0.0881	0.0211	0.0685	-0.0111	0.1235	-0.0142	-0.0727
5x10^{-2}	0.0350	0.0397	-0.0145	0.0238	-0.0157	0.1327	-0.0316	-0.0453
5x10^{-4}	0.0464	0.0576	0.0131	0.0422	0.0229	0.1427	-0.0131	0.0031
5x10^{-6}	0.0408	0.0437	0.0330	0.0287	0.0215	0.1487	0.0035	-0.0001
5x10^{-8}	0.0420	0.0452	0.0427	0.0303	0.0270	0.1494	0.0093	0.0031

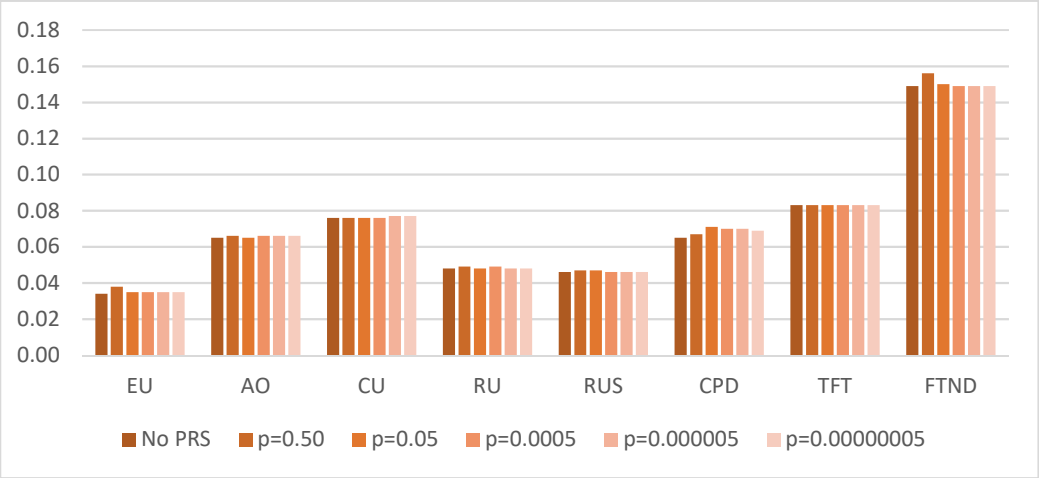
Note: EU = ever use; AO = age of onset; CU = current use; RU = regular use; RUS = regular use among smokers; CPD = cigarettes per day; TFT = time to first tobacco use; FTND = Fagerström Test for Nicotine Dependence

We used five sets of risk scores with p-value thresholds ranging from $5x10^{-8}$ to $5x10^{2}$ to predict ever tobacco use (EU), age of initiation (AO), current use (CU), regular use (RU), regular use among smokers (RUS), cigarettes per day (CPD), time to first tobacco use after waking (TFT), and the Fagerström Test for Nicotine Dependence (FTND) using either logistic regression, where the outcome was binary, or linear regression, where the outcome of interest was continuous.

Polygenic risks scores only significantly predicted: ever tobacco use (using the threshold p-value $\leq 5 \times 10^{-4}$) and time to first tobacco use after waking (using the threshold p-value $\leq 5 \times 10^{-8}$) within the European ancestry group (estimates of the effects, z-value, and model fit statistics are shown in Supplemental Tables 10.4a-b; information also provided in Figure 10.1a-b). Within these models, PRS using the threshold of p-value $\leq 5 \times 10^{-4}$ accounted for 8.0% of the variance of ever tobacco use and the PRS using the threshold

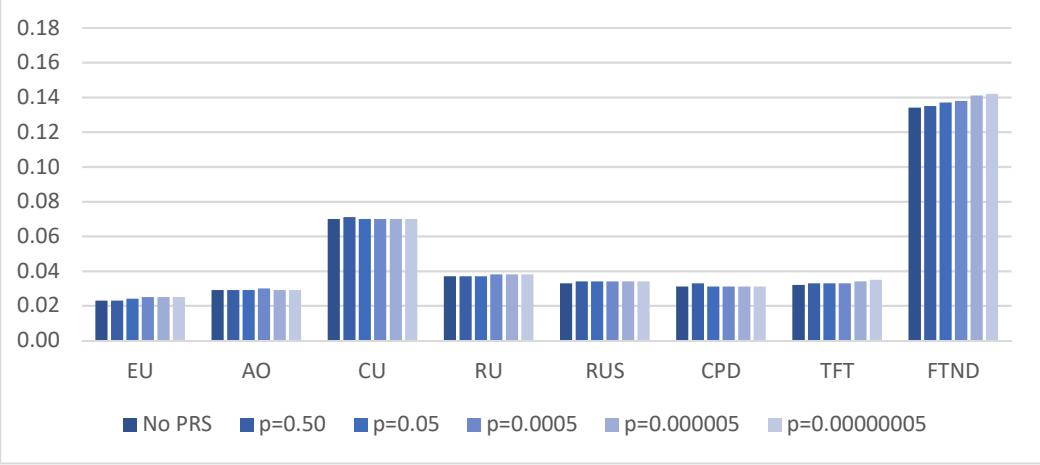
of p-value ≤ 5 x 10^{-8} accounted for 8.5% of the variance of time to first tobacco use within the European ancestry group.

Figure 10.1a. PRS using Ever vs. Never Regular Tobacco Use in the Tobacco and Genetics Consortium Predicting Tobacco Use Phenotypes in Spit for Science (AFR)



Note: EU = ever use; AO = age of onset; CU = current use; RU = regular use; RUS = regular use among smokers; CPD = cigarettes per day; TFT = time to first tobacco use; FTND = Fagerström Test for Nicotine Dependence

Figure 10.1b. PRS using Ever vs. Never Regular Tobacco Use in the Tobacco and Genetics Consortium Predicting Tobacco Use Phenotypes in Spit for Science (EUR)



Note: EU = ever use; AO = age of onset; CU = current use; RU = regular use; RUS = regular use among smokers; CPD = cigarettes per day; TFT = time to first tobacco use; FTND = Fagerström Test for Nicotine Dependence

Main Effects of Environment. Since PRS were only significant for ever tobacco use and time to first tobacco use after waking within those of European ancestry, we went on to

test whether measures of the parental environment and the experience of stressful life events prior to university enrollment were significant in predicting those two variables within those of European ancestry (see Supplementary Tables 10.5a-b). Parental autonomy granting, parental involvement, as well as experiencing an accident, physical assault, or other sexual assault prior to university enrollment were predictive of ever tobacco use. Meanwhile, parental involvement and the experience of physical assault, sexual assault, or other sexual assault prior to enrollment at university predicted time to first tobacco use after waking. These variables were used for subsequent analyses to test PRS by environment interactions, using p-value thresholds that explain the most variance in ever tobacco use (p-value ≤ 5×10^{-4}) and time to first tobacco use after waking (p-value ≤ 5×10^{-8}). Data for the main effects of the environment (and subsequent PRS x environment interactions) on tobacco use among the African ancestry group are not shown, since PRS did not significantly predict tobacco use behaviors within this sample, and because we were interested in testing PRS x environment interactions where main effects of both PRS and the environmental measures are significant.

Interaction Effects of the Environment. Main and interaction effects of PRS and measures of the parental environment on ever tobacco use and time to first tobacco use are shown in Supplementary Tables 10.5a-b, while the main and interaction effects of stressful life events experienced prior to university enrollment and ever tobacco use and time to first tobacco use are shown in Supplemental Tables 10.6a-10.3b. Only the interaction between effect of polygenic risk score (p-value = 5×10^{-4}) and environment was significant for the experience of physical assault, predicting ever tobacco use within the European ancestry group (Supplementary Table 10.6a).

DISCUSSION

Tobacco use behaviors are complex traits that are highly polygenic in nature. Many genetic variants, each of small effect (e.g. $R^2 < 0.005$) contributing to the development of each specific phenotype. Unfortunately, what this means is that it is unlikely that genome wide association studies will lead to straightforward results to be replicated in independent samples[16] at current sample sizes. Results from the TAG Consortium meta-analysis demonstrates these difficulties, as even with large sample sizes (greater than 70,000 individuals) no genome wide significant results were obtained for either smoking initiation or age at smoking initiation. Modeling the additive or cumulative effects of associated variants works to get around this problem and has the potential to explain a higher proportion of variation relative to any single genetic variant[17]. Thus, in efforts to investigate the cumulative effects of associated variants contributing to tobacco use behaviors within our sample, we calculated polygenic risk scores derived from the TAG Consortium meta-analyses of GWAS results for ever vs. never regular use and conducted association analyses to determine whether we could predict: ever tobacco use, age of initiation, current use, regular use among smokers, time to first tobacco use, cigarettes per day, and the Fagerström Test for Nicotine Dependence.

The standard PRS approach requires testing over a range of p-value thresholds[10], which are often chosen arbitrarily. Potential limitations of this approach include the possibility of including variants that could be false positives and raw estimates of effect sizes being subject to selection bias. One alternative to this approach is to consider linkage disequilibrium (LD) among markers, using a reference LD panel. However, potential limitations of the approach using LDpred includes the method's reliance on LD information

from a reference panel and potential heterogeneity hindering prediction accuracy[11]. This approach might not be appropriate for the sample described here, since LDpred performs best with an LD reference panel of at least 1,000 individuals. Additionally, its flagship paper applied LDpred to GWAS summary statistics for large sample sizes ranging from 27,000 to 86,000 individuals. Our LD reference panel, available from the TAG Consortium data, includes a sample of 69,409 and meets these requirements. However, it is possible that our power to detect effects is still limited, since our target sample included less than 10,000 individuals – or more specifically, 3,018 of European ancestry and 1,339 of African ancestry for these genetic analyses. Given that our target sample size was small, the relatively small number of individuals may have limited power to detect effects across all p-value thresholds.

LDpred might also not perform well with our sample, because of the requirement that the target sample needs to have LD patterns like the discovery set. The discovery sample from the TAG Consortium is primarily made up of older adults of European ancestry. Although one of the subsets of our target sample was comprised of individuals of European ancestry, it is possible that our sample has higher levels of admixture which could also result in differences in allele frequencies across the discovery and target samples. Additionally, our sample is younger relative to the discovery set, which is important if some genetic variants do not play a large role until later in life, as demonstrated in previously published twin studies[18].

Other potential limitations exist in the interpretation of findings from the current study and those employing polygenic risk score methodologies more broadly. Firstly, adequate sample size is necessary for score estimates within discovery populations to be precise

and optimal p-value thresholds used for selecting score variants, depends on the size of the discovery sample. Generally, the variance explained is dependent upon the sample size of the discovery set, such that when the sample size is large, the effects detected within the GWAS contains less noise, which can lead to more accurate predictions in the target sample[19]. Disparate patterns of LD and differences in marker allele frequencies between discovery and target samples are also thought to attenuate effects of PRS analyses[17]. Additionally, population stratification and differential patterns of LD across racial and ethnic groups, like those found in the current sample, may bias results within genetic association studies[20]. Since GWAS have traditionally been conducted on individuals of European descent, and methods for computing polygenic risk scores are dependent upon GWAS, risk alleles identified from these studies may be specific to the ancestral group or include tag SNPs not found in other populations[21].

Along those same lines, reported values of R^2 might not directly reflect the degree of missing heritability, but could perhaps reflect the effect of sampling variation on the variance, as explained by estimated scores. Since the effects of individual SNPs are very small, they are estimated with a great deal of error. Thus, the prediction of a phenotype using estimated SNP effects may suffer from sampling variance with which the effect is estimated. Added to this, the measures of tobacco use are crude within the discovery sample, and the worse the estimate of the effect size of the variant in the discovery sample, the worse the variance will be explained by the predictor in the target/validation sample. Nevertheless, the TAG meta-analysis is the largest GWAS meta-analysis for tobacco use behavior that exists. The chances of success of polygenic risk score analyses are dependent upon the size of the discovery set, so if the sample size is too

small (which is still potentially the case with the TAG Consortium data), the risk profiles could be based upon random noise and are not expected to explain variance in the target set.

However, what was encouraging about these analyses, was that they provided further evidence that different measures of tobacco use behaviors were influenced by overlapping genetic factors. Ever tobacco use (EUR: $h^2 = 0.28$, p-value = 0.0133), age of initiation (EUR: $h^2 = 0.30$, p-value = 0.0288), and regular use among smokers (EUR: $h^2 = 0.18$, p-value = 0.0041) yielded significant SNP-based heritability estimates (at least, prior to correction for multiple testing) using GCTA within individuals of European ancestry, as demonstrated in Chapter 9. PRS derived from TAG Consortium data on ever vs. never regular use significantly predicted both ever tobacco use and time to first tobacco use within individuals of European ancestry, suggesting that at least some of the SNPs influencing ever tobacco use were the same as those influencing the time to first tobacco use after waking. And as expected, the amount of variance explained by PRS were lower than previously reported heritability estimates from twin studies.

Adding parental environmental variables (parental autonomy granting and parental involvement) and their interactions with polygenic risk scores explained more of the variance within the tobacco use behaviors of ever tobacco use and time to first tobacco use after waking within those of European ancestry, relative to including polygenic risk scores and covariates alone (see Supplementary Tables 10.5a and 10.5b). This suggests that the environments provided by parents, as well as the interaction between the parental environment and PRS, were significant contributors to the variance in ever tobacco use and time to first tobacco use within this population. Although we also could detect

significant main effects of stressful life events experienced prior to university enrollment on these two measures of tobacco use, none of the interactions between stressful life events and PRS were significant.

PRS did not significantly predict any of the measures of tobacco use behaviors within individuals of African ancestry within this study. The lack of predictive power for the PRS among individuals of African ancestry might be the result of differences in patterns of LD between African and European-ancestry individuals, which has important implications for future studies that intend to use cumulative measures of risk to predict a given phenotype, especially when many large-scale GWAS derive their discovery datasets from individuals of European ancestry. Future studies may require newer methodologies that are able to incorporate more accurate measures of LD from the target population, particularly within ancestry groups that are ad-mixed[16].

These results must be interpreted in the context of certain limitations. This study does provide evidence for a genetic architecture of multiple common variants with small individual effect sizes (as evident from 6 SNPs contributing to PRS for time to first tobacco use after waking and 194 SNPs contributing to PRS for ever tobacco use, among individuals of the European ancestry group) influencing tobacco use behaviors. Additionally, this study explored the contribution of genetic risk across a range of tobacco use behaviors across both European and African ancestral groups. Using an aggregate measure of genetic risk by way of polygenic risk scores, we did not need to require the rigorous statistical test corrections that are required for genome wide association studies. We were also able to examine specific environmental risk factors, focused on risk prior to enrollment in university, and how genetic influences might change as a function of

environmental risk. This information is potentially useful for public health interventions, since it allows for the identification for potentially modifiable risk factors[13]. Also, polygenic risk scores were based on a large discovery sample – the largest meta-analyses to date on tobacco use behaviors – which should add to the accuracy of the polygenic risk scores. However, to be sure that these polygenic risk scores are accurate, more studies need to be conducted across a variety of different ancestral groups with larger sample sizes - especially since it is important to validate the predictive ability of polygenic risk scores for tobacco use phenotypes, while also gaining a better understanding of the genetic contributions to these behaviors.

CONCLUSION

In conclusion, we investigated the genetic architecture of tobacco use behaviors within a sample of young adults attending university. We constructed polygenic risk scores that predicted ever tobacco use and time to first tobacco use after waking, within individuals of European ancestry. As hypothesized, PRS did significantly predict tobacco use behaviors, and PRS explained more of the variance in tobacco use behaviors within individuals of European ancestry relative to African ancestry. However, PRS only predicted ever tobacco use and time to first tobacco use after waking within individuals of European ancestry. Limited predictive power of the PRS may be attributed to three key issues: differences in phenotype definitions, potential differences in marker allele frequencies between the discovery and target samples, and issues related to sample size (potentially in both the discovery and target samples) and power.

Despite these potential limitations, significant additive interactions were observed between polygenic risk scores and aspects of individual environments prior to university

enrollment – particularly, parental autonomy granting, parental involvement, and the experience of an accident, physical, sexual, or other sexual assault prior to university enrollment. Thus, this study provides support for further study of polygenic risk scores and gene-environment interactions. This study also suggests that the time prior to university enrollment may be useful for prevention and intervention strategies. Specifically, the findings of this study suggest that interventions should look towards parental environment and the experience of stressful life events prior to university enrollment to diminishing genetic vulnerability to tobacco use behaviors, such as ever use and time to first tobacco use after waking.

Supplementary Table 10.4a: PLINK-based PRS Predicting Measures of Tobacco Use Behaviors in S4S AFR Ancestry Group

Ever Tobacco Use (AFR)	***Base Model*** Estimate (z-value)		***p-value = 5 x 10^{2}*** Estimate (z-value)		***p-value = 5 x 10^{-2}*** Estimate (z-value)		***p-value = 5 x 10^{-4}*** Estimate (z-value)		***p-value = 5 x 10^{-6}*** Estimate (z-value)		***p-value = 5 x 10^{-8}*** Estimate (z-value)	
PGRS			0.11 (1.71)		0.02 (0.34)		0.05 (0.84)		0.04 (0.75)		0.04 (0.75)	
Sex	-0.41 (-3.24)	**	-0.42 (-3.26)	.	-0.41 (-3.24)	**	-0.41 (-3.23)	**	-0.41 (-3.24)	**	-0.41 (-3.24)	**
Age	0.12 (2.55)	*	0.12 (2.53)	*	0.12 (2.54)	*	0.12 (2.53)	*	0.12 (2.53)	*	0.12 (2.53)	*
PC1	7.71 (3.62)	***	5.94 (2.52)	*	7.50 (3.39)	**	7.37 (3.40)	**	7.42 (3.43)	**	7.40 (3.41)	**
PC3	-3.54 (-1.67)	.	-3.52 (-1.66)	.	-3.54 (-1.67)	.	-3.59 (-1.69)	.	-3.58 (-1.69)	.	. -3.58 (-1.69)	
PC8	3.06(1.48)		3.16 (1.52)		3.08 (1.49)		3.16 (1.52)		3.16 (1.53)		3.16 (1.52)	
Intercept	-1.44 (-1.51)		-1.41 (-1.48)		-1.42 (-1.49)		-1.42 (-1.49)		-1.42 (-1.49)		-1.42 (-1.49)	
Nagelkerke R^{2}	0.034		0.038		0.035		0.035		0.035		0.035	
Change in R^{2}			0.004		0.001		0.001		0.001		0.001	
Chi-Square: Deviance, p			2.9304 (0.08)	.	0.1132 (0.73)		0.7017 (0.40)		0.5572 (0.45)		0.5614 (0.45)	
N	1329		1329		1329		1329		1329		1329	

Age of Initiation (AFR)	***Base Model*** Estimate (z-value)		***p-value = 5 x 10^{2}*** Estimate (z-value)		***p-value = 5 x 10^{-2}*** Estimate (z-value)		***p-value = 5 x 10^{-4}*** Estimate (z-value)		***p-value = 5 x 10^{-6}*** Estimate (z-value)		***p-value = 5 x 10^{-8}*** Estimate (z-value)	
PGRS			0.03 (0.63)		0.00 (0.03)		0.02 (0.66)		0.01 (0.42)		0.02 (0.44)	
Sex	-0.31 (-4.00)	***	-0.31 (-4.00)	***	-0.31 (-4.00)	***	-0.31 (-3.99)	***	-0.31 (-4.00)	***	-0.31 (-4.00)	***
Age	0.08 (2.59)	**	0.08 (2.53)	*	0.08 (2.59)	**	0.08 (2.55)	*	0.08 (2.57)	*	0.08 (2.56)	*
PC1	4.61 (3.62)	***	4.18 (2.88)	**	4.60 (3.43)	***	4.43 (3.39)	***	4.51 (3.46)	***	4.49 (3.44)	***
PC2	1.83 (1.44)	.	1.60 (1.21)		1.83 (1.41)		1.81 (1.42)		1.83 (1.44)		1.82 (1.43)	
PC3	-1.91 (-1.47)		-1.91 (-1.47)		-1.91 (-1.47)		-1.91 (-1.48)		-1.92 (-1.48)		-1.92 (-1.48)	
Intercept	-0.12 (-0.21)		-0.09 (-0.15)		-0.12 (-0.20)		-0.10 (-0.17)		-0.10 (-0.18)		-0.11 (-0.18)	
Nagelkerke R^{2}	0.065		0.066		0.065		0.066		0.066		0.066	
Change in R^{2}			0.001		0.000		0.001		0.001		0.001	
Chi-Square: Deviance, p			0.5127 (0.53)	.	0.0012 (0.97)	.	0.5597 (0.51)	.	0.2206 (0.68)	.	0.2566 (0.65)	.
N	1039		1039		1039		1039		1039		1039	

Current Use (AFR)	***Base Model*** *Estimate (z-value)*		***p-value = 5 x 10^{2}*** Estimate (z-value)		***p-value = 5 x 10^{-2}*** Estimate (z-value)		***p-value = 5 x 10^{-4}*** Estimate (z-value)		***p-value = 5 x 10^{-6}*** Estimate (z-value)		***p-value = 5 x 10^{-8}*** Estimate (z-value)	
PGRS			0.003 (0.04)		-0.01 (-0.17)		0.03 (0.39)		0.06 (0.70)		0.07 (0.84)	
Sex	-0.49 (-2.79)	**	-0.49 (-2.79)	**	-0.49 (-2.79)	**	-0.49 (-2.77)	**	-0.49 (-2.77)	**	-0.49 (-2.76)	**
Age	-0.30 (-4.61)	***	-0.30 (-4.61)	***	-0.30 (-4.61)	***	-0.30 (-4.62)	***	-0.30 (-4.63)	***	-0.30 (-4.63)	***
PC1	4.57 (2.93)		4.51 (1.36)		4.72 (1.53)		4.31 (1.43)		4.13 (1.38)		4.00 (1.33)	
PC2	-5.26 (-1.77)	.	-5.29 (-1.72)	.	-5.19 (-1.72)	.	-5.25 (-1.77)	.	-5.19 (-1.75)	.	-5.17 (-1.74)	.
PC7	-5.74 (-2.87)	*	-5.74 (-2.00)	*	-5.72 (-2.00)	*	-5.76 (-2.00)	*	-5.76 (2.01)	*	-5.76 (-2.01)	*
Intercept	7.38 (5.54)	***	7.38 (5.54)	***	7.38 (5.53)	***	7.38 (5.54)	***	7.41 (5.55)	***	7.42 (5.55)	***
Nagelkerke R^2	0.076		0.076		0.076		0.076		0.077		0.077	
Change in R^2			0.000		0.000		0.000		0.001		0.001	
Chi-Square: Deviance, p			0.0016 (0.97)		0.0275 (0.86)		0.1517 (0.70)		0.4865 (0.49)		0.7121 (0.40)	
N	742		742		742		742		742		742	

Regular Use (AFR)	***Base Model*** Estimate (z-value)		***p-value = 5 x 10^{2}*** Estimate (z-value)		***p-value = 5 x 10^{-2}*** Estimate (z-value)		***p-value = 5 x 10^{-4}*** Estimate (z-value)		***p-value = 5 x 10^{-6}*** Estimate (z-value)		***p-value = 5 x 10^{-8}*** Estimate (z-value)	
PGRS			0.02 (0.81)		0.00 (0.00)		0.01 (0.60)		0.01 (0.34)		0.01 (0.38)	
Sex	-0.20 (-5.50)	***	-0.20 (-5.51)	***	-0.20 (-5.50)	***	-0.20 (-5.50)	***	-0.20 (-5.50)	***	-0.20 (-5.50)	***
Age	0.04 (3.22)	**	0.04 (3.18)	**	0.04 (3.20)	**	0.04 (3.20)	**	0.04 (3.21)	**	0.04 (3.21)	**
PC1	2.27 (3.82)	***	2.02 (3.03)	**	2.27 (3.66)	***	2.20 (3.63)	***	2.23 (3.69)	***	2.22 (3.67)	***
PC2	-0.38 (-0.65)	.	-0.51 (-0.84)	.	-0.38 (-0.64)		-0.39 (-0.66)		-0.38 (-0.65)		-0.38 (-0.65)	
Intercept	-0.09 (-0.32)		-0.08 (0.29)		0.10 (0.18)		-0.08 (-0.30)		-0.08 (-0.31)		-0.08 (-0.31)	
Nagelkerke R^2	0.048		0.049		0.048		0.049		0.048		0.048	
Change in R^2			0.001		0.000		0.001		0.00		0.00	
Chi-Square: Deviance, p			0.2228 (0.42)		0.0000 (0.99)		0.1226 (0.55)		0.0405 (0.73)		0.0483 (0.71)	
N	1329		1329		1329		1329		1329		1329	

Regular Use Among Smokers (AFR)	***Base Model*** Estimate (z-value)		***p-value = 5 x 10^{2}*** Estimate (z-value)		***p-value = 5 x 10^{-2}*** Estimate (z-value)		***p-value = 5 x 10^{-4}*** Estimate (z-value)		***p-value = 5 x 10^{-6}*** Estimate (z-value)		***p-value = 5 x 10^{-8}*** Estimate (z-value)	
PGRS			-0.11 (-0.82)		-0.06 (-0.50)		0.01 (0.05)		0.004 (0.04)		0.02 (0.13)	
Sex	-0.81 (-3.27)	**	-0.81 (-3.26)	**	-0.81 (-3.26)	**	-0.81 (-3.27)	**	-0.81 (-3.27)	**	-0.81 (-3.27)	**
Age	0.12 (1.34)		0.12 (1.36)		0.12 (1.37)		0.12 (1.33)		0.12 (1.33)		0.12 (1.32)	
PC1	6.06 (1.50)		7.84 (1.71)		6.63 (1.58)		6.01 (1.45)		6.03 (1.45)		5.94 (1.42)	
PC2	-3.71 (-1.04)		-2.76 (-0.74)		-3.36 (-0.92)		-3.71 (-1.04)		-3.71 (-1.04)		-3.70 (-1.04)	
Intercept	0.83 (1.37)	.	-3.38 (-1.82)	.	-3.39 (-1.82)	.	-3.33 (-1.79)	.	-3.33 (-1.79)	.	-3.32 (-1.79)	.
Nagelkerke R^{2}	0.046		0.047		0.047		0.046		0.046		0.046	
Change in R^{2}			0.001		0.001		0.001		0.001		0.001	
Chi-Square: Deviance, p			0.6775 (0.41)		0.2502 (0.62)		0.0029 (0.96)		0.0014 (0.97)		0.0163 (0.89)	
N	744		744		744		744		744		744	

