Washington University in St. Louis
Washington University Open Scholarship

Arts & Sciences Electronic Theses and Dissertations | Arts & Sciences

Spring 5-15-2016

Refining Associations between Targeted Genes and the Development of Substance Use Disorders

Emily Olfson
Washington University in St. Louis

Follow this and additional works at: https://openscholarship.wustl.edu/art_sci_etds

Part of the Epidemiology Commons, Genetics Commons, and the Public Health Education and Promotion Commons

Recommended Citation
Olfson, Emily, "Refining Associations between Targeted Genes and the Development of Substance Use Disorders" (2016). *Arts & Sciences Electronic Theses and Dissertations*. 792.
https://openscholarship.wustl.edu/art_sci_etds/792

This Dissertation is brought to you for free and open access by the Arts & Sciences at Washington University Open Scholarship. It has been accepted for inclusion in Arts & Sciences Electronic Theses and Dissertations by an authorized administrator of Washington University Open Scholarship. For more information, please contact digital@wumail.wustl.edu.

WASHINGTON UNIVERSITY IN ST. LOUIS

Division of Biology & Biomedical Sciences

Human and Statistical Genetics

Dissertation Examination Committee:
Laura J. Bierut, Chair
Arpana Agrawal
Donald Conrad
Christina Gurnett
Jay Piccirillo
John Rice

Refining Associations between Targeted Genes and the Development of Substance Use Disorders

by

Emily Olfson

A dissertation presented to the
Graduate School of Arts and Sciences
of Washington University in
partial fulfillment of the
requirements for the degree
of Doctor of Philosophy

May 2016
St. Louis, Missouri

© 2016, Emily Olfson

TABLE OF CONTENTS

LIST OF FIGURES v
LIST OF TABLES vi
ACKNOWLEDGEMENTS vii
ABSTRACT ix
CHAPTER 1: INTRODUCTION: SUBSTANCE USE DISORDERS ARE COMPLEX DISEASES WITH IMPORTANT PUBLIC HEALTH IMPLICATIONS 2
1.1 The public health burden of substance use disorders 3
1.2 Genes influence substance use disorders 4
1.3 Scope of dissertation 7
1.4 References 10
CHAPTER 2: CONVERGENGE OF GWA AND CANDIDATE GENE STUDIES FOR ALCOHOLISM 19
2.1 Abstract 20
2.2 Introduction 22
2.3 Materials and Methods 25
2.4 Results 29
2.5 Discussion 32
2.6 Acknowledgements 39
2.7 References 40
2.8 Tables 46
2.9 Figures 50
CHAPTER 3: AN ADH1B VARIANT AND PEER DRINKING IN PROGRESSION TO ADOLESCENT DRINKING MILESTONES: EVIDENCE OF A GENE-BY-ENVIRONMENT INTERACTION 51
3.1 Abstract 52
3.2 Introduction 54
3.3 Materials and Methods 56
3.4 Results 62
3.5 Discussion 66

3.6 Acknowledgements 71

3.7 References 72

3.8 Tables 79

3.9 Figures 81

CHAPTER 4: *CYP2A6* METABOLISM IN THE DEVELOPMENT OF NICOTINE DEPENDENCE IN YOUNG ADULTS 83

4.1 Abstract 84

4.2 Introduction 85

4.3 Materials and Methods 87

4.4 Results 93

4.5 Discussion 97

4.6 Acknowledgements 102

4.7 References 103

4.8 Tables 110

4.9 Figures 112

4.10 Supplemental Tables 115

4.11 Supplemental Figures 119

CHAPTER 5: COMMON, LOW FREQUENCY, AND RARE CODING VARIANTS IN *CHRNA5* CONTRIBUTE TO NICOTINE DEPENDENCE IN EUROPEAN AND AFRICAN AMERICANS 121

5.1 Abstract 122

5.2 Introduction 123

5.3 Materials and Methods 125

5.4 Results 130

5.5 Discussion 135

5.6 Acknowledgements 140

5.7 References 143

5.8 Tables 150

5.9 Figures 153

5.10 Supplemental Methods 155

5.11 Supplemental Replication Study Descriptions 157

5.12 Supplemental Tables 169

5.13 Supplemental Figure 175
5.14 Supplemental Acknowledgements..176
CHAPTER 6: FUTURE STEPS TO UNDERSTAND THE ROLE OF TRAGETED GENES IN SUBSTANCE USE DISORDERS 183
6.1 Summary of the Dissertation 184
6.2 Future Directions 185
6.3 References...188

LIST OF FIGURES

Figure 2.1: Characteristics of genes associated with Alcoholism in the Human Genome Epidemiology (HuGE) Navigator 50

Figure 3.1: Sample selection for study analyses....................81

Figure 3.2: Survival estimates of drinking milestones stratified by ADH1B rs1229984 genotype and perceived best friends drinking....................82

Figure 4.1: Primary COGA sample selection....................112

Figure 4.2: Association between predicted metabolism and smoking behaviors in two studies of European American young adult daily smokers....................113

Figure 4.3: A theoretical framework of the development of smoking behaviors in relation to *CYP2A6* variation....................114

Supplemental Figure 4.1: Distribution of *CYP2A6* metabolism metric among COGA young adult daily smokers....................119

Supplemental Figure 4.2: Proportion of slow metabolizers among 4 categories of time to first cigarette after waking among young adult daily smokers....................120

Figure 5.1: Protein schematic of *CHRNA5* nonsynonymous and frameshift variants....................153

Figure 5.2: Forest plots showing the primary sample, replication samples, and random effects meta-analyses from *Multivariable model set 1*. 154

Supplemental Figure 5.1: Twenty-six *CHRNA5* coding variants ordered based on minor allele frequency....................175

LIST OF TABLES

Table 2.1: Characteristics of alcohol dependent cases and non-dependent controls in SAGE.....46

Table 2.2: Examination of top ten candidate loci for Alcoholism in SAGE 47

Table 2.1: SNPs in frequently studied candidate genes associated with Alcoholism 48

Table 3.1: Characteristics of samples used in analyses. 79

Table 3.2: Cox proportional hazards regression models of adolescent drinking milestones........80

Table 4.1: Characteristics of primary and replication samples of European American young adults 110

Table 4.2: Logistic regression models of smoking milestones in young adults 111

Supplemental Table 4.1: Predicted metabolism metric and *CYP2A6* diplotypes based on copy number and 5 SNPs. 115

Supplemental Table 4.2: Examination of dichotomous metabolism status as a predictor of smoking milestones in logistic regression models in young adults.........................116

Supplemental Table 4.3: Logistic regression models of the ordinal 4 level variable of time to first cigarette after waking (>60, 31-60, 6-30, ≤5 minutes) 117

Supplemental Table 4.4: Logistic regression models of nicotine dependence that explore the importance of time to first cigarette after waking 118

Table 5.1: Characteristics of primary sample 150

Table 5.2: The effect of common, low frequency, and rare *CHRNA5* coding variants on nicotine dependence in primary sample 151

Table 5.3: Variation in nicotine dependence risk explained by selected variants and multivariate models in primary sample 152

Supplemental Table 5.1: Characteristics of 26 high quality *CHRNA5* coding variants identified by targeted sequencing of 1 582 Nicotine dependent cases and 1 238 controls.. 169

Supplemental Table 5.2: Random effects meta-analysis results for multivariate model set 1 in European Americans 171

Supplemental Table 5.3: Random effects meta-analysis results for multivariate model set 1 in African Americans. 172

Supplemental Table 5.4: Random effects meta-analysis results for multivariate model set 2 in European Americans 173

Supplemental Table 5.5: Random effects meta-analysis results for multivariate model set 2 in African Americans. 174

ACKNOWLEDGEMENTS

First, I would like to express my deepest gratitude to my thesis advisor, Dr. Laura Bierut. Her rigorous scientific thinking, attentive teaching style, and thoughtful career advice has provided me with an ideal atmosphere to conduct research and grow as a scientist.

I would also like to thank all of the members of my thesis and defense committees: Drs. John Rice (thesis committee chair), Arpana Agrawal, Donald Conrad, Christina Gurnett, Mario Schootman, and Jay Piccirillo. Their incisive questioning and suggestions have improved the quality of my work. I appreciate all of their input both during and outside of committee meetings.

I also thank all of the current and past members of the Bierut group. In particular, I would like to thank Drs. Li-Shiun Chen, Rob Culverhouse, Sarah Hartz, and Nancy Saccone for providing me with constructive feedback and assistance in lab meetings and on an as needed basis. I would also like to thank Louis Fox and Joseph Mullaney for teaching me about the various datasets and providing me with technical support for my analyses. Finally, I thank Sherri Fisher for providing editorial, managerial, and scientific advice on my projects.

Thank you to all of my co-authors and collaborators who made this work possible. In particular, I thank the members of The Collaborative Study on the Genetics of Alcoholism (COGA) and the Collaborative Genetic Study of Nicotine Dependence (COGEND) for giving me the opportunity to work with these datasets and for providing critical feedback on manuscripts and presentations. I would also like to thank Dr. Alison Goate and her lab members for their contribution to many of the projects.

I thank the Washington University Medical Scientist Training Program for accepting me and providing funding. Many thanks to the MSTP staff of Brian Sullivan, Christy Durbin, Liz

Bayer, and Linda Perniciaro for keeping me on track and answering questions. Thank you to Dr. Wayne Yokoyama for his steadfast guidance and support of my medical and scientific training.

I would also like to thank the TL1 Predoctoral Clinical Research Training program. This experience introduced me to translational research methods, which complemented my graduate studies. In particular, thank you to Dr. Jay Piccirillo for serving as an exemplary program director and TL1 advisor.

Importantly, I would like to express my gratitude to my family for their love and encouragement. I thank my father, Mark Olfson, for nurturing my passion for math and science from a young age, and providing critical career advice and feedback along the way. I thank my mother, Melinda Hunt, for encouraging me to think creatively about problems and to apply for new opportunities. I would also like to thank my stepmom Anne Marie Albano, grandma Elaine Glickman, and my sister Rachel Olfson for inspiring me to work hard with their best wishes.

Most of all, I would like to thank my fiancé, Andrew Loza. His love of science is infectious, and his continual support is endless. During my tenure in graduate school, Andrew's companionship brightened each day, and I am looking forward to the next chapters of our lives.

My training in graduate school was funded through grants T32GM07200, UL1TR000448, TL1TR000449, and F30AA023685 from the NIH.

Emily Olfson

Washington University in St. Louis

May 2016

ABSTRACT OF THE DISSERTATION

Refining Associations between Targeted Genes and the Development of Substance Use Disorders

by

Emily Olfson

Doctor of Philosophy in Biology and Biomedical Sciences

Human and Statistical Genetics

Washington University in St. Louis, 2016

Professor Laura J. Bierut, Chair

Recent genome-wide association studies (GWAS) provide strong evidence for the contribution of a few specific genes to alcohol and nicotine dependence. Chapter 2 explores numerous previously identified candidate genes for alcohol dependence using a publically available GWAS. I found that many candidate loci do not replicate, highlighting the utility of GWAS for focusing on disease associated genes. Chapters 3-5 dissect associations between three genome-wide significant genes and substance use disorders. Chapter 3 focuses on a functional variant in alcohol dehydrogenase (ADH) 1B. Through examining 1,550 adolescent drinkers in the Collaborative Study on the Genetics of Alcoholism (COGA), I extended adult findings by showing that this *ADH1B* variant protected against early drinking milestones. Furthermore, I provided evidence for a gene-by-environment interaction where best friends drinking eliminated this genetic protective effect, illustrating the important interplay between genetic and environmental factors in the development of drinking behaviors. Chapter 4 examines variation in the nicotine metabolizing cytochrome P450 gene *CYP2A6*. Previous studies show slow metabolizers smoke fewer cigarettes, but provide conflicting results on the role of *CYP2A6* in

nicotine dependence. Using a COGA young adult sample, I found that *CYP2A6* metabolism was not associated with smoking initiation or daily smoking, but among daily smokers, slow metabolism was associated with increased risk of dependence. This association was replicated in an independent sample from the Collaborative Study of Nicotine Dependence, adding insight into the complex role of *CYP2A6* across stages of smoking behaviors. Chapter 5 focuses on coding variation in the α5 nicotinic receptor subunit gene (*CHRNA5*), which harbors a nonsynonymous common variant robustly associated with nicotine dependence. I examined targeted sequence data of *CHRNA5* from approximately 3,000 nicotine dependent cases and controls, with independent replication of common and low frequency variants in 12 studies. I found that common, low frequency, and rare *CHRNA5* coding variants were independently associated with increased nicotine dependence risk. Incorporating coding variants beyond the well-studied common variant increased the variance in nicotine dependence explained by *CHRNA5*. Overall, this dissertation advances our understanding of targeted genes for substance use disorders by incorporating important environments, critical developmental periods, and rare variants.

“The ideal art, the noblest of art: working with the complexities of life, refusing to simplify, to ‘overcome’ doubt.”

-Joyce Carol Oates

CHAPTER ONE:

Introduction: Substance use disorders are complex diseases with important public health implications

1.1 THE PUBLIC HEALTH BURDEN OF SUBSTANCE USE DISORDERS

Substance use disorders are a leading cause of preventable death in both the United States and worldwide. Each year, 3.3 million people die due to the harmful effects of alcohol, representing 5.9% of all deaths across the world (WHO, 2014b). From 2006-2010, approximately 88,000 deaths and 2.5 million years of potential life were lost in the United States due to alcohol consumption (Stahre et al., 2014). Heavy drinking can have immediate health related effects through injuries, violence, alcohol poisoning, and risky sexual behaviors. Over time, excessive alcohol use can lead to several chronic diseases, including heart disease, liver disease, cancer, and mental health problems. Tobacco smoking similarly causes many chronic diseases, including heart disease, lung cancer, and respiratory illnesses. Nearly 6 million people die each year from tobacco products worldwide (WHO, 2014a). In the United States alone, cigarette smoking causes about one of every five deaths each year, accounting for approximately 480,000 deaths annually (CDC, 2014). Beyond excessive morbidity and mortality associated with these behaviors, economically society pays a high price for substance use. Approximately 11% of the total federal and state government budget is spent on the consequences of alcohol, tobacco, and other drug use (CASAColumbia, 2009).

1.2 GENES INFLUENCE SUBSTANCE USE DISORDERS

Genetic factors have long been recognized to influence the development of substance use disorders. Twin studies estimate that the heritability of substance dependence is approximately 50%-60% (Heath et al., 1997, Kendler et al., 2003, Knopik et al., 2004, Li, 2006). Candidate gene studies have identified hundreds of genes potentially associated with substance use disorders (Yu et al.). More recently, large-scale genome wide association (GWA) studies have confirmed the contribution of a few specific genes to alcoholism and smoking (Wang et al., 2012, Rietschel and Treutlein, 2013).

Genes with the clearest associations with alcoholism produce metabolizing enzymes (Edenberg and Foroud, 2013, Hurley and Edenberg, 2012). Alcohol is primarily metabolized in the liver, and the first step is the oxidation of ethanol to acetaldehyde. Acetaldehyde is a toxic intermediate, and systemic build-up leads to unpleasant feelings, such as dizziness, nausea, and tachycardia. This process is catalyzed by alcohol dehydrogenases (ADH), which are a class of enzymes encoded by seven genes on chromosome 4. The enzyme encoded by *ADH1B* has the highest concentration in adult livers, and the *ADH1B* rs1229984 variant has reached genome-wide significance levels for alcohol dependence across different ancestry populations (Li et al., 2011, Li et al., 2012, Bierut et al., 2012, Gelernter et al., 2014). The minor A allele of rs1229984 causes an amino acid change at position 48 that increases the rate of oxidation of alcohol and leads to transient increases in acetaldehyde. Given the toxicity of acetaldehyde, negative effects are experienced by people with this *ADH1B* variant when they consume alcohol, which discourages heavy drinking.

Similar to alcoholism, nicotine metabolism genes are important for the development of smoking behaviors. The cytochrome P450 enzyme CYP2A6 is responsible for the majority of

oxidation of nicotine to cotinine, which is the primary pathway of nicotine metabolism (Hukkanen et al., 2005). The region on chromosome 19 that harbors the *CYP2A6* gene is genome-wide significant in large meta-analyses focused on cigarettes per day (Thorgeirsson et al., 2010, TAG, 2010). Among nicotine dependent adults, the majority of studies support that genetically slower metabolizers smoke fewer cigarettes per day (Benowitz, 2008), reflecting the fact that smokers titrate their cigarette consumption to maintain certain nicotine levels. However, studies in youth present conflicting results on the effect of nicotine metabolism on the development of nicotine dependence and other smoking behaviors (Audrain-McGovern et al., 2007, Huang et al., 2005, Moolchan et al., 2009, O'Loughlin et al., 2004, Rubinstein et al., 2013, Rubinstein et al., 2008). One challenge to studying *CYP2A6* is that the locus is highly polymorphic with functionally diverse alleles. Recently, Bloom et al. (2011) developed a metabolism metric to predict nicotine metabolism based on different *CYP2A6* haplotypes. This metric predicts approximately 70% of the variance in metabolism of orally administered nicotine to cotinine in European Americans.

The strongest genetic contribution to nicotine dependence comes from variation in nicotinic receptor subunit genes. The physiological effects of nicotine are produced through the neuronal nicotinic acetylcholine receptors (nAChRs) (Dani and De Biasi, 2001), which are pentamiric catonic channels primarily composed of combinations of α and β subunits. Several independent studies have demonstrated that the nonsynonymous rs16969968 variant in the α5 subunit gene (*CHRNA5*) is associated with a variety of smoking behaviors (Saccone et al., 2007, Berrettini et al., 2008, Weiss et al., 2008, Stevens et al., 2008, Sherva et al., 2008, Baker et al., 2009, Keskitalo et al., 2009). Subsequently, this association has been reported as the most significant in genome-wide meta-analyses of cigarettes per day ($p=5.57x10^{-72}$) (Thorgeirsson et

al., 2010, TAG, 2010, Liu et al., 2010). Additional studies have extended this association with rs16969968 to smoking-related illnesses, including lung cancer and chronic obstructive pulmonary disease (Amos et al., 2008, Hung et al., 2008, Pillai et al., 2009, Thorgeirsson et al., 2008). This association likely reflects greater exposure to carcinogens in tobacco smoke in response to higher levels of nicotine dependence. Nicotinic receptors containing α5 subunits normally activate the interpeduncular pathway in response to nicotine intake, which discourages high levels of cigarette consumption (Fowler et al., 2011). The rs169669968 variant causes an amino acid change at position 398, and functional studies show that this change decreases receptor function (Bierut et al., 2008, Kuryatov et al., 2011). Decreased function of α5 containing receptors is hypothesized to prevent negative feedback in response to cigarette consumption, leading to heavy smoking.

1.3 SCOPE OF THE DISSERTATION

The primary goal of my dissertation is to improve our understanding of how targeted genes contribute to substance use disorders.

Chapter 2 examines well-studied candidate genes for alcohol dependence using a GWA study comparing alcohol dependent cases and controls. These targeted candidate genes were selected using the Human Genome Epidemiology Navigator, which catalogues published genetic association studies (Yu et al., 2008). Our findings suggest that several extensively studied candidate loci do not strongly contribute to risk of developing alcohol dependence. This chapter has been published in the Journal *Alcoholism: Clinical and Experimental Research* (Olfson and Bierut, 2012).

Olfson E, Bierut LJ. Convergence of genome-wide association and candidate gene studies for alcoholism. *Alcohol Clin Exp Res* 2012 36(12):2086-2094. PMCID: PMC3521088

Chapter 3 focuses on the interplay between the *ADH1B* rs1229984 variant and the critical social environment of peer drinking in the development of adolescent drinking behaviors. Through examining 1,550 European and African American youth enrolled in the Collaborative Study on the Genetics of Alcoholism (COGA), we found that this *ADH1B* variant was protective for early drinking milestones, but under the high risk environment of best friends drinking, this genetic protection was eliminated. These findings illustrate the important interplay between genes and environments in the development of drinking behaviors. This chapter has been published in the Journal *Alcoholism: Clinical and Experimental Research* (Olfson et al., 2014).

Olfson E, Edenberg HJ, Nurnberger J Jr, Agrawal A, Bucholz KK, Almasy LA, Chorlian D, Dick DM, Hesselbrock VM, Kramer JR, Kuperman S, Porjesz B, Schuckit MA, Tischfield JA, Wang JC, Wetherill L, Foroud TM, Rice J, Goate A, Bierut LJ. An *ADH1B* variant and peer drinking in progression to adolescent drinking milestones: Evidence of a gene-by-environment interaction. *Alcohol Clin Exp Res* 2014 Sept 24 NIHMSID#936008

Chapter 4 assesses the role of *CYP2A6* metabolism on the development of smoking behaviors during the critical developmental period of young adulthood. By examining over 1,000 European American young adults enrolled in COGA, we found that the *CYP2A6* metabolism metric was not associated with smoking initiation or the development of daily smoking, but among daily smokers, decreased metabolism was associated with an increased risk of nicotine dependence. This finding was replicated in an independent sample of young adult daily smokers enrolled in the Collaborative Study of Nicotine Dependence. These results demonstrate the complex role of *CYP2A6* variation across different developmental stages of smoking behaviors. At the time of dissertation defense, this chapter was in preparation for submission.

Olfson E, Bloom J, Bertelsen S, Breslau N, Budde J, Chen LS, Culverhouse R, Chorlian D, Dick DM, Edenberg HJ, Hatsukami D, Hesselbrovck VM, Kramer JR, Kuperman S, Porjesz B, Saccone NL, Schuckit MA, Stitzel J, Tischfield JA, Goate A, Bierut LJ. *CYP2A6* metabolism in the development of nicotine dependence in young adults.

Chapter 5 examines whether *CHRNA5* coding variants, beyond the well-studied common rs16969968 variant, contribute to nicotine dependence risk. Next-generation

sequencing of approximately 3,000 nicotine dependent cases and controls identified the only known common variant, 3 low frequency, and 22 rare variants. Our results showed that these newly identified variants independently contribute to nicotine dependence risk. Replication of common and low frequency variants using 12 independent studies with exome chip data in over 10,000 heavy and 10,000 light smokers further supported this conclusion. These newly identified low frequency and rare variants may have important health implications by influencing risk for smoking-related diseases and response to cessation therapies. At the time of dissertation defense, this chapter was in preparation for submission.

Olfson E, Saccone NL, Johnson EO, Chen LS, Culverhouse R, Doheny K, Foltz SM, Fox L, Gogarten SM, Hartz S, Hetrick K, Laurie CC, Marosy B, Amin N, Arnett D, Bartz TM, Bertelsen S, Borecki IB, Brown MR, Chasman DI, van Duijn CM, Feitosa MF, Fox ER, Franceschini N, Franco OH, Grove ML, Guo X, Hofman A, Kardia SLR, Morrison AC, Musani SK, Psaty BM, Rao DC, Reiner AP, Rice K, Ridker PM, Rose LM, Rotter JI, Schick UM, Schwander K, Uitterlinden AG, Vojinovic D, Wang JC, Ware EB, Wilson G, Yao J, Zhao W, Breslau N, Hatsukami D, Stitzel J, Rice J, Goate A, Bierut LJ. Common, low frequency, and rare coding variants in *CHRNA5* contribute to nicotine dependence in European and African Americans.

Overall, this work illustrates that although only a limited number of genes clearly contribute to substance use, hypothesis-driven analyses can advance our understanding of the specific mechanisms by which these genes alter substance use behaviors. In particular, these findings illustrate that incorporating environmental factors, critical developmental periods, and rare variants may refine associations between robust genes and substance use disorders.

1.4 REFERENCES

Amos CI, Wu X, Broderick P, Gorlov IP, Gu J, Eisen T, Dong Q, Zhang Q, Gu X, Vijayakrishnan J, Sullivan K, Matakidou A, Wang Y, Mills G, Doheny K, Tsai YY, Chen WV, Shete S, Spitz MR, Houlston RS (2008) Genome-wide association scan of tag SNPs identifies a susceptibility locus for lung cancer at 15q25.1. Nature genetics 40:616-622.

Audrain-McGovern J, Al Koudsi N, Rodriguez D, Wileyto EP, Shields PG, Tyndale RF (2007) The role of CYP2A6 in the emergence of nicotine dependence in adolescents. Pediatrics 119:e264-274.

Baker TB, Weiss RB, Bolt D, von Niederhausern A, Fiore MC, Dunn DM, Piper ME, Matsunami N, Smith SS, Coon H, McMahon WM, Scholand MB, Singh N, Hoidal JR, Kim SY, Leppert MF, Cannon DS (2009) Human neuronal acetylcholine receptor A5-A3-B4 haplotypes are associated with multiple nicotine dependence phenotypes. Nicotine & tobacco research : official journal of the Society for Research on Nicotine and Tobacco 11:785-796.

Benowitz NL (2008) Clinical pharmacology of nicotine: implications for understanding, preventing, and treating tobacco addiction. Clinical pharmacology and therapeutics 83:531-541.

Berrettini W, Yuan X, Tozzi F, Song K, Francks C, Chilcoat H, Waterworth D, Muglia P, Mooser V (2008) Alpha-5/alpha-3 nicotinic receptor subunit alleles increase risk for heavy smoking. Molecular psychiatry 13:368-373.

Bierut LJ, Goate AM, Breslau N, Johnson EO, Bertelsen S, Fox L, Agrawal A, Bucholz KK, Grucza R, Hesselbrock V, Kramer J, Kuperman S, Nurnberger J, Porjesz B, Saccone NL,

Schuckit M, Tischfield J, Wang JC, Foroud T, Rice JP, Edenberg HJ (2012) ADH1B is associated with alcohol dependence and alcohol consumption in populations of European and African ancestry. Molecular psychiatry 17:445-450.

Bierut LJ, Stitzel JA, Wang JC, Hinrichs AL, Grucza RA, Xuei X, Saccone NL, Saccone SF, Bertelsen S, Fox L, Horton WJ, Breslau N, Budde J, Cloninger CR, Dick DM, Foroud T, Hatsukami D, Hesselbrock V, Johnson EO, Kramer J, Kuperman S, Madden PA, Mayo K, Nurnberger J, Jr., Pomerleau O, Porjesz B, Reyes O, Schuckit M, Swan G, Tischfield JA, Edenberg HJ, Rice JP, Goate AM (2008) Variants in nicotinic receptors and risk for nicotine dependence. The American journal of psychiatry 165:1163-1171.

Bloom J, Hinrichs AL, Wang JC, von Weymarn LB, Kharasch ED, Bierut LJ, Goate A, Murphy SE (2011) The contribution of common CYP2A6 alleles to variation in nicotine metabolism among European-Americans. Pharmacogenetics and genomics 21:403-416.

CASAColumbia. The National Center on Addiction and Substance Abuse at Columbia University . Shoveling up II: The impact of substance abuse on federal, state and local budgets. . Available at: http://www.casacolumbia.org/addiction-research/reports/shoveling-ii-impact-substance-abuse-federal-state-and-local-budgets. Accessed October 16, 2014.

CDC. Tobacco-Related Mortality. Centers for disease control and preventionFebrary 6, 2014. Available at: http://www.cdc.gov/tobacco/data_statistics/fact_sheets/health_effects/tobacco_related_mortality/. Accessed October 16, 2014.

Dani JA, De Biasi M (2001) Cellular mechanisms of nicotine addiction. Pharmacology, biochemistry, and behavior 70:439-446.

Edenberg HJ, Foroud T (2013) Genetics and alcoholism. Nature reviews. Gastroenterology & hepatology.

Fowler CD, Lu Q, Johnson PM, Marks MJ, Kenny PJ (2011) Habenular alpha5 nicotinic receptor subunit signalling controls nicotine intake. Nature 471:597-601.

Gelernter J, Kranzler HR, Sherva R, Almasy L, Koesterer R, Smith AH, Anton R, Preuss UW, Ridinger M, Rujescu D, Wodarz N, Zill P, Zhao H, Farrer LA (2014) Genome-wide association study of alcohol dependence:significant findings in African- and European-Americans including novel risk loci. Molecular psychiatry 19:41-49.

Heath AC, Bucholz KK, Madden PA, Dinwiddie SH, Slutske WS, Bierut LJ, Statham DJ, Dunne MP, Whitfield JB, Martin NG (1997) Genetic and environmental contributions to alcohol dependence risk in a national twin sample: consistency of findings in women and men. Psychological medicine 27:1381-1396.

Huang S, Cook DG, Hinks LJ, Chen XH, Ye S, Gilg JA, Jarvis MJ, Whincup PH, Day IN (2005) CYP2A6, MAOA, DBH, DRD4, and 5HT2A genotypes, smoking behaviour and cotinine levels in 1518 UK adolescents. Pharmacogenetics and genomics 15:839-850.

Hukkanen J, Jacob P, 3rd, Benowitz NL (2005) Metabolism and disposition kinetics of nicotine. Pharmacological reviews 57:79-115.

Hung RJ, McKay JD, Gaborieau V, Boffetta P, Hashibe M, Zaridze D, Mukeria A, Szeszenia-Dabrowska N, Lissowska J, Rudnai P, Fabianova E, Mates D, Bencko V, Foretova L, Janout V, Chen C, Goodman G, Field JK, Liloglou T, Xinarianos G, Cassidy A, McLaughlin J, Liu G, Narod S, Krokan HE, Skorpen F, Elvestad MB, Hveem K, Vatten L, Linseisen J, Clavel-Chapelon F, Vineis P, Bueno-de-Mesquita HB, Lund E, Martinez C, Bingham S, Rasmuson T, Hainaut P, Riboli E, Ahrens W, Benhamou S, Lagiou P,

Trichopoulos D, Holcatova I, Merletti F, Kjaerheim K, Agudo A, Macfarlane G, Talamini R, Simonato L, Lowry R, Conway DI, Znaor A, Healy C, Zelenika D, Boland A, Delepine M, Foglio M, Lechner D, Matsuda F, Blanche H, Gut I, Heath S, Lathrop M, Brennan P (2008) A susceptibility locus for lung cancer maps to nicotinic acetylcholine receptor subunit genes on 15q25. Nature 452**:**633-637.

Hurley TD, Edenberg HJ (2012) Genes encoding enzymes involved in ethanol metabolism. Alcohol research : current reviews 34**:**339-344.

Kendler KS, Prescott CA, Myers J, Neale MC (2003) The structure of genetic and environmental risk factors for common psychiatric and substance use disorders in men and women. Archives of general psychiatry 60**:**929-937.

Keskitalo K, Broms U, Heliovaara M, Ripatti S, Surakka I, Perola M, Pitkaniemi J, Peltonen L, Aromaa A, Kaprio J (2009) Association of serum cotinine level with a cluster of three nicotinic acetylcholine receptor genes (CHRNA3/CHRNA5/CHRNB4) on chromosome 15. Human molecular genetics 18**:**4007-4012.

Knopik VS, Heath AC, Madden PA, Bucholz KK, Slutske WS, Nelson EC, Statham D, Whitfield JB, Martin NG (2004) Genetic effects on alcohol dependence risk: re-evaluating the importance of psychiatric and other heritable risk factors. Psychological medicine 34**:**1519-1530.

Kuryatov A, Berrettini W, Lindstrom J (2011) Acetylcholine receptor (AChR) alpha5 subunit variant associated with risk for nicotine dependence and lung cancer reduces (alpha4beta2)(2)alpha5 AChR function. Molecular pharmacology 79**:**119-125.

Li D, Zhao H, Gelernter J (2011) Strong association of the alcohol dehydrogenase 1B gene (ADH1B) with alcohol dependence and alcohol-induced medical diseases. Biological psychiatry 70:504-512.

Li D, Zhao H, Gelernter J (2012) Further clarification of the contribution of the ADH1C gene to vulnerability of alcoholism and selected liver diseases. Human genetics 131:1361-1374.

Li MD (2006) The genetics of nicotine dependence. Current psychiatry reports 8:158-164.

Liu JZ, Tozzi F, Waterworth DM, Pillai SG, Muglia P, Middleton L, Berrettini W, Knouff CW, Yuan X, Waeber G, Vollenweider P, Preisig M, Wareham NJ, Zhao JH, Loos RJ, Barroso I, Khaw KT, Grundy S, Barter P, Mahley R, Kesaniemi A, McPherson R, Vincent JB, Strauss J, Kennedy JL, Farmer A, McGuffin P, Day R, Matthews K, Bakke P, Gulsvik A, Lucae S, Ising M, Brueckl T, Horstmann S, Wichmann HE, Rawal R, Dahmen N, Lamina C, Polasek O, Zgaga L, Huffman J, Campbell S, Kooner J, Chambers JC, Burnett MS, Devaney JM, Pichard AD, Kent KM, Satler L, Lindsay JM, Waksman R, Epstein S, Wilson JF, Wild SH, Campbell H, Vitart V, Reilly MP, Li M, Qu L, Wilensky R, Matthai W, Hakonarson HH, Rader DJ, Franke A, Wittig M, Schafer A, Uda M, Terracciano A, Xiao X, Busonero F, Scheet P, Schlessinger D, St Clair D, Rujescu D, Abecasis GR, Grabe HJ, Teumer A, Volzke H, Petersmann A, John U, Rudan I, Hayward C, Wright AF, Kolcic I, Wright BJ, Thompson JR, Balmforth AJ, Hall AS, Samani NJ, Anderson CA, Ahmad T, Mathew CG, Parkes M, Satsangi J, Caulfield M, Munroe PB, Farrall M, Dominiczak A, Worthington J, Thomson W, Eyre S, Barton A, Mooser V, Francks C, Marchini J (2010) Meta-analysis and imputation refines the association of 15q25 with smoking quantity. Nature genetics 42:436-440.

Moolchan ET, Parzynski CS, Jaszyna-Gasior M, Collins CC, Leff MK, Zimmerman DL (2009) A link between adolescent nicotine metabolism and smoking topography. Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology 18:1578-1583.

O'Loughlin J, Paradis G, Kim W, DiFranza J, Meshefedjian G, McMillan-Davey E, Wong S, Hanley J, Tyndale RF (2004) Genetically decreased CYP2A6 and the risk of tobacco dependence: a prospective study of novice smokers. Tobacco control 13:422-428.

Olfson E, Bierut LJ (2012) Convergence of genome-wide association and candidate gene studies for alcoholism. Alcoholism, clinical and experimental research 36:2086-2094.

Olfson E, Edenberg HJ, Nurnberger J, Jr., Agrawal A, Bucholz KK, Almasy LA, Chorlian D, Dick DM, Hesselbrock VM, Kramer JR, Kuperman S, Porjesz B, Schuckit MA, Tischfield JA, Wang JC, Wetherill L, Foroud TM, Rice J, Goate A, Bierut LJ (2014) An ADH1B Variant and Peer Drinking in Progression to Adolescent Drinking Milestones: Evidence of a Gene-by-Environment Interaction. Alcoholism, clinical and experimental research.

Pillai SG, Ge D, Zhu G, Kong X, Shianna KV, Need AC, Feng S, Hersh CP, Bakke P, Gulsvik A, Ruppert A, Lodrup Carlsen KC, Roses A, Anderson W, Rennard SI, Lomas DA, Silverman EK, Goldstein DB (2009) A genome-wide association study in chronic obstructive pulmonary disease (COPD): identification of two major susceptibility loci. PLoS genetics 5:e1000421.

Rietschel M, Treutlein J (2013) The genetics of alcohol dependence. Annals of the New York Academy of Sciences 1282:39-70.

Rubinstein ML, Benowitz NL, Auerback GM, Moscicki AB (2008) Rate of nicotine metabolism and withdrawal symptoms in adolescent light smokers. Pediatrics 122:e643-647.

Rubinstein ML, Shiffman S, Moscicki AB, Rait MA, Sen S, Benowitz NL (2013) Nicotine metabolism and addiction among adolescent smokers. Addiction (Abingdon, England) 108:406-412.

Saccone SF, Hinrichs AL, Saccone NL, Chase GA, Konvicka K, Madden PA, Breslau N, Johnson EO, Hatsukami D, Pomerleau O, Swan GE, Goate AM, Rutter J, Bertelsen S, Fox L, Fugman D, Martin NG, Montgomery GW, Wang JC, Ballinger DG, Rice JP, Bierut LJ (2007) Cholinergic nicotinic receptor genes implicated in a nicotine dependence association study targeting 348 candidate genes with 3713 SNPs. Human molecular genetics 16:36-49.

Sherva R, Wilhelmsen K, Pomerleau CS, Chasse SA, Rice JP, Snedecor SM, Bierut LJ, Neuman RJ, Pomerleau OF (2008) 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 (Abingdon, England) 103:1544-1552.

Stahre M, Roeber J, Kanny D, Brewer RD, Zhang X (2014) Contribution of Excessive Alcohol Consumption to Deaths and Years of Potential Life Lost in the United States. Preventing Chronic Disease 11:E109.

Stevens VL, Bierut LJ, Talbot JT, Wang JC, Sun J, Hinrichs AL, Thun MJ, Goate A, Calle EE (2008) Nicotinic receptor gene variants influence susceptibility to heavy smoking. Cancer epidemiology, biomarkers & prevention : a publication of the American Association for

Cancer Research, cosponsored by the American Society of Preventive Oncology 17:3517-3525.

TAG (2010) Genome-wide meta-analyses identify multiple loci associated with smoking behavior. Nature genetics 42:441-447.

