The Evolution of Addiction: A Case Study of Nicotine Dependence

by

Brooke Sadler

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved April 2014 by the
Graduate Supervisory Committee:

Dr. Ana Hurtado, Co-Chair
Dr. Alison Goate, Co-Chair
Dr. Kim Hill
Dr. Craig Nagoshi

ARIZONA STATE UNIVERSITY

May 2014

ABSTRACT

A variety of studies have shown that the tendency toward nicotine dependence has a genetic component. The work described in this thesis addresses three separate questions: i) are there unidentified SNPs in the nicotinic receptors or other genes that contribute to the risk for nicotine dependence; ii) is there evidence of ongoing selection at nicotinic receptor loci; and, iii) since nicotine dependence is unlikely to be the phenotype undergoing selection, is a positive effect on memory or cognition the selected phenotype. I first undertook a genome-wide association scan of imputed data using samples from the Collaborative Study of the Genetics of Nicotine Dependence (COGEND). A novel association was found between nicotine dependence and SNPs at 13q31. The genes at this newly associated locus on chromosome 13 encode a group of micro-RNAs and a member of the glypican gene family. These are among the first findings to implicate a non-candidate gene in risk for nicotine dependence. I applied several complimentary methods to sequence data from the 1000 Genomes Project to test for evidence of selection at the nicotinic receptor loci. I found strong evidence for selection for alleles in the nicotinic receptor cluster on chromosome 8 that confer risk of nicotine dependence. I then used the dataset from the Collaborative Studies on the Genetics of Alcoholism (COGA) and looked for an association between neuropsychological phenotypes and SNPs conferring risk of nicotine dependence. One SNP passed multiple test correction for association with WAIS digit symbol score. This SNP is not itself associated with nicotine dependence but is in reasonable ($r^2 = 0.75$) LD with SNPs that are associated with nicotine dependence. These data suggest at best, a weak correlation between nicotine dependence and any of the tested cognitive phenotypes. Given the reproducible finding of an inverse relationship

between SNPs associated with risk for nicotine dependence and cocaine dependence, I hypothesize that the apparently detrimental phenotype of nicotine dependence may confer decreased risk for cocaine dependence. As cocaine use impairs the positive rewards associated with social interactions, reducing the risk of cocaine addiction may be beneficial to both the individual and the group.

ACKNOWLEDGMENTS

I would like to thank my committee, especially my chair, Dr. Magdalena Hurtado, for supporting me from the moment I got to graduate school and encouraging me to follow my interests, even though they turned out to be completely different from where they started. Dr. Hurtado always believed in me and could always make me smile no matter how I was feeling. I would also like to thank my co-chair, Dr. Alison Goate, for believing that a cultural anthropologist could learn genetics, even when I had little previous training. Dr. Goate was extremely helpful throughout my experience, and provided invaluable guidance, training, editing, and encouragement. She taught me to be a better scientist and a more critical thinker. I could not have done this without either of these wonderful advisors. I would also like to thank the members of the Goate lab who helped me along the way and provided both scientific assistance and moral support on more than one occasion.

I need to thank my wonderful parents, Drs. J. Evan Sadler and Linda J. Pike, for never giving up on me and for never letting me give up on myself. I would not be here if it weren't for them, both literally and figuratively. I am lucky that they are both scientists as well, because their help with brainstorming ideas and editing assistance was part of the reason I was able to do this at all. I also need to thank my friends Beth Suda and Brigid Daly, for being there for me no matter what, and for sometimes knowing me better than I know myself. They always told me it would all be okay in the end because they believed in me.

Lastly, I need to thank my husband, Dr. Gabe Haller. Not only did he provide endless computer assistance and troubleshooting, he routinely edited drafts, helped me

talk through problems with the data, listened to my existential crises, and talked me down off of more than one metaphorical cliff. I literally truly could not have done this without him.

I am so lucky to have such wonderful advisors, friends and family. When I first got to graduate school, I felt confused and lost. I could never have predicted that this would be the topic of my dissertation. The completion of this thesis is proof that anything is doable when you are surrounded by people who want you to succeed.

TABLE OF CONTENTS

Page

LIST OF TABLES

Table Page

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

THE PUBLIC HEALTH IMPACT OF TOBACCO USE ACROSS POPULATIONS

Drug addiction is a pervasive problem across cultures and is both an economic and psychological struggle for the individuals and families involved. Today, tobacco use and the diseases resulting from its use are the number one cause of preventable deaths in the United States, accounting for approximately one out of every five deaths annually (Garrett, 2009). In fact, tobacco use is the single largest preventable cause of death and disease in the world today (WHO, 2011). Recent studies estimate that 6 million people die due to tobacco related illnesses every year and this figure will increase to 8.3 million deaths by 2030, possibly even increasing to 1 billion deaths in the 21st century (Mathers, 2012).

The health impact of smoking takes many forms. More people die from lung cancer each year than from any other type of cancer (ACS, 2009). It is estimated that 15-30% of lung cancer cases are linked to smoking, although smoking can cause cancer almost anywhere in the body (CDC, 2013). Even among those who quit smoking, there is an elevated risk of lung cancer, although this risk is less than in those who do not quit (Huang et al., 2008). Smoking also increases risk of stroke and coronary heart disease. For women, smoking can make it more difficult to conceive, and increase the risk of preterm birth, low birth weight and stillbirth, although it also affects men's sperm (CDC, 2013). Smoking affects things like bone density, teeth and gum health, and make diabetes harder to control (CDC, 2013). Chronic obstructive pulmonary disease (COPD), a serious lung disease often caused by smoking, is also among the leading causes of death

(N. L. Saccone, Culverhouse, et al., 2010). People who begin smoking when they are young are at higher risk of smoking related complications in middle age, and almost half of adolescents who initiate tobacco use will likely die from causes related to tobacco use (Arora, Mathur, & Singh, 2013). For children exposed to tobacco smoke *in utero* either through maternal smoking or second hand smoke, there is a risk of increased respiratory complications and decreased lung function, as well as increased risk of cancers and heart disease (Svanes et al., 2004).

The World Health Organization reports a wide distribution of smoking rates across nations. For example, Senegal has only a 4.6% rate of adult smoking, and Haiti only 9.7%, while Argentina has a rate of 40.4% and Turkey has a rate of 44% (WHO, 2011). Regionally, as of 2004, 3% of all deaths in Africa were attributable to tobacco, while 16% of all deaths in European countries were attributable to tobacco (Mathers, 2012).

Tobacco burden is increasing disproportionately in low and middle-income countries. The WHO estimates that within a few decades, more than 80% of tobacco-related deaths will occur in the developing world (Mathers, 2012). This increase in the developing world is due, in part, to consumption of both cigarettes and smokeless forms of tobacco. For example in Mumbai, India, 56% of women chew tobacco (Mackay, 2002) while only approximately 6% of adults, both men and women, smoke cigarettes (Mitchell, 2011). The increased prevalence of smokeless tobacco in India relative to cigarettes is due to cultural reasons. The most common form of smokeless tobacco use in India is betel quid chewing. Betel quid chewing is an ancient practice that has always been a part of religious, social, and cultural rituals. Originally, betel quid consisted of

betel leaf, pieces of areca nut, a few drops of lime (calcium hydroxide), several condiments, sweetening, and flavoring agents. After tobacco was introduced in India in the 17th century, it became an ingredient of the betel quid. Because it is associated with a socially accepted practice, smokeless tobacco use became widespread (NCI, 2010). Rates of dependence are different between smokers and users of chewing tobacco. Among smokers, ~22% are not addicted, ~33% show moderate symptoms of addiction, and ~44% show high levels of addiction (Bierut et al., 2007). With smokeless tobacco, rates of subjective dependence based on questionnaires seem similar to those of smoking, with the notable difference that smokeless tobacco users found their habit more enjoyable than smokers, and rated enforced abstinence as more unpleasant (Jarvis, 1991). Thus it is clear that cultural and ethnic differences play a role in the use and likely the dependence on nicotine.

In addition to cultural traditions, the wide variety of smoking and smokeless tobacco products available in low and middle-income countries make tobacco a very accessible commodity for adolescents. This leads to a high rate of smoking in the young. Approximately 80,000−100,000 adolescents initiate smoking every day and the WHO estimates that 25% of smokers had their first cigarette before age 10 (Mackay, 2002) Some children begin as young as 6 years of age as documented in a study in India (Arora et al., 2013).

Approximately 1 billion men worldwide are smokers, though this number is slowly declining (Mackay, 2002). Among male smokers, 35% live in developed countries and 50% live in developing countries. Approximately 250 million women smoke or about 1 out of every 5 smokers worldwide (Mackay, 2002). Among women,

smoking is declining in developed countries but it is still increasing in developing countries. This is due in part to the tobacco industry promoting certain brands specifically to women through images of slimness or sophistication (Mackay, 2002). The effects of these trends have become apparent in low and middle-income countries where one study found a much smaller than expected difference in the prevalence of smoking between boys and girls in India (Babar et al., 2010).

In the United States, there are significant ethnic differences in the rates of cigarette smoking. American Indians have the highest rate of smoking of any ethnic or minority group. American Indian males have a smoking rate of 42.4%, and American Indian females have a smoking rate of 42%. Among youth 12-17 years old, the rates of smoking are highest among American Indian females (17.8%), followed by American Indian males (16.7%), white females (12.4%), and white males (11.3%). Among black youth, the values are 5.6% for females and 6.1% for males, indicating that on average, white youth smoke more than black youth and females smoke more than males. By adulthood, these trends are reversed so that by cohort, there are higher rates of smoking in males than females and higher rates in blacks than whites. Other at risk groups include persons of low socio-economic status, persons with mental health and substance abuse issues, the gay/lesbian/transgender community, and persons living in the South and Midwest United States (Garrett, 2009).

As will be discussed below, the observed population and ethnic differences in rates of smoking and addiction can be partially explained by genetic differences. The fact that smoking is so prevalent and that it is highly comorbid with other disorders such as mental illness, heart disease and lung cancer, places a heavy health burden on the

community, particularly among poorer communities. Not surprisingly, cessation rates also vary by ethnicity. In a 2010 survey, 68% of current smokers expressed a desire to quit, and 52.4% had made a quit attempt in the past year. People over 65 had less desire to quit (53.8%) than those under 65 (70.2%). When broken down by ethnicity, blacks had the most interest in quitting (75.6%), followed by whites (69.1%), then persons of other or mixed race (62.5%), and lastly Hispanics (61.0%) (Malarcher, 2011). Overall this same survey showed that a recent cessation was most prevalent in whites (6%) than blacks (3.3%). Thus, although blacks had a higher rate of a desire to quit, they have lower rates of cessation. Illustrating just how difficult it is to quit smoking when addicted, one study found that pregnant women with the minor allele at rs1051730, a SNP in the *CHRNA5-A3-B4* cluster, in have an increased likelihood of continuing to smoke during pregnancy, a time when most women are more likely to quit than any other time in their lives (Freathy et al., 2009). Since so many more people try to quit than actually succeed, there is a definite need for intervention programs focused both on prevention and treatment. It is therefore important to understand the genetic basis for nicotine dependence to aid in population specific treatment and prevention programs.

EVIDENCE FOR A GENETIC BASIS FOR ADDICTION

A variety of approaches have been used to demonstrate that addiction, to any of a number of substances, has a genetically heritable basis. These approaches range from traditional family or twin studies to more molecularly-based studies that employ population genetics and genome sequencing to identify susceptibility loci. The

application of these different approaches to studies of the genetic basis of addiction are discussed below.

Family/Twin/Adoption studies of addiction. Differences in any trait must be due to either genetic or environmental factors or both. Family studies are one way in which the relative contributions of genes and environment to phenotypic variability can be measured. The idea behind family studies is that trait similarities among relatives (typically parent/offspring or between siblings) must be due to shared genes or environment. By comparing the phenotypic similarities across relatives with varying degrees of relatedness, the relative contribution of genes and environment to the phenotype can be inferred.

Family studies have shown significantly higher rates of drug abuse among siblings (especially those with affected parents) than in the general population. Merikangas et al. (Merikangas et al., 1998) interviewed drug-dependent individuals and their first-degree relatives. They found an 8-fold increase in the risk of drug dependence among the relatives of drug-dependent individuals compared to controls, indicating that family history represents a major risk factor for developing substance dependence. This familial aggregation applies to both generalized substance abuse as well as specific drugs (Gelernter & Kranzler, 2010). For example, one study found that the rates of alcohol, tobacco, marijuana and cocaine dependence were increased in the siblings of alcoholics as compared to controls (Bierut et al., 1998). The authors concluded that these dependencies are familial and that they are due to both common and specific additive factors. This means that these individuals have a genetically increased risk for substance dependence in general but also have an increased susceptibility to addiction to a specific

substance. Although family studies are consistent with the idea of a genetic component in addiction phenotypes, they cannot fully disentangle genetics from environment.

Unlike family studies, twin and adoption studies can distinguish between the effects of genes and environment on phenotypes. A common study design is to compare identical and fraternal twins reared together. Identical (monozygotic) twins share all of their genes and the same family environment. Fraternal (dizygotic) twins share half of their genetic inheritance, but like monozygotic twins, they share the same family environment. Therefore, if monzygotic twins are phenotypically more similar for the trait of interest than dizygotic twins, it can be assumed that this is due to the fact that monozygotic twins share twice as much genetic material as dizygotic twins. The heritability of a phenotype can be estimated from the difference between the correlation of the monozygotic twins with the phenotype and the correlation of the dizygotic twins with the phenotype (Nagoshi, 2011).

Supporting the findings from family studies, twin studies have found that addiction has a genetic basis. As far back as 1958, it was reported in a German population, that the concordance for smoking was significantly higher in monozygotic male twin pairs than in dizygotic male twin pairs (Fisher, 1958). This finding was subsequently replicated in twin studies in Finland and Sweden (Kaprio et al., 1982), the United States ((Carmelli, Swan, Robinette, & Fabsitz, 1990; Edwards, Austin, & Jarvik, 1995; Kendler et al., 1999), and Australia (Heath et al., 1993).

Twin studies with larger sample sizes have found significant genetic influences on specific aspects of smoking behavior. For example, (Heath & Martin, 1993) found that 53% of the variance in smoking persistence, defined as whether or not a smoker quits

smoking, was accounted for by genetic factors. In that same cohort, age of initiation was also strongly influenced by genetic factors (Heath, Kirk, Meyer, & Martin, 1999). A meta-analysis of several twin studies also identified a role for genetics in the initiation of smoking but found a smaller role for genetics on dependence in females than in males (M. D. Li, Cheng, Ma, & Swan, 2003). As was true in the family studies, twin studies have found evidence for a genetic basis for multiple substance abuse or dependency. Tsuang et al. (Tsuang et al., 1998) found that among 3372 male twin pairs, abuse of one drug was associated with a significant increase in abuse of other drugs than in the general population. Twin studies are useful because they can provide an estimate of the heritability of a phenotype, however they are still unable to adequately separate the effects of genetics and environment. To do this, one must use adoption studies.

Adoption studies attempt to compare the phenotypic similarities of biological parents and their adopted-away children with the phenotypic similarities of adoptive parents and their adopted children. Adoption studies may also compare the phenotypic similarities between sibling pairs of which one is adopted and the other is not (Nagoshi, 2011). In theory, adoption studies should be able to separate the effects of genetics on phenotype from those of shared environment.

As with family and twin studies, the results from twin and adoption studies support the idea that genetics plays a large role in substance abuse and dependence. For example, King et al. (King et al., 2009) found that the biological children of alcoholics had greater rates of alcoholism than the adopted children of alcoholics. An adoption study by Osler et al. (Osler, Holst, Prescott, & Sorensen, 2001) found that the main genetic influence on smoking behavior in adult adoptees was within the same generation

(i.e. most significant between adopted adults and their biological siblings). While these studies support a role for genetics in addictive behavior, the size of that contribution varies in different studies. For nicotine dependence, estimates of heritability have varied from 44-72% (Carmelli et al., 1990; Lessov, Swan, Ring, Khroyan, & Lerman, 2004).

Phenotypes. There are multiple phenotypes commonly used in the genetic studies of nicotine dependence. The main phenotype is the Fagerström Test of Nicotine Dependence (FTND) score. This questionnaire contains 6 questions, including two questions that are out of three points each, for a total possible score of 10 (Figure 1) (Heatherton, Kozlowski, Frecker, & Fagerstrom, 1991).

	0	1	2	3
1. How soon after you wake up do you smoke your first cigarette?	After 60 Minutes	31 – 60 minutes	6-30 minutes	Within 5 minutes
2. Do you find it difficult to refrain from smoking in places where it is forbidden, e.g., in church, at the library, cinema, etc?	No	Yes		
3. Which cigarette would you hate most to give up?	All others	The first one in the morning		
4. How many cigarettes/day do you smoke?	10 or less	11-20	21-30	31 or more
5. Do you smoke more frequently during the first hours of waking than during the rest of the day?	No	Yes		
6. Do you smoke if you are so ill that you are in bed most of the day?	No	Yes		

Figure 1. FTND questionnaire. Note that items 1 (time to first cigarette) and 4 (cigarettes per day) are worth up to 3 points each.

FTND score can be used either as an ordinal or dichotomous (case/control) variable. Typically a score of 4 or above is counted as a case, and under 4 as a control. Importantly, items 1 (time to first cigarette in the morning) and 4 (cigarettes per day) are

the two three point questions on the test, and cigarettes per day (CPD) is also often used in several studies as the primary phenotype when FTND data are not available. A person is given a CPD score of 0 if they smoke 0-10 cigarettes per day, a score of 1 for 11-20 cigarettes, 2 for 21-30 cigarettes, and a 3 for over 30 cigarettes per day.

Table 1.

DSM-IV dependence criteria

Criterion 1	Tolerance, as defined by either of the following: a) a need for markedly increased amounts of the substance to achieve intoxication or the designed effect, or b) markedly diminished effect with continued use of the same amount of the substance
Criterion 2	Withdrawal, as manifested by either of the following: a) the characteristic withdrawal syndrome for the substance, or b) the same (or a closely related) substance is taken to relieve or avoid withdrawal symptoms (4 or more of the following: depressed or dysphoric mood; insomnia; irritability, frustration or anger; anxiety; difficulty concentrating; worry or impatience; decreased heart rate, increased appetite or weight gain)
Criterion 3	The substance is often taken in larger amounts or over a longer period than was intended
Criterion 4	There is a persistent desire or unsuccessful efforts to cut down or control substance use.
Criterion 5	A great deal of time is spent in activities necessary to obtain the substance, use the substance or recover from its effects
Criterion 6	Important social, occupational or recreational activities are given up or reduced because of substance use.
Criterion 7	The substance use is continued despite knowledge of having a persistent or recurrent physical or psychological problem that is likely to have been caused or exacerbated by the substance,
Specifications	With physiological dependence. This specification should be used when substance dependence is accompanied by evidence of tolerance (Criterion 1) or withdrawal (Criterion 2) Without physiological dependence. This specification should be used when there is no evidence of tolerance (Criterion 1) or withdrawal (Criterion 2). In these subjects substance dependence is characterized by a pattern of compulsive use (at least 3 of Criteria 3-7)

Another measure of nicotine dependence sometimes used is the DSM-IV criteria.

Dependence is indicated by the presence of 3 or more of the 7 criteria within 12 months (Table 1). Like the FTND, these criteria can either be used as ordinal or dichotomous values. When used as ordinal values, this is what is known as DSM-IV symptom count. Overall, FTND score is the most comprehensive phenotype, since it includes CPD and was designed specifically for nicotine dependence, whereas the DSM-IV criteria are the same for all drugs of abuse (cocaine, alcohol, opium, etc.).

Candidate Gene Studies. Once it has been determined that a trait of interest can be partially explained by genetic differences between individuals, i.e. that it is heritable, the next step is to determine which genes and in particular which genetic variants in or near those genes alter an individual's disease risk. One approach for identifying susceptibility genes is the candidate gene approach. In these studies, the researcher uses knowledge of the pathogenesis of a disease or the likely biological mechanisms underlying a trait to select genes that are likely to be associated with the phenotypic trait(s) being measured. For example, the cholinergic nicotinic receptors (*CHRNs*) were chosen for the list of candidate genes for nicotine dependence studies as they are the major receptors for nicotine in the central nervous system. Similarly, drug metabolism genes such as *CYP2A6* and *ADH2* are candidates for nicotine and alcohol dependence respectively because of their roles in the metabolism of these drugs.

Once a list of genes is created, genotyping is performed on a set of known single nucleotide polymorphisms (SNPs) in the genes of interest. SNPs are single base changes in DNA sequence that have previously been identified as genetic variants in a particular region in the genome. SNPs that are in or near the chosen candidate genes are selected for sequencing under the assumption that these variants are most likely to affect gene

function or alter gene expression or be correlated with SNPs that do. If there are many SNPs in these regions often only one of several highly correlated SNPs will be chosen for genotyping. SNPs are the most abundant form of variation, although there are other types such as insertions and deletions (indels), repeats, and copy number variants (CNVs) that can be important as well. Follow-up sequencing of promising candidate genes can narrow down the causative SNP(s) as well. For example, Yu et al. (Yu et al., 2006) followed up a study identifying DOPA decarbolayse (*DDC*) on chromosome 7 by genotyping several SNPs in the gene and finding an association with a particular SNP and FTND score. They sequenced this region and discovered that this SNP is in the same intron as an intronic splicing enhancer for a neuronal isoform lacking exons 10-15.

Candidate gene studies are useful because they facilitate comparison across populations by allowing investigators to determine if the same variants are seen within and between populations. They allow for genotyping large numbers of individuals with few SNPs to analyze. In a candidate gene study using FTND as the phenotype, Ma et al. (Ma, Payne, Nussbaum, & Li, 2010)analyzed 25 SNPs in the glutamate receptor gene ***(****GRIN3A),* which regulates ion flow in the brain, using a total of 2,037 individuals from 602 nuclear families. They found two SNPs in the pooled European-American and African American sample significant after multiple test correction, and one SNP that was significant only in the European-American sample. These data, among other candidate gene studies indicate ethnic differences in the genetic basis for nicotine addiction.

The candidate gene approach has intrinsic limitations because the studies are biased by prior knowledge of the physiological role of the candidate genes in relation to the phenotype. If the biology of the phenotype is not fully understood, it is possible to

unintentionally omit genes that are important in explaining the genetics of that phenotype. Therefore, unbiased approaches such as genome-wide linkage and association scans have been developed to uncover unpredictable relationships.

Linkage Studies of Addiction. Linkage studies are a genome-wide and unbiased method of determining genetic regions contributing to a trait. In a linkage study, each individual is assessed for presence or absence of the phenotype of interest as well as presence or absence of genetic markers (DNA sequence that is known to vary in size or sequence) with known locations. The main goal is to link the phenotype with a genetic marker close to the disease gene. If the phenotype of interest consistently co-occurs with a particular allele of a genetic marker in a family, then those two traits are not independently assorting and must be close to each other on the same chromosome (Nagoshi 2011). Using the primary phenotype of FTND score, Swan et al. (Swan et al., 2006) discovered a linkage peak on chromosome 6, whose support interval is very close to the opioid receptor *OPRM1*, and includes mitogen-activated protein kinase 4 (*MAP3K4*) and lysophosphatidic acid acyltransferase, delta (*LPAAT*-delta), both previously described candidate genes for nicotine dependence (P. F. Sullivan et al., 2004). In a study of 368 Dutch sibling pairs, using the phenotype of age at first cigarette, Vink et al. (Vink, Posthuma, Neale, Eline Slagboom, & Boomsma, 2006) discovered a linkage peak on chromosome 5 encompassing the dopamine receptor 1 (*DRD1*), a known candidate gene for smoking. In a study of 634 pedigrees totaling 2,881 people, Gelernter et al. (Gelernter et al., 2007) completed a genome-wide linkage scan and discovered a risk locus for nicotine dependence on chromosome 5 on the basis of FTND score in the region of peptidylglycine alpha-amidating monooxygenase (*PAM*) and

cocaine and amphetamine regulated transcript (*CART*). Interestingly, the association found by Gelernter et al. (2007) was significant in their African-American sample (p=0.001) but not in their European-American sample. These are just a few examples of the many linkage studies on nicotine dependence. However, an even more precise non-biased method is the genome-wide association scan.

The Power of GWAS Studies. A genome-wide association scan (GWAS) takes advantage of linkage disequilibrium in a population to detect all common single nucleotide variations in an individual. If a SNP on the GWAS chip is associated with a disease phenotype in your sample, it could be because that SNP is causing the disease, or more likely, the SNP is in linkage disequilibrium with the functional SNP causing the disease. As mentioned previously, linkage disequilibrium is the tendency for two or more alleles to be inherited together more often than expected by chance, due to reduced genetic recombination (or distance) between them. A GWAS compares DNA variation at millions of SNPs in individuals that do (affected) or do not (unaffected) exhibit the trait of interest. The frequency of each SNP is then compared between affected and unaffected individuals to determine if there is a significant association between genotype and phenotype. Alternatively, quantitative traits may be examined to test for association between genotype and the levels of the quantitative trait.

To genotype the large number of genetic variants spread out across the genome, genotyping microarrays were invented which allow one to genotype millions of SNPs simultaneously. The microarrays are glass slides with immobilized pieces of DNA complementary to the DNA immediately 5' of the SNP position. The individuals' DNA is hybridized to the microarray and the immobilized DNA probe is extended one base

using nucleotides, each of which has a unique fluorescent dye, using the individuals DNA as the template. An individual's genotype can then be determined based on the fluorescent signal at each position on the array.

The SNPs on the array are selected such that the majority of common SNPs (frequency > 5%) in the genome are either genotyped or in high linkage disequilibrium with a genotyped SNP (Visscher, Brown, McCarthy, & Yang, 2012). The association between genotypes obtained from the array and the trait of interest is then determined. For example, if one allele of a SNP is found more often in affected individuals than unaffected individuals, then it is associated with the disease phenotype. The p-values that are generally accepted for a genome-wide significant association in a GWAS are $< 5x10^{-8}$. This is equal to 0.05/1,000,000. A value of 0.05 is the p-value for significance of 1 SNP. It is divided by 1,000,000 (the standard Bonferroni correction) because it is conservatively estimated that there are 1,000,000 independent regions of linkage disequilibrium in the genome. This makes the number of samples needed to achieve significance large.

GWAS have been used to identify novel genetic variants associated with many disease phenotypes. For example, Ripke et al. (Ripke et al., 2013) performed a GWAS for schizophrenia and then meta-analyzed their results with those of other GWAS. They were able to identify 13 novel loci genome-wide significantly associated with schizophrenia. In a two-stage meta-analysis of GWAS for Alzheimer's disease, Lambert et al. (Lambert et al., 2013) were able to identify 11 novel loci associated with the disease at the genome-wide significant level. Thus, these examples demonstrate that GWAS are useful for identifying SNPs associated with the disease phenotype that otherwise would

not have been considered.

GWAS are largely a discovery tool and are not necessarily meant to capture all possible information. Rather they can be used to give hints as to what genes might be important for the trait of interest. GWAS identify loci which may contain a single gene or many genes depending on LD. Recent studies using pathway analyses or co-regulation analyses have demonstrated that GWAS genes fall into distinct pathways, providing biological insight into the disease even if the functional variant is not identified. By combining family and twin studies with candidate gene studies and GWAS, significant progress has been made in identifying genetic factors that contribute to addictive behavior in general and to addiction to specific substances in particular.

GENETICS OF NICOTINE DEPENDENCE

Candidate gene studies. The most comprehensive candidate gene study of nicotine dependence is that of Saccone et al. (S. F. Saccone et al., 2007). Here, over 300 candidate genes and 3713 SNPs were examined in 1050 cases and 879 controls. The strongest association with FTND was a SNP in *CHRNB3* ($p=9.4x10^{-5}$). Multiple SNPs in the *CHRNB3-A6* and *CHRNA5-A3-B4* clusters were among the top hits, including the on-synonymous change, rs16969968. The other two genes with p-values of less than 0.001 in this study were the potassium inwardly rectifying channel 6 (*KCNJ6)* and gamma-aminobutyric acid receptor a4 (*GABRA4)*.

Because of the enhancing role nicotine has on the actions of dopamine, several candidate gene studies have focused on dopamine receptors. Most drugs of abuse act on the mesolimbic dopaminergic neurons in the brains of humans and many other mammals.

The neurotransmitter dopamine is involved in neural activity related to motivation, emotion, food intake, liking, learning, wanting and cognition. Activation and reinforcement of this system is a necessary part of drug abuse (Koob, 1996). The dopaminergic system is therefore crucial to addiction and it is hypothesized that mutations in these pathways are associated with risk for dependence on multiple drugs, including nicotine.

There are two main dopamine receptor subtypes, *DRD1*-like and *DRD2*-like receptors. Linkage scans have associated the *DRD1* region with cigarette consumption (N. L. Saccone, Neuman, Saccone, & Rice, 2003) and age of initiation in a study of Dutch twin and sibling pairs (Vink et al., 2006). Furthermore, two *DRD1* polymorphisms, A48G (DdeI aka rs4532) and T1403C (rs686), have been associated with heavy smoking behavior and high scores on the Fagerstrom Test for Nicotine Dependence in a pooled sample of 2037 participants from the United States of African-American or European-American origins (Huang et al., 2008). Thus, *DRD1* appears to be associated with smoking behavior and nicotine addiction.

DRD2 is associated with pleasure, and has specifically been called a "reward gene" (Blum et al., 2011). The most studied polymorphism in the *DRD2* region is TaqIA, a C >T substitution (rs1800497) on chromosome 11q22–q23. This SNP has been associated with pathological gambling, overeating, schizophrenia, heroin addiction, nicotine dependence, alcoholism, and other psychiatric disorders (Comings et al., 1996; Epstein et al., 2007; Monakhov, Golimbet, Abramova, Kaleda, & Karpov, 2008; Noble et al., 1998; Smith, Watson, Gates, Ball, & Foxcroft, 2008; Xu et al., 2004). This SNP was originally thought to be in the 3' untranslated region of *DRD2*, but has more recently

been shown to be located in exon 8 of a neighboring gene (10 kb downstream of the *DRD2* gene), named *ANKK1* (ankyrin repeat and kinase domain containing 1) (Blum et al., 2011). The SNP causes a nonsynonymous coding change (Glu713Lys) that can affect *DRD2* receptor expression and synthesis of dopamine in the brain. Confirming the relevance of *ANKK1* in nicotine dependence, Huang et al. (Huang et al., 2009) found a significant correlation with nicotine dependence and a second non-synonymous functional SNP (rs2734849) in the *ANKK1* gene. This correlation was seen only in African-American and not European-American smokers, again implying ethnic differences in the genetic basis for nicotine addiction. The A1 allele was also associated with a higher consumption of heroin in Spanish individuals, particularly males (Perez de los Cobos et al., 2007) as well as increased levels of craving after heroin exposure in Chinese individuals (Y. Li et al., 2006). However, despite some promising results, *DRD2* has not been well replicated over time.

Several studies link the *DRD2* gene to specific smoking related phenotypes. As early as 1996, Taq1A had been implicated as a risk factor for smoking, as one group found a significantly higher portion of A1 alleles in smokers than non-smokers ($p=10^{-8}$) (Comings et al., 1996). Frequencies of A1 and the more 5' B1 alleles have been associated with age of smoking initiation and fewer attempts to quit (Spitz et al., 1998). The Taq1A polymorphism has been associated with smoking progression in adolescents, especially among those with depressive symptoms (Audrain-McGovern, Lerman, Wileyto, Rodriguez, & Shields, 2004). Genotype status at a *DRD2* intron 2 simple tandem repeat was related to cigarettes per day (P=0.035) and heaviness of smoking index (P=0.049) (Vandenbergh et al. 2007). Two common *DRD2* haplotypes were

associated with the quantity of smoking and drinking in a sample of alcoholics who were habitual smokers (Preuss, Zill, Koller, Bondy, & Sokya, 2007). A significant association between TaqA1 genotype and maximum duration of quit time among male Egyptian smokers has been reported (Radwan et al., 2007). Together, these findings suggest that although there are mixed results with association in this region, that the dopaminergic pathway is important to addiction. These genes are likely associated with dependencies to several different drugs and probably confer a general predisposition to addiction, rather than addiction to a particular substance. It is curious that GWAS do not seem to pick up any association with nicotine dependence in the *DRD* genes. Perhaps they do have a nominally significant p-value, but it is not mentioned because it does not pass genome-wide significance, or perhaps the effect of *DRD* genotypes is too downstream of the effect of nicotine to change how nicotine is perceived and thus change a person's likelihood of becoming addicted to it.

GWAS Studies. Several genetic variants that modify susceptibility or resistance to nicotine dependence have been identified by GWAS (Berrettini et al., 2008; Bierut et al., 2007; Thorgeirsson et al., 2008). Perhaps not surprisingly, the loci identified by GWAS as associated with nicotine addiction mainly include genes encoding neuronal cholinergic nicotinic receptors (*CHRNs*) and nicotine metabolizing genes.

Neuronal cholinergic nicotinic receptors *(CHRNs)* are a heterogeneous class of cation (positively charged) channels expressed in the central and peripheral nervous system. There are 11 neuronal *CHRN* genes, each of which encodes a receptor subunit. The neuronally expressed nicotinic receptors consist of combinations of alpha and beta subunits, encoded in humans by 8 alpha (α2-α 7, α9-α10) and 3 beta (β2-β4)

genes (Bierut, 2009). These subunits form homo- or hetero-pentameric subtypes, which are present in various regions throughout the nervous system. To form a receptor, five subunits must be combined within the cell and the specific combination of these subunits defines the receptor subtype.

In the body, the opening of these channels is controlled by the endogenous ligand, acetylcholine, a chemical produced by neurons to activate other nearby neurons. Nicotine, the major psychoactive chemical present in tobacco smoke is a chemical present in the environment that can also stimulate the opening of these nicotinic acetylcholine receptor ion channels (Gotti et al., 2007)Nicotine has differing effects on the brain depending on the receptor subtype and location. For example, when nicotine is bound to *α3ß4* receptors in the medial habenula and interpeduncular nucleus, there is a higher sensitivity to the aversive effects of nicotine. By contrast, α4ß2 receptors in the ventral tegmental area play a major role in nicotine self-administration in mice (Frahm et al., 2011).

A number of GWAS studies have been performed that demonstrate an association between the nicotinic receptors and smoking. The strongest association between nicotinic receptors and nicotine addiction is a non-synonymous change (rs16969968, D398N) in the gene encoding the α5 subunit of the nicotinic receptor *(CHRNA5)* on chromosome 15 (Bierut et al., 2008; N. L. Saccone et al., 2009; S. F. Saccone et al., 2007; Spitz, Amos, Dong, Lin, & Wu, 2008; Thorgeirsson et al., 2008; Weiss et al., 2008). When cells are made to express nicotinic receptors containing the minor allele form of this SNP (398N), agonists induce less channel opening and cell activation than in cells that express receptors containing the major allele (398D). Thus, this SNP results in

a significant functional change in the behavior of this ion channel, causing more nicotine to be needed in individuals with the minor allele to produce the same effect.

The SNP rs16969968 is highly associated with CPD (OR = 1.9, p = 5.96 e^{-31}) (N. L. Saccone, Schwantes-An, et al., 2010). The association between rs16969968 and smoking behavior has been replicated in several studies and across several populations. Interestingly, the minor allele (meaning the less common base pair at this SNP) of rs16969968 is present at a frequency of 37% in European populations, but is almost absent in African populations. Although it is rare in African-Americans, the odds ratio is similar to that of European-Americans, further increasing the confidence in its effect on nicotine dependence (Bierut et al., 2008). While the association of SNP rs6969968 with nicotine addiction is highly significant, it does not explain a large percentage of the overall heritability of nicotine dependence. In fact, it accounts for less than 5% of the variance observed in nicotine dependence (Bierut, 2011). This suggests that other loci also contribute to a susceptibility to nicotine addiction.

Interestingly, with regard to cocaine dependence, it has been shown that the chromosome 15 variant in the *CHRNA5* nicotinic receptor, rs16969968, that influences the development of nicotine dependence, independently contributes to cocaine dependence as well. In a European-American sample, the minor allele of this variant increased the risk for nicotine dependence, but decreased risk for cocaine dependence (Grucza et al., 2008). The reason for this reversed effect has been suggested to be because nicotinic receptors are involved in both excitatory and inhibitory modulation of dopamine-medicated reward pathways (Bierut, 2011). The fact that variants can affect multiple substance dependencies and in different directions underscores the complexity in

investigating these behavioral phenotypes due to the comorbidity that is so common with substance abuse.

Chromosome 8 represents another target for studies of nicotine addiction because the genes encoding the ß3 *(CHRNB3)* and α6 *(CHRNA6)* nicotinic receptor subunits reside there. The nicotinic receptor gene cluster on chromosome 8 that includes the nicotinic receptor subunit gene cluster *CHRNB3-CHRNA6* is correlated with smoking behavior. The same GWAS that identified the association between nicotine addiction and the *CHRNA5* risk allele also identified several SNPs in high linkage disequilibrium around the gene *CHRNB3* (S. F. Saccone et al., 2007). Associations were later found with 3 SNPs in *CHRNB3,* including the one previously identified by Saccone et al. 2007 (Hoft et al., 2009; Zeiger et al., 2008). A SNP in *CHRNA6* was also found to be associated with nicotine dependence (Hoft et al., 2009). These results indicate that there are genes on chromosome 8 that are promising targets for discovering some of the missing heritability for nicotine dependence.