CPD (AFR)	***Base Model*** Estimate (z-value)		***p-value = 5 x 10^{2}*** Estimate (z-value)		***p-value = 5 x 10^{-2}*** Estimate (z-value)		***p-value = 5 x 10^{-4}*** Estimate (z-value)		***p-value = 5 x 10^{-6}*** Estimate (z-value)		***p-value = 5 x 10^{-8}*** Estimate (z-value)	
PGRS			0.02 (0.81)		0.04 (1.55)		0.03 (1.37)		0.03 (1.27)		0.03 (1.23)	
Sex	-0.14 (-2.70)	**	-0.14 (-2.73)	**	-0.15 (-2.78)	**	-0.14 (-2.71)	**	-0.14 (-2.73)	**	-0.14 (-2.73)	**
Age	0.00 (0.01)		-0.00 (-0.02)		-0.001 (-0.09)		-0.00 (-0.03)		-0.00 (-0.04)		-0.00 (0.03)	
PC1	1.64 (2.05)	*	1.27 (1.37)		1.26 (1.51)		1.35 (1.64)		1.39 (1.69)	.	1.38 (1.66)	*
PC2	-1.87 (-2.29)	.*	-1.87 (-2.29)	.*	-2.06 (-2.49)	.*	-1.87 (-2.29)	.*	-1.84 (-2.25)	.*	-1.84 (-2.25)	.*
PC7	-2.20 (-2.48)	*	-2.03 (-2.41)	*	-2.30 (-2.59)	*	-2.28 (-2.57)	*	-2.26 (-2.55)	*	-2.26 (-2.54)	*
Intercept	1.38 (3.58)	***	-2.24 (-2.52)	***	1.42 (3.68)	***	1.38 (3.61)	***	1.39 (3.63)	***	1.39 (3.62)	***
Nagelkerke R^{2}	0.065		0.067		0.071		0.070		0.070		0.069	
Change in R^{2}			0.001		0.005		0.004		0.004		0.003	
Chi-Square: Deviance, p			0.1440 (0.42)		0.5242 (0.12)		0.4102 (0.17)		0.3539 (0.20)		0.3340 (0.22)	
N	369		369		369		369		369		369	

TFT (AFR)	***Base Model*** Estimate (z-value)		***p-value = 5 x 10^{2}*** Estimate (z-value)		***p-value = 5 x 10^{-2}*** Estimate (z-value)		***p-value = 5 x 10^{-4}*** Estimate (z-value)		***p-value = 5 x 10^{-6}*** Estimate (z-value)		***p-value = 5 x 10^{-8}*** Estimate (z-value)	
PGRS			-0.01 (-0.24)		-0.004 (-0.12)		-0.004 (-0.13)		-0.002 (-0.07)		-0.0002 (-0.01)	
Sex	-0.29 (-3.72)	***	-0.29 (-3.69)	***	-0.29 (-3.70)	***	-0.29 (-3.71)	***	-0.29 (-3.71)	***	-0.29 (-3.71)	**
Age	-0.00 (-0.01)		-0.00 (-0.00)		-0.00 (-0.01)		-0.00 (-0.01)		-0.00 (-0.01)		-0.00 (-0.01)	***
PC1	0.68 (0.56)		0.85 (0.61)		0.73 (0.57)		0.72 (0.58)		0.70 (0.56)		0.68 (0.54)	
PC2	-2.53 (-2.18)	*	-2.46 (-2.03)	*	-2.51 (-2.12)	*	-2.54 (-2.18)	*	-2.54 (-2.18)	*	-2.54 (-2.17)	.
PC3	2.24 (1.69)	.	2.26 (1.70)	.	2.25 (1.70)	.	2.24 (1.70)	.	2.25 (1.69)	.	2.24 (1.69)	*
PC5	-2.07 (-1.47)		-2.05 (-1.46)		-2.05 (-1.45)		-2.05 (-1.45)		-2.06 (-1.46)		-2.07 (-1.46)	
PC7	-2.54 (-1.93)	.	-2.52 (-1.91)	.	-2.53 (-1.92)	.	-2.53 (-1.92)	.	-2.54 (-1.93)	.	-2.54 (-1.93)	
Intercept	0.83 (1.37)		0.82 (1.35)		0.82 (1.35)		0.83 (1.36)		0.83 (1.37)		0.83 (1.36)	
Nagelkerke R^2	0.083		0.083		0.083		0.083		0.083		0.083	
Change in R^2			0.000		0.000		0.000		0.000		0.000	
Chi-Square: Deviance, p			0.0332 (0.81)		0.0086 (0.90)		0.010 (0.89)		0.0024 (0.95)		0.0001 (0.99)	
N	427		427		427		427		427		742	

FTND (AFR)	***Base Model*** Estimate (z-value)		***p-value = 5 x 10^{2}*** Estimate (z-value)		***p-value = 5 x 10^{-2}*** Estimate (z-value)		***p-value = 5 x 10^{-4}*** Estimate (z-value)		***p-value = 5 x 10^{-6}*** Estimate (z-value)		***p-value = 5 x 10^{-8}*** Estimate (z-value)	
PGRS			-0.10 (-1.02)		-0.03 (-0.33)		0.0008 (0.01)		-0.01 (-0.10)		-0.005 (-0.06)	
Sex	0.004 (0.02)		0.01 (0.07)		0.01 (0.04)		0.004 (0.02)		0.005 (0.03)		0.005 (0.02)	
Age	0.19 (2.71)	**	0.19 (2.70)	**	0.19 (2.71)	**	0.19 (2.70)	**	0.19 (2.71)	**	0.19 (2.71)	**
PC1	0.40 (0.13)		2.02 (0.58)		0.70 (0.22)		0.39 (0.12)		0.47 (0.15)		0.44 (0.14)	
PC2	-6.78 (-2.24)	.*	-5.93 (-1.89)	.	-6.59 (-2.15)	.*	-6.78 (-2.24)	.*	-6.78 (-2.24)	.*	-6.78 (-2.24)	.*
PC3	5.17 (1.63)		5.29 (1.67)	.	5.23 (1.65)		5.17 (1.63)		5.19 (1.63)		5.19 (1.63)	
PC7	-4.80 (-1.51)		-4.64 (-1.45)		-4.76 (-1.49)		-4.81 (-1.50)		-4.79 (-1.50)		-4.80 (-1.50)	
Intercept	-2.67 (-1.81)	.	-2.66 (-1.81)		-2.68 (-1.82)	.	-2.67 (-1.81)	.	-2.67 (-1.81)	.	-2.67 (-1.81)	.
Nagelkerke R^2	0.149		0.156		0.150		0.149		0.149		0.149	
Change in R^2			0.007		0.001		0.000		0.000		0.000	
Chi-Square: Deviance, p			4.6852 (0.31)		0.4919 (0.74)		0.0004 (0.99)		0.0458 (0.92)		0.0140 (0.96)	
N	527		527		527		527		527		527	

Supplementary Table 10.4b: PLINK-based PRS Predicting Measures of Tobacco Use Behaviors in S4S EUR Ancestry Group

Ever Tobacco Use (EUR)	***Base Model*** Estimate (z-value)		***p-value = 5 x 10^{2}*** Estimate (z-value)		***p-value = 5 x 10^{-2}*** Estimate (z-value)		***p-value = 5 x 10^{-4}*** Estimate (z-value)		***p-value = 5 x 10^{-6}*** Estimate (z-value)		***p-value = 5 x 10^{-8}*** Estimate (z-value)	
PGRS			0.0002 (0.01)		-0.04 (-0.91)		-0.09 (-2.31)	*	-0.07 (-1.70)	.	-0.06 (-1.60)	
Sex	-0.47 (-5.71)	***	-0.47 (-5.71)	***	-0.47 (-5.71)	***	-0.47 (-5.70)	***	-0.47 (-5.71)	***	-0.47 (-5.70)	***
Age	0.12 (3.50)	***	0.12 (3.50)	***	0.12 (3.49)	***	0.12 (3.59)	***	0.12 (3.54)	***	0.12 (3.54)	***
PC1	-0.94 (-0.42)		-0.94 (-0.41)		-0.64 (-0.28)		-0.71 (-0.32)		-0.90 (-0.40)		-0.86 (-0.39)	
PC2	-2.17 (-0.97)		-2.17 (-0.97)		-2.14 (-0.95)		-2.00 (-0.89)		-2.06 (-0.92)		-2.06 (-0.92)	
PC5	-4.57 (-2.01)	*	-4.58 (-2.01)	*	-4.58 (-2.01)	*	-4.45 (-1.94)	*	-4.48 (-1.96)	*	-4.48 (-1.96)	*
Intercept	-0.77 (-1.14)		-0.77 (-1.14)		-0.77 (-1.13)		-0.84 (-1.23)		-0.80 (-1.18)		-0.80 (-1.18)	
Nagelkerke R^2	0.023		0.023		0.024		0.025		0.025		0.025	
Change in R^2			0.000		0.001		0.002		0.002		0.002	
Chi-Square: Deviance, p			0.00002 (0.99)		0.8213 (0.36)		5.3679 (0.02)	*	2.8822 (0.09)	.	2.5536 (0.11)	
N	2980		2980		2980		2980		2980		2980	

Age of Initiation (EUR)	***Base Model*** Estimate (z-value)		***p-value = 5 x 10^{2}*** Estimate (z-value)		***p-value = 5 x 10^{-2}*** Estimate (z-value)		***p-value = 5 x 10^{-4}*** Estimate (z-value)		***p-value = 5 x 10^{-6}*** Estimate (z-value)		***p-value = 5 x 10^{-8}*** Estimate (z-value)	
PGRS			0.002 (0.08)		-0.002 (-0.08)		-0.02 (-1.22)		-0.01 (-0.51)		-0.01 (-0.61)	
Sex	-0.24 (-5.14)	***	-0.24 (-5.14)	***	-0.24 (-5.14)	***	-0.24 (-5.12)	***	-0.24 (-5.13)	***	-0.24 (-5.13)	***
Age	0.06 (3.47)	***	0.06 (3.47)	***	0.06 (3.47)	***	0.07 (3.51)	***	0.06 (3.48)	***	0.06 (3.48)	***
PC1	0.78 (0.60)		0.75 (0.56)		0.79 (0.61)		0.84 (0.65)		0.78 (0.61)		0.79 (0.61)	
PC2	-2.13 (-1.68)	.	-2.13 (-1.69)	.	-2.13 (-1.68)	.	-2.07 (-1.64)	.	-2.11 (-1.67)	.	-2.10 (-1.66)	.
PC5	-2.07 (-1.63)		-2.07 (-1.63)		-2.07 (-1.63)		-2.01 (-1.58)		-2.05 (-1.60)		-2.04 (-1.61)	
Intercept	0.47 (1.26)		0.47 (1.26)		0.47 (1.26)		0.45 (1.20)		0.47 (1.24)		0.46 (1.24)	
Nagelkerke R^2	0.029		0.029		0.029		0.030		0.029		0.029	
Change in R^2			0.000		0.000		0.001		0.000		0.000	
Chi-Square: Deviance, p			0.0082 (0.94)		0.0086 (0.94)		2.0460 (0.22)		0.3632 (0.61)		0.5100 (0.54)	
N	2632		2632		2632		2632		2632		2632	

Current Use (EUR)	***Base Model*** Estimate (z-value)		***p-value = 5 x 10^{2}*** Estimate (z-value)		***p-value = 5 x 10^{-2}*** Estimate (z-value)		***p-value = 5 x 10^{-4}*** Estimate (z-value)		***p-value = 5 x 10^{-6}*** Estimate (z-value)		***p-value = 5 x 10^{-8}*** Estimate (z-value)	
PGRS			0.06 (1.24)		0.04 (0.81)		-0.02 (-0.35)		-0.03 (-0.60)		-0.04 (-0.83)	
Sex	-0.54 (-5.46)	***	-0.54 (-5.44)	***	-0.54 (-5.43)	***	-0.54 (-5.46)	***	-0.54 (-5.46)	***	-0.54 (-5.46)	***
Age	-0.30 (-7.51)	***	-0.29 (-7.45)	***	-0.29 (-7.47)	***	-0.29 (-7.50)	***	-0.29 (-7.50)	***	-0.29 (-7.50)	***
PC1	-6.17 (-2.18)	*	-7.08 (-2.42)	*	-6.52 (-2.28)	*	-6.12 (-2.16)	*	-6.15 (-2.17)	*	-6.10 (-2.16)	*
PC2	-1.21(-0.45)		-1.32 (-0.48)		-1.29 (-0.47)		-1.19 (-0.44)		-1.17 (-0.43)		-1.14 (-0.42)	
PC6	5.18 (1.88)	.	4.99 (1.80)	.	5.06 (1.82)	.	5.21 (1.89)	.	5.21 (1.88)	.	5.23 (1.89)	.
Intercept	7.45 (9.31)	***	7.41 (9.24)	***	7.43 (9.26)	***	7.45 (9.30)	***	7.45 (9.30)	***	7.45 (9.30)	***
Nagelkerke R^2	0.070		0.071		0.070		0.070		0.070		0.070	
Change in R^2			0.001		0.000		0.000		0.000		0.000	
Chi-Square: Deviance, p			1.5399 (0.21)		0.6606 (0.42)		0.1226 (0.73)		0.3594 (0.55)		0.6815 (0.41)	
N	2063		2063		2063		2063		2063		2063	

Regular Use (EUR)	***Base Model*** Estimate (z-value)		***p-value = 5 x 10^{2}*** Estimate (z-value)		***p-value = 5 x 10^{-2}*** Estimate (z-value)		***p-value = 5 x 10^{-4}*** Estimate (z-value)		***p-value = 5 x 10^{-6}*** Estimate (z-value)		***p-value = 5 x 10^{-8}*** Estimate (z-value)	
PGRS			0.003 (0.21)		-0.01 (-0.75)		-0.02 (-1.72)		-0.02 (-1.43)		-0.02 (-1.46)	
Sex	-0.17 (-6.05)	***	-0.17 (-6.05)	***	-0.17 (-6.05)	***	-0.17 (-6.04)	***	-0.17 (-6.04)	***	-0.17 (-6.04)	***
Age	0.08 (6.76)	***	0.08 (6.76)	***	0.08 (6.76)	***	0.08 (6.82)	***	0.08 (6.80)	***	0.08 (6.79)	***
PC1	0.99 (1.29)		0.94 (1.19)		1.07 (1.38)		1.04 (1.36)		1.00 (1.30)		1.01 (1.32)	
PC2	-0.71 (-0.92)		-0.71 (-0.92)		-0.70 (-0.91)		-0.67 (-0.86)		-0.68 (-0.88)		-0.68 (-0.88)	
PC6	1.40 (1.82)	.	1.39 (1.81)	.	1.41 (1.84)	.	1.42 (1.85)	.	1.42 (1.84)	.	1.42 (1.84)	.
Intercept	-0.40 (-1.77)	.	-0.41 (-1.78)	.	-0.40 (-1.77)	.	-0.42 (-1.83)	.	-0.41 (-1.81)	.	-0.41 (-1.80)	.
Nagelkerke R^2	0.037		0.037		0.037		0.038		0.038		0.038	
Change in R^2			0.000		0.000		0.001		0.001		0.001	
Chi-Square: Deviance, p			0.0255 (0.83)		0.3324 (0.45)		1.7218 (0.09)	.	1.1992 (0.15)		1.2415 (0.14)	
N	2980		2980		2980		2980		2980		2980	

Regular Use Among Smokers (EUR)	***Base Model*** Estimate (z-value)		***p-value = 5 x 10^{2}*** Estimate (z-value)		***p-value = 5 x 10^{-2}*** Estimate (z-value)		***p-value = 5 x 10^{-4}*** Estimate (z-value)		***p-value = 5 x 10^{-6}*** Estimate (z-value)		***p-value = 5 x 10^{-8}*** Estimate (z-value)	
PGRS			0.02 (0.45)		-0.01 (-0.22)		-0.02 (-0.51)		-0.04 (-0.87)		-0.02 (-0.95)	
Sex	-0.41 (-4.35)	***	-0.41 (-4.34)	***	-0.41 (-4.36)	***	-0.41 (-4.35)	***	-0.41 (-4.36)	***	-0.41 (-4.36)	***
Age	0.18 (4.85)	***	0.18 (4.86)	***	0.18 (4.84)	***	0.18 (4.86)	***	0.18 (4.86)	***	0.18 (4.86)	***
PC1	3.09 (1.17)		2.77 (1.02)		3.17 (1.18)		3.14 (1.19)		3.10 (1.17)		3.13 (1.19)	
PC2	-2.66 (-1.04)		-2.69 (-1.05)		-2.65 (-1.04)		-2.63 (-1.03)		-2.60 (-1.01)		-2.59 (-1.04)	
PC6	6.97 (2.65)	**	6.90 (2.62)	**	6.99 (2.65)	**	7.00 (2.66)	**	7.00 (2.66)	**	7.00 (2.67)	**
Intercept	-3.68 (-4.84)	***	-3.71 (-4.87)	***	-3.68 (-4.83)	***	-3.69 (-4.86)	***	-3.69 (-4.85)	***	-3.69 (-4.85)	***
Nagelkerke R^2	0.033		0.034		0.034		0.034		0.034		0.034	
Change in R^2			0.001		0.001		0.001		0.001		0.001	
Chi-Square: Deviance, p			0.2053 (0.65)		0.0488 (0.83)		0.2612 (0.61)		0.7486 (0.39)		0.8922 (0.34)	
N	2066		2066		2066		2066		2066		2066	

CPD (EUR)	***Base Model*** Estimate (z-value)		***p-value = 5 x 10^{2}*** Estimate (z-value)		***p-value = 5 x 10^{-2}*** Estimate (z-value)		***p-value = 5 x 10^{-4}*** Estimate (z-value)		***p-value = 5 x 10^{-6}*** Estimate (z-value)		***p-value = 5 x 10^{-8}*** Estimate (z-value)	
PGRS			-0.03 (-1.51)		-0.01 (-0.46)		0.01 (0.39)		0.01 (0.31)		0.01 (0.37)	
Sex	-0.14 (-4.38)	***	-0.15 (-4.37)	***	-0.15 (-4.39)	***	-0.15 (-4.38)	***	-0.14 (-4.38)	***	-0.15 (-4.38)	***
Age	0.04 (3.33)	***	0.04 (3.24)	***	0.04 (3.30)	***	0.04 (3.33)	***	0.04 (3.33)	***	0.04 (3.33)	***
PC1	-1.78 (-1.76)	.	-1.40 (-1.35)	.	-1.07 (-1.66)	.	-1.81 (-1.79)	.	-1.79 (-1.77)	.	-1.80 (-1.78)	.
PC2	-0.08 (-0.08)		-0.03 (-0.03)		-0.06 (-0.07)		-0.09 (-0.10)		-0.09 (-0.10)		-0.09 (-0.11)	
Intercept	0.64 (2.48)	*	0.66 (2.56)	*	0.64 (2.50)	*	0.64 (2.48)	*	0.64 (2.48)	*	0.64 (2.48)	*
Nagelkerke R^2	0.031		0.033		0.031		0.031		0.031		0.031	
Change in R^2			0.000		0.000		0.000		0.000		0.000	
Chi-Square: Deviance, p			0.8216 (0.13)		0.0781 (0.64)		0.0553 (0.70)		0.0345 (0.76)		0.0489 (0.71)	
N	1277		1277		1277		1277		1277		1277	

TFT (EUR)	***Base Model*** Estimate (z-value)		***p-value = 5 x 10^{2}*** Estimate (z-value)		***p-value = 5 x 10^{-2}*** Estimate (z-value)		***p-value = 5 x 10^{-4}*** Estimate (z-value)		***p-value = 5 x 10^{-6}*** Estimate (z-value)		***p-value = 5 x 10^{-8}*** Estimate (z-value)	
PGRS			0.03 (1.13)		0.03 (1.24)		0.03 (1.46)		0.04 (1.77)	.	0.05 (2.02)	*
Sex	-0.20 (-4.44)	***	-0.20 (-4.44)	***	-0.20 (-4.40)	***	-0.20 (-4.44)	***	-0.20 (-4.44)	***	-0.20 (-4.44)	***
Age	0.05 (3.01)	**	0.05 (3.08)	**	0.05 (3.09)	**	0.05 (3.00)	**	0.05 (3.01)	**	0.05 (3.01)	**
PC1	-0.78 (-0.61)		-1.20 (-0.90)		-1.10 (-0.83)		-0.88 (-0.69)		-0.83 (-0.65)		-0.87 (-0.68)	
PC2	-0.03 (-0.03)		-0.10 (-0.81)		-0.09 (-0.07)		-0.11 (-0.10)		-0.14 (-0.12)		-0.16 (-0.13)	
PC5	1.89 (1.60)		1.86 (1.57)		1.87 (1.58)		1.86 (1.57)		1.86 (1.57)		1.86 (1.58)	
PC7	1.96 (1.64)		1.97 (1.65)		1.93 (1.62)		1.97 (1.65)		2.00 (1.67)		1.98 (1.66)	
Intercept	-0.22 (-0.62)		-0.24 (-0.69)		-0.24 (-0.70)		-0.21 (-0.61)		-0.21 (-0.62)		-0.22 (-0.62)	
Nagelkerke R^2	0.032		0.033		0.033		0.033		0.034		0.035	
Change in R^2			0.001		0.001		0.001		0.002		0.003	
Chi-Square: Deviance, p			0.9353 (0.26)		1.1245 (0.22)		1.575 (0.14)		2.2932 (0.08)	.	2.9995 (0.04)	*
N	1476		1476		1476		1476		1476		1476	

FTND (EUR)	***Base Model*** Estimate (z-value)		***p-value = 5 x 10^{2}*** Estimate (z-value)		***p-value = 5 x 10^{-2}*** Estimate (z-value)		***p-value = 5 x 10^{-4}*** Estimate (z-value)		***p-value = 5 x 10^{-6}*** Estimate (z-value)		***p-value = 5 x 10^{-8}*** Estimate (z-value)	
PGRS			0.03 (0.61)		0.06 (1.15)		0.08 (1.39)		0.09 (1.72)	.	0.10 (1.92)	.
Sex	-0.33 (-2.97)	**	-0.32 (-2.96)	**	-0.32 (-2.91)	**	-0.32 (-2.95)	**	-0.32 (-2.93)	**	-0.32 (-2.93)	**
Age	0.22 (5.26)	***	0.22 (5.28)	***	0.22 (5.31)	**	0.22 (5.24)	***	0.22 (5.25)	***	0.22 (5.25)	***
PC1	-1.68 (-0.54)		-2.18 (-0.68)		-2.26 (-0.72)		-1.90 (-0.62)		-1.76 (-0.57)		-1.84 (-0.60)	
PC2	4.04 (1.40)		3.99 (1.37)		3.93 (1.36)		3.91 (1.35)		3.90 (1.34)		3.88 (1.34)	
PC5	-7.61 (-2.62)	**	-7.66 (-2.63)	**	-7.67 (-2.64)	**	-7.70 (-2.65)	**	-7.70 (-2.65)	**	-7.71 (-2.65)	**
Intercept	-2.51 (-2.92)	**	-2.55 (-2.95)	**	-2.56 (-2.98)	**	-2.50 (-2.91)	**	-2.51 (-2.92)	**	-2.51 (-2.92)	**
Nagelkerke R^2	0.134		0.135		0.137		0.138		0.141		0.142	
Change in R^2			0.001		0.003		0.004		0.007		0.008	
Chi-Square: Deviance, p			1.8633 (0.54)		6.6392 (0.25)		9.5950 (0.17)		14.7980 (0.08)	.	18.2930 (0.05)	.
N	1690		1690		1690		1690		1690		1690	

Supplementary Table 10.5a: Main and Interaction Effects of PLINK-based PRS and Parental Environment Variables on Ever Tobacco Use in European Ancestry Group within Spit for Science

Ever Tobacco Use (EUR)	***Base Model*** Estimate (z-value)		***Autonomy Granting Main Effect Only*** Estimate (z-value)		***Involvement Main Effect Only*** Estimate (z-value)	
Parental Environment (PE)			0.18 (4.24)	***	-0.09 (-2.15)	*
Sex	-0.46 (-5.53)	***	-0.45 (-5.37)	***	-0.45 (-5.37)	***
Age	0.11 (3.31)	***	0.11 (3.27)	**	0.11 (3.31)	***
PC1	-0.85 (-0.37)		-0.48 (-0.21)		-0.70 (-0.31)	
PC2	-2.10 (-0.93)		-1.53 (-0.68)	.	-2.04 (-0.91)	
PC5	-4.63 (-2.01)	*	-4.65 (-2.00)		-4.60 (2.00)	*
Intercept	-0.67 (-0.99)		-0.69 (-0.98)		-0.70 (-1.02)	
Nagelkerke R^2	0.022		0.032		0.024	
Change in R^2			0.010		0.002	
Chi-Square: Deviance, p			18.2980 (0.00005)	***	4.7177 (0.03)	*
N	2940		2940		2940	