Thorgeirsson TE, Geller F, Sulem P, Rafnar T, Wiste A, Magnusson KP, Manolescu A, Thorleifsson G, Stefansson H, Ingason A, Stacey SN, Bergthorsson JT, Thorlacius S, Gudmundsson J, Jonsson T, Jakobsdottir M, Saemundsdottir J, Olafsdottir O, Gudmundsson LJ, Bjornsdottir G, Kristjansson K, Skuladottir H, Isaksson HJ, Gudbjartsson T, Jones GT, Mueller T, Gottsater A, Flex A, Aben KK, de Vegt F, Mulders PF, Isla D, Vidal MJ, Asin L, Saez B, Murillo L, Blondal T, Kolbeinsson H, Stefansson JG, Hansdottir I, Runarsdottir V, Pola R, Lindblad B, van Rij AM, Dieplinger B, Haltmayer M, Mayordomo JI, Kiemeney LA, Matthiasson SE, Oskarsson H, Tyrfingsson T, Gudbjartsson DF, Gulcher JR, Jonsson S, Thorsteinsdottir U, Kong A, Stefansson K (2008) A variant associated with nicotine dependence, lung cancer and peripheral arterial disease. Nature 452:638-642.

Thorgeirsson TE, Gudbjartsson DF, Surakka I, Vink JM, Amin N, Geller F, Sulem P, Rafnar T, Esko T, Walter S, Gieger C, Rawal R, Mangino M, Prokopenko I, Magi R, Keskitalo K, Gudjonsdottir IH, Gretarsdottir S, Stefansson H, Thompson JR, Aulchenko YS, Nelis M, Aben KK, den Heijer M, Dirksen A, Ashraf H, Soranzo N, Valdes AM, Steves C, Uitterlinden AG, Hofman A, Tonjes A, Kovacs P, Hottenga JJ, Willemsen G, Vogelzangs N, Doring A, Dahmen N, Nitz B, Pergadia ML, Saez B, De Diego V, Lezcano V, Garcia-Prats MD, Ripatti S, Perola M, Kettunen J, Hartikainen AL, Pouta A, Laitinen J, Isohanni M, Huei-Yi S, Allen M, Krestyaninova M, Hall AS, Jones GT, van Rij AM, Mueller T,

Dieplinger B, Haltmayer M, Jonsson S, Matthiasson SE, Oskarsson H, Tyrfingsson T, Kiemeney LA, Mayordomo JI, Lindholt JS, Pedersen JH, Franklin WA, Wolf H, Montgomery GW, Heath AC, Martin NG, Madden PA, Giegling I, Rujescu D, Jarvelin MR, Salomaa V, Stumvoll M, Spector TD, Wichmann HE, Metspalu A, Samani NJ, Penninx BW, Oostra BA, Boomsma DI, Tiemeier H, van Duijn CM, Kaprio J, Gulcher JR, McCarthy MI, Peltonen L, Thorsteinsdottir U, Stefansson K (2010) Sequence variants at CHRNB3-CHRNA6 and CYP2A6 affect smoking behavior. Nature genetics 42**:**448-453.

Wang JC, Kapoor M, Goate AM (2012) The genetics of substance dependence. Annual review of genomics and human genetics 13**:**241-261.

Weiss RB, Baker TB, Cannon DS, von Niederhausern A, Dunn DM, Matsunami N, Singh NA, Baird L, Coon H, McMahon WM, Piper ME, Fiore MC, Scholand MB, Connett JE, Kanner RE, Gahring LC, Rogers SW, Hoidal JR, Leppert MF (2008) A candidate gene approach identifies the CHRNA5-A3-B4 region as a risk factor for age-dependent nicotine addiction. PLoS genetics 4**:**e1000125.

WHO. Media center: tobacco. World health organization. Available at: http://www.who.int/mediacentre/factsheets/fs339/en/. Accessed October 16, 2014.

WHO. Media centre: Alcohol. World health organization. Available at: http://www.who.int/mediacentre/factsheets/fs349/en/. Accessed October 16, 2014.

Yu W, Gwinn M, Clyne M, Yesupriya A, Khoury MJ (2008) A navigator for human genome epidemiology. Nature genetics 40**:**124-125.

CHAPTER TWO:

Convergence of GWA and candidate gene studies for alcoholism

2.1 ABSTRACT

Background: Genome-wide association (GWA) studies have led to a paradigm shift in how researchers study the genetics underlying disease. Many GWA studies are now publicly available and can be used to examine whether or not previously proposed candidate genes are supported by GWA data. This approach is particularly important for the field of alcoholism because the contribution of many candidate genes remains controversial.

Methods: Using the Human Genome Epidemiology (HuGE) Navigator, we selected candidate genes for alcoholism that have been frequently examined in scientific articles in the past decade. Specific candidate loci as well as all the reported SNPs in candidate genes were examined in the Study of Alcohol Addiction: Genetics and Addiction (SAGE), a GWA study comparing alcohol dependent and non-dependent subjects.

Results: Several commonly reported candidate loci, including rs1800497 in *DRD2*, rs698 in *ADH1C,* rs1799971 in *OPRM1* and rs4680 in *COMT*, are not replicated in *SAGE* (p> .05). Among candidate loci available for analysis, only rs279858 in *GABRA2* (p=0.0052, OR=1.16) demonstrated a modest association. Examination of all SNPs reported in SAGE in over 50 candidate genes revealed no SNPs with large frequency differences between cases and controls and the lowest p value of any SNP was .0006.

Discussion: We provide evidence that several extensively studied candidate loci do not have a strong contribution to risk of developing alcohol dependence in European and African Ancestry

populations. Due to lack of coverage, we were unable to rule out the contribution of other variants and these genes and particular loci warrant further investigation. Our analysis demonstrates that publicly available GWA results can be used to better understand which if any of previously proposed candidate genes contribute to disease. Furthermore, we illustrate how examining the convergence of candidate gene and GWA studies can help elucidate the genetic architecture of alcoholism and more generally complex diseases.

2.2 INTRODUCTION

Genome-wide association (GWA) studies have revolutionized the search for common genetic variants that influence individual risk for complex diseases. Before this revolution, the discovery of genetic associations was dominated by candidate gene studies that used targeted gene approaches. Examination of these previous gene association studies demonstrates that most reported associations are not consistently replicated (Hirschhorn et al., 2002) and the strength of genetic associations in initial studies commonly erodes in subsequent research (Ioannidis et al., 2001). Despite this suggested irreproducibility, many candidate gene association studies continue to be published annually (Yu et al., 2008).

GWA studies rapidly evaluate millions of single nucleotide polymorphisms (SNPs) throughout the genome and therefore have the potential for identifying key variants in complex diseases. Since the publication of the first GWA study in 2005 (Klein et al., 2005), over 1000 GWA studies have established genetic associations of more than 200 traits, many of which are complex diseases. SNP-trait associations from published GWA studies are readily available to investigators through "A Catalog of Genome-Wide Association Studies" by the National Human Genome Research Institute (www.genome.gov/gwastudies). More recently, several datasets from GWA studies have also become available to the scientific community through the database of Genotypes and Phenotypes (dbGaP) maintained by NCBI (Mailman et al., 2007). These online scientific databases provide opportunities for investigators to access GWA data.

Online databases can specifically be used to evaluate whether genes that were previously suggested in candidate gene studies are replicated in GWA studies. Research by Siontis et al. demonstrates that only a few of previously proposed candidate loci of common diseases reached

genome-wide significance in GWA studies (Siontis et al., 2010). The loci that did replicate, however, had important genetic effects and included variants implicated in Alzheimer's disease and statin-induced myopathy. Similarly, a recent analysis by Obeidat et al. examined genetic associations with lung function measures to evaluate the role of previously associated genes in a large GWA study and clarified the role of many controversial associations (Obeidat et al., 2011). This approach of comparing candidate gene and GWA studies is powerful because it highlights which findings have consistent scientific evidence and therefore merit being pursued in future studies. These findings prompted us to examine whether proposed candidate genes associated with alcohol dependence are supported by GWA data.

Genetic and environmental factors contribute to individual susceptibility to alcohol dependence. Twin studies estimate that heritable influences explain 47-64% of variance in risk for alcohol dependence (Heath et al., 1997; Knopik et al., 2004). Several past research efforts have focused on targeted gene approaches to shed light on the genes that underlie these heritable influences. This has led to the proposal of hundreds of candidate genes that contribute to the development of alcohol dependence (Yu et al., 2008). A few GWA studies have also explored genes potentially involved in alcohol dependence (Bierut et al., 2010; Edenberg et al., 2010; Farrer et al., 2009; Heath et al., 2011; Hodgkinson et al., 2010; Joslyn et al., 2010; Treutlein et al., 2009; Zlojutro et al., 2011). Despite extensive candidate gene studies and several GWA studies, little consensus exists over which if any genes contribute to the genetic basis of alcohol dependence.

The existence of many controversial candidate genes for alcoholism highlights the need for further research on whether or not these genes replicate in large datasets. Results from the Study of Alcohol Addiction: Genetics and Addiction (SAGE) have recently become available

through dbGaP. SAGE compares DSM-IV alcohol dependent individuals and non-dependent, unrelated control subjects of European and African American descent. Using the SAGE data, we examined differences in SNP frequencies between cases and controls within previously reported candidate genes. These targeted candidate genes were selected using the Human Genome Epidemiology (HuGE) Navigator, a publicly searchable database established in 2001 of published genetic association and human genome epidemiological studies (Yu et al., 2008). The HuGE Navigator along with the SAGE results facilitated the systematic examination of candidate genes considered in many alcoholism studies over the last decade.

2.3 MATERIALS AND METHODS

Selection of Candidate Genes

The HuGE Navigator was developed using PubMed abstracts as the core data source and using data and text mining algorithms to develop a knowledge database (Yu et al., 2008). Each week since 2001, articles are systematically deposited in the database and represent a comprehensive list of recent articles. An automatic literature program screens PubMed for abstracts and then a genetic epidemiologist selects abstracts meeting inclusion criteria and indexes them. Phenopedia of the HuGE Navigator gives a disease-centered view of genetic association studies by providing information about genes studied in relation to a queried phenotype (Yu et al., 2010). Phenopedia was queried in July 2011 for Alcoholism and 584 genes were retrieved.

We focused our study on genes that have been frequently characterized by candidate gene studies. In primary analysis, over 90% of the genes associated with alcoholism in the HuGE database have 5 or fewer publications (528 out of 584 genes). The 56 candidate genes that have more than 5 publications vary substantially in the number of publications (6-103 publications). **Figure 2.1A** highlights that many genes have one or a few reported publications and there are some outliers that have been examined in many papers. This distribution may be explained in part by the fact that many of the genes with a low number of publications have been primarily identified in a GWA study and are not well characterized in targeted candidate gene studies. **Figure 2.1B** demonstrates that for almost 50% (176/386) of the genes with one publication that publication is itself a GWA study. Based on these preliminary observations, we narrowed our investigation to genes with more than 5 publications to focus our analysis on well-studied genes.

Since the X chromosome is not included in the publicly available SAGE results, the two candidate genes on the X chromosome *MAOA* and *HTR2C*, which have 26 and 8 publications respectively, were excluded from the analysis. The 54 autosomal genes that had more than 5 publications were pursued using the SAGE dataset. For the remainder of this paper, we will only refer to the 54 autosomal candidate genes.

SAGE Data

SAGE is a case-control study that analyzed genetic data on over 3,800 phenotyped subjects funded as part of the Gene Environment Association Studies (GENEVA) initiative supported by the National Human Genome Research Institute (Bierut et al., 2010). Alcohol-dependent cases and controls were selected from three large datasets: the Collaborative Study on the Genetic of Alcoholism (COGA), the Family Study of Cocaine dependence (FSCD) and the Collaborative Genetic Study of Nicotine Dependence (COGEND). Cases are required to have a lifetime history of DSM-IV alcohol dependence. Controls are required to have been exposed to alcohol because alcohol use is necessary to develop dependence, but not to have met lifetime diagnosis criteria for alcohol dependence or dependence for illicit drugs. A common assessment was performed for all cases and controls in the three studies that was based on the Semi-Structured Assessment for the Genetics of Alcoholism (SSAGA) (Bucholz et al., 1994). The common methodology of interview administration, question format and queried domains enabled phenotypic standardization across the three studies (Bierut et al., 2010). Characteristics of the cases and controls in the SAGE dataset are listed in **Table 2.1** and additional information is available at http://www.ncbi.nlm.nih.gov/projects/gap/cgi-

bin/study.cgi?study_id=phs000092.v1.p1. The SAGE dataset is publicly searchable through the Genome Brower under the Analysis tab on this website.

The power Calculator for Association with Two Stage design (CATS) was used to determine what effect sizes the SAGE dataset is able to detect (Skol et al., 2006). Using a sample size of 1900 cases and 1900 controls and an alpha level of .05, we calculated different allele frequencies and risk ratios.

Examination of SNPs in Candidate Genes

The HuGE database was used to survey articles on the ten candidate genes that had the most publications (listed in **Table 2.2**). The most well established loci based on expert opinion of the literature for each of the top ten candidate genes was searched in the genome browser to test whether candidate loci that had been highly reported in candidate gene studies replicated in the SAGE dataset. Since allele A9 for *SLC6A3* is a VNTR, we examined the two SNPs rs27072 and rs27048 as proxies because they have been found to be associated with similar withdrawal symptoms and are roughly in the same region of the gene as the VNTR (Le Strat et al., 2008). As the originators of the SAGE dataset, we were also able to compare the odds ratios and p values within the original three datasets (COGA, FSCD and COGEND) to verify whether there was any heterogeneity across the three contributing studies.

The 54 candidate genes with more than 5 publications were identified and chromosomal regions containing the gene plus 10 kb both 5' and 3' of the gene were examined. These expanded regions were searched using the SAGE genome browser and SNPs within these regions with $p<.05$ were recorded.

For each candidate gene, all SNPs with $p<.05$ were queried together in SNP Annotation and Proxy Search (SNAP) to assess linkage disequilibrium (Johnson et al., 2008). These searches were performed using the 1000 genomes pilot 1 SNP dataset, an $r^2 \geq .8$, and a distance limit of 500. This analysis was performed with both the CEU and YPI population panels separately because of allele frequency differences between European American and African American subsets. All SNPs that had an r^2 greater than 0.8 and at least one other variant in a group were considered a cluster.

2.4 RESULTS

The SAGE dataset contains half of the most commonly reported variants associated with the ten most well studied candidate genes (**Table 2.2**). Of the 5 candidate variants reported in SAGE, the only variant with a p< .05 is rs279858 in *GABRA2* (p=.0052, OR=1.16). The commonly reported variants, rs1800497 in *DRD2,* rs698 in *ADH1C*, rs1799971 in *OPRM1* and rs4680 in *COMT*, have p> .05. The minor allele for rs17999971 in *OPRM1* trends towards being protective (OR=.88) while the minor alleles of rs1800497 in *DRD2*, rs698 in *ADH1C* and rs4680 in COMT trend toward being associated with alcohol dependence (OR=1.11,1.08,1.02 respectively). The effects of these associations are in the expected direction based on previous candidate gene studies (Blum et al., 1990; Bond et al., 1998; Hendershot et al., 2011; Ponce et al., 2008; Thomasson et al., 1991; Tiihonen et al., 1999; Tolstrup et al., 2008; Zhang et al., 2006). In addition, these effects in SAGE were similar to the findings in the individual three studies that contributed to SAGE: COGA, FSCD and COGEND. Across the three contributing studies, the odds ratios ranged from 1.07-1.13 for rs1800497 in *DRD2*, 1.06-1.11 for rs698 in *ADH1C*, 0.82-0.95 for rs1799971 in *OPRM1*, 1.09-1.17 for rs279858 in *GABRA2* and 1.02-1.09 for rs4680 in *COMT* (data not shown).

Several commonly reported variants associated with alcoholism are not on the Illumina chip that was used to generate the SAGE dataset. These SNPs include rs671 in *ALDH2*, rs1229984 in *ADH1B*, rs4795541 in *SLC6A4* and rs3813867 in *CYP2E1*. Since the A9 allele in *SLC6A3* is a VNTR and therefore also not reported in SAGE, we examined two proxy SNPs (Le Strat et al., 2008). Neither of these two SNPs show a significant difference between the cases and controls (p=.8646 for rs27072 and p=.3842 for rs27048).

In every gene with more than 5 publications, few SNPs had impressive differences between cases and controls. Of the 2175 SNPs reported in the 54 genes with more than 5 publications, approximately 5% have a $p<.05$ (116/2175) and approximately 1% have a $p<.01$ (16/2175) (**Table 2.3**). The lowest p value of any variant was 0.0006 for rs925946, which is a SNP upstream of *BDNF*.

In a few genes, a large proportion of the SNPs have modest frequency differences between cases and controls. In 10 out of the 54 genes examined, more than 10% of the SNPs have $p<.05$ and in 3 genes this portion exceeds 20%. Specifically, the proportion of SNPs in SAGE with $p<.05$ is 55% (16/29) in *GABRA2*, 24% (10/29) in *BDNF* and 44% (4/9) in *HTR1A* (**Table 2.3**). To test whether the large proportion of SNPs with small p values in these genes could be explained by linkage disequilibrium, we performed SNAP analyses.

Many variants clustered as defined by $r^2> 0.8$ within the genes but the proportion of clusters containing SNPs with $p<.05$ remained quite similar with SNAP analyses in both CEU and YPI populations (data not shown). Of the variants with linkage disequilibrium data available in SNAP for the CEU population, 27 SNPs in *GABRA2* broke down into 10 clusters of which 5 clusters had SNPs with $p<.05$ (50%), 24 SNPs in *BDNF* broke down into 9 clusters of which 4 clusters had SNPs with $p<.05$ (44%), and 6 SNPs in *HTR1A* broke down into 3 clusters of which 1 cluster had SNPs with $p<.05$ (33%). Generally fewer SNPs clustered in the YRI population than in the CEU population but the proportion of clusters containing SNPs with $p<.05$ was comparable between the two populations. In the YRI population, 25 SNPs in *GABRA2* broke down into 16 clusters of which 12 clusters had SNPs with $p<.05$ (75%), 25 SNPs in *BDNF* broke down into 19 clusters of which 10 clusters had SNPs with $p<.05$ (53%), and 9 SNPs in *HTR1A* broke down into 5 clusters of which 2 cluster had SNPs with $p<.05$ (40%).

Power calculations demonstrate that the SAGE dataset has 90% power with an alpha level of .05 to detect a genetic variant with a minor allele frequency of .10 and an odds ratio of 1.25 or greater. The dataset also has 90% power with an alpha level of .05 to detect a variant with a minor allele frequency of .40 and an odds ratio of 1.15 or greater.

2.5 DISCUSSION

Over the last decade, hundreds of candidate genes have been proposed for alcoholism. We used local and global approaches to specifically investigate variants within the most widely studied of previously proposed candidate genes. Our primary finding is that most of these candidate genes are not strongly supported by GWA data. This observation reduces the likelihood that these previously proposed genes individually have a strong effect on the genetic risk of alcohol dependence. The results mirror prior work that most candidate loci in common diseases are not strongly replicated in GWA studies except for a few biologically important variants (Siontis et al., 2010; Obeidat et al., 2011).

Analysis of well-characterized loci that were previously proposed in candidate gene studies in a large GWA study on alcoholism, SAGE, reveals unimpressive differences between cases and controls at most loci. The frequently studied variants associated with alcoholism in *DRD2*, *ADH1C*, *OPRM1* and *COMT* demonstrate insignificant frequency differences in SAGE ($p>.05$, **Table 2.2**). Although several studies implicate a biological role of these variants in alcoholism (Blum et al., 1990; Bond et al., 1998; Hendershot et al., 2011; Ponce et al., 2008; Thomasson et al., 1991; Tiihonen et al., 1999; Tolstrup et al., 2008; Zhang et al., 2006), our results reveal that these variants are not strongly associated with alcoholism in European and African ancestry populations. The only candidate that modestly replicated in SAGE, rs279858 in *GABRA2*, had a p-value of 0.0052 (OR=1.572). This finding was anticipated because a previous GWA study on the SAGE dataset demonstrated a similar association (Bierut et al., 2010). The replication of rs279858 in SAGE provides some support for future studies focused on the function of this variant and associated variants in *GABRA2* (Edenberg et al., 2004).

When examined globally, none of the well-studied candidate genes demonstrate impressive variant differences between cases and controls. More specifically, only one SNP reported in SAGE (rs925946 upstream of *BDNF*, p=0.0006) in the 54 candidate genes had a p value less than 0.0009, a corrected p value for the number of genes (.05/54= 0.0009). Additionally, the overall number of variants with $p<.05$ and $p<.01$ is close to that predicted by chance considering the total number of SNPs examined in all proposed candidate genes. Although the individual p values for variants in the examined candidate genes are modest, a few candidate genes have a large portion of SNPs with $p<.05$ (**Table 2.3**). The results support further research into whether *GABRA2*, which was the candidate gene with the largest proportion of SNPs with $p< .05$ (55%), contributes to risk of developing alcohol dependence. *BDNF* and *HTR1A* also had more than one fifth of SNPs with $p<.05$, indicating that these genes merit further investigation to elucidate their potential contribution to alcohol dependence.

Lack of replication in SAGE does not exclude the possibility that some previously proposed candidate genes and specific loci are biologically important. Several of the most well studied candidate loci for alcoholism were not available in SAGE, including rs671 in *ALDH2*, rs1229984 in *ADH1B*, rs4795541 in *SLC6A4* and rs3813867 in *CYP2E1*. A recent study that specifically genotyped rs1229984 in SAGE reported that the minor allele has a significant protective effect on alcohol dependence ($p=6.6\times 10^{-10}$) (Bierut et al., 2011). Because rs1229984 is common in Asians but rare in European Americans, this variant in *ADH1B* was not genotyped in the original GWA study. This highlights that GWA studies may miss important variants because of lack of coverage of SNPs that are uncommon in European American populations. Additionally, GWA studies cannot assess all forms of inheritance that can be associated with candidate genes such as insertion/deletion mutations, copy number repeats and epigenetic

changes. Although SAGE is a valuable tool, it cannot exclude the possibility that aspects of genes contribute to genetic risk of alcohol dependence.

Even though the well studied candidate variants in *DRD2*, *ADH1C*, *OPRM1* and *COMT* were not significantly associated with alcohol dependence in SAGE, their odds ratios were in the expected direction based on previous candidate gene studies. More specifically, the odds ratio of 0.088 for rs1799971 in *OPRM1* supports previous studies that the minor allele variant is protective against alcohol dependence (Bond et al., 1998; Zhang et al., 2006) while the odds ratios greater than 1 for rs1800497 in *DRD2*, rs698 in *ADH1C*, and rs4680 in *COMT* supports previous studies that the minor allele of these variants are more common in alcohol dependent individuals (Blum et al., 1990; Hendershot et al., 2011; Ponce et al., 2008; Thomasson et al., 1991; Tiihonen et al., 1999; Tolstrup et al., 2008). The fact that these odds ratios are in the expected direction but did not pass a threshold of .05 for significance may suggest that these variants have a small contribution to alcohol dependence and this study lacked the power to detect the association.

Our study design had several strengths. First, the literature search for candidate genes included all genetic associations irrespective of ethnicity and criteria for alcoholism. By including all genes with the most genetic association study publications, we comprehensively examined previously identified genes associated with alcoholism in a large GWA study on alcoholism. Second, the SAGE dataset has the power to detect associations of small magnitude. SAGE included more than 3,800 subjects and had 90% power to detect a genetic variant with an odds ratio of 1.25 for a risk locus with 10% minor allele frequency. Third, our findings in SAGE regarding the well-characterized loci were found to be very similar to the results in the three independent datasets that contributed to SAGE, which indicates that there is no heterogeneity

across these datasets. Fourth, our approach used data that is available to the scientific community and can be easily replicated in future studies of other phenotypes.

Despite these strengths, the selection of candidate loci and genes based on number of publications retrieved by the HuGE Navigator Phenopedia has some limitations. One limitation is that no data suggests that the potential significance of a given gene is directly proportional to the number of publications. Despite this, we felt that the number of publications is an indicator of research efforts devoted to a given gene. By selecting genes with the most publications, we sought to capture well-studied genes that had been the focus of the field in the past. A second potential limitation is that we did not exclude publications based on the same datasets. Because we used a low threshold of greater than 5 publications in the initial analysis, however, we are confident that we did not exclude any genes that have been examined in many studies. Additionally, the most well-studied loci of the ten genes with the most publications were selected based on expert opinion and were felt to be unambiguously widely studied even if the exact order may not be reflective of the number of data sets published on the genes.

Beyond limitations in our selection of candidate genes, the SAGE dataset has limitations that restrict the interpretation of our results. First, some of the most well studied variants were not covered in SAGE and therefore could not be assessed. Second, the X chromosome is not included in the publically available SAGE results so we were unable to investigate genes on the X chromosome. Specifically, two candidate genes on the X chromosome, *MAOA* and *HTR2C* that had 26 and 8 publications respectively, were not assessed. Third, SAGE is limited in its power to identify genotyped variants on the GWA chip that have small effect sizes. Despite the fact that the SAGE dataset was relatively large when it was originally published, identifying common variants with small effect sizes (<1.1) remains challenging and we are unable to rule out

the possibility of real but modest effects of these genes. Forth, variants that are uncommon (1%-5%) or rare (<1%) in the study population may also not be detected in SAGE because of their individually small contribution to overall alcoholism. Fifth, the SAGE dataset primarily consists of European Americans (69.5%), African Americans (30.3%) and a few Hispanics (3.4%) (**Table 2.1**) and association findings may be different in other populations such as Asians. Some of the genes and variants examined in this analysis are more well studied and have a higher frequency in Asian Ancestry than in European and African Ancestry populations, such as the Asp40 allele of the candidate variant rs1799971 in *OPRM1* (Arias et al., 2006), and therefore may have a more impressive effect in studies that focus on Asian ancestry populations. Sixth, our analysis did not examine the effects of combinations of genes or the effect of different environmental factors. Analysis of multiple genes and populations enriched for specific environmental risk factors will likely explain a greater degree of the genetic risk of alcoholism. Despite these limitations, this analysis demonstrates that GWA studies are a powerful technique for verifying the importance of genes and particular variants that have been previously identified in the candidate gene era.

In summary, we provide evidence that for alcohol dependence, several extensively studied candidate loci and genes are not replicated in a large GWA study, indicating that these variants do not individually have a large contribution to risk of developing alcohol dependence in European and African ancestry populations. Our analysis was unable to rule out the possibility that some variants and genes are important for risk of alcoholism due to lack of coverage. Recent work demonstrates that at least one highly reported variant rs1229984 in *ADH1B* that is not reported in SAGE is significantly associated with alcoholism (Bierut et al., 2011), suggesting the possible importance of further research on highly supported variants that cannot be assessed in

SAGE. Our approach may also have missed variants that have a real but small individual contribution to overall inheritance of alcoholism.

This analysis demonstrates that targeted candidate gene studies and GWA studies each provide important information and studying the convergence of these two experimental designs has the potential to advance understanding of the etiology of alcohol dependence and more generally complex diseases. While GWA studies provide important information about the genetic contribution of common variants to complex diseases across populations, hypothesis driven candidate gene studies are also important to assess variants of lesser significance that may be missed because of the strict p value thresholds required for the large number of comparisons in GWA studies. Incorporating knowledge from both GWA and candidate gene studies will help clarify the role of genetics in complex disease and guide future research.

Our study also shows how the HuGE Navigator and dbGaP databases can be used as tools by researchers to easily access and analyze information on candidate genes and GWA data. Beyond alcoholism, the HuGE Navigator provides an easy way for investigators to search over 2,000 diseases and 10,000 genes for summary information and primary articles about genetic associations and human genome epidemiology (Yu et al., 2008). The dbGaP database provides access to results of over 100 studies examining phenotype and genotype associations, including 40 GWAS studies on different diseases. Since dbGaP currently contains a limited number of GWA studies, researchers examining phenotypes not available in dbGaP may benefit from directly contacting the authors of relevant GWA studies and meta analysis. Because of this easy accessibility, researchers who intend to perform future candidate gene studies should reference the HuGE navigator to assess background information and use dbGaP and existing GWA data to test whether their gene of interest is replicated in GWA data. Candidate gene studies need

replication to meet scientific standards. Simple dbGaP analyses may help to focus future research on genes that are supported by GWA data and therefore more likely to be biologically important for human disease.

2.6 ACKNOWLEDGEMENTS

I would like to acknowledge my co-author Laura J Bierut.

We thank Nancy Saccone for comments and criticisms and Louis Fox for technical assistance. This work was supported by a grant from the National Institute on Drug Abuse K02 DA021237 to Laura J Bierut. Emily Olfson was supported by T32 GM007200: National Research Service Award Medical Scientist Training Grant from the National Institute of General Medical Sciences.

2.7 REFERENCES

Arias A, Feinn R, Kranzler HR (2006) Association of an Asn40Asp (A118G) polymorphism in the mu-opioid receptor gene with substance dependence: a meta-analysis. Drug Alcohol Depend 83(3):262-8.

Bierut LJ, Agrawal A, Bucholz KK, Doheny KF, Laurie C, Pugh E, Fisher S, Fox L, Howells W, Bertelsen S, Hinrichs AL, Almasy L, Breslau N, Culverhouse RC, Dick DM, Edenberg HJ, Foroud T, Grucza RA, Hatsukami D, Hesselbrock V, Johnson EO, Kramer J, Krueger RF, Kuperman S, Lynskey M, Mann K, Neuman RJ, Nothen MM, Nurnberger JI, Jr., Porjesz B, Ridinger M, Saccone NL, Saccone SF, Schuckit MA, Tischfield JA, Wang JC, Rietschel M, Goate AM, Rice JP (2010) A genome-wide association study of alcohol dependence. Proc Natl Acad Sci U S A 107(11):5082-7.

Bierut LJ, Goate AM, Breslau N, Johnson EO, Bertelsen S, Fox L, Agrawal A, Bucholz KK, Grucza R, Hesselbrock V, Kramer J, Kuperman S, Nurnberger J, Porjesz B, Saccone NL, Schuckit M, Tischfield J, Wang JC, Foroud T, Rice JP, Edenberg HJ (2011) ADH1B is associated with alcohol dependence and alcohol consumption in populations of European and African ancestry. Mol Psychiatry.

Blum K, Noble EP, Sheridan PJ, Montgomery A, Ritchie T, Jagadeeswaran P, Nogami H, Briggs AH, Cohn JB (1990) Allelic association of human dopamine D2 receptor gene in alcoholism. JAMA 263(15):2055-60.

Bond C, LaForge KS, Tian M, Melia D, Zhang S, Borg L, Gong J, Schluger J, Strong JA, Leal SM, Tischfield JA, Kreek MJ, Yu L (1998) Single-nucleotide polymorphism in the

human mu opioid receptor gene alters beta-endorphin binding and activity: possible implications for opiate addiction. Proc Natl Acad Sci U S A 95(16):9608-13.

Bucholz KK, Cadoret R, Cloninger CR, Dinwiddie SH, Hesselbrock VM, Nurnberger JI, Jr., Reich T, Schmidt I, Schuckit MA (1994) A new, semi-structured psychiatric interview for use in genetic linkage studies: a report on the reliability of the SSAGA. J Stud Alcohol 55(2):149-58.

Edenberg HJ, Dick DM, Xuei X, Tian H, Almasy L, Bauer LO, Crowe RR, Goate A, Hesselbrock V, Jones K, Kwon J, Li TK, Nurnberger JI, Jr., O'Connor SJ, Reich T, Rice J, Schuckit MA, Porjesz B, Foroud T, Begleiter H (2004) Variations in GABRA2, encoding the alpha 2 subunit of the GABA(A) receptor, are associated with alcohol dependence and with brain oscillations. Am J Hum Genet 74(4):705-14.

Edenberg HJ, Koller DL, Xuei X, Wetherill L, McClintick JN, Almasy L, Bierut LJ, Bucholz KK, Goate A, Aliev F, Dick D, Hesselbrock V, Hinrichs A, Kramer J, Kuperman S, Nurnberger JI, Jr., Rice JP, Schuckit MA, Taylor R, Todd Webb B, Tischfield JA, Porjesz B, Foroud T (2010) Genome-wide association study of alcohol dependence implicates a region on chromosome 11. Alcohol Clin Exp Res 34(5):840-52.

Farrer LA, Kranzler HR, Yu Y, Weiss RD, Brady KT, Anton R, Cubells JF, Gelernter J (2009) Association of variants in MANEA with cocaine-related behaviors. Arch Gen Psychiatry 66(3):267-74.

Heath AC, Bucholz KK, Madden PA, Dinwiddie SH, Slutske WS, Bierut LJ, Statham DJ, Dunne MP, Whitfield JB, Martin NG (1997) Genetic and environmental contributions to alcohol dependence risk in a national twin sample: consistency of findings in women and men. Psychol Med 27(6):1381-96.

Heath AC, Whitfield JB, Martin NG, Pergadia ML, Goate AM, Lind PA, McEvoy BP, Schrage AJ, Grant JD, Chou YL, Zhu R, Henders AK, Medland SE, Gordon SD, Nelson EC, Agrawal A, Nyholt DR, Bucholz KK, Madden PA, Montgomery GW (2011) A Quantitative-Trait Genome-Wide Association Study of Alcoholism Risk in the Community: Findings and Implications. Biol Psychiatry.

Hendershot CS, Lindgren KP, Liang T, Hutchison KE (2011) COMT and ALDH2 polymorphisms moderate associations of implicit drinking motives with alcohol use. Addict Biol.

Hirschhorn JN, Lohmueller K, Byrne E, Hirschhorn K (2002) A comprehensive review of genetic association studies. Genet Med 4(2):45-61.

Hodgkinson CA, Enoch MA, Srivastava V, Cummins-Oman JS, Ferrier C, Iarikova P, Sankararaman S, Yamini G, Yuan Q, Zhou Z, Albaugh B, White KV, Shen PH, Goldman D (2010) Genome-wide association identifies candidate genes that influence the human electroencephalogram. Proc Natl Acad Sci U S A 107(19):8695-700.

Ioannidis JP, Ntzani EE, Trikalinos TA, Contopoulos-Ioannidis DG (2001) Replication validity of genetic association studies. Nat Genet 29(3):306-9.

Johnson AD, Handsaker RE, Pulit SL, Nizzari MM, O'Donnell CJ, de Bakker PI (2008) SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap. Bioinformatics 24(24):2938-9.

Joslyn G, Ravindranathan A, Brush G, Schuckit M, White RL (2010) Human variation in alcohol response is influenced by variation in neuronal signaling genes. Alcohol Clin Exp Res 34(5):800-12.

Klein RJ, Zeiss C, Chew EY, Tsai JY, Sackler RS, Haynes C, Henning AK, SanGiovanni JP, Mane SM, Mayne ST, Bracken MB, Ferris FL, Ott J, Barnstable C, Hoh J (2005) Complement factor H polymorphism in age-related macular degeneration. Science 308(5720):385-9.

Knopik VS, Heath AC, Madden PA, Bucholz KK, Slutske WS, Nelson EC, Statham D, Whitfield JB, Martin NG (2004) Genetic effects on alcohol dependence risk: re-evaluating the importance of psychiatric and other heritable risk factors. Psychol Med 34(8):1519-30.

Le Strat Y, Ramoz N, Pickering P, Burger V, Boni C, Aubin HJ, Ades J, Batel P, Gorwood P (2008) The 3' part of the dopamine transporter gene DAT1/SLC6A3 is associated with withdrawal seizures in patients with alcohol dependence. Alcohol Clin Exp Res 32(1):27-35.

Mailman MD, Feolo M, Jin Y, Kimura M, Tryka K, Bagoutdinov R, Hao L, Kiang A, Paschall J, Phan L, Popova N, Pretel S, Ziyabari L, Lee M, Shao Y, Wang ZY, Sirotkin K, Ward M, Kholodov M, Zbicz K, Beck J, Kimelman M, Shevelev S, Preuss D, Yaschenko E, Graeff A, Ostell J, Sherry ST (2007) The NCBI dbGaP database of genotypes and phenotypes. Nat Genet 39(10):1181-6.

Obeidat M, Wain LV, Shrine N, Kalsheker N, Soler Artigas M, Repapi E, Burton PR, Johnson T, Ramasamy A, Zhao JH, Zhai G, Huffman JE, Vitart V, Albrecht E, Igl W, Hartikainen AL, Pouta A, Cadby G, Hui J, Palmer LJ, Hadley D, McArdle WL, Rudnicka AR, Barroso I, Loos RJ, Wareham NJ, Mangino M, Soranzo N, Spector TD, Glaser S, Homuth G, Volzke H, Deloukas P, Granell R, Henderson J, Grkovic I, Jankovic S, Zgaga

L, Polasek O, Rudan I, Wright AF, Campbell H, Wild SH, Wilson JF, Heinrich J, Imboden M, Probst-Hensch NM, Gyllensten U, Johansson A, Zaboli G, Mustelin L, Rantanen T, Surakka I, Kaprio J, Jarvelin MR, Hayward C, Evans DM, Koch B, Musk AW, Elliott P, Strachan DP, Tobin MD, Sayers I, Hall IP (2011) A comprehensive evaluation of potential lung function associated genes in the SpiroMeta general population sample. PLoS One 6(5):e19382.

Ponce G, Hoenicka J, Jimenez-Arriero MA, Rodriguez-Jimenez R, Aragues M, Martin-Sune N, Huertas E, Palomo T (2008) DRD2 and ANKK1 genotype in alcohol-dependent patients with psychopathic traits: association and interaction study. Br J Psychiatry 193(2):121-5.