Adding support for the necessity of further study of this region, a SNP tagging a region of linkage disequilibrium was discovered in the *CHRNB3-A6* region (N. L. Saccone et al., 2009). Subsequent follow-up produced evidence for two distinct regions of association: one within *CHRNB3* and the other in a nearby non-genic region upstream of the *B3-A6* cluster (N. L. Saccone, Culverhouse, et al., 2010). A separate GWAS discovered additional correlations between nicotine dependence and the *CHRN* cluster on chromosome 8, as well as the nicotine metabolizing gene, *CYP2A6* on chromosome 19 (Thorgeirsson et al., 2010). This region showed genome-wide significant association with nicotine dependence (rs6474412 $p = 1.4\ e^{-8}$) (Thorgeirsson et al., 2010). These

results have been replicated and additional variants discovered (Rice et al., 2012). Although a few studies found no association with smoking behavior and the *B3-A6* region, this may have been due to small sample size (N=277, (Etter et al., 2009); N=485, (Keskitalo-Vuokko et al., 2011). Together, these findings strongly implicate the nicotinic cholinergic receptor genes on chromosome 8 in nicotine addiction. However, further study is needed to fully characterize the genetic variation this region as it relates to nicotine dependence. Rare variants at this locus have also been associated with alcohol and cocaine dependence (Haller et al., 2013), and low-frequency variants have also been associated with risk for cocaine dependence (Sadler et al., 2014). Although biological mechanisms in this region have been elusive, one study identified a GWAS signal of smoking behavior in the region to be strongly associated with changes in a DNase I sensitivity site in the region (Degner et al., 2012). Most recently, a GWAS of imputed data using FTND discovered a genome-wide significant association near a group of micro-RNAs (*MIR17HG*) and a member of the glypican gene family (*GPC5*) on chromosome 13. Notably, the *GPC5* gene is expressed mainly in the adult brain and was previously shown to be involved in the behavioral response to alcohol. These are among the first findings to implicate a non-candidate gene in risk for nicotine dependence, however replication to date has been difficult (unpublished data).

NICOTINIC RECEPTORS IN MEMORY AND LEARNING

Nicotinic receptors are distributed throughout the nervous system and clearly have a role in attention, memory and learning. Agonists of nicotinic receptors have been shown to improve, and receptor antagonists to impair, performance in cognitive tasks

(Dajas-Bailador & Wonnacott, 2004). Nicotine has been shown to improve working memory, although reference memory is not affected by either acute or chronic nicotine administration (Levin & Simon, 1998). The α4β*2,* α*3*β*2,* and α*7* receptors in the hippocampus appear to be important for working memory functions and Greenwood et al. (Greenwood, Parasuraman, & Espeseth, 2012) found substantial evidence in the literature to support the hypothesis that the minor allele of rs1044396, a SNP in *CHRNA4*, is associated with the ability to focus attention very strongly on a target at the expense of all outside stimuli.

Also supporting a role of nicotine in attention, the study of Thiel et al. (Thiel, Zilles, & Fink, 2005) showed that nicotine enhanced the reorientation of attention in visuospatial tasks in a German, nonsmoking population. Murphy & Klein (Murphy & Klein, 1998) also showed that nicotine enhanced visuospatial reorientation in casual smokers immediately after smoking a cigarette. Studies using fMRI to examine behavioral performance and regional brain activity have shown altered neuronal activity responsible for increased attention and arousal with nicotine as compared to placebo (Kumari et al., 2003). Ernst et al. (Ernst et al., 2001) found that administration of nicotine improved reaction time in focused attention tasks in both smokers and non-smokers.

Consistent with a role of nicotine in cognition, several recent studies have demonstrated cognitive differences between smokers and non-smokers. Winterer et al. (Winterer et al., 2010) found an association between nicotine dependence risk variants in *CHRNA5* and lower cognitive performance scores. The authors suggested that these variants may increase the risk for nicotine dependence because the individuals seek out

nicotine for "reasons of cognitive enhancement". Yakir et al. (Yakir et al., 2007) reported that non-smokers had better performance than current smokers on cognitive tests involving sustained attention, control of impulsivity, and planning. The authors suggested that improvement by nicotine of cognitive function in these domains might predispose young women who initiate cigarette smoking to maintain their smoking behaviors for purposes of self-medication if they had deficits in these domains. A similar explanation could account for the results of Hong et al. (Hong et al., 2010) who showed that the rs16969968 risk allele was associated with decreased resting state functional connectivity of the *CHRNA5* nicotinic receptor-expressing regions in the dorsal anterior cingulate–ventral striatum/extended amygdala circuit.

A study attempting to determine the association between SNPs in nicotinic receptors and cognitive function found association between cognitive function and variants in the receptors *CHRNA2, A4, A5, A7, A9, A10, B2* and *B3*, as well as with several related haplotypes (Rigbi et al., 2008). As a group, smokers performed worse overall than non-smokers. However, smokers who carried a particular combination of genetic variants at these loci performed better than non-smokers who carried a different combination at these loci (Rigbi et al., 2008). Given that nicotine improves cognition in the domains investigated in the study, the authors suggested that these variants would render a person more or less likely to smoke in accordance with the direction of effect. Another study found significant impairments in smokers in visual attention and cognitive impulsivity, regardless of smoking quantity. The authors argued for an *a priori* cognitive deficit in smokers and suggest these deficits be considered as phenotypes for future research (Wagner et al., 2013). Together, these data are consistent with the hypothesis

that a predisposition to nicotine dependence could be due to the effects of SNPs either in the nicotinic receptors or in other loci causing variability in memory and learning, for which other SNPs in nicotinic receptors could be attempting to compensate.

THE EVOLUTION OF NICOTINE DEPENDENCE

Addictive drugs are habit forming because they act on brain circuits that subvert more natural and biologically significant rewards (Nesse & Berridge, 1997). These pathways evolved in the adaptive context of inducing positive emotions in the presence of a fitness benefit, and negative emotions in situations where defenses may have been necessary. Selective pressures in our ancestral environments were likely not on addiction, but rather on behaviors or characteristics that enhanced survival and reproductive success.

As discussed above, numerous studies indicate that nicotine enhances cognitive performance and that individuals with subtle differences in cognition or memory may be predisposed to addiction or become nicotine addicted in order to improve their performance. Therefore, it is possible that the genes associated with nicotine addiction were selected for their ability to improve cognition or to counteract negative effects on cognition mediated by other loci.

Population differences in frequencies of alleles related to memory and cognition could answer questions about where and when these alleles arose. For example, rs16969968 in *CHRNA5,* which is so strongly associated with nicotine dependence, is present at a frequency of 37% in Europeans, yet is nearly absent in Africans and Asians. From the frequencies of this allele in different populations worldwide, it is clear that it

had to have arisen after the out-of-Africa migration, due to the fact that it is at nearly zero percent in African populations. Frequency distribution shows that this allele had to have arisen in the Middle East or Europe and followed the pattern of the peopling of the rest of the world (Figure 2 from (Bierut et al., 2008)).

It is possible that after the out of Africa migration, the rs16969968 mutation arose by chance in the population of humans living in Europe. However, given the effects of nicotine on memory and cognition, it is also possible that the mutation conferred some fitness benefit related to brain function. This would have provided a selective pressure to

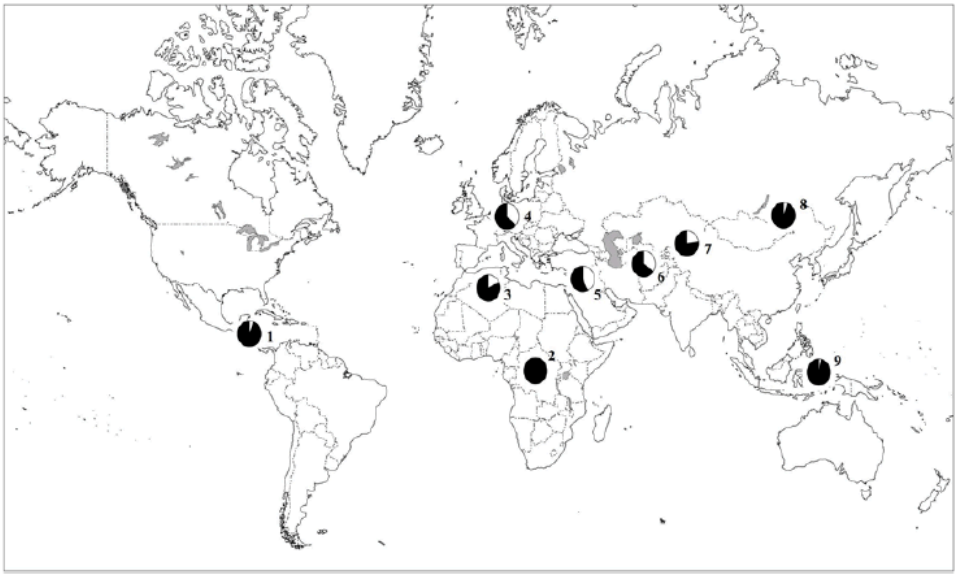

Figure 2. Allele frequency differences of rs16969968 in different ethnic populations. The frequency of the A allele is the white segment of the circles in the figure. Geographic regions: 1. America; 2. Africa; 3. North Africa; 4. Europe; 5. Middle East; 6. Central/South Asia; 7. Central/South Asia; 8. East Asia; 9. Oceania. From (Bierut et al., 2008).

enhance its frequency in Europeans. The addiction phenotype would have been an evolutionarily neutral consequence of this allele selection because the opportunity for prolonged use of purified drugs was absent since there were no highly concentrated sources of nicotine in the ancestral environment. This phenotype would only have

become apparent, and problematic, in the current environment where there is ready access to nicotine-containing products.

The hypothesis that there are variants in the nicotinic receptors that are associated with increased memory and cognition might lead some to ask the question of why these allegedly beneficial variants have not reached fixation (i.e. appear at population frequencies of nearly 100%). However, there are many reasons that a variant conferring a benefit would not reach fixation in a population. For example, the strength of selection on that allele may not be strong enough to exact such a drastic frequency shift, particularly if other variants at the locus are not deleterious but rather neutral. It could also be the case that not enough generations have passed for fixation to occur, especially if these beneficial variants are in linkage disequilibrium with other unknown deleterious polymorphisms.

The heritability of addiction is approximately 50%, yet the confirmed genetic contributions to nicotine dependence and other drug dependencies explain only a small fraction of this heritability. Current explanations for the missing variance are: 1) rare variation not tagged on current GWAS chips; 2) many genes of small effect; and 3) non-additive effects such as gene x gene or gene x environment interactions. If SNPs in the nicotinic receptors involved in nicotine addiction can be shown to be associated with differences in cognitive function, this would suggest a third explanation for missing variance-- unnoticed linkage between phenotypes due to evolutionary reasons.

CENTRAL HYPOTHESIS AND AIMS

The overall hypothesis I propose to test is whether specific genotypes, either

known or discovered by association analyses I will run, within the regions of the *CHRN* clusters on chromosomes 8 and 15 that are significantly associated with either nicotine or cocaine phenotypes, have actually been selected upon due to their effects on memory and learning. This would mean that the primary phenotype would be these more ancient and evolutionarily beneficial phenotypes, and any effect on smoking or cocaine use has hitchhiked along with these older phenotypes. The prevalence of maladaptive behaviors, namely nicotine dependence and cocaine addiction, suggests other factors may be at play.

In chapter 2, I discuss running a GWAS using imputed data from the Collaborative Genetic Study of Nicotine Dependence (COGEND) dataset and perform association testing using the phenotypes of case (nicotine dependent smoker) and control (non-dependent smoker) based on FTND scores to see whether any novel imputed SNPs within these *CHRN* clusters are associated with nicotine addiction. Key SNPs identified in these analyses of the imputed data were selected for follow-up genotyping to confirm association. These analyses reveal new associations, including evidence for more than one signal of association.

My third chapter discusses an in depth association analysis with the nicotinic cluster on chromosome 8 on which less is known about, using genotypic data from a GWAS of the Study of Addiction: Genetics and Environment (SAGE) dataset, to determine if there is any association with cocaine dependence. The basis for this is that the same SNP on chromosome 15 that is most highly associated with risk for nicotine dependence, is also protective for cocaine dependence, so it is reasonable to suspect there may in fact be an association here.

All SNPs discovered in chapters 2 and 3, as well as previously discovered

associations in the regions, are the substrate for chapter 4. In this chapter, I use three separate tests of natural selection on the 1000 Genomes dataset, which I group into European, African and Asian populations. I predict that regions of these genes that have functional importance will show evidence of selection. I also perform an association analysis in a dataset of approximately 500 individuals, between genotype at these SNPs, and performance on assessments of memory and learning. Lastly, I will discuss these results in the context of potential anthropological explanations for these results.

In summary, this work will bring about a more detailed understanding, both biologically and evolutionarily, of the variation in the nicotinic receptors on chromosomes 8 and 15, and their relationship with drug dependence, memory and learning and human sociality.

CHAPTER 2

COMMON VARIANTS NEAR GPC5 AFFECT RISK OF NICOTINE DEPENDENCE IDENTIFIED THROUGH IMPUTATION

ABSTRACT

Previous findings have demonstrated that variants in or near nicotinic receptor genes and the nicotine-metabolizing enzyme *CYP2A6* are associated with nicotine dependence. We conducted genome-wide association analyses using samples from the Collaborative Study of the Genetics of Nicotine Dependence (COGEND) to identify novel genes involved in the development of nicotine dependence. In order to increase power to detect novel associations, we imputed genotypes from the 1000 Genomes project using the IMPUTE2 software package. Variants at two loci were associated with nicotine dependence as defined by a score of ≥4 on the Fagerstrom Test of Nicotine Dependence (FTND). One group included the previously identified SNPs at 15q25 represented by rs55853698. The second locus is a novel association between nicotine dependence and SNPs at 13q31, represented by rs7995715. This association was confirmed by genotyping the top SNPs using Sequenom assays. We also attempted to replicate this finding in an independent dataset using the Study of Addiction, Genes and Environment (SAGE) dataset minus the overlapping COGEND individuals. Among the genes at this newly associated locus are a group of micro-RNAs (*MIR17HG*) and a member of the glypican gene family (*GPC5*). Notably, the *GPC5* gene is expressed mainly in the adult brain and was previously shown to be involved in the behavioral response to alcohol. These are among the first findings to implicate a non-candidate gene in risk for nicotine dependence, as well as the first time

that the known association in *CHRNA5* has been linked at the genome-wide significant level to FTND nicotine dependence.

INTRODUCTION

Tobacco use and related diseases are the number one cause of preventable deaths in the United States, accounting for approximately one out of every five deaths annually (CDC, 2008). The impact of smoking on health is very broad. More people die from lung cancer each year than from any other type of cancer (ACS, 2009). Even among those who quit smoking, there is an elevated risk of lung cancer, although this risk is less than in those who do not quit (Huang et al., 2008). Chronic obstructive pulmonary disease (COPD), a serious lung disease often caused by smoking, is also among the leading causes of death (N. L. Saccone, Culverhouse, et al., 2010).

Smoking rates are elevated in at risk populations such as those with mental illness. In 2000, it was estimated that this population consumes 44.3% of all cigarettes in the country and have approximately twice the risk of becoming a smoker than other individuals (Lasser et al., 2000). Mentally ill individuals have a harder time quitting, and can attempts to quit can cause severe depression. There is recent evidence linking smoking in this population to an underlying biological mechanism, especially schizophrenia, for which nicotine seems to normalize deficits in the *CHRNA7* nicotinic receptor (Leonard et al., 2001). Most recently, an epidemiological study of drug use in those with psychotic illness found that the odds of substance abuse are even higher, putting the odds ratio for smoking in this group at 4.6 (95% CI 4.3-4.9) (Hartz et al., 2014).

The percentage of people in the United States living with a substance abuse disorder is high: 13.5% with nicotine dependence, 13% with alcohol dependence, and 6.1% with all other drug dependencies (Grant, Hasin, Chou, Stinson, & Dawson, 2004; Regier et al., 1990). Adding to the genetic complexity of these phenotypes, persons with substance disorders are often comorbid for multiple substance dependencies as well as mental illnesses. For example, the percentage of nicotine dependence in the general population is 13%, compared with 30-70% in persons with mental illnesses (Hartz & Bierut, 2010). This necessarily means that the genetic basis of these disorders will be complex, and that there will likely be variants that contribute to risk for more than one disorder.

Susceptibility to drug use, abuse, and dependence has been shown by several studies to have a moderate to high genetic component (Bierut, 2011; J. C. Wang, Kapoor, & Goate, 2012). The heritability of nicotine dependence has been estimated to be from 44-72% (Carmelli et al., 1990; Lessov et al., 2004). Genome-wide association scans (GWAS) are a powerful tool for identifying genetic loci with a wide range of effect sizes on traits of interest. GWAS take advantage of linkage disequilibrium in a population to detect all common single nucleotide variations in an individual. If a SNP on the GWAS genotyping array is associated with a disease phenotype in a sample, it could be because that SNP is causing the disease, but more likely, that SNP is in linkage disequilibrium with the real SNP causing the disease. Linkage disequilibrium is the tendency for two or more alleles to be inherited together more often than expected by chance, due to reduced genetic recombination (or distance) between them. A GWAS compares DNA variation at millions of SNPs in individuals that do (affected) or do not (unaffected) exhibit the trait

of interest. SNP allele frequencies between affected and unaffected individuals are then compared to disease status or any other phenotypic data collected to determine if there is a significant association between genotype and phenotype. An important advantage of GWAS is that they do not require prior hypothesis about the underlying biological mechanisms of a trait. However, they do have the disadvantage of not being able to tag all variation in the genome, especially portions of the heritability of the trait that are caused by rare variants.

Several smoking related phenotypes are commonly studied, that get at different aspects of nicotine dependence. The Fagerström Test for Nicotine Dependence (FTND) is a widely accepted quantitative measure of nicotine dependence (ND). This tool has 6 questions including how many cigarettes are smoked per day (CPD), which is one of two three-point questions on the test, as well as time to first cigarette (TTF). The highest score possible is a 10 indicating maximum nicotine dependence (Heatherton et al. 1991). FTND is therefore a more comprehensive measure of ND since it encompasses CPD, as well as other aspects of ND. Nevertheless, CPD is often used by itself as a measure of nicotine dependence, and on the FTND is broken down into 4 possible answers: 0-10, 11-20, 21-30 and 30+ cigarettes per day. The pros of using CPD as the phenotype are that it is a direct measure of consumption and it is simple to collect, however the cons are that it is affected by cultural observations such as African-Americans smoking fewer cigarettes a day (Johnson, Morgan-Lopez, Breslau, Hatsukami, & Bierut, 2008). Additionally, CPD is affected by legislation banning smoking in places such as public buildings that has had a dramatic effect on smoking levels. FTND has the benefit of including CPD, as well as the other questions on the test such as item 3: which cigarette would you most hate to

give up? Rice et al. (Rice et al., 2012) describe an association between early morning smoking and the genome-wide significant bin on chromosome 8, suggesting that using just CPD would miss this aspect of nicotine dependence. However the main disadvantage of the FTND is that because it is more complicated than just how many cigarettes does one smoke per day, it is likely hard to administer with perfect inter-rater reliability. Compared to several of the measures of DSM-IV nicotine dependence, FTND seems to pick up more direct measures of addiction (CPD, TTF, desire to smoke where prohibited, smoking while ill), while the DSM-IV measures aspects related to quality of life and how smoking interferes with leading a normal life. Perhaps the reason is that the DSM-IV uses the same criteria for all drug dependencies, while the FTND is specifically for smoking.

GWAS have discovered several variants that modify susceptibility or resistance to cigarette consumption (Berrettini et al., 2008; Bierut et al., 2007; Thorgeirsson et al., 2010). The loci identified by GWAS as associated with cigarettes per day (CPD) include genes encoding neuronal cholinergic nicotinic receptors (*CHRNs*) and the nicotine metabolizing gene *CYP2A6*. Two regions, one containing the *CHRNA5, CHRNA3* and *CHRNB4* genes on chromosome 15 and one containing the *CHRNB3* and *CHRNA6* genes on chromosome 8 are the most replicated loci associated with nicotine dependence (N. L. Saccone, Culverhouse, et al., 2010; N. L. Saccone et al., 2009; S. F. Saccone et al., 2007; Spitz et al., 2008; Thorgeirsson et al., 2008; Thorgeirsson et al., 2010; Weiss et al., 2008). The strongest association ($p = 5.96 \times 10^{-31}$) between nicotinic receptors and nicotine addiction is a non-synonymous change (rs16969968, D398N) in the gene encoding the α5 subunit of the nicotinic receptor, *CHRNA5* on chromosome 15 (Bierut et al., 2008; N. L.

Saccone et al., 2009; S. F. Saccone et al., 2007; Spitz et al., 2008; Thorgeirsson et al., 2008; Weiss et al., 2008). The phenotype previously found to be associated with this variant at the genome-wide significant level is cigarettes per day (CPD). The minor allele of this SNP has a frequency of ~30% in European-Americans but is rare in other populations including African-Americans. When nicotinic receptors containing the minor allele form of this SNP (398N), are overexpressed in cultured cells, agonists induce less channel opening and cell activation than in cells that express receptors containing the major allele (398D). Thus, this SNP results in a significant functional change in the behavior of this ion channel. Additionally, variants in *CHRNA5* have been shown to alter the levels of expression of *CHRNA5* mRNA, (J. C. Wang et al., 2009) and influence risk for nicotine dependence (N. L. Saccone, Culverhouse, et al., 2010). Associations with cigarette consumption have also been found with SNPs in or near the nicotinic receptors on chromosome 8 (Hoft et al., 2009; Johnson et al., 2010; Rice et al., 2012; N. L. Saccone, Culverhouse, et al., 2010; S. F. Saccone et al., 2007; Thorgeirsson et al., 2010; Zeiger et al., 2008), but the mechanism underlying the association on chromosome 8 is not well understood. Lastly, the candidate gene *CYP2A6*, which plays a role in nicotine metabolism, has been associated at the genome-wide significant level with both CPD (Thorgeirsson et al., 2010) and COPD (Cho et al., 2012). However, no genome-wide significant associations with nicotine dependence related phenotypes have been found outside of these genes. So far, the only association with FTND nicotine dependence has been the genome-wide significant association with rs1451240 near *CHRNB3* described by Rice et al. (Rice et al., 2012). They report that FTND better captured the association on chromosome 8 for ND than CPD and that this held true even when comparing their

sample of 4200 subjects with a meta-analysis of over 75,000 subjects. They attribute part of this improvement with a different phenotype to the fact that African-Americans smoke fewer cigarettes per day than do European-Americans, so CPD would not fully capture ND in this group. Their study also highlights the importance of precise phenotypes like FTND, which has previously been shown to be more consistent across populations than CPD (Johnson et al., 2008). Large-scale meta-analyses might miss associations since increasing sample size does not always increase power when the phenotypes are either imprecise or non-specific.

To date, no GWAS have identified non-candidate gene loci associated with nicotine dependence phenotypes. We report here for the first time, a genome-wide significant locus on chromosome 13q31 near the glypican 5 *(GPC5)* gene, associated with FTND case control status. The second important discovery in this study is the association of the known risk variant in *CHRNA5* with FTND for the first time. Given the previous discussion of the benefits and drawbacks of the various nicotine-related phenotypes, it is possible that the mechanism underlying the association near *GPC5* is one that is better captured by FTND than CPD, and that the increased power in the previous association in *CHRNA5* supports this hypothesis.

METHODS

COGEND Dataset. Subjects in this study have self-identified their ethnicity as either African-American or European-American. Cases and controls were ascertained on the basis of their FTND score (maximum 10 points), where cases had scores of four and above and controls had scores of less than 4. Controls also had to have smoked at least

100 cigarettes in their lifetime and yet not be dependent. Subjects were all recruited from St. Louis and Detroit metropolitan areas. 2646 blood samples were collected for study purposes. The sample consists of 1936 European American individuals (995 nicotine dependent (ND) cases and 941 smokers with no symptoms of dependence (controls)) as well as 710 African Americans (461 ND cases and 249 controls). All data was collected in accordance with the protocols approved by the Institutional Review Board at each institution involved. Informed consent was obtained from all subjects for their DNA to be used in genetic studies and their phenotypic information to be shared with appropriate investigators.

Phenotypes. Subjects were placed in either the case or control category, based on their FTND score. Cases were nicotine dependent individuals who scored four or higher on the FTND, and controls were those who had smoked at least 100 cigarettes in their lifetime, but had not become dependent, and whose FTND scores were less than 4. We also used for comparison the ordinal traits of cigarettes per day (CPD) and time to first cigarette in the morning (TTF). CPD has 4 categories; 0-10, 10-20, 20-30, and 31+. Time to first cigarette has 4 as well; 0-5 minutes, 6-30 minutes, 31-60 minutes, and 61+ minutes.

Genotyping and Quality Control. The GWAS SNP dataset was obtained partially on the Illumina Human 1M-Duo beadchip genotyping array and partially on the 2.5 SNP genotyping array. Those samples that were genotyped on the Illumina Human 1M-Duo beadchip were typed by the Center for Inherited Disease Research (CIDR) at Johns Hopkins University. Of the 1,049,008 SNP assay probes on the chip, 948,658 passed the quality control process, which included investigation of hidden relatedness

(IBD>10%), HapMap controls, batch effects, gender/chromosomal abnormalities, HWE $=10x^{-4}$, genotyping call rate of 97% and Mendelian error or duplication error detection. Additionally, there are 23,812 intensity only probes on the chip (P. Lin et al., 2011). Individuals typed on the Human Omni 2.5M chip, which has 2,379,514 probes, were also typed by CIDR. All data was cleaned by GENEVA. These processes are described in more detail in the data cleaning report freely available on the GENEVA website at (http://www.genevastudy.org/docs/GENEVA_Alcohol_QC_report_8Oct2008.pdf). Follow-up genotyping of top novel association results was done using the Sequenom Mass Array genotyping platform (J. C. Wang et al., 2009). All SNPs with an r^2 >0.8 with rs7995715, and any SNP with an r^2 >0.8 with the first group of SNPs were chosen for follow up genotyping to maximize coverage of potentially interesting LD bins.

Imputation Quality Control. In order to eliminate unnecessary risk of spurious associations that would result from the use of two different platforms, only the SNPs present on both genotyping platforms were considered, which consisted of 605,735 SNPs. This was a particularly important issue because in this case, most of the controls were typed on one chip and most of the cases were typed on the other. Additionally, a minor allele frequency cutoff of 5%, HWE=$10x^{-4}$, a chromosomal missingness cutoff of 5% and genotyping call rate of 97% were employed. Relatedness was evaluated for 3rd degree or higher, based on IBD>10%. We identified a total of 23 2nd-3rd degree relative pairs, involving 12 EAs and 26 AAs. We proceeded without excluding any subjects due to this lesser degree of relatedness. No 1st degree relative pairs were identified. The subjects were then imputed to the 1000 Genomes cosmopolitan reference panel using IMPUTE2.

Imputation Software. Haplotypes were pre-phased using SHAPE-IT (Delaneau, Marchini, & Zagury, 2012), and imputation was performed using the program IMPUTE2 (B. Howie, Marchini, & Stephens, 2011; B. N. Howie, Donnelly, & Marchini, 2009), using all available reference genomes from the 1000 Genomes Project (Durbin, Altshuler, Abecasis, Bentley, & Consortium, 2010). SHAPE-IT allows computational time to be minimized by using multi-threaded phasing. IMPUTE2 was chosen as the imputation program based on studies that have shown it to perform better for situations like ours in comparison to other programs available (Chanda et al., 2012; Ellinghaus, Schreiber, Franke, & Nothnagel, 2009).

Population Stratification. Subjects in the study had self-identified as European-American or African-American. Since admixture across these populations needs to be accounted for, we used the program EIGENSTRAT (Price et al., 2006) to calculate principal components. The first two principal components were included in association analyses to control for population stratification.

GWAS Association Analysis. Genotype/phenotype association analyses were performed using PLINK, with the imputed genotype probabilities data from IMPUTE2 using SNPs genotyped on both platforms and SNPs imputed from those sites. FTND diagnosis of nicotine dependence was the primary phenotype. Time to first cigarette and maximum lifetime cigarettes per day were analyzed in order to dissect the effect of associated SNPs on the FTND phenotype. Logistic regression was run for the dichotomous trait of FTND, and linear regression for continuous traits of TTF and CPD. The covariates used were age at interview, sex, PC1, and PC2 for the combined sample.

RESULTS

Imputed Data. In order to identify novel loci associated with FTND nicotine dependence, we performed a genome-wide association study in individuals from the Collaborative Study of the Genetics of Nicotine Dependence (COGEND). To increase power to detect novel associations, we performed imputation for all samples using phased haplotype data from the 1000 Genomes project as the reference. After filtering SNPs with minor allele frequency < 5% and genotyping or imputation call rate < 97%, we analyzed association data for ~2,500,000 imputed and genotyped autosomal SNPs using logistic regression with age at interview, sex, PC1 and PC2 as covariates.

Mega-analysis. We find a genome-wide significant association in the mega-analysis of the combined sample (N=1952 EA, 709 AA) with SNPs at 15q25, represented by rs114205691 (OR = 1.41, 95% CI= 1.29-1.53, MAF=0.31, P = 4.35 x 10^{-8}) and a novel genome-wide significant association at 13q31 in the combined sample represented by rs7995715 (OR=1.41, 95% CI= 1.29-1.53, MAF=0.33, P = 3.27 x 10^{-8}), with FTND nicotine dependence (Figure 3).

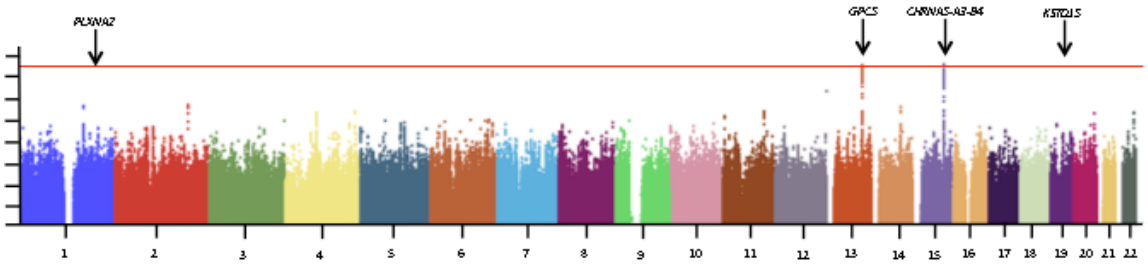

Figure 3. Loci for susceptibility to FTND nicotine dependence detected by GWAS for the combined sample (EAs and AAs). Manhattan plot of association test results of GWAS data showing the chromosomal position of genotyped or imputed SNPs plotted against −log10P. The red horizontal line represents the threshold for genome-wide significance ($P < 5 \times 10^{-8}$).

A regional association plot is shown for the results of the association testing near *GPC5* (Figure 4) and near *CHRNA5-A3-B4* (Figure 5), and a table with the top 50 SNPs from the genome-wide association analysis is shown in Appendix A. This association with the bin tagged by rs7995715 is also significant with the secondary phenotype of CPD ($p=1.40 \times 10^{-4}$ $\beta=0.12$) and genome-wide significant with the secondary phenotype of TTF ($p=4.72 \times 10^{-8}$, $\beta= 0.21$). Overall, there was no inflation in the association results as demonstrated in the QQ-plot of the combined association results ($\lambda=1.02$, SE= 2.38×10^{-6}) (Figure 6).

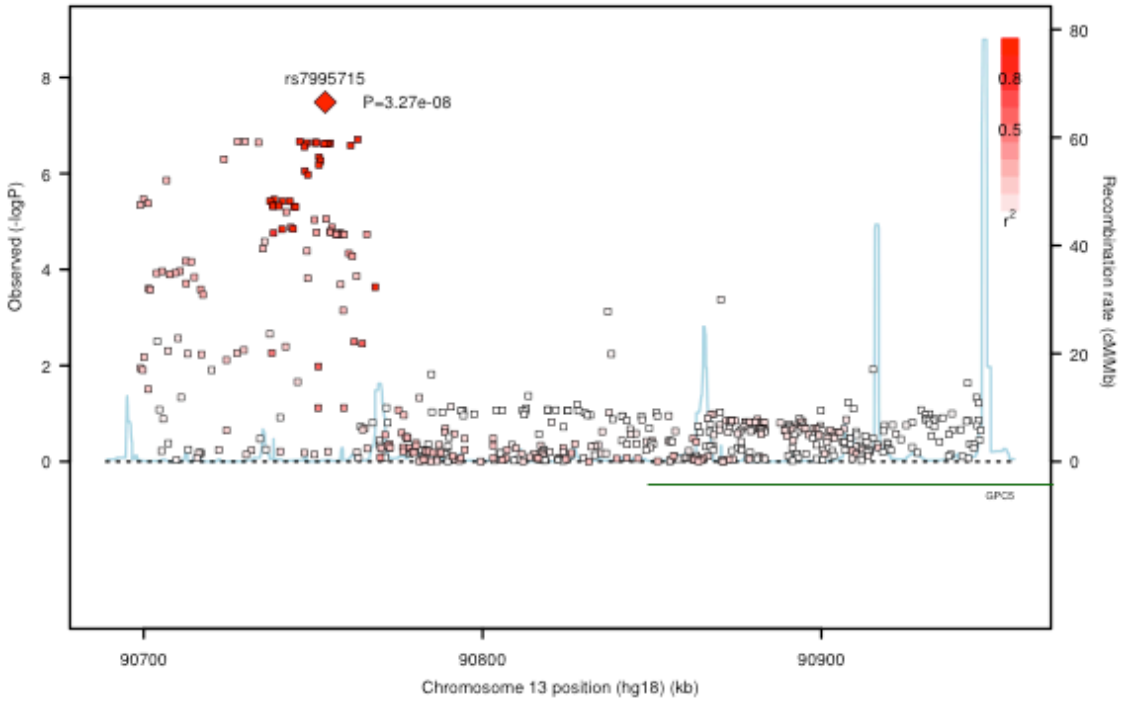

Figure 4. Regional association plot for the area on chromosome 13 surrounding *GPC5*

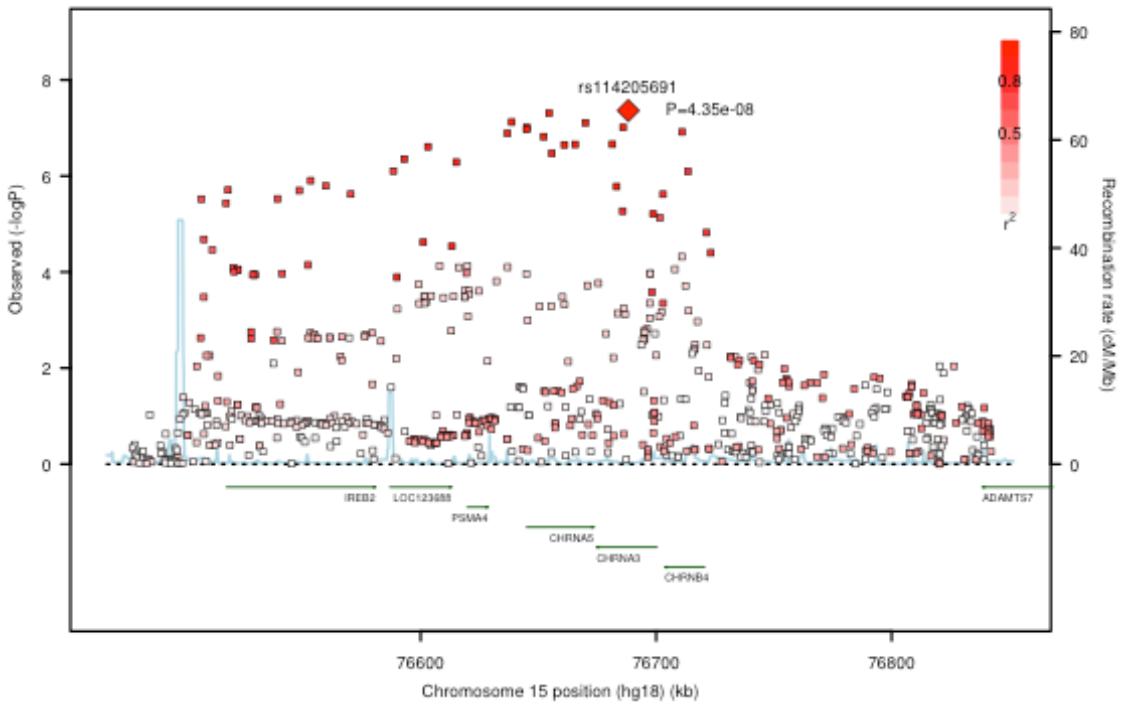

Figure 5. Regional association plot for the area on chromosome 15 surrounding *CHRNA5-A3-B4*.

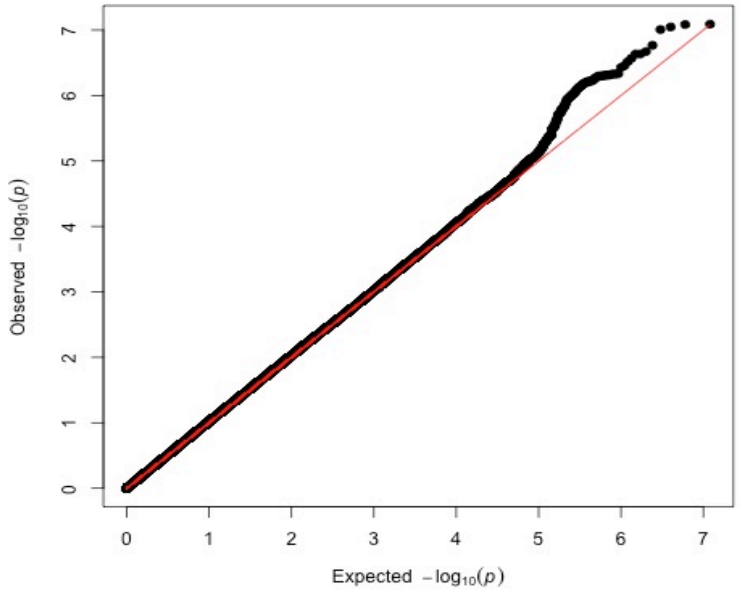

Figure 6. QQ-Plot of Combined EA and AA Whole Genome Association Results. (λ=1.02, SE= 2.38×10^{-6})

Of interest is that 4 genes other than the cluster of nicotinic receptors on chromosome 15 and the *GPC5* region on chromosome 13 had SNPs that made it into the top 50 most significant SNPs in the genome-wide association wit FTND nicotine dependence (Appendix A). Perhaps the most interesting is neurotrimin *(NTM)* on chromosome 11. This gene encodes a protein that promotes neurite outgrowth and adhesion, and is closely linked to opioid binding protein/cell adhesion molecule-like *(OPCML)*. Polymorphisms in this gene have even been found to influence intelligence in the COGA dataset (Pan, Wang, & Aragam, 2011).