Ever Tobacco Use (EUR)	***Base Model*** Estimate (z-value)		***Autonomy Granting Main and Interaction Effects with PGRS*** Estimate (z-value)		***Involvement Main and Interaction Effects with PGRS*** Estimate (z-value)	
PGRS (p-value = 5 x 10^{-4})			-0.09 (-2.33)	*	-0.10 (-2.41)	*
Parental Environment (PE)			0.17 (4.19)	***	-0.09 (-2.14)	*
PGRS x PE			0.06 (1.42)		0.01 (0.33)	
Sex	-0.46 (-5.53)	***	-0.45 (-5.35)	***	-0.45 (-5.37)	***
Age	0.11 (3.31)	***	0.12 (3.41)	*	0.12 (3.40)	***
PC1	-0.85 (-0.37)		-0.08 (-0.03)	***	-0.47 (-0.21)	
PC2	-2.10 (-0.93)		-1.34 (-0.59)		-1.87 (-0.83)	
PC5	-4.63 (-2.01)	*	-4.52 (-1.95)	.	-4.46 (-1.93)	.
Intercept	-0.67 (-0.99)		-1.27 (-1.32)		-0.76 (-1.11)	
Nagelkerke R^2	0.022		0.035		0.027	
Change in R^2			0.013		0.005	
Chi-Square: Deviance, p			26.234 (0.000008)	***	10.575 (0.01)	*
N	2940		2940		2940	

Supplementary Table 10.5b: Main and Interaction Effects of PLINK-based PRS and Parental Environment Variables on Time to First Tobacco Use After Waking in European Ancestry Group within Spit for Science

Time to First Tobacco Use After Waking (EUR)	***Base Model*** Estimate (z-value)		***Autonomy Granting Main Effect Only*** Estimate (z-value)		***Involvement Main Effect Only*** Estimate (z-value)	
Parental Environment (PE)			0.07 (3.18)	**	-0.11 (-5.06)	***
Sex	-0.19 (-4.28)	***	-0.18 (-4.09)	***	-0.19 (-4.23)	***
Age	0.05 (2.97)	**	0.05 (2.91)	**	0.05 (3.14)	**
PC1	-0.86 (-0.66)		-0.74 (-0.57)		-0.85 (-0.66)	***
PC2	-0.17 (-0.14)		0.11 (0.09)		-0.06 (-0.05)	
PC5	1.86 (1.57)		1.84 (1.56)		1.93 (1.64)	
PC7	1.83 (1.57)		1.85 (1.55)		1.87 (1.58)	
Intercept	-0.21 (-0.62)		-1.27 (-1.32)		-0.28 (-0.80)	
Nagelkerke R^2	0.030		0.040		0.054	
Change in R^2			0.001		0.024	
Chi-Square: Deviance, p			7.3368 (0.001)	**	18.3730 (0.0000004)	***
N	1455		1455		1455	

Time to First Tobacco Use After Waking (EUR)	***Base Model*** Estimate (z-value)		***Autonomy Granting Main and Interaction Effects with PGRS*** Estimate (z-value)		***Involvement Main and Interaction Effects with PGRS*** Estimate (z-value)	
PGRS (p-value = 5 x 10^{-8})			0.04 (1.95)	.	0.05 (2.14)	*
Parental Environment (PE)			0.07 (3.20)	**	-0.11 (-5.09)	***
PGRS x PE			0.01 (0.33)		0.01 (0.46)	
Sex	-0.19 (-4.28)	***	-0.18 (-4.08)	***	-0.19 (-4.23)	***
Age	0.05 (2.97)	**	0.05 (2.92)	**	0.05 (3.14)	**
PC1	-0.86 (-0.66)		-0.83 (-0.64)		-0.95 (-0.74)	
PC2	-0.17 (-0.14)		-0.02 (-0.01)		-0.18 (-0.15)	
PC5	1.86 (1.57)		1.81 (1.53)		1.93 (1.63)	
PC7	1.83 (1.57)		1.88 (1.58)		1.91 (1.61)	
Intercept	-0.21 (-0.62)		-0.21 (-0.62)		-0.28 (-0.80)	
Nagelkerke R^2	0.030		0.039		0.058	
Change in R^2			0.009		0.028	
Chi-Square: Deviance, p			10.2520 (0.003)	**	21.6880 (0.000001)	***
N	1455		1455		1455	

Supplementary Table 10.6a: Main Effects of Stressful Life Events on Measures of Ever Tobacco Use and Time to First Tobacco Use After Waking within European Ancestry Group within Spit for Science

Ever Tobacco Use (EUR)	***Base Model*** Estimate (z-value)		***Accident*** Estimate (z-value)		***Physical Assault*** Estimate (z-value)		***Sexual Assault*** Estimate (z-value)		***Other Sexual Assault*** Estimate (z-value)		***Natural Disaster*** Estimate (z-value)	
Environment			0.57 (2.65)	**	0.90 (3.17)	**	0.96 (1.89)	.	0.69 (2.24)	*	0.01 (0.03)	
Sex	-0.50 (-2.35)	*	-0.52 (-2.41)	*	-0.46 (-2.11)	*	-0.57 (-2.67)	**	-0.61 (-2.78)	**	-0.50 (-2.35)	*
Age	0.15 (1.64)		0.17 (1.74)	.	0.18 (1.82)	.	0.16 (1.69)	.	0.15 (1.54)		0.15 (1.63)	
PC1	-2.72 (-0.53)		-3.79 (-0.73)		-3.29 (-0.64)		-2.64 (-0.51)		-3.83 (-0.74)		-2.73 (-0.53)	
PC2	3.31 (0.73)		2.96 (0.65)		3.17 (0.69)		2.99 (0.66)		3.22 (0.71)		3.31 (0.73)	
PC5	2.23 (0.36)		3.57 (0.58)		3.90 (0.61)		2.54 (0.41)		2.42 (0.39)		2.24 (0.36)	
Intercept	-1.21 (-0.64)		-1.69 (-0.88)		-1.93 (-1.00)		-1.31 (-0.69)		-0.99 (-0.53)		-1.21 (-0.64)	.
Nagelkerke R^2	0.027		0.048		0.060		0.040		0.043		0.027	
Change in R^2			0.021		0.033		0.013		0.016		0.000	
Chi-Square: Deviance, p			7.2471 (0.0007)	**	11.3400 (0.0008)	***	4.2864 (0.04)	*	5.4930 (0.02)	*	0.0007 (0.98)	
N	495		495		495		495		495		495	

Time to First Tobacco Use After Waking (EUR)	***Base Model*** Estimate (z-value)		***Accident*** Estimate (z-value)		***Physical Assault*** Estimate (z-value)		***Sexual Assault*** Estimate (z-value)		***Other Sexual Assault*** Estimate (z-value)		***Natural Disaster*** Estimate (z-value)	
Environment			0.09 (0.78)		0.46 (3.81)	***	0.84 (4.49)	***	0.63 (4.78)	***	0.08 (0.70)	
Sex	-0.28 (-2.45)	*	-0.28 (-2.50)	*	-0.26 (-2.30)	*	-0.40 (-3.56)	***	-0.43 (-3.77)	***	-0.27 (-2.42)	*
Age	0.08 (2.27)	*	0.08 (2.25)	*	0.09 (2.46)	*	0.09 (2.45)	*	0.09 (2.48)	*	0.08 (2.31)	*
PC1	-5.87 (-1.90)	.	-6.00 (-1.93)	.	-6.36 (-2.11)	*	-5.54 (-1.86)	.	-7.92 (-2.65)	**	-6.08 (-1.95)	.
PC2	2.51 (0.96)		2.38 (0.90)		2.34 (0.92)		1.79 (0.71)		1.30 (0.52)		2.73 (1.03)	
PC5	-1.37 (-0.42)		-1.22 (-0.37)		0.38 (0.12)		-1.15 (-0.36)		-1.02 (-0.33)		-1.10 (-0.33)	
PC7	-1.48 (-0.47)		-1.64 (-0.52)		-1.51 (-0.50)		-1.67 (-0.55)		-1.72 (-0.58)		-1.46 (-0.46)	
Intercept	-0.65 (-0.86)	.	-0.67 (-0.89)		-0.92 (-1.25)		-0.61 (-0.85)		-0.66 (-0.92)		-0.74 (-0.96)	
Nagelkerke R^2	0.097		0.100		0.172		0.199		0.211		0.099	
Change in R^2			0.003		0.075		0.102		0.114		0.002	
Chi-Square: Deviance, p			0.4461 (0.44)		10.0330 (0.0001)	***	13.6450 (0.000007)	***	15.2780 (0.000002)	***	0.3623 (0.48)	
N	230		230		230		230		230		230	

Supplementary Table 10.6b: Main and Interaction Effects of PGRS and Stressful Life Events on Ever Tobacco Use and Time to First Tobacco Use After Waking within European Ancestry Group within Spit for Science

Ever Tobacco Use (EUR)	***Base Model*** Estimate (z-value)		***Accident Main and Interaction Effects*** Estimate (z-value)		***Physical Assault Main and Interaction Effects*** Estimate (z-value)		***Other Sexual Assault Main and Interaction Effects*** Estimate (z-value)	
PGRS (p-value = 5 x 10^{-4})			-0.39 (-2.66)	**	-0.40 (-3.06)	**	-0.29 (-2.31)	*
Environment			0.53 (2.44)	*	0.94 (3.22)	**	0.78 (2.41)	*
PGRS x Environment			1.29 (1.27)		0.58 (1.99)	*	-0.13 (-0.41)	
Sex	-0.50 (-2.35)	*	-0.51 (-2.35)	*	-0.45 (-2.06)	*	-0.61 (-2.76)	**
Age	0.15 (1.64)		0.18 (1.86)	.	0.20 (2.01)	*	0.17 (1.73)	
PC1	-2.72 (-0.53)		-3.49 (-0.66)		-2.56 (-0.48)		-3.71 (-0.71)	
PC2	3.31 (0.73)		4.69 (1.02)		5.14 (1.11)		5.06 (1.10)	
PC5	2.23 (0.36)		2.57 (0.41)		4.71 (0.72)		1.92 (0.31)	
Intercept	-1.21 (-0.64)		-1.92 (-0.99)		-2.37 (-1.20)		-1.42 (-0.74)	
Nagelkerke R^2	0.027		0.068		0.089		0.064	
Change in R^2			0.021		0.033		0.016	
Chi-Square: Deviance, p			14.44 (0.002)	**	21.8240 (0.00007)	***	13.0650 (0.005)	**
N	495		495		495		495	

Time to First Tobacco Use After Waking (EUR)	***Base Model*** Estimate (z-value)		***Physical Assault Main and Interaction Effects*** Estimate (z-value)		***Sexual Assault Main and Interaction Effects*** Estimate (z-value)		***Other Sexual Assault Main and Interaction Effects*** Estimate (z-value)	
PGRS (p-value = 5 x 10^{-8})			0.04 (0.56)		0.07 (1.14)		0.03 (0.48)	
Environment			0.47 (3.85)	***	0.90 (4.84)	***	0.60 (4.42)	***
PGRS x Environment			0.17 (1.48)		0.30 (1.91)	.	0.10 (0.87)	
Sex	-0.28 (-2.45)	*	-0.24 (-2.13)	*	-0.39 (-3.46)	***	-0.40 (-3.54)	***
Age	0.08 (2.27)	*	0.08 (2.23)	*	0.08 (2.37)	*	0.08 (2.34)	*
PC1	-5.87 (-1.90)	.	-6.44 (-2.15)	*	-6.01 (-2.04)	*	-7.98 (-2.66)	**
PC2	2.51 (0.96)		2.06 (0.81)		1.12 (0.45)		1.09 (0.43)	
PC5	-1.37 (-0.42)		0.99 (0.31)		-1.40 (-0.45)		-0.84 (-0.27)	
PC7	-1.48 (-0.47)		-1.19 (-0.39)		-1.80 (-0.60)		-1.42 (-0.47)	
Intercept	-0.65 (-0.86)	.	-0.77 (-1.06)		-0.56 (-0.79)		-0.59 (-0.82)	
Nagelkerke R^2	0.097		0.200		0.237		0.220	
Change in R^2			0.103		0.140		0.123	
Chi-Square: Deviance, p			10.0330 (0.0001)	***	13.6450 (0.000007)	***	15.2780 (0.00002)	***
N	230		230		230		230	

CHAPTER 11:
A MOVING TARGET: THE EMERGENCE OF NICOTINE DELIVERY SYSTEMS AND ITS POTENTIAL PUBLIC HEALTH IMPACTS

Elizabeth K. Do

Tobacco use remains the leading cause of preventable morbidity and mortality within the United States, accounting for over 400,000 deaths per year[1]. Despite an overall decrease in the use of cigarettes, increases in the use of cigars, e-cigarettes, and hookah/waterpipe among adolescents and young adults have been reported[2,3]. It has also been suggested that the overall increase in use of electronic cigarettes and hookah between 2011 and 2014 has offset the overall decrease in more traditional tobacco products, like cigarettes and cigars[4]. Additionally, the use of alternative tobacco products and nicotine delivery systems are evolving rapidly. Estimates from a national survey conducted in 2014 demonstrate that: one in four current cigarette smokers use alternative tobacco products and nicotine delivery systems, and almost one-third of current users of alternative tobacco products are ex-smokers and never smokers[5]. To some extent, this growth in the availability and use of products is driven by market forces: the tobacco industry and financial markets are capitalizing on this period of tobacco product innovation and transformation[6,7].The potential impact of these products on public health is faced with uncertainty, though a report by the 2014 Surgeon General suggests that additional “endgame strategies” are needed to reduce the projected and sustained pattern of tobacco-related morbidity and mortality[1]. This is especially important, given that other tobacco products and nicotine delivery also have negative health consequences, despite most tobacco-related morbidity and mortality being attributed to the use of cigarettes. It

is plausible that the severity of negative consequences may vary across products; but, overemphasizing this point may counter public health efforts to reduce harm.

Categories of Alternative Tobacco Products and Nicotine Delivery Systems

Alternative tobacco products include two types of products: smoked and smokeless. Smoked products include cigars, cigarillos, and hookah/waterpipe, while smokeless products refer to chewing tobacco, snuff (smokeless tobacco), chaw, and dip. Like these alternative tobacco products, electronic nicotine delivery systems (ENDS) – such as electronic cigarettes, or e-cigarettes – are highly available in the US market and are being promoted as potentially less harmful alternatives to cigarettes. They are also being marketed as products to be used as substitutes for cigarette smoking. However, these products occur more often in combination with the use of cigarettes, rather than in isolation. Multiple studies of tobacco use have reported concurrent use of tobacco products among college students[8] – a population that is prone to experimentation and is actively developing patterns of tobacco use behaviors that could affect their risk for tobacco-related morbidity and mortality. Studies demonstrate that concurrent users of tobacco products and nicotine delivery systems experience higher intermediate levels of mortality, tend to ingest more nicotine daily, and have more difficulty when trying to stop the use of tobacco[9,10].

Why Alternative Tobacco Products/ Nicotine Delivery Systems are Growing in Popularity

Despite the harms involved with any tobacco use, alternative tobacco products and electronic nicotine delivery systems are growing in popularity – especially among

adolescents and young adults. The growing popularity of these products is attributed to multiple factors, though the most common explanations are: the widespread availability, lax regulation, targeted advertising, and perceptions that these products are safer, relative to cigarettes.

Widespread availability of alternative tobacco products. As new tobacco products are being developed, the availability of alternative tobacco products and nicotine delivery systems will grow over time[11]. By January 2014, there were more than 460 brands and 7760 unique flavors of electronic nicotine delivery systems available for purchase online[12]. Reviews of tobacco industry documents demonstrate that tobacco companies are using flavoring to make tobacco products more palatable and attractive to new users of tobacco products[13]. The widespread availability of alternative tobacco products may be driven by the changing preferences that exist among current and former users of tobacco products, who are either looking for a product to use instead of or in addition to cigarettes[14].

Targeted advertising. Tobacco product marketing has emphasized the point that alternative tobacco products may facilitate reduction or cessation of cigarette use[15]. Meanwhile, some advertisements use messages that position alternative tobacco products as a modern substitute to be used in places where smoking is banned or made inconvenient by smoking bans[15,16]. Given the declining cigarette consumption, major US tobacco companies have tried to maintain their profits through the marketing (and development) of alternative tobacco products and nicotine delivery systems[17,18]. Much of the current research has focused on how tobacco companies have focused on the increased use among youth, who are more price sensitive and interested in tobacco-

masking flavors relative to older adults[14]. Manufacturers of alternative tobacco products and electronic nicotine delivery systems have utilized lax regulation over the marketing of alternative tobacco products to their advantage in that they have aggressively marketed products to youth through product flavoring, promotional materials, and the distribution of free samples[19].

Lax regulation. The FDA currently holds immediate authority over cigarettes, smokeless tobacco products, and roll-your-own tobacco. As of October 2015, the Food and Drug Administration lacks the authority to regulate novel tobacco products, such as electronic vapor products marketed for non-therapeutic purposes[20], even though the FDA can promulgate regulations extending regulatory authority over all other products meeting the definition of tobacco product under the Family Smoking Prevention and Tobacco Control Act (FSPTCA)[21]. The FSPTCA was originally enacted in 2009, and provided the FDA with the power to regulate the manufacture, sale, marketing and distribution of tobacco products. Under the FSPTCA, the FDA has the power to restrict marketing and sales of tobacco products to youth, mandate reporting of ingredients and additives, ban cigarette flavorings, and review manufacturers' claims to lower-risk[19]. In 2016, the FDA finalized a rule (also referred to as the "Deeming Rule") extending the agency's authority to regulate electronic cigarettes and related vapor products as tobacco products. This rule, makes it illegal for any e-cigarettes, e-liquids, and other tobacco products without FDA approval to remain on the market and takes effect on August 8, 2018. What this means is that manufacturers must submit a package of research required by the FDA prior to this date, to be considered for approval to stay on the market[22].

Perception that products are safer relative to cigarettes. Research has suggested that marketing focused on potential-harm reduction aspects of alternative tobacco products may cause tobacco users to reject or ignore health messages regarding the potential dangers of these products[23]. There is also a common belief among users of tobacco products that the government evaluates most tobacco products for safety. This is problematic in that the belief that the government evaluates certain products for safety (when many tobacco products are unregulated by the government) may contribute to continued use[24]. It becomes even more problematic when we consider that polytobacco users report the least perceived dangers of various tobacco products[25], despite heightened risk for escalated use and addiction among individuals who use more than one tobacco product concurrently. Polytobacco users may view alternative tobacco products as less dangerous, perhaps because they see alternative tobacco products as an effective means for cessation, or as an acceptable substitute when smoking is publicly prohibited[26].

Who Uses Alternative Tobacco Products and Nicotine Delivery Systems?

Some research has been done to identify which individuals are at higher risk of using alternative tobacco products and nicotine delivery systems. Generally, individuals with peers or family members supporting tobacco use, who hold favorable beliefs about smoking, are younger in age and male are at higher risk of using alternative tobacco products. Some recent studies suggest that the uptake of certain alternative tobacco products may differ by race/ethnicity, but the evidence is mixed. Whereas some studies have indicated that individuals who self-identify as White endorse using alternative tobacco products relative to other race/ethnicities[27], others have demonstrated that the

prevalence is higher among those who self-identify as Black[28]. Furthermore, those who are never smokers are at highest risk of using alternative tobacco products, relative to former, non-daily smokers and non-smokers. The reason for this might be that young adults who use alternative tobacco products might not (want to) perceive themselves to be smokers[28]. Rather, initial users of alternative tobacco products and electronic nicotine delivery systems may self-identify as non-users who are experimenting, or only users within social settings. Data from focus groups conducted to assess themes related to e-cigarette experimentation and discontinuation support this viewpoint by showing that the top reasons for experimenting with or initiating the use of electronic nicotine delivery systems were curiosity, appealing flavors, and peer influences[29]. Since almost a third of current users report that they are nonsmokers, the use of alternative tobacco products and electronic nicotine delivery systems are contributing to nicotine addiction and renormalizing tobacco use[30,31]. The extent to which alternative tobacco products and electronic nicotine delivery systems are contributing to nicotine addiction and tobacco-related morbidity and mortality is not yet clear, though there is evidence demonstrating that concurrent use of tobacco products may sustain nicotine dependence and potentially postpone cessation[32]. More studies need to be conducted to quantify the risks involved with using alternative tobacco products and electronic nicotine delivery systems.

Future Directions of Research

The main challenge to public health regarding tobacco control is how to address the growing use of alternative tobacco products and nicotine delivery systems and balance public health messaging about the use of these products. It remains unclear whether these alternative tobacco products and nicotine delivery systems will be replacing

traditional tobacco products without expanding patterns of nicotine use among adolescents and young adults – the main target of tobacco company advertising and public health harm reduction and prevention efforts. As information is collected regarding how much and how these products are being used, the availability and marketing of these products will continue to grow and change with user preferences[5] – especially since young adulthood is an ideal period during which tobacco use can be introduced and solidified[33].

One area of research that needs more attention is the relationship between the use of alternative tobacco products and electronic nicotine delivery systems, and traditional cigarette use. Alternative tobacco products might serve as gateway products for adolescents, leading to the use of other tobacco products – including cigarettes. A study conducted by Soneji et al. (2015), demonstrates this by providing data from a national survey of young adults aged 15-23 years, showing that baseline water pipe tobacco and snus use were associated with increased interim cigarette initiation, current cigarette smoking, and high-intensity cigarette use after a 2-year follow-up period[34]. Added to this, half of adolescent tobacco users within the US are dual or poly-tobacco users[35]. Yet, their tobacco use trajectories remain poorly characterized. We know from other studies, that several young adults have used multiple tobacco products in their lifetime, and are current dual and poly-tobacco users[36]. More work needs to be done to better understand the increasing complexity of tobacco use among adolescents and young adults to promote effective public health planning and to ensure that the regulation of alternative tobacco products (or potential lack thereof) does not undermine current anti-tobacco regulatory

efforts[19] which have predominantly been focused on consumption changes, successful quitting, as well as attitude and belief changes among current users[37].

To reduce and/or prevent the use of alternative tobacco products, more data needs to be collected on the use of all tobacco products as they are being developed, rather than just focusing on cigarettes[38]. It also seems to be the case that a two-pronged approach needs to be taken in the research field: one focused on how to alter existing social norms regarding tobacco use that may be undermining current tobacco control efforts, and one focused on the reduction of harm of products that are currently and/or will be made available in the future. Much of this research will be driven by changes in consumer-driven preferences already being studied by tobacco companies, who pay close attention to how products are being used and adapted[39].

Thus, in addition to determining the prevalence of use of different tobacco products, better characterization of tobacco products in terms of composition and toxicity should be undertaken to inform regulators to develop guidelines for safety and focus should be placed on minimizing risks associated with tobacco products and nicotine delivery systems[12]. It is also important to determine what the overall morbidity and mortality contributions of different tobacco products are, relative to cigarettes and in combination with cigarette use – given that concurrent tobacco use of multiple products is increasingly common, especially among adolescents and young adults who are the moving targets of both the tobacco industry and public health messaging.

CHAPTER 12:
PLANS FOR A PILOT RANDOMIZED CONTROL TRIAL OF AN INTERNET-BASED EDUCATIONAL INTERVENTION FOR THE REDUCTION OF TOBACCO USE (AND NICOTINE DEPENDENCE)

Elizabeth K. Do, Juan Lu, and Hermine H. Maes

INTRODUCTION

Tobacco use results in $130 billion annually in direct adult medical care costs. It is projected that 5.6 million of Americans under the age of 18 years will die prematurely from tobacco-related illnesses. Thus, developing methods to reduce tobacco use among individuals is necessary[1]. However, before this is possible, a better understanding of the etiology of tobacco use and nicotine dependence is required and examining risk factors for tobacco use behaviors may be useful for the identification of potential areas of intervention and prevention.

Tobacco use behaviors involve the interplay of genetic and environmental factors. Adding to this complexity, the influence of genes and the environment seems to change over the life course**.** Twin studies of adult samples show that genetic and shared environmental influences contribute significantly to the liability of tobacco initiation, regular use, and nicotine dependence[2–4], with significant overlap in the genetic and/or environmental risk factors at each stage[5]. Adolescent studies suggest that the impact of the shared environment is more pronounced during mid-adolescence when many initiate use, but the influence of genes - though present to some degree early on - comes to play a larger role by late adolescence, when the etiological structure of tobacco use resembles that of adults[5–7], such that initiation is explained by genetic, shared, and unique environmental

factors in early adolescence and genetic and unique environmental factors in young adulthood[8].

Despite the growing evidence that genetic factors play an important role in tobacco use, genome wide association studies (GWAS) have only identified a few genes associated with smoking quantity and nicotine dependence so far. Identification and exploration of genetic loci influencing smoking behavior have primarily been conducted in adults and in populations of European ancestry[9]. The most robust finding to emerge from GWAS studies of smoking behavior is the association between genetic variant *rs16969968,* located within the α5-α3-β4 nicotinic receptor gene cluster on chromosome 15, and smoking quantity[10–13]. A study of African Americans confirmed this region as an important susceptibility locus for smoking quantity in men and women. There is also a reported association between *rs1051730* and nicotine dependence (and two tobacco-related diseases, lung cancer and peripheral arterial disease)[13]. Yet, larger studies are needed to validate other suggestive loci not reaching genome-wide significance[9].

Consortia-based GWAS meta-analyses of individuals of European ancestry with sample sizes approaching 10,000 individuals have identified several novel genomic regions associated with a range of smoking phenotypes. A nonsynonymous SNP (*rs6265*) located on the brain-derived neurotrophic factor (*BDNF*) gene on chromosome 11 has been associated with smoking initiation, while a gene variant on chromosome 9 near the dopamine beta hydroxylase (*DBH*) gene was associated with smoking cessation[12,14]. Two nicotinic receptor sub-unit genes (*CHRNB3/CHRNA6*) on chromosome 8 and *CYP2A6* on chromosome 19 have also been associated with smoking quantity[12]. However, the proportion of phenotypic variance explained by *rs16969968-rs1051730* SNPs is less than

1%[15]. A recent genome-wide meta-analysis of an objective marker of smoking has observed an association between multiple variants within the 15q24 region, in strong linkage disequilibrium with *rs16969968*, associated with cotinine level[16] and explaining a larger proportion of the phenotypic variance.