Siontis KC, Patsopoulos NA, Ioannidis JP (2010) Replication of past candidate loci for common diseases and phenotypes in 100 genome-wide association studies. Eur J Hum Genet 18(7):832-7.

Skol AD, Scott LJ, Abecasis GR, Boehnke M (2006) Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. Nat Genet 38(2):209-13.

Thomasson HR, Edenberg HJ, Crabb DW, Mai XL, Jerome RE, Li TK, Wang SP, Lin YT, Lu RB, Yin SJ (1991) Alcohol and aldehyde dehydrogenase genotypes and alcoholism in Chinese men. Am J Hum Genet 48(4):677-81.

Tiihonen J, Hallikainen T, Lachman H, Saito T, Volavka J, Kauhanen J, Salonen JT, Ryynanen OP, Koulu M, Karvonen MK, Pohjalainen T, Syvalahti E, Hietala J (1999) Association between the functional variant of the catechol-O-methyltransferase (COMT) gene and type 1 alcoholism. Mol Psychiatry 4(3):286-9.

Tolstrup JS, Nordestgaard BG, Rasmussen S, Tybjaerg-Hansen A, Gronbaek M (2008) Alcoholism and alcohol drinking habits predicted from alcohol dehydrogenase genes. Pharmacogenomics J 8(3):220-7.

Treutlein J, Cichon S, Ridinger M, Wodarz N, Soyka M, Zill P, Maier W, Moessner R, Gaebel W, Dahmen N, Fehr C, Scherbaum N, Steffens M, Ludwig KU, Frank J, Wichmann HE, Schreiber S, Dragano N, Sommer WH, Leonardi-Essmann F, Lourdusamy A, Gebicke-Haerter P, Wienker TF, Sullivan PF, Nothen MM, Kiefer F, Spanagel R, Mann K, Rietschel M (2009) Genome-wide association study of alcohol dependence. Arch Gen Psychiatry 66(7):773-84.

Yu W, Clyne M, Khoury MJ, Gwinn M (2010) Phenopedia and Genopedia: disease-centered and gene-centered views of the evolving knowledge of human genetic associations. Bioinformatics 26(1):145-6.

Yu W, Gwinn M, Clyne M, Yesupriya A, Khoury MJ (2008) A navigator for human genome epidemiology. Nat Genet 40(2):124-5.

Zhang H, Luo X, Kranzler HR, Lappalainen J, Yang BZ, Krupitsky E, Zvartau E, Gelernter J (2006) Association between two mu-opioid receptor gene (OPRM1) haplotype blocks and drug or alcohol dependence. Hum Mol Genet 15(6):807-19.

Zlojutro M, Manz N, Rangaswamy M, Xuei X, Flury-Wetherill L, Koller D, Bierut LJ, Goate A, Hesselbrock V, Kuperman S, Nurnberger J, Jr., Rice JP, Schuckit MA, Foroud T, Edenberg HJ, Porjesz B, Almasy L (2011) Genome-wide association study of theta band event-related oscillations identifies serotonin receptor gene HTR7 influencing risk of alcohol dependence. Am J Med Genet B Neuropsychiatr Genet 156(1):44-58.

2.8 TABLES

Table 2.1. Characteristics of alcohol dependent cases and non-dependent controls in SAGE

Characteristic	Cases n=1,897	Controls n=1,932	Total n=3,829
Sex, n (%)			
Males	1.155 (60.9)	606 (31.4)*	1,761 (46.0)
Females	742 (39.1)	1,326 (68.6)	2,068 (54.0)
Age, years			
Mean ± SD	39.0 ± 9.3	39.3 ± 9.1	19.2 ± 9.2
Range	18.0-77.0	18.0-65.0	18.0-77.0
Self-reported race, n (%)			
European-American	1,235 (65.1)	1,433 (74.2)*	2,668 (69.5)
African-American	662 (34.9)	499 (25.8)	1,161 (30.3)
Self-reported ethnicity, n (%)			
Hispanic	76 (4.0)	56 (2.8)	132 (3.4)
Alcohol dependence			
Diagnosis, n (%)	1,897 (100.0)	0 (0.0)*	1,897 (49.5)

Sex, age, race, ethnicity and alcohol dependence characteristics of cases and controls in the Study of Alcohol Addiction: Genetics and Addiction (SAGE) dataset (Bierut et al., 2010).

* $p<0.0001$ for difference between cases and controls

Table 2.2. Examination of Top Ten Candidate Loci for Alcoholism in SAGE

Candidate Genes	Publications on Alcoholism association	Commonly reported SNP	Common Name of SNP	P value in SAGE	Odds Ratio in SAGE (CI)
ALDH2	103	rs671	ALDH2*2 (Harada, 1982)	-	-
ADH1B	89	rs1229984	ADH1B*2/ADH2*2 (Thomasson, 1992)	-	-
DRD2	83	rs1800497	TaqIA (Blum, 1990)	0.09	1.1053 (.9845-1.2408)
SLC6A4	83	rs4795541	5-HTTLPR/S allele (Sander, 1997)	-	-
ADH1C	51	rs698	ADH1C*2 (Thomasson, 1992)	0.1452	1.0819 (.9732-1.2028)
OPRM1	38	rs1799971	Ala118Gly (Bond, 1998)	0.1372	.8823 (.7481-1.0407)
CYP2E1	35	rs3813867	CYPE1*c2 (Hayashi, 1991)	-	-
GABRA2	27	rs279858*	(Edenberg, 2004)	0.0052	1.1572 (1.0445-1.2821)
COMT	26	rs4680	Val158Met (Tiihonen, 1999)	0.6328	1.0244 (.9278-1.1311)
SLC6A3	25	**	A9 (VNTR) (Dobashi, 1997)		

In the ten most frequently studied genes associated with alcoholism according to the Human Genome Epidemiology (HuGE) Navigator, the most well studied variants were examined in Study of Alcohol Addiction: Genetics and Addiction (SAGE).

*One of over 20 SNPs significantly associated with alcohol dependence (Edenberg et al., 2004). This SNP was examined because it was the only one in an exon.

**Examined 2 SNPs, rs27072 and rs27048, as proxies (Le Strat et al., 2008)

Table 2.3. SNPs in frequently studied candidate genes associated with Alcoholism

Candidate Genes	Publications on Alcoholism	SNPs recorded in dbSNP	Total SNPs in SAGE	SNPs in SAGE			
				p<.05	p<.01	p<.005	p<.001
ALDH2	103	407	24	-	-	-	-
ADH1B	89	337	21	-	-	-	-
DRD2	83	826	41	7	1	-	-
SLC6A4	83	637	21	2	-	-	-
ADH1C	51	522	29	1	-	-	-
OPRM1	38	3568	122	4	-	-	-
CYP2E1	35	210	49	-	-	-	-
GABRA2	27	1692	29	16	5	1	-
COMT	26	752	55	-	-	-	-
SLC6A3	25	1322	38	-	-	-	-
HTR2A	22	1036	61	8	-	-	-
HTR1B	18	60	12	2	1	-	-
DRD4	18	184	9	-	-	-	-
BDNF	16	624	29	10	2	2	1
NPY	15	241	19	2	-	-	-
DRD3	14	693	31	3	-	-	-
APOE	13	106	12	1	-	-	-
MTHFR	13	324	49	-	-	-	-
GABRA6	13	215	18	-	-	-	-
TPH1	13	277	14	-	-	-	-
GRIN2B	12	5233	245	11	4	3	-
CNR1	12	2778	27	5	2	-	-
TPH2	11	1415	54	1	-	-	-
ADH4	10	830	62	1	-	-	-
CHRM2	10	1992	62	3	-	-	-
CRHR1	9	1183	26	1	-	-	-
ANKK1	9	218	26	4	-	-	-
ALDH1A1	9	739	122	3	-	-	-
DRD1	9	82	19	-	-	-	-
GABRG2	9	991	30	-	-	-	-
GABRB2	9	2699	80	1	-	-	-
HTR1A	9	45	9	4	-	-	-
GSTM1	9	123	3	-	-	-	-
OPRD1	9	585	20	1	-	-	-
OPRK1	9	226	37	-	-	-	-
GABRB3	8	23	104	9	1	-	-
GABRA1	8	583	20	-	-	-	-
DBH	8	562	47	1	-	-	-
ADH1A	8	293	19	2	-	-	-
ADH5	7	584	38	-	-	-	-
GAD1	7	692	31	1	-	-	-
HFE	7	188	28	1	-	-	-
GRIN1	6	513	18	-	-	-	-
GAD2	6	784	49	1	-	-	-
GABRB1	6	4354	111	2	-	-	-
ADH7	6	288	32	-	-	-	-
ADRA2A	6	46	9	-	-	-	-
CHRNA5	6	254	16	-	-	-	-
POMC	6	102	10	-	-	-	-
SLC6A2	6	837	49	1	-	-	-
CCKBR	6	300	20	4	-	-	-
CCKAR	6	167	15	-	-	-	-
TNF	6	177	36	2	-	-	-
CCK	6	461	18	1	-	-	-
Total (n=54)			2175	116	16	6	1

All genes with more than 5 publications in Human Genome Epidemiology (HuGE) Navigator were examined in Study of Alcohol Addiction: Genetics and Addiction (SAGE). Genes were expanded by 10 kb on both sides before they were queried in the SAGE database. *MAOA* and *HTR2C* were excluded because they were on the X chromosome.

2.9 FIGURES

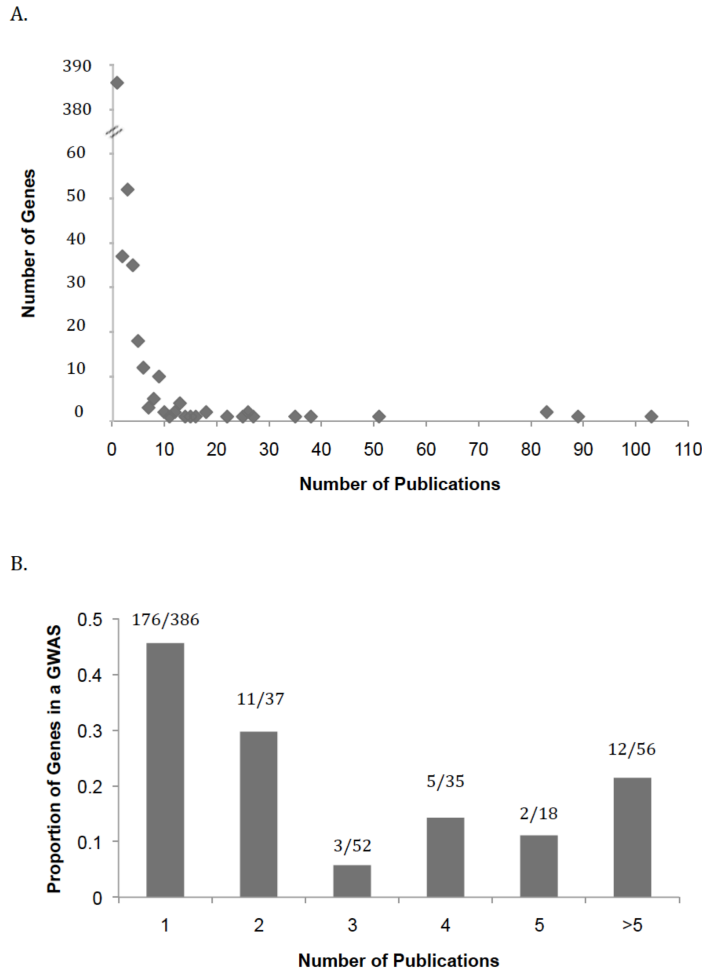


Figure 2.1. Characteristics of genes associated with Alcoholism in the Human Genome Epidemiology (HuGE) Navigator (**A**) Distribution of number of publications on genes; (**B**) Proportion of genes observed in at least one GWA study stratified based on number of publications. 4 genes were identified in 2 GWA study (these genes had 2, 2, 10 and 51 publications). All other genes were found in 1 or none GWA study. A total of 8 GWA studies on Alcoholism are listed in the HuGE database (Bierut et al., 2010; Edenberg et al., 2010; Farrer et al., 2009; Heath et al., 2011; Hodgkinson et al., 2010; Joslyn et al., 2010; Treutlein et al., 2009; Zlojutro et al., 2011).

CHAPTER THREE:

An *ADH1B* variant and peer drinking in progression to adolescent drinking milestones:

Evidence of a gene-by-environment interaction

3.1 ABSTRACT

Background: Adolescent drinking is an important public health concern, one that is influenced by both genetic and environmental factors. The functional variant rs1229984 in alcohol dehydrogenase 1B (*ADH1B*) has been associated at a genome-wide level with alcohol use disorders in diverse adult populations. However, few data are available regarding whether this variant influences early drinking behaviors and whether social context moderates this effect. This study examines the interplay between rs1229984 and peer drinking in the development of adolescent drinking milestones.

Methods: 1,550 European and African American individuals who had a full drink of alcohol before age 18 were selected from a longitudinal study of youth as part of the Collaborative Study on the Genetics of Alcoholism (COGA). Cox proportional hazards regression, with GxE product terms in the final models, was used to study two primary outcomes during adolescence: age of first intoxication and age of first DSM-5 alcohol use disorder symptom.

Results: The minor A allele of rs1229984 was associated with a protective effect for first intoxication (HR=0.56, 95% CI 0.41-0.76) and first DSM-5 symptom (HR=0.45, 95% CI 0.26-0.77) in the final models. Reporting that most or all best friends drink was associated with a hazardous effect for first intoxication (HR=1.81, 95% CI 1.62-2.01) and first DSM-5 symptom (HR=2.17, 95% 1.88-2.50) in the final models. Furthermore, there was a significant GxE interaction for first intoxication (p=.002) and first DSM-5 symptom (p=.01). Among individuals reporting none or few best friends drinking, the *ADH1B* variant had a protective effect for

adolescent drinking milestones, but for those reporting most or all best friends drinking, this effect was greatly reduced.

Conclusions: Our results suggest that the risk factor of best friends drinking attenuates the protective effect of a well-established *ADH1B* variant for two adolescent drinking behaviors. These findings illustrate the interplay between genetic and environmental factors in the development of drinking milestones during adolescence.

3.2 INTRODUCTION

By age 17, most U.S. adolescents (54%-78%) have consumed alcohol, and a significant proportion (15%) meet the criteria for alcohol abuse (Merikangas et al., 2010; NSDUH, 2012; Swendsen et al., 2012). Patterns of alcohol use that begin in adolescence are important determinants for the development of alcohol use disorders during adulthood (Grant et al., 2006; Pitkanen et al., 2005). Therefore, understanding factors that contribute to early drinking behaviors is critical for disease prevention.

For decades, twin studies have recognized that both genetic and environmental factors influence individual risk for alcoholism (Heath et al., 1997; Kendler et al., 1994; Pickens et al., 1991; Prescott and Kendler, 1999). Recently, large-scale genetic studies have provided strong evidence for the contribution of specific genetic variants to alcohol use disorders in adults (Rietschel and Treutlein, 2013; Wang et al., 2012). An important next step in the translation of genetic findings identified in adults is to test whether these genetic variants also affect adolescent drinking behaviors and whether environmental risk factors moderate this role.

Among the most biologically well-understood genetic variants associated with alcohol use disorders is the polymorphism rs1229984 in the enzyme alcohol dehydrogenase 1B (*ADH1B*). The minor A allele (in the coding strand) of rs1229984 causes an amino acid change at position 48 by replacing arginine with histidine, which increases the activity of the *ADH1B* enzyme that oxidizes ethanol to acetaldehyde (Edenberg and Foroud, 2013; Hurley and Edenberg, 2012). After consuming alcohol, elevated *ADH1B* activity has been hypothesized to transiently increase the level of acetaldehyde, leading to unpleasant effects that limit further drinking. Meta-analysis of this variant in Asian populations, where the rs1229984 A allele is

common (allele frequency=0.7 in 1000 Genomes)(Abecasis et al., 2012), has demonstrated strong effects on the risk of developing alcohol-related disorders (OR 0.45: $p=7x10^{-42}$) (Li et al., 2011). Recently, this polymorphism was shown to have a similar effect on risk of alcohol dependence in European and African Americans (African and European OR 0.34: $p=6.6x10^{-10}$ (Bierut et al., 2012); European: $p=1.17x10^{-31}$(Gelernter et al., 2014)), where the rs1229984 A allele is less common (European American frequency=0.05; African American frequency=0.02 in Exome Variant Server)(http://evs.gs.washington.edu/EVS/).

Other studies suggest that social environments that encourage drinking may diminish the protective genetic effects of alcohol metabolizing variants (Hasin et al., 2002; Higuchi et al., 1994; Irons et al., 2007; Irons et al., 2012). However, to our knowledge, no study has explored the interplay of the *ADH1B* rs1229984 variant and the important social context of peer drinking during the critical developmental period of adolescence when alcohol use is initiated and drinking patterns are established. Peer drinking has long been recognized as a strong risk factor for adolescent drinking problems (Curran et al., 1997; Reifman et al., 1998), and recently, twin studies have provided evidence that peer drinking modifies heritable variation in adolescent alcohol involvement (Agrawal et al., 2010; Dick et al., 2007; Guo et al., 2009; Harden et al., 2008).

This study tests the interaction between a genome-wide significant functional *ADH1B* variant and the risk environment of peer drinking in the development of two adolescent drinking milestones: first intoxication and first DSM-5 alcohol use disorder symptom. Examining hypothesis-driven gene-by-environment (GxE) interactions using robust genetic and environmental risks during developmental transitions provides an important approach for untangling the complex etiology of alcohol use disorder.

3.3 MATERIALS AND METHODS

COGA Sample Description

Study participants were enrolled in the Collaborative Study on the Genetics of Alcoholism (COGA), a large, multi-center, family study designed to identify genes that contribute to alcohol use disorders in high-risk (defined as recruited through alcohol dependent probands) and community comparison families (Begleiter et al., 1995). Since 2005, the adolescent and young adult study in COGA has used a longitudinal design to examine the development of alcohol use disorders in young participants from these families. Individuals aged 12 to 22 were recruited from six sites across the US and interviewed every two years. Institutional review boards at all sites approved the study design. Adult participants provided informed consent, parents provided consent for all children younger than 18, and children provided assent.

Assessment of Phenotypes

Interview assessment was performed using the Semi-Structured Assessment for the Genetics of Alcoholism (SSAGA) to gather reliable and valid information on alcohol use behaviors (Bucholz et al., 1995; Bucholz et al., 1994; Hesselbrock et al., 1999). Participants 18 years and older were assessed with the Phase IV SSAGA, and those less than 18 years were assessed with an age appropriate adolescent version called the Phase IV C-SSAGA (Kuperman et al., 2001).

Drinking Milestones

Two adolescent drinking milestones were used as primary outcomes among adolescent ever-drinkers: age of first intoxication, a common and clinically relevant variable, and first DSM-5 alcohol use disorder symptom, a heritable characteristic associated with future alcohol-related problems (Rhee et al., 2003; Young et al., 2006). These outcomes commonly occur during adolescence and therefore coincide with the environment of adolescent peer drinking. Age of first intoxication was derived from responses to the question "How old were you the first time you got drunk, that is, your speech was slurred or you were unsteady on your feet?" Age of first DSM-5 symptom was developed from examining the youngest age that individuals first experienced one of the 11 symptoms of alcohol use disorder. Given the longitudinal design of this study with multiple assessments over time, the earliest interview in which the participant endorsed first intoxication or first DSM-5 symptom was selected to assign the age of onset.

Peer Drinking

The environment of adolescent peer drinking was derived from participant responses to questions addressing the proportion of best friends who drink. With the longitudinal design of the study, 88% (1366/1550) of participants received at least one adult SSAGA assessment at age 18 years or older. Assignment of the level of peer drinking in these participants was determined from the first adult SSAGA interview with the question "When you were 12-17, how many of your best friends used alcohol?" and the 4 possible answers of none, few, most, or all. For participants who had not reached age 18 at the last assessment, peer drinking was evaluated with the maximum value from all C-SSAGA answers to the question "How many of your best friends

use alcohol?" For the primary analyses, peer drinking was dichotomized into low peer drinking (few or no best friends drink) and high peer drinking (most or all best friends drink) as done in previous studies (Kuperman et al., 2013). The four level peer drinking variable (none, few, most, or all best friends) was also investigated in secondary analyses to assess a possible dose response, but interaction effects are not presented because of the small number of individuals in some groups.

To assess the concordance of the retrospective SSAGA interview peer drinking responses for ages 12-17 with current peer drinking reported in C-SSAGA assessments, we compared the first adult SSAGA response and the maximum value from all C-SSAGA assessments among individuals with at least one adult and one child questionnaire. For the 996 participants with both adult and child interviews, 73% of peer drinking assignments had the same dichotomous variable (none/few vs most/all best friends). This concordance demonstrates that our retrospective approach of using the first SSAGA interview when available is a reasonable strategy to assess peer drinking across adolescence. It also shows that for the 12% of participants without a single adult SSAGA assessment, using the maximum value from C-SSAGA assessments reasonably estimates the proportion of best friends drinking from ages 12-17.

Genotyping

Blood samples were obtained for genetic analysis. The *ADH1B* rs1229984 variant was genotyped with Sequenom MassArray technology (Sequenom, San Diego, CA, USA) following standard procedures. Several quality control measures were employed. Genetic variants had a genotyping rate of greater than 99% and were in Hardy-Weinberg equilibrium in both the European and African American groups. The program PEDCHECK (O'Connell and Weeks,

1998) was used to examine Mendelian inheritance, and only individuals with no Mendelian inconsistencies were included in the rs1229984 genotyped sample (N=2580, **Figure 3.1**).

A set of 64 ancestry informative markers was genotyped as part of a 96 SNP Biorepository Panel by the Rutgers University Cell and DNA Repository. These markers were used in SNPrelate, a function in R, to assign ancestry groups. HapMap populations were included as reference groups. There was high concordance (97%) between self-reported and genetically determined ethnicity for European and African American individuals, and only concordant individuals were used in the analyses.

Sample Selection

In the COGA adolescent and young adult study, 2,580 individuals with a first interview age of 12 to 22 were genotyped for the *ADH1B* rs1229984 variant, and participants for the analyses were drawn from this group (**Figure 3.1**). Focusing on European and African American subjects and excluding individuals with missing or unreliable data left 2,410 individuals (entire sample described in **Table 3.1**). The samples used for the primary analyses of first intoxication and first DSM-5 symptom consisted of 1,550 ever-drinkers before age 18 (also described in **Table 3.1**). Ever-drinkers were targeted because the *ADH1B* variant is only expected to exhibit a protective effect in response to alcohol consumption. Because the peer drinking variable examined the age-range of 12-17, the primary analyses focused on events that occurred during this time.

Data Analysis

Data were analyzed using the Statistical Analysis System (SAS 9.3, Cary, NC, USA). Cox-Proportional Hazards Regression (SAS PROC PHREG) was used to model drinking milestones and all individuals who did not experience an event in adolescence were censored at their age of last interview or 18. Participants with rs1229984 GA genotype (N=96) and AA genotype (N=2) were collapsed into one group for comparison with the GG genotype participants (N=1,452), as done in previous studies (Bierut et al., 2012). Models were checked for violations of the proportional hazards assumption and Schoenfeld residuals were examined. The option COVSANDWICH (AGGREGATE) was used to statistically adjust for the non-independence of correlated familial data in all analyses, as done in previous studies (Kuperman et al., 2013).

Models in Primary Analyses

Main effects of the *ADH1B* variant and peer drinking were examined in univariate and multivariate models of age of first intoxication and first DSM-5 symptom in the sample of adolescent ever-drinkers (N=1,550, called *univariate model set* and *multivariate model set*, **Table 3.2**). All models presented in the tables employed STRATA statements for gender and ethnicity to adjust for differences in baseline hazards in these groups. The interplay between the *ADH1B* variant and peer drinking was assessed by adding product interaction terms to models of drinking milestones (called *interaction model set*, **Table 3.2**). This final proportional hazards model was $\lambda(t)=\lambda(t)\exp(\beta_1*(rs1229984) + \beta_2*(peer_drinking) + \beta_3*(rs1229984*peer_drinking))$. The possibility of a gene-environment correlation between *ADH1B* rs1229984 and peer drinking was also assessed because genetic factors influence selection of peers who drink (Fowler et al.,

2007) and inadequate control of this correlation could produce false interactions. Using logistic regression, the outcome peer drinking was modeled with the variables of the *ADH1B* variant, gender, and ethnicity.

Secondary Analyses

Secondary analyses were performed to test the robustness of our primary findings. First, association of the *ADH1B* rs1229984 variant with the milestone of age of drinking initiation was examined in the entire sample, which included adolescent never-drinkers (N=2,410). Second, analyses stratified by ancestry were performed to examine the main and interaction effects within the subpopulations of European and African Americans.

3.4 RESULTS

Participant Characteristics

Demographic, behavioral, and genotypic characteristics of the study samples are presented in **Table 3.1**. The sample of ever-drinkers before age 18 used in the primary analyses consisted of 1,550 individuals from 1,151 nuclear families (defined by full-siblings) and 645 extended families. The mean first interview age was 17, 49% were female, and the majority came from high-risk families (89%) and were European American (73%). Before age 18, 74% had a first intoxication and 44% experienced a first DSM-5 symptom of alcohol use disorder. From ages 12 to 17, 39% reported that most or all of their best friends drank alcohol. Consistent with the expected population frequencies of the *ADH1B* variant, 6% carried at least one copy of the protective A allele (8% in European Americans and 3% in African Americans).

Effect of Peer Drinking

Most/all best friends drinking compared to none/few best friends drinking between ages 12-17 was associated with a main hazardous effect in univariate and multivariate models of early drinking behaviors (**Table 3.2**). In the final *interaction model set* with GxE product terms, self-reported peer drinking had a robust effect on first intoxication (Hazards ratio (HR)=1.81, 95% CI 1.62-2.01) and first DSM-5 symptom (HR=2.17, 95% CI 1.88-2.50). In secondary analyses examining all four responses for best friends drinking (none, few, most, all), an increase in the number of best friends drinking was similarly related to the first intoxication (*multivariate model set* with *none* as the reference; *few* HR=1.72, 95% CI 1.44-2.05; *most* HR=2.65, 95% CI 2.20-3.18; *all* HR=3.69, 95% CI 2.93-4.64) and first DSM-5 symptom (*multivariate model set* with

none as the reference; *few* HR=2.43, 95% CI 1.77-3.33; *most* HR=4.29, 95% CI 3.12-5.92; *all* HR=5.84, 95% CI 4.16-8.21). These results indicate a “dosage effect” where the reported proportion of best friends drinking was positively associated with higher risk for developing adolescent drinking milestones.

Effect of *ADH1B* rs1229984 Variant

During adolescence, presence of the *ADH1B* variant (GA/AA genotypes) was associated with a protective main effect among ever-drinkers for first intoxication and first DSM-5 symptom in univariate and multivariate models (**Table 3.2**). In the final *interaction model set* with GxE product terms, the effect of the *ADH1B* variant was strong for both first intoxication (HR=0.56, 95% CI 0.41-0.76) and first DSM-5 symptom (HR=0.45, 95% CI 0.26-0.77). In secondary analyses of the entire sample that included never-drinkers, presence of the variant exhibited no effect on drinking initiation (HR in univariate model=1.12, 95% CI 0.92-1.36), consistent with the mechanism of the variant of only exhibiting an effect in response to alcohol consumption.

Interaction between *ADH1B* rs1229984 and Peer Drinking

The interaction between the *ADH1B* variant and peer drinking was tested by adding GxE product term to models of drinking milestones in adolescent drinkers (N=1,550), which illustrated a significant statistical interaction for first intoxication (p=.002) and first DSM-5 symptom (p=.01) (**Table 3.2**). Among individuals who reported none/few best friends drinking, the *ADH1B* GA/AA genotypes had a strong protective effect for first intoxication (HR=0.56, 95% CI 0.41-0.76) and first DSM-5 symptom (HR=0.45, 95% CI 0.26-0.77). In individuals who

reported most/all best friends drinking, however, this protective effect was not observed for either first intoxication (HR=1.16, 95% CI 0.82-1.65) or first DSM-5 symptom (HR=1.03, 95% CI 0.73-1.45), as illustrated by the point estimates close to 1. **Figure 3.2** more clearly illustrates this GxE interaction by presenting the survival estimates.

Association between *ADH1B* Variant and Peer Drinking

No evidence of a gene-environment correlation between the *ADH1B* variant and peer drinking was observed. Specifically, the independent variable of the *ADH1B* rs1229984 variant was not significant in the logistic regression model of perceived peer drinking controlling for sex and ethnicity as covariates (most/all vs none/few best friends drink, Odds Ratio=1.19, 95% CI 0.78-1.83).

Assessment of Robustness of Results

The proportional hazards assumption was satisfied in first DSM-5 symptom models. Violations were noted in a subset of first intoxication analyses. Examination of Scholenfeld residuals indicated that the group of 17 year olds was driving this violation, perhaps reflecting important transitions at this age. Censoring at age 17 instead of 18 satisfied the proportional hazards assumption without substantially altering the parameter estimates, supporting our conclusions.

Ancestry-stratified analyses demonstrated consistent main and interaction effects in the European American subpopulation (N=1,130). In the *interaction model set* for European American individuals, peer drinking had a hazardous effect on first intoxication (HR=1.87, 95% 1.66-2.11) and first DSM-5 symptom (HR=2.23, 95% CI 1.89-2.63); rs1229984 had a protective

effect on first intoxication (HR=0.60, 95% CI 0.44-0.82) and first DSM-5 symptom (HR=0.47, 95% CI 0.27-0.82); and interaction terms were significant ($p<.02$). The *ADH1B* GA/AA genotypes were protective among individuals reporting none/few best friends drinking, but not among those reporting most/all best friends drinking, corroborating our findings in the overall sample.

Stratified analyses of African Americans (N=420) provided trending evidence of main effects. In the *interaction model set* with GxE product terms, peer drinking had a hazardous effect on first intoxication (HR=1.62, 95% 1.27-2.08) and first DSM-5 symptom (HR=1.98, 95% CI 1.50-2.61); rs1229984 had a trending protective effect on first intoxication (HR=0.32, 95% CI 0.08-1.27) and first DSM-5 symptom (HR=0.35, 95% CI 0.05-2.28); and interaction terms were insignificant ($p>.7$). The limited sample size of African Americans combined with the low frequency of the rs1229984 minor allele limits power to detect interactions in this analysis. Nonetheless, the robust effect of peer drinking in both ancestry groups and the well-established role of rs1229984 across ancestry groups lends support for our conclusions drawn from the combined sample.

3.5 DISCUSSION

Alcohol use behaviors established during adolescence are important contributing factors for the later progression to alcohol dependence (Grant et al., 2006; Pitkanen et al., 2005). These data provide an example of the important interplay of genetic and environmental risks in the development of drinking milestones during this critical period of adolescence. Using a longitudinal sample of European and African American adolescent drinkers, we demonstrate that the *ADH1B* rs1229984 minor A allele is associated with a protective effect for early drinking behaviors, and in the environmental high-risk context of most or all best friends drinking, this genetic protection is negated.

The observation that the *ADH1B* variant is associated with a decreased risk of first intoxication and first DSM-5 symptom during adolescence (**Table 3.2**) extends previous findings that this variant protects against alcohol-related health problems in adulthood (Bierut et al., 2012; Gelernter et al., 2014; Li et al., 2011). Despite having an early role in the trajectory of drinking behaviors, the *ADH1B* variant was not associated with drinking initiation, consistent with the hypothesized mechanism of action that requires alcohol exposure (Edenberg and Foroud, 2013; Hurley and Edenberg, 2012). This specific example of a genetic variant that influences early drinking milestones, but not initiation, builds on twin and adoption study findings that genetic factors contribute to the development of adolescent alcohol-related problems, and environmental factors more strongly drive drinking initiation (Hopfer et al., 2003; Lynskey et al., 2010).

Beyond demonstrating an early protective role of the *ADH1B* GA/AA genotypes in the development of these drinking behaviors, the results illustrate that reporting most or all best

friends drinking was associated with attenuation of this genetic protection (**Figure 3.2**). The observation that social context modifies the effect of an *ADH1B* variant extends previous studies on alcohol metabolizing variants. Higuchi et al. (1994) found that the proportion of alcohol dependent adults in Japan with one copy of a protective aldehyde dehydrogenase 2 (*ALDH2*) variant increased between 1979 and 1992, following the increased cultural pressure to drink alcohol. Similarly, Irons et al. (2007) reported that the high-risk environment of sibling substance use was associated with a diminished effect of this *ALDH2* variant in East Asian adolescent adoptees, and more recently, this group demonstrated that high parental alcohol use and misuse reduced the effect of the *ALDH2* protective allele (Irons et al., 2012). For the *ADH1B* rs1229984 variant, Hasin et al. (2002) observed a weaker protective role in certain groups, which was hypothesized to reflect differences in environmental exposure to heavy drinking. Our findings expand on these earlier observations by demonstrating that the critical high-risk social context of adolescent peer drinking is associated with the loss of the protective genetic effect of the *ADH1B* variant in European and African Americans.

Previous studies of metabolizing variants have focused on Asian populations where the *ADH1B* rs1229984 A allele is common, and only recently was this variant associated with alcoholism at a genome-wide level in an European and African American sample ($p=6.6x10^{-10}$) (Bierut et al., 2012). A recent GWAS of alcohol dependence further supports a strong effect of this variant in European Americans ($p=1.17x10^{-31}$) (Gelernter et al., 2014). To our knowledge, this study is the first to examine the effect of the *ADH1B* rs1229984 variant on adolescent drinking behaviors and incorporate environmental moderation in European and African Americans.

One challenge of studying the influence of the *ADH1B* rs1229984 variant in populations of European and African ancestry is the low frequency of the protective A allele. Although over 1,500 adolescent drinkers were examined in this analysis, only 98 (6%) carried an A allele (of which 36 reported most/all best friends drinking). Nonetheless, the influence of this variant and the GxE interaction was persistently strong in models of first intoxication and first DSM-5 symptom (**Table 3.2**). Secondary ancestry-stratified analyses also demonstrated consistent main and interaction effects in the European American subpopulation (N=1,130) and provided trending evidence of main effects in the African American subpopulation (N=420), where power was limited. These analyses, combined with previous studies supporting the protective role of rs1229984 across ancestry groups as well as the moderating effect of social environments, support our conclusion that this variant is associated with a protective effect for early drinking behaviors in European and African Americans, but this genetic protection may be eliminated by adolescent peer drinking.

The findings reported here have several limitations. First, studying a specific genetic variant provides limited information on the general genetic underpinnings of complex diseases such as alcohol use disorder (Dick and Kendler, 2012). Nevertheless, examination of specific robust variants provides important insight into underlying biological mechanisms that are not assessed by traditional studies of latent genetic influences. Second, other genetic variants may influence associations between *ADH1B* rs1229984 and drinking behaviors (Meyers et al., 2013; Toth et al., 2011). Third, self-reported peer drinking was viewed as an environmental risk factor in this study, but research suggests that genetic factors contribute to peer alcohol involvement (Fowler et al., 2007). Gene-environment correlations can arise when an individual's heritable behavior evokes an environmental response (evocative rGE) or when an individual possesses a

heritable propensity to select an environment (active rGE). In this study, the *ADH1B* rs1229984 variant was not associated with self-reported peer drinking, supporting our interpretation that peer drinking acts as an environmental modifier, but other gene-environment correlations may still contribute to the observed effects. Fourth, the temporal ordering of peer drinking and the onset of drinking behaviors could not be assessed in this study (**Table 3.1**). It is possible that other risk factors correlated with peer drinking, such as parental monitoring or genetic risk for anti-social behavior, may account for the observed associations. Fifth, peer drinking was assessed by respondent report and may not reflect the actual proportion of best friends drinking. Finally, the majority of participants were from high-risk families, which may limit the generalizability of the findings. It is possible that only individuals at high-risk for alcohol use disorders lose the protective effect of the *ADH1B* rs1229984 variant under environments that encourage drinking. Replication of these findings in independent samples is a critical next step.

Despite these limitations, this study has several strengths. First, the analysis focused on a genetic variant with strong statistical and biological evidence for alcohol-related measures, which addresses common criticisms of GxE studies (Duncan and Keller, 2011; Joober et al., 2007; Risch et al., 2009). Second, focusing on a youth population and employing a longitudinal study design reduced recall bias, enabling more accurate assessment of drinking behaviors during the critical period of adolescence. Third, the robust environment of respondent report of best friends drinking from ages 12-17 coincided with the timing of the primary outcomes under study. This analysis focused on drinking behaviors that are common in adolescence and therefore are more likely to be directly influenced by peer drinking during this period. Finally, studying adolescent drinking milestones facilitated the characterization of the unfolding of genetic and environmental risks across development. Recent studies further support the discovery potential

of examining genetic variants during important behavioral transitions in at-risk youth (Belsky et al., 2013; Dick et al., 2013). Future research on alcohol use disorders may benefit from similar hypothesis-driven study designs that examine well-established genes and environments during critical developmental periods.