This confirms the previous association of rs16969968 at 15q25 (r^2=0.9 with rs55853698 in HapMap CEU samples) and nicotine related behavior. However, this is the first time that this SNP been associated with FTND nicotine dependence at a genome-wide significant level. Previous studies have demonstrated genome-wide significant associations between this SNP and the number of cigarettes smoked per day (CPD), only one component of the FTND score.

Meta-analysis. We performed a meta-analysis of the European-American and African-American samples and compared the results to the mega-analysis (Appendix A). The results were nearly identical to those of the mega-analysis. The LD bin on chromosome 15 tagged by rs114205691 remained genome-wide significant ($p=4.14 \times 10^{-8}$, $\beta=0.71$), and the LD bin on chromosome 13 tagged by rs7995715 was nearly genome-wide significant ($p=9.20 \times 10^{-8}$, $\beta=0.72$).

European-Americans. Genome-wide association results for European-Americans with FTND nicotine dependence still pick up the signals on chromosomes 13 and 15, although not as strongly as in the combined sample (Figure 7).

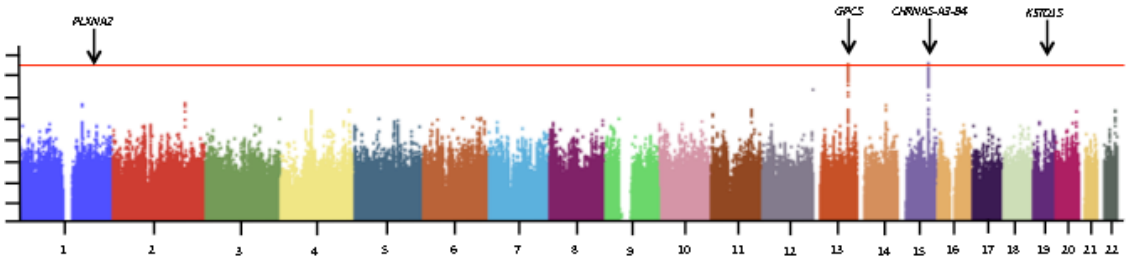

Figure 7. Loci for susceptibility to FTND nicotine dependence detected by GWAS among European Americans. Manhattan plot of association test results of GWAS data showing the chromosomal position of genotyped or imputed SNPs plotted against −log10P. The red horizontal line represents the threshold for genome-wide significance ($p < 5 \times 10^{-8}$).

The top 50 SNPs genome-wide are included in Table 2. Within EAs separately, the strongest association with FTND nicotine dependence in the *GPC5* region of chromosome 13 is the bin tagged by rs7995715 ($p=2.01x10^{-6}$, OR=1.43, 95% CI= 1.27-1.59). However, a locus on chromosome 19 containing the gene *KTD15* has three SNPs with more significant p-values than *GPC5*. *KTD15*, which is expressed in the central nervous system, has been associated with obesity in a meta-analysis (Willer et al. 2008). However to date it does not appear to be associated with any psychiatric conditions.

With the secondary phenotype of CPD, rs7995715 is still highly associated ($p=3.38 \times 10^{-5}$, $\beta=0.17$), similarly for TTF ($p=1.85 \times 10^{-6}$, $\beta=0.22$). The association on chromosome 15 remains strong as well. The bin tagged by rs114205691 is still the top hit in the association analysis with FTND nicotine dependence ($p=1.32x10^{-6}$, OR=1.40, 95% CI=1.26-1.54). Covariates used in all cases were age, sex, PC1 and PC2.

Table 2

Top 50 SNPs in the GWAS in the European-American sample

Genome-wide Top 50 SNPs in European-Americans from FTND Nicotine Dependence Association Analysis										
CHR	Position	Gene	rs number	A1	A2	Freq	INFO	OR	SE	P-Value
19	34341157	*KCTD15*	rs10404267	C	T	0.9684	0.9493	0.3442	0.2168	8.71E-07
19	34342004	*KCTD15*	rs10424551	C	T	0.9683	0.9482	0.3448	0.2166	8.83E-07
19	34343918	*KCTD15*	rs10413064	C	A	0.9682	0.9463	0.3464	0.216	9.13E-07
13	91953042	*GPC5*	rs7994634	C	T	0.7418	1.0408	0.6943	0.0753	1.27E-06
19	34349137	*KCTD15*	rs73926943	C	T	0.9675	0.9407	0.3564	0.2132	1.30E-06
13	91952459	*GPC5*	rs9301726	T	C	0.7417	1.0403	0.6945	0.0753	1.30E-06
15	78901113	*CHRNA3-A5-B4*	rs114205691	C	T	0.6456	1.0039	0.7139	0.0697	1.32E-06
15	78898932	*CHRNA3-A5-B4*	rs55676755	C	G	0.6468	1.0033	0.7141	0.0698	1.39E-06
13	91953719	*GPC5*	rs1332216	T	C	0.7421	1.0418	0.6966	0.0753	1.56E-06
13	91954199	*GPC5*	rs10161911	C	T	0.7413	1.0394	0.6967	0.0753	1.57E-06
13	91955192	*GPC5*	rs9523299	G	A	0.7423	1.0436	0.697	0.0752	1.59E-06
15	78906177	*CHRNA3-A5-B4*	rs146009840	A	T	0.6472	1.0026	0.7153	0.0698	1.59E-06
19	34340136	*KCTD15*	rs10421416	G	T	0.9691	0.9538	0.3529	0.2175	1.68E-06
19	34340137	*KCTD15*	rs10423005	C	T	0.9691	0.9537	0.3529	0.2175	1.68E-06
13	91957035	*GPC5*	rs7335045	A	G	0.7423	1.0422	0.6975	0.0753	1.69E-06
13	91956188	*GPC5*	rs9589183	C	T	0.7423	1.0424	0.6976	0.0752	1.71E-06
13	91956038	*GPC5*	rs7989842	G	T	0.7423	1.0425	0.6977	0.0752	1.71E-06
13	91953721	*GPC5*	rs1332217	T	G	0.7417	1.0406	0.6976	0.0753	1.72E-06
13	91965134	*GPC5*	rs9515908	C	T	0.7501	0.9802	0.6872	0.0784	1.73E-06
13	91955287	*GPC5*	chr13:91955287:I	A	AT	0.7423	1.0426	0.6978	0.0752	1.73E-06
15	78867482	*CHRNA3-A5-B4*	rs17486278	A	C	0.6465	1.0055	0.7171	0.0697	1.80E-06
13	91949562	*GPC5*	rs9515905	A	G	0.7417	1.039	0.698	0.0753	1.82E-06
13	91949444	*GPC5*	rs9523296	G	A	0.7417	1.0389	0.6981	0.0753	1.83E-06
13	91950114	*GPC5*	rs7986895	C	A	0.7416	1.0385	0.6982	0.0753	1.86E-06
15	78882925	*CHRNA3-A5-B4*	rs16969968	G	A	0.6482	1.0079	0.7175	0.0696	1.86E-06
13	91963080	*GPC5*	rs1475655	A	T	0.7503	0.9823	0.6884	0.0783	1.88E-06
13	91952853	*GPC5*	chr13:91952853:D	CA	C	0.7423	1.0423	0.6987	0.0752	1.89E-06
13	91940278	*GPC5*	rs28620036	C	T	0.76	0.9996	0.6875	0.0788	1.95E-06
13	91948047	*GPC5*	rs9523295	G	A	0.7416	1.038	0.6989	0.0753	1.98E-06
13	91955562	*GPC5*	rs7995715	T	G	0.7399	1.0426	0.7002	0.075	2.01E-06
13	91950403	*GPC5*	rs9515906	G	C	0.7479	1.0313	0.6962	0.0762	2.03E-06
13	91946788	*GPC5*	rs12708388	C	T	0.7415	1.0372	0.6992	0.0753	2.05E-06
13	91946343	*GPC5*	rs12867738	G	A	0.7415	1.0368	0.6994	0.0754	2.08E-06
13	91946092	*GPC5*	rs9515904	G	T	0.7414	1.0366	0.6994	0.0754	2.10E-06
13	91945089	*GPC5*	rs9523293	C	T	0.7414	1.0359	0.6996	0.0754	2.14E-06
13	91942919	*GPC5*	rs34165267	C	T	0.7414	1.0357	0.6996	0.0754	2.15E-06
13	91942808	*GPC5*	rs35921784	T	C	0.7414	1.0357	0.6996	0.0754	2.16E-06
13	91941936	*GPC5*	rs9583908	T	C	0.7414	1.0357	0.6996	0.0754	2.16E-06
13	91940083	*GPC5*	rs7984992	T	C	0.7414	1.0356	0.6997	0.0754	2.17E-06
13	91940484	*GPC5*	rs7994469	G	A	0.7414	1.0356	0.6997	0.0754	2.17E-06
15	78886947	*CHRNA3-A5-B4*	rs4887067	G	A	0.6479	1.0069	0.719	0.0697	2.17E-06
13	91940169	*GPC5*	rs7985179	T	A	0.7414	1.0356	0.6997	0.0754	2.17E-06
15	78886198	*CHRNA3-A5-B4*	rs8192482	C	T	0.6479	1.0069	0.719	0.0697	2.17E-06
13	91939270	*GPC5*	rs9583907	C	T	0.7414	1.0355	0.6997	0.0754	2.17E-06
15	78894339	*CHRNA3-A5-B4*	rs1051730	G	A	0.6469	1.0038	0.7189	0.0697	2.20E-06
15	78868636	*CHRNA3-A5-B4*	rs72740964	G	A	0.6485	1.006	0.7191	0.0697	2.24E-06
15	78922638	*CHRNA3-A5-B4*	rs2869548	G	A	0.617	0.9761	0.7202	0.0694	2.27E-06
15	78849034	*CHRNA3-A5-B4*	rs58365910	T	C	0.6394	1.0026	0.7202	0.0694	2.28E-06
15	78857939	*CHRNA3-A5-B4*	rs55853698	T	G	0.6436	0.9999	0.7193	0.0697	2.29E-06
15	78862453	*CHRNA3-A5-B4*	rs7172118	C	A	0.648	1.0026	0.7192	0.0698	2.35E-06

African-Americans. Within AAs separately, the strongest association with FTND nicotine dependence in the *GPC5* region of chromosome 13 is not the bin tagged by rs7995715, although this SNP is still marginally significant (p= 0.01, OR=1.33, 95% CI=1.09-1.57). The gene that dominates the top fifty genome-wide significant results with FTND nicotine dependence is *PLXNA2* (Table 3 and Figure 8 for Manhattan plot). This gene is a semaphorin co-receptor that is expressed during nervous system development, and has been found at elevated levels in patients with schizophrenia (Eastwood, Law, Everall, & Harrison, 2003). A GWAS of schizophrenia resulted in several SNPs in this gene associated with disease status (Mah et al., 2006). One group found an association between a SNP in this gene and anxiety, depression, neuroticism and psychological distress (Wray et al., 2007). It would make sense that this gene is also associated with addiction phenotypes, since addiction is highly comorbid with other psychiatric disorders. Interestingly, the GWAS study was in European-Americans, yet we find association with variants in *PLXNA2* and FTND nicotine dependence predominantly in African- Americans.

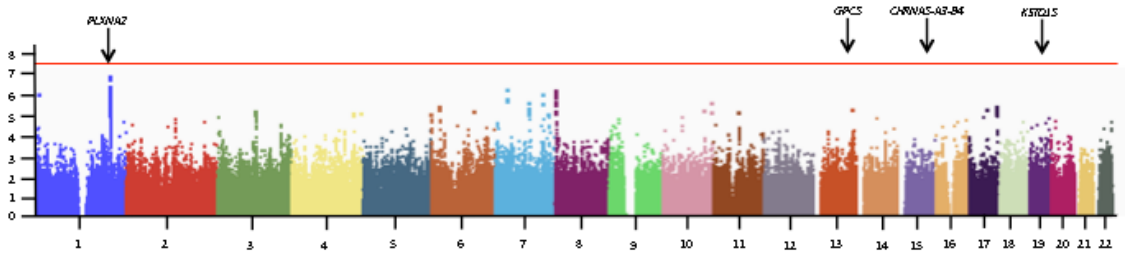

Figure 8. Loci for susceptibility to FTND nicotine dependence detected by GWAS among African Americans. Manhattan plot of association test results of GWAS data showing the chromosomal position of genotyped or imputed SNPs plotted against −log10P. The red horizontal line represents the threshold for genome-wide significance ($p < 5 \times 10^{-8}$).

Table 3

Top 50 SNPs in the GWAS in the African-American sample

Genome-wide Top 50 SNPs in African-Americans from FTND Nicotine Dependence Association Analysis										
CHR	Position	Gene	rs number	A1	A2	Freq	INFO	OR	SE	P-Value
1	208909872	*PLXNA2*	rs11119123	G	T	0.6294	0.9785	0.5157	0.1269	1.80E-07
1	208904565	*PLXNA2*	chr1:208904565:D	GATAA	G	0.6227	0.9705	0.5188	0.1267	2.23E-07
1	208905388	*PLXNA2*	rs73090281	A	G	0.623	0.9707	0.519	0.1267	2.27E-07
1	208905914	*PLXNA2*	rs75097667	C	T	0.6232	0.971	0.5192	0.1267	2.31E-07
1	208906343	*PLXNA2*	rs2297940	G	A	0.6233	0.9712	0.5194	0.1267	2.33E-07
1	208907775	*PLXNA2*	rs11119122	C	T	0.6211	0.9646	0.5274	0.1266	4.31E-07
1	208913341	*PLXNA2*	rs12063346	T	C	0.6341	0.9668	0.5279	0.1277	5.61E-07
7	33543550	*BBS9*	rs2392241	G	A	0.1266	0.9966	0.4246	0.1712	5.67E-07
8	5443069	*CSMD1*	rs11776337	G	A	0.8926	0.991	2.5085	0.1843	6.03E-07
1	208874955	*PLXNA2*	chr1:208874955:D	CAGCT	C	0.5573	0.9289	1.8505	0.1233	6.04E-07
8	5443259	*CSMD1*	rs11787025	A	T	0.8888	0.9626	2.5042	0.1841	6.18E-07
8	5444800	*CSMD1*	rs7815374	A	G	0.892	0.9877	2.5013	0.1842	6.47E-07
1	208881060	*PLXNA2*	rs189457972	G	A	0.6126	0.8835	0.5174	0.1327	6.82E-07
1	208912269	*PLXNA2*	rs61434781	A	T	0.6239	0.9703	0.5344	0.1262	6.94E-07
1	208882605	*PLXNA2*	rs73088247	A	G	0.5662	0.9653	0.542	0.1234	6.98E-07
1	208885796	*PLXNA2*	rs6540486	G	A	0.5956	0.9476	0.5345	0.1265	7.43E-07
1	208903281	*PLXNA2*	rs2275912	G	T	0.6372	0.9626	0.5305	0.1284	7.88E-07
8	5443046	*CSMD1*	rs11774009	T	C	0.8919	0.9892	2.4739	0.1838	8.27E-07
1	17552875	*PLXNA2*	rs2977290	A	G	0.34	0.9413	0.5387	0.1256	8.42E-07
1	208897416	*PLXNA2*	rs1166879	C	T	0.5334	0.9998	1.7847	0.1177	8.51E-07
8	5441318	*CSMD1*	rs2408064	A	G	0.8928	0.9907	2.4771	0.1843	8.61E-07
1	208912736	*PLXNA2*	rs12062092	T	C	0.6318	0.9463	0.5311	0.1286	8.62E-07
1	208907024	*PLXNA2*	rs6669474	C	A	0.6176	0.9683	0.5389	0.1257	8.71E-07
1	208906804	*PLXNA2*	rs2297941	A	G	0.6175	0.9681	0.5389	0.1257	8.73E-07
1	208905917	*PLXNA2*	rs1166882	C	T	0.5267	0.9734	1.796	0.1192	8.93E-07
7	129555243	*UBE2H*	rs141050514	G	A	0.9502	0.6538	4.9996	0.3276	8.98E-07
1	208888351	*PLXNA2*	rs1770207	C	T	0.578	0.9546	1.8191	0.1218	9.04E-07
1	208881174	*PLXNA2*	rs138249513	G	A	0.5846	0.9362	0.5387	0.1266	1.02E-06
1	208885801	*PLXNA2*	rs6540487	T	C	0.5851	0.9444	0.5409	0.126	1.08E-06
1	208876405	*PLXNA2*	rs12081558	T	C	0.5648	0.9617	0.5471	0.1237	1.09E-06
8	5444781	*CSMD1*	rs7831116	G	A	0.8931	0.9899	2.4529	0.1846	1.17E-06
1	208880999	*PLXNA2*	rs112296483	A	G	0.6159	0.8745	0.5237	0.1331	1.18E-06
1	208879549	*PLXNA2*	chr1:208879549:D	CAA	C	0.5838	0.9391	0.5423	0.1263	1.26E-06
7	33549368	*BBS9*	rs6972695	G	C	0.1327	1.0127	0.4459	0.1668	1.28E-06
1	208881463	*PLXNA2*	rs148752960	T	C	0.5744	0.9477	0.5484	0.1247	1.44E-06
1	208877502	*PLXNA2*	rs144386006	C	T	0.5991	0.9101	0.5375	0.1292	1.55E-06
2	131637355	*AK127124*	chr2:131637355:I	T	TA	0.9468	0.8888	3.6251	0.2684	1.60E-06
7	33546619	*BBS9*	rs2392243	G	A	0.133	1.0193	0.451	0.166	1.60E-06
8	5445144	*CSMD1*	rs67617814	G	C	0.8939	0.991	2.4253	0.1849	1.66E-06
7	90674192	*CDK14*	rs962281	C	T	0.8435	1.0193	2.0918	0.1546	1.79E-06
10	133808669	*BNIP3*	rs11146478	T	G	0.6216	0.9022	0.5351	0.1311	1.83E-06
8	5454722	*CSMD1*	rs4875594	C	T	0.8071	0.8839	2.0981	0.1556	1.90E-06
7	129543199	*UBE2H*	rs17559441	C	G	0.9554	0.69	4.9778	0.3373	1.95E-06
8	5435605	*CSMD1*	rs4875588	A	G	0.9109	0.9721	2.6245	0.2034	2.10E-06
8	5441053	*CSMD1*	rs4875593	G	A	0.9017	0.9925	2.4725	0.1913	2.22E-06
1	208886608	*PLXNA2*	rs73090226	A	G	0.5772	0.9546	0.555	0.1245	2.25E-06
1	208887322	*PLXNA2*	rs11811442	C	T	0.5793	0.9571	0.5558	0.1246	2.43E-06
1	208886470	*PLXNA2*	rs17013108	A	G	0.5794	0.9575	0.556	0.1246	2.45E-06
1	208886512	*PLXNA2*	chr1:208886512:D	CT	C	0.5794	0.9575	0.556	0.1246	2.45E-06
1	208886536	*PLXNA2*	chr1:208886536:I	A	AT	0.5794	0.9575	0.556	0.1246	2.45E-06

However, with the phenotype of CPD, rs7995715 in AAs alone is not at all significant ($p=0.95$, $\beta=0.00$). This is interesting because this could be why the signal on chromosome 8 was not picked up in a large meta-analysis using CPD, but was picked up as GWS using FTND nicotine dependence as the phenotype (Rice et al., 2012). Lastly with TTF, rs7995715 is again more significant ($p=0.02$, $\beta=0.17$). The association on chromosome 15 with FTND nicotine dependence is high although not as strong as in European-Americans ($p=8.83x10^{-3}$, OR=1.5, 95% CI= 1.2-1.8 in AAs). Again, covariates used in all cases were age, sex, PC1 and PC2. For full regional results for the combined sample as well as separate samples, refer to Table 4 for FTND nicotine dependence near *GPC5,* Table 5 for CPD near *GPC5,* Table 6 for TTF near *GPC5,* and Table 7 for FTND nicotine dependence near *CHRNA5-A3-B4*.

Table 4

Top 30 SNPs in the GPC5 region associated with FTND nicotine dependence

			FTND-Combined Sample				***FTND - EAs Only***				***FTND - AAs Only***			
SNP	*A1*	*A2*	*Freq*	*OR*	*SE*	*P-Value*	*Freq*	*OR*	*SE*	*P-Value*	*Freq*	*OR*	*SE*	*P-Value*
rs7995715	T	G	0.68	0.71	0.06	3.27E-08	0.74	0.70	0.08	2.01E-06	0.50	0.75	0.12	1.27E-02
rs9515908	C	T	0.71	0.71	0.07	1.96E-07	0.75	0.69	0.08	1.73E-06	0.59	0.79	0.12	4.65E-02
rs9523295	G	A	0.70	0.72	0.06	2.15E-07	0.74	0.70	0.08	1.98E-06	0.60	0.80	0.12	4.74E-02
rs9523288	C	T	0.82	0.67	0.08	2.15E-07	0.86	0.69	0.10	7.92E-05	0.69	0.67	0.13	2.51E-03
rs67147421	A	C	0.82	0.67	0.08	2.17E-07	0.86	0.69	0.10	1.24E-04	0.70	0.66	0.13	1.54E-03
rs9523289	G	A	0.81	0.67	0.08	2.26E-07	0.86	0.68	0.10	6.10E-05	0.69	0.68	0.13	3.51E-03
rs9301726	T	C	0.70	0.72	0.06	2.28E-07	0.74	0.69	0.08	1.30E-06	0.59	0.81	0.11	6.80E-02
rs7994634	C	T	0.70	0.72	0.06	2.28E-07	0.74	0.69	0.08	1.27E-06	0.59	0.81	0.11	6.91E-02
rs7986895	C	A	0.70	0.72	0.06	2.31E-07	0.74	0.70	0.08	1.86E-06	0.59	0.80	0.11	5.31E-02
rs7335045	A	G	0.70	0.72	0.06	2.35E-07	0.74	0.70	0.08	1.69E-06	0.59	0.81	0.11	5.93E-02
rs9589183	C	T	0.70	0.72	0.06	2.38E-07	0.74	0.70	0.08	1.71E-06	0.59	0.81	0.11	5.93E-02
rs7989842	G	T	0.70	0.72	0.06	2.38E-07	0.74	0.70	0.08	1.71E-06	0.59	0.81	0.11	5.93E-02
rs9523299	G	A	0.70	0.72	0.06	2.42E-07	0.74	0.70	0.08	1.59E-06	0.59	0.81	0.11	6.00E-02
rs1475655	A	T	0.71	0.72	0.06	2.60E-07	0.75	0.69	0.08	1.88E-06	0.60	0.80	0.12	5.68E-02
rs9523296	G	A	0.70	0.72	0.06	2.77E-07	0.74	0.70	0.08	1.83E-06	0.59	0.81	0.11	6.24E-02
chr13: 91955287:I	A	AT	0.70	0.73	0.06	3.05E-07	0.74	0.70	0.08	1.73E-06	0.59	0.81	0.11	6.93E-02
chr13: 91952853:D	CA	C	0.70	0.73	0.06	4.15E-07	0.74	0.70	0.08	1.89E-06	0.59	0.82	0.11	8.36E-02
rs1332216	T	C	0.67	0.73	0.06	4.57E-07	0.74	0.70	0.08	1.56E-06	0.49	0.83	0.11	1.11E-01
rs73599638	G	A	0.82	0.68	0.08	5.06E-07	0.86	0.70	0.10	1.61E-04	0.69	0.68	0.13	3.03E-03
rs10161911	C	T	0.69	0.73	0.06	5.41E-07	0.74	0.70	0.08	1.57E-06	0.54	0.84	0.11	1.18E-01
rs1332217	T	G	0.68	0.73	0.06	6.63E-07	0.74	0.70	0.08	1.72E-06	0.50	0.84	0.11	1.29E-01
rs9515905	A	G	0.69	0.74	0.06	8.87E-07	0.74	0.70	0.08	1.82E-06	0.57	0.85	0.11	1.50E-01
rs9515906	G	C	0.69	0.74	0.06	1.06E-06	0.75	0.70	0.08	2.03E-06	0.53	0.85	0.11	1.52E-01
rs68126334	C	T	0.82	0.69	0.08	1.39E-06	0.86	0.69	0.10	1.08E-04	0.69	0.72	0.13	1.25E-02
rs59920274	C	G	0.80	0.70	0.08	2.36E-06	0.86	0.69	0.10	1.21E-04	0.64	0.74	0.12	1.46E-02
rs9583905	T	A	0.81	0.70	0.08	3.43E-06	0.86	0.69	0.10	1.16E-04	0.67	0.75	0.13	2.71E-02
rs7994469	G	A	0.72	0.75	0.06	3.50E-06	0.74	0.70	0.08	2.17E-06	0.65	0.89	0.12	3.19E-01
rs9583907	C	T	0.72	0.75	0.06	3.76E-06	0.74	0.70	0.08	2.17E-06	0.65	0.89	0.12	3.24E-01
rs9523293	C	T	0.72	0.75	0.06	3.77E-06	0.74	0.70	0.08	2.14E-06	0.65	0.89	0.12	3.27E-01
rs34165267	C	T	0.72	0.75	0.06	3.79E-06	0.74	0.70	0.08	2.15E-06	0.65	0.89	0.12	3.27E-01

Data shown for the combined sample and EAs/AAs alone. Table sorted by p-value in the combined sample

Table 5

Top 30 SNPs in the GPC5 region associated with cigarettes per day (CPD)

			CPD - Combined Sample				CPD - EAs Only				CPD - AAs Only			
SNP	**A1**	**A2**	**Freq**	**BETA**	**SE**	**P-Value**	**Freq**	**BETA**	**SE**	**P-Value**	**Freq**	**BETA**	**SE**	**P-Value**
rs1475655	**A**	**T**	**0.71**	**-0.14**	**0.03**	**4.09E-05**	**0.75**	**-0.18**	**0.04**	**1.18E-05**	**0.60**	**-0.02**	**0.05**	**6.60E-01**
rs9523299	**G**	**A**	**0.70**	**-0.13**	**0.03**	**8.23E-05**	**0.74**	**-0.17**	**0.04**	**3.27E-05**	**0.59**	**-0.02**	**0.05**	**6.41E-01**
rs7994634	**C**	**T**	**0.70**	**-0.13**	**0.03**	**8.57E-05**	**0.74**	**-0.17**	**0.04**	**3.31E-05**	**0.59**	**-0.02**	**0.05**	**6.65E-01**
rs9301726	**T**	**C**	**0.70**	**-0.13**	**0.03**	**8.66E-05**	**0.74**	**-0.17**	**0.04**	**3.35E-05**	**0.59**	**-0.02**	**0.05**	**6.65E-01**
rs7986895	**C**	**A**	**0.70**	**-0.13**	**0.03**	**9.19E-05**	**0.74**	**-0.17**	**0.04**	**4.29E-05**	**0.59**	**-0.03**	**0.05**	**6.14E-01**
rs12561118	**T**	**C**	**0.90**	**0.23**	**0.06**	**9.43E-05**	**0.87**	**0.24**	**0.07**	**2.01E-04**	**0.97**	**0.02**	**0.18**	**9.06E-01**
chr13: 91955287:I	**A**	**AT**	**0.70**	**-0.13**	**0.03**	**9.44E-05**	**0.74**	**-0.17**	**0.04**	**3.65E-05**	**0.59**	**-0.02**	**0.05**	**6.67E-01**
rs9523295	**G**	**A**	**0.70**	**-0.13**	**0.03**	**1.04E-04**	**0.74**	**-0.17**	**0.04**	**4.45E-05**	**0.60**	**-0.02**	**0.05**	**6.42E-01**
rs9523296	**G**	**A**	**0.70**	**-0.13**	**0.03**	**1.07E-04**	**0.74**	**-0.17**	**0.04**	**4.25E-05**	**0.59**	**-0.02**	**0.05**	**6.72E-01**
chr13: 91952853:D	**CA**	**C**	**0.70**	**-0.12**	**0.03**	**1.09E-04**	**0.74**	**-0.16**	**0.04**	**4.53E-05**	**0.59**	**-0.02**	**0.05**	**6.64E-01**
rs7335045	**A**	**G**	**0.70**	**-0.12**	**0.03**	**1.14E-04**	**0.74**	**-0.17**	**0.04**	**3.61E-05**	**0.59**	**-0.02**	**0.05**	**7.50E-01**
rs9589183	**C**	**T**	**0.70**	**-0.12**	**0.03**	**1.15E-04**	**0.74**	**-0.17**	**0.04**	**3.62E-05**	**0.59**	**-0.02**	**0.05**	**7.50E-01**
rs7989842	**G**	**T**	**0.70**	**-0.12**	**0.03**	**1.15E-04**	**0.74**	**-0.17**	**0.04**	**3.63E-05**	**0.59**	**-0.02**	**0.05**	**7.50E-01**
rs9515908	**C**	**T**	**0.71**	**-0.13**	**0.03**	**1.17E-04**	**0.75**	**-0.18**	**0.04**	**1.11E-05**	**0.59**	**0.01**	**0.05**	**8.93E-01**
rs7995715	**T**	**G**	**0.68**	**-0.12**	**0.03**	**1.40E-04**	**0.74**	**-0.17**	**0.04**	**3.38E-05**	**0.50**	**0.00**	**0.05**	**9.49E-01**
rs9523293	**C**	**T**	**0.72**	**-0.12**	**0.03**	**3.94E-04**	**0.74**	**-0.16**	**0.04**	**4.61E-05**	**0.65**	**0.01**	**0.05**	**8.48E-01**
rs34165267	**C**	**T**	**0.72**	**-0.12**	**0.03**	**3.94E-04**	**0.74**	**-0.16**	**0.04**	**4.62E-05**	**0.65**	**0.01**	**0.05**	**8.49E-01**
rs9583907	**C**	**T**	**0.72**	**-0.12**	**0.03**	**3.99E-04**	**0.74**	**-0.16**	**0.04**	**4.64E-05**	**0.65**	**0.01**	**0.05**	**8.47E-01**
rs7994469	**G**	**A**	**0.72**	**-0.12**	**0.03**	**4.07E-04**	**0.74**	**-0.16**	**0.04**	**4.64E-05**	**0.65**	**0.01**	**0.05**	**8.12E-01**
rs12873378	**T**	**A**	**0.69**	**-0.13**	**0.04**	**4.20E-04**	**0.73**	**-0.15**	**0.05**	**2.12E-03**	**0.59**	**-0.10**	**0.06**	**8.23E-02**
rs7985179	**T**	**A**	**0.72**	**-0.11**	**0.03**	**4.22E-04**	**0.74**	**-0.16**	**0.04**	**4.64E-05**	**0.65**	**0.01**	**0.05**	**8.14E-01**
rs7984992	**T**	**C**	**0.72**	**-0.11**	**0.03**	**4.26E-04**	**0.74**	**-0.16**	**0.04**	**4.64E-05**	**0.65**	**0.01**	**0.05**	**8.10E-01**
rs9583908	**T**	**C**	**0.72**	**-0.11**	**0.03**	**4.26E-04**	**0.74**	**-0.16**	**0.04**	**4.63E-05**	**0.65**	**0.01**	**0.05**	**8.10E-01**
rs12867738	**G**	**A**	**0.71**	**-0.11**	**0.03**	**4.43E-04**	**0.74**	**-0.16**	**0.04**	**4.56E-05**	**0.61**	**0.01**	**0.05**	**7.89E-01**
rs12708388	**C**	**T**	**0.71**	**-0.11**	**0.03**	**4.60E-04**	**0.74**	**-0.16**	**0.04**	**4.52E-05**	**0.64**	**0.02**	**0.05**	**7.64E-01**
rs9515905	**A**	**G**	**0.69**	**-0.11**	**0.03**	**4.96E-04**	**0.74**	**-0.17**	**0.04**	**4.24E-05**	**0.57**	**0.02**	**0.05**	**7.00E-01**
rs10161911	**C**	**T**	**0.69**	**-0.11**	**0.03**	**5.12E-04**	**0.74**	**-0.17**	**0.04**	**3.72E-05**	**0.54**	**0.02**	**0.05**	**6.41E-01**
rs1332217	**T**	**G**	**0.68**	**-0.11**	**0.03**	**5.60E-04**	**0.74**	**-0.17**	**0.04**	**4.24E-05**	**0.50**	**0.03**	**0.05**	**6.22E-01**
rs1332216	**T**	**C**	**0.67**	**-0.11**	**0.03**	**7.13E-04**	**0.74**	**-0.17**	**0.04**	**3.83E-05**	**0.49**	**0.04**	**0.05**	**4.90E-01**
rs9515906	**G**	**C**	**0.69**	**-0.11**	**0.03**	**9.12E-04**	**0.75**	**-0.16**	**0.04**	**8.83E-05**	**0.53**	**0.02**	**0.05**	**6.86E-01**

Data shown for the combined sample and EAs/AAs alone. Table sorted by p-value in the combined sample.

Table 6

Top 30 SNPs in the GPC5 region associated with time to first cigarette (TTF)

			TTF - Combined Sample				TTF - EAs Only				TTF - AAs Only			
SNP	A1	A2	Freq	BETA	SE	P-Value	Freq	BETA	SE	P-Value	Freq	BETA	SE	P-Value
rs7995715	T	G	0.68	-0.21	0.04	4.72E-08	0.74	-0.22	0.05	1.85E-06	0.50	-0.17	0.07	1.83E-02
rs1475655	A	T	0.71	-0.22	0.04	5.27E-08	0.75	-0.23	0.05	1.24E-06	0.60	-0.17	0.07	2.05E-02
rs9515908	C	T	0.71	-0.22	0.04	6.28E-08	0.75	-0.23	0.05	1.14E-06	0.59	-0.17	0.07	2.41E-02
rs9523289	G	A	0.81	-0.26	0.05	6.46E-08	0.86	-0.23	0.06	9.34E-05	0.69	-0.27	0.08	6.38E-04
rs9523295	G	A	0.70	-0.21	0.04	8.01E-08	0.74	-0.22	0.05	1.69E-06	0.60	-0.16	0.07	2.24E-02
rs7994634	C	T	0.70	-0.21	0.04	8.44E-08	0.74	-0.23	0.05	1.03E-06	0.59	-0.15	0.07	3.41E-02
rs9301726	T	C	0.70	-0.21	0.04	8.47E-08	0.74	-0.23	0.05	1.05E-06	0.59	-0.15	0.07	3.36E-02
rs7335045	A	G	0.70	-0.21	0.04	8.59E-08	0.74	-0.22	0.05	1.38E-06	0.59	-0.16	0.07	2.86E-02
rs9589183	C	T	0.70	-0.21	0.04	8.67E-08	0.74	-0.22	0.05	1.39E-06	0.59	-0.16	0.07	2.86E-02
rs7989842	G	T	0.70	-0.21	0.04	8.69E-08	0.74	-0.22	0.05	1.40E-06	0.59	-0.16	0.07	2.86E-02
rs9523299	G	A	0.70	-0.21	0.04	9.36E-08	0.74	-0.22	0.05	1.29E-06	0.59	-0.15	0.07	3.04E-02
rs7986895	C	A	0.70	-0.21	0.04	9.64E-08	0.74	-0.22	0.05	1.59E-06	0.59	-0.16	0.07	2.71E-02
rs9523296	G	A	0.70	-0.21	0.04	1.13E-07	0.74	-0.22	0.05	1.57E-06	0.59	-0.15	0.07	3.17E-02
chr13:91955287:I	A	AT	0.70	-0.21	0.04	1.15E-07	0.74	-0.22	0.05	1.41E-06	0.59	-0.15	0.07	3.45E-02
chr13:91952853:D	CA	C	0.70	-0.21	0.04	1.33E-07	0.74	-0.22	0.05	1.78E-06	0.59	-0.15	0.07	3.26E-02
rs9523288	C	T	0.82	-0.25	0.05	1.40E-07	0.86	-0.22	0.06	1.68E-04	0.69	-0.27	0.08	7.50E-04
rs67147421	A	C	0.82	-0.25	0.05	1.72E-07	0.86	-0.22	0.06	2.45E-04	0.70	-0.27	0.08	6.14E-04
rs73599638	G	A	0.82	-0.24	0.05	3.89E-07	0.86	-0.21	0.06	3.65E-04	0.69	-0.26	0.08	9.84E-04
rs10161911	C	T	0.69	-0.20	0.04	3.90E-07	0.74	-0.22	0.05	1.37E-06	0.54	-0.12	0.07	9.54E-02
rs9515905	A	G	0.69	-0.20	0.04	4.35E-07	0.74	-0.22	0.05	1.56E-06	0.57	-0.12	0.07	9.25E-02
rs1332216	T	C	0.67	-0.20	0.04	4.44E-07	0.74	-0.22	0.05	1.34E-06	0.49	-0.12	0.07	1.09E-01
rs1332217	T	G	0.68	-0.19	0.04	6.70E-07	0.74	-0.22	0.05	1.48E-06	0.50	-0.11	0.07	1.29E-01
rs68126334	C	T	0.82	-0.23	0.05	7.97E-07	0.86	-0.22	0.06	2.20E-04	0.69	-0.23	0.08	3.73E-03
rs9515906	G	C	0.69	-0.19	0.04	1.59E-06	0.75	-0.22	0.05	3.36E-06	0.53	-0.11	0.07	1.38E-01
rs9583905	T	A	0.81	-0.23	0.05	1.64E-06	0.86	-0.22	0.06	2.38E-04	0.67	-0.21	0.08	7.16E-03
rs9301724	C	T	0.80	-0.22	0.05	1.72E-06	0.85	-0.21	0.06	2.12E-04	0.67	-0.21	0.08	7.81E-03
rs9523293	C	T	0.72	-0.18	0.04	2.58E-06	0.74	-0.22	0.05	1.81E-06	0.65	-0.08	0.07	2.50E-01
rs34165267	C	T	0.72	-0.18	0.04	2.60E-06	0.74	-0.22	0.05	1.82E-06	0.65	-0.08	0.07	2.50E-01
rs9583907	C	T	0.72	-0.18	0.04	2.61E-06	0.74	-0.22	0.05	1.83E-06	0.65	-0.08	0.07	2.50E-01
rs59920274	C	G	0.80	-0.22	0.05	2.92E-06	0.86	-0.22	0.06	2.40E-04	0.64	-0.20	0.08	9.36E-03

The data shown are for the combined sample and EAs and AAs alone. The table is sorted by p-value in the combined sample.