Despite studies demonstrating that genes contribute to tobacco use behaviors, current prevention and intervention approaches do not take the contribution of genetic factors to tobacco use into account. Instead, focus is placed on mean behavior change, either through advice and behavioral counseling at the individual level[17] or through population-level tobacco control measures such as tobacco advertising bans seeking to reduce smoking initiation and prevalence among minors and young adults or smoke-free policies aimed to reduce secondhand smoke exposure to nonsmokers and create an environment that aids smokers to quit[18]. Using datasets incorporating measured aspects of the environment into genetically informed studies allows for a greater understanding how specific environments are related to tobacco use outcomes, interact with genetic liability, and are helpful in identifying potential points of intervention. For example, studies have identified parental environment, parental monitoring[19,20], maternal smoking during pregnancy[21], peer smoking[22], average neighborhood social cohesion[23], traumatic events[23], self-rated religiousness[24], and marketing and vending machine restrictions[22] as moderators of genetic and/or environmental factors influencing smoking behaviors, as summarized in two recently published systematic reviews[25,26].

Other studies suggest that notification of susceptibility for tobacco-related illness may influence individual-level tobacco use. Interestingly, awareness of the health hazards associated with cigarette consumption appears not to be sufficient in the initiation of

smoking cessation[27]. This could be due to tobacco users' underestimation of personal risk of tobacco-related illness[28]. Research suggests that improvements in cessation rates may be achieved by providing personal feedback on susceptibility for tobacco-related illness[29]. Short-term benefits of this approach include: positive change in perceptions of risk and beliefs about quitting[29–31] and increased attempts to quit/enhanced cessation rates at six months[32]. However, empirical data have not provided sufficient evidence that knowledge of genetic variants conferring susceptibility to tobacco-related illness yields long-term benefit in terms of quit rate[33]. Additionally, there are concerns that genotypic notification could demotivate high-risk individuals to change their behavior, due to feelings of fatalism or reduced sense of personal control over chances of getting smoking-related diseases, while individuals receiving low-risk results may be falsely reassured and become complacent[34]. Further research is needed to delineate the direction of effect of genetic notification of susceptibility for smoking-related outcomes.

The current study tries to build on this existing research by providing information on the planning of a feasibility study involving an Internet-based educational intervention examining how providing college students with information on the influence of genes and the environment on tobacco use behaviors and nicotine dependence impacts subsequent patterns of tobacco use behavior. To date, few interventions are specifically aimed at young adult smokers, even though tobacco use is common among college students. College is a critical time in the development of tobacco use behaviors[35], especially since young adults are at risk for experimenting with tobacco use and establishing tobacco use patterns. Furthermore, their current use may be predictive of tobacco use in years to come. As such, tobacco use prevention, reduction, and nicotine dependence treatment

efforts aimed at college students could have the potential to yield considerable benefits in reducing the overall health burden of nicotine dependence[36].

This study will apply the principles of the Health Belief Model (HBM), which assumes that health behavior is determined by perceptions of perceived threat, perceived susceptibility and severity, perceived benefits, and perceived barriers and the strategies available to decrease its occurrence[37]. By applying the HBM concept and constructs to our intervention, which seeks to reduce tobacco use (and risk for nicotine dependence) in young adult tobacco users, we seek to increase their perceived threat, susceptibility, and severity of nicotine dependence while decreasing their perceived barriers to reducing tobacco use among college student participants by providing knowledge of genetic and environmental risks for nicotine dependence and means to decrease barriers to reducing tobacco use. We hypothesize that personalizing the information presented (e.g. using the pronoun “you” in explanations of risk) will improve outcomes, relative to individuals in the control group and those receiving generalized risk information about tobacco use behaviors and risks for nicotine dependence. We also predict that this reduction of tobacco use will be the same or higher for those who also receive information on their genetic risk for developing nicotine dependence.

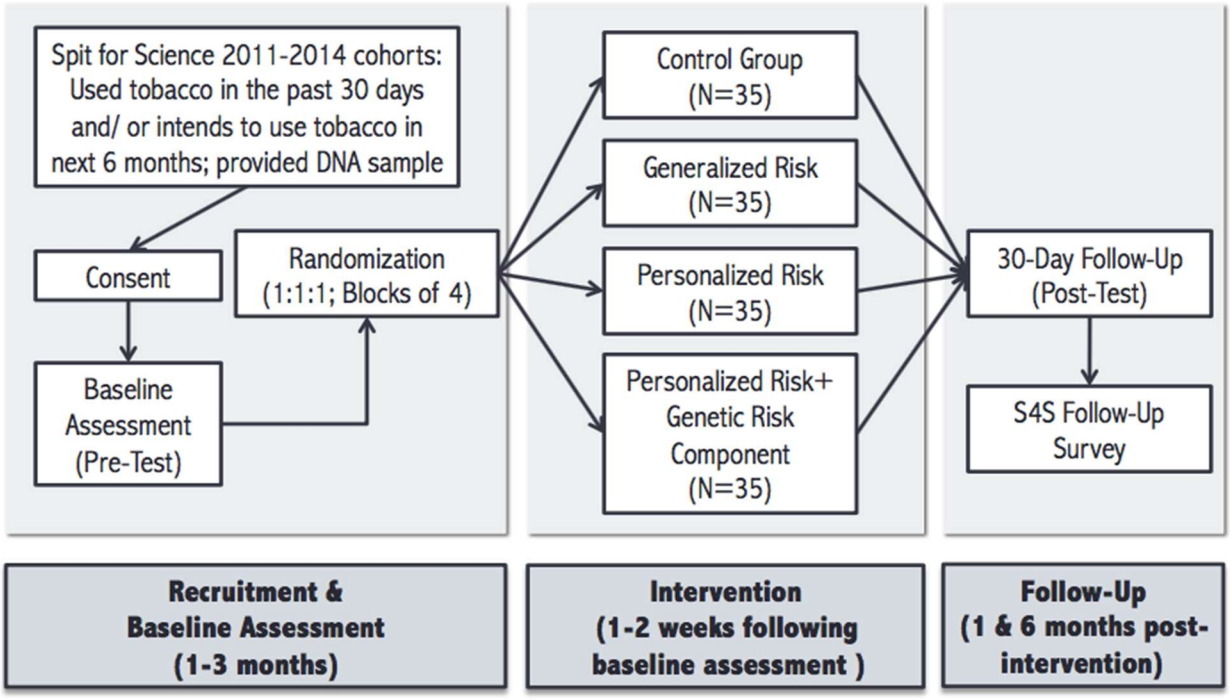
METHODS

Design. Participants will be recruited into a four-arm randomized control trial pilot study, from an ongoing study that investigates the role of genes and the environment on emotional health and substance use within a university setting. The four-arms of the study include a control, generalized risk information, personalized risk information, and

personalized risk information with added genetic component group. Baseline characteristics and outcomes of interest will be compared across each of the four arms of the study groups. Recruitment, study consent and enrollment, and data collection will all be conducted online, with no face-to-face contact. The registry of the parent study will initially contact eligible participants to determine interest in the study, and report back to this research study team. The research team will then ask those who indicated their interest in participation in the study to fill out an electronic consent form and continue with baseline assessment. The baseline assessment will be found as a survey link that participants can click on. However, before providing consent, participants will need to read study information that outlines procedures, inclusion criteria, and will be encouraged to contact the research team regarding any questions/comments they have regarding the study protocol. After providing electronic consent, participants will complete the baseline assessment and then be randomized to receive one of the study conditions. Participants who indicate that they are not comfortable with receiving any genetic risk information will not be allocated to the personalized risk with genetic component group. A few weeks after the completion of the baseline assessment, participants will be automatically followed-up through e-mail and asked to complete the internet-based educational intervention, which includes: online self-assessment questionnaires and viewing a set of three-minute videos. Participants who do not complete the internet-based educational intervention within a couple of weeks will be re-contacted via reminder e-mail. Thirty days following the completion of the internet-based educational intervention, participants are re-contacted to complete an online self-assessment follow-up questionnaire. Again, participants who do not complete the online follow-up were re-contacted via e-mail. Figure 12.1 depicts

study design and recruitment strategy. The university Institutional Review Board approved the study.

Figure 12.1: Study Protocol



Participants and Eligibility Criteria. We aim to recruit 140 participants (35 participants per arm). This estimate is based upon the available budget and expected recruitment rates, from previously conducted spin-off studies. To be eligible for the study, participants had to be: at least 18 years of age, previously enrolled in the parent study, Spit for Science, provided a DNA sample, and indicated that they had used any tobacco product (e.g. cigarettes, cigars/cigarillos, hookah/waterpipe, smokeless tobacco, and/or electronic cigarettes or other nicotine delivery systems) in the past 30 days on their most recent Spit for Science follow-up survey. Eligible participants must confirm that they either currently

use or intend to use tobacco products in the next six months (i.e. the duration of the study) during the consent process.

Compensation. Participants will be compensated for participating in the study, which requires: one baseline assessment, an internet-based educational intervention (which includes a set of three-minute videos and surveys), and one follow-up assessment. Participants will be compensated $10 for each of these assessments and will be eligible to receive an additional $10 for the completion of all assessments.

Recruitment Duration. Recruitment will end once we reach our desired sample size of 140 participants, which is estimated to take about two months. Recruitment will be extended by one month, if the sample size is not reached. If, after three months, the desired sample size is not reached, study procedures will continue with whatever sample size is obtained and the analysis procedure will be adapted accordingly.

Consent Process. Participants will be asked to answer questions to determine eligibility for this study and were told that by participating, they will be randomly assigned to one of four educational intervention groups: control group, generalized risk information, personalized risk information, and personalized risk information with genetic component. Participants will not be told what their assignment is, but will be given the opportunity to obtain information from the alternative condition(s) to which they were assigned, after the termination of the study. All participants will be told that they will be asked to watch a set of 2-3 minute videos and complete online survey questions regarding their own tobacco use behavior and life experiences. Participants will also be notified that by consenting to participate in this study, they will be granting researchers access to data from Spit for

Science[38] surveys and the DNA sample that they provided previously. However, all responses will be stripped of identifiers, coded, and linked with survey data and the DNA sample provided for Spit for Science using only coded numbers. Research analyses will be done without names attached to the data. Participants will not receive any direct benefit from completing the study, other than compensation for their participation, and the study involves no more than minimal risks. Finally, participants will be reminded that as voluntary participants, they have the option to withdraw from the study at any time.

Data Capture and Management. Participant data will be collected and maintained in REDCap[39], the online survey tool that will send survey invitations and receive the survey data. All private identifiable information and data collected will be secured in a REDCap database that is accessible only to the research coordinator and staff. After the study is complete, REDCap data will be de-identified by the research coordinator and extracted for analysis. Only de-identified data will be shared with investigators.

Outcome Measures of Interest. This study aims to explore the feasibility and efficacy of the proposed intervention. Feasibility outcomes included: recruitment and retention rates, acceptability of the intervention, and follow-up and adherence/compliance rates. Although many measures will be collected as a part of this project, primary efficacy outcomes included: tobacco use behaviors of participants (e.g. current use, frequency of use, and FTND following the intervention). Secondary outcome measures include measures that may influence the primary efficacy outcomes, such as: self-efficacy to quit using tobacco, perceived benefits to reducing tobacco use, perceived susceptibility to tobacco use, and perceived barriers to tobacco use. An overview of the collected measures is shown below, in Figure 12.2.

Figure 12.2: Overview of Survey Measures

Survey Item Set	# Items	Baseline Assessment	Intervention	Follow-Up (30-Day)	Spit for Science Follow-Up
Family and Peer History of Tobacco Use	4	x			
(Past and) Current Tobacco Use	15	x	x	x	x
Fagerstrom Test for Nicotine	6	x		x	x
Feelings about Tobacco Use	2	x	x		
Self-Efficacy to Quit Using Tobacco	2	x	x	x	
Perceived Benefits to Reducing Use	3	x	x	x	
Perceived Susceptibility to Tobacco	3	x	x	x	
Perceived Barriers to Reduced Tobacco Use	1	x	x	x	
Perceived Severity of Tobacco Use Risks	10	x	x	x	
Motivations to Obtain Genetic Testing	3	x	x	x	
Knowledge of Nicotine Dependence Risk Factors	16	x	x	x	
Demographic Characteristics	5	x			
Video Module: Genes and Nicotine Dependence	1		x		
Video Module: Family Tobacco Use, Family Environment, and Risk for Tobacco Use and Nicotine Dependence	2		x		
Video Module: Peers, Early Experiences with Tobacco, and Symptoms of Nicotine Dependence	1		x		
Video Module: Stress, Tobacco Use, and Nicotine Dependence	1		x		
Motivation to Change Tobacco Use Behavior	1		x		
Confidence in Knowledge of Risks for Nicotine Dependence	3		x		
Program Evaluation	3		x		
Total Number of Items	189	169	67	59	21

Baseline Assessment. Eligible participants will be asked to fill out a questionnaire immediately after consent is provided via a survey linked to REDCap that includes measures of: demographic characteristics (i.e. race/ethnicity, sex, year in college, health insurance status), family history of tobacco use, current tobacco use (i.e. tobacco products used, quantity, frequency, and quit attempts), FTND, self-efficacy to quit using tobacco, perceived benefits to reducing tobacco use, perceived susceptibility to tobacco use, perceived barriers to tobacco use, perceived severity of risks due to tobacco use, motivations to obtain genetic testing, and knowledge of nicotine dependence risk factors.

Randomization. Randomization will be conducted using the Randomization Module in REDCap using a 1:1:1:1 ratio, in blocks of 4 following the collection of consents and baseline assessments.

Intervention Descriptions. Two weeks following randomization, participants will be contacted via e-mail to let them know that the internet-based educational intervention is ready to be completed. Again, the participants will not know to which intervention arm they are assigned. All groups will receive a set of surveys to complete, along with a set of four 2-3 minute videos. All groups, but the control group, will watch educational video modules describing risk factors for nicotine dependence. The control group will be shown a set of videos not related to risk factors for nicotine dependence.

The main difference between the generalized risk group and the personalized risk groups (one with genetic component and the other without) will be the way the information is presented. For the personalized risk groups, information will be framed using the word "you" (e.g. You are at slight/mild/moderate/severe increased risk for nicotine dependence) and for the generalized risk group, information will be framed in terms of "an individual" (e.g. "Individuals who ____ are at slight/mild/moderate/severe increased risk of nicotine dependence"). In addition to receiving personalized feedback on risks for nicotine dependence, participants receiving the personalized risk with genetic component condition will also receive information regarding genetic variant(s) associated with nicotine dependence. Participants will be advised that having this genetic variant does not guarantee that they will develop nicotine dependence, but rather, having certain genetic variants increases their likelihood of developing nicotine dependence and this information should be regarded as preliminary research that is ongoing in the field of

behavior genetics. Participants will have one week to complete the educational intervention and brief questionnaires that go along with the educational intervention. Both groups will also be given the contact information of a genetic counselor, in case they have questions regarding the interpretation of genetic and environmental risks related to nicotine dependence.

Video Modules and Questionnaires. Video modules will cover the following topics: *(1) genetics and nicotine dependence*, which explains how genetic factors contribute to tobacco use behavior*, (2) family history and parental environment in adolescence*, which explains how individual tobacco use is shaped by family history and environment, *(3) early experiences with tobacco use*, which explains how early experiences shape tobacco use patterns, and *(4) environmental stressors*, which provides information on the influence that stress may have on tobacco use. The short questionnaires following each module will be used to gauge comprehension and knowledge of genetic and environmental risk information provided in the online video modules. They will also be used to determine if a participant completed the educational modules and will be like quizzes commonly used in college-level courses. Participants will also be asked to answer a few questions on: how they define their tobacco use, potential exposure to nicotine in-utero, family tobacco use, symptoms they had when they first started using tobacco, parental monitoring, and stressful life events. Questionnaires will also contain measures on self-efficacy to quit using tobacco, perceived benefits to reducing tobacco use, perceived susceptibility to tobacco use, perceived barriers to tobacco use, perceived severity of risks due to tobacco use, and motivations to obtain genetic testing for comparison to baseline. It is estimated

that this one-time intervention will take 35-45 minutes (including educational videos and questionnaires).

30-Day Follow-Up Assessment. Approximately 30 days following the completion of the educational intervention, participants will be invited to take a 30-minute follow-up survey via email with a link to the questionnaire which includes questions on: knowledge of nicotine dependence risk factors, current tobacco use, self-efficacy to quit using tobacco, perceived benefits to reducing tobacco use, perceived susceptibility to tobacco use, perceived barriers to tobacco use, perceived severity of risks due to tobacco use, and motivations to obtain genetic testing. This will serve as the second point of assessment, since the questions will be the same as those asked at baseline, and immediately following the educational intervention. Participants will also be asked to answer several questions regarding their perceptions of the proposed intervention and its acceptability, to serve as process evaluation measures. The follow-up survey will be available for two weeks.

6-Month Follow-Up Assessment. As a third point of assessment, information will be gathered from the Spit for Science Spring follow-up survey regarding current tobacco use and the Fagerström Test for Nicotine Dependence.

Data Analysis. Analyses regarding acceptability and feasibility outcomes will be mainly descriptive. Recruitment rate (i.e. ratio of individuals who complete the eligibility screener to those that are randomized) and retention rate (i.e. ratio of number of retained participants to number of participants enrolled in the study) will be reported. To assess whether outcomes are affected by attrition bias, we will examine the baseline

characteristics (obtained from parent survey registry data; e.g. sex, race/ethnicity, year in school, average frequency/quantity of tobacco use) of participants who were lost to follow-up and those remaining and compare rates of loss to follow-up between the arms of the intervention. Baseline characteristics of participants' tobacco use will be reported as means/standard deviations for continuous variables and frequencies/percentages for categorical variables. Chi-square tests, Mann-Whitney tests, or t-tests will be used to examine baseline differences between intervention arms in terms of demographic characteristics, tobacco use measures, and other assessment measures between the different intervention conditions. Regression models will be used to determine if there are differences between the outcomes of the intervention arms. Potential covariates included in these analyses include: sex, race/ethnicity, and type/ number of tobacco products used.

The data will be analyzed on an intention-to-treat (ITT) basis, as if participants who do not complete the trial had not reduced tobacco use. Under ITT, data for all subjects randomized to a treatment is analyzed according to the treatment to which subjects were allocated, regardless of whether they received the treatment or not. Its purpose is to preserve the theoretical basis for the validity of statistical results, particularly by eliminating subjects with prognostic factors, which could systematically influence a participant's selection to a given treatment[40]. This approach was selected because a previous study demonstrates that ITT gives unbiased estimates of treatment effects (e.g. effectiveness)[41], while also preserving the sample size. However, the estimate of treatment effects is generally more conservative, relative to alternative approaches such as-treated (or per-protocol) analysis[42].

As-treated analysis involves analyzing study data from only subjects who complete the study and adhere to protocol requirements. The main advantage of this approach is that only participants with complete data are examined, and the resulting information can be used to determine what the effect of treatment is, when taken in an optimal manner (e.g. efficacy). However, as-treated analysis could lead to a significant reduction in sample size, especially if many participants do not follow protocol, and is prone to potential bias, since participants might deviate from the protocol for non-random reasons, such as experiencing unpleasant side effects from treatment, or failure to see improvement in health[43].

The main outcome measure calculated would be the reduction in tobacco use incidence ratio, using two-tailed Fisher's exact tests to compare the proportion of participants who reduced tobacco use between the intervention groups. Two-tailed paired t-tests will also be used to compare the change in the average use of tobacco overall (and if applicable, across each tobacco product). Tobacco reduction incidences will be computed at 30-day and 6-month follow-up. All analyses will be conducted using SAS 9.4.

DISCUSSION

Potential Limitations. Participant recruitment is essential to the success of this project. However, we are confident that we will be able to recruit the desired number of participants since procedures to recruit these participants have already been established and utilized successfully in the past. For example, of those invited to the most recent spin-off study, 56.5% expressed interest and 40.4% were enrolled. Thus, a potential strategy to overcome this obstacle of participant recruitment is to contact many participants during

the recruitment period. There might also be concerns over whether there will be adequate power to conduct the planned analyses**.** Fortunately, the sample size of the brief intervention proposed is within the range of sample sizes found in related studies (range: 61-697)[35]. However, since the suggested sample size for this study is towards the lower end of the range, results from this study should be regarded as preliminary and replication is needed for any effects that we identify. Related to sample size, are the concerns regarding participant drop-out. Short-term follow-up duration does protect the study against possible participant dropout; however, it may also mask the possibility of longer-term effects. To get around this issue, it would be possible to look at long-term effects from responses provided in follow-up surveys conducted by the parent study, collected every spring semester. One last potential limitation can be found in our reliance on self-reported data. Some individuals may feel uncomfortable sharing information about their substance use and/or answering questions about their previous tobacco use. However, since this study is being conducted over the Internet and there is minimal to no personal contact, there is unlikely to be any differential biases by intervention group[44].

*Study Strengths and Potential Contributions***.** The contributions of this proposed project are potentially significant due to its focus on young adulthood, a critical period of change in the lifespan and the period when the use and abuse of substances peaks. Furthermore, this proposed project will be one of only a few tobacco reduction interventions aimed at non-treatment-seeking college students, a population with the highest prevalence of substance use and the greatest likelihood to quit following intervention. This is an especially important task to undertake, given that few studies examine how presenting genetic and environmental risk information to tobacco users will influence patterns of

tobacco use through increased knowledge of risk factors for nicotine dependence, as well as motivations and barriers to the reduction of tobacco use. Given that tobacco use among young adults is both common and associated with several adverse consequences (i.e. tobacco-related illness and addiction, cancer, and death), the potential impact of intervention on the overall disease burden of this group is large. This emphasizes the need to understand factors contributing to substance use in young adulthood, to which this study contributes by including longitudinal, repeated measures of tobacco use prevalence and behavior across multiple tobacco products. Furthermore, Spit for Science is one of just a few samples asking survey questions about more recent forms of tobacco use, such as electronic cigarettes and hookah use. Finally, studying college students allows for a unique opportunity to intervene, since all major life activities are concentrated in a single setting.

Future Implications

The present study will provide information on whether providing college students information about the risk factors for tobacco use behaviors and nicotine dependence impacts tobacco use behaviors, such as frequency and quantity of use, current use, and symptoms of nicotine dependence as measured by the Fagerström Test for Nicotine Dependence, through self-efficacy, perceived benefits to quitting, perceived susceptibility to risks, and perceived barriers to reducing tobacco use. Even if we find the effects of the intervention to be small, the information gained from the survey measures collected can be used to characterize motivations for tobacco use within young adults. Process evaluation measures, such as program acceptability, can be used to inform follow-up studies and future interventions.

CHAPTER 13:
GLOBAL DISCUSSION - THEMES, RESEARCH FINDINGS, AND FUTURE DIRECTIONS

This dissertation thesis presents a comprehensive set of studies investigating genetic and environmental risk for tobacco use behaviors and nicotine dependence, using samples of adolescents and young adults, covering a broad range of methodologies. Altogether, this set of studies contributes to the existing literature by providing information on how genes, the environment, and their potential interactions may influence tobacco use behaviors.

THEMES

In conducting these studies, several overall themes emerge regarding the complexity of tobacco use behavior phenotypes specifically, and how we investigate risk factors for complex traits, such as tobacco use, more broadly. The themes resulting from this research, which integrates literature reviews, twin/family studies, population-based studies, epidemiology, and genomic analyses include:

1. Behaviors are complex, and we need to develop better strategies for the harmonization and standardization of phenotypes.
2. Although twin methods have been found to obtain higher heritability estimates than genomic approaches for complex traits, such as tobacco use behaviors, they are still very useful in determining genetic and environmental architecture, and identifying potential covariation and interaction between genetic and environmental influences.

3. Polygenic risk scores might account for more variance in traits than single genetic variants do individually, but the clinical utility of these genomic approaches are still to be determined.
4. Public health planning needs to consider, and where possible, incorporate, knowledge regarding social and genetic influences on complex behaviors, such as tobacco use.

RESEARCH FINDINGS

The themes listed above were derived from examining each of the chapters presented in this thesis. Under each thematic heading are more specific details about how research findings from singular studies in this dissertation are related to the listed themes.

Developing better strategies for harmonization and standardization of phenotypes

Phenotype heterogeneity across studies represents a continuing obstacle that limits successful identification of (replicable) associations[1]. The narrative and systematic literature reviews conducted in chapters 2 and 3 of this dissertation which assess the existing literature on contributions of genes, the environment, and their interactions on tobacco use phenotypes, support this claim. By reviewing the choice of measurement for environmental variables, how main and interaction effects are tested and reported across existing studies, how covariates are treated in analyses, and how gene-by-environment correlations are reported, chapter 3 demonstrates that differences in the approach taken by studies seeking to answer similar research questions make it difficult to interpret and summarize findings.

Harmonization and standardization of phenotypes is especially important in the realm of genome wide association studies, like those described in chapters 9 and 10. As explained elsewhere, phenotype harmonization requires identifying common phenotypes, determining the feasibility of cross-study analysis for each, and preparing common definitions[1]. These concerns are not limited to epidemiological studies, but extend to genetic studies as well. Although considerable time and effort has gone into reducing genotype measurement error and ensuring accuracy and consistency of results[2], phenotype heterogeneity still represents a huge challenge to successful GWAS analyses of complex traits and replication studies. Despite continual efforts, some differences are inevitable[3] due to differences in study design, what data is collected, and how data was collected. The implications of developing better strategies for harmonization and standardization of phenotypes are many; increasing sample size (and consequently, greater power to detect effects), allows for the optimization of power for discovering new associations[3].