From a public health perspective, this study provides a genetic argument in support of early social interventions to decrease affiliation with peer drinkers. Specifically, these findings support the use of a screening tool for practitioners to identify at-risk youth, developed by the National Institute on Alcohol Abuse and Alcoholism and the American Academy of Pediatrics, in which the first question addresses friends' drinking (NIAAA, 2011). Under the high-risk environment of best friends drinking, all adolescents were at increased risk for early drinking problems, and particularly, those at lower genetic risk experienced the greatest added risk. This study serves as a model of how understanding the interplay between genes and environments may increase etiological knowledge of alcohol use disorders and potentially inform interventions that aim to disrupt progression to alcoholism.

3.6 ACKNOWLEDGEMENTS

I would like to acknowledge all of my co-authors: Edenberg HJ, Nurnberger J Jr, Agrawal A, Bucholz KK, Almasy LA, Chorlian D, Dick DM, Hesselbrock VM, Kramer JR, Kuperman S, Porjesz B, Schuckit MA, Tischfield JA, Wang JC, Wetherill L, Foroud TM, Rice J, Goate A, and Bierut LJ.

This work was supported by U10AA008401 from the National Institutes of Health. EO was supported by T32GM07200, UL1TR000448, and TL1TR000449. AA was supported by K02DA032573 and R21AA021235. DMD was supported by K02AA018755.

Disclosures:

AA has previously received peer-reviewed funding from ABMRF/Foundation for Alcohol Research, which receives support from brewers. VMH is a member of the scientific advisory board of D&A Pharma, Paris, France. JCW, AG, and LJB are listed as inventors on Issued U.S. Patent 8,080,371,"Markers for Addiction" covering the use of certain SNPs in determining the diagnosis, prognosis, and treatment of addiction. The other authors declare no conflict of interest.

3.7 REFERENCES

Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, Handsaker RE, Kang HM, Marth GT, McVean GA (2012) An integrated map of genetic variation from 1,092 human genomes. Nature 491:56-65.

Agrawal A, Balasubramanian S, Smith EK, Madden PA, Bucholz KK, Heath AC, Lynskey MT (2010) Peer substance involvement modifies genetic influences on regular substance involvement in young women. Addiction (Abingdon, England) 105:1844-1853.

Begleiter H, Reich T, Hesselbrock V, Porjesz B, Li T, Schuckit M, Edenberg H, Rice J (1995) The Collaborative Study on the Genetics of Alcoholism. Alcohol Health Res World 19:228–236.

Belsky DW, Moffitt TE, Baker TB, Biddle AK, Evans JP, Harrington H, Houts R, Meier M, Sugden K, Williams B, Poulton R, Caspi A (2013) Polygenic risk and the developmental progression to heavy, persistent smoking and nicotine dependence: evidence from a 4-decade longitudinal study. JAMA psychiatry 70:534-542.

Bierut LJ, Goate AM, Breslau N, Johnson EO, Bertelsen S, Fox L, Agrawal A, Bucholz KK, Grucza R, Hesselbrock V, Kramer J, Kuperman S, Nurnberger J, Porjesz B, Saccone NL, Schuckit M, Tischfield J, Wang JC, Foroud T, Rice JP, Edenberg HJ (2012) ADH1B is associated with alcohol dependence and alcohol consumption in populations of European and African ancestry. Molecular psychiatry 17:445-450.

Bucholz KK, Cadoret R, Cloninger CR, Dinwiddie SH, Hesselbrock VM, Nurnberger JI, Jr., Reich T, Schmidt I, Schuckit MA (1994) A new, semi-structured psychiatric interview

for use in genetic linkage studies: a report on the reliability of the SSAGA. Journal of studies on alcohol 55:149-158.

Bucholz KK, Hesselbrock VM, Shayka JJ, Nurnberger JI, Jr., Schuckit MA, Schmidt I, Reich T (1995) Reliability of individual diagnostic criterion items for psychoactive substance dependence and the impact on diagnosis. Journal of studies on alcohol 56:500-505.

Curran PJ, Stice E, Chassin L (1997) The relation between adolescent alcohol use and peer alcohol use: a longitudinal random coefficients model. Journal of consulting and clinical psychology 65:130-140.

Dick DM, Cho SB, Latendresse SJ, Aliev F, Nurnberger JI, Jr., Edenberg HJ, Schuckit M, Hesselbrock VM, Porjesz B, Bucholz K, Wang JC, Goate A, Kramer JR, Kuperman S (2013) Genetic influences on alcohol use across stages of development: GABRA2 and longitudinal trajectories of drunkenness from adolescence to young adulthood. Addiction biology.

Dick DM, Kendler KS (2012) The impact of gene-environment interaction on alcohol use disorders. Alcohol research : current reviews 34:318-324.

Dick DM, Pagan JL, Viken R, Purcell S, Kaprio J, Pulkkinen L, Rose RJ (2007) Changing environmental influences on substance use across development. Twin research and human genetics : the official journal of the International Society for Twin Studies 10:315-326.

Duncan LE, Keller MC (2011) A critical review of the first 10 years of candidate gene-by-environment interaction research in psychiatry. The American journal of psychiatry 168:1041-1049.

Edenberg HJ, Foroud T (2013) Genetics and alcoholism. Nature reviews. Gastroenterology & hepatology 10**:**487-494.

Exome Variant Server, NHLBI GO Exome Sequencing Project (ESP), Seattle, WA. Available at: URL: http://evs.gs.washington.edu/EVS/. Accessed May, 2014.

Fowler T, Shelton K, Lifford K, Rice F, McBride A, Nikolov I, Neale MC, Harold G, Thapar A, van den Bree MB (2007) Genetic and environmental influences on the relationship between peer alcohol use and own alcohol use in adolescents. Addiction (Abingdon, England) 102**:**894-903.

Gelernter J, Kranzler HR, Sherva R, Almasy L, Koesterer R, Smith AH, Anton R, Preuss UW, Ridinger M, Rujescu D, Wodarz N, Zill P, Zhao H, Farrer LA (2014) Genome-wide association study of alcohol dependence:significant findings in African- and European-Americans including novel risk loci. Molecular psychiatry 19**:**41-49.

Grant JD, Scherrer JF, Lynskey MT, Lyons MJ, Eisen SA, Tsuang MT, True WR, Bucholz KK (2006) Adolescent alcohol use is a risk factor for adult alcohol and drug dependence: evidence from a twin design. Psychological medicine 36**:**109-118.

Guo G, Elder GH, Cai T, Hamilton N (2009) Gene–environment interactions: Peers' alcohol use moderates genetic contribution to adolescent drinking behavior. Social Science Research 38**:**213-224.

Harden KP, Hill JE, Turkheimer E, Emery RE (2008) Gene-environment correlation and interaction in peer effects on adolescent alcohol and tobacco use. Behavior genetics 38**:**339-347.

Hasin D, Aharonovich E, Liu X, Mamman Z, Matseoane K, Carr L, Li TK (2002) Alcohol and ADH2 in Israel: Ashkenazis, Sephardics, and recent Russian immigrants. The American journal of psychiatry 159:1432-1434.

Heath AC, Bucholz KK, Madden PA, Dinwiddie SH, Slutske WS, Bierut LJ, Statham DJ, Dunne MP, Whitfield JB, Martin NG (1997) Genetic and environmental contributions to alcohol dependence risk in a national twin sample: consistency of findings in women and men. Psychological medicine 27:1381-1396.

Hesselbrock M, Easton C, Bucholz KK, Schuckit M, Hesselbrock V (1999) A validity study of the SSAGA--a comparison with the SCAN. Addiction (Abingdon, England) 94:1361-1370.

Higuchi S, Matsushita S, Imazeki H, Kinoshita T, Takagi S, Kono H (1994) Aldehyde dehydrogenase genotypes in Japanese alcoholics. Lancet 343:741-742.

Hopfer CJ, Crowley TJ, Hewitt JK (2003) Review of twin and adoption studies of adolescent substance use. Journal of the American Academy of Child and Adolescent Psychiatry 42:710-719.

Hurley TD, Edenberg HJ (2012) Genes encoding enzymes involved in ethanol metabolism. Alcohol research : current reviews 34:339-344.

Irons DE, Iacono WG, Oetting WS, McGue M (2012) Developmental trajectory and environmental moderation of the effect of ALDH2 polymorphism on alcohol use. Alcoholism, clinical and experimental research 36:1882-1891.

Irons DE, McGue M, Iacono WG, Oetting WS (2007) Mendelian randomization: a novel test of the gateway hypothesis and models of gene-environment interplay. Development and psychopathology 19:1181-1195.

Joober R, Sengupta S, Schmitz N (2007) Promoting measured genes and measured environments: on the importance of careful statistical analyses and biological relevance. Archives of general psychiatry 64:377-378; author reply 378-379.

Kendler KS, Neale MC, Heath AC, Kessler RC, Eaves LJ (1994) A twin-family study of alcoholism in women. The American journal of psychiatry 151:707-715.

Kuperman S, Chan G, Kramer JR, Wetherill L, Bucholz KK, Dick D, Hesselbrock V, Porjesz B, Rangaswamy M, Schuckit M (2013) A model to determine the likely age of an adolescent's first drink of alcohol. Pediatrics 131:242-248.

Kuperman S, Schlosser SS, Kramer JR, Bucholz K, Hesselbrock V, Reich T, Reich W (2001) Risk domains associated with an adolescent alcohol dependence diagnosis. Addiction (Abingdon, England) 96:629-636.

Li D, Zhao H, Gelernter J (2011) Strong association of the alcohol dehydrogenase 1B gene (ADH1B) with alcohol dependence and alcohol-induced medical diseases. Biological psychiatry 70:504-512.

Lynskey MT, Agrawal A, Heath AC (2010) Genetically informative research on adolescent substance use: methods, findings, and challenges. Journal of the American Academy of Child and Adolescent Psychiatry 49:1202-1214.

Merikangas KR, He JP, Burstein M, Swanson SA, Avenevoli S, Cui L, Benjet C, Georgiades K, Swendsen J (2010) Lifetime prevalence of mental disorders in U.S. adolescents: results from the National Comorbidity Survey Replication--Adolescent Supplement (NCS-A). Journal of the American Academy of Child and Adolescent Psychiatry 49:980-989.

Meyers JL, Shmulewitz D, Aharonovich E, Waxman R, Frisch A, Weizman A, Spivak B, Edenberg HJ, Gelernter J, Hasin DS (2013) Alcohol-Metabolizing Genes and Alcohol

Phenotypes in an Israeli Household Sample. Alcoholism, clinical and experimental research. Alcoholism, clinical and experimental research 37:1872-1881.

NIAAA (2011) National Institute on Alcohol Abuse and Alcoholism. Alcohol Screening and Brief Intervention for Youth: A Practitioner's Guide, National Institute on Alcohol Abuse and Alcoholism Publications Distribution Center, Rockville, MD.

NSDUH. Results from the 2012 National Survey on Drug Use and Health: Detailed Tables. Available at: http://www.samhsa.gov/data/NSDUH/2012SummNatFindDetTables/Index.aspx. Accessed February 28, 2014.

O'Connell JR, Weeks DE (1998) PedCheck: a program for identification of genotype incompatibilities in linkage analysis. American journal of human genetics 63:259-266.

Pickens RW, Svikis DS, McGue M, Lykken DT, Heston LL, Clayton PJ (1991) Heterogeneity in the inheritance of alcoholism. A study of male and female twins. Archives of general psychiatry 48:19-28.

Pitkanen T, Lyyra AL, Pulkkinen L (2005) Age of onset of drinking and the use of alcohol in adulthood: a follow-up study from age 8-42 for females and males. Addiction (Abingdon, England) 100:652-661.

Prescott CA, Kendler KS (1999) Genetic and environmental contributions to alcohol abuse and dependence in a population-based sample of male twins. The American journal of psychiatry 156:34-40.

Reifman A, Barnes GM, Dintcheff BA, Farrell MP, Uhteg L (1998) Parental and peer influences on the onset of heavier drinking among adolescents. Journal of studies on alcohol 59:311-317.

Rhee SH, Hewitt JK, Young SE, Corley RP, Crowley TJ, Stallings MC (2003) Genetic and environmental influences on substance initiation, use, and problem use in adolescents. Archives of general psychiatry 60:1256-1264.

Rietschel M, Treutlein J (2013) The genetics of alcohol dependence. Annals of the New York Academy of Sciences 1282:39-70.

Risch N, Herrell R, Lehner T, Liang KY, Eaves L, Hoh J, Griem A, Kovacs M, Ott J, Merikangas KR (2009) Interaction between the serotonin transporter gene (5-HTTLPR), stressful life events, and risk of depression: a meta-analysis. JAMA : the journal of the American Medical Association 301:2462-2471.

Swendsen J, Burstein M, Case B, Conway KP, Dierker L, He J, Merikangas KR (2012) Use and abuse of alcohol and illicit drugs in US adolescents: results of the National Comorbidity Survey-Adolescent Supplement. Archives of general psychiatry 69:390-398.

Toth R, Fiatal S, Petrovski B, McKee M, Adany R (2011) Combined effect of ADH1B RS1229984, RS2066702 and ADH1C RS1693482/ RS698 alleles on alcoholism and chronic liver diseases. Disease markers 31:267-277.

Wang JC, Kapoor M, Goate AM (2012) The genetics of substance dependence. Annual review of genomics and human genetics 13:241-261.

Young SE, Rhee SH, Stallings MC, Corley RP, Hewitt JK (2006) Genetic and environmental vulnerabilities underlying adolescent substance use and problem use: general or specific? Behavior genetics 36:603-615.

3.8 TABLES

Table 3.1. Characteristics of samples used in analyses

Characteristic	Entire sample (N=2,410)	Ever-drinkers before age 18 (N=1,550)
Ancestry, N (%)		
European	1,648 (68.4)	1,130 (72.9)
African	762 (31.6)	420 (27.1)
Sex, N (%)		
Males	1,182 (49.1)	784 (50.6)
Females	1,228 (51.0)	766 (49.4)
Age at first interview, years		
Mean ± sd	16.3 ± 3.2	16.7 ± 3.0
Range	12-22	12-22
No. of interviews		
Mean ± sd	3.2 ± 1.1	3.2 ± 1.1
Range	(1-5)	(1-5)
Family status, N (%)		
From high-risk families	2,096 (87.0)	1,384 (89.3)
From comparison families	314 (13.0)	166 (10.7)
No. of extended families	781	645
No. of nuclear families		
Only full-siblings	1,629	1,151
Including half-siblings	1,438	1,044
No. of individuals per extended family, median (range)	2 (1-24)	2 (1-17)
Drinking milestones reached before age 18, N (%)		
First drink	1,573 (65.3)	1,550 (100.0)
First intoxication	1,170 (48.6)	1,147 (74.0)
First DSM-5 symptom	702 (29.1)	683 (44.1)
Among those who exhibit a first intoxication before age 18		
Mean age ± sd	15.3 (1.5)	15.4 (1.4)
Age range	8-17	12-17
Among those who exhibit a first DSM-5 symptom before age 18		
Mean age ± sd	15.6 (1.3)	15.6 (1.2)
Age range	10-17	12-17
rs1229984, N (%)		
GG	2,270 (94.2)	1,452 (93.7)
GA	137 (5.7)	96 (6.2)
AA	3 (0.1)	2 (0.1)
Reported proportion of best friends who use alcohol between ages 12-17, N (%)		
None	746 (31.0)	239 (15.4)
Few	981 (40.7)	708 (45.7)
Most	513 (21.3)	453 (29.2)
All	170 (7.1)	150 (9.7)

Table 3.2. Cox proportional hazards regression models of adolescent drinking milestones

	Drinking milestones in ever-drinkers before age 18 (N=1,550)			
	Models of first intoxication		Models of first DSM-5 symptom	
	Hazard ratio (95% CI)	χ^2 p value	Hazard ratio (95% CI)	χ^2 p value
Univariate model set				
rs1229984[a]	0.72 (0.56-0.91)	.006	0.69 (0.50-0.94)	.02
peer drinking[b]	1.89 (1.70-2.10)	<.0001	2.27 (1.98-2.60)	<.0001
Multivariate model set				
rs1229984	0.76 (0.61-0.96)	.02	0.73 (0.54-0.97)	.03
peer drinking	1.88 (1.69-2.09)	<.0001	2.26 (1.97-2.60)	<.0001
Interaction model set				
rs1229984	0.56 (0.41-0.76)	.0002	0.45 (0.26-0.77)	.004
peer drinking	1.81 (1.62-2.01)	<.0001	2.17 (1.88-2.50)	<.0001
rs1229984*peer drinking	2.10 (1.32-3.32)	.002	2.29 (1.21-4.30)	.01
Examination of GxE term in interaction model set				
None/few best friends drink (GA/AA vs GG)	0.56 (0.41-0.76)	.0002	0.45 (0.26-0.77)	.004
Most/all best friends drink (GA/AA vs GG)	1.16 (0.82-1.65)	.39	1.03 (0.73-1.45)	.87

[a] Reference *ADH1B* rs1229984 genotype GG was compared to GA/AA; [b] Reference peer drinking status none/few best friends drink was compared to most/all best friends drink; All models adjusted for gender and ethnicity.

3.9 FIGURES

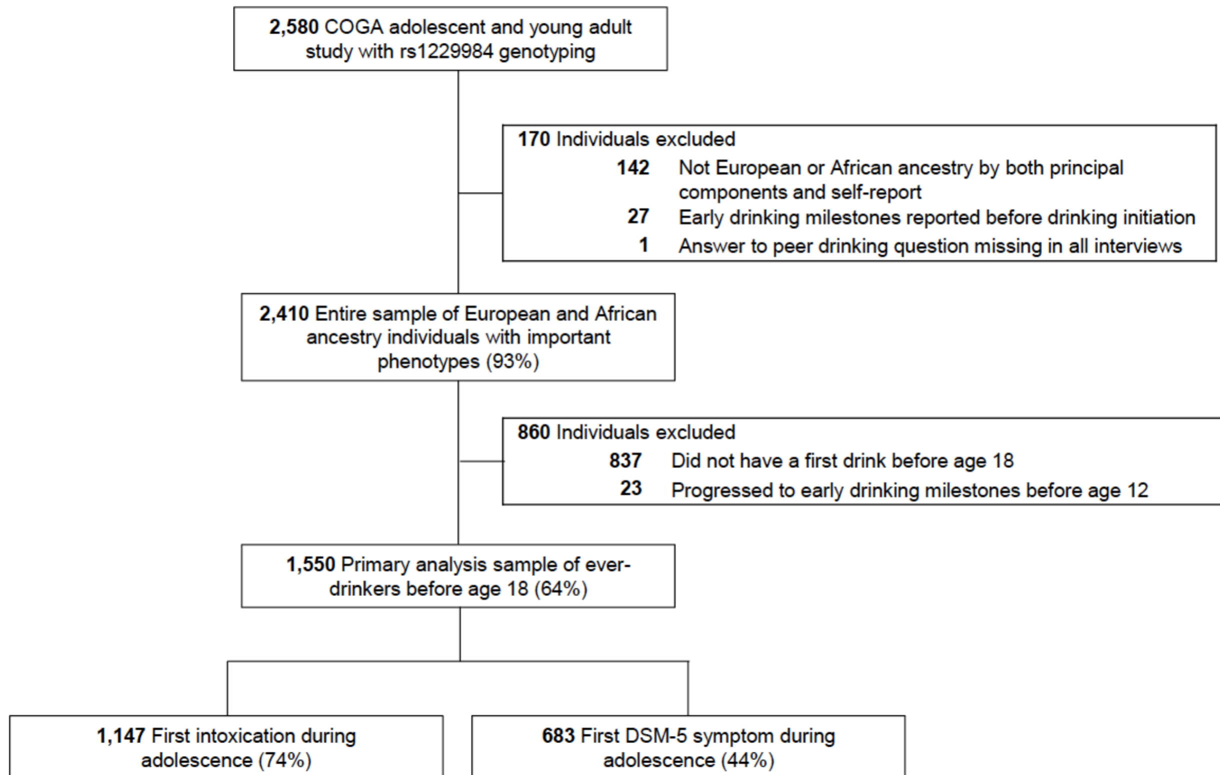


Figure 3.1. Sample selection for study analyses. European and African American adolescent ever-drinkers with *ADH1B* rs1229984 genotyping were drawn from the Collaborative Study on the Genetics of Alcoholism (COGA) for the primary analyses of two early drinking milestones.

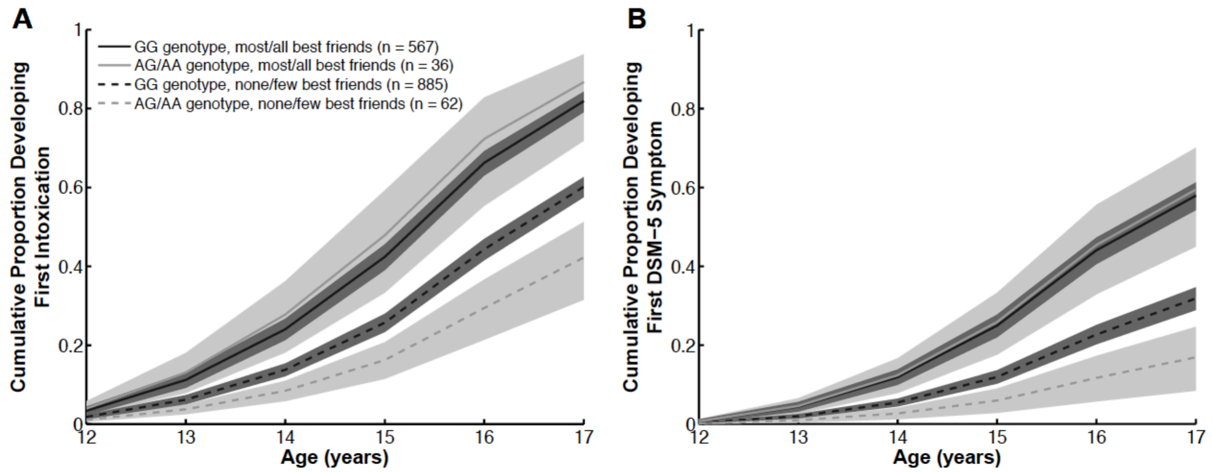


Figure 3.2. Survival estimates of drinking milestones stratified by ADH1B rs1229984 genotype and perceived best friends drinking. Cox proportional hazards regression survival estimates of (A) first intoxication and (B) first DSM-5 alcohol use disorder symptom in adolescent ever-drinkers (N=1,550) with the variables of *ADH1B* genotype, best friends drinking, and GXE interaction term.

CHAPTER FOUR:

CYP2A6 metabolism in the development of nicotine dependence in young adults

4.1 ABSTRACT

The gene *CYP2A6* encodes the enzyme responsible for the majority of nicotine metabolism. Previous studies support that slow metabolizers smoke fewer cigarettes once nicotine dependent, but provide conflicting results on the role of *CYP2A6* in the development of dependence. By focusing on the critical developmental period of young adulthood, this study examines the role of variation in *CYP2A6* on different smoking milestones. A total of 1,102 European Americans with a last interview age from 19-30 years enrolled in the Collaborative Study on the Genetics of Alcoholism were genotyped for *CYP2A6* variants to calculate a previously well-validated nicotine metabolism metric. This metric was not associated with smoking initiation or the development of daily smoking ($p>0.5$), but among daily smokers ($n=468$), decreased metabolism was associated with increased risk of nicotine dependence (defined as Fagerstrom Test for Nicotine Dependence score ≥4) after controlling for sex, study site, age of last interview, and familial relatedness ($p=0.03$). This finding was replicated in 335 daily smokers ages 25-30 years enrolled in the Collaborative Study of Nicotine Dependence ($p=0.02$). Meta-analysis indicated that slow metabolizers (defined by a metric≤0.85) had a 1.64 increased odds (95% CI 1.17-2.28, $p=0.004$) of developing nicotine dependence as compared to normal metabolizers (metric>0.85). Overall, these findings add important knowledge about the complex role of *CYP2A6* variation across different developmental stages of smoking behaviors. Although slow metabolism may be protective for cigarette consumption among nicotine dependent adults, we show that slow metabolism is associated with an increased risk of developing nicotine dependence in young adult daily smokers.

4.2 INTRODUCTION

The development of nicotine dependence requires smoking initiation, conversion from experimental to daily use, and finally the development of advanced smoking behaviors (Belsky et al., 2013; Bierut, 2011). Although the majority of adult smokers initiate smoking during adolescence, rates of daily smoking substantially increase during young adulthood among ever-smokers (22% at ages 12-17, 52% at ages 18-34, 62% at ages 35 or more) (NSDUH, 2012). Furthermore, among those who report smoking within the past 30 days, the proportion of individuals that smoke more than approximately a pack a day also dramatically increases with age (6% at ages 12-17, 23% at ages 18-34, and 44% at ages 35 or more) (NSDUH, 2012). Increasing our knowledge of what factors drive some young adults and not others to transition from initiation to daily smoking and then later smoking behaviors is important for effectively preventing the progression of nicotine dependence.

One genetic factor that may play an important role in the development of smoking behaviors is variation in the gene *CYP2A6*, which encodes a cytochrome P450 enzyme. This enzyme is responsible for the majority of oxidation of nicotine to cotinine, which is the primary pathway of nicotine metabolism in humans (Hukkanen et al., 2005). The *CYP2A6* locus is highly polymorphic, and alleles with reduced function have been associated with slower rates of nicotine metabolism. Common variants define multiple *CYP2A6* haplotypes in European ancestry individuals (Haberl et al., 2005), and the majority of inter-individual variation in the metabolism of nicotine to cotinine is explained by targeted polymorphisms in European Americans (Bloom et al., 2011).

In adults, the region on chromosome 19 encompassing *CYP2A6* is genome-wide significant in large meta-analyses focused on cigarettes per day in European ancestry populations (TAG, 2010; Thorgeirsson et al., 2010). Among nicotine dependent adults, the majority of studies demonstrate that genetically slower metabolizers smoke fewer cigarettes per day. This observation is thought to reflect the fact that smokers naturally titrate their cigarette consumption to maintain steady nicotine levels.

Studies in youth present conflicting results on the effect of nicotine metabolism on the development of nicotine dependence and other smoking behaviors (Audrain-McGovern et al., 2007; Huang et al., 2005; Moolchan et al., 2009; O'Loughlin et al., 2004; Rubinstein et al., 2008; Rubinstein et al., 2013). Some studies suggest that slow nicotine metabolism is associated with an *increased* risk for acquisition of nicotine dependence (O'Loughlin et al., 2004; Rubinstein et al., 2013), possibly reflecting an increased sensitivity to initial nicotine exposure among youth that metabolize nicotine more slowly. In contrast, other studies suggest that slower metabolizers have a *decreased* risk for dependence and related symptoms (Audrain-McGovern et al., 2007; Rubinstein et al., 2008), paralleling findings in adults regarding heaviness of smoking.

Our goal was to investigate how variation in *CYP2A6* influences the development of nicotine dependence and other smoking behaviors during the critical period of young adulthood in European Americans. A better understanding of how variation in nicotine metabolism contributes to the acquisition of smoking milestones will add to our fundamental knowledge of the developmental processes that lead to nicotine dependence and has the potential to identify individuals at increased susceptibility during this critical period.

4.3 MATERIALS AND METHODS

Primary Sample Description

The Collaborative Study on the Genetics of Alcoholism (COGA) is a United States multi-center, family study that aims to identify genes that contribute to alcohol use disorders and related phenotypes (Begleiter et al., 1995). Since 2005, the adolescent and young adult study in COGA has used a longitudinal design to examine the development of substance use disorders in youth from high-risk (defined as recruited through alcohol dependent probands with two or more dependent first degree relatives) and community comparison families. Members aged 12 to 22 were recruited from six sites across the US and interviewed every two years.

Smoking Behaviors in COGA

Interview assessments were performed using the Semi-Structured Assessment for the Genetics of Alcoholism (SSAGA), which comprehensively gathers detailed information on substance use with high reliability and validity (Bucholz et al., 1994; Bucholz et al., 1995; Hesselbrock et al., 1999). Smoking initiation was evaluated with the question "Have you ever smoked a full cigarette?" Daily smoking, defined as smoking at least 4 days per week for at least a month, was assessed among individuals who had initiated smoking using the questions "How many cigarettes did you usually smoke a day" and "For how long, did you smoke this many cigarettes at that rate?"

Among individuals who reported daily smoking, several measures of late smoking behaviors were assessed that focused on the period of heaviest smoking. Time to first cigarette after waking was derived from the question "During this period when you were smoking the

most, about how many minutes after you woke up did you smoke your first cigarette?" and the 4 answers: more than one hour, 31-60 minutes, 6-30 minutes, and within 5 minutes. For the primary analyses, time to first cigarette was dichotomized into >5 minutes and ≤5 minutes after waking. Cigarettes per day was evaluated with the question "During the period of time when you were smoking the most, about how many cigarettes did you usually have per day?" and the 4 answers: 10 or fewer, 11-20, 21-30, and 31 or more cigarettes. Cigarettes per day was dichotomized into ≤20 and >20 cigarettes in the primary analyses as done in previous studies (Belsky et al., 2013). A total Fagerstrom Test for Nicotine Dependence (FTND) score during the heaviest period of smoking was calculated at each interview using responses to these 2 questions as well as responses to questions assessing the four remaining criteria (details in Heatherton et al., 1991). For the primary analyses, nicotine dependence was defined as a FTND score of 4 or more.

Given the longitudinal design of this study, an endorsement of smoking initiation or daily smoking at any interview was used to capture these behaviors. The highest FTND score across available interviews was chosen to capture the lifetime maximum, and the variables time to first cigarette as well as cigarettes per day were set at these same interviews.

Genotyping

Recently, Bloom et al. (2011) developed a metric based on several genetic variants in *CYP2A6* to estimate nicotine metabolism. Cross-validation estimates that this metric predicts approximately 70% of the variance in metabolism of orally administered nicotine to cotinine in European Americans. Our goal was to use this *CYP2A6* metabolism metric to test whether *CYP2A6* variation predicts cigarette smoking behaviors.

Blood samples from COGA were obtained for genetic analysis. Five *CYP2A6* single nucleotide polymorphisms (SNPs) (rs1801272, rs28399442, rs28399433, rs1137115, rs28399435) were genotyped using the LGC Genomics Competitive Allele-Specific PCR (KASP), a FRET-based endpoint genotyping assay (http://www.lgcgenomics.com). PCR reactions were run on an ABI GeneAmp PCR System 9700, and fluorescence measurements taken on an ABI 7900HT Fast Real-Time PCR System (http://www.lifetechnologies.com). The *CYP2A6* copy number variant (CNV) was genotyped with TaqMan 5' Nuclease Assays (Hs00010002_cn and Hs07545275_cn, Life Technologies) using a standard qPCR protocol on the ABI 7900HT System. The CNV assay was run in duplicate, and genotype calls were made using CopyCaller software. The program PEDCHECK (O'Connell and Weeks, 1998) was used to examine Mendelian inheritance, and only individuals with no Mendelian inconsistencies were included in the genotyped sample. The metabolism metric was calculated based on the genotypes of the five *CYP2A6* SNPs and the CNV using an algorithm described in **Supplemental Table 4.1** (adapted from Bloom et al., 2012).

A set of 64 ancestry informative markers was genotyped as part of a 96 SNP Biorepository Panel by the Rutgers University Cell and DNA Repository. These markers were used in SNPrelate, a function in R, to assign ancestry groups. HapMap populations were included as reference groups. There was high concordance (98%) between self-reported and genetically determined ethnicity among European Americans. Only genetically determined European Americans were included in the analysis because the metric was optimized for this population (Bloom et al., 2011).

Primary Sample Selection

In the COGA adolescent and young adult study, 1,102 European ancestry individuals with age of last interview between 19 and 30 were genotyped for the *CYP2A6* variants, and participants for the analyses were drawn from this group (**Figure 1**). The analysis was restricted to individuals who had reached young adulthood because we were interested in transitions to daily smoking and late smoking behaviors, outcomes that often occur during this time. The sample used to analyze daily versus non-daily smokers consisted of 706 (64%) individuals who had initiated smoking. For transitions to late smoking behaviors, we focused on the sample of 468 (66%) daily smokers (described in **Table 4.1**).

Replication COGEND sample

The Collaborative Study of Nicotine Dependence (COGEND) is a multi-center case-control study designed to identify genes that contribute to nicotine dependence (Saccone et al., 2007). Community based recruitment enrolled participants ages 25-45 years old. Cases were required to be current smokers and have an FTND score of 4 or more. Controls were required to have smoked at least 100 cigarettes and have a lifetime maximum FTND score of 1. For this analysis, only subjects who self-identified as being of European ancestry were examined, and previous analyses using EIGENSTRAT have shown a high correspondence with genetically determined ancestry groups (Saccone et al., 2009). Genotyping of variants to calculate the metabolism metric in COGEND has been previously described (Bloom et al., 2012). We focused on the subsample of 377 COGEND young adults ages 25-30, which overlapped with the age range of the COGA young adult sample. From this group, 335 (89%) reported smoking every

day or nearly every day for at least 2 months and were considered daily smokers. Replication sample characteristics of these daily smokers are described in **Table 4.1**.

Primary Data Analysis

Data were analyzed using the Statistical Analysis System. Logistic regression was used to model dichotomous outcomes of daily smoking, nicotine dependence, time to first cigarette, and cigarettes per day. In the primary analyses in COGA and COGEND, the continuous metabolism metric, sex, study site, and last interview age were included as variables. In COGA, family structure was accounted for using generalized estimating equations via PROC GENMOD. Results from the COGEND replication sample were meta-analyzed with the primary COGA results (**Table 4.2**) using a publically available SAS macro (http://www.hsph.harvard.edu/spiegelman/metaanal.html). Meta-analyses results were based on fixed effect models to determine the evidence for association within the collected samples. In these analyses, we did not observe heterogeneity between the two studies based on the Q statistic ($p>0.1$).

Secondary Data Analyses

Several secondary analyses were performed to test the robustness of our primary findings. First, individuals were divided into slow and normal metabolizers using a cut-off of ≤0.85 on the metabolism metric as done in previous studies (Chen et al., 2014). This cut-off represents approximately the lowest quartile of metabolizers and this dichotomous variable was examined in logistic regression models of smoking behaviors. Second, the 4 level variable of time to first cigarette after waking (>60, 31-60, 6-30, ≤5 minutes) was also investigated in

cumulative logistic regression models to assess whether the metabolism metric predicted response across these four ordinal categories. Third, because FTND contains information about time to first cigarette, we also examined the equation predicting dichotomous nicotine dependence with this added variable of time to first cigarette. Finally, we calculated FTND scores without the question regarding time to first cigarette after waking and explored the association between the metabolism metric and this new alternative nicotine dependence measure. Without the important question of time to first cigarette after waking, this alternative score was ≥4 for only 25% (115/461) of young adult daily smokers in COGA and 26% (87/335) in COGEND.

4.4 RESULTS

Participant Characteristics

Demographic, behavioral, and metabolism metric characteristics of the COGA and COGEND samples are presented in **Table 4.1**. The primary COGA sample of young adult daily smokers consisted of 468 European American individuals from 401 nuclear families and 293 extended families. The mean age at last interview was 24: 44% were female, and the majority came from families at high-risk for alcoholism (92%). Among these daily smokers, 59% were nicotine dependent, 32% smoked within 5 minutes after waking, and 26% smoked greater than 20 cigarettes per day (**Figure 4.1** and **Table 4.2**). Twenty-six percent of the young adults were slow metabolizers, and the distribution of the metabolism metric (**Supplemental Figure 4.1**) was similar to that seen in other samples (Bloom et al., 2012; Chen et al., 2014).

The COGEND replication sample of young adult daily smokers consisted of 335 European Americans with an average age at interview of 28, and the majority were female (61%). Among COGEND young adult daily smokers, 50% were nicotine dependent, 24% smoked within 5 minutes after waking, 25% smoked greater than 20 cigarettes per day, and 30% were slow metabolizers (distribution in **Supplemental Figure 4.1**) .

CYP2A6 Metabolism Metric and Early Smoking Behaviors

The *CYP2A6* metabolism metric was not associated with smoking initiation (p=0.51) and the development of daily smoking (p=0.57) in the COGA young adults (**Table 4.2**). Therefore, subsequent analyses of late smoking milestones focused on the 468 daily smokers. Of the 238 young adults who initiated smoking but did not transition to daily smoking, essentially all of

them (99.6%) failed to develop any of the late smoking behaviors. This supports the notion that daily smoking is a prerequisite for the development of late smoking behaviors.

CYP2A6 **Metabolism Metric and Late Smoking Behaviors in Daily Smokers**

CYP2A6 haplotypes predictive of slower metabolism were associated with an increased risk of nicotine dependence in both the primary COGA and replication COGEND samples of young adult daily smokers (**Table 4.2, Figure 4.2**). In multivariate models adjusting for age, sex, and study site, the continuous *CYP2A6* metabolism metric had a significant effect in COGA (p =0.03) and COGEND (p=0.02), where a slow predicted metabolism was associated with an increased risk of nicotine dependence defined by an FTND score ≥4 (**Table 4.2**). Secondary analyses showed that slow metabolizers (defined by a metric of ≤0.85) had a 1.64 increased odds (95% CI 1.17-2.28, p=0.004) of developing nicotine dependence as compared to normal metabolizers (metric>0.85) in meta-analyses of COGA and COGEND studies (**Supplemental Table 4.2**). **Figure 4.2** illustrates this association by showing that a larger proportion of slow metabolizers in both COGA and COGEND developed nicotine dependence as compared to normal metabolizers.