Table 7

Top 30 SNPs in the CHRNA5-A3-B4 region associated with FTND nicotine dependence

			FTND - Combined Sample				FTND - EAs Only				FTND - AAs Only			
SNP	**A1**	**A2**	**Freq**	**OR**	**SE**	**P-Value**	**Freq**	**OR**	**SE**	**P-Value**	**Freq**	**OR**	**SE**	**P-Value**
chr15: 78874842:D	**AG**	**A**	**0.73**	**0.69**	**0.07**	**2.93E-08**	**0.67**	**0.72**	**0.07**	**4.87E-06**	**0.89**	**0.43**	**0.22**	**1.31E-04**
rs114205691	**C**	**T**	**0.69**	**0.71**	**0.06**	**4.35E-08**	**0.65**	**0.71**	**0.07**	**1.32E-06**	**0.82**	**0.67**	**0.15**	**8.83E-03**
rs17486278	**A**	**C**	**0.67**	**0.72**	**0.06**	**4.91E-08**	**0.65**	**0.72**	**0.07**	**1.80E-06**	**0.72**	**0.72**	**0.13**	**9.28E-03**
rs147499554	**C**	**T**	**0.73**	**0.69**	**0.07**	**5.33E-08**	**0.68**	**0.72**	**0.07**	**6.91E-06**	**0.86**	**0.52**	**0.19**	**5.31E-04**
rs141518190	**A**	**G**	**0.73**	**0.69**	**0.07**	**5.33E-08**	**0.68**	**0.72**	**0.07**	**6.91E-06**	**0.86**	**0.52**	**0.19**	**5.31E-04**
rs2036527	**G**	**A**	**0.68**	**0.72**	**0.06**	**7.54E-08**	**0.64**	**0.72**	**0.07**	**2.96E-06**	**0.78**	**0.68**	**0.14**	**5.36E-03**
rs16969968	**G**	**A**	**0.73**	**0.70**	**0.07**	**7.82E-08**	**0.65**	**0.72**	**0.07**	**1.86E-06**	**0.94**	**0.43**	**0.29**	**3.31E-03**
rs55781567	**C**	**G**	**0.67**	**0.72**	**0.06**	**9.70E-08**	**0.64**	**0.72**	**0.07**	**2.66E-06**	**0.74**	**0.72**	**0.13**	**1.29E-02**
rs55676755	**C**	**G**	**0.70**	**0.71**	**0.06**	**9.74E-08**	**0.65**	**0.71**	**0.07**	**1.39E-06**	**0.83**	**0.70**	**0.16**	**2.34E-02**
rs11633958	**C**	**T**	**0.73**	**0.70**	**0.07**	**9.95E-08**	**0.65**	**0.72**	**0.07**	**2.35E-06**	**0.94**	**0.44**	**0.28**	**3.73E-03**
rs8192482	**C**	**T**	**0.73**	**0.70**	**0.07**	**1.03E-07**	**0.65**	**0.72**	**0.07**	**2.17E-06**	**0.94**	**0.43**	**0.29**	**3.96E-03**
rs4887067	**G**	**A**	**0.73**	**0.70**	**0.07**	**1.03E-07**	**0.65**	**0.72**	**0.07**	**2.17E-06**	**0.94**	**0.43**	**0.29**	**3.97E-03**
rs55853698	**T**	**G**	**0.72**	**0.70**	**0.07**	**1.08E-07**	**0.64**	**0.72**	**0.07**	**2.29E-06**	**0.92**	**0.53**	**0.23**	**5.87E-03**
rs17487223	**C**	**T**	**0.69**	**0.71**	**0.07**	**1.21E-07**	**0.62**	**0.72**	**0.07**	**3.08E-06**	**0.89**	**0.58**	**0.21**	**1.07E-02**
rs72740955	**C**	**T**	**0.71**	**0.71**	**0.06**	**1.29E-07**	**0.64**	**0.72**	**0.07**	**3.54E-06**	**0.88**	**0.60**	**0.18**	**5.71E-03**
rs140330585	**G**	**A**	**0.71**	**0.71**	**0.06**	**1.36E-07**	**0.65**	**0.72**	**0.07**	**3.13E-06**	**0.87**	**0.63**	**0.18**	**1.11E-02**
rs2869548	**G**	**A**	**0.70**	**0.70**	**0.07**	**1.48E-07**	**0.62**	**0.72**	**0.07**	**2.27E-06**	**0.94**	**0.46**	**0.29**	**6.30E-03**
rs17486195	**A**	**G**	**0.71**	**0.71**	**0.06**	**1.52E-07**	**0.65**	**0.72**	**0.07**	**2.42E-06**	**0.88**	**0.65**	**0.18**	**1.63E-02**
rs7172118	**C**	**A**	**0.71**	**0.71**	**0.06**	**1.74E-07**	**0.65**	**0.72**	**0.07**	**2.35E-06**	**0.89**	**0.64**	**0.19**	**1.62E-02**
rs1051730	**G**	**A**	**0.71**	**0.72**	**0.06**	**2.18E-07**	**0.65**	**0.72**	**0.07**	**2.20E-06**	**0.88**	**0.66**	**0.18**	**2.28E-02**
rs951266	**G**	**A**	**0.71**	**0.71**	**0.06**	**2.21E-07**	**0.65**	**0.72**	**0.07**	**2.65E-06**	**0.89**	**0.64**	**0.19**	**1.82E-02**
rs56077333	**C**	**A**	**0.70**	**0.72**	**0.06**	**2.24E-07**	**0.66**	**0.72**	**0.07**	**6.55E-06**	**0.82**	**0.66**	**0.16**	**7.79E-03**
rs7180002	**A**	**T**	**0.71**	**0.72**	**0.06**	**2.25E-07**	**0.65**	**0.72**	**0.07**	**2.71E-06**	**0.89**	**0.64**	**0.19**	**1.84E-02**
rs8031948	**G**	**T**	**0.70**	**0.72**	**0.06**	**2.46E-07**	**0.65**	**0.73**	**0.07**	**8.71E-06**	**0.85**	**0.64**	**0.17**	**8.41E-03**
rs56390833	**C**	**A**	**0.71**	**0.72**	**0.06**	**2.67E-07**	**0.65**	**0.72**	**0.07**	**3.10E-06**	**0.89**	**0.64**	**0.19**	**1.88E-02**
rs138544659	**T**	**G**	**0.73**	**0.70**	**0.07**	**2.84E-07**	**0.69**	**0.73**	**0.07**	**1.84E-05**	**0.87**	**0.54**	**0.19**	**1.37E-03**
rs147144681	**C**	**T**	**0.71**	**0.71**	**0.07**	**3.18E-07**	**0.66**	**0.72**	**0.07**	**4.24E-06**	**0.85**	**0.67**	**0.17**	**1.78E-02**
rs72740964	**G**	**A**	**0.72**	**0.71**	**0.07**	**3.38E-07**	**0.65**	**0.72**	**0.07**	**2.24E-06**	**0.93**	**0.59**	**0.25**	**3.28E-02**
rs58365910	**T**	**C**	**0.67**	**0.73**	**0.06**	**3.75E-07**	**0.64**	**0.72**	**0.07**	**2.28E-06**	**0.75**	**0.76**	**0.13**	**4.01E-02**
rs146009840	**A**	**T**	**0.72**	**0.71**	**0.07**	**4.16E-07**	**0.65**	**0.72**	**0.07**	**1.59E-06**	**0.93**	**0.64**	**0.24**	**6.80E-02**

Data shown for the combined sample and EAs/AAs alone. Table sorted by p-value in the combined sample.

Genotype Confirmation. Using the Sequenom platform, we performed follow-up genotyping on 28 SNPs in the region of the novel *GPC5* signal (Table 8). All SNPs with an r^2 >0.8 with rs7995715, and any SNP with an r^2 >0.8 with the first group of SNPs were chosen for follow up genotyping. We confirm the results from the imputed data of a

GWS signal near *GPC5* associated with nicotine dependence phenotypes. In a logistic analysis on chromosome 13, rs9515908 (r^2= 0.96 with rs7995715) had an N=2614; p-value = 8.42 x 10^{-8} and OR=1.4 with the phenotype of FTND nicotine dependence.

Table 8

Results from the follow-up genotyping of SNPs in LD with rs7995715

Sequenom Genotyping Confirmation						
CHR	**SNP**	**BP**	**N**	**Freq**	**OR**	**P-Value**
13	**rs9515908**	**91965134**	**2614**	**0.30**	**1.40**	**8.42E-08**
13	**rs7995715**	**91955562**	**2567**	**0.32**	**1.39**	**1.28E-07**
13	**rs9523295**	**91948047**	**2565**	**0.29**	**1.39**	**1.75E-07**
13	**rs7994634**	**91953042**	**2613**	**0.30**	**1.38**	**2.38E-07**
13	**rs9523296**	**91949444**	**2604**	**0.30**	**1.38**	**2.60E-07**
13	**rs9523299**	**91955192**	**2614**	**0.30**	**1.38**	**3.30E-07**
13	**rs9301726**	**91952459**	**2612**	**0.30**	**1.38**	**3.49E-07**
13	**rs1475655**	**91963080**	**2607**	**0.29**	**1.37**	**5.73E-07**
13	**rs9515905**	**91949562**	**2491**	**0.30**	**1.36**	**1.55E-06**
13	**rs7139676**	**91970313**	**2611**	**0.29**	**1.32**	**9.40E-06**
13	**rs869544**	**91956987**	**2606**	**0.49**	**0.78**	**1.44E-05**
13	**rs78375372**	**91976200**	**2612**	**0.12**	**0.83**	**0.03625**
13	**rs74357547**	**92027055**	**2611**	**0.03**	**0.78**	**0.1375**
13	**rs16945778**	**91978780**	**2609**	**0.18**	**0.91**	**0.1881**
13	**rs9634624**	**92029241**	**2611**	**0.22**	**0.93**	**0.3163**
13	**rs7332464**	**92065093**	**2606**	**0.18**	**1.08**	**0.3563**
13	**rs72640378**	**92106136**	**2613**	**0.01**	**0.75**	**0.361**
13	**rs17556509**	**91995570**	**2610**	**0.07**	**0.91**	**0.4026**
13	**rs74622835**	**92058305**	**2612**	**0.07**	**0.93**	**0.4742**
13	**rs7325427**	**91930464**	**2613**	**0.05**	**0.90**	**0.4897**
13	**rs9556077**	**91945836**	**2581**	**0.15**	**0.95**	**0.493**
13	**rs9589195**	**91974565**	**2614**	**0.09**	**1.07**	**0.4959**
13	**rs9556074**	**91904181**	**2611**	**0.00**	**1.35**	**0.5288**
13	**rs12232047**	**91928228**	**2612**	**0.00**	**1.31**	**0.5921**
13	**rs9589196**	**91974740**	**2600**	**0.43**	**0.98**	**0.7133**
13	**rs7324710**	**92085325**	**2597**	**0.14**	**1.03**	**0.7254**
13	**rs79977572**	**92095067**	**2613**	**0.04**	**0.95**	**0.7509**
13	**rs7318578**	**92005469**	**2612**	**0.35**	**0.99**	**0.8457**

Replication Datasets. We have attempted to replicate these findings in independent datasets. The first of which was in the SAGE dataset, subtracting those individuals that overlap between SAGE and COGEND. However with the phenotype of FTND nicotine dependence we could not replicate the results from the COGEND dataset for either chromosome 13 or chromosome 15. In European-Americans, the p-value for rs7995715 was 0.08, OR=1.40, 95% CI=1.02-1.78; in African-Americans p=0.94, OR=1.01, 95% CI=0.67-1.39; meta-analysis p=0.12, OR=1.22. For rs114205691, in European-Americans, the p-value was 0.73, OR=1.06, 95% CI= 0.70-1.42; in African-Americans p=0.57, OR=1.14, 95% CI= 0.70-1.58; meta-analysis p=0.93, OR=1.10. There was 62% power to detect an association with a p-value less than 0.05 in the COGA dataset and a power of 47% in the SAGE-COGEND dataset. See Table 9 for full results. We also examined differing CPD thresholds to see if a different phenotype was more able to detect an association. No matter what the threshold, no SNP in LD with either rs7995715 or rs114205691 was significant.

We also attempted replication in the individuals in the COGA dataset that are not in SAGE. However rs7995715 was not typed in the COGA dataset. The only two SNPs in COGA that are in high LD with rs7995715 that were genotyped in COGA are rs9523299 (r^2=1) and rs9583907 (r^2=1). We examined both SNPs with the phenotypes of CPD as an ordinal trait and dichotomous over 20 CPD vs under 20 CPD. In both cases for both SNPs, neither was significant ($0.20<p<0.82$, $-0.78<\beta<0.00$).

Table 9

Results of SAGE minus COGEND replication attempt for the top SNPs on chromosomes 13 and 15

SAGE minus COGEND Replication Results				
	P-Value rs7995715	**Odds Ratio**	**P-Value rs114205691**	**Odds Ratio**
FTND_DX				
EA	**0.08**	**1.40**	**0.73**	**1.06**
AA	**0.94**	**1.01**	**0.57**	**1.14**
Meta	**0.93**	**1.10**	**0.20**	**1.18**
CPD Over 20/Under 10				
EA	**0.94**	**1.02**	**0.61**	**1.13**
AA	**0.65**	**1.10**	**0.79**	**1.07**
Meta	**0.76**	**1.06**	**0.58**	**1.10**
CPD Over 20/Under 20				
EA	**0.62**	**1.07**	**0.61**	**1.06**
AA	**0.88**	**1.03**	**0.73**	**1.08**
Meta	**0.75**	**1.04**	**0.77**	**1.03**
CPD Over 30/Under 10				
EA	**0.87**	**1.05**	**0.83**	**1.06**
AA	**0.26**	**1.11**	**0.53**	**1.20**
Meta	**0.69**	**1.08**	**0.56**	**1.12**
CPD Over 30/Under 30				
EA	**0.33**	**1.16**	**0.80**	**1.03**
AA	**0.86**	**1.04**	**0.73**	**1.09**
Meta	**0.36**	**1.12**	**0.95**	**1.01**

DISCUSSION

We report here the largest GWAS for FTND nicotine dependence resulting in two genome-wide significant hits. This is the first time that the known bin on

chromosome 15 has been associated with FTND nicotine dependence at the genome-wide significant level and this is also the first time there has been an association with nicotine dependence and a SNP near a non-candidate gene, in this case *GPC5* on chromosome 13. Both associations with FTND nicotine dependence are clearly detected in the combined analysis as well as the meta-analysis (Appendix A). The signal on chromosome 15 is much stronger in European-Americans ($p=1.32x10^{-6}$, OR=1.4, 95% CI= 1.33-1.47) than African-Americans ($p=8.83x10^{-3}$, OR=1.5, 95% CI=1.2-1.8), partially due to the larger EA sample size, however it should be noted that the point estimate of the odds ratios are similar even though the p-values are different. The lower p-value in African-Americans is also due to the fact that it has been shown that CPD is not as effective a measure of addiction in African-Americans, and CPD is a large component of FTND score. In this dataset, CPD in African-Americans is not significantly associated with rs7995715 ($p=0.95$, $\beta=0.00$), concurring with previous work showing that FTND nicotine dependence can succeed where CPD failed in picking up this association (Rice et al. 2012). However for the signal on chromosome 13, in addition to FTND nicotine dependence, TTF is also genome-wide significant in the combined sample ($p=4.72 \times 10^{-8}$, $\beta=-0.21$), suggesting that it may be picking up on some aspect of craving that is important to the biology of this association at *GPC5* that CPD does not pick up as strongly ($p=1.40 \times 10^{-4}$, $\beta=-0.12$).

The glypican gene family contains 6 members (*GPC1* to *GPC6).* This gene family is composed of heparin sulphate proteoglycans (HSPGs). *GPC5* variants have been associated with several types of tumors including lymphomas, neurological tumors, and breast tumors (Y. Li & Yang, 2011). The *GPC5* gene contains eight exons encoding

572 amino acids and spans a region of 1.47 Mb (Zheng et al., 2012). In a GWAS of never smokers rs2352028/rs2352029 were identified as associated with lung cancer (Y. Li et al., 2010), although the p-value of 5.94×10^{-6} was not genome-wide significant. Although this has yet to be replicated, Zheng et al. (Zheng et al., 2012) reported a nominal p-value of 0.04 in a Han-Chinese population when a recessive model was assumed, as well as when cases were narrowed down to only those patients with adenocarcinoma. In a GWAS of alcohol dependence, several SNPs within *GPC5* have been associated with AD (top hit rs148154304, $p=7.80 \times 10^{-6}$, Dr. Amy Adkins, personal communication).

These results are encouraging in that smoking-related phenotypes have been previously significantly associated with genotypes in this gene region. If common gene family of the glypicans is any indication of common function, the fact that both *GPC4* and *GPC6* have been associated with the psychiatric conditions of attention deficit hyperactivity disorder (Lesch et al., 2008) and neuroticism (Calboli et al., 2010) respectively is also encouraging, due to the comorbidity of psychiatric conditions.

Although we have only imputed from ~600,000 SNPs, we have found a new genome-wide significant association, as well as replicated a previous association. Imputation can be problematic, and result in misleading or spurious associations in some cases due to insufficient filtering or improperly combined platforms. However in this case, we have taken the most conservative approach possible in that we only used the intersection of these two platforms, discarding ~2 million SNPs in the process. This is preferable to simply combining the two datasets, as this incurs many problems of its own, including spurious associations that result from a SNP only having been genotyped on

one or the other platform, or from having two separate populations imputed together. We also filtered the SNPs with MAF less than 5% and genotyping or imputed call rate of 97% or higher. Because of this, we are confident in our results, as well as because the Q-Q plots show no inflation after removal of the two genome-wide significant SNPs.

Although we were not able to replicate our results within either SAGE (without COGEND) or in all of COGA, this lack of replication could reflect differences in ascertainment of the samples or differences in phenotype: Only the COGEND study has FTND measured. In the other studies we relied on CPD which shows a weaker association even in the original dataset. We dichotomized the CPD categories to approximate the FTND categories but this may not be entirely accurate in samples like COGA that have multigenerational families and thus the CPD corresponding to a particular FTND may be difference across generations. Additionally, the lack of replication could be due to poor controls. For example just because a person does not smoke many cigarettes a day does not mean they are not addicted to nicotine. A more powerful replication sample would be one that has measured FTND nicotine dependence and that was ascertained in a similar manner to the discovery dataset.

Importantly, we also show that although large sample sizes are useful when using the GWAS approach, they are not necessary if precise phenotypes are used. Here we had a sample size of less than 3,000; however we have discovered previously unknown addiction associations by using a uniformly collected dichotomous phenotype. It is our hope that with this new information, new pathways involved in addiction will be studied and will lead to novel strategies for cessation therapy and initiation prevention. Like all approaches, imputation and GWAS have limitations, including the inability to detect rare

variants. But if performed carefully and interpreted conservatively, imputation has its place in association analyses. GWAS can identify genes involved in underlying pathways, which is promising for increasing our understanding of the biology of addiction and addictive behavior. In the future, we would like to re-genotype all of the samples on the 2.5M chip to increase the number of SNPs from which to impute, as this may lead to additional discoveries of loci associated with nicotine dependence.

CHAPTER 3

LOW FREQUENCY AND COMMON VARIANTS NEAR CHRNB3-CHRNA6 ARE ASSOCIATED WITH MULTIPLE SUBSTANCE DEPENDENCIES

ABSTRACT

Drug and alcohol dependence are pervasive problems that affect millions of individuals across the world every year. Numerous studies have demonstrated the presence of drug specific and multi-drug genetic influences. One such genetic factor, the *CHRNA3-B4-A5* nicotinic receptor gene cluster on chromosome 15, was recently identified as a locus contributing to alcohol, cocaine and nicotine dependence, each independently of the other. Similarly, our group recently demonstrated an association between rare coding variants in *CHRNB3* with alcohol and cocaine dependence without an effect on nicotine dependence while common variants within the *CHRNB3-A6* gene cluster have been associated with cigarette consumption in several genome-wide association studies. These data suggest that other genetic variants in or near nicotinic receptor genes may play a role in one or more of these substance dependencies. Generally, these receptors represent intriguing candidate genes for the study of cocaine and alcohol dependence because nicotinic receptors are thought to be involved in generalized addiction pathways in addition to nicotine specific pathways. Using genotypic data from a GWAS of the Study of Addiction: Genetics and Environment (SAGE) dataset including 1976 European-Americans, we tested for association of *CHRNB3-A6* SNPs with DSM-5 cocaine use disorder and DSM-5 alcohol use disorder. Multiple SNPs in the region were significantly associated with increased risk of cocaine use disorder, but none were significantly associated with alcohol use disorder after

multiple test correction. Further, inclusion of the most significant SNP as a covariate in a linear regression model provided evidence for an additional independent signal within this locus for DSM-5 cocaine use disorder, in European Americans. Interestingly, the SNPs associated with increased risk for cocaine use disorder, are also associated with decreased risk for nicotine dependence in this dataset. When the previously identified nicotine dependence risk variant (rs1451240) is included in the model, the newly identified SNPs remain associated with DSM-5 cocaine use disorder but are no longer associated with nicotine dependence, suggesting that the SNPs in this region affecting risk for these two disorders are at least partially independent and that the *CHRNB3-A6* locus contains multiple variants affecting risk for vulnerability to cocaine and nicotine dependence. This locus is the second nicotinic receptor gene cluster containing SNPs that show opposing directions of effect for nicotine and cocaine dependence risk.

INTRODUCTION

Differences in any trait must be due to either genetic or environmental factors or both, and addiction is no exception. From twin studies, we have found that different substances have common and specific genetic liabilities. Numerous twin studies indicate a high degree of overlap among genetic factors influencing the liability to a variety of substance use disorders (Kendler, Myers, & Prescott, 2007; Tsuang, Bar, Harley, & Lyons, 2001). Genomic studies have also suggested that there are genetic loci that have substance-specific effects but also that loci exist that affect risk for the development of dependence on multiple substances (see (J. C. Wang et al., 2012) for review). Loci that have been largely implicated to specifically influence a single substance use disorder

include those that exert metabolic influence on the substance of abuse. For instance, SNPs in cytochrome P450 2A6 (*CYP2A6)*, the gene encoding the major nicotine-metabolizing enzyme, affect cigarette consumption (Mwenifumbo & Tyndale, 2009; Thorgeirsson et al., 2010) and a SNP in the Alcohol Dehydrogenase 1B (*ADH1B*) gene affects levels of alcohol consumption (Bierut et al., 2012) and risk for alcohol dependence (J. C. Wang et al., 2012) via regulation of conversion of alcohol to acetaldehyde.

On the other hand, there are numerous examples of receptor encoding loci whose effects extend across multiple substance dependence phenotypes (Sherva et al., 2010). One such example is the SNP rs16969968 (D398N) in the cholinergic nicotinic receptor subunit α5 (*CHRNA5*) that both increases nicotine dependence risk and decreases cocaine dependence risk (Sherva et al., 2010). The minor allele of this SNP is the most significant and widely replicated variant associated with cigarette consumption and is also associated with protection against cocaine dependence (Consortium, 2010; Grucza et al., 2008; Thorgeirsson et al., 2010). The protective effect of rs16969968 with CD has been replicated in both European and African-Americans (Sherva et al., 2010). The same study also found that another SNP in *CHRNA5* (rs684513) is associated with risk for cocaine dependence in African-Americans (OR=1.43, P=0.0004).

In fact, prior studies of the nicotinic receptors on chromosomes 8 and 15 show that variants within or near these receptors are associated with nicotine (Grucza et al., 2008; S. F. Saccone et al., 2007), alcohol (Haller et al., 2013), and cocaine dependencies (Haller et al., 2013). In addition, a cluster of nicotinic receptors on chromosome 8 including *CHRNA6-B3* was also previously shown to reduce risk for nicotine-related

phenotypes in several GWAS of nicotine dependence and cigarettes smoked per day (Hoft et al., 2009; Mwenifumbo & Tyndale, 2009; Rice et al., 2012; N. L. Saccone, Culverhouse, et al., 2010; N. L. Saccone et al., 2009; S. F. Saccone et al., 2007; Thorgeirsson et al., 2010; Zeiger et al., 2008). The role of these specific SNPs in the etiology of other drug dependencies remains unexplored but several rare variants in *CHRNB3* have been associated with increased risk for both cocaine and alcohol dependence (Haller et al., 2013). Together, these results suggest that nicotinic receptors are good candidate genes for susceptibility to multiple substance dependence vulnerability and that investigation of the role of common and low frequency variants within the *CHRNB3-A6* locus in drug dependence is warranted.

Drug addiction is a pervasive problem across cultures and is both an economic and psychological burden for the individuals and families involved. Susceptibility to drug use, abuse, and dependence has been shown by several studies to have a moderate to high genetic component (Goldman, Oroszi, & Ducci, 2005; J. C. Wang et al., 2012). The one year point prevalence, or the proportion of people possessing a phenotype within a one year timeframe, for substance use disorders in the USA, excluding nicotine has been estimated to be 9.35% (Goldman et al., 2005). The heritability of cocaine dependence (CD) has been estimated from twin studies to be 63-79% (Grucza et al. 2008), and that of alcohol dependence (AD) has been estimated to be 40-60% (J. C. Wang et al., 2012). However as it is common for an individual to have a dependence on more than one drug, as well as to have comorbidity with mental disorders (Goldman et al., 2005), the loci associated with one substance, have a greater chance of also being associated with multiple substances and other psychiatric disorders.

In this study, we describe a novel association between DSM-5 cocaine use disorder and genotyped SNPs (~24kb) upstream of the *CHRNB3* transcription start site that contains the locus previously discovered to be genome-wide significant with nicotine dependence. We show that these SNPs remain significant after adjusting for genotype at the SNP previously reported to be associated with nicotine dependence in GWAS, suggesting that the cocaine association is not simply due to the nicotine association.

METHODS

Samples. Subjects were members of the Study of Addiction: Genetics and Environment (SAGE) dataset, part of the Gene Environment Association Studies (GENEVA) program of the National Institutes of Health (NIH) Genes, Environment, and Health Initiative (Laurie et al., 2010). SAGE was designed to study alcohol dependence, and as a result is composed largely of unrelated alcohol-dependent cases (n = 1048) and non-alcohol-dependent control subjects (n = 928). The SAGE dataset was ascertained from 3 large substance dependence datasets: the Collaborative Genetic Study of Nicotine Dependence (COGEND), the Collaborative Study on the Genetics of Alcoholism (COGA) and the Family Study of Cocaine Dependence (FSCD) (Bierut et al., 2010). For the purpose of the current analyses, there were 1976 European-Americans as defined both by self-report and principal components from the GWAS data (See Table 10 for a summary of comorbidity within the sample).

The DSM-5 was published on May 18, 2013, and supersedes the DSM-IV text revision published in 2000. In the DSM-5, the DSM-IV criteria cocaine abuse and cocaine dependence have been combined into a single cocaine use disorder. Cocaine use

Table 10

Comorbidity in the sample

	COGEND	COGA	FSCD	Total
ND Cases	189	429	196	814
ND Controls	892	172	98	1162
AUD Cases	769	483	273	1525
AUD Controls	312	118	21	451
CUD Cases	76	212	218	506
CUD Controls	1005	389	76	1470
ND, AUD	184	373	189	746
ND, CUD	54	170	173	397
ND, AUD, CUD	54	170	168	392

disorder is now divided into mild (2-3 criteria), moderate (4-5 criteria) and severe (6 or more criteria). A further difference is that whereas in the DSM-IV, cocaine abuse diagnostic criteria required only one symptom, in the DSM-5 diagnostic criteria, a diagnosis of mild cocaine use disorder requires at least two criteria to be met. Lastly, the DSM-IV recurrent legal problems criterion for cocaine abuse was replaced with the new criterion of craving (Hasin et al., 2013). These same changes are also true for the

phenotype of alcohol use disorder. We recoded DSM-IV values in SAGE to DSM-5 for both cocaine and alcohol use, since we examine both phenotypes in our analyses.

COGEND Sample. COGEND was designed as a community based case–control study of nicotine dependence. COGEND contains current smokers with nicotine dependence defined by a Fagerstrom Test for Nicotine Dependence (FTND) score ≥ 4 (maximum score of 10) and non-nicotine dependent subjects who had smoked at least 100 cigarettes and had a lifetime FTND score of zero or one. All subjects were ascertained from Detroit and St Louis. Out of the 53,000 subjects who were screened by telephone, 2,800 were interviewed in person and approximately 2,700 donated blood samples for genetic studies (Rice et al. 2012).

COGA Sample. Out of more than 11,000 subjects who participated in COGA, a case-control series of unrelated individuals was selected for SAGE. COGA recruited subjects in Hartford, Connecticut; Indianapolis, Indiana; Iowa City, Iowa; New York City, New York; San Diego, California; St Louis, Missouri; and Washington, DC. For inclusion in SAGE, cases had to meet lifetime criteria for DSM-IV alcohol dependence, the majority of cases were recruited from alcoholism treatment centers. Control subjects, were both biologically unrelated to cases, and had consumed alcohol but never experienced any significant alcohol or drug-related problems, according to the Semi-Structured Assessment for the Genetics of Alcoholism (SSAGA) (Rice et al., 2012).

FSCD Sample. Subjects in the FSCD were specifically recruited for cocaine use from chemical dependency treatment units in the greater St Louis metropolitan area. The Missouri Family Registry identified community-based control subjects and matched them by age, race, gender and residential zip code. Controls were biologically unrelated

individuals from the same communities who consumed alcohol, but had no lifetime history of dependence on any substance.

Genotyping and quality control. All DNA samples were genotyped on the Illumina Human 1M-Duo beadchip by the Center for Inherited Disease Research (CIDR) at Johns Hopkins University. After thorough genotype quality control process, 948,758 of the 1,049,008 genotyped SNPs were available for genetic analysis. Sixty-five of these genotyped SNPs fell within the region containing the *CHRNA6* and *CHRNB3* genes on chromosome 8. Of the 65, only SNPs with a minor allele frequency (MAF) >1% and a genotyping call rate >0.98 were considered (47 SNPs). Full details regarding the quality control procedures are provided in the data cleaning report posted on the GENEVA website (http://www.genevastudy.org/docs/GENEVA_Alcohol_QC_report_8Oct2008.pdf) and in related publications (Bierut et al., 2010; Laurie et al., 2010).

Phenotypes. FTND nicotine dependence for all members of the SAGE dataset was calculated by adding together all point totals from the FTND questionnaire. Here we used an FTND score of 4 or above as a case (N=814) and below 0-3 as a control (N=1162). Alcohol use disorder for all members of the SAGE dataset was measured using the DSM-5 criteria (Hasin et al., 2013). As outlined in the manual, 11 criteria (3 abuse, 7 dependence and craving) were combined and alcohol use disorder was scored as the endorsement of 2 or more of these 11 criteria (N=1525). Unaffected individuals met zero or one of the DSM-5 criteria (N=451). Cocaine use disorder for all members of the SAGE dataset was measured using the same DSM-5 criteria. As outlined in the manual, 11 criteria (3 abuse, 7 dependence and craving) were combined and cocaine use disorder

was scored as the endorsement of 2 or more of these 11 criteria (N=506). Unaffected individuals met zero or one of the DSM-5 criteria (N=1470).

Statistical Analyses. All analyses were performed on genotyped data. Association analyses were conducted in PLINK (Purcell et al. 2007) for SNPs in the region on chromosome 8 encoding the *α6* and *β3* subunits of nicotinic receptors (42,600,000 kb to 42,800,000 kb). Logistic regression with DSM-5 cocaine use disorder as the dependent variable was performed. Covariates included were age at interview as a continuous variable, gender, study, maximum lifetime FTND score (0-10, based on the Fagerström Test of Nicotine Dependence) to control for smoking status, and DSM-5 alcohol use disorder. Study was coded using two dummy variables (yes/no for two of the three studies) in order to control for differences in ascertainment. Haploview was run using the genotypes of the study population to determine the number of independent linkage disequilibrium (LD) bins in the region using a threshold of $r^2 \geq 0.8$. The Bonferroni correction used in this study is p=0.002 (0.05/22), as the number of LD bins in the region examined is 22. A conditional analysis was conducted including allele dosage for the top associated SNP as a covariate in the logistic model.

In a case/control division of subjects based on presence or absence of cocaine use disorder, logistic regressions were run both using as controls only those who had been exposed to cocaine but had not become dependent (i.e. have used cocaine at least once in their lifetime) and all non-cocaine-dependent individuals in the sample, regardless of exposure status.

To improve our understanding of observed associations, the top SNPs identified in the whole SAGE dataset were examined using the same models described above in

strata of the data defined by study (COGEND, COGA, FSCD), smoking status (FTND cases and FTND controls), and alcohol use disorder (DSM-5 cases and DSM-5 controls). A two-SNP haplotype analysis was run in R using the top SNP and the SNP tagging the bin previously found to be genome-wide significant with nicotine dependence (rs1451240) (Rice et al. 2012). This model included the covariates age, sex, study, DSM-5 alcohol symptom count, FTND total (in the cocaine haplotypes) and DSM-5 cocaine use disorder (in the FTND haplotypes), and examined the association with each haplotype with the phenotype compared to homozygotes for the reference allele at both SNPs. Finally, we used conditional analyses to examine the extent of independence between these cocaine-associated SNPs and the previous association in the region with nicotine dependence tagged by rs1451240.

RESULTS

The 47 SNPs within the *CHRNA6-B3* region constitute 22 LD bins using an r^2 cutoff ≥ 0.8, requiring a p-value of 0.002 after Bonferroni correction. Eleven SNPs, representing four LD bins met this cutoff and are associated with DSM-5 cocaine use disorder (Table 11). Overall, a total of thirty-one SNPs were nominally significant ($2.34x10^{-4}<p<4.66x10^{-2}$) in this single SNP analysis. Consistent with previous results in an overlapping dataset, we saw a protective effect of rs16969968 in *CHRNA5* on risk for DSM-5 cocaine use disorder (Grucza et al., 2008). Inclusion of rs16969968 as a covariate had no effect on the association of the top SNP within *CHRNA6-CHRNB3*, rs9298626 with DSM-5 cocaine use disorder. When we run the analysis with DSM-5 alcohol use disorder, no SNPs in the region pass the Bonferroni correction. The most

significant SNP with this phenotype is rs7844566 (OR=1.64, p=0.01). With FTND nicotine dependence, the most highly associated SNP is rs4950 (OR=0.66, p=7.32 x 10^{-5}), Table 11.

Top Association Results for the Linear Models Run for DSM-5 Cocaine Use Disorder

SNP	LD Bin	bp	N	OR	L95	U95	Freq	P
rs9298626	**1**	**42,647,165**	**1970**	**2.618**	**1.568**	**4.372**	**0.04**	**2.34E-04**
rs7844824	**1**	**42,672,170**	**1970**	**2.652**	**1.575**	**4.464**	**0.04**	**2.43E-04**
rs4305884	**2**	**42,637,880**	**1966**	**2.133**	**1.419**	**3.205**	**0.06**	**2.69E-04**
rs7824160	**1**	**42,705,413**	**1969**	**2.502**	**1.494**	**4.19**	**0.04**	**4.88E-04**
rs11986893	**4**	**42,772,016**	**1971**	**1.564**	**1.216**	**2.011**	**0.2**	**4.92E-04**
rs7002907	**1**	**42,702,998**	**1970**	**2.494**	**1.49**	**4.174**	**0.04**	**5.03E-04**
rs6997994	**1**	**42,702,328**	**1971**	**2.494**	**1.49**	**4.175**	**0.04**	**5.04E-04**
rs7815274	**1**	**42,701,740**	**1967**	**2.469**	**1.469**	**4.149**	**0.04**	**6.41E-04**
rs4952	**1**	**42,706,222**	**1971**	**2.427**	**1.444**	**4.078**	**0.04**	**8.12E-04**
rs10107450	**5**	**42,749,052**	**1969**	**1.504**	**1.178**	**1.918**	**0.22**	**1.04E-03**
rs1868859	**2**	**42,634,958**	**1971**	**1.847**	**1.269**	**2.688**	**0.07**	**1.36E-03**
rs892413	3	42,733,535	1971	1.475	1.151	1.889	0.2	2.11E-03
rs4950	6	42,671,790	1957	1.422	1.107	1.827	0.22	5.93E-03
rs13280604	6	42,678,743	1971	1.406	1.095	1.804	0.22	7.49E-03
rs1530848	6	42,672,065	1964	1.399	1.094	1.79	0.22	7.53E-03
rs2196128	3	42,737,443	1971	1.374	1.083	1.744	0.23	8.83E-03
rs6997909	6	42,679,406	1971	1.388	1.083	1.78	0.22	9.67E-03
rs6474414	6	42,679,493	1971	1.388	1.083	1.78	0.22	9.67E-03
rs4736835	6	42,666,190	1971	1.388	1.081	1.782	0.22	1.03E-02
rs9298628	3	42,725,148	1968	1.376	1.078	1.756	0.21	1.04E-02
rs6474415	6	42,682,095	1970	1.383	1.079	1.774	0.22	1.05E-02
rs1451240	6	42,665,868	1970	1.381	1.075	1.774	0.22	1.14E-02
rs7004381	6	42,670,318	1971	1.376	1.072	1.767	0.22	1.24E-02
rs13273442	6	42,663,174	1970	1.376	1.071	1.767	0.22	1.25E-02
rs1955185	6	42,668,804	1971	1.371	1.069	1.759	0.22	1.30E-02
rs6474413	6	42,670,221	1971	1.371	1.068	1.759	0.22	1.31E-02
rs16891620	7	42,744,820	1970	1.424	1.057	1.918	0.13	2.00E-02
rs10958726	6	42,655,066	1971	1.342	1.043	1.726	0.21	2.20E-02
rs6474421	12	42,776,255	1969	1.497	1.049	2.137	0.07	2.62E-02
rs10958725	6	42,643,741	1968	1.31	1.021	1.682	0.22	3.37E-02
rs7012713	13	42,711,460	1970	1.648	1.008	2.695	0.04	4.66E-02

Bolded SNPs passed multiple test correction (p≥0.002). Maximum FTND is the score from 0-10, L95 and U95 is the 95% confidence interval, and the frequency is in the SAGE dataset.

a SNP that tags the previously discovered GWS signal in the region. Overall, 29 SNPs within this region are associated with FTND nicotine dependence with a p-value equal to or less than 0.002. This is consistent with previous studies showing an association in this

region. (See Appendix B for the entire association results in the region with each substance examined).