Utility of twin analyses in determining the genetic architecture of traits

Twin studies use correlations between pairs of relatives to parses the individual differences in a trait to latent (unmeasured) genetic and environmental influences, allowing for the estimation of heritability or the percentage of variance due to genetic influences. In the context of tobacco use behaviors, twin studies have been useful in determining the heritability of a range of phenotypes, including those examined in chapters 4 and 5: namely, smoking initiation and current quantity smoked. Results from the study conducted in chapter 4 demonstrate that various factors contribute to smoking initiation and current quantity smoked across mid-adolescence into early adulthood, such

that smoking initiation and current quantity smoked have independent liabilities until adulthood when liabilities are shared. This change might be attributed to access and availability to cigarettes, such that as access and availability of cigarettes increase, the expression of genetic predispositions towards increased smoking frequency and potential addiction may also increase, after initiation.

Results from chapter 5 provide evidence for shared genetic and environmental liability for the association between stressful life events and smoking initiation and that this structure differs in early adolescence and young adulthood, but not by sex. This suggests that the same genes and environments are influencing both stressful life events and smoking initiation in males and females, such that the underlying factors influencing stressful life events and smoking initiation are influenced by partly shared environmental factors in young adolescents and by partly shared genetic risk factors in young adulthood. This might suggest that once an individual is exposed to the effects of stressful life events, genetic factors come into play, and that only individuals with a certain set of genes will choose to initiate tobacco use.

The utility of studies is found in the fact that twin models allow for the testing of various hypotheses regarding the genetic architecture of traits, while also considering the potential influence of the environment. Twin studies such as the one conducted in chapter 5, demonstrate that both genetic and environmental factors can have different effects on tobacco-related phenotypes, such as smoking initiation, across time. For example, there exists little evidence currently of common genetic effects that influence both initiation and persistence, implying that there are genetic processes contributing to experimentation and the initiation of tobacco use that are distinct from those influencing the development

and maintenance of tobacco use patterns. Although these studies do not provide information about what specific genes or environments are involved with these processes, they are useful in generating hypotheses for future studies.

Clinical utility of (aggregated) measures of genetic risk

The genetic analyses presented in chapters 9 and 10 demonstrate that genetic factors are involved with the liability of tobacco use behaviors among university students, at both the level of individual variants and at the aggregate level. Estimates of SNP-based heritability using genome wide complex trait analyses (GCTA)[4] indicate that tobacco use behaviors are moderately heritable, and lower than those estimated from twin studies. Yet, estimates of heritability were non-zero and significant amongst those of European ancestry, suggesting that sample sizes for other ancestry groups found in Spit for Science were too small to detect significant heritability. At the level of individual variants, no findings from larger meta-analyses could be replicated. It was suggested in chapter 9, and described previously, that the failure to replicate might be attributed to variability of phenotype, inadequate sample size, false positive results, and population specific effects. Given that novel loci were identified in the Spit for Science sample, and no loci was replicated from the TAG Consortium data, we would deduce that either: the genetic architecture of tobacco use behaviors within these two studies are different, or that the current study is underpowered, making it less likely to find significant hits while also making it more likely that the significant hits found are false positives. In either case, these findings contribute to the existing literature by demonstrating the polygenic nature of the phenotypes investigated.

Chapter 10 expands upon the aims of chapter 9 in efforts to find evidence for common genetic risk variation through the aggregation of genetic effects that do not individually achieve significance in large-scale GWAS. In conducting the analyses for chapter 10, polygenic risk scores were constructed and then put into predictive models of tobacco use behavior. Regression analyses indicated that polygenic risk scores were predictive of some tobacco use behavior phenotypes. The analyses from this chapter also indicate that environmental variables, as well as interactions between polygenic risk and environmental variables (e.g. parental autonomy granting, parental involvement, physical abuse prior to university enrollment) also contribute to the variability in tobacco use behavior phenotypes.

Although polygenic risk scores can be useful for examining the cumulative predictive ability of genetic variation of a trait, there is limited clinical utility for these scores beyond risk prediction due to concerns regarding predictive accuracy, as well as the cost and ability of clinicians and patients to effectively use this information[5]. In addition to varied predictive power based upon methodological approach, we do not yet have enough information regarding specific variants contributing to tobacco use behaviors for clinical risk prediction. Expanding the number of replicated variants associated with tobacco use behaviors, such as nicotine dependence, would improve risk prediction models. However, a prediction tool based upon genetic information alone would probably not be sufficient for clinical use, especially since environmental factors have been found to account for a considerable portion of variability in tobacco use behaviors. Additionally, many of the previous studies have only been conducted within those of European ancestry and still need to be tested within other ancestry groups – especially since the genetic architecture

of tobacco-related phenotypes may differ across populations (and ages). Thus, more research needs to be done to identify genetic factors associated with subclinical phenotypes, especially among non-European ancestry groups[6].

Public health planning and future directions

As explained in chapter 6, despite public health successes in reducing the consumption of cigarettes, the increasing popularity of alternative tobacco products and nicotine delivery systems poses new challenges for researchers and policymakers alike. This is especially true, given that these products contribute to negative outcomes similar to cigarette use and may make it more difficult to quit if used concurrently with cigarettes. As described in chapter 11, national survey data collected in 2014 demonstrate that one in four current cigarette smokers use other tobacco products and nicotine delivery systems[7]. Thus, one of the main challenges for public health research and planning is how to address the growing availability and use of alternative tobacco products and nicotine delivery systems. To reduce and/or prevent the use of these products, more data needs to be collected on the use of all tobacco products as they are being developed, rather than solely focusing on cigarette use[8]. Furthermore, a two-pronged approach may be necessary: one focused on how to alter existing social norms and perceptions regarding alternative tobacco products and nicotine delivery systems and one focused on the reduction of harm caused by products currently and/or will be brought to market. One such direction to take is to educate young adults – a population with the highest prevalence of substance use and the greatest likelihood to quit following intervention – on genetic and environmental risk factors contributing to certain tobacco use behaviors and nicotine dependence, as outlined in chapter 12.

Another approach would be to consider what we know about the correlates and predictors of tobacco use behaviors, including nicotine dependence, and identify potential areas of prevention and intervention. For example, from chapter 7, we know that initial reactions to tobacco differ by tobacco type, and by sex and that age of onset, sex, and positive initial experiences predict both recent use and meeting criteria for nicotine dependence. Thus, it would be beneficial to conduct further research to identify genetic and biological pathways influencing initial experiences with nicotine and the social contexts that influence initial experiences with tobacco use in efforts to delay the overall age of onset for tobacco use and reduce individual risk for nicotine dependence. The information gained from such studies can be used to characterize motivating factors for (alternative) tobacco product use and use reduction within young adults and inform the planning and implementation of future studies.

LIMITATIONS OF THE CURRENT SET OF STUDIES

Like all other studies, this dissertation is not free from limitations. To some extent, these limitations are attributed to study design, which are described in greater detail below.

Concerns regarding power to detect effects of genetic factors influencing tobacco

Power to detect effects is influenced by the size of the effect and the size of the sample used to detect it. In the case where prevalence is low, like in chapter 4 where the prevalence of smoking behavior among early adolescents is low, the power to detect effects is also low. The issue of power is also of importance in the genomic studies conducted in chapters 9 and 10. Previous studies investigating the genetic architecture of tobacco use behaviors that have reported significant SNP associations with tobacco

use phenotypes report much larger sample sizes than those found in the current study. This is a concern since low-powered studies tend to produce more false negatives relative to higher-powered studies and have reduced probabilities of observing effects that pass the required threshold of significance. And, when true effects are found within low powered studies, it is likely that the estimate of the magnitude of the effect is exaggerated. However, after conducting power analyses, it was determined that we *did* have sufficient power (e.g. >80%) to detect aggregate effects of all SNPs, using GCTA – but only for sample sizes of >1500 individuals.

Interpretation of heritability

GCTA yielded significant, non-zero estimates for SNP-based heritability (at p-value ≤0.05), but only amongst those of European ancestry for ever tobacco use and age of initiation. However, once we corrected for multiple testing, these effects were no longer significant. This suggests that the total variance explained by all SNPs is zero; however, this does not necessarily mean that genes are unimportant for the specific traits studied. Rather, it might suggest that the genetic markers used in these studies might not be able to explain the existing phenotypic differences in the population that we are examining or that factors, other than the genetic variants being investigated contribute more to the variation in phenotypic differences.

Inability to infer causality

Due to cross-sectional nature, none of the current studies can infer causality. To get around this issue, the exposure time for the experience of stressful life events, parental

autonomy granting, and parental involvement were limited to prior to university enrollment (chapter 6).

Self-reported data and the potential for recall bias

Across each of the studies, participants were asked to answer questions about their experiences in the past. For example, students were asked to retrospectively report their parental environment and stressful life events prior to university enrollment (chapter 6). Since recall of information is solely dependent on memory, self-reports may be imperfect and potentially unreliable. It is possible that using self-reported data underestimates the prevalence of smoking behaviors across each of the studies included in this dissertation, because of social desirability bias, which has downstream effects on genetic analyses.

Variability in phenotype definitions. Since measurement protocols differ across studies, it is inevitable that there will be variability in phenotype definitions. In efforts to address this problem, we attempted harmonize phenotypes in related analyses, and make measures as comparable as possible.

Generalizability of findings. Many of the studies included in this dissertation have focused on samples of White/Caucasian Americans and Black/African Americans due to concerns regarding sample size and power to detect effects in other race/ethnicity and ancestral groups. Thus, more research needs to be done to assess whether the findings of these studies are generalizable to other groups of young adults who do not fall within these race/ethnicities or ancestral groups. These concerns are also shared across genetic studies, since disparate patterns of linkage disequilibrium and differences in marker allele frequencies between discovery and target samples could attenuate genetic

effects. This is a critical point, since our genomic analyses are conducted on an ethnically-diverse population that is very different from a majority of genomic studies conducted on individuals of European descent.

CLOSING REMARKS

Despite these limitations, the findings from this dissertation contribute to the literature by providing a better comprehensive understanding of how genes, the environment, and their potential interactions influence many tobacco use behavior phenotypes. A key strength of this dissertation project is the variety in the methodologies explored in untangling the influences of genes and the environment on tobacco use behaviors in young adulthood – an understudied population – and the amount of training opportunities afforded from the conducting analyses for and writing up this dissertation.

The studies included in this project focused primarily on the transition from adolescence to young adulthood, to capture the critical period of change during the lifespan and a time at which the use and abuse of substances peak. Given that tobacco use among young adults is both prevalent and associated with several adverse consequences, it is important to understand factors contributing to tobacco use behaviors. This dissertation accomplishes this through multiple methodologies, including: reviews of existing literature on genes, environment, and tobacco use; twin studies of genetic and environmental influences on tobacco use behavior phenotypes; epidemiological studies of prevalence, correlates, and predictors of tobacco use behaviors; genomic analyses of tobacco use behaviors; a commentary on the emergence of alternative nicotine delivery systems and its public health impacts; and, plans for an internet-based educational intervention seeking to reduce tobacco use (and nicotine dependence) by providing students

attending university with information on genetic and environmental risk factors for nicotine dependence.

Of course, although the work accomplished in this set of studies is extensive, it is far from exhaustive. With the changing climate of tobacco use behaviors and available products, further research is needed to not only uncover new environmental risk factors contributing to tobacco use in a broader sense, but also to investigate more deeply the genetic and environmental influences on tobacco use that are already known – which I hope comes across clearly from reading this dissertation thesis.

LIST OF REFERENCES

Chapter 1: Global Introduction

1. Centers for Disease Control and Prevention. Current cigarette smoking among adults - United States 2005-2014. *Morb. Mortal. Wkly. Rep.* **64,** 1233–1240 (2015).

2. U.S. Department of Health and Human Services. *The health consequences of smoking - 50 years of progress: A report of the Surgeon General*. (2014).

3. Centers for Disease Control and Prevention. Smoking-attributable mortality, years of potential life lost, and productivity losses - United States, 2000-2004. *Morb. Mortal. Wkly. Rep.* **57,** 1226–1228 (2008).

4. Centers for Disease Control and Prevention. Current cigarette smoking among adults - United States, 2011. *Morb. Mortal. Wkly. Rep.* **61,** 889–894 (2012).

5. Hu, S. S. *et al.* Tobacco product use among adults - United States, 2013-2014. *Morb. Mortal. Wkly. Rep.* **65,** 685–691 (2016).

6. Lee, Y. O., Herbert, C. J., Nonnemaker, J. M. & Kim, A. E. Multiple tobacco product use among adults in the United States: Cigarettes, cigars, electronic cigarettes, hookah, smokeless tobacco, and snus. *Prev. Med.* **62,** 14–19 (2014).

7. McMillen, R., Maduka, J. & Winickoff, J. Use of emerging tobacco products in the United States. *J. Environ. Public Health* **2012,** 1–8 (2012).

8. Soneji, S., Sargent, J. & Tanski, S. Multiple tobacco product use among US adolescents and young adults. *Tob. Control* **25,** 174–180 (2016).

9. Doll, R., Peto, R., Boreham, J. & Sutherland, I. Mortality in relation to smoking: 50 years' observations on male British doctors. *Br. Med. J.* **328,** 1519–1528 (2004).

10. Chassin, L., Presson, C. C., Rose, J. S. & Sherman, S. J. The natural history of cigarette smoking from adolescence to adulthood: demographic predictors of continuity and change. *Health Psychol.* **15,** 478–484 (1996).

11. Lantz, P. M. Smoking on the rise among young adults: implications for research and policy. *Tob. Control* **12,** i60–i70 (2003).

12. Hammond, D. Smoking behavior among young adults: beyond youth prevention. *Tob. Control* **14,** 181–185 (2005).

13. Kaplan, C. P., Napoles-Springer, A., Stewart, S. L. & Perez-Stable, E. J. Smoking acquisition among adolescents and young Latinas: the role of socioenvironmental and personal factors. *Addict. Behav.* **26,** 531–550 (2001).

14. Bares, C. B., Kendler, K. S. & Maes, H. H. M. Racial differences in heritability of cigarette smoking in adolescents and young adults. *Drug Alcohol Depend.* **166,** 75–84 (2016).

Chapter 2: Genes, Environment, and Tobacco Use

1. Xu, X., Bishop, E. E., Kennedy, S. M., Simpson, S. A. & Pechacek, T. F. Annual healthcare spending attributable to cigarette smoking: an update. *Am. J. Prev. Med.* **48,** 326–333 (2014).

2. Hines, L. A., Morley, K. I., Mackie, C. & Lynskey, M. Genetic and environmental interplay in adolescent substance use disorders. *Curr. Addict. Rep.* **2,** 122–129 (2015).

3. Kendler, K. S. & Eaves, L. J. Models for the joint effect of genotype and environment on liability to psychiatric illness. *Am. J. Psychiatry* **143,** 279–289 (1986).

4. Rasmussen, S. R., Prescott, E., Sorensen, T. I. & Sogaard, J. The total lifetime health cost savings of smoking cessation to society. *Eur. J. Public Health* **15,** 601–606 (2005).

5. Dick, D. M. Gene-environment interaction in psychological traits and disorders. *Annu. Rev. Clin. Psychol.* **7,** 383–409 (2011).

6. Young-Wolff, K. C., Enoch, M. & Prescott, C. A. The influence of gene-environment interactions on alcohol consumption and alcohol use disorders: a comprehensive review. *Clin. Psychol. Rev.* **31,** 800–816 (2011).

7. Timberlake, D. S. *et al.* The moderating effects of religiosity on the genetic and environmental determinants of smoking initiation. *Nicotine Tob. Res.* **8,** 123–133 (2006).

8. Dick, D. M. *et al.* Parental monitoring moderates the importance of genetic and environmental influences on adolescent smoking. *J. Abnorm. Child Psychol.* **116,** 213–218 (2007).

9. Hiemstra, M., Kleinjan, M., Schayck, O. C. P. van, Engels, R. C. M. E. & Otten, R. Environmental Smoking and Smoking Onset in Adolescence: The Role of Dopamine-Related Genes. Findings from Two Longitudinal Studies. *PLOS ONE* **9,** e86497 (2014).

10. Hiemstra, M., Engels, R. C. M. E., Barker, E. D., Schayck, O. C. P. van & Otten, R. Smoking-Specific Parenting and Smoking Onset in Adolescence: The Role of Genes from the Dopaminergic System (DRD2, DRD4, DAT1 Genotypes). *PLOS ONE* **8,** e61673 (2013).

11. Tobacco and Genetics Consortium. Genome-wide meta-analysis identify multiple loci associated with smoking behavior. *Nat. Genet.* **42,** 441–447 (2010).

12. Fagerström, K. The epidemiology of smoking: health consequences and benefits of cessation. *Drugs* **62 Suppl 2,** 1–9 (2002).

13. Chassin, L., Curran, P. J., Presson, C. C., Sherman, S. J. & Wirth, R. J. in *Phenotypes and endophenotypes: foundations for genetic studies of nicotine use and dependence* (eds. Swan, G. E. et al.) (US Department of Health and Human Services, National Institutes of Health, National Cancer Institute, 2009).

14. Piper, M. E. *et al.* Refining the tobacco dependence phenotype using the Wisconsin Inventory of Smoking Dependence Motives. *J. Abnorm. Psychol.* **117,** 747–761 (2008).

15. Hall, W., Madden, P. & Lynskey, M. The genetics of tobacco use: methods, findings and policy implications. *Tob. Control* **11,** 119–124 (2002).

16. Sullivan, P. F. & Kendler, K. S. The genetic epidemiology of smoking. *Nicotine Tob. Res.* **1,** S51–S57 (1999).

17. Maes, H. H. *et al.* A twin study of genetic and environmental influences on tobacco initiation, regular tobacco use, and nicotine dependence. *Psychol. Med.* **34,** 1251–1261 (2004).

18. Li, M. D., Cheng, R., Ma, J. Z. & Swan, G. E. A meta-analysis of estimated genetic and environmental effects on smoking behavior in male and female adult twins. *Addict. Abingdon Engl.* **98,** 23–31 (2003).

19. Schnoll, R. A., Johnson, T. A. & Lerman, C. Genetics and smoking behavior. *Curr. Psychiatry Rep.* **9,** 349–357 (2007).

20. Kaprio, J. Genetic epidemiology of smoking behavior and nicotine dependence. *J. Chronic Obstr. Pulm. Dis.* **6,** 304–306 (2009).

21. Kendler, K. S., Schmitt, E., Aggen, S. H. & Prescott, C. A. Genetic and environmental influences on alcohol, caffeine, cannabis, and nicotine use from early adolescence to middle adulthood. *Arch Gen Psychiatry* **65,** 674–682 (2008).

22. Hartz, S. M. *et al.* Increased genetic vulnerability to smoking at CHRNA5 in early-onset smokers. *Arch. Gen. Psychiatry* **69,** 854–860 (2012).

23. Kendler, K. S. Psychiatric genetics: a methodologic critique. *Am. J. Psychiatry* **162,** 3–11 (2005).

24. Hamdani, N., Ades, J. & Gorwood, P. [Heritability and candidate genes in tobacco use]. *L'Encephale* **32,** 966–975 (2006).

25. Caporaso, N. *et al.* Genome-wide and candidate gene association of cigarette smoking behaviors. *PLoS ONE* **4,** e4653 (2009).

26. Loukola, A., Hallfors, J., Korhonen, T. & Kaprio, J. Genetics and smoking. *Curr. Addict. Rep.* **1,** 75–82 (2014).

27. Chen, X. *et al.* Variants in nicotinic acetylcholine receptors alpha5 and alpha3 increase risks to nicotine dependence. *Am. J. Med. Genet. Neuropsychiatr. Genet.* **150,** 926–933 (2009).

28. Stevens, V. L. *et al.* Nicotinic receptor gene variants influence susceptibility to heavy smoking. *Cancer Epidemiol. Biomark. Prev. Publ. Am. Assoc. Cancer Res. Cosponsored Am. Soc. Prev. Oncol.* **17,** 3517–3525 (2008).

29. Saccone, S. F., Hinnchs, A. L. & Saccone, N. L. Cholinergic nicotinic receptor genes implicated in nicotine dependence association study targeting 348 candidate genes with 3713 SNPs. *Hum. Mol. Genet.* **16,** 36–49 (2007).

30. Rossing, M. A. Genetic influences on smoking: candidate genes. *Environ. Health Perspect.* **106,** 231–238 (1998).

31. Weiss, R. B. *et al.* A candidate gene approach identifies the CHRNA5-A3-B4 region as a risk factor for age-dependent nicotine addiction. *PLoS Genet.* **4,** e1000125 (2008).

32. Chen, L., Johnson, E. O. & Breslau, N. Interplay of genetic risk factors and parent monitoring in risk for nicotine dependence. *Addiction* **104,** 1731–1740 (2009).

33. Chen, L.-S. *et al.* Interplay of Genetic Risk Factors (CHRNA5-CHRNA3-CHRNB4) and Cessation Treatments in Smoking Cessation Success. *Am. J. Psychiatry* **169,** 735–742 (2012).

34. Ducci, F., Kaakinen, M. & Pouta, A. TTC12-ANKK1-DRD2 and CHRNA5-CHRNA3-CHRNB4 influence different pathways leading to smoking behavior from adolescence to mid-adulthood. *Biol. Psychiatry* **69,** 650–660 (2011).

35. Thorgeirsson, T. E., Geller, F. & Sulem, P. A variant associated with nicotine dependence, lung cancer and peripheral arterial disease. *Nature* **452,** 638–642 (2008).

36. Hong, L. E. *et al.* A CHRNA5 allele related to nicotine addiction and schizophrenia. *Genes Brain Behav.* **10,** 530–535 (2011).

37. Conlon, M. S. & Bewick, M. A. Single nucleotide polymorphisms in CHRNA5 rs16969968, CHRNA3 rs578776, and LOC123688 rs8034191 are associated with heaviness of smoking in women in Northeastern Ontario, Canada. *Nicotine Tob. Res. Off. J. Soc. Res. Nicotine Tob.* **13,** 1076–1083 (2011).

38. Daw, J. *et al.* Genetic sensitivity to peer behaviors: 5HTTLPR, smoking, and alcohol consumption. *J. Health Soc. Behav.* **54,** (2013).

39. Ohmoto, M., Hirakoshi, M. & Mitsumoto, Y. Effects of moderating factors including serotonin transporter polymorphisms on smoking behavior: a systematic review and meta-analysis update. *Nicotine Tob. Res. Off. J. Soc. Res. Nicotine Tob.* **15,** 572–582 (2013).

40. Evangelou, E. & Ioannidis, J. P. A. Meta-analysis methods for genome-wide association studies and beyond. *Nat. Rev. Genet.* **14,** 379–389 (2013).

41. Kaprio, J. Twins and the mystery of missing heritability: the contribution of gene-environment interactions. *J. Intern. Med.* **272,** 440–448 (2012).

42. Audrain, J. *et al.* Genetic susceptibility testing in smoking-cessation treatment: one-year outcomes of a randomized trial. *Addict. Behav.* **22,** 741–751 (1997).

43. Kendler, K. S. *et al.* Recent advances in the genetic epidemiology and molecular genetics of substance use disorders. *Nat. Neurosci.* **15,** 181–189 (2012).

44. Lerman, C. & Niaura, R. Applying genetic approaches to the treatment of nicotine dependence. *Oncogene* **21,** 7412–7420 (2002).

45. Hiscock, R., Bauld, L., Amos, A., Fidler, J. A. & Munafò, M. Socioeconomic status and smoking: a review. *Ann. N. Y. Acad. Sci.* **1248,** 107–123 (2012).

46. Royal College of Physicians. *Nicotine addiction in Britain: a report of the tobacco advisory group of the Royal College of Physicians*. (2000).

47. Senol, Y., Donmez, L., Turkay, M. & Aktekin, M. The incidence of smoking and risk factors for smoking initiation in medical faculty students: cohort study. *BMC Public Health* **6,** 128 (2006).

48. Botvin, G. J., Epstein, J. A., Schinke, S. P. & Diaz, T. Predictors of cigarette smoking among inner-city minority youth. *J. Dev. Behav. Pediatr. JDBP* **15,** 67–73 (1994).

49. Substance Abuse and Mental Health Services Administration. *Results from the 2010 national survey on drug use and health: summary of national findings*. (Substance Abuse and Mental Health Services Administration, 2011).

50. Hedman, L. *et al.* Factors related to tobacco use among teenagers. *Respir. Med.* **101,** 496–502 (2007).

51. Okoli, C., Greaves, L. & Fagyas, V. Sex differences in smoking initiation among children and adolescents. *Public Health* **127,** 3–10 (2013).

52. Chen, P. & Jacobson, K. C. Developmental trajectories of substance use from early adolescence to young adulthood: gender and racial/ethnic differences. *J. Adolesc. Health Off. Publ. Soc. Adolesc. Med.* **50,** 154–163 (2012).

53. in *Psychiatry* (eds. Tasman, A., Kay, J., Lieberman, J. A., First, M. B. & Maj, M.) (John Wiley & Sons).

54. Hill, K. G., Hawkins, J. D., Catalano, R. F., Abbott, R. D. & Guo, J. Family influences on the risk of daily smoking initiation. *J. Adolesc. Health* **37,** 202–210 (2005).

55. Avenevoli, S. & Merikangas, K. R. Familial influences on adolescent smoking. *Addict. Abingdon Engl.* **98 Suppl 1,** 1–20 (2003).

56. Mayhew, K. P., Flay, B. R. & Mott, J. A. Stages in the development of adolescent smoking. *Drug Alcohol Depend.* **59 Suppl 1,** S61-81 (2000).

57. Tilson, E. C., McBride, C. M., Lipkus, I. M. & Catalano, R. F. Testing the interaction between parent–child relationship factors and parent smoking to predict youth smoking. *J. Adolesc. Health* **35,** 182–189 (2004).

58. Flay, B. R., Phil., D., Hu, F. B. & Richardson, J. Psychosocial Predictors of Different Stages of Cigarette Smoking among High School Students. *Prev. Med.* **27,** A9–A18 (1998).

59. Biglan, A., Duncan, T. E., Ary, D. V. & Smolkowski, K. Peer and parental influences on adolescent tobacco use. *J. Behav. Med.* **18,** 315–330 (1995).