Consistent with the nicotine dependence results, a lower metabolism metric was associated with an increased risk of smoking within 5 minutes after waking (**Table 4.2, Figure 4.2**). The continuous *CYP2A6* metabolism metric had a trending effect in COGA (p=0.07) and a significant effect in COGEND (p=0.01). In meta-analysis, slow metabolizers had a 1.61 increased odds (95% CI 1.15-2.26, p=0.006) of smoking within 5 minutes after waking compared to normal metabolizers (**Supplemental Table 4.2, Figure 4.2**). The *CYP2A6* metabolism metric was not associated with smoking more than 20 cigarettes per day in both

samples (**Table 4.2**), and **Figure 4.2** illustrates that a similar proportion of slow and fast metabolizers reported smoking more than 20 cigarettes per day.

Robustness of effect of *CYP2A6* Metabolism Metric on Time to First Cigarette after Waking

Examination of all 4 responses of time to first cigarette after waking (≥60, 31-60, 6-30, ≤5 minutes) in secondary analyses demonstrated a slightly more robust effect of the metabolism metric in both COGA (p=0.06) and COGEND (p=0.004) as compared to the dichotomous variable of time to first cigarette (>5 and ≤5 minutes) (**Supplemental Table 4.3** and **Table 4.2**, respectively). **Supplemental Figure 4.2** illustrates that across the 4 categories, there was an increased proportion of slow metabolizers at shorter times to first cigarette after waking among COGA daily smokers. In COGEND daily smokers, we observed a similar trend, except in the category of 31-60 minutes that only had 19 individuals (6% of sample, **Table 4.1**). Taken together these results support a possible "dosage effect" where predicted slower metabolism was correlated with smoking sooner after waking.

Since time to first cigarette after waking contributes to the calculation of FTND score, we also examined the effect of the metabolism metric on nicotine dependence after controlling for time to first cigarette after waking as a variable. In COGEND, the metabolism metric was no longer a significant predictor of nicotine dependence (p=0.23) and in COGA the significance was diminished (p=0.04) (**Supplemental Table 4.4**). Furthermore, when FTND scores were calculated excluding the time to first cigarette after waking question, there was essentially no association between the metabolism metric and this adjusted measure of nicotine dependence (COGA p=0.51 and COGEND p=0.69, **Supplemental Table 4.4**). These results suggest that our

primary association between the *CYP2A6* metabolism metric and nicotine dependence is driven by time to first cigarette after waking.

4.5 DISCUSSION

Young adulthood is a critical developmental period for the progression from initiation to late smoking milestones (NSDUH, 2012). This study links variation in a genome-wide significant gene, *CYP2A6*, with the development of smoking behaviors in two independent samples of European American young adults. Using specific *CYP2A6* polymorphisms, we calculated a nicotine metabolism metric, which has been previously shown to account for approximately 70% of the variance in metabolism of orally administered nicotine to cotinine in European Americans (Bloom et al., 2012; Bloom et al., 2011). Our primary finding is that decreased predicted nicotine metabolism is associated with an elevated risk of developing nicotine dependence among young adult daily smokers, adding important insight into the role of variation in *CYP2A6* across stages of smoking development, as illustrated in **Figure 4.3**.

Despite having an important role in the development of nicotine dependence, variation in *CYP2A6* was not associated with smoking initiation or the progression to daily smoking (**step 1 in Figure 4.3**). Previous twin studies support that environmental influences primarily drive early adolescent nicotine use, and that the role of heritable factors on smoking behaviors increases throughout young adulthood (Kendler et al., 2008; Koopmans et al., 1999). Our results are consistent with this model by providing evidence of a gene that impacts the transition from daily smoking to nicotine dependence, without influencing initiation and daily smoking.

The observation that decreased predicted nicotine metabolism is associated with increased risk of nicotine dependence in young adults also builds on previous studies conducted in adolescents (**step 2 in Figure 4.3**). O'Loughlin et al. (2004) followed 228 non-dependent smokers in grade 7 over approximately 30 months and found that those with inactive genetic

variants in *CYP2A6* were more likely to develop nicotine dependence by the international classification of diseases (ICD-10) criteria, but smoked fewer cigarettes per day once dependent. Huang et al. (2005) examined variation in *CYP2A6* in 1,518 adolescents enrolled in a longitudinal study in the United Kingdom and similarly found that individuals with inactive variants associated with slower metabolism were more likely to be current versus former smokers at age 18 compared to normal metabolizers. More recently, Rubinstein et al. (2013) assessed a biomarker of the rate of nicotine metabolism (the nicotine metabolite ratio) in 164 adolescent smokers and found that slower metabolizers showed greater symptoms of dependence on the modified Fagerstrom Tolerance Questionnaire. Our findings expand on these earlier results by demonstrating that during young adulthood, when many late smoking behaviors develop, slow metabolizers continue to have a greater risk of dependence.

The increased susceptibility to developing nicotine dependence encountered by youth slow metabolizers has been hypothesized to reflect prolonged exposure to nicotine during initial smoking experiences (Chenoweth et al., 2013; Malaiyandi et al., 2005; Rubinstein et al., 2013). Although accumulating evidence support this role, it is important to note that a few studies show the opposite effect where slow metabolism is associated with *decreased risk* of smoking behaviors in youth (Audrain-McGovern et al., 2007; Moolchan et al., 2009; Rubinstein et al., 2008). For example, Audrain-McGovern (2007) examined 222 European ancestry adolescent ever-smokers and found that normal *CYP2A6* metabolizers developed symptoms of dependence on the modified Fagerstrom Tolerance Questionnaire at a faster rate than slower *CYP2A6* metabolizers. Many possible explanations exist for these discrepant results, including differences in measures of nicotine metabolism and dependence.

Our results suggest that time to first cigarette after waking is a critical contributor to the association between the *CYP2A6* metabolism metric and nicotine dependence assessed by the FTND criteria. Little consensus exists on the best measure of nicotine dependence, but research supports that two items from the FTND score, time to first cigarette after waking and cigarettes per day, are strong, valid, reliable predictors of quitting behaviors, which are key indicators of dependence (Baker et al., 2007; Borland et al., 2010; Hyland et al., 2006). Studies also suggest that these two measures are distinct predictors of addiction (Borland et al., 2010; Lessov et al., 2004), suggesting the possibility that different genetic factors may contribute to urgency to smoke and levels of cigarette consumption. In a sample of over 1000 young adults, Haberstick et al. (2007) found that time to first cigarette was the most informative measure of heritable factors from the FTND score. Our results complement these findings by illustrating that necessity to smoke measured by time to first cigarette after waking drives the association of the *CYP2A6* metabolism metric and nicotine dependence in young adults.

These findings in young adults should be considered in the context of the adult literature. Previous studies of adults demonstrate that once dependent, genetically slower metabolizers smoke fewer cigarettes to reach target blood nicotine levels (Benowitz, 2008) (**step 3 in Figure 4.3**). Although we did not observe an effect of slow metabolism on risk of smoking more than 20 cigarettes per day among daily smokers (**Table 4.2, Figure 4.2**), only 26% of these young adults were heavy smokers, and heaviness of smoking continues to increase throughout adulthood (NSDUH). In the entire COGEND sample ages 25-45, previous work suggests that once dependent, slower metabolism is associated with decreased cigarette consumption (Bloom et al., 2012). These findings underscore that variation in *CYP2A6* has a variety of effects on smoking behaviors across stages of development in the COGEND sample: slow metabolism leads to

increased risk for developing nicotine dependence in young adult daily smokers, but once dependent, slow metabolism is protective for heaviness of smoking.

Another important consideration is that the fraction of slow metabolizers in the population of smokers has been observed to decrease with age, suggesting that slow metabolizers are more likely to quit smoking (Benowitz, 2008) (**step 4 in Figure 4.3**). Among COGEND dependent current smokers ages 25-30 years, we found that 36% (60/166) were slow metabolizers. However, among COGEND current dependent smokers over 30 years old, only 28% (250/883) were slow metabolizers, supporting that proportionally more slow metabolizers have quit by this time. Furthermore, other studies directly support that slow nicotine metabolism, measured by *CYP2A6* genotypes or the nicotine metabolite ratio, is associated with increased cessation rates in both youth (Chenoweth et al., 2013) and adults enrolled in clinical trials (Ray 2009; Chen 2014). Taken together, these findings suggest that across development, slow metabolizers may quit smoking more easily. Therefore, the observation that slow metabolism is associated with increased risk of nicotine dependence may be most pronounced in samples of youth when symptoms of dependence are first developing and before cessation attempts occur.

The findings reported here have limitations. First, this study focused on European Ancestry individuals because the metabolism metric was optimized for this population (Bloom et al., 2011). Second, the temporal ordering of smoking behaviors could not be examined in this analysis because the smoking questions did not assess age of onset in COGA, and COGEND is a cross-sectional study. Third, the majority of the COGA participants were from families at high risk for alcoholism, which may limit the generalizability of the findings. Replication of the

primary findings in a community based recruitment sample (COGEND), however, supports that the findings are not specific to a high-risk population.

In summary, using a validated *CYP2A6* metabolism metric, this study demonstrates that slower nicotine metabolism is associated with an increased risk of nicotine dependence in two independent samples of young adult daily smokers. These findings add important knowledge about the complex role of *CYP2A6* variation across different developmental stages of smoking stages. From a public health perspective, these findings and others (Belsky et al., 2013) provide a genetic argument in support of early interventions before the development of nicotine dependence.

4.6 ACKNOWLEDGEMENTS:

I would like to acknowledge my co-authors: Bloom J, Bertelsen S, Breslau N, Budde J, Chen LS, Culverhouse R, Chorlian D, Dick DM, Edenberg HJ, Hatsukami D, Hesselbrovck VM, Kramer JR, Kuperman S, Porjesz B, Saccone NL, Schuckit MA, Stitzel J, Tischfield JA, Goate A, and Bierut LJ.

COGA was supported by U10AA008401 and COGEND was supported by P01CA89392. EO was supported by T32GM07200, UL1TR000448, TL1TR000449, and F30AA023685.

Disclosure/Conflict of Interest:

LJB, AG, and the spouse of NLS are listed as inventors on Issued U.S. Patent 8,080,371,“Markers for Addiction” covering the use of certain SNPs in determining the diagnosis, prognosis, and treatment of addiction. The other authors declare no conflict of interest.

4.7 REFERENCES

Audrain-McGovern, J., Al Koudsi, N., Rodriguez, D., Wileyto, E.P., Shields, P.G., Tyndale, R.F., 2007. The role of CYP2A6 in the emergence of nicotine dependence in adolescents. Pediatrics 119, e264-274.

Baker, T.B., Piper, M.E., McCarthy, D.E., Bolt, D.M., Smith, S.S., Kim, S.Y., Colby, S., Conti, D., Giovino, G.A., Hatsukami, D., Hyland, A., Krishnan-Sarin, S., Niaura, R., Perkins, K.A., Toll, B.A., 2007. Time to first cigarette in the morning as an index of ability to quit smoking: implications for nicotine dependence. Nicotine & tobacco research : official journal of the Society for Research on Nicotine and Tobacco 9 Suppl 4, S555-570.

Begleiter, H., Reich, T., Hesselbrock, V., Porjesz, B., Li, T., Schuckit, M., Edenberg, H., Rice, J., 1995. The Collaborative Study on the Genetics of Alcoholism. Alcohol Health Res World 19, 228–236.

Belsky, D.W., Moffitt, T.E., Baker, T.B., Biddle, A.K., Evans, J.P., Harrington, H., Houts, R., Meier, M., Sugden, K., Williams, B., Poulton, R., Caspi, A., 2013. Polygenic risk and the developmental progression to heavy, persistent smoking and nicotine dependence: evidence from a 4-decade longitudinal study. JAMA psychiatry 70, 534-542.

Benowitz, N.L., 2008. Clinical pharmacology of nicotine: implications for understanding, preventing, and treating tobacco addiction. Clinical pharmacology and therapeutics 83, 531-541.

Bierut, L.J., 2011. Genetic vulnerability and susceptibility to substance dependence. Neuron 69, 618-627.

Bloom, A.J., Harari, O., Martinez, M., Madden, P.A., Martin, N.G., Montgomery, G.W., Rice, J.P., Murphy, S.E., Bierut, L.J., Goate, A., 2012. Use of a predictive model derived from in vivo endophenotype measurements to demonstrate associations with a complex locus, CYP2A6. Human molecular genetics 21, 3050-3062.

Bloom, J., Hinrichs, A.L., Wang, J.C., von Weymarn, L.B., Kharasch, E.D., Bierut, L.J., Goate, A., Murphy, S.E., 2011. The contribution of common CYP2A6 alleles to variation in nicotine metabolism among European-Americans. Pharmacogenetics and genomics 21, 403-416.

Borland, R., Yong, H.H., O'Connor, R.J., Hyland, A., Thompson, M.E., 2010. The reliability and predictive validity of the Heaviness of Smoking Index and its two components: findings from the International Tobacco Control Four Country study. Nicotine & tobacco research : official journal of the Society for Research on Nicotine and Tobacco 12 Suppl, S45-50.

Bucholz, K.K., Cadoret, R., Cloninger, C.R., Dinwiddie, S.H., Hesselbrock, V.M., Nurnberger, J.I., Jr., Reich, T., Schmidt, I., Schuckit, M.A., 1994. A new, semi-structured psychiatric interview for use in genetic linkage studies: a report on the reliability of the SSAGA. Journal of studies on alcohol 55, 149-158.

Bucholz, K.K., Hesselbrock, V.M., Shayka, J.J., Nurnberger, J.I., Jr., Schuckit, M.A., Schmidt, I., Reich, T., 1995. Reliability of individual diagnostic criterion items for psychoactive substance dependence and the impact on diagnosis. Journal of studies on alcohol 56, 500-505.

Chen, L.S., Bloom, A.J., Baker, T.B., Smith, S.S., Piper, M.E., Martinez, M., Saccone, N., Hatsukami, D., Goate, A., Bierut, L., 2014. Pharmacotherapy effects on smoking

cessation vary with nicotine metabolism gene (CYP2A6). Addiction (Abingdon, England) 109, 128-137.

Chenoweth, M.J., O'Loughlin, J., Sylvestre, M.P., Tyndale, R.F., 2013. CYP2A6 slow nicotine metabolism is associated with increased quitting by adolescent smokers. Pharmacogenetics and genomics 23, 232-235.

Haberl, M., Anwald, B., Klein, K., Weil, R., Fuss, C., Gepdiremen, A., Zanger, U.M., Meyer, U.A., Wojnowski, L., 2005. Three haplotypes associated with CYP2A6 phenotypes in Caucasians. Pharmacogenetics and genomics 15, 609-624.

Haberstick, B.C., Timberlake, D., Ehringer, M.A., Lessem, J.M., Hopfer, C.J., Smolen, A., Hewitt, J.K., 2007. Genes, time to first cigarette and nicotine dependence in a general population sample of young adults. Addiction (Abingdon, England) 102, 655-665.

Heatherton, T.F., Kozlowski, L.T., Frecker, R.C., Fagerstrom, K.O., 1991. The Fagerstrom Test for Nicotine Dependence: a revision of the Fagerstrom Tolerance Questionnaire. British journal of addiction 86, 1119-1127.

Hesselbrock, M., Easton, C., Bucholz, K.K., Schuckit, M., Hesselbrock, V., 1999. A validity study of the SSAGA--a comparison with the SCAN. Addiction (Abingdon, England) 94, 1361-1370.

Huang, S., Cook, D.G., Hinks, L.J., Chen, X.H., Ye, S., Gilg, J.A., Jarvis, M.J., Whincup, P.H., Day, I.N., 2005. CYP2A6, MAOA, DBH, DRD4, and 5HT2A genotypes, smoking behaviour and cotinine levels in 1518 UK adolescents. Pharmacogenetics and genomics 15, 839-850.

Hukkanen, J., Jacob, P., 3rd, Benowitz, N.L., 2005. Metabolism and disposition kinetics of nicotine. Pharmacological reviews 57, 79-115.

Hyland, A., Borland, R., Li, Q., Yong, H.H., McNeill, A., Fong, G.T., O'Connor, R.J., Cummings, K.M., 2006. Individual-level predictors of cessation behaviours among participants in the International Tobacco Control (ITC) Four Country Survey. Tobacco control 15 Suppl 3, iii83-94.

Kendler, K.S., Schmitt, E., Aggen, S.H., Prescott, C.A., 2008. Genetic and environmental influences on alcohol, caffeine, cannabis, and nicotine use from early adolescence to middle adulthood. Archives of general psychiatry 65, 674-682.

Koopmans, J., Slutske, W., Heath, A., Neale, M., Boomsma, D., 1999. The Genetics of Smoking Initiation and Quantity Smoked in Dutch Adolescent and Young Adult Twins. Behavior genetics 29, 383-393.

Lessov, C.N., Martin, N.G., Statham, D.J., Todorov, A.A., Slutske, W.S., Bucholz, K.K., Heath, A.C., Madden, P.A., 2004. Defining nicotine dependence for genetic research: evidence from Australian twins. Psychological medicine 34, 865-879.

Malaiyandi, V., Sellers, E.M., Tyndale, R.F., 2005. Implications of CYP2A6 genetic variation for smoking behaviors and nicotine dependence. Clinical pharmacology and therapeutics 77, 145-158.

Moolchan, E.T., Parzynski, C.S., Jaszyna-Gasior, M., Collins, C.C., Leff, M.K., Zimmerman, D.L., 2009. A link between adolescent nicotine metabolism and smoking topography. Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology 18, 1578-1583.

Munafo, M.R., Timofeeva, M.N., Morris, R.W., Prieto-Merino, D., Sattar, N., Brennan, P., Johnstone, E.C., Relton, C., Johnson, P.C., Walther, D., Whincup, P.H., Casas, J.P., Uhl,

G.R., Vineis, P., Padmanabhan, S., Jefferis, B.J., Amuzu, A., Riboli, E., Upton, M.N., Aveyard, P., Ebrahim, S., Hingorani, A.D., Watt, G., Palmer, T.M., Timpson, N.J., Davey Smith, G., 2012. Association between genetic variants on chromosome 15q25 locus and objective measures of tobacco exposure. Journal of the National Cancer Institute 104, 740-748.

NSDUH, 2012. United States Department of Health and Human Services. Substance Abuse and Mental Health Services Administration. Center for Behavioral Health Statistics and Quality. National Survey on Drug Use and Health, 2012. ICPSR34933-v2. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2014-10-06. http://doi.org/10.3886/ICPSR34933.v2. http://www.samhsa.gov/data/NSDUH/2012SummNatFindDetTables/Index.aspx.accessed on.

O'Connell, J.R., Weeks, D.E., 1998. PedCheck: a program for identification of genotype incompatibilities in linkage analysis. American journal of human genetics 63, 259-266.

O'Loughlin, J., Paradis, G., Kim, W., DiFranza, J., Meshefedjian, G., McMillan-Davey, E., Wong, S., Hanley, J., Tyndale, R.F., 2004. Genetically decreased CYP2A6 and the risk of tobacco dependence: a prospective study of novice smokers. Tobacco control 13, 422-428.

Ray, R., Tyndale, R.F., Lerman, C., 2009. Nicotine dependence pharmacogenetics: role of genetic variation in nicotine-metabolizing enzymes. Journal of neurogenetics 23, 252-261.

Rubinstein, M.L., Benowitz, N.L., Auerback, G.M., Moscicki, A.B., 2008. Rate of nicotine metabolism and withdrawal symptoms in adolescent light smokers. Pediatrics 122, e643-647.

Rubinstein, M.L., Shiffman, S., Moscicki, A.B., Rait, M.A., Sen, S., Benowitz, N.L., 2013. Nicotine metabolism and addiction among adolescent smokers. Addiction (Abingdon, England) 108, 406-412.

Saccone, N.L., Wang, J.C., Breslau, N., Johnson, E.O., Hatsukami, D., Saccone, S.F., Grucza, R.A., Sun, L., Duan, W., Budde, J., Culverhouse, R.C., Fox, L., Hinrichs, A.L., Steinbach, J.H., Wu, M., Rice, J.P., Goate, A.M., Bierut, L.J., 2009. The CHRNA5-CHRNA3-CHRNB4 nicotinic receptor subunit gene cluster affects risk for nicotine dependence in African-Americans and in European-Americans. Cancer research 69, 6848-6856.

Saccone, S.F., Hinrichs, A.L., Saccone, N.L., Chase, G.A., Konvicka, K., Madden, P.A., Breslau, N., Johnson, E.O., Hatsukami, D., Pomerleau, O., Swan, G.E., Goate, A.M., Rutter, J., Bertelsen, S., Fox, L., Fugman, D., Martin, N.G., Montgomery, G.W., Wang, J.C., Ballinger, D.G., Rice, J.P., Bierut, L.J., 2007. Cholinergic nicotinic receptor genes implicated in a nicotine dependence association study targeting 348 candidate genes with 3713 SNPs. Human molecular genetics 16, 36-49.

TAG, 2010. Genome-wide meta-analyses identify multiple loci associated with smoking behavior. Nature genetics 42, 441-447.

Thorgeirsson, T.E., Gudbjartsson, D.F., Surakka, I., Vink, J.M., Amin, N., Geller, F., Sulem, P., Rafnar, T., Esko, T., Walter, S., Gieger, C., Rawal, R., Mangino, M., Prokopenko, I., Magi, R., Keskitalo, K., Gudjonsdottir, I.H., Gretarsdottir, S., Stefansson, H., Thompson,

J.R., Aulchenko, Y.S., Nelis, M., Aben, K.K., den Heijer, M., Dirksen, A., Ashraf, H., Soranzo, N., Valdes, A.M., Steves, C., Uitterlinden, A.G., Hofman, A., Tonjes, A., Kovacs, P., Hottenga, J.J., Willemsen, G., Vogelzangs, N., Doring, A., Dahmen, N., Nitz, B., Pergadia, M.L., Saez, B., De Diego, V., Lezcano, V., Garcia-Prats, M.D., Ripatti, S., Perola, M., Kettunen, J., Hartikainen, A.L., Pouta, A., Laitinen, J., Isohanni, M., Huei-Yi, S., Allen, M., Krestyaninova, M., Hall, A.S., Jones, G.T., van Rij, A.M., Mueller, T., Dieplinger, B., Haltmayer, M., Jonsson, S., Matthiasson, S.E., Oskarsson, H., Tyrfingsson, T., Kiemeney, L.A., Mayordomo, J.I., Lindholt, J.S., Pedersen, J.H., Franklin, W.A., Wolf, H., Montgomery, G.W., Heath, A.C., Martin, N.G., Madden, P.A., Giegling, I., Rujescu, D., Jarvelin, M.R., Salomaa, V., Stumvoll, M., Spector, T.D., Wichmann, H.E., Metspalu, A., Samani, N.J., Penninx, B.W., Oostra, B.A., Boomsma, D.I., Tiemeier, H., van Duijn, C.M., Kaprio, J., Gulcher, J.R., McCarthy, M.I., Peltonen, L., Thorsteinsdottir, U., Stefansson, K., 2010. Sequence variants at CHRNB3-CHRNA6 and CYP2A6 affect smoking behavior. Nature genetics 42, 448-453.

4.8 TABLES

Table 4.1. Characteristics of primary and replication samples of European American young adults

Characteristic	COGA Young Adult European American Daily Smokers (N=468)	COGEND Young Adult European American Daily Smokers (N=335)
Sex, N (%)		
Males	264 (56%)	129 (39%)
Females	204 (44%)	206 (61%)
Age at last interview, years		
Mean ± sd	23.6 ± 3.0	27.8 ± 1.7
Range	19-30	25-30
No. of interviews		
Mean ± sd	3.6 ± 1.2	-
Range	1-5	1
Family status, N (%)		
From high-risk families	431 (92%)	-
From comparison families	37 (8%)	-
No. of extended families	293	-
No. of nuclear families (full siblings)	401	-
FTND score		
Mean ± sd	4.1 ± 2.6	3.0 ± 3.3
Range	0-10	0-10
Nicotine dependence (FTND≥4), N (%)	276 (59%)	166 (50%)
Time to first cigarette after waking		
More than 1 hour	86 (18%)	168 (50%)
31-60 minutes	67 (15%)	19 (6%)
6-30 minutes	167 (36%)	67 (20%)
Within 5 minutes	148 (32%)	81 (24%)
Cigarettes per day		
10 or fewer	171 (37%)	171 (51%)
11-20	169 (37%)	78 (23%)
21-30	72 (16%)	45 (13%)
31 or more	48 (10%)	41 (12%)
Metabolism metric*		
Mean ± sd	0.86 ± 0.07	0.86 ± 0.07
Range	0.44-0.90	0.44-0.90
Metabolism status		
Low (Metric ≤ .85)	123 (26%)	103 (31%)
Normal (Metric > .85)	345 (74%)	232 (69%)

*Distribution of metabolism metric in COGA and COGEND young adult daily smokers provided in Supplemental Figure 4.1.

Table 4.2. Logistic regression models of smoking milestones in young adults

	Metabolism Metric in COGA Young Adults			Metabolism Metric in COGEND Yong Adults			Meta-analysis of results		
	Estimate	SE	p-value	Estimate	SE	p-value	Estimate	SE	p-value
Among all young adults (COGA n=1,102)									
Smoking initiation	0.67	1.03	0.51	-	-	-	-	-	-
Among young adult ever-smokers (COGA n=706)									
Daily smoking	-0.68	1.22	0.57	-	-	-	-	-	-
Among young adult daily smokers (COGA n=468; COGEND n=335)									
Nicotine dependence	3.94	1.76	0.03	4.36	1.86	0.02	4.14	1.28	0.001
Smoked within 5 minutes after waking	2.44	1.35	0.07	4.63	1.82	0.01	3.22	1.09	0.003
Smoked 20 cigarettes per day	-0.58	1.49	0.70	1.53	1.85	0.41	0.25	1.16	0.83

All models include sex, study site, and age of last interview as covariates; Analyses with COGA were also adjusted for familial clustering;

4.9 FIGURES

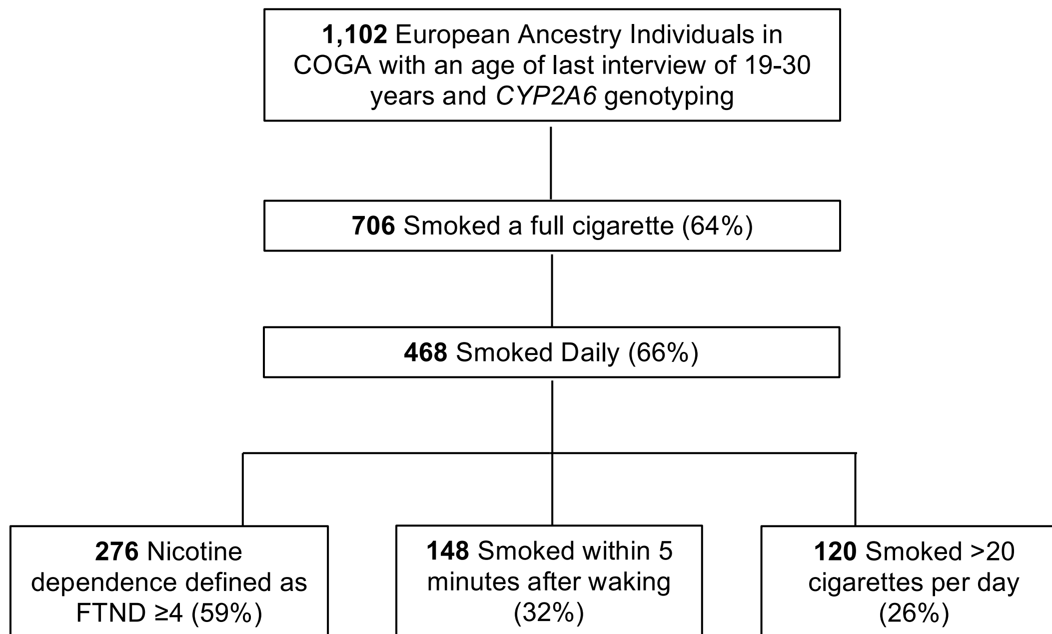


Figure 4.1. Primary COGA sample selection

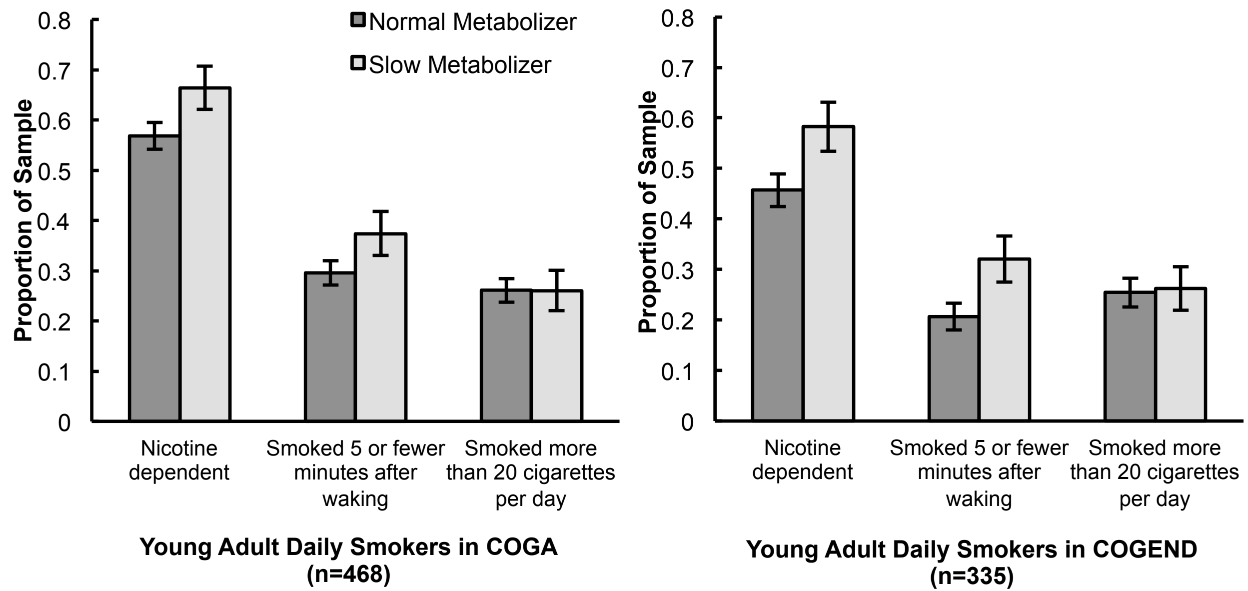


Figure 4.2. Association between predicted metabolism and smoking behaviors in two studies of European American young adult daily smokers. Error bars reflect standard errors.

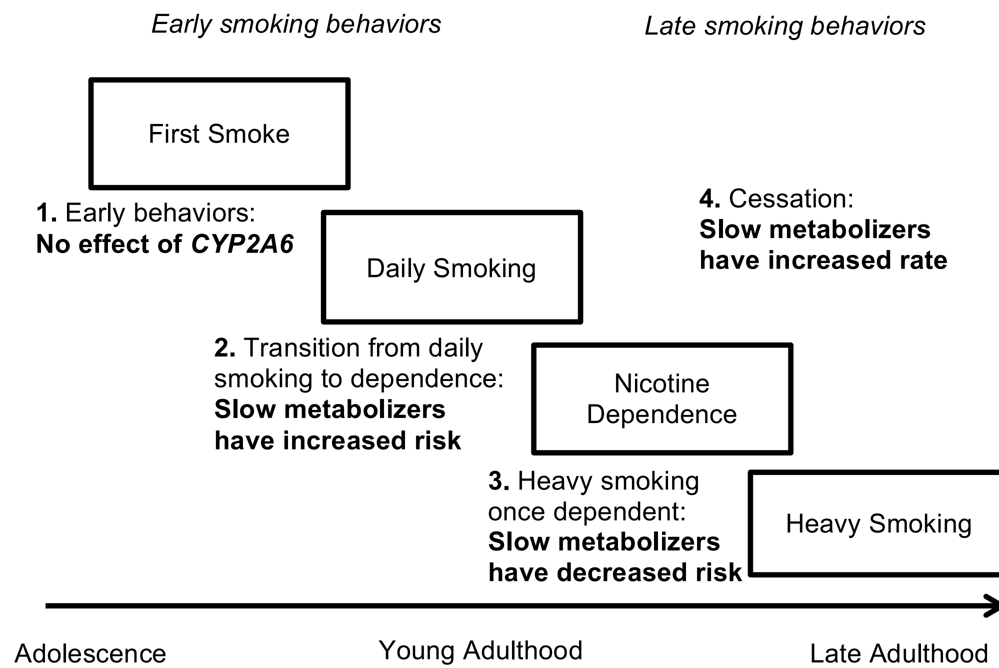


Figure 4.3. A theoretical framework of the development of smoking behaviors in relation to *CYP2A6* variation

4.10 SUPPLEMENTAL TABLES

Supplemental Table 4.1. Predicted metabolism metric and *CYP2A6* diplotypes based on copy number and 5 SNPs

CYP2A6 Copy number	rs1801272+ rs28399442	rs28399433	rs1137115	rs8399435	Haplotype 1	Haplotype 2	Predicted Metabolism metric
0					*4	*4	0.44
1	≥1				*4	*2/*12	0.44
	0	≥1			*4	*9	0.64
		0	≥1	0	*4	*1A	0.68
				≥1	*4	other	0.76
			0		*4	other	0.76
≥2	2				*2/*12	*2/*12	0.44
	1	≥1			*2/*12	*9	0.64
		0	≥1	0	*2/*12	*1A	0.68
				≥1	*2/*12	other	0.76
			0		*2/*12	other	0.76
	0	2			*9	*9	0.76
		1	1	0	*9	*1A	0.79
				1	*9	other	0.85
			0		*9	other	0.85
		0	2	0	*1A	*1A	0.82
				1	*1A	other	0.87
				2	other	other	0.90
			1	0	*1A	other	0.87
				1	other	other	0.90
			0		other	other	0.90

Adapted from Bloom *et al.* 2012;

Supplemental Table 4.2. Examination of dichotomous metabolism status as a predictor of smoking milestones in logistic regression models in young adults

	Metabolizer status in COGA Young Adults		Metabolizer status in COGEND Young Adults		Meta-analysis of results	
	Odds Ratio (95% CI)	p value	Odds ratio (95% CI)	p value	Odds ratio (95% CI)	p value
Among all young adults (COGA n=1,102)						
Smoking initiation	0.97 (0.72-1.30)	0.82	-	-	-	-
Among young adults ever-smokers (COGA n=706)						
Daily smoking	0.78 (0.54-1.04)	0.18	-	-	-	-
Among young adults who smoked daily (COGA n=468; COGEND n=335)						
Nicotine dependence	1.51 (0.96-2.38)	0.08	1.79 (1.10-2.91)	0.02	1.64 (1.17-2.28)	0.004
Smoke within 5 minutes	1.44 (0.93-2.23)	0.10	1.89 (1.11-3.22)	0.01	1.61 (1.15-2.26)	0.006
Smoke greater than 20	0.98 (0.62-1.56)	0.94	1.13 (0.76-2.29)	0.67	1.04 (0.73-1.48)	0.82

All models include sex, study site, and age of last interview as covariates; Analyses with COGA were adjusted for familial clustering.

Supplemental Table 4.3. Logistic regression models of the ordinal 4 level variable of time to first cigarette after waking (>60, 31-60, 6-30, ≤5 minutes)

	Metabolism Metric in COGA Young Adult Daily smokers (n=468)			Metabolism Metric in COGEND Yong Adult Daily Smokers (n=335)			Meta-analysis of results of Daily Smokers		
	Beta	SE	p value	Beta	SE	p value	Beta	SE	p value
Time to first cigarette after waking	2.43	1.28	0.06	4.55	1.58	0.004	3.27	0.99	0.001

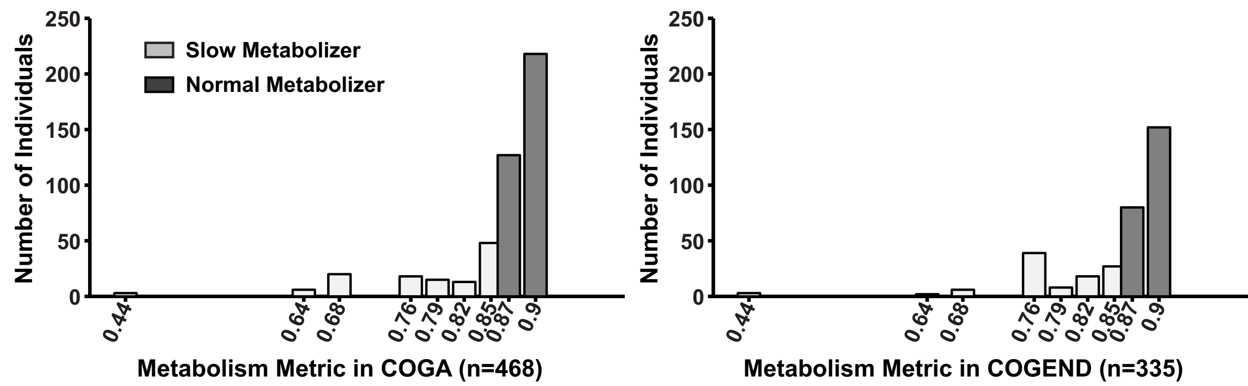
All models include sex, study site, and age of last interview as covariates; Analyses with COGA were also adjusted for familial clustering.