Conditional analyses suggest at least two independently associated SNPs with DSM-5 cocaine use disorder. To determine whether there was evidence for multiple independently associated variants at this locus contributing to risk for DSM-5 cocaine use disorder, the most significant SNP in the region (rs9298626) was added to the model as a covariate. Conditioning on rs9298626 eliminated the association with SNPs in LD bins 1 and 2, but the association remained for SNPs in bins 4 and 5. After including rs9298626 as a covariate, the top SNP associated with this phenotype was rs892413 ($p=3.57x10^{-3}$, OR=1.58, CI=1.23-2.04).

Examination of LD shows that the r^2 between rs9298626 and rs892413 is low, suggesting that these SNPs represent independent association signals ($r^2 = 0.01$; D'= 0.85). Because the minor allele frequency for rs9298626 is low, the r^2 will never be high but the D' indicates that the minor allele of this SNP is usually but not always on the background of one allele of rs892413. Neither of these SNPs is in significant LD with the previously identified genome-wide significant signal (rs1451240) associated with cigarette consumption and nicotine dependence (Rice et al., 2012; Thorgeirsson et al., 2010) in this region ($r^2 = 0.14$ between rs9298626 and rs1451240, $r^2= 0.35$ between rs892413 and rs1451240; Figure 9), although rs9298626 is in high LD ($r^2 = 0.94$) with rs4952, another SNP previously reported to be associated with nicotine dependence (Saccone et al. 2007). We find that rs9298626, rs892413 and rs1451240 correspond to different LD bins using $r^2=0.8$ as the threshold for defining the bins (Table 11). This is

consistent with recent results from our group showing, in an overlapping dataset, that the genome-wide significant signal in this region, tagged by rs1451240, is solely associated with nicotine dependence. Taken together, this suggests that these three SNPs represent different association signals in this region.

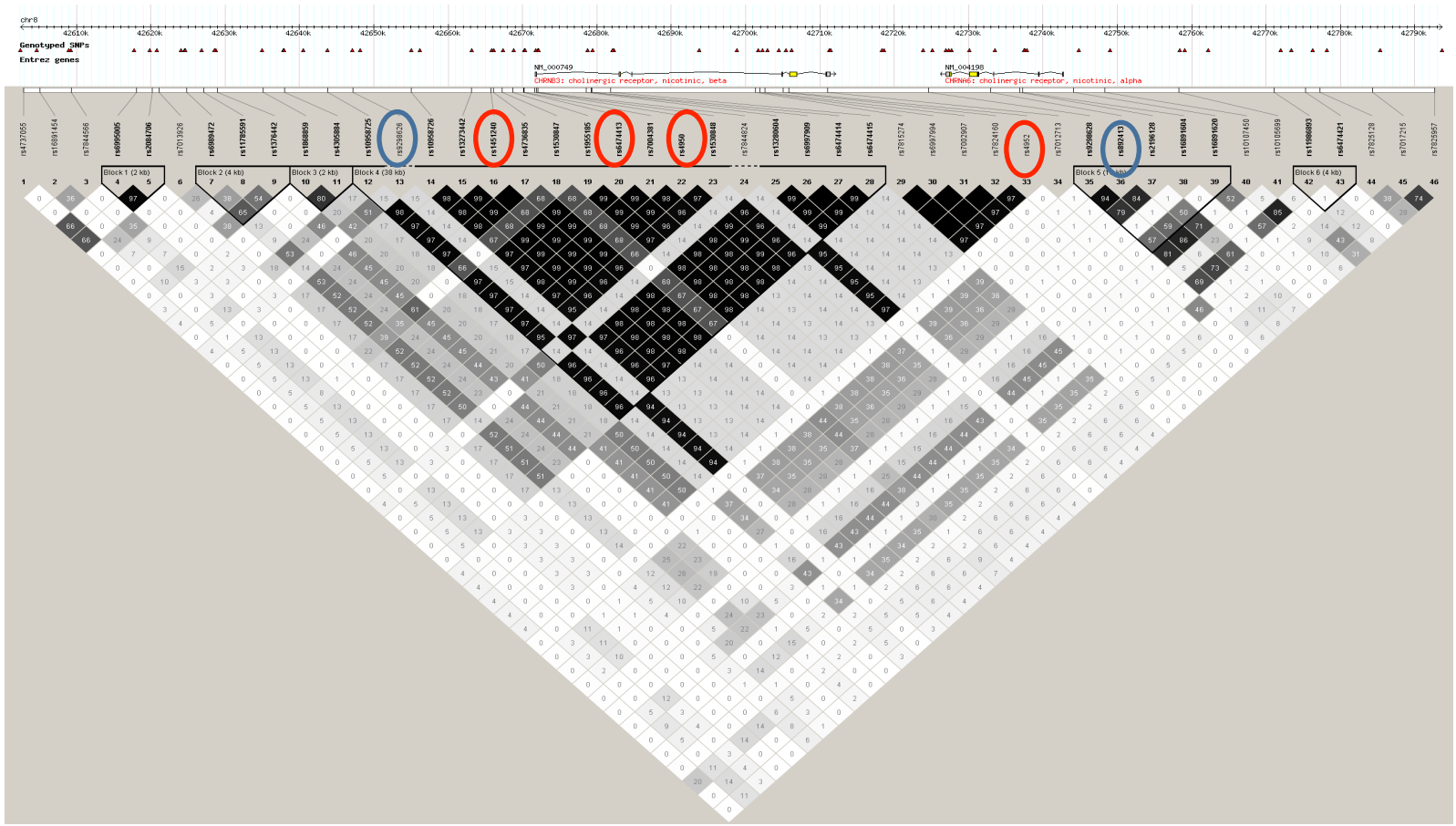

Figure 9. LD plot of the region. Values and colors reflect r-squared. Important SNPs include rs9298626, the top SNP associated with DSM-5 cocaine use disorder, the SNP that becomes significant when putting this SNP into the respective model as a covariate (rs892413), and SNPs representing the previously discovered association with nicotine dependence (rs4950, rs6474413, rs1451240 and rs4952). Red circles represent SNPs previously found associated with ND. Blue circles represent SNPs we find associated with DSM-5 cocaine use disorder .

Effects of cocaine exposure in the control population on association with rs9298626. To determine whether there is a significant effect of lifetime cocaine exposure on these associations, we compared the frequency of the minor allele of rs9298626 in those with DSM-5 cocaine use disorder, non-exposed non-dependent individuals and those who were exposed to cocaine but were unaffected (N=393 with 0-1 DSM-5 criteria), as well as those who were never exposed (N=1077). The minor allele frequency of rs9298626 was 9.9% among those with DSM-5 cocaine use disorder and 5.3% among subjects who have been exposed to cocaine but did not progress to cocaine use disorder. Non-exposed controls have an intermediate minor allele frequency of 7.7%, suggesting that both those with DSM-5 cocaine use disorder and those who were cocaine-exposed non-dependent controls show allele frequency differences from unselected controls (Table 12), although the difference in frequency of rs9298626 between exposed but unaffected controls and unexposed controls was not significant. When control subjects were restricted to those who had been exposed to cocaine but were unaffected (n=899 vs. 1976), the significance of the association between rs9298626 and DSM-5 cocaine use disorder was reduced but the odds ratio was unchanged ($p= 3.12x10^{-3}$, OR=2.68, CI=1.40-5.16), supporting the role of this SNP, or another SNP in LD with rs9298626, in risk for DSM-5 cocaine use disorder, even after accounting for cocaine exposure. We conclude that the minor allele of rs9298626 is correlated with cocaine use disorder, which is strengthened by the fact that there remains an association even when not considering unexposed individuals in the analysis.

Table 12

Characteristics of the sample broken down by: DSM-5 cocaine use disorder, exposed but unaffected, and non-exposed unaffected

	DSM-5 Cocaine Use Disorder EAs N=506	Exposed Unaffected EAs N=393	Non-exposed Unaffected EAs N=1077	Total N=1976
Age (years)				
mean age	36	38	39	38
<35	37%	26%	35%	34%
35-39	28%	27%	22%	25%
40-44	23%	35%	25%	26%
≥45	12%	12%	18%	13%
Male	62%	50%	34%	44%
Female	38%	50%	66%	56%
Nicotine Dependence				
FTND_0,1	10%	47%	68%	49%
FTND 2,3	12%	9%	9%	10%
FTND≥4	78%	44%	23%	41%
DSM-5 Alcohol Use Disorder				
Case	99%	88%	37%	42%
Control	1%	12%	63%	58%
rs9298626 minor allele carrier				
Yes	9.9%	5.3%	7.7%	7.8%
No	90.1%	94.7%	92.3%	92.2%

Stratified analysis show robustness of association with DSM-5 cocaine use disorder. Because SAGE is composed of individuals from three independent studies, each ascertained for a different substance dependence, we performed stratified analyses both by study and by nicotine dependence and alcohol dependence to determine if there existed a subset of subjects in which the association was most pronounced. The top SNP

associated with DSM-5 cocaine use disorder (rs9298626) in the whole SAGE dataset was significantly associated with DSM-5 cocaine use disorder in the COGA subset and showed a trend in the same direction in the FSCD and COGEND subsets. Furthermore, when individuals from the whole dataset were stratified by DSM-5 alcohol use disorder, or FTND nicotine dependence there was evidence of association between rs9298626 and cocaine use disorder in both groups (Table 13). This suggests that the observed associations are not an artifact of ascertainment and supports the hypothesis that this SNP is associated with DSM-5 cocaine use disorder and that the *CHRNB3-A6* locus is robustly associated with DSM-5 cocaine use disorder, regardless of comorbidity or ascertainment.

Table 13.

European-American DSM-5 Cocaine Use Disorder

	N	OR	95% CI	p-Value for rs9298626
Study				
COGEND	1077	1.77	0.67-4.62	0.25
COGA	599	2.62	1.25-5.52	0.01
FSCD	294	6.35	0.67-60.43	0.11
Smoking Status				
FTND Cases	814	2.31	1.06-5.02	0.03
FTND Controls	1156	2.59	1.22-5.50	0.01
Alcohol Status				
DSM-5 Cases	1522	2.41	1.46-3.98	0.0006
DSM-5 Controls	448	7.05	0.52-94.60	0.14

Results of stratified analyses in these groups for rs9298626 – controls included here are both those who are exposed and unaffected, as well as those who are unexposed.

Haplotype analysis suggest functional allele responsible for association with DSM-5 cocaine use disorder is in high LD with rs9298626. To further examine the relationship between our top variant identified for DSM-5 cocaine use disorder and the group of variants known to be associated with smoking, tagged by rs1451240, we performed haplotype-based association testing using rs9298626 and rs1451240. These two SNPs occur on three haplotypes that occur with a frequency >1% (Table 14).

Table 14.

Haplotypes observed in the SAGE GWAS European-American sample for DSM-5 cocaine use disorder.

Haplotypes			DSM-5 Cocaine Use Disorder	
rs9298626	rs1451240	Frequency	Odds Ratio	p-value
A	A	0.04	3.19	$1.35x10^{-4}$
A	G	<0.01	9.89	0.89
C	A	0.18	1.23	0.14
C	G	0.78	-	-

The grey box indicates the major allele for that SNP. P-values are denoting significance of that haplotype relative to the reference haplotype. SNPs are arranged in the order they occur on the chromosome. Covariates used are age, sex, study, DSM-5 alcohol symptom count, FTND total.

We chose rs1451240 because it was found to be genome-wide significant for nicotine dependence in a previous study using the SAGE GWAS data (Rice et al., 2012). The most common haplotype, composed of the major alleles of both SNPs, has a frequency of 78%. The haplotype associated with the highest risk for DSM-5 cocaine use disorder, has

a frequency of 4% and is composed of the minor alleles at both rs9298626 and rs1451240 (OR=3.19 p= $1.35x10^{-4}$, 95% CI=1.64-4.73). A haplotype composed of the major allele at rs9298626 and the minor allele at rs1451240 has a frequency of 18% but was not associated with DSM-5 cocaine use disorder (p=0.14). Because the frequencies and odds ratio of the haplotype with both minor alleles is nearly identical to that of the single SNP analysis for rs9298626 and the fact that the other haplotype containing the minor allele of rs1451240 is not associated with DSM-5 cocaine use disorder, we conclude that the functional allele responsible for this association is in high LD with the low frequency variant, rs9298626.

Association with rs9298626 and DSM-5 cocaine use disorder is independent of association with nicotine dependence. To further examine the relationship between nicotine dependence and DSM-5 cocaine use disorder associations in this region, we performed additional conditional analyses. In a linear regression model using age, sex, study, DSM-5 alcohol symptom count, total FTND score and rs1451240 genotype as covariates, the association with DSM-5 cocaine use disorder remained significant (Table 15). Lastly, the DSM-5 cocaine use disorder signal remains significant when conditioning on the two rare variants (rs35327613 and rs149775276) recently identified by our group to be associated with DSM-IV alcohol and cocaine dependence symptom count (Haller et al., 2013). This is not surprising given that these rare missense variants are present on the haplotypes containing the major allele of rs9298626, whereas the association reported here is with the minor allele. The fact that the association with cocaine use disorder remains when conditioning on the genome-wide significant signal

with nicotine dependence in the region, suggests that the association is independent and not acting through nicotine dependence.

Table 15

DSM-5 Cocaine Use Disorder with Age, Sex, DSM-5 Alcohol Symptom Count Study, FTND Total and rs1451240 as Covariates

SNP	N	OR	95% CI	P	P(age)	P(sex)	P(FTND Total)	P(DSM-5 Alc Sx)	P (COGEND)	P(COGA)	P (rs1451240)
rs9298626	1969	2.28	(1.32,3.95)	$3.22x10^{-3}$	$2.56x10^{-10}$	$8.07x10^{-3}$	$2.12x10^{-9}$	$1.55x10^{-27}$	$9.31x10^{-41}$	$1.43x10^{-21}$	0.17
rs4952	1970	2.10	(1.21,3.64)	$8.24x10^{-3}$	$1.71x10^{-10}$	0.01	$4.10x10^{-9}$	$7.67x10^{-28}$	$1.35x10^{-40}$	$2.98x10^{-21}$	0.13
rs892413	1970	1.35	(0.99,1.83)	$5.26x10^{-2}$	$8.69x10^{-11}$	$8.83x10^{-3}$	$6.43x10^{-9}$	$6.39x10^{-28}$	$4.28x10^{-40}$	$7.46x10^{-21}$	0.34

All SNPs remain associated with DSM-5 cocaine use disorder after conditioning on rs1451240.

DISCUSSION

We have shown, in genotyped data from European-Americans in the SAGE dataset, that there are at least two statistically independent signals associated with increased risk for DSM-5 cocaine use disorder in the region of the *CHRNB3-A6* nicotinic receptors on chromosome 8. Several SNPs representing the rs9298626 LD bin surpass the multiple test correction for the region with DSM-5 cocaine use disorder (p=0.002). rs9298626 is also associated with reduced risk for nicotine dependence (OR=0.47, 95% CI= 0.30-0.76, $p=1.80x10^{-3}$) in a univariate genetic analysis. This may be due in part to the fact that, in European ancestry populations, the minor allele of rs9298626 (MAF=0.04) occurs almost exclusively on the background of the more frequent minor allele (MAF=0.22) for the variant (rs1451240) previously reported to be genome-wide significantly associated with nicotine dependence (Table 14). Conditioning on rs1451240 had no effect on the association with DSM-5 cocaine use disorder (Table 15). This is not surprising because rs1451240 is not associated with DSM-5 cocaine use disorder in our data (Table 11). We find no association with DSM-5 alcohol use disorder in this dataset, which could indicate that the association at this locus is unique to DSM-5 cocaine use disorder and nicotine dependence. However, the sample size of those with DSM-5 alcohol use disorder could also be too small to detect an association.

LD bins tagged by rs9298626 and rs892413 each show association with DSM-5 cocaine use disorder in joint SNP analysis. Analyses conditioning on rs9298626 reveal that rs892413 is independently associated with DSM-5 cocaine use disorder. rs892413 is also associated with DSM-5 cocaine use disorder independent of the previously identified genome-wide significant association in the region with nicotine dependence (represented by rs1451240), providing

support for a direct effect of this SNP on higher DSM-5 cocaine use disorder risk, as opposed to acting through nicotine dependence risk (Tables 14 and 15).

The LD bin containing rs9298626 also contains rs4952 and rs4953, two low frequency synonymous variants in *CHRNB3* that have previously been reported to be associated with reduced risk for nicotine dependence (S. F. Saccone et al., 2007) and increased risk for bipolar disorder EAs (OR=1.7, 95%, CI= 1.2-2.4, p=0.001) (Hartz et al., 2011). Interestingly the association of rs4952/rs4953 with cocaine use disorder is in the same direction as the association with bipolar disorder (risk) but in the opposite direction to the association with nicotine dependence (protective) suggesting that *CHRNB3* variants have pleiotropic effects on substance use disorders and other psychiatric diseases. Many epidemiological studies have reported the common co-occurrence of bipolar disorder and substance dependence (Goodwin, Zvolensky, Keyes, & Hasin, 2012; Kenneson, Funderburk, & Maisto, 2013; Leverich & Post, 2006). Studies have also implicated shared genes with substance dependence and bipolar disorder (P. I. Lin et al., 2006; Post & Kalivas, 2013). It is therefore possible that the high frequency of bipolar disorder and substance dependence comorbidity is in part due to common underlying genetic risk factors such as the risk alleles in the *CHRNB3-A6* locus reported here.

Our group has previously reported that rare missense variants in *CHRNB3* increase risk for cocaine dependence (Haller et al., 2013). The results reported here demonstrate that low frequency and common alleles within the *CHRNB3* locus are also associated with increased risk of DSM-5 cocaine use disorder. Cocaine dependence has now been associated with SNPs in two different nicotinic receptor gene clusters, on chromosomes 8 and 15 (Grucza et al., 2008; Haller et al., 2013). It is interesting, however, that the variant on chromosome 15, within *CHRNA5* is

associated with decreased risk for cocaine dependence, while rs9298626 and other variants in the *CHRNB3-A6* region are associated with higher (OR =2.62) risk for DSM-5 cocaine use disorder.

Furthermore, similar to the observation on chromosome 15, the chromosome 8 locus is associated with opposing effects on the risk for cocaine dependence and nicotine dependence. The *CHRNB3-A6* locus is associated with decreased risk for nicotine dependence and increased risk for DSM-5 cocaine use disorder. In contrast, in *CHRNA5,* the same variant*, D398N* (rs16969968), increases risk for nicotine dependence and decreases risk for cocaine dependence. Furthermore, our data suggest that different but overlapping SNPs may explain the cocaine and nicotine dependence associations in *CHRNB3-A6* rather than a single SNP causing opposing effects as was seen on chromosome 15. These results suggest that *CHRNA5* and *CHRNB3* demonstrate pleiotropic effects on substance dependence risk.

Nicotinic acetylcholine receptors (nAChR) are expressed in multiple types of neurons, and have been shown to modulate reward response for several substances (Grucza et al. 2008). For example, work in animals suggests that activation of α3β4 nAChR can increase cocaine self-administration (Hansen & Mark, 2007). Because comorbidity between substance dependencies is so high, it is plausible that these receptors could play a role in addiction to multiple substances.

Most drugs of abuse act on the mesolimbic dopamine-containing receptors in the brains of humans and many other mammals. Among other functions, this system is known to regulate motivation (Koob, 1996; Wise, 1996) and has similar effects across mammalian species (Tanda, Pontieri, & Di Chiara, 1997). Activation and reinforcement of this system is a necessary part of drug abuse (Koob, 1996). The dopaminergic system is therefore crucial to addiction, however other neurotransmitters besides dopamine affect the mesolimbic system, especially acetylcholine (Hansen & Mark, 2007).

The biological connection between these two systems could be related to the reversal of the odds ratio for rs16969968, which is protective for cocaine dependence but a risk factor for nicotine dependence, as well as our observation that in the *CHRNB3-A6* locus on chromosome 8, there are variants associated with protection against nicotine dependence, in addition to variants associated with risk for DSM-5 cocaine use disorder and bipolar disorder. Since the finding in *CHRNA5* on chromosome 15 is the only association with cocaine dependence to be successfully replicated, it would be interesting to examine the *CHRNB3-A6* region in other datasets that have assessed cocaine dependence phenotypes, as well as to analyze datasets of other ethnicities.

Currently, there is no evidence that either of the variants reported here are correlated with SNPs that have known functional consequences. However, rs4952 and rs4953 are both synonymous variants in *CHRNB3* and may therefore have some, as yet unknown effect on transcription or translation of *CHRNB3* mRNA. Overall, our findings underscore the comorbidity among drug dependencies and corroborate the role of nicotinic receptors in cocaine-related phenotypes. This study represents one of only a few to implicate specific variants in cocaine dependence phenotypes and the first to implicate low frequency variants within the *CHRNB3-A6* locus in risk for DSM-5 cocaine use disorder.

CHAPTER 4

EVOLUTION, NATURAL SELECTION, AND NICOTINE DEPENDENCE

ABSTRACT

Much of the evolution of human behavior remains a mystery, including how certain disadvantageous behaviors have become so prevalent. Nicotine addiction is one such phenotype. Several loci have been implicated in nicotine related phenotypes including the nicotinic receptor clusters (*CHRN*s) on chromosomes 8 and 15, and the nicotine metabolizing gene *CYP2A6*. Here we use 1000 Genomes sequence data from 3 populations (Africans, Asians and Europeans) to examine whether natural selection has occurred at these loci. Further, we test the hypothesis that any selection that has occurred at these loci is not related to nicotine addiction, but rather is associated with cognitive phenotypes such as memory and learning. To test for selection, we have used multiple complimentary methods that include Tajima's D, integrated haplotype score (iHS) and Ka/Ks ratio. While each method has its own strengths and weaknesses, together they capture selection at multiple time-depths. Our results from these statistics provide evidence for strong selection in the nicotinic receptor cluster on chromosome 8, previously found to be significantly associated with nicotine and cocaine dependencies. This selection is occurring at certain loci associated with increased risk for nicotine dependence but decreased risk for cocaine dependence. This is intriguing given recent studies that have shown that cocaine addicts have a dampened, and therefore maladaptive, reward response to social interaction. This suggests the possibility that selection is acting to decrease risk of cocaine addiction at the expense of an increased risk for nicotine dependence. We also find evidence of weaker, but still detectable selection, acting on the region containing the *CHRNA5* nicotinic receptor gene on chromosome 15 that is genome wide significant for risk for nicotine dependence. To examine the possibility

that this selection is related to memory and learning, we performed an association in exome chip data from the Collaborative Studies on the Genetics of Alcoholism (COGA) dataset with neuropsychological phenotypes. We find one SNP that passes multiple test correction for the phenotype of WAIS digit symbol. This test captures aspects of reaction time and memory, suggesting that this locus is associated with both nicotine dependence and cognition.

INTRODUCTION

Nicotine dependence is the leading cause of preventable death in the USA. It has been noted that some populations experience higher levels of addiction than others but the reason for this is not understood. Multiple studies have demonstrated a genetic component to nicotine addiction (Berrettini et al., 2008; Bierut et al., 2007; S. F. Saccone et al., 2007; Thorgeirsson et al., 2010) but little is known about the role of natural selection in shaping the genetic components of nicotine addiction. Such knowledge could help us understand the genetic and behavioral nature of addiction and ultimately facilitate the design and delivery of appropriate interventions to reduce nicotine addiction.

It has been estimated that approximately 10% of the genome has been affected by linkage due to recent selective sweeps (Williamson et al., 2007). However it is often difficult to determine the actual phenotype that was the target of selection. This is particularly true when the phenotype being examined has no obvious beneficial impact or has an apparently deleterious effect but is nonetheless undergoing positive selection. In this case, additional mechanisms and/or alternative explanations must be sought for the existence of selection on the gene of interest.

An example of such a situation occurs in the gene for hemoglobin. In homozygous form, the 'sickle cell' allele, HbS, drives the formation of malformed red cells, which aggregate to cause blockages of blood flow to numerous organs including the brain. This results in organ damage and strokes, severely shortening the lifespan of the individual. Nonetheless, the HbS allele is maintained in the gene pool in regions where malaria is endemic because in heterozygous form it provides protection against malaria (for a review see (Ashley-Koch, Yang, & Olney, 2000)).

A second example derives from a SNP in a p53 binding site in the *KITLG* gene. This SNP has undergone positive selection in Caucasians despite its association with an increased risk of several types of cancer (Zeron-Medina et al., 2013). The authors hypothesize that this is due to the role of *KITLG* in the tanning response, which provides a protective effect against UV light. Thus, for purposes of selection, a beneficial effect of the gene in one setting can over-ride an apparently deleterious effect of that gene in another context.

The case of nicotine addiction represents a similar conundrum. Several genetic variants that modify susceptibility or resistance to nicotine dependence have been identified by genome-wide association studies (GWAS) (Berrettini et al., 2008; Bierut et al., 2007; Thorgeirsson et al., 2010). Perhaps not surprisingly, the loci identified in these studies mainly include genes encoding neuronal nicotinic cholinergic receptors (*CHRNs*) and a nicotine-metabolizing gene (*CYP2A6*).

But why would nature seemingly select for this trait, especially given the fact that it is believed that nicotine has not been a part of our evolutionary history long enough, and in large enough quantities, for its effects to be visible in our genomes? One hypothesis is that selection

acted on a more primary phenotype and the effect on nicotine addiction was secondary and incidental, a genetic phenomenon termed hitchhiking.

Nicotine is known to have an enhancing effect on cognitive performance. For example one study showed that nicotine enhanced the reorientation of attention in visuospatial tasks (Thiel et al., 2005). A second study used fMRI to show that nicotine altered neuronal activity responsible for increased attention and arousal (Kumari et al., 2003). Nicotinic receptors are also important in the functional impairments found in Alzheimer's disease (AD), as it has been shown that AD patients have a reduction in nicotinic receptor binding sites (Newhouse, Potter, Kelton, & Corwin, 2001). Furthermore, epidemiological evidence suggests that smokers have a significantly lower incidence of symptoms and diagnoses of AD and Parkinson's disease than non-smokers (Fratiglioni & Wang, 2000; Tyas, 1996). With regard to genetics, Rigbi et al (Rigbi et al., 2008) found an association between cognitive function and variants within the genes encoding *A2, A4, A5, A7, A9, A10, B2* and *B3* nicotinic receptors, as well as with several related haplotypes. More recently, Winterer et al (Winterer et al., 2010) found an association between risk variants for nicotine dependence in *CHRNA5* and lower cognitive performance scores. They suggested that these individuals would choose to use nicotine more often than non-risk allele carriers to overcome this lower cognitive performance.

Evidence from nicotinic receptor knockout mice also supports a role for these receptors in memory and learning, as well as anxiety levels. *CHRNA7* knockout mice have impaired reaction times (Hoyle, Genn, Fernandes, & Stolerman, 2006) and decreased procedural learning (Young, Meves, Tarantino, Caldwell, & Geyer, 2011). Interestingly, *CHRNA6* knockout mice show that this receptor plays a role in nicotinic modulation of dopaminergic transmission. These knockout mice lose high-affinity binding of alpha-conotoxin-MII (αCtxMII), a compound that

blocks nicotine-induced dopamine release. Combined with data showing that α3 knockout mice do not show changes in αCtxMII binding, this suggests that *CHRNA6*, and not *CHRNA3*, preferentially combines with the β2 subunit in dopaminergic neurons (Drago, McColl, Horne, Finkelstein, & Ross, 2003). Based on the foregoing observations, we hypothesize that the nicotinic receptors may have been targets of recent selection and that this selection is related to the role of nicotinic receptors in memory and learning.

Selective forces leave informative signatures in the human genome. There are several tests designed to measure departures from neutrality that can be indicative of selection at a locus or loci that have sufficiently high linkage disequilibrium with the target site. Each test gives the most accurate results when functioning within optimal parameters for variables such as time depth and allele frequencies. In this chapter, three different methods were used for detecting natural selection at loci relevant to nicotine dependence, specifically the *CHRNA5-A3-B4* region on chromosome 15q25 and the *CHRNB3-A6* region on chromosome 8p11. The data provide strong evidence for selection in the *CHRNB3-A6* region and moderate evidence for selection in the *CHRNA5-A3-B4* region. However, there is only a modest correlation between nicotine dependence and score on the Wechsler Adult Intelligence Scale (WAIS) Digit Symbol test. We discuss the alternate possibility that because the effects of SNPs associated with risk of nicotine dependence are independently associated with protection from cocaine dependence that it is the latter phenotype that may be driving selection.

METHODS

To determine whether the nicotinic receptor loci are under selection, we used Tajima's D, integrated haplotype score (iHS), and the ratio of nonsynonymous to synonymous substitutions

(Ka/Ks) to examine the landscape of natural selection at three loci previously demonstrated to harbor genetic variants contributing to the risk of nicotine dependence. These tests have different but complementary strengths. Tajima's D test functions best on recently completed selective sweeps. There are many variables that contribute to how far in the past a sweep can be detected, such as how extreme the sweep was in the first place. Both the mutation rate and the recombination rate affect it as well and vary widely across the genome making generalizations difficult. By contrast, integrated haplotype score iHS functions best for detecting sweeps in progress with alleles at intermediate frequencies, mainly in the range of or after the separation of European, Asian and African populations, during the agricultural phase of human evolution (Voight, Kudaravalli, Wen, & Pritchard, 2006). Ka/Ks can detect older selection in orthologous protein coding regions that has occurred between lineages. Together, these tests should be able to detect selection at multiple time-depths, as well as both sequence-based and haplotype-based selection.

We utilized 1000 Genomes data for Tajima's D and iHS analyses. The populations were grouped into EUR (GBR, TSI, CEU, FIN), ASN (CHS, CHB, JPT), and AFR (YRI, LWK, ASW). For Ka/Ks we used Genbank reference mRNA sequences for human (hg19) and chimp (PanTro4) as the outgroup. All methods were calculated for the same regions: the *CHRNA5-A3-B4* region on chromosome 15q25, the *CHRNB3-A6* region on chromosome 8p11, the *LCT* region as a positive control on chromosome 2q21, and several intergenic negative control regions where applicable.

Tajima's D test. Tajima's D was calculated using the program Variscan (Hutter, Vilella, & Rozas, 2006). To run Variscan, one must input certain parameters such as how large of a sliding window to use, and what bp increment to move each time. After an exploratory data

analysis of window size, we concluded a sliding window size of 1000 bp, and window increments of 100 bp were best for these data. This is because any smaller of a window and there were often too few SNPs in a window to calculate properly, and any larger of a window and it made it much harder to narrow down specific SNPs that may be the ultimate target of selection. Variscan then outputs a file giving the Tajima's D value for every window of the specified bp size on the sliding scale (Hutter et al., 2006). These values were then superimposed onto graphs of the regions.

Integrated Haplotype Score. iHS is a measure of whether a SNP is on an unusually long haplotype carrying the ancestral or derived allele. In other words, it compares the rate of haplotype decay between haplotypes carrying either the ancestral or derived allele at a given site, called the core SNP. The haplotype decay is calculated until the extended haplotype homozygosity (EHH) reaches 0.05. EHH is defined as "the probability that two randomly chosen chromosomes carrying the core haplotype of interest are identical by descent for the entire interval from the core region to point x" (Sabeti et al., 2002). This can be thought of as haplotype transmission with no recombination. Haplotypes whose core SNP is under selection will be unusually long compared to those evolving neutrally. Long haplotypes with derived alleles are indicated by negative iHS values and those with ancestral alleles are indicated by positive iHS values. Under neutrality, extreme scores are distributed throughout the genome, however under selection, they are clustered across the selected region (Voight et al., 2006). iHS can be clearer than Tajima's D, but also somewhat nuanced to work with because of the dependence on local recombination rate. However, it is a good method for detecting directional selection, especially a sweep that is in its early phases. We used the program WHAMM to calculate this statistic (Voight et al., 2006).

First, we extracted the desired regions from the 1000 Genomes dataset. We then selected known SNPs within each region, and extracted a region of plus or minus 2000 SNPs around that SNP, except in the case of *CHRNB3-A6* where we selected plus or minus 2500 SNPs. We constructed recombination maps using cM maps provided by the SHAPEIT2 program (Delaneau et al., 2012). Ancestral alleles were determined using the latest version of Seattleseq (http://snp.gs.washington.edu/SeattleSeqAnnotation137/). Phased haplotypes were coded as number of copies of the derived allele. All positions in which the derived allele could not be determined unambiguously (i.e. C/G or A/T SNPs) as well as those without known chimp alleles were removed from further analyses. All analyses were run on each population separately. As iHS is greatly influenced by SNP allele frequency, iHS values from WHAMM were standardized using the average and standard deviation of all SNPs on chromosome 15 and 8 binned by allele frequency such that the average iHS value for each bin after standardization was identical. We excluded SNPs with a minor allele frequency less than 5% because low frequency SNPs are difficult to normalize accurately. After removing problematic SNPs discussed above, extracting just the desired gene regions, and removing those with MAF of <0.05, there were ~150-350 SNPs per region, depending on the population. Standardization was done separately for each population using population specific averages and standard deviations. iHS values were then superimposed onto graphs of the regions.

The haplotype on which a beneficial allele resides tends to be significantly longer than the other haplotypes at the same frequency in the population when adjusted for the recombination background. However, long haplotypes tend to occur in regions with low recombination, and these can be confused with genuine genomic signals of positive selection (Liu et al., 2013). This is why WHAMM attempts to control for recombination by requiring the

input of a cM map. The map we used here was the cM map for imputation available on the website for the program SHAPEIT2.

Candidate regions of positive selection were defined as genomic regions containing an uncharacteristic clustering of SNPs with high iHS statistics. This was quantified as the proportion of SNPs with |iHS| > 2 in the four regions of interest. Candidate regions of positive selection were identified as containing any SNP with an iHS score of |iHS| > 2, as this corresponds to the top ~5% of all scores. The iHS value at a SNP "measures the strength of evidence for selection acting at or near that SNP" however does not provide a formal significance test (Voight et al. 2006).

Ka/Ks_Calculator. Nonsynonymous and synonymous substitution rates (denoted as Ka and Ks, respectively), or rather their ratio (Ka/Ks), is indicative of neutral evolution when the two are equal, negative (purifying) selection when Ka is less than Ks, and positive selection when Ka is greater than Ks (Zhang et al., 2006). This is a way of comparing selective pressures at homologous genes. However since not all datasets have the same degree of substitutions, having only one possible model is not optimal. This is why the Ka/Ks_Calculator program (freely available at https://code.google.com/p/kaks-calculator/) provides a model selection step to choose a best-fit model. Here we have selected the Li-Wu-Luo (LWL) approximate method because unlike other similar methods such as the Jukes-Cantor (JC) method which makes the simplifying assumption that nucleotide substitution occurs randomly, the LWL method essentially weights Ka (number of nonsynonymous substitutions per nonsynonymous site) and Ks (number of synonymous substitutions per synonymous site) by whether or not the site is fourfold degenerate (always synonymous), twofold degenerate (one third synonymous and two-thirds nonsynonymous), or nondegenerate (always nonsynonymous).

A site is degenerate if all possible changes are nonsynonymous or nonsense. A twofold degenerate site is one where one of the three possible changes is synonymous, and a fourfold degenerate site is one where all possible positions at the third codon result in a synonymous change. The purpose of this classification is to estimate nonsynonymous and synonymous rates of substitution separately (W. H. Li, Wu, & Luo, 1985). The LWL method involves three steps: 1) count the numbers of synonymous and nonsynonymous sites; 2) calculate the numbers of synonymous and nonsynonymous substitutions and 3) correct for multiple substitutions (Zhang et al., 2006).

Ka/Ks does not provide a measure of selective pressure at each nucleotide within a gene, but rather selective pressure on the gene as a whole. Thus, it cannot give detailed information on which SNPs in the gene are associated with selection. However, this statistic can give us information on selection events that have occurred in the more distant past than either Tajima's D or iHS, such as changes that have occurred between species lineages. This is somewhat intuitive because determining Ka/Ks for multiple sequences sampled from the same population will not yield information about fixation events along independent lineages, but rather polymorphisms segregating in a population (Kryazhimskiy & Plotkin, 2008). Limitations to this approach are that only selection in protein coding regions can be calculated and since selective pressures are averaged over the region, it can take quite a strong selective signal to be detected. Also, since balancing selection does not cause an amino acid change, it cannot be detected (Yang & Bielawski, 2000). Human and chimp sequences were aligned using ClustalW (Larkin et al., 2007) and input into the Ka/Ks calculator as pairwise alignments. The pairwise alignments for each gene analyzed were used to calculate Ka and Ks.

Association Analyses. DNA samples were collected as part of the Collaborative Study of the Genetics of Alcoholism (COGA). All members of the COGA sample underwent a semi-structured interview, the SSAGA, which assessed alcohol, cocaine and nicotine use as well as comorbid psychiatric conditions. The COGA sample utilized in this study consisted of family GWAS data from 2102 European-Americans.

COGA administered a variety of neuropsychological tests to its subjects including the three used here: Wechsler Adult Intelligence Scale (WAIS) Block Design, WAIS Digit Symbol, and WAIS Information. In total there were 1247 European-Americans with these neuropsychological phenotypes. However, the overlap between this number and those with family GWAS data was 492. Therefore, our analyses were done using 492 subjects.

SNPs in the region of the nicotinic receptor clusters on chromosomes 8 and 15 were tested for association with the neuropsychological phenotypes in European-Americans from the COGA study using linear regression as implemented in the GWAF package in R using age, sex and FTND score as covariates.

RESULTS

Tajima's D Analyses. Tajima's D is a method of addressing the frequencies of variant sites, based on the expectation that under neutrality, different estimates of expected diversity (θ) should be equal. Tajima's D tests for a skew in the frequency spectrum by comparing two estimates of θ – the number of segregating sites (S), and pairwise nucleotide diversity (π) (Tajima, 1989). Extreme positive values can indicate either balancing selection or population subdivision, and extreme negative values can indicate positive selection or population growth (Jobling, 2004). If the same skew is detected across the genome, the effect is likely due to

demography, whereas if the skew is localized to a few loci, selection is more likely to be occurring.

In a review, Garrigan & Hammer (Garrigan & Hammer, 2006) have combined published data for Tajima's D values from 65 autosomal loci. They find the mean value for Africans is slightly negative (-0.20) and for non-Africans is slightly positive (0.13). Overall, the values range from -2 to 2. As such, we have taken Tajima's D values above 2 or below -2 to count as extreme values, as this represents the 95% confidence interval of values in our data.