60. Chassin, L. *et al.* Parenting style and smoking-specific parenting practices as predictors of adolescent smoking onset. *J. Pediatr. Psychol.* **30,** 333–344 (2005).

61. Simons-Morton, B. G. The protective effect of parental expectations against early adolescent smoking initiation. *Health Educ. Res.* **19,** 561–569 (2004).

62. von Bothmer, M. I. K., Mattsson, B. & Fridlund, B. Influences on adolescent smoking behaviour: siblings' smoking and norms in the social environment do matter. *Health Soc. Care Community* **10,** 213–220 (2002).

63. Castrucci, B. C., Gerlach, K. K., Kaufman, N. J. & Orleans, C. T. The association among adolescents' tobacco use, their beliefs and attitudes, and friends' and parents' opinions of smoking. *Matern. Child Health J.* **6,** 159–167 (2002).

64. Gritz, E. R. *et al.* Cigarette Smoking in a Multiethnic Population of Youth: Methods and Baseline Findings. *Prev. Med.* **27,** 365–384 (1998).

65. Sargent, J. D. & Dalton, M. Does parental disapproval of smoking prevent adolescents from becoming established smokers? *Pediatrics* **108,** 1256–1262 (2001).

66. Bricker, J. B. *et al.* Prospective prediction of children's smoking transitions: role of parents' and older siblings' smoking. *Addiction* **101,** 128–136 (2006).

67. Li, C., Pentz, M. A. & Chou, C.-P. Parental substance use as a modifier of adolescent substance use risk. *Addict. Abingdon Engl.* **97,** 1537–1550 (2002).

68. Slomkowski, C., Rende, R., Novak, S., Lloyd-Richardson, E. & Niaura, R. Sibling effects on smoking in adolescence: evidence for social influence from a genetically informative design. *Addict. Abingdon Engl.* **100,** 430–438 (2005).

69. Simons-Morton, B. G. & Farhat, T. Recent findings on peer group influences on adolescent smoking. *J. Prim. Prev.* **31,** 191–208 (2010).

70. Björkqvist, K., Båtman, A. & Aman-Back, S. Adolescents' use of tobacco and alcohol: correlations with habits of parents and friends. *Psychol. Rep.* **95,** 418–420 (2004).

71. West, P., Sweeting, H. & Ecob, R. Family and friends' influences on the uptake of regular smoking from mid-adolescence to early adulthood. *Addict. Abingdon Engl.* **94,** 1397–1411 (1999).

72. de Vries, H., Engels, R., Kremers, S., Wetzels, J. & Mudde, A. Parents' and friends' smoking status as predictors of smoking onset: findings from six European countries. *Health Educ. Res.* **18,** 627–636 (2003).

73. de Vries, H., Candel, M., Engels, R. & Mercken, L. Challenges to the peer influence paradigm: results for 12-13 year olds from six European countries from the European Smoking Prevention Framework Approach study. *Tob. Control* **15,** 83–89 (2006).

74. Kobus, K. Peers and adolescent smoking. *Addict. Abingdon Engl.* **98 Suppl 1,** 37–55 (2003).

75. Bauman, K. E. & Ennett, S. T. On the importance of peer influence for adolescent drug use: commonly neglected considerations. *Addict. Abingdon Engl.* **91,** 185–198 (1996).

76. Leatherdale, S. T. & Manske, S. The relationship between student smoking in the school environment and smoking onset in elementary school students. *Cancer Epidemiol. Biomark. Prev. Publ. Am. Assoc. Cancer Res. Cosponsored Am. Soc. Prev. Oncol.* **14,** 1762–1765 (2005).

77. Simons-Morton, B. G. Prospective analysis of peer and parent influences on smoking initiation among early adolescents. *Prev. Sci. Off. J. Soc. Prev. Res.* **3,** 275–283 (2002).

78. Forster, J., Chen, V., Blaine, T., Perry, C. & Toomey, T. Social exchange of cigarettes by youth. *Tob. Control* **12,** 148–154 (2003).

79. DiFranza, J. R., Savageau, J. A. & Aisquith, B. F. Youth access to tobacco: the effects of age, gender, vending machine locks, and 'it's the law' programs. *Am. J. Public Health* **86,** 221–224 (1996).

80. Centers for Disease Control and Prevention. Current cigarette smoking among adults - United States, 2011. *Morb. Mortal. Wkly. Rep.* **61,** 889–894 (2012).

81. Kendler, K. S., Myers, J., Damaj, M. I. & Chen, X. Early smoking onset and risk for subsequent nicotine dependence: a monozygotic co-twin control study. *Am. J. Psychiatry* **170,** 408–413 (2013).

82. Klein, H., Sterk, C. E. & Elifson, K. W. Initial smoking experiences and current smoking behaviors and perceptions among current smokers. *J. Addict.* **2013,** (2013).

83. Breslau, N. & Peterson, E. L. Smoking cessation in young adults: age at initiation of cigarette smoking and other suspected influences. *Am. J. Public Health* **86,** 214–220 (1996).

84. Britton, J. & Bogdanovica, I. Tobacco control efforts in Europe. *Lancet Lond. Engl.* **381,** 1588–1595 (2013).

85. Ross, H. & Chaloupka, F. J. The effect of cigarette prices on youth smoking. *Health Econ.* **12,** 217–230 (2003).

86. Fichtenberg, C. M. & Glantz, S. A. Effect of smoke-free workplaces on smoking behaviour: systematic review. *BMJ* **325,** 188 (2002).

87. Wilson, L. M. *et al.* Impact of tobacco control interventions on smoking initiation, cessation, and prevalence: a systematic review. *J. Environ. Public Health* **2012,** 961724 (2012).

88. Weaver, A. J., Flannelly, K. J. & Strock, A. L. A review of research on the effects of religion on adolescent tobacco use published between 1990 and 2003. *Adolescence* **40,** 761–776 (2005).

89. Sinha, J. W., Cnaan, R. A. & Gelles, R. J. Adolescent risk behaviors and religion: findings from a national study. *J. Adolesc.* **30,** 231–249 (2007).

90. Nonnemaker, J., McNeely, C. A. & Blum, R. W. Public and private domains of religiosity and adolescent smoking transitions. *Soc. Sci. Med. 1982* **62,** 3084–3095 (2006).

91. Ford, J. A. & Hill, T. D. Religiosity and adolescent substance use: evidence from the national survey on drug use and health. *Subst. Use Misuse* **47,** 787–798 (2012).

92. Institute of Medicine (US) Committee on Assessing Interactions Among Social Behavioral, and Genetic Factors in Health. in *Genetic, behavior, and the social environment: moving beyond the nature/nurture debate* (eds. Hernandez, L. M. & Blazer, D. G.) (National Academies Press, 2006).

93. Lerman, C. E., Schnoll, R. A. & Munafò, M. R. Genetics and smoking cessation improving outcomes in smokers at risk. *Am. J. Prev. Med.* **33,** S398-405 (2007).

94. Boardman, J. D., Saint Onge, J. M., Haberstick, B. C., Timberlake, D. S. & Hewitt, J. K. Do schools moderate the genetic determinants of smoking? *Behav. Genet.* **38,** 234–246 (2008).

95. Boardman, J. D. State-level moderation of genetic tendencies to smoke. *Am. J. Public Health* **99,** 480–486 (2009).

96. Vandenbergh, D. J. *et al.* An Adolescent Substance Prevention Model Blocks the Effect of CHRNA5 Genotype on Smoking During High School. *Nicotine Tob. Res. Off. J. Soc. Res. Nicotine Tob.* **18,** 212–220 (2016).

97. Meyers, J. L. *et al.* Interaction between polygenic risk for cigarette use and environmental exposures in the Detroit neighborhood health study. *Transl. Psychiatry* **3,** e290 (2013).

98. Public Health Law Center at Mitchell Hamline School of Law. Tobacco Control. (2015). Available at: http://publichealthlawcenter.org/topics/tobacco-control. (Accessed: 13th January 2016)

99. Vink, J. M. & Boomsma, D. I. Interplay between heritability of smoking and environmental conditions? A comparison of two birth cohorts. *BMC Public Health* **11,** 316 (2011).

100. Boardman, J. D., Blalock, C. L. & Pampel, F. C. Trends in the Genetic Influences on Smoking. *J. Health Soc. Behav.* **51,** 108–123 (2010).

101. Fletcher, J. M. Why Have Tobacco Control Policies Stalled? Using Genetic Moderation to Examine Policy Impacts. *PLoS ONE* **7,** (2012).

102. Boardman, J. D. *et al.* Population composition, public policy, and the genetics of smoking. *Demography* **48,** 1517–1533 (2011).

103. Hutchison, K. E., LaChance, H., Niaura, R., Bryan, A. & Smolen, A. The DRD4 VNTR polymorphism influences reactivity to smoking cues. *J. Abnorm. Psychol.* **111,** 134–143 (2002).

104. Rose, R. J., Broms, U., Korhonen, T., Dick, D. M. & Kaprio, J. in *Handbook of Behavior Genetics* (ed. Kim) 411–432 (Springer, 2009).

105. Vink, J. M., Willemsen, G. & Boomsma, D. I. Heritability of smoking initiation and nicotine dependence. *Behav. Genet.* **35,** 397–406 (2005).

106. Do, E. K. *et al.* Genetic and environmental influences on smoking behavior across adolescence and young adulthood in the Virginia Twin Study of Adolescent Behavioral Development and the Transitions to Substance Abuse Follow-Up. *Twin Res. Hum. Genet.* **18,** 43–51 (2015).

107. Sullivan, P. F., Jiang, Y., Neale, M. C., Kendler, K. S. & Straub, R. E. Association of the tryptophan hydroxylase gene with smoking initiation but not progression to nicotine dependence. *Am. J. Med. Genet. Neuropsychiatr. Genet.* **105,** 479–484 (2001).

108. Kremer, I., Bachner-Melman, R. & Reshef, A. Association of the serotonin transporter gene with smoking behavior. *Am. J. Psychiatry* **162,** 924–930 (2005).

109. Gu, D. F., Hinks, L. J., Morton, N. E. & Day, I. N. The use of long PCR to confirm three common alleles at the CYP2A6 locus and the relationship between genotype and smoking habit. *Ann. Hum. Genet.* **64,** 383–390 (2000).

110. Pianezza, M. L., Sellers, E. M. & Tyndale, R. F. Nicotine metabolism defect reduces smoking. *Nature* **393,** 750 (1998).

111. Lerman, C. *et al.* Evidence suggesting the role of specific genetic factors in cigarette smoking. *Health Psychol. Off. J. Div. Health Psychol. Am. Psychol. Assoc.* **18,** 14–20 (1999).

112. Kaprio, J., Pulkkinen, L. & Rose, R. J. Genetic and environmental factors in health-related behaviors: studies on Finnish twins and twin families. *Twin Res. Off. J. Int. Soc. Twin Stud.* **5,** 366–371 (2002).

113. Lantz, P. M. *et al.* Investing in youth tobacco control: a review of smoking prevention and control strategies. *Tob. Control* **9,** 47–63 (2000).

Chapter 3: Genotype X Environment Interaction in Smoking Behaviors: A Systematic Review

1. WHO Media Centre. Tobacco. 339 (2015).

2. Rose, R. J., Broms, U., Korhonen, T., Dick, D. M. & Kaprio, J. in *Handbook of Behavior Genetics* (ed. Kim) 411–432 (Springer, 2009).

3. Maes, H. H. *et al.* A twin study of genetic and environmental influences on tobacco initiation, regular tobacco use, and nicotine dependence. *Psychol. Med.* 34, 1251–1261 (2004).

4. Kaprio, J. Genetic epidemiology of smoking behavior and nicotine dependence. *J. Chronic Obstr. Pulm. Dis.* 6, 304–306 (2009).

5. Kendler, K. S., Schmitt, E., Aggen, S. H. & Prescott, C. A. Genetic and environmental influences on alcohol, caffeine, cannabis, and nicotine use from early adolescence to middle adulthood. *Arch Gen Psychiatry* 65, 674–682 (2008).

6. Schnoll, R. A., Johnson, T. A. & Lerman, C. Genetics and smoking behavior. *Curr. Psychiatry Rep.* 9, 349–357 (2007).

7. Lessov-Schlaggar, C. N., Pergadia, M. L., Khroyan, T. V. & Swan, G. E. Genetics of nicotine dependence and pharmacotherapy. *Biochem. Pharmacol.* 75, 178–195 (2008).

8. Uhl, G. R. *et al.* Molecular genetics of nicotine dependence and abstinence: whole genome association using 520,000 SNPs. *BMC Genet.* 8, 10 (2007).

9. Saccone, S. F., Hinnchs, A. L. & Saccone, N. L. Cholinergic nicotinic receptor genes implicated in nicotine dependence association study targeting 348 candidate genes with 3713 SNPs. *Hum. Mol. Genet.* 16, 36–49 (2007).

10. Lou, X.-Y. *et al.* Gene-based analysis suggests association of the nicotinic acetylcholine receptor β1 subunit (CHRNB1) and M1 muscarinic acetylcholine receptor (CHRM1) with vulnerability for nicotine dependence. *Hum. Genet.* 120, 381–389 (2006).

11. Zhang, L., Kendler, K. S. & Chen, X. The μ-opioid receptor gene and smoking initiation and nicotine dependence. *Behav. Brain Funct.* 2, 28 (2006).

12. Fagerström, K. The epidemiology of smoking: health consequences and benefits of cessation. *Drugs* 62 Suppl 2, 1–9 (2002).

13. Royal College of Physicians. *Nicotine addiction in Britain: a report of the tobacco advisory group of the Royal College of Physicians*. (2000).

14. Flay, B. R., Phil., D., Hu, F. B. & Richardson, J. Psychosocial Predictors of Different Stages of Cigarette Smoking among High School Students. *Prev. Med.* 27, A9–A18 (1998).

15. Gritz, E. R. *et al.* Cigarette Smoking in a Multiethnic Population of Youth: Methods and Baseline Findings. *Prev. Med.* 27, 365–384 (1998).

16. Tilson, E. C., McBride, C. M., Lipkus, I. M. & Catalano, R. F. Testing the interaction between parent–child relationship factors and parent smoking to predict youth smoking. *J. Adolesc. Health* 35, 182–189 (2004).

17. Bricker, J. B. *et al.* Prospective prediction of children's smoking transitions: role of parents' and older siblings' smoking. *Addiction* 101, 128–136 (2006).

18. West, P., Sweeting, H. & Ecob, R. Family and friends' influences on the uptake of regular smoking from mid-adolescence to early adulthood. *Addict. Abingdon Engl.* 94, 1397–1411 (1999).

19. Björkqvist, K., Båtman, A. & Aman-Back, S. Adolescents' use of tobacco and alcohol: correlations with habits of parents and friends. *Psychol. Rep.* 95, 418–420 (2004).

20. Meyers, J. L. *et al.* Interaction between polygenic risk for cigarette use and environmental exposures in the Detroit neighborhood health study. *Transl. Psychiatry* 3, e290 (2013).

21. Boardman, J. D. State-level moderation of genetic tendencies to smoke. *Am. J. Public Health* 99, 480–486 (2009).

22. Boardman, J. D., Blalock, C. L. & Pampel, F. C. Trends in the Genetic Influences on Smoking. *J. Health Soc. Behav.* 51, 108–123 (2010).

23. Daw, J. *et al.* Genetic sensitivity to peer behaviors: 5HTTLPR, smoking, and alcohol consumption. *J. Health Soc. Behav.* 54, (2013).

24. Dick, D. M. *et al.* Parental monitoring moderates the importance of genetic and environmental influences on adolescent smoking. *J. Abnorm. Child Psychol.* 116, 213–218 (2007).

25. Vink, J. M., Willemsen, G. & Boomsma, D. I. Heritability of smoking initiation and nicotine dependence. *Behav. Genet.* 35, 397–406 (2005).

26. Ducci, F., Kaakinen, M. & Pouta, A. TTC12-ANKK1-DRD2 and CHRNA5-CHRNA3-CHRNB4 influence different pathways leading to smoking behavior from adolescence to mid-adulthood. *Biol. Psychiatry* 69, 650–660 (2011).

27. Chen, L., Johnson, E. O. & Breslau, N. Interplay of genetic risk factors and parent monitoring in risk for nicotine dependence. *Addiction* 104, 1731–1740 (2009).

28. Boardman, J. D., Saint Onge, J. M., Haberstick, B. C., Timberlake, D. S. & Hewitt, J. K. Do schools moderate the genetic determinants of smoking? *Behav. Genet.* 38, 234–246 (2008).

29. Hiemstra, M., Kleinjan, M., Schayck, O. C. P. van, Engels, R. C. M. E. & Otten, R. Environmental Smoking and Smoking Onset in Adolescence: The Role of Dopamine-Related Genes. Findings from Two Longitudinal Studies. *PLOS ONE* 9, e86497 (2014).

30. Hiemstra, M., Engels, R. C. M. E., Barker, E. D., Schayck, O. C. P. van & Otten, R. Smoking-Specific Parenting and Smoking Onset in Adolescence: The Role of Genes from the Dopaminergic System (DRD2, DRD4, DAT1 Genotypes). *PLOS ONE* 8, e61673 (2013).

31. Belsky, D. W. *et al.* Polygenic risk accelerates the developmental progression to heavy, persistent smoking and nicotine dependence: Evidence from a 4-Decade Longitudinal Study. *JAMA Psychiatry Chic. Ill* 70, 534–542 (2013).

32. Xie, P. *et al.* Childhood Adversity Increases Risk for Nicotine Dependence and Interacts with α5 Nicotinic Acetylcholine Receptor Genotype Specifically in Males. *Neuropsychopharmacology* 37, 669–676 (2012).

33. Fletcher, J. M. Why Have Tobacco Control Policies Stalled? Using Genetic Moderation to Examine Policy Impacts. *PLoS ONE* 7, (2012).

34. Chen, L.-S. *et al.* Interplay of Genetic Risk Factors (CHRNA5-CHRNA3-CHRNB4) and Cessation Treatments in Smoking Cessation Success. *Am. J. Psychiatry* 169, 735–742 (2012).

35. Timberlake, D. S. *et al.* The moderating effects of religiosity on the genetic and environmental determinants of smoking initiation. *Nicotine Tob. Res.* 8, 123–133 (2006).

36. D'Souza, M. S. & Markou, A. Neuronal Mechanisms Underlying Development of Nicotine Dependence: Implications for Novel Smoking-Cessation Treatments. *Addict. Sci. Clin. Pract.* 6, 4–16 (2011).

37. Vink, J. M., Smit, A. B. & de Geus, E. J. C. Genome-wide association study of smoking initiation and current smoking. *Am. J. Hum. Genet.* 84, 367–379 (2009).

38. Belsky, D. W., Moffitt, T. E. & Caspi, A. Genetics in population health science: strategies and opportunities. *Am. J. Public Health* 103 Suppl 1, S73-83 (2013).

39. Duncan, L. E. & Keller, M. C. A Critical Review of the First 10 Years of Candidate Gene-by-Environment Interaction Research in Psychiatry. *Am. J. Psychiatry* 168, 1041–1049 (2011).

40. Duncan, L. E., Pollastri, A. R. & Smoller, J. W. Mind the gap: why many geneticists and psychological scientists have discrepant views about gene-environment interaction (G×E) research. *Am. Psychol.* 69, 249–268 (2014).

41. Meyers, J. L. & Dick, D. M. Genetic and environmental risk factors for adolescent-onset substance use disorders. *Child Adolesc. Psychiatr. Clin. N. Am.* 19, 465–477 (2010).

42. Ottman, R. Gene–Environment Interaction: Definitions and Study Designs. *Prev. Med.* 25, 764–770 (1996).

43. Eaves, L. & Eysenck, H. Genotype x age interaction for neuroticism. *Behav. Genet.* 6, 359–362 (1976).

44. Dempfle, A. *et al.* Gene-environment interactions for complex traits: definitions, methodological requirements and challenges. *Eur. J. Hum. Genet. EJHG* 16, 1164–1172 (2008).

45. Moffitt, T. E., Caspi, A. & Rutter, M. Measured Gene-Environment Interactions in Psychopathology: Concepts, Research Strategies, and Implications for Research, Intervention, and Public Understanding of Genetics. *Perspect. Psychol. Sci. J. Assoc. Psychol. Sci.* 1, 5–27 (2006).

46. Hartz, S. M. *et al.* Increased genetic vulnerability to smoking at CHRNA5 in early-onset smokers. *Arch. Gen. Psychiatry* 69, 854–860 (2012).

47. Plomin, R., DeFries, J. C. & Loehlin, J. C. Genotype-environment interaction and correlation in the analysis of human behavior. *Psychol. Bull.* 84, 309–322 (1977).

48. Purcell, S. Variance components models for gene-environment interaction in twin analysis. *Twin Res. Off. J. Int. Soc. Twin Stud.* 5, 554–571 (2002).

Chapter 4: Genetic and Environmental Influences on Smoking Behavior across Adolescence and Young Adulthood in the Virginia Twin Study of Adolescent Behavioral Development and the Transitions to Substance Abuse Follow-Up

1. Centers for Disease Control and Prevention. Smoking-attributable mortality, years of potential life lost, and productivity losses - United States, 2000-2004. *Morb. Mortal. Wkly. Rep.* **57,** 1226–1228 (2008).

2. Centers for Disease Control and Prevention. Current cigarette smoking among adults - United States, 2011. *Morb. Mortal. Wkly. Rep.* **61,** 889–894 (2014).

3. U.S. Department of Health and Human Services. *Preventing tobacco use among young people: surgeon general's report*. (U.S. Department of Health and Human Services, 1994).

4. Substance Abuse and Mental Health Services Administration. *Results from the 2008 national survey on drug use and health*. (Substance Abuse and Mental Health Services Administration, 2014).

5. Centers for Disease Control and Prevention. *State data highlights, 2006*. (Office on Smoking and Health, 2008).

6. Koopmans, J. R., Slutske, W. S., Heath, A. C., Neale, M. C. & Boomsma, D. I. The Genetics of Smoking Initiation and Quantity Smoked in Dutch Adolescent and Young Adult Twins. *Behav. Genet.* **29,** 383–393 (1999).

7. Lyons, M. *et al.* A Twin Study of Smoking, Nicotine Dependence, and Major Depression in Men. *Nicotine Tob. Res.* **10,** 97–108 (2008).

8. Madden, P. A. *et al.* The genetics of smoking persistence in men and women: a multicultural study. *Behav. Genet.* **29,** 423–431 (1999).

9. Koopmans, J. R., van Doornen, L. J. & Boomsma, D. I. Association between alcohol use and smoking in adolescent and young adult twins: a bivariate genetic analysis. *Alcohol. Clin. Exp. Res.* **21,** 537–546 (1997).

10. Slomkowski, C., Rende, R., Novak, S., Lloyd-Richardson, E. & Niaura, R. Sibling effects on smoking in adolescence: evidence for social influence from a genetically informative design. *Addict. Abingdon Engl.* **100,** 430–438 (2005).

11. Maes, H. H. *et al.* Tobacco, alcohol and drug use in eight- to sixteen-year-old twins: the Virginia Twin Study of Adolescent Behavioral Development. *J. Stud. Alcohol* **60,** 293–305 (1999).

12. Vink, J. M., Willemsen, G. & Boomsma, D. I. Heritability of smoking initiation and nicotine dependence. *Behav. Genet.* **35,** 397–406 (2005).

13. Kendler, K. S., Schmitt, E., Aggen, S. H. & Prescott, C. A. Genetic and environmental influences on alcohol, caffeine, cannabis, and nicotine use from early adolescence to middle adulthood. *Arch Gen Psychiatry* **65,** 674–682 (2008).

14. Karp, I., O'Loughlin, J., Paradis, G., Handley, J. & DiFranza, J. Smoking trajectories of adolescent novice smokers in a longitudinal study of tobacco use. *Ann. Epidemiol.* **15,** 445–452 (2005).

15. Edwards, K. L., Austin, M. A. & Jarvik, G. P. Evidence for genetic influences on smoking in adult women twins. *Clin. Genet.* **47,** 236–244 (1995).

16. Heath, A. C. *et al.* Genetic contribution to risk of smoking initiation: comparisons across birth cohorts and across cultures. *J. Subst. Abuse* **5,** 221–246 (1993).

17. Heath, A. C., Kirk, K. M., Meyer, J. M. & Martin, N. G. Genetic and social determinants of initiation and age at onset of smoking in Australian twins. *Behav. Genet.* **29,** 395–407 (1999).

18. Kendler, K. S. *et al.* A population-based twin study in women of smoking initiation and nicotine dependence. *Psychol. Med.* **29,** 299–308 (1999).

19. True, W. R. *et al.* Genetic and environmental contributions to smoking. *Addict. Abingdon Engl.* **92,** 1277–1287 (1997).

20. Broms, U., Silventoinen, K., Madden, P. A. F., Heath, A. C. & Kaprio, J. Genetic architecture of smoking behavior: a study of Finnish adult twins. *Twin Res. Hum. Genet. Off. J. Int. Soc. Twin Stud.* **9,** 64–72 (2006).

21. Zavos, H. M. S. *et al.* Genetic and environmental etiology of nicotine use in Sri Lankan male twins. *Behav. Genet.* **42,** 798–807 (2012).

22. Heath, A. C., Martin, N. G., Lynskey, M. T., Todorov, A. A. & Madden, P. A. F. Estimating two-stage models for genetic influences on alcohol, tobacco or drug use initiation and dependence vulnerability in twin and family data. *Twin Res. Off. J. Int. Soc. Twin Stud.* **5,** 113–124 (2002).

23. Li, M. D., Cheng, R., Ma, J. Z. & Swan, G. E. A meta-analysis of estimated genetic and environmental effects on smoking behavior in male and female adult twins. *Addict. Abingdon Engl.* **98,** 23–31 (2003).