Supplemental Table 4.4. Logistic regression models of nicotine dependence that explore the importance of time to first cigarette after waking

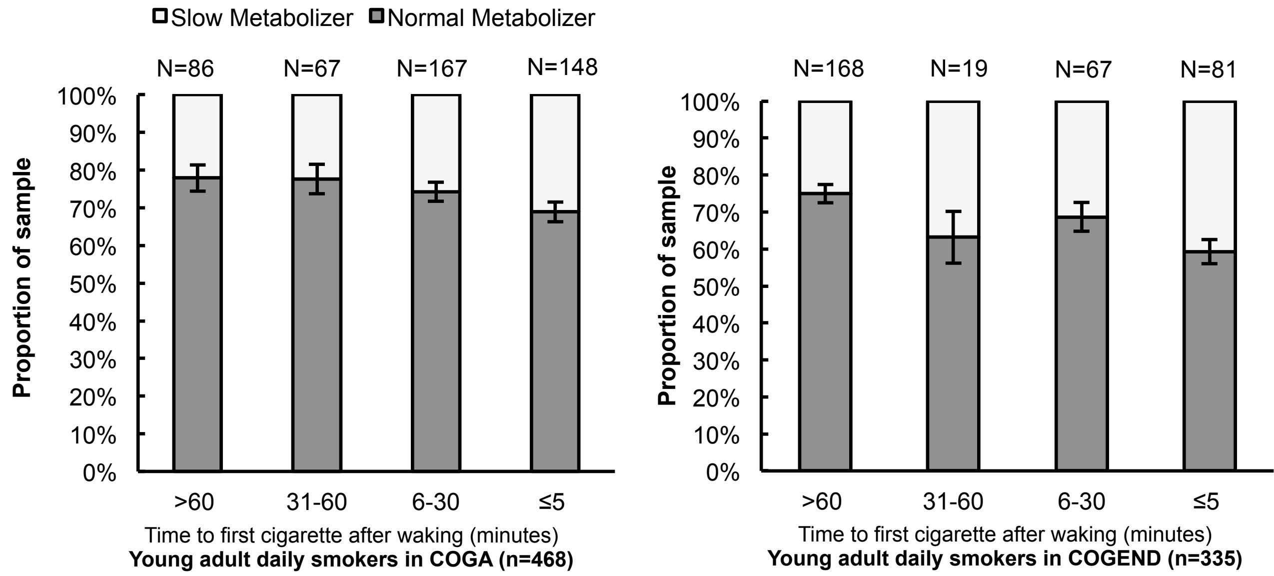
	Metabolism Metric in COGA Young Adult Daily smokers (n=468)			Metabolism Metric in COGEND Yong Adult Daily Smokers (n=335)			Meta-analysis of results of Daily Smokers		
	Beta	SE	p value	Beta	SE	p value	Beta	SE	p value
Nicotine dependence conditioned on time to first cigarette after waking	5.25	2.54	0.04	-18.66	15.7	0.23	4.63	2.25	0.07
Nicotine dependence based on adjusted FTND scores calculated without the time to first cigarette after waking question	-1.09	1.65	0.51	0.74	2.86	0.69	-0.28	1.23	0.82

All models include sex, study site, and age of last interview; Analyses with COGA were also adjusted for familial clustering;

4.11 SUPPLEMENTAL FIGURES



Supplemental Figure 4.1. Distribution of *CYP2A6* metabolism metric among COGA and COGEND young adult daily smokers.



Supplemental Figure 4.2. Proportion of slow metabolizers among 4 categories of time to first cigarette after waking among young adult daily smokers

CHAPTER FIVE:

Common, low frequency, and rare coding variants in *CHRNA5* contribute to nicotine dependence in European and African Americans

5.1 ABSTRACT

The functional nonsynonymous variant rs16969968 in the α5 nicotinic receptor subunit gene (*CHRNA5*) is the strongest genetic risk factor for nicotine dependence in European Americans (MAF=0.35), and contributes to risk in African Americans (MAF=0.06). To comprehensively examine whether other *CHRNA5* coding variation influences nicotine dependence risk, we performed targeted sequencing on 1 582 nicotine dependent cases (Fagerstrom Test for Nicotine Dependence score≥4) and 1 238 controls ages 25-45, with independent replication of common and low frequency variants using 12 studies with exome chip data. Next-generation sequencing with 180X coverage identified 24 nonsynonymous variants and 2 frameshift deletions in *CHRNA5*, including 9 novel variants. Nicotine dependence was examined using logistic regression and the variables sex, age, ancestry PCs, individual common variants (MAF≥0.05), aggregate low frequency variants (0.05>MAF≥0.005), and aggregate rare variants (MAF<0.005). Meta-analysis of primary results with replication studies containing 12 174 heavy and 11 290 light smokers confirmed robust independent risk effects of the only common variant (rs16969968, European: OR=1.3, p=3.5×10^{-11}; African: OR=1.3, p=0.01) and 3 low frequency variants (aggregate term, European: OR=1.3, p=0.005; African: OR=1.4, p=0.0006). The remaining 22 rare coding variants were associated with increased risk in the European American primary sample (OR=12.9, p=0.01) and in the risk direction in African Americans (OR=1.5, p=0.37). Beyond the well-studied rs16969968, we show that low frequency and rare *CHRNA5* coding variants are independently associated with nicotine dependence risk. These newly identified variants may have important health implications by influencing risk for smoking-related diseases and response to cessation therapies.

5.2 INTRODUCTION

Nicotine is the primary addictive component of tobacco products, and its physiological effects are mediated largely through neuronal nicotinic acetylcholine receptors (Dani and De Biasi, 2001). The α5/α3/β4 nicotinic subunit gene cluster on chromosome 15 harbors the strongest and most replicated genetic risk factor for several smoking related traits. Specifically, many independent studies demonstrated that rs16969968, a single nucleotide polymorphism in the α5 subunit gene (*CHRNA5*), is associated with nicotine dependence, cigarettes per day, smoking cessation, chronic obstructive pulmonary disease, and lung cancer (Amos *et al*, 2008; Berrettini *et al*, 2008; Bierut *et al*, 2008; Caporaso *et al*, 2009; Chen *et al*, 2012a; Chen *et al*, 2009; Hung *et al*, 2008; Pillai *et al*, 2009; Saccone *et al*, 2007; Thorgeirsson *et al*, 2008). Subsequent large-scale meta-analyses of European ancestry populations, where the rs16969968 minor allele is common (minor allele frequency (MAF)=0.35 in Exome Variant Server), identified this region as unequivocally associated with heaviness of smoking ($p=5.57x10^{-72}$) (Liu *et al*, 2010; TAG, 2010; Thorgeirsson *et al*, 2010). Recently, rs16969968 was shown to have a similar effect in African ancestry populations (Chen *et al*, 2012b; Saccone *et al*, 2009), where the minor allele is less common (MAF=0.06 in Exome Variant Server)(http://evs.gs.washington.edu/EVS/).

Beyond robust association studies across ancestry groups, functional studies support the biological role of *CHRNA5* and rs16969968 in the development of nicotine dependence. The highest density of α5 subunits has been reported in the interpeduncular nucleus in the brain, which receives input from the medial habenula (Hsu *et al*, 2013; Marks *et al*, 1992). Fowler et al. (2011) demonstrated that mice with a null mutation for *Chrna5* exhibited increased nicotine

intake, which was “rescued” by re-expression of α5 in the medial habenula. Their findings support the hypothesis that nicotine activates the habenulo-interpeduncular pathway through α5 containing receptors, limiting further nicotine intake. Genetic alterations that decrease the function of α5 would therefore be expected to increase liability towards nicotine dependence. Consistent with this hypothesis, the risk allele of rs16969968 has been shown to reduce receptor function (Bierut *et al*, 2008; Kuryatov *et al*, 2011). Specifically, the A allele of rs16969968 causes an aspartic acid to asparagine change at position 398 in the α5 subunit, and expression of this risk allele leads to decreased response to nicotine agonists in cell culture (Bierut *et al*, 2008) as well as lower Ca^{2+} permeability and increased short term desensitization when incorporated into certain neuronal nicotinic receptors (Kuryatov *et al*, 2011).

We hypothesized that additional low frequency and rare α5 coding variants may alter risk for nicotine dependence. To comprehensively assess the relationship between *CHRNA5* coding variation and liability to nicotine dependence, we analyzed targeted sequence data from approximately 3 000 nicotine dependent cases and non-dependent controls of European and African descent. In addition, we used 12 studies with exome chip data for replication analysis of the associations of common and low frequency variants with smoking behaviors found in our primary data.

5.3 MATERIALS AND METHODS

Primary Sample Ascertainment and Description

Subjects were recruited from the St Louis MO, Detroit MI, and Chicago IL metropolitan areas through the Collaborative Genetic Study of Nicotine Dependence and the Genetic Study of Nicotine Dependence in African Americans (Bierut *et al*, 2007; Saccone *et al*, 2007). Community-based recruitment enrolled subjects aged 25-45 years old. All subjects underwent comprehensive phenotypic assessments of smoking behaviors, including the Fagerstrom Test for Nicotine Dependence (FTND). Nicotine dependent cases were required to be current smokers and have an FTND score of 4 or higher. Non-dependent controls had smoked at least 100 cigarettes (to ensure exposure to nicotine) but had a lifetime maximum FTND score of 1.

Targeted sequencing of *CHRNA5*

DNA samples were derived from blood. The Center for Inherited Disease Research (CIDR) performed next-generation targeted sequencing on genes strongly associated with smoking, including *CHRNA5*. Details of the sequencing procedures and quality control measures are provided in the **Supplemental Methods**. The mean on-target coverage was 180X, and greater than 96% of on-target bases had a depth greater than 20X.

Evaluation of *CHRNA5* coding variants

Genotypic data that passed initial quality control at CIDR were released to the Quality Assurance/Quality Control analysis team at the University of Washington Genetics Coordinating Center. *CHRNA5* coding variants were identified by ANNOVAR (Wang *et al*, 2010) and then

manually reviewed. This review involved examining summary statistics of the quality control metrics, comparing the quality of novel variants with known variants from dbSNP and HapMap, as well as inspecting alignments of selected samples with non-reference calls to pass or fail variant sites. Seven samples were identified as low quality and omitted. Large genetic databases (Abecasis *et al*, 2012) and protein prediction programs (Gonzalez-Perez and Lopez-Bigas, 2011) were also used to assess identified coding variants.

Previously, Haller *et al* (2012) performed pooled sequencing of *CHRNA5* in a sample that also contributed 511 participants to targeted sequencing, identifying 4 *CHRNA5* coding variants beyond the well-studied rs16969968. Targeted sequencing found these 4 coding variants in the same 34 people as pooled sequencing, demonstrating high concordance. Furthermore, targeted sequencing also identified 6 additional singleton variants among the 511 people included in both analyses. The high quality of the targeted sequencing data was further verified using the HumanExome-12v1-1 array. All 2 820 individuals included in our primary analysis were genotyped using this array, and the concordance for the common and low frequency coding variants was 99.9%.

Data Analysis

A total of 1 432 European and 1 388 African Americans with targeted sequencing of *CHRNA5* and available smoking behaviors were examined in this analysis. Data were analyzed using the Statistical Analysis System (SAS 9.3, Cary, NC, USA). Logistic regression was used to model case-control status. European and African Americans were analyzed separately. Ancestry groups were verified using EIGENSTRAT (Price *et al*, 2006) and previously collected genome-wide arrays. HapMap populations were included as reference groups and linkage disequilibrium

filters were applied. Ten ancestry-specific principal components (PCs) were also developed. Examination of eigenvalues led us to include the first PC in our statistical analyses of both ancestry groups. All models included the standard covariates of sex, age, and first ancestry-specific PC.

We analyzed all variants identified by ANNOVAR as functional. These coding variants were divided into three classes based on the derived MAF in the entire sample: rare ($MAF<0.005$), low frequency ($0.05>MAF\geq0.005$), and common ($MAF\geq0.05$). Visual examination of the distribution of the allele frequencies in the sample (**Figure S1**) highlights a natural grouping of these three frequency classes in this dataset.

In the primary analytic model, low frequency and rare variants were collapsed into an aggregate low frequency variant term and aggregate rare variant term, respectively. Specifically, individuals with at least one copy of the minor allele for any of the nonsynonymous and frameshift variants were coded as 1 in each variant class (rare or low frequency) and individuals without any minor allele copies in this class were coded as 0. This collapsing method was based on a burden test (Li and Leal, 2008) to increase power to detect the cumulative effect of these variant classes.

Main effects of the one common rs16969968 coding variant, aggregate low frequency variants, and aggregate rare variants were analyzed together in a multivariate model of case-control status (*multivariate model set 1*). This approach was used to examine the effect of low frequency and rare variants conditioned on the effect of the well-established common rs16969968 variant. The primary logistic regression model was $\text{logit}(p)=\beta_0+\beta_{G1}\text{rs16969968}+\beta_{G2}\text{LowFrequencyTerm}+\beta_{G3}\text{RareTerm}+\beta_C C$ where C is the vector of standard covariates.

In secondary analyses, we examined the three low frequency variants (rs2229961,

rs80087508, rs79109919) as individual terms together with the common rs16969968 variant and aggregate rare variants (*multivariate model set 2*). The secondary logistic regression model was $\text{logit}(p)=\beta_0+\beta_{G1}\text{rs16969968}+\beta_{G2}\text{rs2229961}+\beta_{G3}\text{rs80087508}+\beta_{G4}\text{rs79109919}+\beta_{G5}\text{ RareTerm}+\beta_C C$. Because very few people were homozygous for the minor allele of the low frequency variants (0-5 individuals per variant), the heterozygous and homozygous individuals for each minor allele were collapsed into a single group and compared to the homozygous individuals of the major allele in these secondary analyses.

Explaining phenotypic variation

To examine the variation in nicotine dependence explained by *CHRNA5* coding variants, we used Nagelkerke's adjusted R^2 from logistic regression of case-control status (Nagelkerke, 1991). The variance in phenotype attributed to selected variants was derived as the R^2 attributable to the full model minus the R^2 attributable to the base model alone, including only age, sex, and first ancestry-specific PC as predictors of outcome. European and African American samples were analyzed separately. Each SNP or aggregate term was first examined individually. We then examined the final *multivariate model sets 1 and 2*.

Replication samples

The common and low frequency *CHRNA5* variants were assessed in 12 independent replication datasets with smoking phenotypes and exome chip genotypes. Cigarettes smoked per day (CPD), a proxy for nicotine dependence, was used as the outcome because FTND scores were not available. Our replication analyses compared light smokers ($CPD\leq10$) to heavy smokers ($CPD>20$) aged 25-80 years old. Previous work has demonstrated that these thresholds

of CPD agree with nicotine dependence defined by FTND (Berrettini *et al*, 2008; Stevens *et al*, 2008). European and African Americans were examined in separate logistic regression models that were similar to the primary sample models without the rare variant term. Specifically, the primary replication model was $\text{logit}(p)=\beta_0+\beta_{G1}\text{rs16969968}+\beta_{G2}\text{LowFrequencyTerm}+\beta_C C$, and the secondary replication model was $\text{logit}(p)=\beta_0+\beta_{G1}\text{rs16969968}+\beta_{G2}\text{rs2229961}+\beta_{G3}\text{rs80087508}+\beta_{G3}\text{rs79109919}+\beta_C C$, where C is the vector of standard covariates sex, age, ancestry-specific PCs, and field center (if applicable). For each ancestry group, all replication studies were required to have at least 50 light and 50 heavy smokers to be included in analyses of that group.

Meta-Analysis

A series of meta-analyses involving the 12 replication datasets were performed using PLINK (Purcell *et al*, 2007). First, beta values for the genetic factors obtained from *multivariable model sets 1 and 2* stratified by ancestry were meta-analyzed using weighting by standard errors. Each of these meta-analyses were then repeated with the addition of the beta from the primary dataset. Although examination of the Q statistic suggested no heterogeneity across studies for any of the genetic factors in any of the meta-analyses ($p>0.1$), except rs16969968 in European Americans ($p=0.02$)(*Multivariable model sets 1* and *2* with and without primary sample), to be consistent and conservative, all reported meta-analysis results are from random effects models. Some of the individual low frequency variants in *multivariable model set 2* were found in a limited number of individuals in certain studies. An ancestry specific sample was excluded from the meta-analysis of one of the low-frequency variants, if the minor allele of that variant occurred less than 5 times in the sample.

5.4 RESULTS

Variants identified in Sequencing

Sequencing initially identified 30 coding variants in *CHRNA5*, including 4 frameshift deletions and 26 non-synonymous variants. Four variants failed the stringent quality control metrics. Specifically, three variants were removed based on the review performed by the Quality Assurance/Quality Control analysis team: p.Ala10fs had low sequencing depth, p.434_435del was located at the wrong position, and p.Ile80Thr had a mean reference allele fraction significantly deviated from 0.5. Finally, p.Ser6Leu was excluded because sequence information was missing for 8% of the sample.

The remaining 26 *CHRNA5* variants (24 nonsynonymous and 2 frameshift) included in this analysis were rated as high quality in the manual review and were available in the entire sample (details of these variants are listed in **Supplemental Table 5.1 and Figure 5.1**). The majority are projected to be deleterious through protein prediction programs (Gonzalez-Perez and Lopez-Bigas, 2011).

Common *CHRNA5* variant

The only common coding variant identified was the previously well-studied rs16969968 located in the cytoplasmic domain of *CHRNA5.* In the primary sample, the rs1696968 minor allele was associated with increased risk for nicotine dependence in European (OR=1.3, p=0.003) and African Americans (OR=1.5, p=0.04) (Multivariate model set 1, **Table 5.2**). Replication results from 12 independent studies support the hypothesis that the A allele of rs16969968 increases risk for heaviness of smoking (**Figure 5.2, Supplemental Tables 5.2-5.3**).

Meta-analyses combining results from the primary and replication datasets demonstrate the robust role of rs16969968 (European: OR=1.3, p=3.7x10^{-11}; African: OR=1.3, p=0.01).

Aggregate low frequency *CHRNA5* variants

Three low frequency non-synonymous *CHRNA5* variants were identified (**Figure 1**). To increase power to detect associations, our primary analyses used an aggregate low frequency variant term, comparing individuals with at least one minor allele of a low frequency variant to those without any (6 individuals in the primary sample had 2 copies of a low frequency variant and 1 individual had 2 different low frequency variants). In the primary *Multivariate Model set 1*, this aggregate low frequency term provided trending evidence for association in both populations (European: OR=1.8, p=0.06; African: OR=1.4, p=0.07) (**Table 5.2**). Results from the replication studies demonstrated a significant combined effect of the three low frequency variants on heaviness of smoking (European: OR=1.2, p=0.02; African: OR=1.4, p=0.004) (**Figure 5.1, Supplemental Tables 5.2-5.3**). The overall meta-analysis from the primary and replication samples further illustrated the robust risk effect of the aggregate low frequency variants in both European (OR=1.3, p=0.005) and African Americans (OR=1.4, p=0.0006).

Individual low frequency *CHRNA5* variants

In secondary analyses using *multivariate model set 2,* we examined the independent contributions of the three low frequency variants to nicotine dependence risk controlling for the effect of other *CHRNA5* coding variants. One of these low frequency variants was found primarily in European Americans, and the other two were found almost exclusively in African Americans (**Table 5.2**).

The first low frequency variant rs2229961 causes a valine to isoleucine change at position 134 in the extracellular domain. The minor allele principally occurred in European Americans (MAF=0.02) and was rare in African Americans (MAF=0.002) (**Table 5.2**). In the primary sample, all 51 individuals with a copy of rs2229961 also possessed at least one copy of the well-established rs16969968 minor allele, suggesting that these two coding variants are transmitted together. Controlling for the effect of rs16969968, the minor allele of rs2229961 was in the risk direction in European (OR=1.7, p=0.1) and African Americans (OR=2.6, p=0.4). Meta-analysis of these primary results and the independent replication samples provided strong evidence that this variant contributed a risk effect in European Americans (OR=1.3, p=0.007) where it predominantly occurs (**Supplemental Table 5.4**).

The minor allele of the second low frequency variant rs80087508 causes a lysine to arginine transition at position 167 in the extracellular domain. This variant occurred exclusively in African Americans in both the primary sequencing sample (MAF=0.01, **Table 5.2**) and in the replication studies (**Supplemental Table 5.5**). This variant co-occurred with the common rs16969968 minor allele in 5 out of 38 individuals in the primary sample. In *multivariable model set 2* controlling for other coding variants*,* the minor allele of rs80087508 trended in the risk direction in African Americans (OR=2.1, p=0.06) (**Table 2**). Meta-analysis of these primary results and the 12 independent replication samples provided evidence that this variant contributed an independent risk effect in African Americans (OR=1.6, p=0.02) where it exclusively occurred (**Supplemental Table 5.5**).

The final low frequency variant, rs79109919, causes a leucine to glutamine change at amino acid position 363, which is located in the cytoplasmic domain (**Figure 5.1**). The minor allele of rs79109919 was common in African Americans (MAF=0.06) and occurred in only one

European American individual (MAF=0.0003) in the primary sample. Of the 158 individuals who possessed at least one copy of the rs79109919 minor allele, 7 also possessed a copy of the common rs16969968 risk allele and 1 possessed a copy of the low frequency rs80087508 variant, suggesting the independent transmission of these variants. In the primary sample, the minor allele of rs79109919 was in the risk direction in African Americans (OR=1.3, p=0.15) (**Table 5.2**). Meta-analysis of the primary and replication results provided strong evidence that this variant contributed an independent risk effect in African Americans (OR=1.4, p=0.03) where it primarily occurred (**Supplemental Table 5.5**).

Aggregate rare *CHRNA5* variants

Sequencing identified 22 rare coding variants (MAF<0.5%), including 20 nonsynonymous variants and 2 frameshift deletions. These variants occurred throughout the protein sequence (**Figure 5.1**). Each variant occurred in 1-4 individuals in the primary sample (**Supplemental Figure 5.1**). Furthermore, 9 of the 22 rare variants were seen in a single individual and were previously unreported in large reference datasets (Abecasis *et al*, 2012) (Exome Variant Server) (**Supplemental Table 5.1**).

Because these variants occurred in only a limited number of individuals, we used a collapsing burden test to assess their cumulative effect. Overall, 37 individuals possessed at least 1 rare variant, including 34 individuals with only one rare variant and 3 individuals (2 cases and 1 control) with 2 rare variants. In the primary sample, the aggregate rare variant term was associated with a risk effect in the European Americans (OR=12.9, p=0.01) as 12/13 (92%) individuals with at least one rare variant were cases (**Table 5.2**). In African Americans, the rare

variant term was in the risk direction but not significant (OR=1.5, p=0.37) as 17/24 (71%) of the individuals with at least one rare variant were cases.

Phenotypic variation accounted for by testing genetic factors

Nagelkerke's adjusted R^2 was used to assess the proportion of nicotine dependence variation explained by individual SNPs and multivariable models in the primary sample (**Table 5.3**). The well-studied rs16969968 gave the single strongest R^2 of 1.0% in European Americans and a lower R^2 of 0.4% in African Americans, where the variant is less common. In African Americans, the two low frequency variants, rs80087508 and rs79109919, which each occurred independently of rs16969968, gave R^2 estimates of 0.3% and 0.1%, respectively. In European Americans, the low frequency variant rs2229961 gave an R^2 of 0.4%. Finally, the aggregate rare variant term had a high R^2 of 1.0% in European Americans and a lower R^2 of 0.1% in African Americans.

In multivariate models with common, low frequency, and rare *CHRNA5* coding variants, the overall phenotypic variance explained by the genetic variants was 2.4% in European Americans and 1.0% in African Americans (*multivariate model set 2*, **Table 5.3**), supporting our conclusion that substantial variation in liability to nicotine dependence is attributable to *CHRNA5* coding variation.

5.5 DISCUSSION

Mounting evidence demonstrates that the functional rs16969968 variant in *CHRNA5* strongly contributes to differences in the risk of developing nicotine dependence across diverse populations (Chen *et al*, 2012b; Saccone *et al*, 2009). Here we show that multiple independent *CHRNA5* coding variants increase risk of nicotine dependence in European and African Americans by examining targeted next-generation sequencing data of approximately 3 000 nicotine dependent cases and non-dependent controls, with replication of low frequency variants using exome chip data in over 20 000 smokers.

Targeted sequencing of *CHRNA5* provided evidence that common, low frequency, and rare coding variants are independently associated with an increased risk of nicotine dependence, extending previous sequencing studies of other smoking-related measures. Wessel *et al* (2010) sequenced exons of *CHRNA5* and other nicotinic receptors in 448 European American participants enrolled in a smoking cessation trial. Using a weighted allele sharing test, this study provided initial evidence that both common and rare variants contribute to level of FTND score (pseudo-F=3.92, p=0.046). Haller *et al* (2012) used pooled sequencing to examine 5 nicotinic receptor subunit genes (including *CHRNA5*) in 400 European and 352 African Americans from a sample, which also contributed 511 participants to our study. Pooled sequencing identified 5 *CHRNA5* coding variants. The well-studied rs16969968 was the only variant that exhibited a significant effect on nicotine dependence ($p<0.05$), yet the minor allele of all but one of the coding variants trended in the risk direction. Doyle *et al* (2014) sequenced 250 African American heavy smokers and identified a few coding variants, including a novel frameshift deletion, which the authors hypothesized leads to nonsense mediated decay. Our results build on these findings

by sequencing a large, diverse, unrelated sample (n=2 820), which enabled us to identify many coding variants of high quality (n=26), including 9 novel variants.

Using exome chip data from 12 studies containing over 10 000 heavy and 10 000 light smokers, we confirmed that common and low frequency *CHRNA5* coding variants identified through targeted sequencing had a robust association with smoking behaviors. Previously, Vrieze *et al* (2014) used the HumanExome BeadChip array to assess the effect of nonsynonymous variants on addiction and behavioral disinhibition in a European ancestry sample. Examining 3 412 individuals from 1 694 families exposed to nicotine, this study identified 8 nonsynonymous *CHRNA5* variants. In single variant tests, rs16969968 (MAF=0.34) and rs2229961 (MAF=0.01) had a trending risk effect with a bonferroni corrected threshold (p=0.015 and p=0.046, respectively) on a composite nicotine dependence phenotype derived from factor analysis of frequency (days per month), quantity (cigarettes per day), and symptoms of dependence. The 6 other nonsynonymous variants were rare (each occurred in 1-5 individuals), and the majority were in the risk direction in single variant tests. When considered together using burden tests, these rare variants did not reach the adjusted significance threshold (SKAT p-value=0.049). Our replication study design of examining phenotypic extremes of smoking quantity (heavy vs light smokers) using a large diverse sample facilitated the detection of strong associations between common and low frequency nonsynonymous *CHRNA5* variants and heaviness of smoking.

An important strength of our study was the large sample of African Americans (n=1 388), a population often under-represented in genetic studies. Differences in the genetic architecture of European and African ancestry groups indicate that distinct genetic factors contribute to nicotine dependence in these populations. These differences are highlighted by the fact that the well-established rs16969968 variant is substantially more common in European (MAF=0.35)

than African Americans (MAF=0.06). We provide new evidence that two low frequency variants primarily occurring in African Americans (rs80087508 and rs79109919) independently contribute to nicotine dependence risk.

Since multiple independent *CHRNA5* signals were identified, a critical question is what proportion of phenotypic variance is explained by coding variation in this gene. Previous genetic studies of complex traits have identified reproducible associations, but these findings often explain only a modest proportion of phenotypic variance (Maher, 2008). For nicotine dependence, rs16969968 is arguably the single strongest genetic risk factor in European ancestry populations, and our results demonstrate that this variant only accounts for 1.0% of variance in European Americans (**Table 3**). However, the addition of low frequency and rare coding variants increased the estimated explained phenotypic variance in European Americans (R^2=2.4%). In African Americans, rs16969968 is less common and therefore explains a smaller proportion of estimated phenotypic variance (R^2=0.4%), and adding low frequency and rare coding variants increased this estimate (R^2=1.0%). These results highlight that low frequency and rare coding variants, beyond the genome-wide significant common variant, increased the estimated variance in nicotine dependence attributable to *CHRNA5*. An important next step is to examine variation explained by these coding variants on biomarkers, which more closely resemble tobacco exposure (Bloom *et al*, 2014; Munafo *et al*, 2012). Bloom *et al* (2014) found that rs16969968 explained four times more of the variance in carbon monoxide levels compared to self-reported cigarette consumption. Low frequency and rare *CHRNA5* coding variants will likely add to the phenotypic variance of biomarkers and long-term disease.

The findings reported here have limitations. This analysis focused on coding variants in *CHRNA5* because the common nonsynonymous rs16969968 variant is associated with changes

in receptor function (Bierut *et al*, 2008; Kuryatov *et al*, 2011) and we hypothesized that other coding variants may have a similar effect. Several noncoding variants tagged by rs588765 have been previously associated with changes in *CHRNA5* mRNA expression levels in the brain (Wang *et al*, 2009a; Wang *et al*, 2009b; Wang *et al*, 2013). Our analysis does not address the contribution of noncoding *CHRNA5* variation to nicotine dependence, which could influence our observed findings. Another limitation is that our analysis is restricted to a single gene. Previous studies (Saccone *et al*, 2010; Wessel *et al*, 2010) suggest that variation in other nicotinic receptors contribute to nicotine dependence. Specifically, recent evidence supports that rare variants in *CHRNA3* and *CHRNA4* are protective for nicotine dependence (Haller *et al*, 2012; Slimak *et al*, 2014; Xie *et al*, 2011). However, *CHRNA5* is clearly associated with nicotine dependence, making it a high priority first gene for study.

In summary, this study provides evidence that common, low frequency, and rare coding variants in *CHRNA5* independently increase risk for nicotine dependence in both European and African Americans as well as explain a substantial proportion of variance of this disease. Importantly, we identify associations with nicotine dependence for three low frequency non-synonymous variants, two of which almost exclusively occur in African Americans. From a public health perspective, these newly identified *CHRNA5* variants may have important prognostic and therapeutic implications on an individual level. Beyond nicotine dependence, previous studies show that rs16969968 is the strong genetic risk factor for lung cancer and chronic obstructive pulmonary disease (Amos *et al*, 2008; Hung *et al*, 2008; Pillai *et al*, 2009; Thorgeirsson *et al*, 2008) as well as influences response to smoking cessation therapies (Chen *et al*, 2012a). An important next step is to test whether these low frequency and rare *CHRNA5* coding variants similarly increase the risk of smoking-related diseases and response to smoking

cessation. Expanding our knowledge of which genetic variants influence risk for long-term diseases and response to treatments will inform personalized medical care for smokers.

5.6 ACKNOWLEDGEMENTS

I would like to acknowledge the co-authors of this study: Saccone NL, Johnson EO, Chen LS, Culverhouse R, Doheny K, Foltz SM, Fox L, Gogarten SM, Hartz S, Hetrick K, Laurie CC, Marosy B, Amin N, Arnett D, Bartz TM, Bertelsen S, Borecki IB, Brown MR, Chasman DI, van Duijn CM, Feitosa MF, Fox ER, Franceschini N, Franco OH, Grove ML, Guo X, Hofman A, Kardia SLR, Morrison AC, Musani SK, Psaty BM, Rao DC, Reiner AP, Rice K, Ridker PM, Rose LM, Rotter JI, Schick UM, Schwander K, Uitterlinden AG, Vojinovic D, Wang JC, Ware EB, Wilson G, Yao J, Zhao W, Breslau N, Hatsukami D, Stitzel J, Rice J, Goate A and Bierut LJ.

Please see supplement for acknowledgements listed by study.

Funding and Disclosure:

NIH grants and contracts supported the following studies and groups**: COGEND** (P01CA89392), **AAND**(R01DA025888), **CIDR** (HHSN268201100011I), **ARIC** (HHSN268201100005C, HHSN268201100006C, HHSN268201100007C, HHSN268201100008C, HHSN268201100009C, HHSN268201100010C, HHSN268201100011C, HHSN268201100012C, 5RC2HL102419, R01 HL118305), **CHS** (HHSN268201200036C, HHSN268200800007C, N01HC55222, N01HC85079, N01HC85080, N01HC85081, N01HC85082, N01HC85083, N01HC85086, U01HL080295, R01HL087652, R01HL105756, R01HL103612, R01HL120393, R01HL085251, R01HL068986, R01AG023629, UL1TR000124, DK063491), **COGA** (U10AA008401), **FamHS** (R01HL118305, R01-DK-089256), **GENOA** (HL054464, HL054457, HL054481, HL071917, NS041558, HL87660,

HL119443, HL118305), **HyperGEN** (HL54471, HL54472, HL54473, HL54495, HL54496, HL54497, HL54509, HL54515, 2 R01 HL55673-12, R01 HL055673, R01 HL118305, U01 HL54473, R01 HL055673, R01 HL118305), **MESA** (N01-HC-95159, N01-HC-95160, N01-HC-95161, N01-HC-95162, N01-HC-95163, N01-HC-95164, N01-HC-95165, N01-HC-95166, N01-HC-95167, N01-HC-95168, N01-HC-95169, UL1-TR-000040, UL1-RR-025005, R01-HL-071051, R01-HL-071205, R01-HL-071250, R01-HL-071251, R01-HL-071252, R01-HL-071258, R01-HL-071259, UL1-RR-025005, N02-HL-6-4278, UL1TR000124, DK063491), **WGHS** (HL043851, HL080467, CA047988), **WHI** (HHSN268201100046C, HHSN268201100001C, HHSN268201100002C, HHSN268201100003C, HHSN268201100004C, HHSN271201100004C, R21HL123677-01, 1R01HL118305-01A1).

ERF was supported by the following grants: European Commission FP6 STRP grant number 018947 (LSHG-CT-2006-01947); European Community's Seventh Framework Program (FP7/2007–2013, HEALTH-F4-2007-201413); Netherlands Organization for Scientific Research and the Russian Foundation for Basic Research (NWO-RFBR 047.017.043); ZonMw grant (project 91111025).

EO was supported by T32GM07200, UL1TR000448, TL1TR000449, and F30AA023685.
LJB was supported by U19CA148172.
LJB, AG, and JCW as well as the spouse of NLS are listed as inventors on Issued U.S. Patent 8,080,371,“Markers for Addiction” covering the use of certain SNPs in determining the diagnosis, prognosis, and treatment of addiction.
JAS has received support from Pfizer, Inc.

NA is supported by the Hersenstichting Nederland (project number F2013(1)-28).

OHF Franco works in ErasmusAGE, a center for aging research across the life course funded by Nestlé Nutrition (Nestec Ltd.); Metagenics Inc.; and AXA. Nestlé Nutrition (Nestec Ltd.); Metagenics Inc.; and AXA had no role in design and conduct of the study; collection, management, analysis, and interpretation of the data; and preparation, review or approval of the manuscript.

BMP serves on the DSMB of a clinical trial funded by the device manufacturer (Zoll LifeCor) and on the Steering Committee of the Yale Open Data Access Project funded by Johnson & Johnson.

The other authors declare no conflicts of interest.

5.7 REFERENCES

Exome Variant Server, NHLBI GO Exome Sequencing Project (ESP). Seattle, WA.

Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, Handsaker RE*, et al* (2012). An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**(7422): 56-65.

Amos CI, Wu X, Broderick P, Gorlov IP, Gu J, Eisen T*, et al* (2008). Genome-wide association scan of tag SNPs identifies a susceptibility locus for lung cancer at 15q25.1. *Nature genetics* **40**(5): 616-622.

Berrettini W, Yuan X, Tozzi F, Song K, Francks C, Chilcoat H*, et al* (2008). Alpha-5/alpha-3 nicotinic receptor subunit alleles increase risk for heavy smoking. *Molecular psychiatry* **13**(4): 368-373.

Bierut LJ, Madden PA, Breslau N, Johnson EO, Hatsukami D, Pomerleau OF*, et al* (2007). Novel genes identified in a high-density genome wide association study for nicotine dependence. *Human molecular genetics* **16**(1): 24-35.

Bierut LJ, Stitzel JA, Wang JC, Hinrichs AL, Grucza RA, Xuei X*, et al* (2008). Variants in nicotinic receptors and risk for nicotine dependence. *The American journal of psychiatry* **165**(9): 1163-1171.

Bloom AJ, Hartz SM, Baker TB, Chen LS, Piper ME, Fox L*, et al* (2014). Beyond cigarettes per day. A genome-wide association study of the biomarker carbon monoxide. *Annals of the American Thoracic Society* **11**(7): 1003-1010.

Caporaso N, Gu F, Chatterjee N, Sheng-Chih J, Yu K, Yeager M*, et al* (2009). Genome-wide and candidate gene association study of cigarette smoking behaviors. *PloS one* **4**(2): e4653.

Chen LS, Baker TB, Piper ME, Breslau N, Cannon DS, Doheny KF*, et al* (2012a). Interplay of genetic risk factors (CHRNA5-CHRNA3-CHRNB4) and cessation treatments in smoking cessation success. *The American journal of psychiatry* **169**(7): 735-742.

Chen LS, Saccone NL, Culverhouse RC, Bracci PM, Chen CH, Dueker N*, et al* (2012b). Smoking and genetic risk variation across populations of European, Asian, and African American ancestry--a meta-analysis of chromosome 15q25. *Genetic epidemiology* **36**(4): 340-351.