Figures 10 through 13 show histograms of the Tajima's D values for our negative controls (Figure 10), our positive control (Figure 11), the chromosome 15 locus which contains the genes encoding the α3, ß4 and α5 subunits of the nicotinic receptor (Figure 12) and the chromosome 8 locus which contains the genes for the α6 and ß3 subunits of the nicotinic receptor (Figure 13). Each figure shows the results for our three populations, AFR, ASN and EUR, in a separate histogram. The positions of relevant SNPs demonstrated in GWAS studies to be associated with risk for (Bierut et al., 2008; N. L. Saccone et al., 2009; S. F. Saccone et al., 2007; Spitz et al., 2008; Thorgeirsson et al., 2008; Weiss et al., 2008) or protection from (Hoft et al., 2009; Rice et al., 2012; S. F. Saccone et al., 2007; Zeiger et al., 2008) nicotine dependence are shown at the bottom of the histogram for the nicotinic receptor loci. We have included on these figures rs1051730, a synonymous change, in *CHRNA3* as it is in the same LD bin as rs16969968 and has in the past had significant p-values, although here it does not quite reach the cutoff threshold.

Controls. Ten 10 kb intergenic regions served as our controls. As shown in Figure 10, most of the values fall between -1.5 to 1.5, with few exceptions, mostly in the ASN population.

Overall, the proportion of sliding windows with extreme Tajima's D values in the control, intergenic regions was less than 2% in all our populations (Table 16). The ASN population had the highest proportion of extreme Tajima's D values with 1.7%, followed by EUR with 1.1% and AFR with 0%.

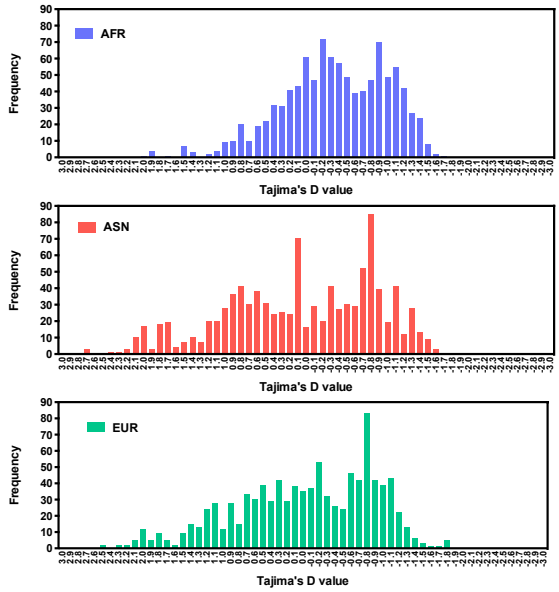

Figure 10. Histogram of the frequencies of Tajima's D values in each population for the ten 10 kb control regions.

For our positive control we used the *LCT* gene. *LCT* encodes the protein, lactase, and mutations in the region give rise to lactase persistence. The lactase persistence phenotype is actually due to changes in a regulatory region that enhances the expression of *LCT*. This enhancer is located in intron 13 of the neighboring gene, *MCM6* (Jones et al., 2013). Multiple SNPs have been found to be enhancers in different populations. We have marked the most common SNP, rs4988235, on the figures. Thus, we examined both genes in this region for evidence of selection. As can be seen from Figure 11, both the ASN and EUR populations showed a significant number of windows with extreme Tajima's D values. For Europeans, the

proportion of windows with extreme values was 8.6 %. The corresponding value for Asians was 6.5 % (Table 16). Essentially all of these values were extreme on the positive, rather than the negative, side. These values are significantly different than the relevant negative controls ($p=1.2x10^{-15}$ and $p=4.2x10^{-8}$ for EUR and ASN, respectively) and are consistent with the occurrence of balancing selection or an ongoing sweep at the *LCT* locus in the EUR and ASN populations.

In contrast to what was seen with EUR and ASN, the histogram for AFR shows few windows with extreme Tajima's D values. Indeed, the proportion of windows with extreme values was <1% and was not significantly different from the negative controls. This indicates a lack of selection at this locus in AFR.

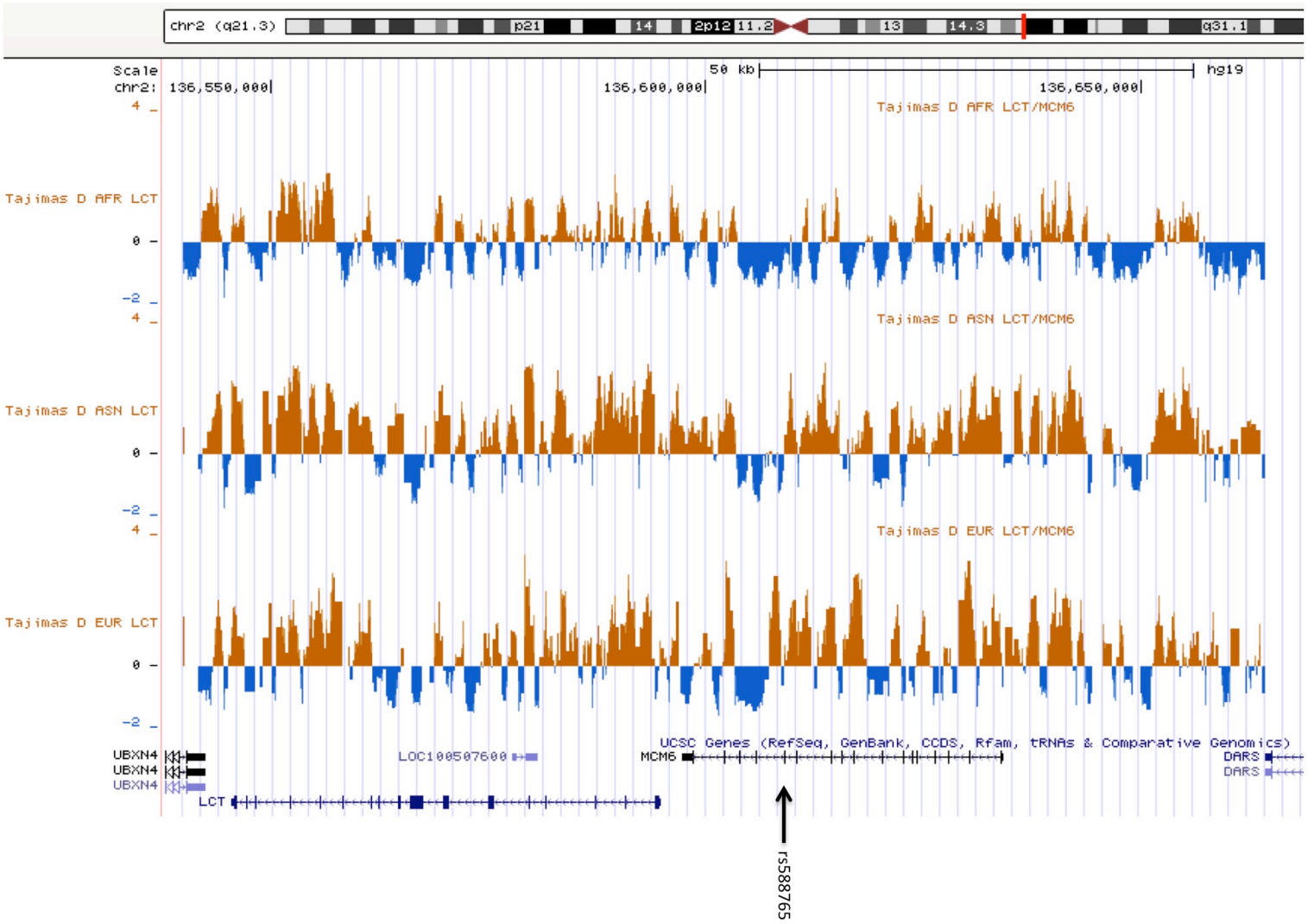

Figure 11. Tajima's D values for the Lactase region on chromosome 2 plotted with genes below

CHRNA3-B4-A5. We next applied the Tajima's D test to the *CHRNA3-B4-A5* gene cluster on chromosome 15. In this region, we analyzed a 120 kb stretch of DNA from bp 78,840,000 to 78,960,000, which includes the nicotinic receptor cluster as well as ~18 kb upstream. This was done to ensure that we included the large region upstream of the cluster that has been associated with regulation of the level of expression of the receptors (J.-C. Wang et al., 2013). We refer to this as the *CHRNA3-B4-A5* region even though it encompasses sequence outside of the nicotinic receptor genes proper.

Figure 12 shows the distribution of Tajima's D values in the sliding windows in the *CHRNA3-B4-A5* region in AFR, ASN and EUR. As can be seen from the figure, all three populations showed numerous windows with extreme values, though the distribution of these windows differed somewhat among the different populations, reflecting population differences in LD. Overall, 2.3% of the sliding windows in the AFR population showed extreme Tajima's D values (Table 16), with all of them being positive. This represents a significant difference from the negative control for the AFR population ($p=2.8x10^{-8}$). Noteworthy is the observation that in AFR, these extreme positive values were concentrated almost exclusively in *CHRNA3* or the intergenic region between *CHRNA3* and *CHRNB4*. This is interesting given the fact that rs1051730, a synonymous change in *CHRNA3,* has often had equally strong associations with nicotine dependence as rs16969968, leading some to wonder about the true importance of this SNP. However at least in these analyses, rs1051730 did not have extreme values.

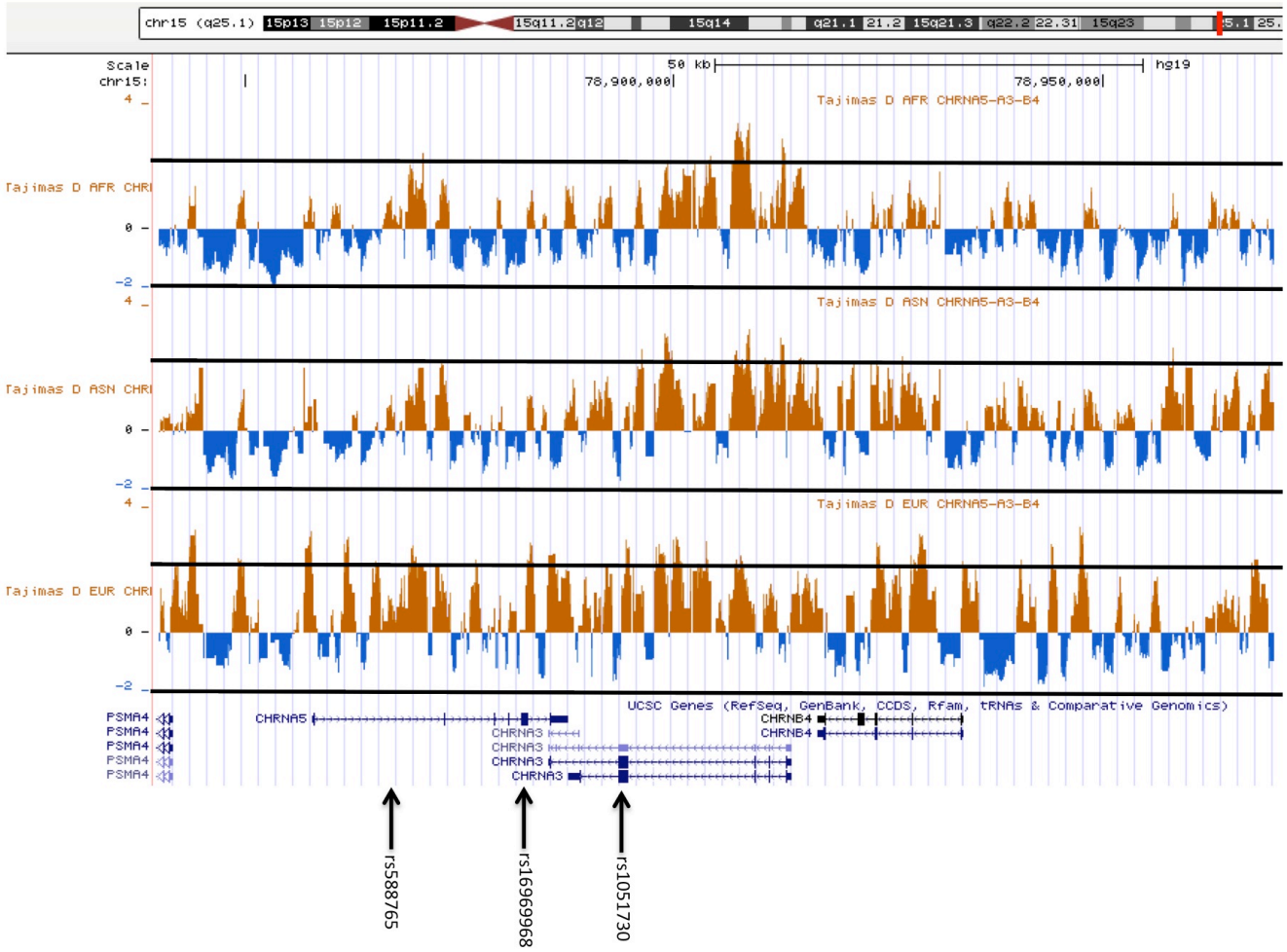

Figure 12. Tajima's D values for the nicotinic receptor cluster on chromosome 15 plotted with gene positions and relevant SNPs below.

The ASN population also had a significant concentration of extreme positive Tajima's D values in the *CHRNA3* and the intergenic region between *CHRNA3* and *CHRNB4*. In the ASN population, 7.1% of the windows overall exhibited extreme values (Table 16), which was again significantly different from the relevant negative controls ($p=1.4x10^{-9}$).

Table 16.

Summary of extreme Tajima's D values in Each Region in Comparison with Negative Controls

Chromosomal Position	Gene(s)	Size (kb)	Population	# windows with TajD>2 or <-2	Proportion	P-value	Most Extreme TajD
78,840,000 -	*CHRNA5-A3-B4*	120	EUR	173/1273	13.60%	***<2x10^{-16}***	3.23
78,960,000			ASN	91/1273	7.10%	***1.4x10^{-9}***	3.14
			AFR	30/1294	2.30%	***2.8x10^{-8}***	3.41
42,520,000 -	*CHRNB3-A6*	110	EUR	13/1066	1.22%	1	2.67
42,630,000			ASN	1/973	<1%	1	2.06
			AFR	23/1097	2.10%	***2.4x10^{-7}***	3.65
136,540,000 -	*LCT/MCM6*	140	EUR	98/1136	8.63%	***1.2x10^{-15}***	3.27
136,664,100			ASN	76/1161	6.55%	***4.2x10^{-8}***	2.82
			AFR	5/1240	<1%	0.051	2.13
Control Regions	n/a	100	EUR	11/982	1.1%	-	2.43
(10 x 10kb)			ASN	17/983	1.7%	-	2.69
			AFR	0/1010	0.0%	-	1.80

P-values are from a Fisher's Exact Test comparing each population for each genic region to the negative control values for that same population. Significant p-values indicate a value that is significantly different than the negative controls.

In the EUR, 13.6% of the sliding windows had extreme Tajima's D values (Table 16), again with the majority being extreme in the positive direction. This is significantly different than the negative control for the EUR population ($p=2x10^{-16}$) (Table 16). Unlike AFR and ASN, in which the extreme Tajima's D values were concentrated in *CHRNA3* or the intergenic region between *CHRNA3* and *CHRNB4*, EUR showed sliding windows with extreme positive Tajima's D values more evenly distributed throughout the region examined, although some windows of extreme values were found in *CHRNA3* that overlapped with similar blocks in AFR and ASN.

Two main LD bins in this region have previously been shown to be associated with risk for nicotine dependence. The first is a large bin of ~86 kb that includes rs16969968, which has been shown in GWAS studies to exhibit the most significant association with risk of nicotine

dependence in the entire genome. As previously mentioned, rs16969968 yields a non-synonymous change in the gene encoding the α5 subunit of the nicotinic receptor on chromosome 15 that gives rise to a decrease in the binding affinity of the α5 receptor for nicotine and acetylcholine (J. C. Wang et al., 2009; J.-C. Wang et al., 2013). This SNP is found at a frequency of ~35% in European-Americans but is nearly absent in African-Americans and Asians.

Table 17

Tajima's D Statistics on SNPs in First LD Bin on Chromosome 15

SNP	dbSNP Func Annot	EUR Frequency A1	Taj D Range - EUR	ASN Frequency A1	Taj D Range - ASN	AFR Frequency A1	Taj D Range - AFR
rs72740955	intergenic	0.37	(1.53, 2.07)	0.03	(0.72, 0.97)	0.09	(0.88, 1.00)
rs2036527	intergenic	0.37	(0.11, 0.11)	0.03	(-0.57, -0.54)	0.16	(-1.15, 0.20)
rs55853698	CHRNA5 5'UTR	0.37	(0.56, 1.93)	0.04	(-0.02, 1.00)	0.06	(0.24, 0.56)
rs17486195	CHRNA5 intronic	0.36	(1.59, 1.64)	0.03	(-0.17, 0.14)	0.11	(-0.10, -0.07)
rs17486278	CHRNA5 intronic	0.36	(0.39, 0.70)	0.31	(0.04, 0.60)	0.28	(0.42, 0.47)
rs72740964	CHRNA5 intronic	0.36	(0.78, 1.63)	0.03	(0.21, 0.98)	0.04	(0.08, 0.56)
rs951266	CHRNA5 intronic	0.36	(-0.22, 0.22)	0.03	(-0.84, -0.37)	0.08	(-0.89, -0.74)
rs16969968	CHRNA5 missense	0.36	**(2.00, 2.33)**	0.03	(0.43, 1.01)	0.02	(0.29, 0.62)
rs1051730	CHRNA3 synonymous	0.36	(1.40, 1.90)	0.03	(0.41, 0.78)	0.09	(-0.29, -0.23)
rs1317286	CHRNA3 intronic	0.36	(0.32, 1.27)	0.09	**(2.00, 2.56)**	0.24	(0.73, 1.36)
rs12914385	CHRNA3 intronic	0.4	**(2.35, 2.55)**	0.32	(1.85, 2.01)	0.2	(1.13, 1.35)
rs114205691	CHRNA3 intronic	0.64	(0.80, 1.65)	0.68	(0.00, 0.81)	0.8	(1.75, 1.97)
rs8040868	CHRNA3 synonymous	0.59	(0.87, 1.37)	0.62	(1.63, 2.17)	0.63	(-0.27, 0.13)
rs55958997	intergenic	0.39	(1.03, 1.66)	0.06	(0.74, 1.23)	0.3	(-0.41, -0.38)
rs72743158	CHRNB4 intronic	0.38	(0.26, 1.27)	0.02	(0.79, 1.34)	0.03	(0.20, 0.20)
rs55988292	CHRNB4 intronic	0.39	(-1.52, -0.24)	0.02	(-0.35, 028)	0.14	(-0.93, -0.61)

Tajima's D ranges for all SNPs in the genome-wide significant LD bin on chromosome 15 tagged by rs16969968 associated with increased risk of nicotine dependence. Significant values are bolded and highlighted in yellow.

As shown in Table 17, this SNP lies in a window of extreme Tajima's D values for the EUR population. A second SNP in this LD bin, rs12914385, is also located in a block of Tajima's D values >2 in EUR. An adjacent SNP in this same LD bin, rs1317286, is located in a block of sliding windows that show extreme positive Tajima's D values in ASN. No SNPs in this LD bin were in regions of extreme Tajima's D values, consistent with the lower overall

signal for selection in this region in AFR. Together, these data are consistent with the possibility that selection at these SNPs related to risk of nicotine dependence may contribute to the signal for selection seen at these locations in EUR and ASN. Because the extreme Tajima's D values in this bin are positive, we can conclude that at least in EUR, there has been recent balancing selection in the region of rs16969968.

A second LD bin in this cluster is tagged by rs588765 (see Figure 12). The minor allele of rs588765 is associated with decreased mRNA expression of *CHRNA5* and a decreased risk for nicotine addiction (J. C. Wang et al., 2009). Other SNPs in this LD bin are listed in Table 18. Nearly all of them are in introns of either *CHRNA5* or *CHRNA3,* suggesting they affect splicing or expression, not the function of these subunits.

Within the EUR population, 5 SNPs in this bin, including 3 consecutive SNPs correlated to *CHRNA5* were in windows with extreme Tajima's D values (Table 18). Two additional SNPs in this bin were also found in regions with high Tajima's D values in EUR, specifically rs6115470 and rs2869546. This latter SNP, rs2869546, which is in an intron of *CHRNA3,* had an extreme value in all three populations examined. This SNP is classified as a DNase hypersensitivity site in B-lymphocyte, medulloblastoma and neuroblastoma cell types. All Tajima's D values were extreme in the positive direction, suggesting that balancing selection, focused on gene expression, may be occurring at this locus across all populations.

Table 18

Tajima's D Statistics on SNPs in Second LD Bin on Chromosome 15

SNP	dbSNP Func Annot	EUR Freq A1	Taj D Range - EUR	ASN Freq A1	Taj D Range - ASN	AFR Freq A1	Taj D Range - AFR
rs4275821	intergenic	0.33	**(2.02, 2.32)**	0.14	(0.61, 088)	0.18	(0.74, 0.99)
rs588765	CHRNA5 intronic	0.4	**(2.18, 2.22)**	0.15	(0.00-0.26)	0.25	(-0.30, -0.30)
rs6495306	CHRNA5 intronic	0.4	**(2.07, 2.32)**	0.15	(-0.42, 0.06)	0.26	(-0.16, 0.31)
rs495090	CHRNA5 intronic	0.35	(1.47, 2.06)	0.2	(1.17, 1.37)	0.4	(0.77, 1.23)
rs680244	CHRNA5 intronic	0.4	(-0.77, 0.83)	0.21	(-0.60, 0.24)	0.4	(-0.63, -0.06)
rs621849	CHRNA5 intronic	0.41	(0.36, 1.35)	0.21	**(2.07, 2.08)**	0.4	(0.47, 1.13)
rs11637635	CHRNA5 intronic	0.65	(1.61, 2.18)	0.86	(-0.28, -0.13)	0.78	(-0.60, -0.35)
rs481134	CHRNA5 intronic	0.6	(0.14, 0.77)	0.85	(-0.48, 0.19)	0.75	(-0.87, -0.61)
rs555018	CHRNA5 intronic	0.4	(-0.84, 0.78)	0.15	(-0.39, 0.04)	0.26	(-0.77, -0.23)
rs647041	CHRNA5 intronic	0.4	(-1.06, -0.14)	0.18	(-1.05, 0.03)	0.24	(-1.08, -0.60)
rs615470	CHRNA5 3' UTR	0.65	**(2.54, 2.57)**	0.83	(-1.20, -0.75)	0.67	(-1.09, -0.60)
rs6495307	CHRNA3 intronic	0.6	(-1.17, 0.14)	0.82	(0.55, 0.84)	0.66	(-0.83, -0.41)
rs62010327	CHRNA3 intronic	0.35	(0.15, 1.60)	0.14	(0.42, 0.85)	0.08	(0.65, 0.97)
rs12901300	CHRNA3 intronic	0.4	(-1.37, -0.26)	0.19	(-0.71, 0.11)	0.34	(-0.47, 0.32)
rs3743077	CHRNA3 intronic	0.4	(1.86, 2.62)	0.18	(-0.25, 0.14)	0.11	(-0.96, -0.77)
rs62010328	CHRNA3 intronic	0.34	(1.86, 1.90)	0.14	(-0.54, -0.25)	0.07	(-1.17, -0.96)
rs2869546	CHRNA3 intronic	0.36	**(2.08, 2.40)**	0.18	**(2.31, 2.63)**	0.29	**(3.17, 3.41)**
rs4366683	CHRNA3 intronic	0.54	(-0.72, 0.13)	0.45	(1.50, 2.00)	0.5	(0.70, 1.54)
rs58643100	CHRNA3 intronic	0.46	(-0.72, 0.13)	0.26	(1.50, 2.00)	0.41	(0.70, 1.54)

Tajima's D ranges for all SNPs in the genome-wide significant LD bin on chromosome 15 tagged by rs588765 and associated with decreased risk of nicotine dependence. Significant values are bolded and highlighted in yellow.

There is a third, smaller LD bin in this region, whose SNPs are also associated with *CHRNA5* expression (J. C. Wang et al., 2009). However, we did not explore this LD bin here due to the fact that after correcting for the first, larger bin associated with *CHRNA5* expression, this third bin was no longer significant.

CHRNB3-A6. For the *CHRNB3- A6* region, we examined a rather larger segment of the genome upstream of the gene cluster. This was done so as to include several upstream SNPs that have previously been shown to exhibit associations with nicotine dependence or cocaine dependence (Sadler et al., 2014). Thus, we have included from bp 42,520,000 to 42,630,000.

Figure 13 shows the histogram of Tajima's D values across the sliding windows in the *CHRNB3-A6* region. In AFR, there were several windows with extreme Tajima's D values upstream of the *CHRNB3* gene and in the intergenic region between *CHRNB3* and *CHRNA6*.

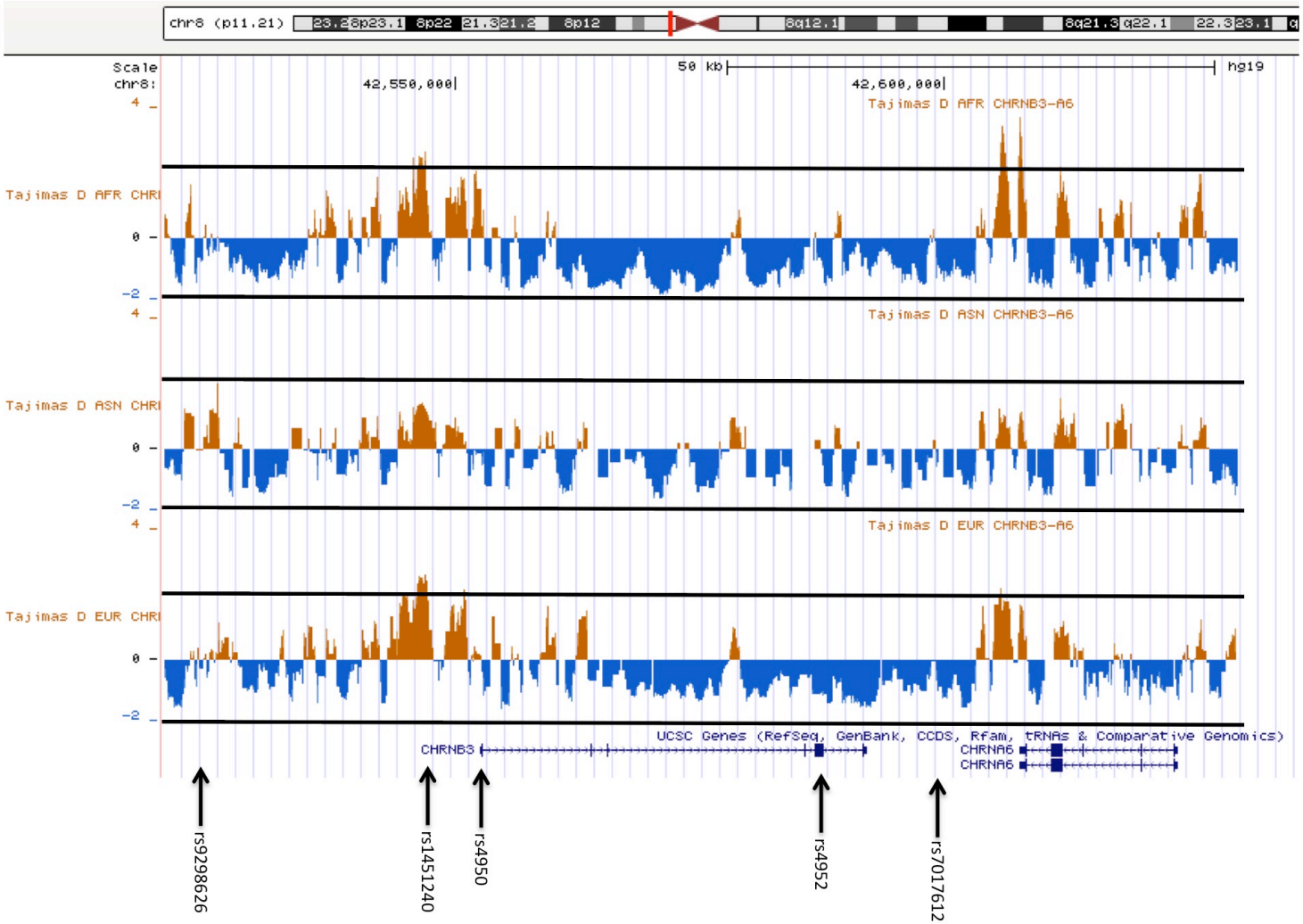

Figure 13. Tajima's D values for the nicotinic receptor cluster on chromosome 8 plotted with gene positions and relevant SNPs below.

While the overall proportion of sliding windows with extreme values was only 2.1% in AFR (Table 16), this was nonetheless significantly different than the relevant negative control ($p=2.4x10^{-7}$).

EUR also exhibit a cluster of sliding windows showing extreme positive Tajima's D values just upstream of *CHRNB3* and overlapping with that cluster in AFR. There was also a cluster of positive iHS values in the same intergenic region between *CHRNB3* and *CHRNA6* as seen in AFR. However, over the entire region, there was no significant difference from the EUR negative control (Table 16). While ASN did have a region of positive Tajima's D values upstream of *CHRNB3*, none of the sliding windows were in the "extreme" range and overall, there was no significant difference between this locus and the negative controls (Table 16).

Overall, the evidence for selection is weak in this cluster. However, the concentration of windows with extreme Tajima's D values upstream of *CHRNB3* is noteworthy in the context of risk for nicotine addiction. A GWAS study by Rice et al. (Rice et al., 2012) found that a SNP in this region, rs1451240, was associated at the genome-wide significant level with protection from nicotine dependence using the score on the Fagerstrom Test for Nicotine Dependence (FTND) as the phenotype. The LD bin tagged by rs1451240 spans ~66 kb and several other SNPs in this bin have been associated with protection from nicotine dependence, although no others are significant at the genome-wide level (Rice et al., 2012). The data in Table 19 show that in EUR and AFR, four adjacent SNPs from this LD bin, including rs1451240, were present in sliding windows with extreme Tajima's D values. These data suggest that these SNPs associated with a nicotine dependence phenotype may be undergoing balancing selection or positive selection in these two populations.

Table 19

Tajima's D Values for GWS LD bin on Chromosome 8 Associated with Nicotine Dependence

SNP	dbSNP Func Annot	EUR Freq A1	Taj D Range - EUR	ASN Freq A1	Taj D Range - ASN	AFR Freq A1	Taj D Range - AFR
rs1979140	intergenic	0.77	(-0.75, -0.34)	0.82	(-1.21, -1.15)	0.28	(-1.26, -0.99)
rs7816726	intergenic	0.77	(0.33, 1.00)	0.81	(-0.02, 0.70)	0.28	(0.08, 0.17)
rs10958726	intergenic	0.23	(0.1, 0.68)	0.18	(-0.85, -0.15)	0.63	(-0.70, 0.18)
rs7842601	intergenic	0.23	(0.17, 0.62)	0.18	(-0.33, -0.17)	0.63	(0.63, 1.33)
rs13273442	intergenic	0.23	(-0.39, 0.69)	0.18	(-0.60, -0.45)	0.63	(-0.75, 0.04)
rs9792277	intergenic	0.77	(1.56, 1.80)	0.81	(1.06, 1.17)	0.33	(1.38, 2.32)
rs1451239	intergenic	0.77	**(2.41, 2.58)**	0.82	(1.34, 1.40)	0.34	**(2.18, 2.51)**
rs1451240	intergenic	0.23	**(2.41, 2.54)**	0.18	(1.33, 1.41)	0.67	**(2.18, 2.46)**
rs1901281	intergenic	0.77	**(2.37, 2.68)**	0.81	(1.26, 1.33)	0.28	**(2.18, 2.19)**
rs4736835	intergenic	0.23	**(2.48, 2.68)**	0.18	(1.17, 1.26)	0.66	(1.47, 2.63)
rs1955185	intergenic	0.77	(1.49, 1.67)	0.82	(0.64, 0.69)	0.28	(1.08, 1.59)
rs13277254	intergenic	0.23	(0.89, 1.27)	0.18	(0.64, 0.97)	0.63	(1.11, 1.29)
rs13277524	intergenic	0.77	(0.64, 0.89)	0.82	(0.24, 0.97)	0.28	(0.43, 1.29)
rs6474412	intergenic	0.77	(1.09, 1.43)	0.82	(0.53, 0.66)	0.34	(1.01, 1.49)
rs6474413	intergenic	0.77	(1.86, 2.19)	0.82	(0.13, 0.22)	0.28	(1.15, 1.33)
rs7004381	intergenic	0.23	(0.68, 1.86)	0.18	(-0.15, 0.22)	0.63	(0.29, 1.15)
rs6985052	intergenic	0.77	(-0.37, 0.68)	0.82	(-0.85, -0.16)	0.29	(-1.04, -0.29)
rs4950	CHRNB3 5'UTR	0.23	(-0.39, 0.12)	0.18	(-0.44, -0.16)	0.8	(0.38, 1.55)
rs9643891	CHRNB3 intronic	0.77	(-0.34, 0.13)	0.82	(0.60, 0.72)	0.15	(-0.65, -0.16)
rs9643853	CHRNB3 intronic	0.77	(-1.04, -0.34)	0.82	(-0.60, 0.60)	0.15	(-1.55, -0.65)
rs13280604	CHRNB3 intronic	0.23	(0.47, 1.68)	0.18	(0.07, 0.50)	0.8	(0.04, 0.83)
rs6997909	CHRNB3 intronic	0.77	(0.70, 0.70)	0.82	(-0.25, 0.11)	0.15	(-1.00, -1.00)
rs6474414	CHRNB3 intronic	0.77	(0.40, 0.40)	0.82	(-0.60, -0.21)	0.15	(-0.76, -0.12)
rs6474415	CHRNB3 intronic	0.23	(1.26, 1.35)	0.18	(-0.12, 0.22)	0.85	(-0.85, -0.52)
rs4236926	CHRNB3 intronic	0.77	(-0.04, -0.03)	0.81	(-0.14, -0.11)	0.15	(-0.54, -0.54)
rs16891561	CHRNB3 intronic	0.77	(-1.28, -0.90)	0.81	(-0.89, 0.72)	0.15	(-1.08, -0.48)
rs55828312	CHRNB3 intronic	0.24	(-(0.89, -0.74)	0.19	(-1.13, -0.31)	0.83	(-0.85, 0.67)

Tajima's D ranges for all SNPs in the genome-wide significant LD bin (r2 = 0.9) on chromosome 8 tagged by rs1451240 and associated with decreased risk of nicotine dependence. Significant values are bolded and highlighted in yellow.

Integrated Haplotype Score (iHS) Analyses. iHS is a measure of whether a SNP is on an unusually long haplotype carrying the ancestral or derived allele. In other words, it compares the rate of haplotype decay between haplotypes carrying either the ancestral or derived allele at a given site, called the core SNP. Haplotypes whose core SNP is under selection will be unusually long compared to those evolving neutrally. Long haplotypes with derived alleles are indicated by negative iHS values and those with ancestral alleles are indicated by positive iHS values. Under neutrality, extreme scores are distributed throughout the genome, however under selection, they

are clustered across the selected region (Voight et al., 2006). iHS can be clearer than Tajima's D, but also somewhat nuanced to work with because of the dependence on local recombination rate. However, it is a good method for detecting directional selection, especially in a sweep that is in its early phases. We used the program WHAMM to calculate this statistic (Voight et al., 2006).

Candidate regions of positive selection were defined as genomic regions containing an uncharacteristic clustering of SNPs with high iHS statistics. A high iHS score was defined as an |iHS| > 2, as this corresponds to the top ~5% of all scores. This was quantified as the proportion of SNPs with |iHS| > 2 in the four regions of interest. The iHS value at a SNP "measures the strength of evidence for selection acting at or near that SNP", however, it does not provide a formal significance test (Voight et al., 2006).

Controls. Our negative controls for the iHS analysis were the same 10 inter-genic regions used as negative controls in the Tajima's D analyses. The histograms for these regions are shown in Figure 14 and the data are summarized in Table 20. For this analysis, AFR had the highest proportion of extreme values, with 4%, followed by ASN with 1.6% and EUR with 1.3%.

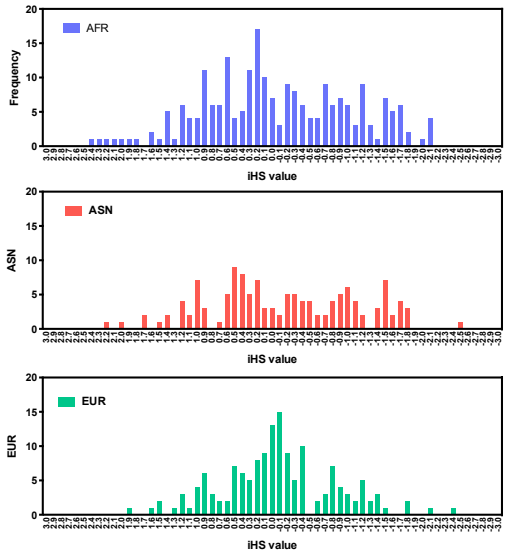

Figure 14. Histogram of the frequencies of iHS values in each population for the ten 10 kb control regions.