24. Maes, H. H. *et al.* A twin study of genetic and environmental influences on tobacco initiation, regular tobacco use, and nicotine dependence. *Psychol. Med.* **34,** 1251–1261 (2004).

25. Gillespie, N. A., Neale, M. C. & Kendler, K. S. Pathways to cannabis abuse: a multi-stage model from cannabis availability, cannabis initiation and progression to abuse. *Addict. Abingdon Engl.* **104,** 430–438 (2009).

26. Fowler, T. *et al.* Exploring the relationship between genetic and environmental influences on initiation and progression of substance use. *Addict. Abingdon Engl.* **102,** 413–422 (2007).

27. Kendler, K. S., Gardner, C., Jacobson, K. C., Neale, M. C. & Prescott, C. A. Genetic and environmental influences on illicit drug use and tobacco use across birth cohorts. *Psychol. Med.* **35,** 1349–1356 (2005).

28. Meyer, J. M., Silberg, J. L., Simonoff, E., Kendler, K. S. & Hewitt, J. K. The Virginia Twin-Family Study of Adolescent Behavioral Development: assessing sample biases in demographic correlates of psychopathology. *Psychol. Med.* **26,** 1119–1133 (1996).

29. Hewitt, J. K. *et al.* Genetics and developmental psychopathology: 1. Phenotypic assessment in the Virginia Twin Study of Adolescent Behavioral Development. *J. Child Psychol. Psychiatry* **38,** 943–963 (1997).

30. Boker, S. *et al.* OpenMx: An Open Source Extended Structural Equation Modeling Framework. *Psychometrika* **76,** 306–317 (2011).

31. Neale, M. C., Boker, S. M., Xie, G. & Maes, H. H. *Statistical modeling*. (Virginia Commonwealth University, 2003).

32. Neale, M. C., Harvey, E., Maes, H. H. M., Sullivan, P. F. & Kendler, K. S. Extensions to the modeling of initiation and progression: applications to substance use and abuse. *Behav. Genet.* **36,** 507–524 (2006).

33. Agrawal, A., Neale, M. C., Jacobson, K. C., Prescott, C. A. & Kendler, K. S. Illicit drug use and abuse/dependence: modeling of two-stage variables using the CCC approach. *Addict. Behav.* **30,** 1043–1048 (2005).

34. Vuong, Q. H. Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses. *Econometrica* **57,** 307–333 (1989).

35. Neyman, J. & Pearson, E. S. On the Use and Interpretation of Certain Test Criteria for Purposes of Statistical Inference: Part I. *Biometrika* **20A,** 175–240 (1928).

36. Akaike, H. Factor analysis and AIC. *Psychometrika* **52,** 317–332 (1987).

37. Williams, L. J. & Holahan, P. J. Parsimony-based fit indices for multiple-indicator models: Do they work? *Struct. Equ. Model. Multidiscip. J.* **1,** 161–189 (1994).

Chapter 5: A Twin Study of the Genetic and Environmental Relationship of Stressful Life Events and Smoking Initiation using the Virginia Twin Studies of Adolescent Behavioral Development

1. U.S. Department of Health and Human Services. *The health consequences of smoking - 50 years of progress: A report of the Surgeon General*. (2014).

2. Brandon, T. H. Negative Affect as Motivation to Smoke. *Curr. Dir. Psychol. Sci.* **3,** 33–37 (1994).

3. Wills, T. A. & Shiffman, S. in *Coping and substance use* 3–24 (Academic Press, 1985).

4. Mates, D. & Allison, K. R. Sources of stress and coping responses of high school students. *Adolescence* **27,** 461–474 (1992).

5. Fiore, M. C. Trends in cigarette smoking in the United States. The epidemiology of tobacco use. *Med. Clin. North Am.* **76,** 289–303 (1992).

6. Koval, J. J., Pederson, L. L., Mills, C. A., McGrady, G. A. & Carvajal, S. C. Models of the relationship of stress, depression, and other psychosocial factors to smoking behavior: a comparison of a cohort of students in grades 6 and 8. *Prev. Med.* **30,** 463–477 (2000).

7. Bonaguro, J. A. & Bonaguro, E. W. Self-concept, stress symptomatology, and tobacco use. *J. Sch. Health* **57,** 56–58 (1987).

8. McKee, S. A., Maciejewski, P. K., Falba, T. & Mazure, C. M. Sex differences in the effects of stressful life events on changes in smoking status. *Addict. Abingdon Engl.* **98,** 847–855 (2003).

9. Kassel, J. D., Stroud, L. R. & Paronis, C. A. Smoking, stress, and negative affect: correlation, causation, and context across stages of smoking. *Psychol. Bull.* **129,** 270–304 (2003).

10. Booker, C. L., Gallaher, P., Unger, J. B., Ritt-Olson, A. & Johnson, C. A. Stressful life events, smoking behavior, and intentions to smoke among and multiethnic sample of sixth graders. *Ethn. Health* **9,** 369–397 (2004).

11. Byrne, D. G. & Mazanov, J. Sources of adolescent stress, smoking and the use of other drugs. *Stress Med.* **15,** 215–227 (1999).

12. Byrne, D. G., Byrne, A. E. & Reinhart, M. I. Personality, stress and the decision to commence cigarette smoking in adolescence. *J. Psychosom. Res.* **39,** 53–62 (1995).

13. Anda, R. F. *et al.* Adverse childhood experiences and smoking during adolescence and adulthood. *JAMA* **282,** 1652–1658 (1999).

14. Patton, G. C. *et al.* The course of early smoking: a population-based cohort study over three years. *Addict. Abingdon Engl.* **93,** 1251–1260 (1998).

15. García-Rodríguez, O. *et al.* Toward a comprehensive developmental model of smoking initiation and nicotine dependence. *Drug Alcohol Depend.* **144,** 160–169 (2014).

16. Simonoff, E. *et al.* The Virginia Twin Study of Adolescent Behavioral Development: influences of age, sex, and impairment on rates of disorder. *Arch. Gen. Psychiatry* **54,** 801–808 (1997).

17. Hewitt, J. K. *et al.* Genetics and developmental psychopathology: 1. Phenotypic assessment in the Virginia Twin Study of Adolescent Behavioral Development. *J. Child Psychol. Psychiatry* **38,** 943–963 (1997).

18. Boker, S. *et al.* OpenMx: An Open Source Extended Structural Equation Modeling Framework. *Psychometrika* **76,** 306–317 (2011).

19. Neale, M. & Cardon, L. *Methodology for Genetic Twin Studies of Twins and Families*. (Springer - Science + Business Media, B.V., 1992).

20. Akaike, H. Factor analysis and AIC. *Psychometrika* **52,** 317–332 (1987).

21. Patrick, D. L. *et al.* The validity of self-reported smoking: a review and meta-analysis. *Am. J. Public Health* **84,** 1086–1093 (1994).

Chapter 6: Prevalence and Correlates of Tobacco Use and Nicotine Delivery Systems among Young Adults in a University Setting

1. U.S. Department of Health and Human Services. *The health consequences of smoking - 50 years of progress: A report of the Surgeon General*. (2014).

2. Rigotti, N. A. & Wechsler, H. US college students' use of tobacco products: results of a national survey. *J. Am. Med. Assoc.* **284,** 609–705 (2000).

3. Wechsler, H., Rigotti, N. A., Gledhill-Hoyt, J. & Lee, H. Increased levels of cigarette use among college students: a cause for national concern. *JAMA* **280,** 1673–1678 (1998).

4. Korte, A. Alternative tobacco products may be just as dangerous as cigarettes. (2016). Available at: http://www.ttac.org/services/college/facts/cessation.html. (Accessed: 13th June 2016)

5. Lee, Y. O., Hebert, C. J., Nonnemaker, J. M. & Kim, A. E. Youth tobacco product use in the United States. *Pediatrics* **135,** 409–415 (2015).

6. Fix, B. V. *et al.* Patterns and correlates of polytobacco use in the United States over a decade: NSDUH 2002-2011. *Addict. Behav.* **39,** 768–781 (2014).

7. Backinger, C. L., Fagan, P. & O'Connell, M. E. Use of other tobacco products among U.S. adult cigarette smokers: prevalence, trends, and correlates. *Addict. Behav.* **33,** 472–489 (2009).

8. Eissenberg, T., Ward, K. D., Smith-Simone, S. & Maziak, W. Waterpipe tobacco smoking on a U.S. College campus: prevalence and correlates. *J. Adolesc. Health Off. Publ. Soc. Adolesc. Med.* **42,** 526–529 (2008).

9. Bombard, J. M., Pederson, L. L., Koval, J. J. & O'Hegarty, M. How are lifetime polytobacco users different than current cigarette-only users? Results from a Canadian young adult population. *Addict. Behav.* **34,** 1069–1072 (2009).

10. Wetter, D. W., McClure, J. B. & de Moor, C. Concommitant use of cigarettes and smokeless tobacco: prevalence, correlates, and predictors of tobacco cessation. *Prev. Med.* **34,** 638–648 (2002).

11. Moran, S., Wechsler, H. & Rigotti, N. A. Social smoking among US college students. *Pediatrics* **114,** 1028–1034 (2004).

12. Center for College Health and Safety. College tobacco prevention resource. (2014). Available at: http://www.ttac.org/services/college/facts/cessation.html. (Accessed: 13th June 2016)

13. Schane, R., Glantz, S. & Ling, P. Social smoking: implications for public health, clinical practice, and intervention research. *Am. J. Prev. Med.* **37,** 124–130 (2009).

14. Mounts, N. S. Adolescents' Perceptions of Parental Management of Peer Relationships in an Ethnically Diverse Sample. *J. Adolesc. Res.* **19,** 446–467 (2004).

15. Simons-Morton, B., Haynie, D. L., Crump, A. D., Eitel, S. P. & Saylor, K. E. Peer and parent influences on smoking and drinking among early adolescents. *Health Educ. Behav. Off. Publ. Soc. Public Health Educ.* **28,** 95–107 (2001).

16. Nichter, M., Carkoglu, A., The Tobacco Etiology Research Network & Nichter, M. Reconsidering stress and smoking: a qualitative study among college students. *Tob. Control* **16,** 211–214 (2007).

17. Iakunchykova, O. P. *et al.* The impact of early life stress on risk of tobacco smoking initiation by adolescents. *Addict. Behav.* **50,** 222–228 (2015).

18. Brandon, T. H. Negative Affect as Motivation to Smoke. *Curr. Dir. Psychol. Sci.* **3,** 33–37 (1994).

19. Naquin, M. R. & Gilbert, G. G. College students' smoking behavior, perceived stress, and coping styles. *J. Drug Educ.* **26,** 367–376 (1996).

20. Siqueira, L. M., Rolnitzky, L. M. & Rickert, V. I. Smoking cessation in adolescents: the role of nicotine dependence, stress, and coping methods. *Arch. Pediatr. Adolesc. Med.* **155,** 489–495 (2001).

21. Wetter, D. W. *et al.* Prevalence and predictors of transitions in smoking behavior among college students. *Health Psychol. Off. J. Div. Health Psychol. Am. Psychol. Assoc.* **23,** 168–177 (2004).

22. Dick, D. M. *et al.* Spit for science: launching a longitudinal study of genetic and environmental influences on substance use and emotional health at a large US university. *Front. Genet.* **5,** 47 (2014).

23. Harris, P. A. *et al.* Research Electronic Data Capture (REDCap) - a metadata-driven methodology and workflow process for providing translational research informatics support. *J. Biomed. Inform.* **42,** 377–381 (2009).

24. Popova, L. & Ling, P. M. Alternative tobacco product use and smoking cessation: a national study. *Am. J. Public Health* **103,** 923–930 (2013).

25. Akl, E. A. *et al.* The effects of waterpipe tobacco smoking on health outcomes: a systematic review. *Int. J. Epidemiol.* **39,** 834–857 (2010).

26. Baker, F. *et al.* Health risks associated with cigar smoking. *JAMA* **284,** 735–740 (2000).

27. Vander Weg, M. W. *et al.* Prevalence of alternative forms of tobacco use in a population of young adult military recruits. *Addict. Behav.* **33,** 69–82 (2008).

28. Caraballo, R. S., Yee, S. L., Gfroerer, J. & Mirza, S. A. Adult tobacco use among racial and ethnic groups living in the United States, 2002-2005. *Prev. Chronic. Dis.* **5,** A78 (2008).

29. King, B. A., Dube, S. R. & Tynan, M. A. Current tobacco use among adults in the United States: findings from the National Adult Tobacco Survey. *Am. J. Public Health* **102,** e93–e100 (2012).

30. Mahabee-Gittens, E. M., Xiao, Y., Gordon, J. S. & Khoury, J. C. The dynamic role of parental influences in preventing adolescent smoking initiation. *Addict. Behav.* **38,** 1905–1911 (2013).

31. Mahabee-Gittens, E. M., Xiao, Y., Gordon, J. S. & Khoury, J. C. The role of family influences on adolescent smoking in different racial/ethnic groups. *Nicotine Tob. Res. Off. J. Soc. Res. Nicotine Tob.* **14,** 264–273 (2012).

32. McKee, S. A., Maciejewski, P. K., Falba, T. & Mazure, C. M. Sex differences in the effects of stressful life events on changes in smoking status. *Addict. Abingdon Engl.* **98,** 847–855 (2003).

33. McKee, S. A. *et al.* Stress decreases the ability to resist smoking and potentiates smoking intensity and reward. *J. Psychopharmacol. Oxf. Engl.* **25,** 490–502 (2011).

34. Muraven, M. & Baumeister, R. F. Self-regulation and depletion of limited resources: does self-control resemble a muscle? *Psychol. Bull.* **126,** 247–259 (2000).

35. U.S. Food and Drug Administration. Overview of the Family Smoking Prevention and Tobacco Control Act: Consumer Fact Sheet. (2009). Available at: www. fda.gov/TobaccoProducts/GuidanceComplianceRegulatoryInformation/ ucm246129.htm.

36. Butler, K. M., Ickes, M. J., Rayens, M. K., Wiggins, A. T. & Hahn, E. J. Polytobacco Use Among College Students. *Nicotine Tob. Res. Off. J. Soc. Res. Nicotine Tob.* **18,** 163–169 (2016).

Chapter 7: Initial Experiences with Nicotine and its Association with Recent Use of Tobacco and Nicotine Dependence

1. Pomerleau, O. F., Pomerleau, C. S. & Namenek, R. J. Early experiences with tobacco among women smokers, ex-smokers, and never smokers. *Addiction* **93,** 595–599 (1998).

2. Pomerleau, O. F., Collins, A. C., Shiffman, S. & Pomerleau, C. S. Why some people smoke and others do not: new perspectives. *J. Consult. Clin. Psychol.* **61,** 723–731 (1993).

3. O'Connor, R. J. *et al.* An examination of early smoking experiences and smoking status in a national cross-sectional sample. *Addiction* **100,** 1352–1357 (2005).

4. Hirschman, R. S., Leventhal, H. & Glynn, K. The development of smoking behavior: conceptualization and supportive cross-sectional survey data. *J. Appl. Soc. Psychol.* **14,** 184–206 (1984).

5. Eissenberg, T. & Balster, R. L. Initial tobacco use episodes in children and adolescents: current knowledge, future directions. *Drug Alcohol Depend.* **59,** S41–S60 (2000).

6. De Wit, H. & Phillips, T. J. Do initial responses to drugs predict future use or abuse? *Neurosci. Biobehav. Rev.* **36,** 1565–1575 (2012).

7. DiFranza, J. R., Saveau, J. A. & Fletcher, K. Recollections and repercussions of the first inhaled cigarette. *Addict. Behav.* **29,** 261–272 (2004).

8. Buchmann, A. F. *et al.* Early smoking onset may promise initial pleasurable sensations and later addiction. *Addict. Biol.* **18,** 947–954 (2011).

9. Dick, D. M. *et al.* Spit for science: launching a longitudinal study of genetic and environmental influences on substance use and emotional health at a large US university. *Front. Genet.* **5,** 47 (2014).

10. Meyer, J. M., Silberg, J. L., Simonoff, E., Kendler, K. S. & Hewitt, J. K. The Virginia Twin-Family Study of Adolescent Behavioral Development: assessing sample biases in demographic correlates of psychopathology. *Psychol. Med.* **26,** 1119–1133 (1996).

11. Harris, P. A. *et al.* Research Electronic Data Capture (REDCap) - a metadata-driven methodology and workflow process for providing translational research informatics support. *J. Biomed. Inform.* **42,** 377–381 (2009).

12. Simonoff, E. *et al.* The Virginia Twin Study of Adolescent Behavioral Development: influences of age, sex, and impairment on rates of disorder. *Arch. Gen. Psychiatry* **54,** 801–808 (1997).

13. Eaves, L. J. *et al.* Genetics and developmental psychopathology: 2. the main effects of genes and environment on behavioral problems in the Virginia Twin Study of Adolescent and Behavioral Development. *J. Child Psychol. Psychiatry* **38,** 965–980 (1997).

14. Rodriguez, D. & Audrain, J. Construct validity analysis of the early smoking experience questionaire for adolescents. *Addict. Behav.* **29,** 1053–1057 (2004).

15. Baggio, S. *et al.* Factor structure of early smoking experiences and associations with smoking behavior: valence or sensitivity model? *Int. J. Environ. Res. Public. Health* **10,** 6305–6318 (2013).

16. Klein, H., Sterk, C. E. & Elifson, K. W. Initial smoking experiences and current smoking behaviors and perceptions among current smokers. *J. Addict.* **2013,** (2013).

17. Baggio, S. *et al.* The relationship between subjective experiences during first tobacco and cannabis and the effect of substance experienced first. *Nicotine Tob. Res.* **16,** 84–92 (2014).

18. Zabor, E. C. *et al.* Initial reactions to tobacco use and risk of future regular use. *Nicotine Tob. Res.* **15,** 509–517 (2013).

19. Wackowski, O. A. & Koob, G. F. The development and maintenance of drug addiction. *Neuropsychopharmacology* **39,** 254–262 (2014).

20. Hu, M., Greisler, P. C., Schaffran, C., Wall, M. M. & Kandel, D. B. Trajectories of criteria of nicotine dependence from adolescence to early adulthood. *Drug Alcohol Depend.* **125,** 283–289 (2012).

21. Pomerleau, C. S., Pomerleau, O. F., Namenek, R. J. & Marks, J. L. Initial exposure to nicotine in college-age women smokers and never smokers: a replication and extension. *J. Addict. Disord.* **18,** 13–19 (1999).

22. Okoli, C. T. C., Richardson, C. G. & Johnson, J. L. An examination of the relationship between adolescents' initial smoking experience and their exposure to peer and family member smoking. *Addict. Behav.* **33,** 1183–1191 (2008).

23. Brown, S. A. Drug effect expectancies and addictive behavior change. *Exp. Clin. Psychopharmacol.* **1,** 55–67 (1993).

24. Schuck, K., Otten, R., Engles, R. C. M. E. & Kleinjan, M. Initial responses to the first dose of nicotine in novel smokers: the role of exposure to environmental smoking and genetic predisposition. *Psychol. Health* **29,** 698–716 (2014).

25. Kozlowski, L. T. & Harford, M. R. On the significance of never using a drug: a example from cigarette smoking. *J. Abnorm. Psychol.* **85,** 433–434 (1976).

26. Haberstick, B. C., Ehringer, M. A., Lessem, J. M., Hopfer, C. & Hewitt, J. K. Dizziness and the genetic influences on subjective experiences to initial cigarette use. *Addiction* **106,** 391–399 (2011).

27. Sherva, R. *et al.* Association of a single nucleotide polymorphism in neuronal acetylcholine receptor subunit alpha 5 (CHRNA5) with smoking status and with 'pleasurable buzz' during early experimentation with smoking. *Addiction* **103,** 1544–1552 (2008).

28. Pomerleau, O. F. & Pomerleau, C. S. Dizziness upon initial experiementation with cigarettes: implications for smoking persistence. *Addiction* **106,**

29. Perkins, K. A. *et al.* VARIABILITY IN INITIAL NICOTINE SENSITIVITY DUE TO SEX, HISTORY OF OTHER DRUG USE, AND PARENTAL SMOKING. *Drug Alcohol Depend.* **99,** 47–57 (2009).

30. Verhagen, M., Kleinjan, M. & Engels, R. C. A systematic review of the A118G (Asn40Asp) variant of OPRM1 in relation to smoking initiation, nicotine dependence and smoking cessation. *Pharmacogenomics* **13,** 917–933 (2012).

31. Perkins, K. A. *et al.* GENE AND GENE BY SEX ASSOCIATIONS WITH INITIAL SENSITIVITY TO NICOTINE IN NONSMOKERS. *Behav. Pharmacol.* **19,** 630–640 (2008).

32. Nobel, E. P. D2 dopamine receptor gene in psychiatric and neurologic disorders and its phenotypes. *Am. J. Med. Genet. Neuropsychiatr. Genet.* **11B,** 103–125 (2003).

Chapter 8: An Exploration of Sex Differences in Responses to Items of the Fagerström Test for Nicotine Dependence

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

2. Haddock, C. K., Lando, H., Klesges, R. C., Talcott, G. W. & Renaud, E. A. A study of the psychometric and predictive properties of the Fagerström Test for Nicotine Dependence in a population of young smokers. *Nicotine Tob. Res.* **1,** 59–66 (1999).

3. Radzius, A. *et al.* A factor analysis of the Fagerström Test for Nicotine Dependence (FTND). *Nicotine Tob. Res.* **5,** 255–260 (2003).

4. Floyd, F. J. & Widaman, K. F. Factor analysis in the development and refinement of clinical assessment instruments. *Psychol. Assess.* **7,** 286–299 (1995).

5. Payne, T. J., Smith, P. O., McCracken, L. M., McSherry, W. C. & Anthony, M. M. Assessing nicotine dependence: a comparison of the Fagerström Tolerance Questionnaire (FTQ) with the Fagerström Test for Nicotine Dependence (FTND) in a clinical sample. *Addict. Behav.* **19,** 307–37 (1994).

6. Richardson, C. G. & Ratner, P. A. A confirmatory factor analysis of the Fagerström Test for Nicotine Dependence. *Addict. Behav.* **30,** 695–709 (2005).

7. Nakajima, M., al'Absi, M., Dokam, A., Alsoofi, M. & Khalil, N. S. An examination of the Fagerström Test for Nicotine Dependence among concurrent and khat users. *J. Psychoactive Drugs* **44,** 437–441 (2012).

8. Furr, R. M. *Scale Construction and Psychometrics for Social and Personality Psychology*. (SAGE, 2010).

9. MacCallum, R. C., Widaman, K. F., Zhang, S. & Hong, S. Sample size in factor analysis. *Psychol. Methods* **4,** 84–99 (1999).

Chapter 9: Genetic Analyses of Tobacco Use Behaviors among an Ethnically Diverse University Sample

1. Kaprio, J. Genetic epidemiology of smoking behavior and nicotine dependence. *J. Chronic Obstr. Pulm. Dis.* **6,** 304–306 (2009).

2. Kendler, K. S., Schmitt, E., Aggen, S. H. & Prescott, C. A. Genetic and environmental influences on alcohol, caffeine, cannabis, and nicotine use from early adolescence to middle adulthood. *Arch Gen Psychiatry* **65,** 674–682 (2008).

3. Schnoll, R. A., Johnson, T. A. & Lerman, C. Genetics and smoking behavior. *Curr. Psychiatry Rep.* **9,** 349–357 (2007).

4. Haberstick, B. C., Timberlake, D. S. & Ehringer, M. A. Genes, time to first cigarette and nicotine dependence in a general popualtion of young adults. *Addiction* **102,** 655–665 (2007).

5. Bergen, S. E., Gardner, C. O. & Kendler, K. S. Age related changes in heritability of behavioral phenotypes over adolescence and young adulthood: a meta-analysis. *Twin Res. Hum. Genet.* **10,** 423–433 (2007).

6. Tobacco and Genetics Consortium. Genome-wide meta-analysis identify multiple loci associated with smoking behavior. *Nat. Genet.* **42,** 441–447 (2010).

7. Dick, D. M. *et al.* Spit for science: launching a longitudinal study of genetic and environmental influences on substance use and emotional health at a large US university. *Front. Genet.* **5,** 47 (2014).

8. Harris, P. A. *et al.* Research Electronic Data Capture (REDCap) - a metadata-driven methodology and workflow process for providing translational research informatics support. *J. Biomed. Inform.* **42,** 377–381 (2009).

9. Webb, B. T. *et al.* Molecular genetic influences on normative and problematic alcohol use in a population-based sample of college students. *Front. Genet.* (2017). doi:10.3389/fgene.2017.00030

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

11. Delaneau, O., Zagury, J. & Marchini, J. Improved whole-chromosome phasing for disease and population genetic studies. *Nat. Methods* **10,** 5–6 (2013).

12. Howie, B. N., Donnelly, P. & Marchini, J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet.* **5,** e1000529 (2009).

13. Price, A. L., Zaitlen, N. A., Reich, D. & Patterson, N. New approaches to population stratification in genome-wide association studies. *Nat. Rev. Genet.* **11,** 459–463 (2010).

14. Price, A. L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* **38,** 904–909 (2006).

15. Patterson, N., Price, A. L. & Reich, D. Population structure and eigenanalysis. *PLoS Genet.* **2,** e190

16. Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: A tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* **88,** 76–82 (2011).

17. Marchini, J. & Howie, B. N. Genotype imputation for genome-wide association studies. *Nat. Rev. Genet.* **11,** 499–511 (2010).

18. Bigdeli, T. B., Neale, B. M. & Neale, M. C. Statistical properties of single-marker tests for rare variants. *Twin Res. Hum. Genet.* **17,** 143–150 (2014).

19. Willer, C. J., Li, Y. & Abecasis, G. R. Fast and efficient meta-analysis of genomewide scans. *Bioinformatics* **26,** 2190–2191 (2010).

20. R Core Team. *R: A Language and Environment for Statistical Computing*. (2016).

21. Huber, W. *et al.* Orchestrating high-throughput genomic analysis with Bioconductor. *Nat. Methods* **12,** 115–121 (2015).