Chen X, Chen J, Williamson VS, An SS, Hettema JM, Aggen SH*, et al* (2009). Variants in nicotinic acetylcholine receptors alpha5 and alpha3 increase risks to nicotine dependence. *American journal of medical genetics Part B, Neuropsychiatric genetics : the official publication of the International Society of Psychiatric Genetics* **150B**(7): 926-933.

Dani JA, De Biasi M (2001). Cellular mechanisms of nicotine addiction. *Pharmacology, biochemistry, and behavior* **70**(4): 439-446.

Doyle GA, Chou AD, Saung WT, Lai AT, Lohoff FW, Berrettini WH (2014). Identification of CHRNA5 rare variants in African-American heavy smokers. *Psychiatric genetics* **24**(3): 102-109.

Fowler CD, Lu Q, Johnson PM, Marks MJ, Kenny PJ (2011). Habenular alpha5 nicotinic receptor subunit signalling controls nicotine intake. *Nature* **471**(7340): 597-601.

Gonzalez-Perez A, Lopez-Bigas N (2011). Improving the assessment of the outcome of nonsynonymous SNVs with a consensus deleteriousness score, Condel. *American journal of human genetics* **88**(4): 440-449.

Haller G, Druley T, Vallania FL, Mitra RD, Li P, Akk G*, et al* (2012). Rare missense variants in CHRNB4 are associated with reduced risk of nicotine dependence. *Human molecular genetics* **21**(3): 647-655.

Hsu YW, Tempest L, Quina LA, Wei AD, Zeng H, Turner EE (2013). Medial habenula output circuit mediated by alpha5 nicotinic receptor-expressing GABAergic neurons in the interpeduncular nucleus. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **33**(46): 18022-18035.

Hung RJ, McKay JD, Gaborieau V, Boffetta P, Hashibe M, Zaridze D*, et al* (2008). A susceptibility locus for lung cancer maps to nicotinic acetylcholine receptor subunit genes on 15q25. *Nature* **452**(7187): 633-637.

Kuryatov A, Berrettini W, Lindstrom J (2011). Acetylcholine receptor (AChR) alpha5 subunit variant associated with risk for nicotine dependence and lung cancer reduces (alpha4beta2)(2)alpha5 AChR function. *Molecular pharmacology* **79**(1): 119-125.

Li B, Leal SM (2008). Methods for detecting associations with rare variants for common diseases: application to analysis of sequence data. *American journal of human genetics* **83**(3): 311-321.

Liu JZ, Tozzi F, Waterworth DM, Pillai SG, Muglia P, Middleton L*, et al* (2010). Meta-analysis and imputation refines the association of 15q25 with smoking quantity. *Nature genetics* **42**(5): 436-440.

Maher B (2008). Personal genomes: The case of the missing heritability. *Nature* **456**(7218): 18-21.

Marks MJ, Pauly JR, Gross SD, Deneris ES, Hermans-Borgmeyer I, Heinemann SF*, et al* (1992). Nicotine binding and nicotinic receptor subunit RNA after chronic nicotine treatment. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **12**(7): 2765-2784.

Munafo MR, Timofeeva MN, Morris RW, Prieto-Merino D, Sattar N, Brennan P*, et al* (2012). Association between genetic variants on chromosome 15q25 locus and objective measures of tobacco exposure. *Journal of the National Cancer Institute* **104**(10): 740-748.

Nagelkerke N (1991). A note on a general definition of teh coefficient of determination. *Biometrika* **78**(3): 691-692.

Pillai SG, Ge D, Zhu G, Kong X, Shianna KV, Need AC*, et al* (2009). A genome-wide association study in chronic obstructive pulmonary disease (COPD): identification of two major susceptibility loci. *PLoS genetics* **5**(3): e1000421.

Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nature genetics* **38**(8): 904-909.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D*, et al* (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *American journal of human genetics* **81**(3): 559-575.

Saccone NL, Schwantes-An TH, Wang JC, Grucza RA, Breslau N, Hatsukami D*, et al* (2010). Multiple cholinergic nicotinic receptor genes affect nicotine dependence risk in African and European Americans. *Genes, brain, and behavior* **9**(7): 741-750.

Saccone NL, Wang JC, Breslau N, Johnson EO, Hatsukami D, Saccone SF*, et al* (2009). The CHRNA5-CHRNA3-CHRNB4 nicotinic receptor subunit gene cluster affects risk for nicotine dependence in African-Americans and in European-Americans. *Cancer research* **69**(17): 6848-6856.

Saccone SF, Hinrichs AL, Saccone NL, Chase GA, Konvicka K, Madden PA*, et al* (2007). Cholinergic nicotinic receptor genes implicated in a nicotine dependence association study targeting 348 candidate genes with 3713 SNPs. *Human molecular genetics* **16**(1): 36-49.

Slimak MA, Ables JL, Frahm S, Antolin-Fontes B, Santos-Torres J, Moretti M*, et al* (2014). Habenular expression of rare missense variants of the beta4 nicotinic receptor subunit alters nicotine consumption. *Frontiers in human neuroscience* **8**: 12.

Stevens VL, Bierut LJ, Talbot JT, Wang JC, Sun J, Hinrichs AL*, et al* (2008). Nicotinic receptor gene variants influence susceptibility to heavy smoking. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology* **17**(12): 3517-3525.

TAG (2010). Genome-wide meta-analyses identify multiple loci associated with smoking behavior. *Nature genetics* **42**(5): 441-447.

Thorgeirsson TE, Geller F, Sulem P, Rafnar T, Wiste A, Magnusson KP*, et al* (2008). A variant associated with nicotine dependence, lung cancer and peripheral arterial disease. *Nature* **452**(7187): 638-642.

Thorgeirsson TE, Gudbjartsson DF, Surakka I, Vink JM, Amin N, Geller F*, et al* (2010). Sequence variants at CHRNB3-CHRNA6 and CYP2A6 affect smoking behavior. *Nature genetics* **42**(5): 448-453.

Vrieze SI, Feng S, Miller MB, Hicks BM, Pankratz N, Abecasis GR*, et al* (2014). Rare nonsynonymous exonic variants in addiction and behavioral disinhibition. *Biological psychiatry* **75**(10): 783-789.

Wang JC, Cruchaga C, Saccone NL, Bertelsen S, Liu P, Budde JP*, et al* (2009a). Risk for nicotine dependence and lung cancer is conferred by mRNA expression levels and amino acid change in CHRNA5. *Human molecular genetics* **18**(16): 3125-3135.

Wang JC, Grucza R, Cruchaga C, Hinrichs AL, Bertelsen S, Budde JP*, et al* (2009b). Genetic variation in the CHRNA5 gene affects mRNA levels and is associated with risk for alcohol dependence. *Molecular psychiatry* **14**(5): 501-510.

Wang JC, Spiegel N, Bertelsen S, Le N, McKenna N, Budde JP*, et al* (2013). Cis-regulatory variants affect CHRNA5 mRNA expression in populations of African and European ancestry. *PloS one* **8**(11): e80204.

Wang K, Li M, Hakonarson H (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic acids research* **38**(16): e164.

Wessel J, McDonald SM, Hinds DA, Stokowski RP, Javitz HS, Kennemer M*, et al* (2010). Resequencing of nicotinic acetylcholine receptor genes and association of common and rare variants with the Fagerstrom test for nicotine dependence.

Neuropsychopharmacology : official publication of the American College of Neuropsychopharmacology **35**(12): 2392-2402.

Xie P, Kranzler HR, Krauthammer M, Cosgrove KP, Oslin D, Anton RF*, et al* (2011). Rare nonsynonymous variants in alpha-4 nicotinic acetylcholine receptor gene protect against nicotine dependence. *Biological psychiatry* **70**(6): 528-536.

5.8 TABLES

Table 5.1. Characteristics of primary sample

	European American (n=1 432)		African American (n=1 388)	
	cases	**controls**	**cases**	**controls**
Sample, n	728	704	854	534
Age, mean (range)	37 (25-45)	36 (25-45)	36 (25-45)	36 (25-45)
Sex				
Female	386 (53%)	482 (68%)	514 (60%)	321 (60%)
Male	342 (47%)	222 (32%)	340 (40%)	213 (40%)
FTND[a] score, mean (range)	6.49 (4-10)	0.02 (0-1)	6.21 (4-10)	0.33 (0-1)
CPD[b] category, mean (range)	1.94 (0-3)	0.01 (0-1)	1.11 (0-3)	0.03 (0-1)

[a]FTND is the Fagerstrom test for Nicotine Dependence

[b]CPD is categorical cigarettes per day (1 is ≤10, 2 is 11-20, 3 is 21-30, 4 is >30);

Table 5.2. The effect of common, low frequency, and rare *CHRNA5* coding variants on nicotine dependence in primary sample

Variant Class	Variant	European Americans (n=1 432)			African Americans (n=1 388)		
		MAF*	OR (95% CI)	p-value	MAF	OR (95% CI)	p-value
Multivariable Model set 1							
Common	rs16969968	0.355	1.27 (1.08-1.49)	**0.003**	0.058	1.46 (1.02-2.07)	**0.04**
Low Frequency	Aggregate term**	0.016	1.81 (0.97-3.42)	**0.06**	0.071	1.35 (0.98-1.87)	**0.07**
Rare	Aggregate term	0.005	12.90 (1.66-100.54)	**0.01**	0.009	1.47 (0.60-3.59)	**0.40**
Multivariable Model set 2							
Common	rs16969968	0.355	1.28 (1.09-1.50)	**0.003**	0.058	1.42 (1.00-2.03)	**0.05**
Low Frequency	rs2229961	0.016	1.71 (0.91-3.23)	**0.10**	0.002	2.57 (0.28-23.91)	**0.40**
	rs80087508	0***	.	**.**	0.014	2.00 (0.94-4.27)	**0.07**
	rs79109919	0.0003 ****	.	**.**	0.057	1.22 (0.86-1.75)	**0.26**
Rare	Aggregate term	0.005	12.91 (1.66-100.66)	**0.01**	0.009	1.51 (0.62-3.68)	**0.37**

Multivariable model set 1 includes rs16969968, the aggregate low frequency variant term, and the aggregate rare variant term:

Multivariable model set 2 includes rs16969968, rs2229961, rs800087508, rs79109919, and the aggregate rare variant term;

All models adjusted for sex, age, and first ancestry-specific PC as covariates;

*MAF stands for minor allele frequency; **for aggregate terms, the MAF was estimated by the dividing the number of people with at least one low frequency/rare variant by 2 times the total number of people; *** rs80087508 is non-polymorphic in European Americans; **** Because the minor allele of rs79109919 occurred less than 5 times in European Americans, the OR and p-value are not presented.

Table 5.3. Variation in nicotine dependence risk explained by selected variants and multivariate models in primary sample

Variant Class	Variant	European Americans (n=1 432)			African Americans (n=1 388)		
		MAF	R^2	p-value	MAF	R^2	p-value
Common	rs16969968	0.355	1.0%	0.001	0.058	0.4%	0.04
Low Frequency	aggregate term	0.016	0.5%	0.02	0.071	0.3%	0.06
	rs2229961	0.016	0.4%	0.03	0.002	0.1%	0.24
	rs80087508	0	.	.	0.014	0.3%	0.07
	rs79109919	0.0003	0.2%	0.15	0.057	0.1%	0.34
Rare	aggregate term	0.005	1.0%	0.0009	0.009	0.1%	0.37
Multivariable model set 1			2.3%	1.2×10^{-5}		0.8%	0.04
Multivariable model set 2			2.4%	5.5×10^{-5}		1.0%	0.07

R^2 is the Nagelkerke's adjusted R^2 difference from logistic regression, comparing the base model with intercept, sex, age, and ancestry specific PCs to models with genetic variants;

Each individual variant and aggregate term was examined first by itself and then we examined the final multivariate model sets 1 and 2;

Multivariable model set 1 includes rs16969968, the aggregate low frequency variant term, and the aggregate rare variant term:

Multivariable model set 2 includes rs16969968, rs2229961, rs800087508, rs79109919, and the aggregate rare variant term;

p-values calculated by taking the difference between the -2logliklihoods in the base model and those with variants as a chi-square statistic.

5.9 FIGURES

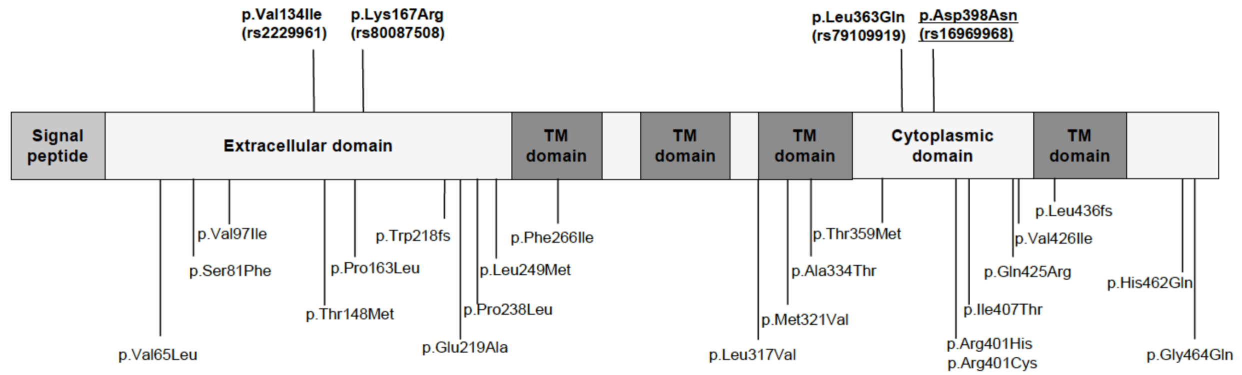


Figure 5.1. Protein Schematic of *CHRNA5* nonsynonymous and frameshift variants. **Bold underline** indicates the only common variant (MAF>5%); **Bold** indicates low frequency variants (5%>MAF≥0.5%); Other variants are rare (MAF<0.5%).

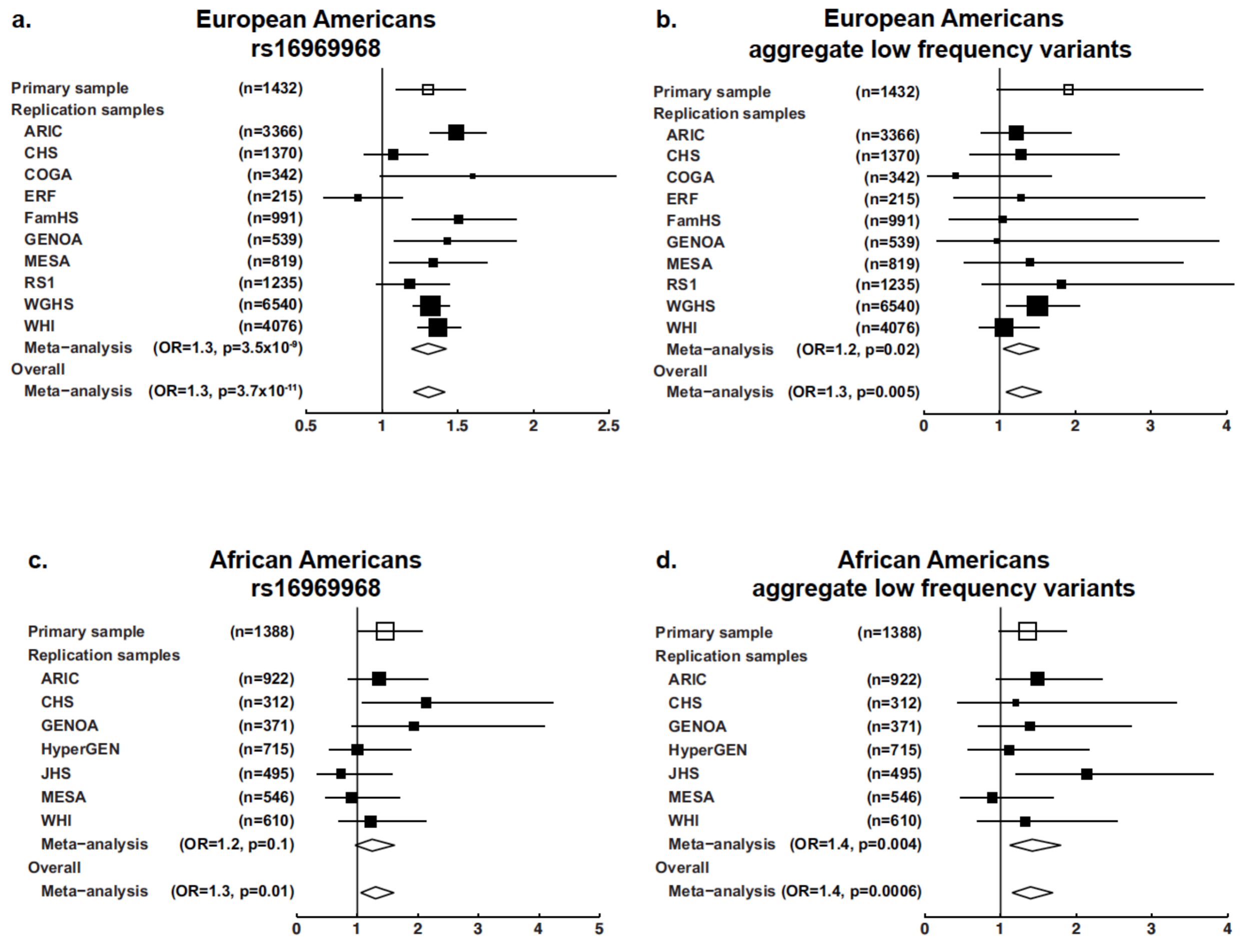


Figure 5.2. Forest plots showing the primary sample, replication samples, and random effects meta-analyses from *Multivariable model set 1*. (a) rs1696998 in European Americans; (b) aggregate low frequency variant term in European Americans; (c) rs1696998 in African Americans; (d) aggregate low frequency variant term in African Americans.

5.10 SUPPLEMENTAL METHODS

Targeted sequencing of *CHRNA5*

Custom baits were designed using Agilent's web-based design tool eArray (https://earray.chem.agilent.com/earray) to capture regions with 50 kilobases on either side of the gene. For library preparation, samples were placed on 96-well plates that were stratified by case/control status, recruitment study site, ethnicity, and sex. A range of 500ng to 1ug of genomic DNA was sheared using the Covaris E-210 instrument using modified parameters for shearing (DutyCycle=10%, Intensity=4, Cycles per Burst=200, time=80sec). Libraries were prepared according to the Agilent protocol (SureSelect^{XT} Target enrichment for Illumina Multiplexed Sequencing Protocol v1.1.1). Amplification of the libraries prior to and post capture were performed using the Kapa Biosystems HiFi HotStart Ready Mix. Samples were clustered for sequencing using the Illumina cBOT Cluster Generation system. One hundred base pairs paired end sequencing was performed on the Illumina HiSeq 2000 platform v3 chemistry. FastQ files were aligned with BWA (Li and Durbin, 2010) to the 1000 genomes phase 2 (GRCh37) human genome reference (Abecasis *et al*, 2012). GATK2.3-9 was used for base quality recalibration and local realignment. Unified Genotyper was used for multi-sample calling and VQSR for variant filtering. All samples had 96-SNP barcode genotyping for sample identity tracking and concordance checking.

Quality Control Measures

Data quality was systematically evaluated using a robust alignment and variant calling workflow implemented by CIDR (http://www.cidr.jhmi.edu/index.html). Over 100 quality

control metrics were evaluated in real time to quickly identify potential errors and implement fixes throughout the sequencing process. Briefly, samples underwent pretesting with 96 SNP barcode panels, and more than 50% of samples also had previously acquired GWAS array data to verify sample identity and quality. Sample sex was confirmed by examining normalized read depth on the X and Y chromosomes. 64 duplicate samples and 65 HapMap controls were also used to assess the quality of variant calls.

References:

Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, Handsaker RE*, et al* (2012). An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**(7422): 56-65.

Li H, Durbin R (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics (Oxford, England)* **26**(5): 589-595

5.11 SUPPLEMENTAL REPLICATION STUDY DESCRIPTIONS

Atherosclerosis Risk Communities Study (ARIC)

The ARIC study has been described in detail previously.[1] Men and women aged 45-64 years at baseline were recruited from four communities: Forsyth County, North Carolina; Jackson, Mississippi; Minneapolis, Minnesota; and Washington County, Maryland. A total of 15,792 individuals, predominantly White and African American, participated in the baseline examination in 1987-1989, with three additional triennial follow-up examinations and a fifth exam in 2011. Informed consent was obtained from all participants, and this project was approved by the Institutional Review Boards of all participating institutions. The current analysis included 3 366 European Americans (1 859 Heavy smokers and 1 507 Light smokers) and 922 African Americans (214 heavy smokers and 708 light smokers) with available Illumina Infinium Human Exome Array v1.0 data.[2]

Reference:

1. The ARIC Investigators. The Atherosclerosis Risk in Communities (ARIC) Study: design and objectives. *Am J Epidemiol* 129, 687-702 (1989).

2. Grove, M.L., Yu, B., Cochran, B.J., Haritunians, T., Bis, J.C., Taylor, K.D., Hansen, M., Borecki, I.B., Cupples, L.A., Fornage, M., et al. (2013). Best practices and joint calling of the HumanExome BeadChip: the CHARGE Consortium. PLoS ONE 8, e68095.

Cardiovascular Health Study (CHS)

This is a population-based cohort study of risk factors for the development and progression of cardiovascular disease in older adults sponsored by the by the NHLBI.[1] Between 1989 and 1990, this study recruited 5201 adults ages 65 and older from four U.S. communities, and recruited an additional predominately African-American cohort of 687 people in 1992-1993. Subjects received annual clinic follow-up and semi-annual phone calls. Informed consent was obtained from all participants, and this project was approved by the Institutional Review Boards of all participating institutions. The current analysis included 1 370 European Americans (513 Heavy smokers and 857 Light smokers) and 312 African Americans (50 heavy smokers and 262 light smokers) with available Illumina HumanExome BeadChip v1.0 data.

Reference:

1. Fried LP, Borhani NO, Enright P, Furberg CD, Gardin JM, Kronmal RA, Kuller LH, Manolio TA, Mittelmark MB, Newman A. The Cardiovascular Health Study: design and rationale. Ann Epidemiol. 1991; 1(3):263-76

Collaborative Study on the Genetics of Alcoholism (COGA)

COGA was initiated in 1989 as a large, family study designed to identify genes that contribute to alcohol use disorders and related behaviors funded by the National Institute on Alcohol Abuse and Alcoholism.[1,2] Alcoholic probands were recruited from treatment facilities from seven sites across the US. Family members of these probands were invited to participate and a set of comparison families was also drawn from the same communities. COGA has gathered detailed, standardized data on study participants, including diagnostic, genetic, and neurophysiological

assessments. COGA continues to enroll young participants from these families in a longitudinal study to examine the development of substance use disorders. Informed consent was obtained from all participants, and this project was approved by the Institutional Review Boards of all participating institutions. The current analysis included 340 European Americans (208 Heavy smokers and 132 Light smokers) with available Affymetrix Axiom Exome 319 array data.

References:

1. Begleiter H, Reich T, Hesselbrock V, Porjesz B, Li T, Schuckit M, Edenberg H, Rice J (1995) The Collaborative Study on the Genetics of Alcoholism. Alcohol Health Res World 19**:**228–236

2. Edenberg, H. J. (2002) The Collaborative Study on the Genetics of Alcoholism:
an update. Alcohol Res Health 26, 214-218.,

Erasmus Rucphen Family study (ERF)

This is a family-base cohort study, which is part of the Genetic Research in isolated Populations (GRIP) program.[1,2] The goal of this study is to identify genetic risk factors in the development of complex disorders within a genetically isolated population in the southwest of the Netherlands. In this study, 22 couples that had at least 6 children baptized in the community church from 1850-1900 were identified. All living descendants and their spouses were invited to participate. Study population includes ~3000 individuals. All data were collected between 2002 and 2005. Subjects received extensive clinical evaluations at a research center within the community. Informed consent was obtained from all participants, and this project was approved by the Institutional Review Boards of all participating institutions. The current analysis focused on 215

Europeans (77 Heavy smokers and 138 Light smokers) with available Illumina HumanExome chip v1.1 data.

References:

1. Aulchenko YS, Heutink P, Mackay I, Bertoli-Avella AM, Pullen J, Vaessen N, Rademaker TA, Sandkuijl LA, Cardon L, Oostra B, van Duijn CM. *Linkage disequilibrium in young genetically isolated Dutch population. Eur J Hum Genet 2004;12:527–34*

2. Henneman P, Aulchenko YS, Frants RR, van Dijk KW, Oostra BA, van Duijn CM. *Prevalence and heritability of the metabolic syndrome and its individual components in a Dutch isolate: the Erasmus Rucphen Family study* J Med Genet 2008*;45:572-577*

Family Heart Study (FamHS)

The FamHS (https://dsgweb.wustl.edu/fhscc/) is a multi-center, population-based study of genetic and nongenetic determinants of coronary heart disease (CHD), atherosclerosis, and cardiovascular risk factors. This study began in 1992 with the ascertainment of 1,200 families, half randomly sampled and half selected because of an excess of CHD or risk factor abnormalities as compared with age- and sex-specific population rates.[1] The families, with approximately 6,000 subjects, were sampled from four population-based parent studies. The participants attended a first clinic visit between the years 1994-1996 and a broad range of phenotypes was assessed in the general domains of CHD, atherosclerosis, cardiac and vascular function, inflammation and hemostasis, lipids and lipoproteins, blood pressure, diabetes and insulin resistance, pulmonary function, diet, habitual physical activity, anthropometry, medical history and medication use. Approximately 8 years later, 2,756 European American (EA)

subjects belonging to the 510 of the largest and most informative pedigrees were invited for a second clinical visit (2002-04). The most important CHD risk factors were measured again. Medical history and medication use was updated. Informed consent was obtained from all participants, and this project was approved by the Institutional Review Boards of all participating institutions. FamHS participants were genotyped on the HumanExome Bead-Chip v.1.0 (Illumina) and jointly called at the University of Texas Health Science Center at Houston. Informed consent was obtained from all participants, and this project was approved by the Institutional Review Boards of all participating institutions. The current analysis included 991 European Americans (550 Heavy smokers and 441 Light smokers) from the first clinical visit in which has a total of 3,868 subjects with exome data.

Reference:

1. Higgins M, Province M, Heiss G, Eckfeldt J, Ellison RC, Folsom AR, Rao DC, Sprafka JM, Williams R. NHLBI Family Heart Study: objectives and design. Am J Epidemiol. 1996 Jun 15; 143(12):1219-28.

Genetic Epidemiology Network of Arteriopathy (GENOA)

This is one of four research networks that form the NHLBI Family Blood Pressure Program.[1,2] The overall goal of GENOA is to elucidate the genetics of hypertension and its arteriosclerotic target-organ damage, including macrovascular and microvascular complications in heart, brain, kidneys, and peripheral arteries. From 1995 to 2000, two cohorts were ascertained through sibships in which at least 2 siblings had essential hypertension diagnosed prior to 60 years of age. All siblings were invited to participate. Approximately 80% of participants received a

follow-up exam between 2000 and 2005. Informed consent was obtained from all participants, and this project was approved by the Institutional Review Boards of all participating institutions. The current analysis included 539 European Americans (258 Heavy smokers and 281 Light smokers) and 371 African Americans (73 Heavy smokers and 298 Light smokers) with available Illumina Infinium HumanExome BeadChip v1.1 data.

References:

1.FBPP Investigators. Multi-center genetic study of hypertension: The Family Blood Pressure Program (FBPP). Hypertension. 2002 Jan; 39(1):3-9.

2. Daniels PR, Kardia SL, Hanis CL, Brown CA, Hutchinson R, Boerwinkle E, Turner ST, Genetic Epidemiology Network of Arteriopathy study. Familial aggregation of hypertension treatment and control in the Genetic Epidemiology Network of Arteriopathy (GENOA) study. Am J Med. 2004 May 15; 116(10):676-81.

Hypertension Genetics Epidemiology Network (HyperGEN)

This is one of four research networks that form the NHLBI Family Blood Pressure Program.[1,2] The goal of HyperGEN is to identify major genetic determinants of hypertension and to study possible interactions between genetic and non-genetic factors in defined populations. African American and non-Hispanic white hypertensive siblings along with available parents and untreated adult offspring were recruited from 5 field centers across the US. Preference in ascertainment and recruitment was given to sibships with a least one subject with severe hypertension. Informed consent was obtained from all participants, and this project was approved by the Institutional Review Boards of all participating institutions. The current analysis included

715 African Americans (97 Heavy smokers and 618 Light smokers) with available Affymetrix Genome-Wide Human SNP Array 6.0 and 5.0 data.

References:

1. FBPP Investigators. Multi-center genetic study of hypertension: The Family Blood Pressure Program (FBPP). Hypertension. 2002 Jan; 39(1):3-9.

2. Williams RR, Rao DC, Ellison RC, Arnett DK, Heiss G, Oberman A, Eckfeldt JH, Leppert MF, Province MA, Mockrin SC, Hunt SC. NHLBI family blood pressure program: methodology and recruitment in the HyperGEN network. Hypertension genetic epidemiology network. Ann Epidemiol. 2000 Aug;10(6):389-400.

Jackson Heart Study (JHS)

This is a large, community-based, observational study to understand factors that influence heart disease and other illnesses in African Americans funded by the NHLBI and the Office of Research on Minority Health at NIH.[1] JHS is an expansion of the ARIC study site in Jackson, Mississippi. Since its inception in 1998, 5 301 African American men and women have been enrolled in this study from urban and rural areas in the Jackson, Mississippi metropolitan area. The study focused on individuals ages 35-84, except in the family cohort where individuals 21 to 34 were also eligible. Three back-to-back cohort clinical exams were performed (2000-2004, 2005-2009, and 2009-2012), providing extensive longitudinal data. Informed consent was obtained from all participants, and this project was approved by the Institutional Review Boards of all participating institutions. The current analysis included 495 African Americans (102 Heavy smokers and 393 Light smokers) with available Illumina HumanExome Chip v1.0 data.

Reference:

1. Taylor HA Jr, Wilson JG, Jones DW, Sarpong DF, Srinivasan A, Garrison RJ, Nelson C, Wyatt SB. Toward resolution of cardiovascular health disparities in African Americans: design and methods of the Jackson Heart Study. Ethn Dis. 2005 Autumn;15(4 Suppl 6):S6-4-17.

Multi-Ethnic Study of Atherosclerosis (MESA)

This population-based cohort study was initiated in July 2000 to investigate the prevalence, correlates, and progression of subclinical cardiovascular disease.[1] A total of 6 814 asymptomatic individuals' ages 45-84 years olds were recruited from 6 field centers across the US. Each participant received an extensive clinical exam and blood samples were collected to test biochemical and genetic risk factors. Informed consent was obtained from all participants, and this project was approved by the Institutional Review Boards of all participating institutions. The current analysis included 819 European Americans (369 Heavy smokers and 450 Light smokers) and 546 African Americans (101 Heavy smokers and 445 Light smokers) with available Illumina HumanExome Chip v1.0
data.

Reference:

1. Bild DE, Bluemke DA, Burke GL, Detrano R, Diez Roux AV, Folsom AR, Greenland P, Jacob DR Jr, Kronmal R, Liu K, Nelson JC, O'Leary D, Saad MF, Shea S, Szklo M, Tracy RP. Multi-ethnic study of atherosclerosis: objectives and design. Am J Epidemiol. 2002 Nov 1; 156(9):871-81.

Rotterdam Study Cohort 1 (RS1)

The Rotterdam Study is a prospective, population-based cohort study to determine the occurrence of cardiovascular, neurological, ophthalmic, endocrine, hepatic, respiratory, and psychiatric diseases in elderly people.[1] Subjects were recruited from Ommoord, a suburb of Rotterdam, in three different cohorts. The initial cohort (RS1) began in 1990 with 7 983 individuals aged 55 and older with follow-up visits in 1994-1995, 1997-1999, 2002-2004, and 2009-2011. All participants were interviewed at home and received an extensive set of clinical examinations, including imaging and sample collection for molecular and genetic analyses. Informed consent was obtained from all participants, and this project was approved by the Institutional Review Boards of all participating institutions. The current analysis included 1 235 Europeans (535 Heavy smokers and 700 Light smokers) with available Illumina HumanExome beadchip v1.0 data.

Reference:

1. Hofman A, Darwish Murad S, van Duijn CM, Franco OH, Goedegebure A, Ikram MA, Klaver CC, Nijsten TE, Peeters RP, Stricker BH, Tiemeier HW, Uitterlinden AG, Vernooij MW. The Rotterdam Study: 2014 objectives and design update. Eur J Epidemiol. 2013 Nov;28(11):889-926. doi: 10.1007/s10654-013-9866-z. Epub 2013 Nov 21. PubMed PMID: 24258680.

Women's Genomic Health Study (WGHS)

The Women's Genome Health Study (WGHS) is a prospective cohort of initially healthy, female North American health care professionals at least 45 years old at baseline representing

participants in the Women's Health Study (WHS) who provided a blood sample at baseline and consent for blood-based analyses. [1] The WHS was a 2x2 trial beginning in 1992-1994 of vitamin E and low dose aspirin in prevention of cancer and cardiovascular disease with about 10 years of follow-up. Since the end of the trial, follow-up has continued in observational mode. Additional information related to health and lifestyle were collected by questionnaire throughout the WHS trial and continuing observational follow-up. Informed consent was obtained from all participants, and this project was approved by the Institutional Review Boards of all participating institutions. The current analysis included 6 540 European Americans (4 924 Heavy smokers and 1 616 Light smokers) that had the common rs16969968 variant genotyped on the Illumina HumanHap300 Duo"+" and the three low frequency *CHRNA5* variants genotyped on the Illumina exome v.1.1 chip.

Reference:

1. Ridker PM, Chasman DI, Zee RY, Parker A, Rose L, Cook NR, Buring JE; Women's Genome Health Study Working Group. Rationale, design, and methodology of the Women's Genome Health Study: a genome-wide association study of more than 25,000 initially healthy american women. Clin Chem. 2008 Feb;54(2):249-55. Epub 2007 Dec 10. PubMed PMID: 18070814.

Women's Health Initiative (WHI)

The WHI is a long-term national health study focused on strategies for preventing heart disease, breast and colorectal cancer, and osteoporotic fractures in postmenopausal women. Between 1993 and 1998, 161 808 postmenopausal women were enrolled from 40 clinical centers in either a clinical trial (68 132) or an observation study (93 676).[1-4] The clinical trials were designed to

test the effects of postmenopausal hormone therapy, diet modification, as well as calcium and vitamin D supplements on disease outcomes. The observation study examined the relationship between lifestyle, environmental, medical, and other risk factors on specific measures of health. Recruitment was done through mass mailing to age-eligible women obtained from voter registration, driver's license and Health Care Financing Administration or insurance lists, with target recruitment of a socio-demographically diverse population. WHI participants were also invited to participate in an extension study with follow-up through 2010. Informed consent was obtained from all participants, and this project was approved by the Institutional Review Boards of all participating institutions. The current analysis included 4 076 European Americans (2 065 Heavy smokers and 2 011 Light smokers) and 610 African Americans (117 Heavy smokers and 493 Light smokers) with available Illumina HumanExome BeadChip v1exome chip data.

References:

1. Design of the women's health initiative clinical trial and observational study. The women's health initiative study group. *Control Clin Trials*. 1998;19:61-109

2. Anderson GL, Manson J, Wallace R, Lund B, Hall D, Davis S, Shumaker S, Wang CY, Stein E, Prentice RL. Implementation of the women's health initiative study design. *Ann Epidemiol*. 2003;13:S5-17

3. Howard BV, Van Horn L, Hsia J, Manson JE, Stefanick ML, Wassertheil-Smoller S, Kuller LH, LaCroix AZ, Langer RD, Lasser NL, Lewis CE, Limacher MC, Margolis KL, Mysiw WJ, Ockene JK, Parker LM, Perri MG, Phillips L, Prentice RL, Robbins J, Rossouw JE, Sarto GE, Schatz IJ, Snetselaar LG, Stevens VJ, Tinker LF, Trevisan M, Vitolins MZ, Anderson GL, Assaf AR, Bassford T, Beresford SA, Black HR, Brunner RL, Brzyski RG, Caan B, Chlebowski RT,

Gass M, Granek I, Greenland P, Hays J, Heber D, Heiss G, Hendrix SL, Hubbell FA, Johnson KC, Kotchen JM. Low-fat dietary pattern and risk of cardiovascular disease: The women's health initiative randomized controlled dietary modification trial. *Jama*. 2006;295:655-666

4. Jackson RD, LaCroix AZ, Gass M, Wallace RB, Robbins J, Lewis CE, Bassford T, Beresford SA, Black HR, Blanchette P, Bonds DE, Brunner RL, Brzyski RG, Caan B, Cauley JA, Chlebowski RT, Cummings SR, Granek I, Hays J, Heiss G, Hendrix SL, Howard BV, Hsia J, Hubbell FA, Johnson KC, Judd H, Kotchen JM, Kuller LH, Langer RD, Lasser NL, Limacher MC, Ludlam S, Manson JE, Margolis KL, McGowan J, Ockene JK, O'Sullivan MJ, Phillips L, Prentice RL, Sarto GE, Stefanick ML, Van Horn L, Wactawski-Wende J, Whitlock E, Anderson GL, Assaf AR, Barad D. Calcium plus vitamin d supplementation and the risk of fractures. *The New England journal of medicine*. 2006;354:669-683

5.12 SUPPLEMENTAL TABLES

Supplemental Table 5.1. Characteristics of 26 high quality *CHRNA5* coding variants identified by targeted sequencing of 1 582 Nicotine dependent cases and 1 238 controls.