Table 20

Summary of Extreme iHS Values in Each Region in Comparison with Negative Controls

Chromosomal Position	Gene(s)	Size (kb)	Population	# SNPs with iHS>2 or <-2	Proportion	P-Value	Most Extreme iHS
78,840,000 -	*CHRNA5-A3-B4*	120	EUR	1/181	<1%	0.91	2.03
78,960,000			ASN	3/224	1.3%	0.74	-2.72
			AFR	16/266	5.6%	0.23	2.72
42,520,000 -	*CHRNB3-A6*	110	EUR	26/110	23.6%	$\mathbf{1.2x10^{-7}}$	3.11
42,630,000			ASN	3/116	2.59%	0.45	2.17
			AFR	93/357	26.1%	$\mathbf{2.3x10^{-10}}$	3.79
136,540,000 -	*LCT/MCM6*	140	EUR	111/138	80.4%	$\mathbf{<2x10^{-16}}$	4.57
136,664,100			ASN	20/219	9.1%	**0.004**	-2.66
			AFR	9/278	3.2%	0.76	-2.44
Control Regions	n/a	100	EUR	2/149	1.3%	-	-2.49
(10 x 10kb)			ASN	2/129	1.6%	-	-2.56
			AFR	9/223	4.0%	-	2.32

Proportion of extreme his values within the three regions sampled, including intergenic regions, as well as the control regions. P-values are from a Fisher's Exact Test comparing each population for each genic region to the negative control values for that same population. Significant p-values indicate a value that is significantly different than the negative controls.

We again used *LCT/MCM6* as our positive control. Figure 15 shows the histograms of iHS values in the *LCT/MCM6* region for EUR, ASN and AFR populations. The AFR population shows few extreme values (only 3.2%) and this does not differ significantly from the negative control. Likewise, in ASN, there are few windows of extreme values although in total, the number is different from the negative control (Table 20). In EUR, the overall average proportion of extreme values for this region is 80.4%. This high proportion of extreme values in genic regions is evident from the histogram. The clustering of extreme iHS values in the genic areas of this region is consistent with what is known about large-scale positive selection at this locus in the EUR population (Jones et al., 2013). This demonstrates the validity of this approach for identifying genes undergoing selection.

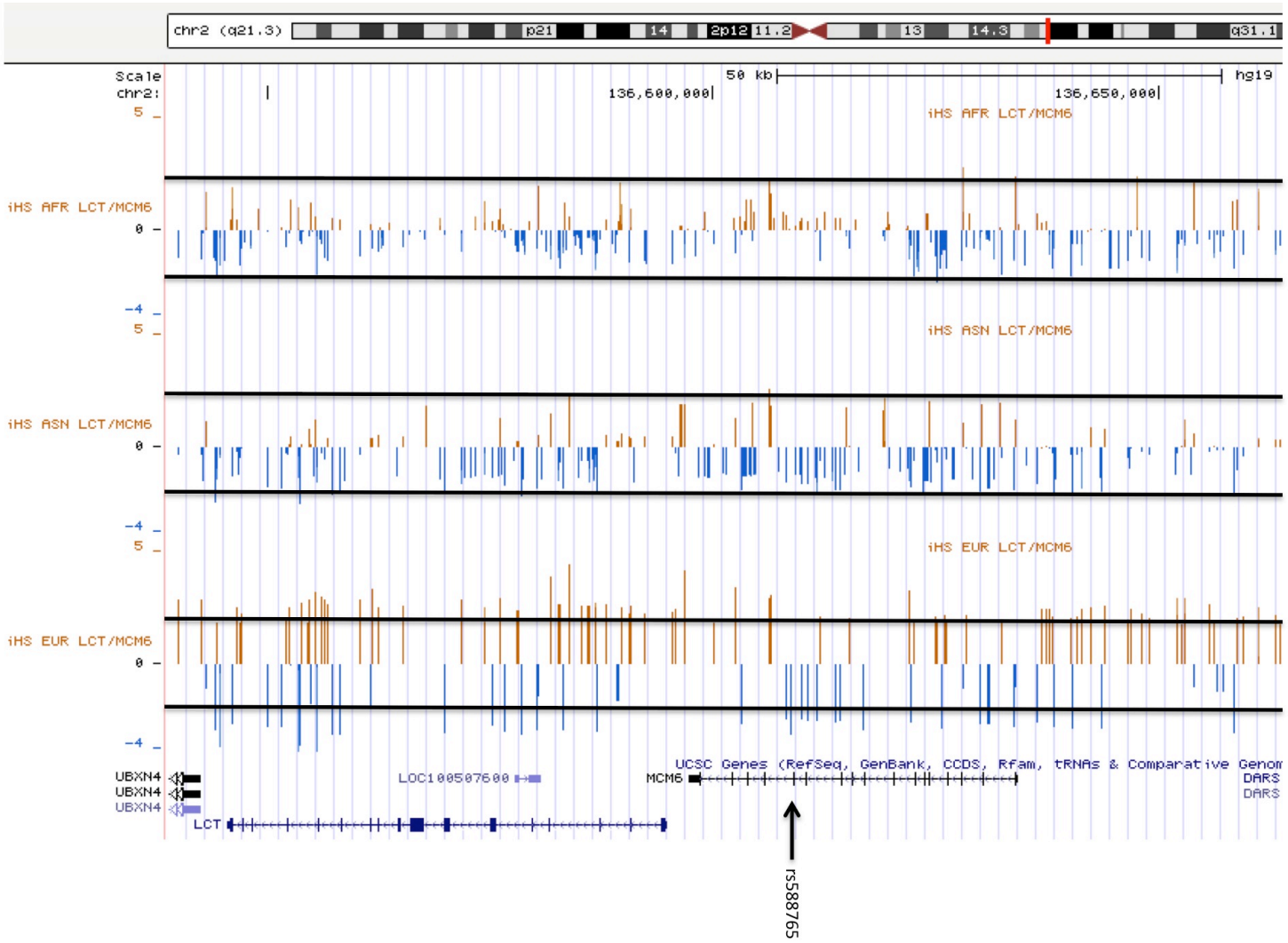

Figure 15. iHS values for the Lactase region on chromosome 2 plotted with genes below.

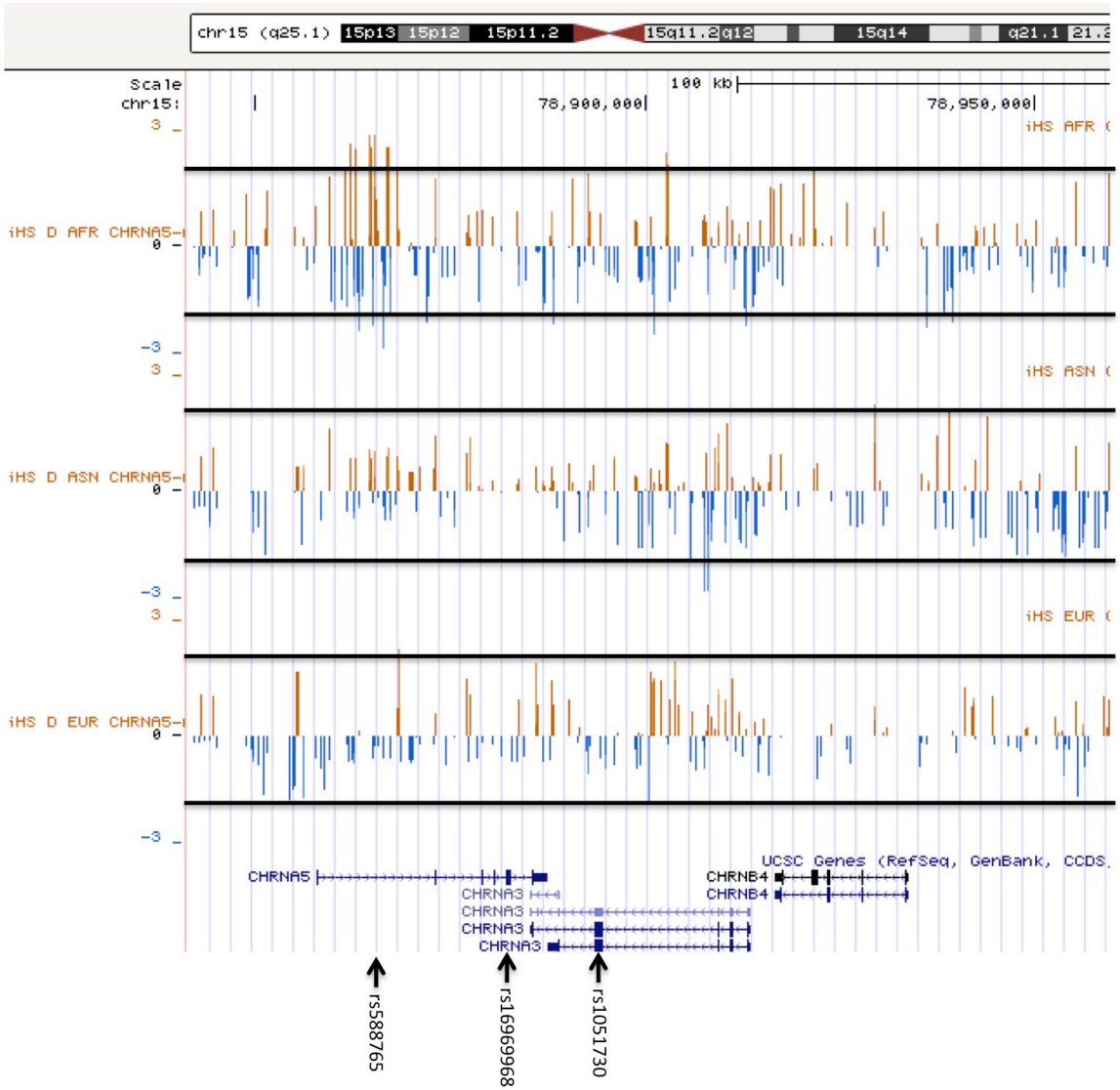

Figure 16. iHS values for the nicotinic receptor cluster on chromosome 15 plotted with gene positions and relevant SNPs below.

CHRNA3-B4-A5. Figure 16 shows the histogram of iHS values across the *CHRNA3-B4-A5* locus on chromosome 15. The summary statistics are given in Table 20. Among the 3 populations studied, none showed a proportion of extreme iHS values that was significantly different from the relevant negative control. In addition, none of the windows with extreme values included any of the SNPs previously found to be associated with nicotine dependence (Tables 21 and 22).

Table 21

iHS Values for SNPs in First LD Bin on Chromosome 15

SNP	dbSNP Func Annot	EUR Frequency A1	iHS value EUR	ASN Frequency A1	iHS value ASN	AFR Frequency A1	iHS value AFR
rs72740955	intergenic	0.37	-0.74	0.03	-1.25	0.09	0.44
rs2036527	intergenic	0.37	-0.5	0.03	-0.85	0.16	1.48
rs55853698	CHRNA5 5'UTR	0.37	-0.65	0.04	-0.31	0.06	1.05
rs17486195	CHRNA5 intronic	0.36	-0.62	0.03	-0.82	0.11	0.48
rs17486278	CHRNA5 intronic	0.36	-0.62	0.31	-0.45	0.28	0.72
rs72740964	CHRNA5 intronic	0.36	-0.73	0.03	-0.57	0.04	0.14
rs951266	CHRNA5 intronic	0.36	-0.65	0.03	-0.5	0.08	0.91
rs16969968	CHRNA5 missense	0.36	-0.71	0.03	-0.44	0.02	-0.01
rs1051730	CHRNA3 synonymous	0.36	-0.64	0.03	-0.67	0.09	0.69
rs1317286	CHRNA3 intronic	0.36	-0.8	0.09	0.78	0.24	-0.43
rs12914385	CHRNA3 intronic	0.4	-0.39	0.32	0.36	0.2	0.39
rs114205691	CHRNA3 intronic	0.64	1.53	0.68	0.15	0.8	0.23
rs8040868	CHRNA3 synonymous	0.59	1.54	0.62	0.25	0.63	0.67
rs55958997	intergenic	0.39	-1.08	0.06	0.23	0.3	-0.45
rs72743158	CHRNB4 intronic	0.38	-0.79	0.02	-0.83	0.03	1.11
rs55988292	CHRNB4 intronic	0.39	-0.26	0.02	-0.94	0.14	-0.7

iHS values for all SNPs in the genome-wide significant LD bin on chromosome 15 associated with increased risk of nicotine dependence. Significant values are bolded and highlighted in yellow.

Voight et al. (Voight et al., 2006) have previously shown that there is definite clustering of extreme iHS values in regions where SNPs show evidence of selection. In our analyses, there was some clustering of extreme iHS values in *CHRNA5* in the AFR population. In this group, there are 5.6% extreme values when the entire *CHRNA3-B4-A5* cluster is considered but 18.7% extreme values for the *CHRNA5* genic region. The latter value is significantly different than the negative control, suggesting that this gene

may be undergoing selection. Unfortunately, none of the SNPs previously associated with nicotine dependence were in windows that showed extreme iHS values so the data do not address the question of whether the selection is related to nicotine addiction phenotype.

Table 22

iHS Values for SNPs in the Second LD Bin on Chromosome 15

SNP	dbSNP Func Annot	EUR Freq A1	iHS value - EUR	ASN Freq A1	iHS value - ASN	AFR Freq A1	iHS value - AFR
rs4275821	intergenic	0.33	-0.39	0.14	-0.02	0.18	-0.13
rs588765	CHRNA5 intronic	0.4	-0.32	0.15	-0.2	0.25	0.44
rs6495306	CHRNA5 intronic	0.4	-0.32	0.15	-0.2	0.26	0.41
rs495090	CHRNA5 intronic	0.35	-0.66	0.2	0.48	0.4	0.06
rs680244	CHRNA5 intronic	0.4	-0.27	0.21	0.62	0.4	-0.11
rs621849	CHRNA5 intronic	0.41	-0.27	0.21	0.57	0.4	-0.21
rs11637635	CHRNA5 intronic	0.65	1.51	0.86	0.99	0.78	0.35
rs481134	CHRNA5 intronic	0.6	1.1	0.85	1.06	0.75	0.19
rs555018	CHRNA5 intronic	0.4	-0.22	0.15	0.01	0.26	0.97
rs647041	CHRNA5 intronic	0.4	-0.24	0.18	0.24	0.24	0.77
rs615470	CHRNA5 3' UTR	0.65	1.46	0.83	0.66	0.67	-0.08
rs6495307	CHRNA3 intronic	0.6	0.96	0.82	0.71	0.66	-0.07
rs62010327	CHRNA3 intronic	0.35	-0.74	0.14	-0.78	0.08	-0.99
rs12901300	CHRNA3 intronic	0.4	-0.3	0.19	0.26	0.34	0.91
rs3743077	CHRNA3 intronic	0.4	-0.27	0.18	0.34	0.11	0.33
rs62010328	CHRNA3 intronic	0.34	-0.8	0.14	-0.69	0.07	-0.54
rs2869546	CHRNA3 intronic	0.36	-0.18	0.18	-0.09	0.29	0.66
rs4366683	CHRNA3 intronic	0.54	0.44	0.45	-0.96	0.5	0.02
rs58643100	CHRNA3 intronic	0.46	0.63	0.26	0.84	0.41	0.24

iHS values for all SNPs in the genome-wide significant LD bin on chromosome 15 tagged by rs588765 and associated with decreased risk of nicotine dependence. Significant values are bolded and highlighted in yellow.

CHRNB3-A6. Figure 17 shows a histogram of the iHS values for the *CHRNB3-A6* cluster on chromosome 8. The summary statistics are shown in Table 20. As can be seen in the figure, the AFR population shows a cluster of extreme iHS values with a highly significant overall proportion of 26.1% extreme values in this genetic region (Table 20). Thus, as with the nicotinic receptor cluster on chromosome 15, the *CHRNB3-*

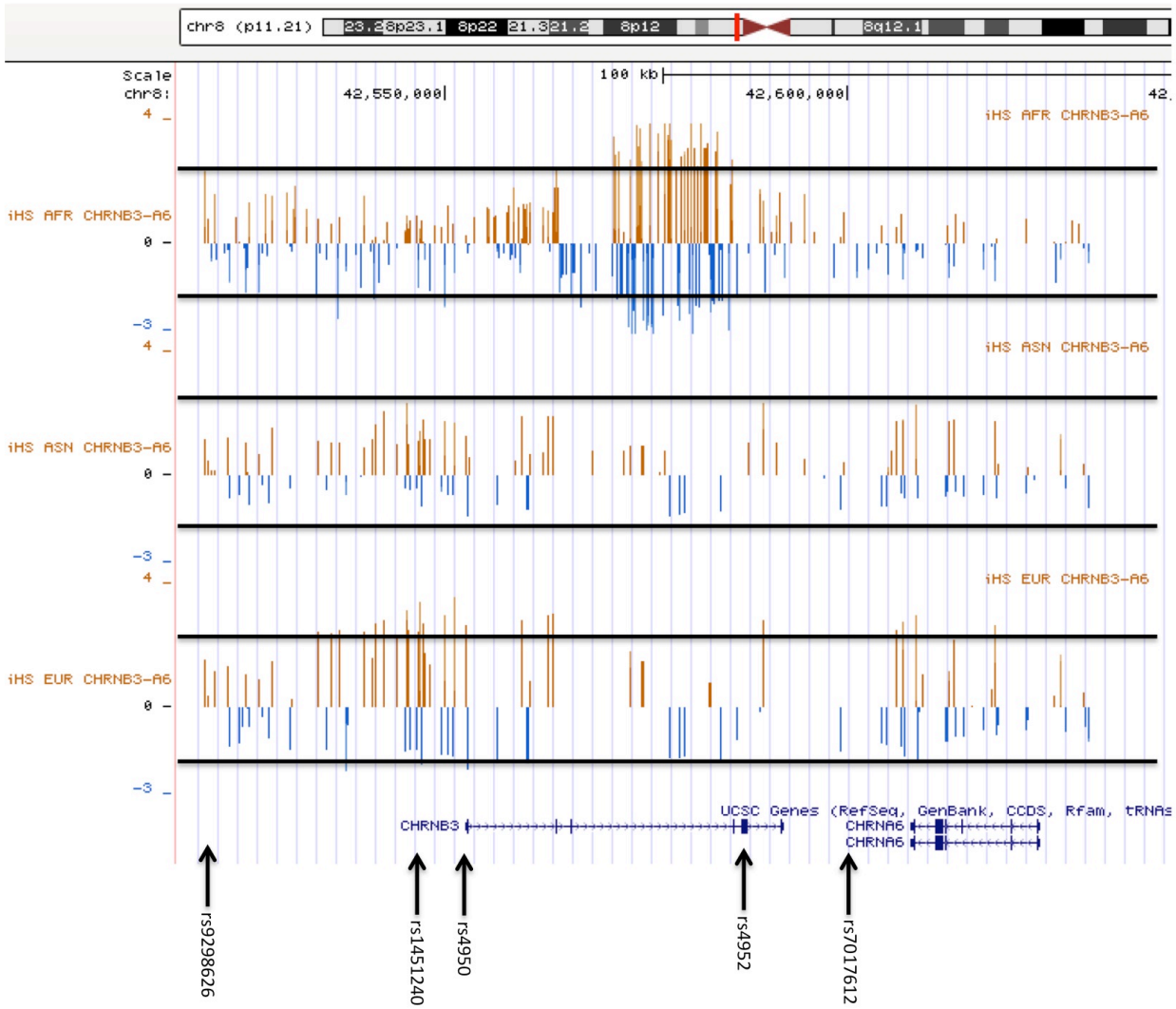

Figure 17. iHS values for the nicotinic receptor cluster on chromosome 8 plotted with gene positions and relevant SNPs below.

A6 locus on chromosome 8 shows evidence for ongoing selection in the AFR population. By contrast, in the ASN population, there were few windows of extreme iHS scores and the overall proportion was not significantly different from the negative control.

Yet a third pattern was observed in the EUR population. As in AFR, there was a highly significant overall proportion of extreme iHS values (23.6%) in EUR (Table 20). Some clustering was observed in the region just upstream of *CHRNB3,* however, in EUR the extreme values were somewhat more dispersed throughout the *CHRNB3-A6* region. This clustering of extreme his values upstream of *CHRNB3* is of high interest as this same region showed extreme values in the Tajima's D analysis.

As noted previously, an LD bin in this region tagged by rs1451240 has been shown in a GWAS study to be significantly associated with a decreased risk of nicotine addiction (Rice et al., 2012). Table 23 lists the SNPs in this LD bin and provides the iHS value for the window that contains that SNP. SNPs with an absolute value of iHS >2 (and hence an extreme value) are in bold and highlighted in yellow. In the EUR population, 13 SNPs in this LD bin, including the tag SNP, rs1451240, are in regions with an extreme iHS value. All but one have positive values, indicating the presence of unusually long haplotypes containing the ancestral allele. This suggests that the ancestral allele, which is associated with a greater risk of nicotine dependence, or one that is hitchhiking with it, is being favored by selection.

Several lines of evidence suggest that an increased risk for nicotine dependence is associated with a reduced risk of cocaine addiction and vice versa (Grucza et al., 2008; Levine et al., 2011; Sadler et al., 2014). We therefore also examined an LD bin in the

Table 23

iHS Values for GWS LD Bin on Chromosome 8 Associated with Nicotine Dependence

SNP	dbSNP Func Annot	EUR Freq A1	iHS value - EUR	ASN Freq A1	iHS value - ASN	AFR Freq A1	iHS value - AFR
rs1979140	intergenic	0.77	-1.44	0.82	-0.47	0.28	-0.15
rs7816726	intergenic	0.77	-1.44	0.81	-0.49	0.28	-0.56
rs10958726	intergenic	0.23	**2.46**	0.18	1.06	0.63	0.65
rs7842601	intergenic	0.23	**2.58**	0.18	1.06	0.63	0.84
rs13273442	intergenic	0.23	**2.32**	0.18	1.02	0.63	-0.21
rs9792277	intergenic	0.77	-1.45	0.81	-0.54	0.33	0.77
rs1451239	intergenic	0.77	-1.45	0.82	-0.46	0.34	0.89
rs1451240	intergenic	0.23	**2.48**	0.18	1.11	0.67	0.09
rs1901281	intergenic	0.77	-1.56	0.81	-0.54	0.28	0.37
rs4736835	intergenic	0.23	**2.6**	0.18	1.11	0.66	-0.08
rs1955185	intergenic	0.77	-1.58	0.82	-0.52	0.28	0.54
rs13277254	intergenic	0.23	**2.69**	0.18	1.32	0.63	-0.6
rs13277524	intergenic	0.77	-1.63	0.82	-0.65	0.28	0.61
rs6474412	intergenic	0.77	-1.63	0.82	-0.65	0.34	1.25
rs6474413	intergenic	0.77	-1.66	0.82	-0.65	0.28	0.62
rs7004381	intergenic	0.23	**2.72**	0.18	1.32	0.63	-0.72
rs6985052	intergenic	0.77	-1.66	0.82	-0.65	0.29	0.71
rs4950	CHRNB3 5'UTR	0.23	**2.71**	0.18	1.31	0.8	0.25
rs9643891	CHRNB3 intronic	0.77	-1.85	0.82	-1.01	0.15	-0.15
rs9643853	CHRNB3 intronic	0.77	-1.89	0.82	-1.01	0.15	-0.1
rs13280604	CHRNB3 intronic	0.23	**2.9**	0.18	1.94	0.8	0.12
rs6997909	CHRNB3 intronic	0.77	-1.88	0.82	-1.19	0.15	0
rs6474414	CHRNB3 intronic	0.77	-1.88	0.82	-1.19	0.15	0
rs6474415	CHRNB3 intronic	0.23	**3.07**	0.18	1.97	0.85	0.84
rs4236926	CHRNB3 intronic	0.77	-1.71	0.81	-1.41	0.15	0.14
rs16891561	CHRNB3 intronic	0.77	-1.67	0.81	-1.26	0.15	0.2
rs55828312	CHRNB3 intronic	0.24	**2.86**	0.19	2.13	0.83	1.12

iHS values for all SNPS in the genome-wide significant (GWS) LD bin ($r^2 = 0.9$) on chromosome 8 tagged by rs1451240 and associated with a decreased risk of nicotine dependence. Significant values are bolded and highlighted in yellow.

CHRNB3-A6 region bin that has been shown to have SNPs significantly associated with increased risk for cocaine dependence (Sadler et al., 2014). This bin is fairly large and spans the entire *CHRNB3-A6* cluster. It contains rs4952 and rs4953, two low frequency synonymous variants in *CHRNB3* that have previously been reported to be associated with a reduced risk for nicotine dependence (S. F. Saccone et al., 2007). All SNPs in the bin are present at around 10% in AFR and 4% in EUR but absent in ASN.

We used the tag SNP from Sadler et al. (2014) - rs9298626 – and included all SNPs with a correlation (r^2) of 0.9 or higher. The data in Appendix C demonstrate that

although rs9298626 itself did not have an extreme iHS value, multiple SNPs in LD with rs9298626 do. Overall, 40% of the SNPs in this bin showed extreme values in AFR and 15% of the SNPs in this bin showed extreme values in EUR. As these SNPs are absent from the ASN population, none showed extreme iHS values. The dense clustering of extreme iHS values in AFR and EUR is a definite indicator that this region is undergoing selection in these populations.

In EUR, all 9 of the SNPs with extreme iHS values were extreme in the positive direction and all were shared with AFR. Thus, in both populations, the ancestral allele with decreased risk for cocaine addiction is being favored. This is consistent with the low frequency of these SNPs in these populations. These data suggest that these SNPs were present in the population that traveled out of Africa to Europe and that both AFR and EUR have continued to select for the ancestral allele. The absence of these SNPs from the ASN population is consistent with the possibility that the population that migrated to Asia either lacked the SNPs initially or underwent a bottleneck that removed the SNPs from the gene pool. The AFR population exhibited an additional 13 SNPs with extreme iHS values, both positive and negative, suggesting that there are a variety of alleles, ancestral and derived, that tag the selected haplotype.

Ka/Ks ratio. The Ka/Ks ratio is a way of comparing selective pressures at homologous genes. When non-synonymous (Ka) and synonymous (Ks) substitution rates are equal, their ratio = 1, and this is indicative of neutral evolution. However, if the Ka/Ks ratio is less than one, this indicates purifying or stabilizing selection. If the Ka/Ks ratio is greater than one, this is indicative of positive selection (Zhang et al., 2006).

Here we compare the human and chimpanzee sequences for the nicotinic receptors to test for evolution. In this analysis, the null hypothesis is Ka=Ks. We can reject the null hypothesis if Ka is significantly greater or less than Ks, as indicated by the p-value for the associated Fisher Exact test.

Table 24

Ka/Ks Ratio for all CHRN Genes Examined

Sequence	Method	Ka	Ks	Ka/Ks	P-Value(Fisher)
CHRNB3	LWL	0.001	0.017	0.055	1.12E-04
CHRNA3	LWL	0.002	0.015	0.113	7.20E-04
CHRNA6	LWL	0.002	0.013	0.134	2.29E-03
CHRNB4	LWL	0.019	0.049	0.393	6.79E-03
CHRNA5	LWL	0.005	0.010	0.462	1.25E-01
LCT	LWL	0.007	0.014	0.466	2.29E-02

Ka/Ks ratio for all genes examined, including the LCT positive control and significance of the ratio according to Fisher Exact test.

Table 24 shows the Ka/Ks ratio for all the nicotinic receptor genes (*CHRNB3*, *CHRNA3*, *CHRNA6, CHRNB4* and *CHRNA5*) plus *LCT* as the positive control. As Ka/Ks can only be calculated for protein coding regions, there are no data for the intergenic negative controls. For all the genes examined, the Ka/Ks ratio was less than 1, indicating fixation of more synonymous than non-synonymous substitutions and thus describing a situation in which there is purifying or stabilizing selection. Our positive control, *LCT* had the highest Ka/Ks ratio of 0.466, which was nonetheless still significant (p=2.29x10-*2*). This relatively weak signal may be due to the fact that it is the expression of lactase that is under selection rather than the protein sequence itself.

Among the nicotinic receptors, *CHRNB3* appears to be under the strongest stabilizing selection, as its Ka/Ks ratio is nearly zero and the Fisher p-value ($p=1.12x10^{-4}$) indicates this is highly significant. *CHRNA3* and *CHRNA6* are similar with a Ka/Ks ratio

of 0.113 ($p=7.20x10^{-4}$) and 0.134 (p=2.29x10-4), respectively. *CHRNB4* has a slightly higher ratio of 0.393 ($p=6.79x10^{-3}$), but this still appears to be significant. Of all the CHRN genes, only *CHRNA5* has a Ka/Ks ratio that is not significantly different from 1 (Ka/Ks =0.462, p=0.125).

Thus, all but one of the nicotinic genes are undergoing stabilizing selection according to this metric. As this comparison is between human and chimp, it is expected that some selection would have been occurring. However, it is interesting that the selection is working to maintain the "ancestral" chimp sequence.

Nicotine addiction and cognitive function. The Tajima's D analysis, integrated haplotype score and Ka/Ks ratio all indicate that the *CHRNB3-A6* cluster is undergoing selection and in particular, the iHS scores suggest that it is the risk allele for nicotine dependence on chromosome 8 that is being selected for. As it seems unlikely that risk of nicotine dependence is the phenotype undergoing selection, and because nicotinic receptors are involved in memory and learning, we hypothesized that a phenotype related to memory or learning, such as attention, might be the phenotype being selected.

To test this possibility, we obtained data on genotype and cognitive phenotype from the Collaborative Study of the Genetics of Alcoholism (COGA). Using this dataset, we tested the association between genotype and three of the most relevant phenotypes, namely scores on Wechsler Adult Intelligence Scale (WAIS) Block Design, WAIS Digit Symbol and WAIS Information tests.

In WAIS Block Design, the subject replicates models or pictures of two-color designs with blocks. In WAIS Information, the subject answers a series of questions

about factual information. In WAIS Digit Symbol, the subject writes down as quickly as possible the symbols that correspond to a series of numbers.

Table 25

SNP Association with Scores on WAIS Digit Symbol Test

SNP	Beta (β)	Std. Error (SE)	P-Value (N=492)
rs7017612	0.43	0.19	**0.003**
rs6982753	0.39	0.20	0.009
rs10958725	0.25	0.18	0.035
rs13273442	0.28	0.18	0.036
rs4950	0.24	0.18	0.038
rs1530848	0.24	0.18	0.038
rs6474413	0.25	0.18	0.038
rs10107450	0.10	0.19	0.064
rs16891620	0.27	0.25	0.066
rs2196128	0.22	0.20	0.066
rs1530847	0.23	0.20	0.104
rs16891530	0.19	0.42	0.414
rs4952	0.19	0.42	0.414
rs7815274	0.19	0.42	0.426
rs10109429	0.07	0.31	0.494
rs13270610	-0.16	0.35	0.632
rs16891604	0.23	0.42	0.826

Beta, standard error and p-value for all SNPs in *CHRNB3-A6* region of the COGA family GWAS and their association with scores on the WAIS Digit Symbol test. Covariates used were age, sex and FTND score.

Table 25 summarizes our findings. Of the 17 SNPs in the *CHRNB3-A6* region on chromosome 8, one SNP – rs7017612 - passed multiple test correction ($p \leq 0.003$) for association with score on WAIS Digit Symbol ($\beta=0.43$, $p=0.003$). Specifically, this test consists of nine pairs of digits and symbols (e.g. 1/-, 2/X….9/*). This is followed by a list of digits, under which the person must write the corresponding symbol as fast as they

can, and the number of correct responses within a time limit is measured (Salthouse, 1992). After this, paired and free recall of the symbols is measured. As shown in Figure 17, rs7017612 lies in the intergenic region between *CHRNB3* and *CHRNA6*. This SNP has an r^2 of 0.75 and a D' of 0.95 with rs6474413, a SNP tagging the genome-wide significant bin for decreased risk for nicotine dependence. These data suggest a modest association between genotype at these SNPs and cognitive function. We also repeated the analysis using CPD as a covariate instead of FTND in order to measure consumption rather than case/control status. The p-values were essentially the same.

A second SNP in the *CHRNB3-A6* region– rs6982753- had a nominal p-value with the WAIS Digit Symbol phenotype before multiple test correction and almost passed the multiple test correction (p=0.009). Interestingly, this SNP has an r^2 of 0.91 with rs892413 (β=0.39, p=0.008), a SNP that has previously been associated with increased risk for cocaine dependence (Sadler et al. 2014).

No other neurocognitive phenotypes besides WAIS digit symbol had SNPs with significant values in the *CHRNB3-A6* region and none of the three neurocognitive phenotypes had a significant association with SNPs in the region of *CHRNA5-A3-B4* on chromosome 15 (not shown).

DISCUSSION

It is clear from a variety of studies that the risk of nicotine addiction has a genetic component. The question is why such a phenotype would exist and possibly be selected for in a population. Selective pressures in our ancestral environments were likely not on addiction, but rather on behaviors that were biologically rewarding (i.e. mate or food finding, avoidance of harmful stimuli). Given the role of nicotine in neurological

function, it is possible that, in the case of nicotine addiction, the phenotype on which natural selection was working was related to enhancements in memory or cognition. The addiction phenotype would have hitchhiked along because it acts through the same or related mechanisms. The addiction phenotype was likely not selected against in ancestral environments because the availability and opportunity for prolonged use of purified drugs was negligible.

In this study, we sought to address this question of the evolution of nicotine addiction by first determining whether the genes for the nicotinic receptor subunits show evidence of ongoing selection and then by determining whether any SNPs in these genes might be associated with a phenotype associated with memory or cognition. We performed three different tests of selection on chromosomal regions containing the genes for five subunits of the nicotinic receptor and all three of these analyses indicate that selection is occurring at the *CHRNB3-A6* locus.

ß3 and α6 are two of the 11 subunits (8 alphas and 3 betas) that combine to form the pentameric nicotinic cholinergic receptor. There are two major subtypes containing these receptors: α6α4β2β3 and α6β2β3. Almost all β3 receptor subunits occur in the presence of α6, forming the complex α6α4β2β3 subtype (Gotti et al., 2005). However, α6β2β3 accounts for only ~ 40-60% of all α6-containing receptors (Gotti et al., 2007).

ß3 mRNA has a limited expression pattern, with high levels being seen in the substantia nigra, medial habenula, ventral tegmental area and thalamus in the brain (Drago et al., 2003). Interestingly β3-containing *CHRNs* appear to play a significant role in dopaminergic neurotransmission (Cui et al., 2003). β3 null mice have altered locomotor activity and prepulse inhibition of acoustic startle (presentation of a tone prior

to an acoustic stimulus). Both of these behaviors are partially mediated by dopaminergic pathways, suggesting a relationship between ß3 and the pathways known to mediate drug addiction.

α6 mRNA is mainly expressed in the substantia nigra, vental tegmental area and locus coeruleus, and to a lesser extent in the retina and thalamic reticular nucleus. In all these regions, it co-localizes with β3 mRNA (Gotti et al., 2007). *CHRNA6* receptors are highly and selectively expressed by mesostriatal dopamine neurons that mediate the behavioral effects of nicotine such as habit learning and reinforcement (Gotti et al., 2010). Selective expression of $\alpha 6^*$ nAChRs in monkey striatum suggests that the α6β2β3, α6α4β2β3, and $\alpha 3\beta 2^*$ nAChR subtypes are present on dopaminergic terminals (Quik et al., 2005). Thus, like ß3 subunits, the a6 subunits appear to be involved in the regulation of dopaminergic pathways associated with drug addiction.

Both the Tajima's D test and iHS point to an ongoing sweep in humans on chromosome 8. In the case of the *CHRNB3-A6* locus, all of the extreme values in the Tajima's D analysis were positive. High positive Tajima's D values occur when there is an excess of variants in a region with intermediate allele frequencies. This can occur in either balancing selection or ongoing positive selection. We also found extreme iHS values in the *CHRNB3-A6* locus. This is the only *CHRN* region to fulfill the criteria for a sweep laid out in Voight et al. (Voight et al., 2006), i.e. clustering of extreme iHS values. iHS values are only extreme in the presence of positive selection. Together, these data imply that the Tajima's D analysis is picking up on ongoing positive selection rather than balancing selection.

Several SNPs in the *CHRNB3-A6* locus on chromosome 8 have previously been associated with a decreased risk of nicotine dependence (Rice et al., 2012). One of these, rs1451240, was present in a window that showed extreme values in both the Tajima's D test and iHS. The extreme positive iHS value in the window including rs1451240 indicates that the haplotype containing the ancestral allele is being selected. As the derived allele provides protection from nicotine addiction, this suggests that it is the allele that is associated with a greater risk of nicotine dependence that is being selected for. Since highly concentrated sources of nicotine were not present in the ancestral environment, it seems likely that this phenotype of nicotine dependence would have hitchhiked along with a more beneficial phenotype, such as improved memory or cognition.

To test this possibility, we assessed the association of SNPs in the *CHRNB3-A6* locus with scores on WAIS tests of memory and cognitive function. Our analysis of the individuals in the COGA dataset suggests that one SNP, rs7017612, that lies in the intergenic region between *CHRNB3* and *CHRNA6,* is associated with increased score on the WAIS Digit Symbol test. This test is thought to measure largely processing speed, but also, to some extent, memory. rs7017612 itself has not been associated with nicotine dependence. However, it is in reasonable LD ($r^2 = 0.75$) with rs6474413, a SNP tagging the genome-wide significant bin for decreased risk for nicotine dependence. Thus, our data are consistent with the possibility that improved performance on this particular cognitive test is modestly associated with a decreased risk for nicotine dependence.

This conclusion is contrary to the bulk of our results relating to the direction of selection bias on the *CHRNB3-A6* locus. Our results largely suggest that positive

selective pressure is being exerted on the ancestral allele, which does not provide protection against nicotine dependence. This suggests that another mechanism may exist to explain the apparent selection for SNPs that fail to confer an obvious selective advantage.

Genetic studies of nicotine addiction have identified an inverse relationship between the risk for nicotine addiction and the risk for cocaine addiction. A variant in the gene for the α5 subunit of the nicotinic receptor, rs16969968, that enhances risk for the development of nicotine dependence, independently decreases risk for cocaine dependence as well (Grucza et al., 2008). This finding was replicated in a subsequent study (Sherva et al., 2010). My studies have identified a similar inverse relationship between the risk for nicotine and cocaine addiction conferred by SNPs in the *CHRNA6-B3* locus on chromosome 8 (Sadler et al., 2014).

Consistent with a functional relationship between cocaine and nicotine, Levine et al. (Levine et al., 2011) showed that the use of cocaine among smokers increases the risk of becoming dependent on cocaine. These workers further showed that pretreatment of mice with nicotine increased the response to cocaine, particularly in the striatum, a brain region involved in addiction-related reward, where α6ß3 containing nicotinic receptors are expressed. Thus, an enhanced responsiveness to nicotine would prime the user to the potentiating effect of cocaine on addiction-related reward pathways. Presumably, a decreased responsiveness to nicotine would limit the effect of cocaine on addiction-related reward.