22. Pruim, R. J. *et al.* LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics* **26,** 2336–2337 (2010).

23. Visscher, P. M. *et al.* Statistical Power to Detect Genetic (Co)Variance of Complex Traits Using SNP Data in Unrelated Samples. *PLOS Genet.* **10,** e1004269 (2014).

24. NCBI. RAB11FIP3 RAB11 family interacting protein 3 [homo sapiens (human)]. (2017). Available at: http://www.ncbi.nlm.nih.gov/gene/9727. (Accessed: 22nd February 2017)

25. Sridhar, S. *et al.* Smoking-induced gene expression changes in the bronchial airway are reflected in nasal and buccal epithelium. *BMC Genomics* **9,** (2008).

Chapter 10: Polygenic Risk Scores for Tobacco Use Behaviors: Are They Predictive Within a University Sample?

1. Uhl, G. R. *et al.* Molecular genetics of nicotine dependence and abstinence: whole genome association using 520,000 SNPs. *BMC Genet.* **8,** 10 (2007).

2. Beirut, L. J., Madden, P. A. & Breslau, N. Novel genes identified in a high-density genome wide association for nicotine dependence. *Hum. Mol. Genet.* **16,** 24–35 (2007).

3. Vink, J. M., Smit, A. B. & de Geus, E. J. C. Genome-wide association study of smoking initiation and current smoking. *Am. J. Hum. Genet.* **84,** 367–379 (2009).

4. Vink, J. M., Hottenga, J. J. & de Geus, E. J. C. Polygenic risk scores for smoking: predictors for alcohol and cannabis use? *Addiction* **109,** 1141–1151 (2014).

5. Thorgeirsson, T. E. *et al.* Sequence variants at CHRNB3-CHRNA6 and CYP2A6 affect smoking behavior. *Nat. Genet.* **42,** 448–453 (2010).

6. Tobacco and Genetics Consortium. Genome-wide meta-analysis identify multiple loci associated with smoking behavior. *Nat. Genet.* **42,** 441–447 (2010).

7. Thorgeirsson, T. E., Geller, F. & Sulem, P. A variant associated with nicotine dependence, lung cancer and peripheral arterial disease. *Nature* **452,** 638–642 (2008).

8. Vink, J. M. Genetics of addiction: future focus on gene x environment interaction? *J. Stud. Alcohol Drugs* **77,** 684–687 (2016).

9. Maher, B. S. Polygenic scores in epidemiology: risk prediction, etiology, and clinical utility. *Curr. Epidemiol. Rep.* **2,** 239–244 (2015).

10. The International Schizophrenia Consortium. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* **460,** 748–752 (2009).

11. Vilhjalmsson, B. J., Yang, J. & Finucane, H. K. Modeling linkage disequilibrium increases accuracy of polygenic risk scores. *Am. J. Hum. Genet.* **97,** 576–592

12. Euesden, J., Lewis, C. M. & O'Reilly, P. F. PRSice: Polygenic Risk Score Software. *Bioinformatics* **31,** 1466–1468 (2015).

13. Meyers, J. L. *et al.* Interaction between polygenic risk for cigarette use and environmental exposures in the Detroit neighborhood health study. *Transl. Psychiatry* **3,** e290 (2013).

14. Do, E. K. & Maes, H. H. Narrative review of genes, environment, and cigarettes. *Ann. Med.* **48,** 337–351 (2016).

15. Dick, D. M. *et al.* Spit for science: launching a longitudinal study of genetic and environmental influences on substance use and emotional health at a large US university. *Front. Genet.* **5,** 47 (2014).

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

17. Wray, N. R. *et al.* Research review: polygenic methods and their application to pyschiatric traits. *J. Child Psychol. Psychiatry* **55,** 1068–1087 (2014).

18. Do, E. K. *et al.* Genetic and environmental influences on smoking behavior across adolescence and young adulthood in the Virginia Twin Study of Adolescent Behavioral Development and the Transitions to Substance Abuse Follow-Up. *Twin Res. Hum. Genet.* **18,** 43–51 (2015).

19. Groen-Blokhuis, M. M., Middeldorp, C. M. & Kan, K. J. Attention-deficit/hyperactivity disorder polygenic risk scores predict attention problems in a population-based sampl of children. *J. Am. Acad. Child Adolesc. Psychiatry* **53,** 1123–1129 (2014).

20. Price, A. L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* **38,** 904–909 (2006).

21. Domingue, B. W. *et al.* Polygenic risk predicts obesity in both white and black young adults. *PLoS ONE* **9,** e101595–e101589 (2014).

Chapter 11: A Moving Target: The Emergence of Nicotine Delivery Systems and its Potential Public Health Impact

1. U.S. Department of Health and Human Services. *The health consequences of smoking - 50 years of progress: A report of the Surgeon General*. (2014).

2. Barnett, T. E., Forest, J. R., Porter, L. & Curbow, B. A. A multiyear assessment of hookah use prevalence among Florida high school students. *Nicotine Tob. Res.* **16,** 373–377 (2014).

3. Centers for Disease Control and Prevention. Current cigarette smoking among adults - United States 2005-2014. *Morb. Mortal. Wkly. Rep.* **64,** 1233–1240 (2015).

4. Marshall, J. R., Lotfipour, S. & Chakravarthy, B. Growing trend of alternative tobacco use among the nation's youth: a new generation of addicts. *West. J. Emerg. Med.* **17,** 139–142 (2016).

5. Weaver, S. R. *et al.* Use of electronic nicotine delivery systems and other tobacco products among USA adults, 2014: results from a national survey. *Int. J. Public Health* **61,** 177–188 (2016).

6. Herzog, B., Gerberi, J. & Scott, A. *Tobacco talk: independent vapor manufacturer survey*. (2015).

7. Philip Morris International. Annual meeting of shareholders. (2015). Available at: http://investors.pmi.com/phoenix.zhtml?c=146476&p=irol-eventDetails&EventId=5125408. (Accessed: 12th February 2016)

8. Rigotti, N. A. & Wechsler, H. US college students' use of tobacco products: results of a national survey. *J. Am. Med. Assoc.* **284,** 609–705 (2000).

9. Backinger, C. L., Fagan, P. & O'Connell, M. E. Use of other tobacco products among U.S. adult cigarette smokers: prevalence, trends, and correlates. *Addict. Behav.* **33,** 472–489 (2009).

10. Wetter, D. W., McClure, J. B. & de Moor, C. Concommitant use of cigarettes and smokeless tobacco: prevalence, correlates, and predictors of tobacco cessation. *Prev. Med.* **34,** 638–648 (2002).

11. Stellman, S. D. & Djordjevic, M. V. Monitoring the tobacco use epidemic II: the agent: current and emerging tobacco products. *Prev. Med.* **48,** S11–S15 (2009).

12. Zhu, S., Sun, J. Y. & Bonnevie, E. Four hundred and sixty brands of e-cigarettes and counting: implications for product regulation. *Tob. Control* **23,** iii3-iii9 (2014).

13. Connolly, G. N. The marketing of nicotine addiction by one oral snuff manufacturer. *Tob. Control* **4,** 73–79 (1995).

14. Delnevo, C. D. *et al.* Examining market trends in the United States smokeless tobacco use: 2005-2011. *Tob. Control* **23,** 107–112 (2014).

15. Meija, A. B. & Ling, P. M. Tobacco industry consumer research on smokeless tobacco users and product development. *Am. J. Public Health* **100,** 78–87 (2010).

16. Timberlake, D. S., Pechmann, C. & Tran, S. Y. A content analysis of camel snus advertisements in print media. *Nicotine Tob. Res.* **13,** 431–439 (2011).

17. Felberbaum, M. Reynolds American 3Q profit falls on charges. *Yahoo! News* (2011).

18. Felberbaum, M. Altria profit falls 9 percent in 4th quarter. *Yahoo! News* (2012).

19. Amrock, S. M. & Weitzman, M. Alternative tobacco products as the second front in the war on tobacco. *J. Am. Med. Assoc.* **314,** 1507–1508 (2015).

20. Chen, I. & Husten, C. G. Introduction to tobacco control supplement. *Tob. Control* **23,** ii1-ii3 (2015).

21. Lindblom, E. N. Effectively regulating e-cigarettes and their advertising - and the first amendment. *Food Drug Law* **70,** 57–94 (2015).

22. Russell, C. Unintended consequences of the FDA's e-cigarette regulations. *The Hill* (2016). Available at: http://thehill.com/blogs/pundits-blog/healthcare/292542-unintended-consequences-of-the-fdas-e-cigarette-regulations. (Accessed: 13th April 2017)

23. Cummings, K. M. *et al.* Are smokers adequately informed about the health risks of smoking and medicinal nicotine. *Nicotine Tob. Res.* **6,** S333–S340 (2004).

24. Kaufman, A. R., Waters, E. A. & Parascandola, M. Food and Drug Administration evaluation and cigarette smoking risk perceptions. *Am. J. Health Behav.* **35,** 766–776 (2011).

25. Latimer, L. A., Batanova, M. & Loukas, A. Prevalence and harm perceptions of various tobacco products among college students. *Nicotine Tob. Res.* **16,** 519–526 (2014).

26. Carpenter, C. M., Connolly, G. N., Ayo-Yusuf, O. A. & Wayne, G. F. Developing smokeless tobacco products for smokers: an examination of tobacco industry documents. *Tob. Control* **18,** 54–59 (2009).

27. Saunders, C. & Geletko, K. Adolescent cigarette smokers' and non-cigarette smokers' use of alternative tobacco products. *Nicotine Tob. Res.* **14,** 977–985 (2012).

28. Enofe, N., Berg, C. J. & Nehl, E. J. Alternative tobacco use among college students: who is at highest risk? *Am. J. Health Behav.* **38,** 180–189 (2014).

29. Kong, G., Morean, M. E., Cavallo, D. A., Camenga, D. R. & Krishnan-Sarin, S. Reasons for electronic cigarette experimentation and discontinuation among adolescents and young adults. *Nicotine Tob. Res.* **17,** 847–854 (2015).

30. McMillen, R., Maduka, J. & Winickoff, J. Use of emerging tobacco products in the United States. *J. Environ. Public Health* **2012,** 1–8 (2012).

31. McMillen, R., Gottlieb, M. A., Whitmore Shaefer, R. M., Winickoff, J. & Klein, J. D. Trends in electronic cigarette use among U.S. adults: use is increasing in both smokers and nonsmokers. *Nicotine Tob. Res.* **17,** 1195–1202 (2015).

32. Henningfield, J., Rose, C. A. & Giovino, G. A. Brave new world of tobacco disease prevention: promoting dual tobacco-product use? *Am. J. Prev. Med.* **23,** 226–228 (2002).

33. Ling, P. M. & Glantz, S. A. Why and how the tobacco industry sells cigarettes to young adults: evidence from industry documents. *Am. J. Public Health* **92,** 908–916 (2002).

34. Soneji, S., Sargent, J., Tanski, S. & Primack, B. A. Associations between initial water pipe tobacco smoking and snus use and subsequent cigarette smoking. *JAMA Pediatr.* **169,** 129–136 (2015).

35. Arrazola, R., Singh, T. & Corey, C. G. Tobacco use among middle and high school students - United States 2011-2014. *Morb. Mortal. Wkly. Rep.* **64,** 381–385 (2015).

36. Rath, J. M., Villanti, A. C., Abrams, D. B. & Vollone, D. M. Patterns of tobacco use and dual use in US young adults: the missing llink betweeen youth prevention and adult cessation. *J. Environ. Public Health* **2012,** 1–9 (2012).

37. International Agency for Research on Cancer. *Methods for Evaluating Tobacco Control Policies*. **12,** (World Health Organization).

38. Cruz, T. B. Monitoring the tobacco use epidemic IV. The vector: Tobacco industry data sources and recommendations for research and evaluation. *Prev. Med.* **48,** S24–S34 (2009).

39. Berg, C. J. *et al.* Perceived harm, addictiveness, and social acceptability of tobacco products and marijuana among young adults: marijuana, hookah, and electronic cigarettes win. *Subst. Use Misuse* **50,** 79–89 (2015).

Chapter 12: Pilot Randomized Control Trial of Internet-Based Educational Intervention for Reduction of Tobacco Use (And Nicotine Dependence)

1. U.S. Department of Health and Human Services. *The health consequences of smoking - 50 years of progress: A report of the Surgeon General*. (2014).

2. Sullivan, P. F. & Kendler, K. S. The genetic epidemiology of smoking. *Nicotine Tob. Res.* **1,** S51–S57 (1999).

3. Maes, H. H. *et al.* A twin study of genetic and environmental influences on tobacco initiation, regular tobacco use, and nicotine dependence. *Psychol. Med.* **34,** 1251–1261 (2004).

4. Kaprio, J. Genetic epidemiology of smoking behavior and nicotine dependence. *J. Chronic Obstr. Pulm. Dis.* **6,** 304–306 (2009).

5. Kendler, K. S., Schmitt, E., Aggen, S. H. & Prescott, C. A. Genetic and environmental influences on alcohol, caffeine, cannabis, and nicotine use from early adolescence to middle adulthood. *Arch Gen Psychiatry* **65,** 674–682 (2008).

6. Karp, I., O'Loughlin, J., Paradis, G., Handley, J. & DiFranza, J. Smoking trajectories of adolescent novice smokers in a longitudinal study of tobacco use. *Ann. Epidemiol.* **15,** 445–452 (2005).

7. Maes, H. H. & Neale, M. C. in *NCI tobacco control monograph series 20: Phenotypes and endophenotypes: Foundations for genetic studies of nicotine use and dependence* 245–288 (US National Institutes of Health, 2009).

8. Do, E. K. *et al.* Genetic and environmental influences on smoking behavior across adolescence and young adulthood in the Virginia Twin Study of Adolescent

Behavioral Development and the Transitions to Substance Abuse Follow-Up. *Twin Res. Hum. Genet.* **18,** 43–51 (2015).

9. David, S. P., Hamidovic, A. & Chen, G. K. Genome-wide meta-analyses of smoking behaviors in African Americans. *Transl. Psychiatry* **2,** e119 (2012).

10. Beirut, L. J. Convergence of genetic findings for nicotine dependence and related diseases with chromsome 15q24-25. *Trends Pharmacol. Sci.* **31,** 46–51 (2010).

11. Saccone, S. F., Hinnchs, A. L. & Saccone, N. L. Cholinergic nicotinic receptor genes implicated in nicotine dependence association study targeting 348 candidate genes with 3713 SNPs. *Hum. Mol. Genet.* **16,** 36–49 (2007).

12. Thorgeirsson, T. E. *et al.* Sequence variants at CHRNB3-CHRNA6 and CYP2A6 affect smoking behavior. *Nat. Genet.* **42,** 448–453 (2010).

13. Thorgeirsson, T. E., Geller, F. & Sulem, P. A variant associated with nicotine dependence, lung cancer and peripheral arterial disease. *Nature* **452,** 638–642 (2008).

14. Tobacco and Genetics Consortium. Genome-wide meta-analysis identify multiple loci associated with smoking behavior. *Nat. Genet.* **42,** 441–447 (2010).

15. Ware, J. J. & Munafo, M. R. in *The Neurobiology and Genetics of Nicotine and Tobacco, Current Topics in Behavioral Neurosciences 23* (eds. Balfour, D. J. K. & Munafo, M. R.) 19–36 (Springer International Publishing, 2015).

16. Ware, J. J., Munafo, M. R. & Smith, G. D. The genetic basis of cotinine levels in daily smokers: results from a genome-wide meta-analysis. *Soc. Res. Nicotine Tob. 2014 Pap. Sess.* (2014).

17. Hartmann-Boyce, J., Lancaster, T. & Stead, L. F. Print-based self-help interventions for smoking cessation. *Cochrane Database Syst. Rev.* (2014). doi:10.1002/14651858.CD001118.pub3

18. Calo, W. A. & Kransy, S. E. Environmental determinants of smoking behaviors: the role of policy and environmental interventions in preventing smoking initiation and supporting cessation. *Curr. Cardiovasc. Risk Rep.* **7,** 446–452 (2013).

19. Chen, L., Johnson, E. O. & Breslau, N. Interplay of genetic risk factors and parent monitoring in risk for nicotine dependence. *Addiction* **104,** 1731–1740 (2009).

20. Dick, D. M. *et al.* Parental monitoring moderates the importance of genetic and environmental influences on adolescent smoking. *J. Abnorm. Child Psychol.* **116,** 213–218 (2007).

21. Ducci, F., Kaakinen, M. & Pouta, A. TTC12-ANKK1-DRD2 and CHRNA5-CHRNA3-CHRNB4 influence different pathways leading to smoking behavior from adolescence to mid-adulthood. *Biol. Psychiatry* **69,** 650–660 (2011).

22. Boardman, J. D. State-level moderation of genetic tendencies to smoke. *Am. J. Public Health* **99,** 480–486 (2009).

23. Meyers, J. L. *et al.* Interaction between polygenic risk for cigarette use and environmental exposures in the Detroit neighborhood health study. *Transl. Psychiatry* **3,** e290 (2013).

24. Timberlake, D. S. *et al.* The moderating effects of religiosity on the genetic and environmental determinants of smoking initiation. *Nicotine Tob. Res.* **8,** 123–133 (2006).

25. Do, E. K. & Maes, H. H. Genotype x Environment Interaction in Smoking Behaviors: A Systematic Review. *Nicotine Tob. Res.* **19,** 387–400 (2017).

26. Do, E. K. & Maes, H. H. Narrative review of genes, environment, and cigarettes. *Ann. Med.* **48,** 337–351 (2016).

27. Kano, M., Goto, Y., Atsuta, Y., Naito, M. & Hamajima, N. Smoking Cessation After Genotype Notification: Pilot Studies of Smokers Employed by a Municipal Government and Those on Nagoya University Medical Campus. *Nagoya J. Med. Sci.* **69,** 149–156 (2007).

28. Bize, R., Burnand, B., Mueller, Y. & Cornuz, J. Effectiveness of biomedical risk assessment as an aid for smoking cessation: a systematic review. *Tob. Control* **16,** 151–156 (2007).

29. Audrain, J. *et al.* Genetic susceptibility testing in smoking-cessation treatment: one-year outcomes of a randomized trial. *Addict. Behav.* **22,** 741–751 (1997).

30. Hishida, A., Terazawa, T. & Mamiya, T. Efficacy of genotype notification to Japanese smokers on smoking cessation - an intervention study. *Cancer Epidemiol.* **43,** 98–100 (2010).

31. Lerman, C. *et al.* Incorporating biomarkers of exposure and genetic susceptibility into smoking cessation treatment: effects on smoking-related cognitions, emotions, and behavior change. *Health Psychol.* **16,** 87–99 (1997).

32. McBride, C. M. *et al.* Incorporating Genetic Susceptibility Feedback into a Smoking Cessation Program for African-American Smokers with Low Income. *Cancer Epidemiol. Biomarkers Prev.* **11,** 521–528 (2002).

33. Carlsten, C., Halperin, A., Crouch, J. & Burke, W. Personalized medicine and tobacco-related health disparities: is there a role for genetics? *Ann. Fam. Med.* **9,** 366–371 (2011).

34. Sanderson, S. C. & Michie, S. Genetic testing for heart disease susceptibility: potential impact on motivation to quit smoking. *Clin. Genet.* **71,** 501–510 (2007).

35. Brown, J. A review of the evidence on technology-based interventions for the treatment of tobacco dependence in college health. *Worldviews Evid. Based Nurs.* **10,** 150–162 (2013).

36. Greene, C. S., Penrod, N. M., Williams, S. M. & Moore, J. H. Failure to replicate a genetic association may provide important clues about genetic architecture. *PLoS ONE* **4,** e5639 (2009).

37. Hayden, J. A. in *Introduction to Health Behaviour Theory* 31–44 (Jones and Barlett Publishers, 2009).

38. Dick, D. M. *et al.* Spit for science: launching a longitudinal study of genetic and environmental influences on substance use and emotional health at a large US university. *Front. Genet.* **5,** 47 (2014).

39. Harris, P. A. *et al.* Research Electronic Data Capture (REDCap) - a metadata-driven methodology and workflow process for providing translational research informatics support. *J. Biomed. Inform.* **42,** 377–381 (2009).

40. Gupta, S. K. Intention-to-treat concept: a review. *Perspect. Clin. Res.* **2,** 109–112

41. Newell, D. J. Intention-to-treat analysis: implications for quantitative and qualitative research. *Int. J. Epidemiol.* **21,** 837–841 (1992).

42. Ranstam, J. *et al.* Alternative analyses for handling incomplete follow-up in the intention-to-treat analysis: the randomized controlled trial of balloon kyphoplasty versus non-surgical care for vertebral compression fracture (FREE). *BMC Med. Res. Methodol.* 35 (2012). doi:10.1186/1471-2288-12-35

43. Irish, W. Making sense of biostatistics: as-treated and intention-to-treat analysis. *J. Clin. Res. Best Pract.* **7,**

44. Benowitz, N. L. *et al.* Biochemical verification of tobacco use and cessation. *Nicotine Tob. Res.* **4,** 159 (2002).

Chapter 13: Global Conclusions and Future Directions

1. Bennet, S. N. *et al.* Phenotype harmonization and cross-study collaboration in GWAS consortia: The GENEVA experience. *Genet. Epidemiol.* **35,** 159–173 (2011).

2. de Bakker, P. I. W. *et al.* Practical aspects of imputation-driven meta-analysis of genome-wide association studies. *Hum. Mol. Genet.* **17,** R122–R128 (2008).

3. Evangelou, E. & Ioannidis, J. P. A. Meta-analysis methods for genome-wide association studies and beyond. *Nat. Rev. Genet.* **14,** 379–389 (2013).

4. Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: A tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* **88,** 76–82 (2011).

5. Jostins, L. & Barrett, J. C. Genetic risk prediction in complex disease. *Hum. Mol. Genet.* **20,** R182–R188 (2011).

6. Smith, J. A., Ware, E. B., Midda, P., Beacher, L. & Kardia, S. L. R. Current applications of genetic risk scores to cardiovascular outcomes and subclinical phenotypes. *Curr. Epidemiol. Rep.* **2,** 180–190 (2015).

7. Weaver, S. R. *et al.* Use of electronic nicotine delivery systems and other tobacco products among USA adults, 2014: results from a national survey. *Int. J. Public Health* **61,** 177–188 (2016).

8. Cruz, T. B. Monitoring the tobacco use epidemic IV. The vector: Tobacco industry data sources and recommendations for research and evaluation. *Prev. Med.* **48,** S24–S34 (2009).

CLARIFICATION OF CONTRIBUTIONS

I am very fortunate to have had the help and support I have received from many individuals during my dissertation work. All work, except that of which are described below and/or are cited within the text, is exclusively my own.

Chapters 2/3: These chapters were reviewed by my pre-doctoral qualifying committee (Drs. Roxann Nay-Roberson, Nathan Gillespie, Brien Riley, Timothy York, Danielle Dick, Hermine Maes). Each member provided suggestions and comments on previous versions, that were incorporated in the present chapters.

Chapter 4/5: Data for these projects was originally collected by Drs. Lindon J. Eaves, Judy L. Silberg, and Donna R. Miles. Each of these individuals, in addition to Drs. Elizabeth C. Prom-Wormley and Hermine H. Maes provided suggestions and comments on previous versions of this manuscript/chapter, which were incorporated into the present versions.

Chapter 6: Data from S4S was made available to me by Drs. Danielle M. Dick and Kenneth S. Kendler, with the help of the Research Coordinator, Kimberly Pedersen. Suggestions, comments, and edits were provided by Megan E. Cooke, and Drs. Elizabeth C. Prom-Wormley and Hermine H. Maes.

Chapter 7: Data from Spit for Science was made available to me by Drs. Danielle M. Dick and Kenneth S. Kendler. Data from the VTSABD was originally collected by Drs. Lindon J. Eaves, Judy L. Silberg, and Donna R. Miles. Suggestions, comments, and edits were provided by Drs. Elizabeth C. Prom-Wormley and Hermine H. Maes.

Chapter 8: Data from S4S was made available to me by Drs. Danielle M. Dick and Kenneth S. Kendler. Dr. Hermine H. Maes helped to provide suggestions, comments, and edits.

Chapters 9/10: Data from Spit for Science was made available to me by Drs. Danielle M. Dick and Kenneth S. Kendler. Data from the Tobacco and Genetics Consortium was available online. Assistance with learning how to navigate the genotypic data in S4S was provided by Arden Moscati. Dr. Bradley T. Webb was responsible for setting up the “genetic analysis pipeline” for S4S used to conduct genetic analyses in this dissertation. This includes, but is not limited to: pre-imputation quality control, and imputation, creation of ancestry group principal components to correct for potential population stratification, genetic based population assignment, and within group quality control. Dr. Roseann Peterson and I worked together to create the within ancestry group principal components. Assistance with learning how to create polygenic risk scores was provided by Jeanne E. Savage and Dr. Roseann E. Peterson. Analytic choices were conducted in consultation with Drs. Bradley T. Webb, Roseann E. Peterson, and Hermine H. Maes. Further suggestions, comments, and edits were provided by Drs. Bradley T. Webb and Hermine H. Maes.

Chapter 12: This chapter was derived from an assignment completed for Dr. Juan Lu’s course in Clinical and Translational Sciences. In addition to teaching me about clinical trial design, Dr. Juan Lu provided feedback and comments. I also consulted with Drs. Kellie Carlyle and John M. Quillin regarding health communication.

VITA

Elizabeth Kieuvan Do was born on October 13, 1988 in Fairfax County, Virginia and is an American citizen. She graduated from Annandale High School, Annandale, Virginia in 2006. She received her Bachelor of Arts in Foreign Affairs from the University of Virginia, Charlottesville, Virginia in 2010 and subsequently received her Masters of Public Health from Virginia Commonwealth University, Richmond, Virginia in 2012.