Position of variants	dbSNP identifier	Coding Change	Protein Change	Study MAF[a], allele count		frequency class in analysis	Exome Sequencing Project MAF, allele count		Phase 1 1000 Genomes MAF, allele count	Condel Score (Class)
Frameshift Deletions				**European Ancestry (n=1,432)**	**African Ancestry (n=1,388)**		**AA**	**EA**		
chr15:78882385..78882386		c.653delG	p.Trp218fs		0.0004, 1(A1)/2775(R)	Rare	0.000, 1(A1)/4263(R)	0.001, 5(A1)/8247(R)		
chr15:78885495..78885500		c.1308_1312del	p.436_438del		0.0004, 1(A1)/2775(R)	Rare				
Nonsynonymous SNVs										
chr15:78873239		c.193G>C	p.Val65Leu	0.0003, 1(C)/2863(G)		Rare				0.712 (deleterious)
chr15:78873288		c.242C>T	p.Ser81Phe		0.0004, 1(T)/2775(C)	Rare				0.861 (deleterious)
chr15:78879017	rs148722844	c.289G>A	p.Val97Ile		0.0007, 2(A)/2774(G)	Rare	0.000, 1(A)/4391(G)			0.386 (neutral)
chr15:78880752	rs2229961	c.400G>A	p. Val134Ile	0.016, 46(A)/2818(G)	0.002, 5(A)/2771(G)	**Low frequency**	0.003, 14(A)/4378(G)	0.019, 165(A)/8421(G)	0.006, 13(A)/2171(G)	0.905 (deleterious)
chr15:78882176	rs201563436	c.443C>T	p.Thr148Met		0.0004, 1(T)/2775(C)	Rare		0.000, 8585(C)/1(T)		0.966 (deleterious)
chr15:78882221	rs55863434	c.488C>T	p.Pro163Leu	0.0003, 1(T)/2863(C)		Rare		0.000, 8583(C)/3(T)		1.000 (deleterious)
chr15:78882233	rs80087508	c.500A>G	p. Lys167Arg		0.014, 39(G)/2737(A)	**Low frequency**	0.019, 4308(A)/84(G)	0.000, 8585(A)/1(G)	0.005, 2174(A)/10(G)	0.790 (deleterious)
chr15:78882389		c.656A>C	p.Glu219Ala	0.0003, 1(C)/2863(A)		Rare				0.449 (neutral)
chr15:78882446	rs61742337	c.713C>T	p.Pro238Leu	0.0007, 2(T)/2862(C)		Rare		0.000, 8584(C)/2(T)		0.883 (deleterious)

chr15:78882478	rs137878726	c.745C>A	p.Leu249Met		0.001, 4(A)/2772(C)	Rare	0.000, 2(A)/4390(C)			0.841 (deleterious)
chr15:78882529	rs138719535	c.796T>A	p.Phe266Ile		0.001, 3(A)/2773(T)	Rare	0.001, 3(A)/4389(T)			0.858 (deleterious)
chr15:78882682	rs116099178	c.949C>G	p. Leu317Val		0.001, 4(G)/2772(C)	Rare	0.001, 4389(C)/3(G)		0.001, 2182(C)/2(G)	0.411 (neutral)
chr15:78882694	rs74865777	c.961A>G	p.Met321Val	0.0007, 2(G)/2862(A)		Rare				0.643 (deleterious)
chr15:78882733		c.1000G>A	p.Ala334Thr		0.0004, 1(A)/2775(G)	Rare				0.473 (deleterious)
chr15:78882809	rs79721430	c.1076C>T	p. Thr359Met	0.0003, 1(T)/2863(C)		Rare		0.000, 8584(C)/2(T)	0.000, 2183(C)/1(T)	0.027 (neutral)
chr15:78882821	rs79109919	c.1088T>A	p.Leu363Gln	0.0003, 1(A)/2863(T)	0.058, 160(A)/2616(T)	**Low frequency**	0.052, 228(A)/4164(T)	0.001, 5(A)/8581(T)	0.021, 46(A)/2138(T)	0.871 (deleterious)
chr15:78882925	rs16969968	c.1192G>A	p.Asp398Asn	0.355, 1016(A)/1848(G)	0.058, 161(A)/2615(G)	**Common**	0.062, 274(A)/4118(G)	0.349, 2993(A)/5593(G)	0.175, 383(A)/1801(G)	0.018 (neutral)
chr15:78882934	rs76766434	c.1202C>T	p.Arg401Cys	0.0003, 1(T)/2863(C)	0.0007, 2(T)/2774(C)	Rare	0.002, 4385(C)/7(T)	0.000, 8585(C)/1(T)	0.000, 2183(C)/1(T)	0.458 (neutral)
chr15:78882935	rs141180754	c.1202G>A	p.Arg401His		0.001, 3(A)/2773(G)	Rare	0.001, 3(A)/4389(G)			0.025 (neutral)
chr15:78882953		c.1220C>T	p.Ile407Thr		0.0004, 1(T)/2775(C)	Rare				0.005 (neutral)
chr15:78885462	rs202052590	c.1274A>G	p.Gln425Arg	0.0003, 1(G)/2863(A)		Rare				0.910 (deleterious)
chr15:78885464	rs150329151	c.1276G>A	p.Val426Ile		0.0004, 1(A)/2775(G)	Rare	0.000, 1(A)/4391(G)			0.746 (deleterious)
chr15:78885574	rs76071148	c.1386T>A	p.His462Gln	0.0007, 2(A)/2862(T)	0.0007, 2(A)/2774(T)	Rare	0.000, 2(A)/4390(T)	0.000, 2(A)/8584(T)	0.076, 166(A)/2018(T)	0.340 (neutral)
chr15:78885579		c.1391A>G	p.Gly464Glu	0.0003, 1(G)/2863(A)		Rare				0.88- (deleterious)

[a]MAF is minor allele frequency.

Supplemental Table 5.2. Random effects meta-analysis results for multivariate model set 1 in European Americans

	Heavy Smokers	Light Smokers	rs16969968			Aggregate low frequency variants		
			MAF*	OR	p-value	MAF**	OR	p-value
Primary sample* **	728	704	0.36	1.27	0.003	0.02	1.81	0.06
Replication Studies								
ARIC	1859	1507	0.33	1.44	1.03E-09	0.02	1.20	0.41
CHS	513	857	0.34	1.07	0.48	0.02	1.25	0.50
COGA	209	133	0.33	1.53	0.06	0.02	0.49	0.24
ERF	138	77	0.36	0.86	0.25	0.02	1.26	0.38
FamHS	550	441	0.34	1.45	0.0006	0.01	1.03	0.95
GENOA	258	281	0.37	1.39	0.01	0.01	0.97	0.96
MESA	369	450	0.33	1.30	0.02	0.01	1.35	0.48
RS1	535	700	0.30	1.16	0.11	0.01	1.73	0.17
WGHS	4924	1616	0.34	1.29	4.17E-09	0.02	1.45	0.01
WHI	2065	2011	0.36	1.33	3.05E-09	0.02	1.05	0.77
Meta-analysis	11420	8073		1.27	3.51E-09		1.23	0.02
Overall******								
Meta-analysis	12148	8777		1.27	3.69E-11		1.27	0.005

Multivariable model set 1 includes rs16969968 and the aggregate low frequency variant term:

All models adjusted for sex, age, PCs, and field center (if appropriate) as covariates;

*MAF stands for minor allele frequency; **For the aggregate low frequency variant terms, the MAF was calculated by the dividing the number of people with at least one low frequency/rare variant by the total number of people by 2; ***Primary sample compared nicotine dependent cases versus controls instead of heavy versus light smokers; ****Overall meta-analysis included the primary and replication samples.

Supplemental Table 5.3. Random effects meta-analysis results for multivariate model set 1 in African Americans

	Heavy Smokers	Light Smokers	rs16969968			Aggregate low frequency variants		
			MAF*	OR	p-value	MAF**	OR	p-value
Primary sample* **	854	534	0.06	1.46	0.04	0.06	1.46	0.04
Replication Studies								
ARIC	214	708	0.06	1.36	0.20	0.07	1.49	0.09
CHS	50	262	0.08	2.14	0.03	0.06	1.20	0.73
GENOA	73	298	0.05	1.93	0.09	0.08	1.38	0.35
HyperGEN	97	618	0.05	1.00	1.00	0.07	1.11	0.76
JHS	102	393	0.06	0.73	0.42	0.07	2.14	0.01
MESA	101	445	0.07	0.89	0.73	0.06	0.89	0.73
WHI	117	493	0.06	1.22	0.49	0.05	1.33	0.39
Meta-analysis	754	3217		1.24	0.10		1.42	0.004
Overall******								
Meta-analysis	1608	3751		1.30	0.01		1.40	0.0006

Multivariable model set 1 includes rs16969968 and the aggregate low frequency variant term:

All models adjusted for sex, age, PCs, and field center (if appropriate) as covariates;

*MAF stands for minor allele frequency; **For the aggregate low frequency variant terms, the MAF was calculated by the dividing the number of people with at least one low frequency/rare variant by the total number of people by 2; ***Primary sample compared nicotine dependent cases versus controls instead of heavy versus light smokers; ****Overall meta-analysis included the primary and replication samples.

Supplemental Table 5.4. Random effects meta-analysis results for multivariate model set 2 in European Americans

	Heavy Smokers	Light Smokers	rs16969968			rs2229961			rs79109919		
			MAF*	OR	p-value	MAF	OR	p-value	MAF	OR	p-value
Primary sample* **	728	704	0.36	1.28	0.003	0.02	1.71	0.10	*0.0003***	.	.
Replication Studies											
ARIC	1859	1507	0.33	1.44	1.07E-09	0.02	1.20	0.41	*0.0003***	.	.
CHS	513	856	0.34	1.07	0.49	0.02	1.27	0.47	0	.	.
COGA	209	133	0.33	1.52	0.06	0.02	0.60	0.43	*0.002***	.	.
ERF	138	77	0.36	0.86	0.25	0.02	1.26	0.38	0	.	.
FamHS	548	438	0.34	1.46	0.0005	0.01	1.02	0.97	0	.	.
GENOA	258	281	0.37	1.39	0.01	0.01	0.97	0.96	0	.	.
MESA	369	450	0.33	1.29	0.02	0.01	1.45	0.40	*0.0008***	.	.
RS1	535	700	0.30	1.16	0.11	0.01	1.73	0.17	0	.	.
WGHS	4924	1616	0.34	1.29	6.15E-09	0.02	1.38	0.03	0.0004	3.50	0.21
WHI	2065	2011	0.36	1.34	2.14E-09	0.02	1.05	0.79	0.0006	2.01	0.45
Meta-analysis	11418	8069		1.27	3.72E-09		1.23	0.02		2.60	0.16
Overall******											
Meta-analysis	12146	8773		1.28	3.60E-11		1.26	0.007		2.60	0.16

Multivariable model set 2 includes rs16969968, rs229961, rs8008750, and rs79109919. However, rs8008750 was non-polymorphic in European Americans in all of the studies so it is not included in this table.

All models adjusted for sex, age, PCs, and field center (if appropriate) as covariates;

*MAF stands for minor allele frequency; ** Because the minor allele of rs79109919 occurred less than 5 times in the primary sample, ARIC, COGA, and MESA, these results were not included in the meta-analyses; ***Primary sample compared nicotine dependent cases versus controls instead of heavy versus light smokers; ****Overall meta-analysis included the primary and replication samples.

Supplemental Table 5.5. Random effects meta-analysis results for multivariate model set 2 in African Americans

	Heavy Smokers	Light Smokers	rs16969968			rs2229961			rs8008750			rs79109919		
			MAF*	OR	p-value	MAF	OR	p-value	MAF	OR	p-value	MAF	OR	p-value
Primary sample* **	854	534	0.06	1.46	0.04	0.002	2.57	0.40	0.01	2.00	0.07	0.06	1.22	0.26
Replication Studies														
ARIC	214	708	0.06	1.49	0.11	0.005	0.51	0.46	0.02	1.04	0.94	0.05	1.70	0.03
CHS	50	262	0.08	2.11	0.04	0.005	1.60	0.73	0.01	1.74	0.54	0.04	0.93	0.92
GENOA	73	298	0.05	1.79	0.14	*0.001***	.	.	0.01	1.35	0.74	0.07	1.10	0.81
HyperGEN	97	618	0.05	1.03	0.92	*0.001***	.	.	0.02	1.13	0.84	0.05	1.12	0.78
JHS	102	393	0.06	0.89	0.78	*0.004***	.	.	0.02	1.31	0.62	0.05	3.16	0.001
MESA	101	445	0.07	0.87	0.68	0.009	1.00	1.00	0.02	2.16	0.14	0.05	0.88	0.75
WHI	117	493	0.06	1.22	0.50	0.005	0.94	0.94	0.01	1.79	0.35	0.04	1.10	0.81
Meta-analysis	754	3217		1.29	0.04		0.85	0.77		1.43	0.12		1.39	0.06
Overall******														
Meta-analysis	1608	3751		1.33	0.005		1.06	0.91		1.56	0.02		1.36	0.03

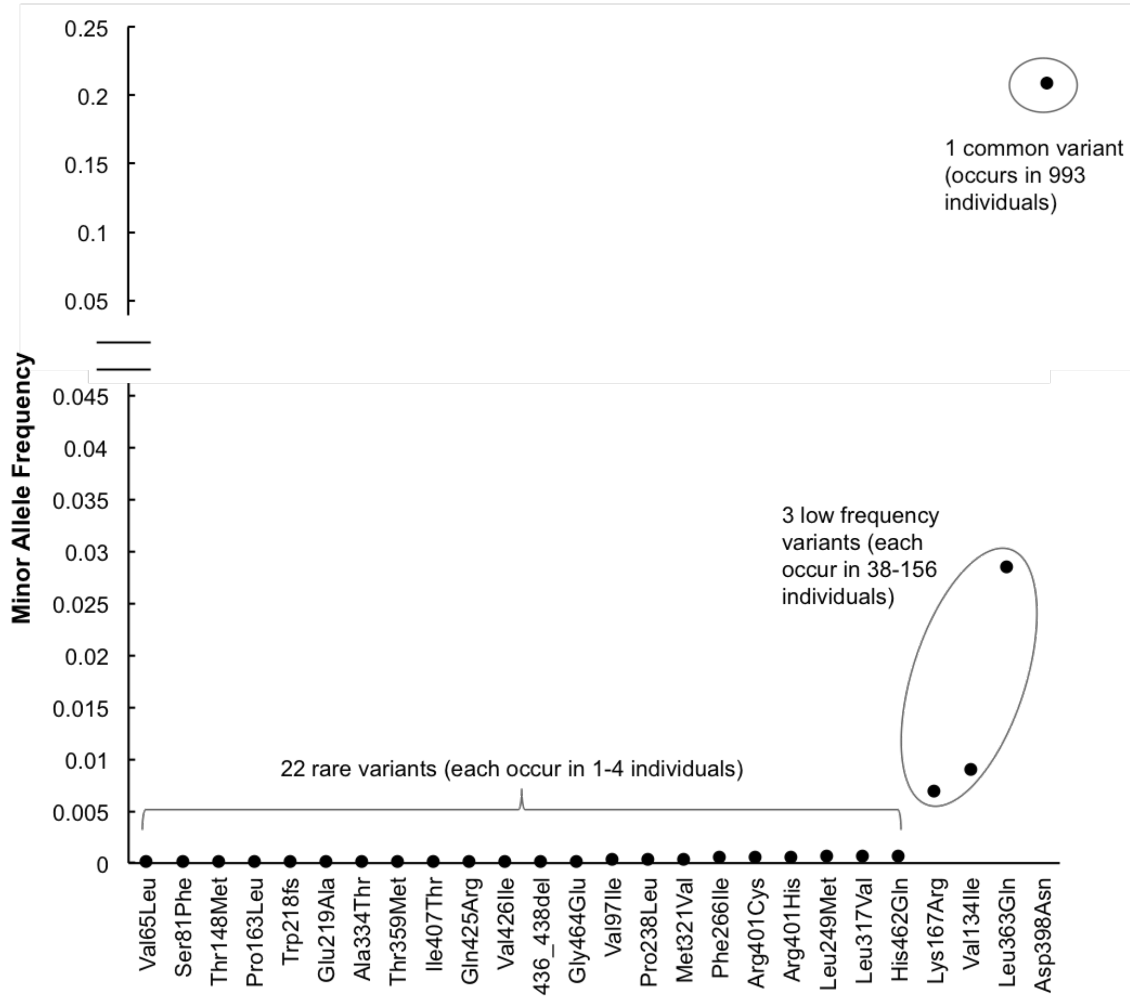
Multivariable model set 2 includes rs16969968, rs229961, rs8008750, and rs79109919.

All models adjusted for sex, age, PCs, and field center (if appropriate) as covariates.

*MAF stands for minor allele frequency; ** Because the minor allele of rs2229961 occurred less than 5 times in GENOA, HyperGEN, and JHS, these results were not included in the meta-analysis; ***Primary sample compared nicotine dependent cases versus controls instead of heavy versus light smokers; ****Overall meta-analysis included the primary and replication samples.

5.13 SUPPLEMENTAL FIGURE

Supplemental Figure 5.1. Twenty-six *CHRNA5* coding variants ordered based on minor allele frequency



5.14 SUPPLEMENTAL ACKNOWLEDGEMENTS

Primary Sample Analysis

The Collaborative Genetic Study of Nicotine Dependence was supported by P01CA89392 from the National Cancer Institute. The Genetic Study of Nicotine Dependence in African Americans was supported by R01DA025888 from the National Institute of Drug Abuse. Emily Olfson was supported by T32GM07200, UL1TR000448, TL1TR000449, and F30AA023685. LJB was supported by U19CA148172.

Center for Inherited Disease Research (CIDR)

Targeted sequencing services were provided by the Center for Inherited Disease Research (CIDR). CIDR is fully funded through a federal contract from the National Institutes of Health to The Johns Hopkins University, contract number HHSN268201100011I.

Atherosclerosis Risk Communities Study (ARIC)

The Atherosclerosis Risk in Communities (ARIC) study is carried out as a collaborative study supported by the National Heart, Lung, and Blood Institute (NHLBI) contracts (HHSN268201100005C, HHSN268201100006C, HHSN268201100007C, HHSN268201100008C, HHSN268201100009C, HHSN268201100010C, HHSN268201100011C, and HHSN268201100012C). The authors thank the staff and participants of the ARIC study for their important contributions. Funding support for "Building on GWAS for NHLBI-diseases: the U.S. CHARGE consortium" was provided by the NIH

through the American Recovery and Reinvestment Act of 2009 (ARRA) (5RC2HL102419). Analyses were additionally supported by R01 HL118305.

Cardiovascular Health Study (CHS)

This CHS research was supported by NHLBI contracts HHSN268201200036C, HHSN268200800007C, N01HC55222, N01HC85079, N01HC85080, N01HC85081, N01HC85082, N01HC85083, N01HC85086; and NHLBI grants U01HL080295, R01HL087652, R01HL105756, R01HL103612, R01HL120393, R01HL085251, R01HL068986 with additional contribution from the National Institute of Neurological Disorders and Stroke (NINDS). Additional support was provided through R01AG023629 from the National Institute on Aging (NIA). A full list of principal CHS investigators and institutions can be found at CHS-NHLBI.org.

The provision of genotyping data was supported in part by the National Center for Advancing Translational Sciences, CTSI grant UL1TR000124, and the National Institute of Diabetes and Digestive and Kidney Disease Diabetes Research Center (DRC) grant DK063491 to the Southern California Diabetes Endocrinology Research Center.

The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Collaborative Study on the Genetics of Alcoholism (COGA)

The Collaborative Study on the Genetics of Alcoholism (COGA), Principal Investigators B. Porjesz, V. Hesselbrock, H. Edenberg, L. Bierut, includes ten different centers: University of

Connecticut (V. Hesselbrock); Indiana University (H.J. Edenberg, J. Nurnberger Jr., T. Foroud); University of Iowa (S. Kuperman, J. Kramer); SUNY Downstate (B. Porjesz); Washington University in St. Louis (L. Bierut, A. Goate, J. Rice, K. Bucholz); University of California at San Diego (M. Schuckit); Rutgers University (J. Tischfield); Texas Biomedical Research Institute (L. Almasy), Howard University (R. Taylor) and Virginia Commonwealth University (D. Dick). Other COGA collaborators include: L. Bauer (University of Connecticut); D. Koller, S. O'Connor, L. Wetherill, X. Xuei (Indiana University); Grace Chan (University of Iowa); N. Manz (SUNY Downstate); J. Rohrbaugh, J-C Wang (Washington University in St. Louis); A. Brooks (Rutgers University); and F. Aliev (Virginia Commonwealth University). A. Parsian and M. Reilly are the NIAAA Staff Collaborators. We continue to be inspired by our memories of Henri Begleiter and Theodore Reich, founding PI and Co-PI of COGA, and also owe a debt of gratitude to other past organizers of COGA, including Ting-Kai Li, currently a consultant with COGA, P. Michael Conneally, Raymond Crowe, and Wendy Reich, for their critical contributions. This national collaborative study is supported by NIH Grant U10AA008401 from the National Institute on Alcohol Abuse and Alcoholism (NIAAA) and the National Institute on Drug Abuse (NIDA).

Erasmus Rucphen Family study (ERF)

The ERF study as a part of EUROSPAN (European Special Populations Research Network) was supported by European Commission FP6 STRP grant number 018947 (LSHG-CT-2006-01947) and also received funding from the European Community's Seventh Framework Program (FP7/2007–2013)/grant agreement HEALTH-F4-2007-201413 by the European Commission under the program 'Quality of Life and Management of the Living Resources' of 5th Framework

Program (no. QLG2-CT-2002-01254). High-throughput analysis of the ERF data was supported by joint grant from Netherlands Organization for Scientific Research and the Russian Foundation for Basic Research (NWO-RFBR 047.017.043). Exome sequencing analysis in ERF was supported by the ZonMw grant (project 91111025). We are grateful to all study participants and their relatives, general practitioners and neurologists for their contributions and to P Veraart for her help in genealogy, J Vergeer for the supervision of the laboratory work, P Snijders for his help in data collection and E.M. van Leeuwen for genetic imputation.

Family Heart Study (FamHS)

This study was supported by RO1HL118305 (D.C. Rao and I. B. Borecki) from NHLBI and R01-DK-089256 (I.B. Borecki) from NIDDK.

Genetic Epidemiology Network of Arteriopathy (GENOA)

Support for the Genetic Epidemiology Network of Arteriopathy (GENOA) was provided by the National Heart, Lung and Blood Institute (HL054464, HL054457, HL054481, HL071917, and HL87660) and the National Institute of Neurological Disorders and Stroke (NS041558) of the National Institutes of Health. Genotyping was performed at the Mayo Clinic (Stephen Turner, Mariza de Andrade, Julie Cunningham) and was made possible by the University of Texas Health Sciences Center (Eric Boerwinkle, Megan L. Grove-Gaona). Support for these analyses was provided by HL119443 and HL118305. We would also like to thank the families that participated in the GENOA study.

Hypertension Genetics Epidemiology Network (HyperGEN)

The hypertension network is funded by cooperative agreements (U10) with NHLBI: HL54471, HL54472, HL54473, HL54495, HL54496, HL54497, HL54509, HL54515, and 2 R01 HL55673-12. The study involves: University of Utah: (Network Coordinating Center, Field Center, and Molecular Genetics Lab); Univ. of Alabama at Birmingham: (Field Center and Echo Coordinating and Analysis Center); Medical College of Wisconsin: (Echo Genotyping Lab); Boston University: (Field Center); University of Minnesota: (Field Center and Biochemistry Lab); University of North Carolina: (Field Center); Washington University: (Data Coordinating Center); Weil Cornell Medical College: (Echo Reading Center); National Heart, Lung, & Blood Institute. For a complete list of HyperGEN Investigators: http://www.biostat.wustl.edu/hypergen/Acknowledge.html

The GWAS in HyperGEN was supported by R01 HL055673 (Donna Arnett, PI), and the analysis for this project was supported by R01 HL118305 (DC Rao and Ingrid Borecki, MPIs).

Multi-Ethnic Study of Atherosclerosis (MESA)

This research was supported by the Multi-Ethnic Study of Atherosclerosis (MESA) contracts N01-HC-95159, N01-HC-95160, N01-HC-95161, N01-HC-95162, N01-HC-95163, N01-HC-95164, N01-HC-95165, N01-HC-95166, N01-HC-95167, N01-HC-95168, N01-HC-95169 and by grants UL1-TR-000040 and UL1-RR-025005 from NCRR . Funding for MESA Family was provided by grants R01-HL-071051, R01-HL-071205, R01-HL-071250, R01-HL-071251, R01-HL-071252, R01-HL-071258, R01-HL-071259, and UL1-RR-025005. Funding for MESA SHARe genotyping was provided by NHLBI Contract N02-HL-6-4278. The provision of genotyping data was supported in part by the National Center for Advancing Translational

Sciences, CTSI grant UL1TR000124, and the National Institute of Diabetes and Digestive and Kidney Disease Diabetes Research Center (DRC) grant DK063491 to the Southern California Diabetes Endocrinology Research Center.

Rotterdam Study Cohort 1 (RS1)

The dedication, commitment, and contribution of inhabitants, general practitioners, and pharmacists of the Ommoord district to the Rotterdam Study are gratefully acknowledged.
The Rotterdam Study is funded by Erasmus MC and Erasmus University, Rotterdam, the Netherlands; the Netherlands Organisation for Scientific Research (NWO); the Netherlands Organisation for the Health Research and Development (ZonMw); the Research Institute for Diseases in the Elderly (RIDE); the Ministry of Education, Culture and Science; the Ministry for Health, Welfare and Sports; the European Commission (DG XII); and the Municipality of Rotterdam.

Women's Genomic Health Study (WGHS)

The WGHS is supported by HL043851 and HL080467 from the National Heart, Lung, and Blood Institute and CA047988 from the National Cancer Institute, the Donald W. Reynolds Foundation and the Fondation Leducq, with collaborative scientific support and funding for genotyping provided by Amgen.

Women's Health Initiative (WHI)

The WHI program is funded by the National Heart, Lung, and Blood Institute, National Institutes of Health, U.S. Department of Health and Human Services through contracts

HHSN268201100046C, HHSN268201100001C, HHSN268201100002C,

HHSN268201100003C, HHSN268201100004C, and HHSN271201100004C. This analysis was

funded by R21HL123677-01 (PI: Franceschini) and 1R01HL118305-01A1 (PI:Rao).

CHAPTER SIX:

Future steps to understand the role of targeted genes in substance use disorders

6.1 SUMMARY OF THE DISSERTATION

Overall, this dissertation illustrates the discovery potential of analyses focused on dissecting associations between genome-wide significant genes and substance use disorders. First, I illustrate in chapter 2 that many previously identified candidate genes are not strongly associated with alcohol dependence in a large genome wide association (GWA) study, highlighting that GWA studies can clarify the role of candidate genes for substance use disorders. I then conducted several hypothesis-driven analyses focused on three functional candidate genes (*ADH1B*, *CYP2A6*, and *CHRNA5*) that have reached genome-wide levels for alcoholism or smoking in GWA studies. By incorporating important environmental factors, critical developmental periods, and rare coding variants, I refined associations between these genes and substance use behaviors. In chapter 3, I showed that the high-risk environment of peer drinking eliminates the protective effect of an *ADH1B* variant on adolescent drinking milestones. In chapter 4, I demonstrated that a *CYP2A6* metabolism metric was not associated with smoking initiation or daily smoking, but slow metabolism was associated with increased risk of nicotine dependence among daily smokers. Finally, in chapter 5, I provided evidence that low frequency and rare *CHRNA5* coding variants contribute an independent risk effect to nicotine dependence. These findings add insight into the biological mechanisms that lead to alcohol and nicotine use disorders, two diseases with substantial public health implications in the US and worldwide (CDC, 2014, Stahre et al., 2014, WHO, 2014b, WHO, 2014a).

6.2 FUTURE DIRECTIONS

Replication in independent samples

These studies provide new evidence for the role of targeted genes in the development of substance use disorders. Although some of these new findings were replicated in independent samples (e.g. role of metabolism metric on nicotine dependence in chapter 4 and effect of low frequency variants on smoking behaviors in chapter 5), an appropriate replication sample was not available to verify all of the association findings in this dissertation. Replication of these findings in independent samples using a variety of different populations would provide additional evidence supporting the robustness and generalizability of our conclusions.

Experiments to understand biological mechanisms

Association findings from the analyses presented in this dissertation have led to new hypotheses about the role of genetic variation in the development of substance use disorders. An important next step is to experimentally test these hypotheses. For example, in chapter 6, I identified several low frequency and rare *CHRNA5* coding variants that confer an independent risk for nicotine dependence beyond the common rs16969968 variant. Previous biological studies of rs16969968 have shown that the risk A allele leads to decreased response to nicotine agonists in cell culture experiments (Bierut et al., 2008) as well as lower Ca^{2+} permeability and increased short term desensitization when incorporated into certain neuronal nicotinic receptors expressed in *Xenopus laevis* oocytes (Kuryatov et al., 2011). Similar electrophysiology patch-clamp experiments could be used to assess the functional effects of the newly identified low frequency and rare *CHRNA5* coding variants. Biological studies could inform our understanding

of the mechanisms by which these identified genetic variants alter risk for substance use disorders.

Test additional hypotheses and build composite risk scores

By focusing on variation in *ADH1B*, *CYP2A6*, and *CHRNA5*, we have developed a framework for understanding how environmental factors, developmental periods, and rare variants influence the roles of targeted genes in substance use disorders. This framework could also be used to test additional hypotheses that build on the findings presented in this dissertation. For example, beyond peer drinking, parental monitoring is a critical environment for the development of adolescent drinking behaviors and twin studies suggest that family factors modify heritable variation in youth alcohol involvement (Kendler et al., 2011, Miles et al., 2005). Based on our observation that the high-risk environment of adolescent peer drinking diminishes the protective effect of an *ADH1B* variant in chapter 2, we would hypothesize that the social context of low parental monitoring would likewise decrease the protective effect of metabolizing variants on early adolescent drinking behaviors. Similar hypothesis-driven studies of robust genetic factors may add to our knowledge of the complex role of these genes on the development of substance use behaviors.

Future studies may also seek to expand on these findings with individual genes to build composite risk scores that predict the development of substance use disorders based on several genetic and environmental factors. Accurate prediction tools could inform intervention strategies that aim to identify at-risk individuals and prevent disease progression.

Translation to long-term disease risk and tailored therapies

In this dissertation, I identified and refined associations between robust genes and substance use disorders. An important next step is to extend these findings to long-term disease risk as well as response to treatments. Excessive alcohol use and persistent tobacco smoking are both associated with increased risk of a variety of chronic diseases, and it is important to understand how observed genetic associations with substance use disorders translate to related disease risk and response to treatments. For example, previous studies show that rs16969968 is the strongest genetic risk factor for lung cancer and chronic obstructive pulmonary disease (Amos et al., 2008, Hung et al., 2008, Pillai et al., 2009, Thorgeirsson et al., 2008) as well as influences response to different smoking cessation therapies (Chen et al., 2012). An important next step is to test whether these low frequency and rare *CHRNA5* coding variants similarly increase the risk of smoking-related diseases and response to cessation treatment. This knowledge could inform personalized medical care of individuals who suffer from substance use disorders.

6.3 REFERENCES

Amos CI, Wu X, Broderick P, Gorlov IP, Gu J, Eisen T, Dong Q, Zhang Q, Gu X, Vijayakrishnan J, Sullivan K, Matakidou A, Wang Y, Mills G, Doheny K, Tsai YY, Chen WV, Shete S, Spitz MR, Houlston RS (2008) Genome-wide association scan of tag SNPs identifies a susceptibility locus for lung cancer at 15q25.1. Nature genetics 40:616-622.

Bierut LJ, Stitzel JA, Wang JC, Hinrichs AL, Grucza RA, Xuei X, Saccone NL, Saccone SF, Bertelsen S, Fox L, Horton WJ, Breslau N, Budde J, Cloninger CR, Dick DM, Foroud T, Hatsukami D, Hesselbrock V, Johnson EO, Kramer J, Kuperman S, Madden PA, Mayo K, Nurnberger J, Jr., Pomerleau O, Porjesz B, Reyes O, Schuckit M, Swan G, Tischfield JA, Edenberg HJ, Rice JP, Goate AM (2008) Variants in nicotinic receptors and risk for nicotine dependence. The American journal of psychiatry 165:1163-1171.

CDC. Tobacco-Related Mortality. Centers for disease control and preventionFebrary 6, 2014. Available at: http://www.cdc.gov/tobacco/data_statistics/fact_sheets/health_effects/tobacco_related_mortality/. Accessed October 16, 2014.

Chen LS, Baker TB, Piper ME, Breslau N, Cannon DS, Doheny KF, Gogarten SM, Johnson EO, Saccone NL, Wang JC, Weiss RB, Goate AM, Bierut LJ (2012) Interplay of genetic risk factors (CHRNA5-CHRNA3-CHRNB4) and cessation treatments in smoking cessation success. The American journal of psychiatry 169:735-742.

Hung RJ, McKay JD, Gaborieau V, Boffetta P, Hashibe M, Zaridze D, Mukeria A, Szeszenia-Dabrowska N, Lissowska J, Rudnai P, Fabianova E, Mates D, Bencko V, Foretova L, Janout V, Chen C, Goodman G, Field JK, Liloglou T, Xinarianos G, Cassidy A,

McLaughlin J, Liu G, Narod S, Krokan HE, Skorpen F, Elvestad MB, Hveem K, Vatten L, Linseisen J, Clavel-Chapelon F, Vineis P, Bueno-de-Mesquita HB, Lund E, Martinez C, Bingham S, Rasmuson T, Hainaut P, Riboli E, Ahrens W, Benhamou S, Lagiou P, Trichopoulos D, Holcatova I, Merletti F, Kjaerheim K, Agudo A, Macfarlane G, Talamini R, Simonato L, Lowry R, Conway DI, Znaor A, Healy C, Zelenika D, Boland A, Delepine M, Foglio M, Lechner D, Matsuda F, Blanche H, Gut I, Heath S, Lathrop M, Brennan P (2008) A susceptibility locus for lung cancer maps to nicotinic acetylcholine receptor subunit genes on 15q25. Nature 452:633-637.

Kendler KS, Gardner C, Dick DM (2011) Predicting alcohol consumption in adolescence from alcohol-specific and general externalizing genetic risk factors, key environmental exposures and their interaction. Psychological medicine 41:1507-1516.

Kuryatov A, Berrettini W, Lindstrom J (2011) Acetylcholine receptor (AChR) alpha5 subunit variant associated with risk for nicotine dependence and lung cancer reduces (alpha4beta2)(2)alpha5 AChR function. Molecular pharmacology 79:119-125.

Miles DR, Silberg JL, Pickens RW, Eaves LJ (2005) Familial influences on alcohol use in adolescent female twins: testing for genetic and environmental interactions. Journal of studies on alcohol 66:445-451.

Pillai SG, Ge D, Zhu G, Kong X, Shianna KV, Need AC, Feng S, Hersh CP, Bakke P, Gulsvik A, Ruppert A, Lodrup Carlsen KC, Roses A, Anderson W, Rennard SI, Lomas DA, Silverman EK, Goldstein DB (2009) A genome-wide association study in chronic obstructive pulmonary disease (COPD): identification of two major susceptibility loci. PLoS genetics 5:e1000421.

Stahre M, Roeber J, Kanny D, Brewer RD, Zhang X (2014) Contribution of Excessive Alcohol Consumption to Deaths and Years of Potential Life Lost in the United States. Preventing Chronic Disease 11:E109.

Thorgeirsson TE, Geller F, Sulem P, Rafnar T, Wiste A, Magnusson KP, Manolescu A, Thorleifsson G, Stefansson H, Ingason A, Stacey SN, Bergthorsson JT, Thorlacius S, Gudmundsson J, Jonsson T, Jakobsdottir M, Saemundsdottir J, Olafsdottir O, Gudmundsson LJ, Bjornsdottir G, Kristjansson K, Skuladottir H, Isaksson HJ, Gudbjartsson T, Jones GT, Mueller T, Gottsater A, Flex A, Aben KK, de Vegt F, Mulders PF, Isla D, Vidal MJ, Asin L, Saez B, Murillo L, Blondal T, Kolbeinsson H, Stefansson JG, Hansdottir I, Runarsdottir V, Pola R, Lindblad B, van Rij AM, Dieplinger B, Haltmayer M, Mayordomo JI, Kiemeney LA, Matthiasson SE, Oskarsson H, Tyrfingsson T, Gudbjartsson DF, Gulcher JR, Jonsson S, Thorsteinsdottir U, Kong A, Stefansson K (2008) A variant associated with nicotine dependence, lung cancer and peripheral arterial disease. Nature 452:638-642.

WHO. Media center: tobacco. World health organization. Available at: http://www.who.int/mediacentre/factsheets/fs339/en/. Accessed October 16, 2014.

WHO. Media centre: Alcohol. World health organization. Available at: http://www.who.int/mediacentre/factsheets/fs349/en/. Accessed October 16, 2014.