Cocaine addiction is characterized by a dampened reward response to social interaction, meaning that it inhibits the positive emotions that accompany social

interaction or feelings of belonging. A recent study demonstrated that cocaine users process social gaze (joint attention on an object) differently than controls, resulting in a reduced activation of the reward system during social interactions (Preller et al., 2014). Using fMRI, these authors showed that cocaine users had decreased activation of the medial orbitofrontal cortex, a region of the brain central for reward processing. These observations could explain why alleles that protect against cocaine dependence could be under positive selection.

Humans have an evolutionary need to form connections with each other, and as such, social bonding is adaptive and rewarding. The opposite, social isolation, is maladaptive and painful (Verdejo-Garcia, 2014). Selecting against the risk for cocaine addiction would maximize the likelihood that individuals would maintain their reward system for social interaction. This would benefit both the individual and the group. Since nicotine sensitizes the animal to the effects of cocaine, which blunts the reward of social interactions, alleles that reduced the ability of nicotine to enhance the effects of cocaine would have undergone positive selection. In this scenario, the nicotine dependence phenotype is not hitchhiking with memory or learning phenotypes, but rather with phenotypes protecting against antisocial and therefore maladaptive behavior.

Like nicotine, a source of highly concentrated cocaine was likely not present in the ancestral environment. Coca, the most abundant source of cocaine, is native to South America and therefore would not have been available to our early hominid ancestors in Africa. However, other stimulants, such as the amphetamine derivatives present in khat, would have been available in the ancestral environment and several lines of evidence

suggest that nicotine has a similar relationship with amphetamines as it does with cocaine.

Work from several groups suggests that, as with cocaine, nicotine potentiates the behavioral effects of amphetamine (Jutkiewicz, Nicolazzo, Kim, & Gnegy, 2008; Santos, Marin, Cruz, Delucia, & Planeta, 2009). Furthermore, humans express an endogenous neuropeptide known as *CART* (cocaine- and amphetamine-regulated transcript) (Jaworski & Jones, 2006). Expression of *CART* is highly upregulated in response to either cocaine or amphetamines and it acts to suppress the effects of both of these drugs, both of which act to enhance the levels of dopamine in the brain (Subhedar, Nakhate, Upadhya, & Kokare, 2014). Finally, administration of either cocaine or amphetamines enhances the release of acetylcholine (Imperato et al. 1993). The latter observation provides a physiological link between these drugs and nicotinic receptors.

It is possible that there is a positively reinforcing cycle between nicotine and a variety of drugs, including but perhaps not limited to, cocaine and amphetamines. Excess nicotinic receptor stimulation, either by nicotine or the endogenous ligand, acetylcholine, in a highly responsive receptor, would enhance the effects of cocaine or amphetamines. This would make the individuals more susceptible to addiction by drugs that impair the rewards of social interactions, a negative trait when survival depends on the entire group contributing to the procurement of food and defense against intruders. Thus, SNPs that limit the efficacy of acetylcholine (or nicotine) for enhancing addiction to drugs that lead to anti-social behavior, without markedly diminishing cognitive function, could represent a trait for positive selection in a social group.

While not as compelling, there was nonetheless some evidence for selection at the *CHRNA3-B4-A5* locus. In particular, rs16969968, the SNP that encodes the missense mutation in α5 that is strongly associated with risk of nicotine dependence, lies in a sliding window exhibiting a high Tajima's D score. However, the iHS analysis of this locus did not provide evidence for selection. Although the evidence for selection at the *CHRN* cluster on chromosome 15 was less convincing, it would be interesting to extend the integrated haplotype score (iHS) analyses another ~110 kb upstream from *CHRNA5* through *IREB2*. SNPs in *IREB2* are in the LD bin that contains rs16969968. *IREB2* is an iron-responsive element binding protein that regulates iron homeostasis. Knockout of this gene in mice results in behavioral abnormalities and neurodegeneration. *IREB2* is of interest as Johnson/Sadler et al. (in preparation) find evidence for an association between *IREB2* transcript levels and rs141518190, a SNP intronic to *CHRNA3*, in African-Americans. The data suggest that rs141518190 acts as a cis-eQTL for *IREB2*, linking this SNP in *CHRNA3* to an important neurological phenotype.

CHAPTER 5

CONCLUSIONS AND FUTURE DIRECTIONS

Evolutionary explanations of addiction have focused on the idea that addictive drugs are habit forming because they act on the dopaminergic brain circuits that mediate natural and biologically significant rewards of evolutionarily beneficial behaviors (Nesse & Berridge, 1997). The hypothesis that drug use may be adaptive in specific cases is perhaps most plausible for the most commonly used drugs, nicotine and caffeine, since the use of these drugs is patterned in ways that would be predicted if adaptive use is assumed. For example, groups under greater physical and emotional stress appear to use these drugs more often, just as the adaptive model would suggest. Some have argued that humans and psychotropic plants, including tobacco, could have a co-evolutionary relationship that is millions of years old, and posited that humans have eaten plants to obtain neurochemicals directly to reduce foraging time (R. J. Sullivan & Hagen, 2002). Indeed, nicotine can suppress appetite and reduce fatigue (Pomerleau, 1997), allowing more effective foraging.

While it is generally agreed that drug use must be adaptive in some way, to date there have been no attempts to uncover evidence for natural selection in genes specifically related to addiction. And few studies have examined the relationship between SNPs in genes that correlate with addiction phenotype and any neuropsychological phenotypes that could be the true targets of natural selection. The work reported in this thesis has addressed this gap in our knowledge by looking at two loci associated with risk for nicotine dependence to seek an answer to the question of the evolutionary basis for the existence of genes for nicotine addiction.

Because the few SNPs that have been associated with nicotine dependence are in genes encoding nicotinic receptors or nicotine metabolizing genes, we initially undertook a GWAS of imputed data from the Collaborative Study of the Genetics of Nicotine Dependence (COGEND) dataset to look for genetic associations outside of the cholinergic receptor loci. This work identified SNPs in micro RNAs and *GPC5* that represent the first non-nicotinic receptor SNPs that are associated with elevated risk of nicotine dependence.

Currently, the genotyping situation on the GWAS performed in this work is less than ideal. With most of the controls genotyped on the 1M SNP chip and most of the cases on the 2.5M SNP chip, the logical and conservative choice was to only impute from the intersection of these two platforms, which was approximately 600,000 SNPs. This approach discards data on nearly 2 million SNPs for some subjects. Future studies should re-genotype all of the subjects from the GWAS of imputed data on the 2.5M SNP chip to increase the number of SNPs from which to impute. This would lead to increased accuracy of imputation, and may lead to additional discoveries of loci associated with nicotine dependence. In addition, the *GPC5* finding needs to be replicated in other larger independent datasets.

Our studies described in chapter 3 identify the second example of a SNP within a *CHRN* locus (rs9298626 on chromosome 8) that has opposing effects on nicotine and cocaine dependencies. Conditional and haplotype analyses discovered that this effect on cocaine dependence was independent of previous effects on nicotine dependence, meaning that when you control for the known nicotine association, the association with cocaine is still significant. To strengthen our conclusion, it would be useful to replicate

the finding on chromosome 8 in other datasets that have assessed cocaine and nicotine dependence phenotypes. In addition, it would be interesting to analyze datasets of other ethnicities to determine if these results are unique to European-Americans or are generally true across all populations. Because no underlying functional mechanisms in this region have been discovered to account for the association with either nicotine or cocaine, it would be interesting to use functional tests to determine functionally why these SNPs may have been under selection. For example, in brain tissue, DNase-seq assays could be run on those SNPs predicted to be DNase hypersensitivity sites and ChIP-seq assays could be run on SNPs predicted to affect transcription factor binding or histone marks. While SNPs in these genes have been tested for their effect on expression and methylation in different types of brain tissues, nothing was found for SNPs in or upstream of *CHRNB3* (Jen Wang, personal communication). Finally, it would be interesting to further investigate the possible inverse relationship between nicotine dependence and cocaine addiction by determining the nicotine-dependence phenotype and genotype in individuals harboring mutations in the *CYFIP2* gene recently shown to be associated with response to cocaine (Kumar et al., 2013).

To determine if any variants previously related to nicotine dependence showed evidence that they were undergoing natural selection, we examined the nicotinic receptor genes on chromosomes 8 and 15 using three statistical tests of selection: Tajima's D; integrated haplotype score (iHS); and, Ka/Ks ratio. Most interestingly, we discovered evidence for positive selection on the ancestral alleles within the *CHRNB3* region on chromosome 8. These ancestral alleles are associated with risk for nicotine dependence but protection from cocaine dependence.

Our analyses of the association between nicotine dependence and neurocognitive features reported in Chapter 4 was more limited than we would have liked. Because of the discrepancy between the number of people in the Collaborative Study of the Genetics of Alcoholism (COGA) dataset with neurocognitive data and the number with GWAS data available, we conducted our analysis on 492 individuals. Even with this small dataset, we found a significant association between a SNP linked to risk of nicotine dependence and score on WAIS digit symbol. In the future, it would be of interest to genotype the rest of the individuals with phenotypes (but no genotypes) to increase the sample size for the association analysis from 492 to 1247. This might increase the strength of the association at the SNPs discovered in my analysis, or possibly lead to associations at new SNPs either on chromosome 8 or 15 with WAIS digit symbol, block design or information.

Given the modest association between performance on WAIS digit symbol and nicotine dependence, it is not possible to draw a strong conclusion regarding whether this represents a phenotype upon which selection might act. However, our findings regarding the inverse relationship between nicotine dependence and cocaine dependence offer another possible explanation for the presence of SNPs conferring risk for nicotine dependence. This possibility could be addressed by additional genetic studies of nicotine and cocaine addiction as well as nicotine and amphetamine addiction. Further, the possible role of *CART*, (cocaine- and amphetamine-regulated transcript) as a genetic basis for cocaine and amphetamine addiction and a possible relationship to nicotine addiction should be investigated. This is particularly true as Gelernter et al. (2007) have previously

identified a nicotine dependence risk locus on chromosome 5 that includes the *CART* gene.

There is still much work to be done in determining both genetic and evolutionary explanations for phenotypes. Behavioral phenotypes are especially difficult in this respect, particularly in humans, because of complex cultural variables. However, I believe that evolutionary explanations must be increasingly considered as genomic information becomes cheaper and easier to obtain. It is my hope that this work has added valuable insight to this field and will inspire others to build on this work or investigate evolutionary explanations for other behaviors.

REFERENCES

ACS. (2009). American Cancer Society. Cancer Facts and Figures. *American Cancer Society.* from http://www.cancer.org/acs/groups/content/@nho/documents/document/500809webpdf.pdf

Arora, M., Mathur, M. R., & Singh, N. (2013). A framework to prevent and control tobacco among adolescents and children: introducing the IMPACT model. *Indian J Pediatr, 80 Suppl 1*, S55-62. doi: 10.1007/s12098-012-0768-y

Ashley-Koch, A., Yang, Q., & Olney, R.S. (2000). Sickle Hemoglobin (HbS) Allele and Sickle Cell Disease: A HuGE Review. *Am J Epidemiol, 151*, 839-845.

Audrain-McGovern, J., Lerman, C., Wileyto, E. P., Rodriguez, D., & Shields, P. G. (2004). Interacting effects of genetic predisposition and depression on adolescent smoking progression. *Am J Psychiatry, 161*(7), 1224-1230.

Babar, A. A., Stigler, M. H., Perry, C. L., Arora, M., Shrivastav, R., & Reddy, K. S. (2010). Tobacco-use psychosocial risk profiles of girls and boys in urban India: implications for gender-specific tobacco intervention development. *Nicotine Tob Res, 12*(1), 29-36. doi: 10.1093/ntr/ntp169

Berrettini, W., Yuan, X., Tozzi, F., Song, K., Francks, C., Chilcoat, H., . . . Mooser, V. (2008). Alpha-5/alpha-3 nicotinic receptor subunit alleles increase risk for heavy smoking. *Mol Psychiatry, 13*(4), 368-373. doi: 10.1038/sj.mp.4002154

Bierut, L. J. (2009). Nicotine dependence and genetic variation in the nicotinic receptors. *Drug Alcohol Depend, 104 Suppl 1*, S64-69. doi: 10.1016/j.drugalcdep.2009.06.003

Bierut, L. J. (2011). Genetic vulnerability and susceptibility to substance dependence. *Neuron, 69*(4), 618-627. doi: 10.1016/j.neuron.2011.02.015

Bierut, L. J., Agrawal, A., Bucholz, K. K., Doheny, K. F., Laurie, C., Pugh, E., . . . Gene, Environment Association Studies Consortium. (2010). A genome-wide association study of alcohol dependence. *Proc Natl Acad Sci U S A, 107*(11), 5082-5087. doi: 10.1073/pnas.0911109107

Bierut, L. J., Dinwiddie, S. H., Begleiter, H., Crowe, R. R., Hesselbrock, V., Nurnberger, J. I., Jr., . . . Reich, T. (1998). Familial transmission of substance dependence: alcohol, marijuana, cocaine, and habitual smoking: a report from the Collaborative Study on the Genetics of Alcoholism. *Arch Gen Psychiatry, 55*(11), 982-988.

Bierut, L. J., Goate, A. M., Breslau, N., Johnson, E. O., Bertelsen, S., Fox, L., . . . Edenberg, H. J. (2012). ADH1B is associated with alcohol dependence and alcohol consumption in populations of European and African ancestry. *Mol Psychiatry, 17*(4), 445-450. doi: 10.1038/mp.2011.124

Bierut, L. J., Madden, P. A., Breslau, N., Johnson, E. O., Hatsukami, D., Pomerleau, O. F., . . . Ballinger, D. G. (2007). Novel genes identified in a high-density genome wide association study for nicotine dependence. *Hum Mol Genet, 16*(1), 24-35. doi: 10.1093/hmg/ddl441

Bierut, L. J., Stitzel, J. A., Wang, J. C., Hinrichs, A. L., Grucza, R. A., Xuei, X., . . . Goate, A. M. (2008). Variants in nicotinic receptors and risk for nicotine dependence. *Am J Psychiatry, 165*(9), 1163-1171. doi: 10.1176/appi.ajp.2008.07111711

Blum, K., Chen, A. L., Oscar-Berman, M., Chen, T. J., Lubar, J., White, N., . . . Bailey, J. A. (2011). Generational association studies of dopaminergic genes in reward deficiency syndrome (RDS) subjects: selecting appropriate phenotypes for reward dependence behaviors. *Int J Environ Res Public Health, 8*(12), 4425-4459. doi: 10.3390/ijerph8124425

Calboli, F. C., Tozzi, F., Galwey, N. W., Antoniades, A., Mooser, V., Preisig, M., . . . Balding, D. J. (2010). A genome-wide association study of neuroticism in a population-based sample. *PLoS One, 5*(7), e11504. doi: 10.1371/journal.pone.0011504

Carmelli, D., Swan, G. E., Robinette, D., & Fabsitz, R. R. (1990). Heritability of substance use in the NAS-NRC Twin Registry. *Acta Genet Med Gemellol (Roma), 39*(1), 91-98.

CDC. (2008). Cigarette smoking among adults United States, 2007. *Morbidity and Mortality Weekly Report, 57*, 1221-1226.

CDC. (2013). Centers for Disease Control . Health Effects of Cigarette Smoking. from http://www.cdc.gov/tobacco/data_statistics/fact_sheets/health_effects/effects_cig_smoking/index.htm

Chanda, P., Yuhki, N., Li, M., Bader, J. S., Hartz, A., Boerwinkle, E., . . . Arking, D. E. (2012). Comprehensive evaluation of imputation performance in African Americans. *J Hum Genet, 57*(7), 411-421. doi: 10.1038/jhg.2012.43

Cho, M. H., Castaldi, P. J., Wan, E. S., Siedlinski, M., Hersh, C. P., Demeo, D. L., . . . Investigators, C. OPDGene. (2012). A genome-wide association study of COPD identifies a susceptibility locus on chromosome 19q13. *Hum Mol Genet, 21*(4), 947-957. doi: 10.1093/hmg/ddr524

Comings, D. E., Ferry, L., Bradshaw-Robinson, S., Burchette, R., Chiu, C., & Muhleman, D. (1996). The dopamine D2 receptor (DRD2) gene: a genetic risk factor in smoking. *Pharmacogenetics, 6*(1), 73-79.

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

Cui, C., Booker, T. K., Allen, R. S., Grady, S. R., Whiteaker, P., Marks, M. J., . . . Heinemann, S. F. (2003). The beta3 nicotinic receptor subunit: a component of alpha-conotoxin MII-binding nicotinic acetylcholine receptors that modulate dopamine release and related behaviors. *J Neurosci, 23*(35), 11045-11053.

Dajas-Bailador, F., & Wonnacott, S. (2004). Nicotinic acetylcholine receptors and the regulation of neuronal signalling. *Trends Pharmacol Sci, 25*(6), 317-324. doi: 10.1016/j.tips.2004.04.006

Degner, J. F., Pai, A. A., Pique-Regi, R., Veyrieras, J. B., Gaffney, D. J., Pickrell, J. K., . . . Pritchard, J. K. (2012). DNase I sensitivity QTLs are a major determinant of human expression variation. *Nature, 482*(7385), 390-394. doi: 10.1038/nature10808

Delaneau, O., Marchini, J., & Zagury, J. F. (2012). A linear complexity phasing method for thousands of genomes. *Nat Methods, 9*(2), 179-181. doi: 10.1038/nmeth.1785

Drago, J., McColl, C. D., Horne, M. K., Finkelstein, D. I., & Ross, S. A. (2003). Neuronal nicotinic receptors: insights gained from gene knockout and knockin mutant mice. *Cell Mol Life Sci, 60*(7), 1267-1280.

Durbin, R.M., Altshuler, D.L., Abecasis, G.R., Bentley, D.R., & 1000 Genomes Consortium, The. (2010). A Map of Human Genome Variation from Population-Scale Sequencing. *Nature, 467*, 1061-. doi: 10.1038/nature09534

Eastwood, S. L., Law, A. J., Everall, I. P., & Harrison, P. J. (2003). The axonal chemorepellant semaphorin 3A is increased in the cerebellum in schizophrenia and may contribute to its synaptic pathology. *Mol Psychiatry, 8*(2), 148-155. doi: 10.1038/sj.mp.4001233

Edwards, K. L., Austin, M. A., & Jarvik, G. P. (1995). Evidence for genetic influences on smoking in adult women twins. *Clin Genet, 47*(5), 236-244.

Ellinghaus, D., Schreiber, S., Franke, A., & Nothnagel, M. (2009). Current software for genotype imputation. *Hum Genomics, 3*(4), 371-380.

Epstein, L. H., Temple, J. L., Neaderhiser, B. J., Salis, R. J., Erbe, R. W., & Leddy, J. J. (2007). Food reinforcement, the dopamine D2 receptor genotype, and energy intake in obese and nonobese humans. *Behav Neurosci, 121*(5), 877-886. doi: 10.1037/0735-7044.121.5.877

Ernst, M., Matochik, J. A., Heishman, S. J., Van Horn, J. D., Jons, P. H., Henningfield, J. E., & London, E. D. (2001). Effect of nicotine on brain activation during performance of a working memory task. *Proc Natl Acad Sci U S A, 98*(8), 4728-4733. doi: 10.1073/pnas.061369098

Etter, J. F., Hoda, J. C., Perroud, N., Munafo, M., Buresi, C., Duret, C., . . . Bertrand, D. (2009). Association of genes coding for the alpha-4, alpha-5, beta-2 and beta-3 subunits of nicotinic receptors with cigarette smoking and nicotine dependence. *Addict Behav, 34*(9), 772-775. doi: 10.1016/j.addbeh.2009.05.010

Fisher, R. A. (1958). Lung cancer and cigarettes. *Nature, 182*(4628), 108.

Frahm, S., Slimak, M. A., Ferrarese, L., Santos-Torres, J., Antolin-Fontes, B., Auer, S., . . . Ibanez-Tallon, I. (2011). Aversion to nicotine is regulated by the balanced activity of beta4 and alpha5 nicotinic receptor subunits in the medial habenula. *Neuron, 70*(3), 522-535. doi: 10.1016/j.neuron.2011.04.013

Fratiglioni, L., & Wang, H. X. (2000). Smoking and Parkinson's and Alzheimer's disease: review of the epidemiological studies. *Behav Brain Res, 113*(1-2), 117-120.

Freathy, R. M., Ring, S. M., Shields, B., Galobardes, B., Knight, B., Weedon, M. N., . . . Hattersley, A. T. (2009). A common genetic variant in the 15q24 nicotinic acetylcholine receptor gene cluster (CHRNA5-CHRNA3-CHRNB4) is associated with a reduced ability of women to quit smoking in pregnancy. *Hum Mol Genet, 18*(15), 2922-2927. doi: 10.1093/hmg/ddp216

Garrett, B., Dube, S., Trosclair, A., Caraballo, R., & Pechacek, T. (2009). Cigarette Smoking - United States, 1965-2008. *CDC Morbidity and Mortality Weekly Report.* Retrieved 01, 60, from http://www.cdc.gov/mmwr/preview/mmwrhtml/su6001a24.htm

Garrigan, D., & Hammer, M. F. (2006). Reconstructing human origins in the genomic era. *Nat Rev Genet, 7*(9), 669-680. doi: 10.1038/nrg1941

Gelernter, J., & Kranzler, H. R. (2010). Genetics of drug dependence. *Dialogues Clin Neurosci, 12*(1), 77-84.

Gelernter, J., Panhuysen, C., Weiss, R., Brady, K., Poling, J., Krauthammer, M., . . . Kranzler, H. R. (2007). Genomewide linkage scan for nicotine dependence:

identification of a chromosome 5 risk locus. *Biol Psychiatry, 61*(1), 119-126. doi: 10.1016/j.biopsych.2006.08.023

Goldman, D., Oroszi, G., & Ducci, F. (2005). The genetics of addictions: uncovering the genes. *Nat Rev Genet, 6*(7), 521-532. doi: 10.1038/nrg1635

Goodwin, R. D., Zvolensky, M. J., Keyes, K. M., & Hasin, D. S. (2012). Mental disorders and cigarette use among adults in the United States. *Am J Addict, 21*(5), 416-423. doi: 10.1111/j.1521-0391.2012.00263.x

Gotti, C., Guiducci, S., Tedesco, V., Corbioli, S., Zanetti, L., Moretti, M., . . . Zoli, M. (2010). Nicotinic acetylcholine receptors in the mesolimbic pathway: primary role of ventral tegmental area alpha6beta2* receptors in mediating systemic nicotine effects on dopamine release, locomotion, and reinforcement. *J Neurosci, 30*(15), 5311-5325. doi: 10.1523/JNEUROSCI.5095-09.2010

Gotti, C., Moretti, M., Clementi, F., Riganti, L., McIntosh, J. M., Collins, A. C., . . . Whiteaker, P. (2005). Expression of nigrostriatal alpha 6-containing nicotinic acetylcholine receptors is selectively reduced, but not eliminated, by beta 3 subunit gene deletion. *Mol Pharmacol, 67*(6), 2007-2015. doi: 10.1124/mol.105.011940

Gotti, C., Moretti, M., Gaimarri, A., Zanardi, A., Clementi, F., & Zoli, M. (2007). Heterogeneity and complexity of native brain nicotinic receptors. *Biochem Pharmacol, 74*(8), 1102-1111. doi: 10.1016/j.bcp.2007.05.023

Grant, B. F., Hasin, D. S., Chou, S. P., Stinson, F. S., & Dawson, D. A. (2004). Nicotine dependence and psychiatric disorders in the United States: results from the national epidemiologic survey on alcohol and related conditions. *Arch Gen Psychiatry, 61*(11), 1107-1115. doi: 10.1001/archpsyc.61.11.1107

Greenwood, P. M., Parasuraman, R., & Espeseth, T. (2012). A cognitive phenotype for a polymorphism in the nicotinic receptor gene CHRNA4. *Neurosci Biobehav Rev, 36*(4), 1331-1341. doi: 10.1016/j.neubiorev.2012.02.010

Grucza, R. A., Wang, J. C., Stitzel, J. A., Hinrichs, A. L., Saccone, S. F., Saccone, N. L., . . . Bierut, L. J. (2008). A risk allele for nicotine dependence in CHRNA5 is a protective allele for cocaine dependence. *Biol Psychiatry, 64*(11), 922-929. doi: 10.1016/j.biopsych.2008.04.018

Haller, G., Kapoor, M., Budde, J., Xuei, X., Edenberg, H., Nurnberger, J., . . . Goate, A. (2013). Rare missense variants in CHRNB3 and CHRNA3 are associated with risk of alcohol and cocaine dependence. *Hum Mol Genet*. doi: 10.1093/hmg/ddt463

Hansen, S. T., & Mark, G. P. (2007). The nicotinic acetylcholine receptor antagonist mecamylamine prevents escalation of cocaine self-administration in rats with extended daily access. *Psychopharmacology (Berl), 194*(1), 53-61. doi: 10.1007/s00213-007-0822-z

Hartz, S. M., & Bierut, L. J. (2010). Genetics of addictions. *Clin Lab Med, 30*(4), 847-864. doi: 10.1016/j.cll.2010.07.005

Hartz, S. M., Lin, P., Edenberg, H. J., Xuei, X., Rochberg, N., Saccone, S., . . . Consortium, Nimh Genetics Initiative Bipolar Disorder. (2011). Genetic association of bipolar disorder with the beta(3) nicotinic receptor subunit gene. *Psychiatr Genet, 21*(2), 77-84. doi: 10.1097/YPG.0b013e32834135eb

Hartz, S. M., Pato, C. N., Medeiros, H., Cavazos-Rehg, P., Sobell, J. L., Knowles, J. A., . . . Genomic Psychiatry Cohort, Consortium. (2014). Comorbidity of severe psychotic disorders with measures of substance use. *JAMA Psychiatry, 71*(3), 248-254. doi: 10.1001/jamapsychiatry.2013.3726

Hasin, D. S., O'Brien, C. P., Auriacombe, M., Borges, G., Bucholz, K., Budney, A., . . . Grant, B. F. (2013). DSM-5 criteria for substance use disorders: recommendations and rationale. *Am J Psychiatry, 170*(8), 834-851. doi: 10.1176/appi.ajp.2013.12060782

Heath, A. C., Cates, R., Martin, N. G., Meyer, J., Hewitt, J.K., Neale, M.C., & Eaves, L.J. (1993). Genetic Contribution to Risk of Smoking Initiation: Comparisons Across Birth Cohorts and Across Cultures. *J Subst Abuse, 5*, 221-246.

Heath, A. C., Kirk, K. M., Meyer, J. M., & Martin, N. G. (1999). Genetic and social determinants of initiation and age at onset of smoking in Australian twins. *Behav Genet, 29*(6), 395-407.

Heath, A. C., & Martin, N. G. (1993). Genetic models for the natural history of smoking: evidence for a genetic influence on smoking persistence. *Addict Behav, 18*(1), 19-34.

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

Hoft, N. R., Corley, R. P., McQueen, M. B., Schlaepfer, I. R., Huizinga, D., & Ehringer, M. A. (2009). Genetic association of the CHRNA6 and CHRNB3 genes with tobacco dependence in a nationally representative sample. *Neuropsychopharmacology, 34*(3), 698-706. doi: 10.1038/npp.2008.122

Hong, L. E., Hodgkinson, C. A., Yang, Y., Sampath, H., Ross, T. J., Buchholz, B., . . . Stein, E. A. (2010). A genetically modulated, intrinsic cingulate circuit supports human nicotine addiction. *Proc Natl Acad Sci U S A, 107*(30), 13509-13514. doi: 10.1073/pnas.1004745107

Howie, B., Marchini, J., & Stephens, M. (2011). Genotype imputation with thousands of genomes. *G3 (Bethesda), 1*(6), 457-470. doi: 10.1534/g3.111.001198

Howie, B. N., Donnelly, P., & Marchini, J. (2009). A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet, 5*(6), e1000529. doi: 10.1371/journal.pgen.1000529

Hoyle, E., Genn, R. F., Fernandes, C., & Stolerman, I. P. (2006). Impaired performance of alpha7 nicotinic receptor knockout mice in the five-choice serial reaction time task. *Psychopharmacology (Berl), 189*(2), 211-223. doi: 10.1007/s00213-006-0549-2

Huang, W., Ma, J. Z., Payne, T. J., Beuten, J., Dupont, R. T., & Li, M. D. (2008). Significant association of DRD1 with nicotine dependence. *Hum Genet, 123*(2), 133-140. doi: 10.1007/s00439-007-0453-9

Huang, W., Payne, T. J., Ma, J. Z., Beuten, J., Dupont, R. T., Inohara, N., & Li, M. D. (2009). Significant association of ANKK1 and detection of a functional polymorphism with nicotine dependence in an African-American sample. *Neuropsychopharmacology, 34*(2), 319-330. doi: 10.1038/npp.2008.37

Hutter, S., Vilella, A. J., & Rozas, J. (2006). Genome-wide DNA polymorphism analyses using VariScan. *BMC Bioinformatics, 7*, 409. doi: 10.1186/1471-2105-7-409

Jarvis, Martin (1991). Dependence on Smokeless Tobacco *Smoking and Tobacco Control Monographs, Chapter 4*, 239-244.

Jaworski, J.N., & Jones, D.C. (2006). The Role of CART in the Reward/Reinforcing Properties of Psychostimulants. *Peptides, 2006*, 1993-2004.

Jobling, M.A., Hurles, M.E., & Tyler-Smith, C. (2004). *Human Evoutionary Genetics: Origins, Peoples and Disease*. New York, NY: Garland Publishing.

Johnson, E. O., Chen, L. S., Breslau, N., Hatsukami, D., Robbins, T., Saccone, N. L., . . . Bierut, L. J. (2010). Peer smoking and the nicotinic receptor genes: an examination of genetic and environmental risks for nicotine dependence. *Addiction, 105*(11), 2014-2022. doi: 10.1111/j.1360-0443.2010.03074.x

Johnson, E. O., Morgan-Lopez, A. A., Breslau, N., Hatsukami, D. K., & Bierut, L. J. (2008). Test of measurement invariance of the FTND across demographic groups:

assessment, effect size, and prediction of cessation. *Drug Alcohol Depend, 93*(3), 260-270. doi: 10.1016/j.drugalcdep.2007.10.001

Jones, B. L., Raga, T. O., Liebert, A., Zmarz, P., Bekele, E., Danielsen, E. T., . . . Swallow, D. M. (2013). Diversity of lactase persistence alleles in Ethiopia: signature of a soft selective sweep. *Am J Hum Genet, 93*(3), 538-544. doi: 10.1016/j.ajhg.2013.07.008

Jutkiewicz, E. M., Nicolazzo, D. M., Kim, M. N., & Gnegy, M. E. (2008). Nicotine and amphetamine acutely cross-potentiate their behavioral and neurochemical responses in female Holtzman rats. *Psychopharmacology (Berl), 200*(1), 93-103. doi: 10.1007/s00213-008-1159-y

Kaprio, J., Hammar, N., Koskenvuo, M., Floderus-Myrhed, B., Langinvainio, H., & Sarna, S. (1982). Cigarette smoking and alcohol use in Finland and Sweden: a cross-national twin study. *Int J Epidemiol, 11*(4), 378-386.

Kendler, K. S., Myers, J., & Prescott, C. A. (2007). Specificity of genetic and environmental risk factors for symptoms of cannabis, cocaine, alcohol, caffeine, and nicotine dependence. *Arch Gen Psychiatry, 64*(11), 1313-1320. doi: 10.1001/archpsyc.64.11.1313

Kendler, K. S., Neale, M. C., Sullivan, P., Corey, L. A., Gardner, C. O., & Prescott, C. A. (1999). A population-based twin study in women of smoking initiation and nicotine dependence. *Psychol Med, 29*(2), 299-308.

Kenneson, A., Funderburk, J. S., & Maisto, S. A. (2013). Substance use disorders increase the odds of subsequent mood disorders. *Drug Alcohol Depend*. doi: 10.1016/j.drugalcdep.2013.06.011

Keskitalo-Vuokko, K., Pitkaniemi, J., Broms, U., Heliovaara, M., Aromaa, A., Perola, M., . . . Kaprio, J. (2011). Associations of nicotine intake measures with CHRN genes in Finnish smokers. *Nicotine Tob Res, 13*(8), 686-690. doi: 10.1093/ntr/ntr059

King, S. M., Keyes, M., Malone, S. M., Elkins, I., Legrand, L. N., Iacono, W. G., & McGue, M. (2009). Parental alcohol dependence and the transmission of adolescent behavioral disinhibition: a study of adoptive and non-adoptive families. *Addiction, 104*(4), 578-586. doi: 10.1111/j.1360-0443.2008.02469.x

Koob, G. F. (1996). Drug addiction: the yin and yang of hedonic homeostasis. *Neuron, 16*(5), 893-896.

Kryazhimskiy, S., & Plotkin, J. B. (2008). The population genetics of dN/dS. *PLoS Genet, 4*(12), e1000304. doi: 10.1371/journal.pgen.1000304

Kumar, V., Kim, K., Joseph, C., Kourrich, S., Yoo, S. H., Huang, H. C., . . . Takahashi, J. S. (2013). C57BL/6N mutation in cytoplasmic FMRP interacting protein 2 regulates cocaine response. *Science, 342*(6165), 1508-1512. doi: 10.1126/science.1245503

Kumari, V., Gray, J. A., ffytche, D. H., Mitterschiffthaler, M. T., Das, M., Zachariah, E., . . . Sharma, T. (2003). Cognitive effects of nicotine in humans: an fMRI study. *Neuroimage, 19*(3), 1002-1013.

Lambert, J.-C., Ibrahim-Verbaas, C.A., Harold, D., Initiative, European Alzheimer's Disease, Disease, Genetic and Environmental Risk in Alzheimer's, Consortium, Alzheimer's Disease Genetic, & Cohorts for, Heart and Aging Research in Genomic Epidemiology. (2013). Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat Genet, 45*(12), 1452-1458. doi: 10.1038/ng.2802

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., . . . Higgins, D. G. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics, 23*(21), 2947-2948. doi: 10.1093/bioinformatics/btm404

Lasser, K., Boyd, J. W., Woolhandler, S., Himmelstein, D. U., McCormick, D., & Bor, D. H. (2000). Smoking and mental illness: A population-based prevalence study. *JAMA, 284*(20), 2606-2610.

Laurie, C. C., Doheny, K. F., Mirel, D. B., Pugh, E. W., Bierut, L. J., Bhangale, T., . . . Investigators, Geneva. (2010). Quality control and quality assurance in genotypic data for genome-wide association studies. *Genet Epidemiol, 34*(6), 591-602. doi: 10.1002/gepi.20516

Leonard, S., Adler, L. E., Benhammou, K., Berger, R., Breese, C. R., Drebing, C., . . . Freedman, R. (2001). Smoking and mental illness. *Pharmacol Biochem Behav, 70*(4), 561-570.

Lesch, K. P., Timmesfeld, N., Renner, T. J., Halperin, R., Roser, C., Nguyen, T. T., . . . Jacob, C. (2008). Molecular genetics of adult ADHD: converging evidence from genome-wide association and extended pedigree linkage studies. *J Neural Transm, 115*(11), 1573-1585. doi: 10.1007/s00702-008-0119-3

Lessov, C. N., Swan, G. E., Ring, H. Z., Khroyan, T. V., & Lerman, C. (2004). Genetics and drug use as a complex phenotype. *Subst Use Misuse, 39*(10-12), 1515-1569.

Leverich, G.S., & Post, R. M. (2006). Course of Bipolar Illness After History of Childhood Trauma. *The Lancet, 367*, 1040-1041.

Levin, E. D., & Simon, B. B. (1998). Nicotinic acetylcholine involvement in cognitive function in animals. *Psychopharmacology (Berl), 138*(3-4), 217-230.

Levine, A., Huang, Y.Y., Drisaldi, B., Griffin Jr., E.A., Pollak, D.D., Xu, S., . . . Kandel, E.R. (2011). Molecular Mechanism for a Gateway Drug: Epigenetic Changes Initiated by Nicotine Prime Gene Expression by Cocaine. *Sci Transl Med, 3*, 107ra109. doi: 10.1126/scitranslmed.3003062

Li, M. D., Cheng, R., Ma, J. Z., & Swan, G. E. (2003). A meta-analysis of estimated genetic and environmental effects on smoking behavior in male and female adult twins. *Addiction, 98*(1), 23-31.

Li, W. H., Wu, C. I., & Luo, C. C. (1985). A new method for estimating synonymous and nonsynonymous rates of nucleotide substitution considering the relative likelihood of nucleotide and codon changes. *Mol Biol Evol, 2*(2), 150-174.

Li, Y., Shao, C., Zhang, D., Zhao, M., Lin, L., Yan, P., . . . Jin, L. (2006). The effect of dopamine D2, D5 receptor and transporter (SLC6A3) polymorphisms on the cue-elicited heroin craving in Chinese. *Am J Med Genet B Neuropsychiatr Genet, 141B*(3), 269-273. doi: 10.1002/ajmg.b.30264

Li, Y., Sheu, C. C., Ye, Y., de Andrade, M., Wang, L., Chang, S. C., . . . Yang, P. (2010). Genetic variants and risk of lung cancer in never smokers: a genome-wide association study. *Lancet Oncol, 11*(4), 321-330. doi: 10.1016/S1470-2045(10)70042-5

Li, Y., & Yang, P. (2011). GPC5 gene and its related pathways in lung cancer. *J Thorac Oncol, 6*(1), 2-5. doi: 10.1097/JTO.0b013e3181fd6b04

Lin, P., Hartz, S. M., Wang, J. C., Krueger, R. F., Foroud, T. M., Edenberg, H. J., . . . Investigators, Geneva. (2011). Copy number variation accuracy in genome-wide association studies. *Hum Hered, 71*(3), 141-147. doi: 10.1159/000324683

Lin, P. I., McInnis, M. G., Potash, J. B., Willour, V., MacKinnon, D. F., DePaulo, J. R., & Zandi, P. P. (2006). Clinical correlates and familial aggregation of age at onset in bipolar disorder. *Am J Psychiatry, 163*(2), 240-246. doi: 10.1176/appi.ajp.163.2.240

Liu, X., Ong, R. T., Pillai, E. N., Elzein, A. M., Small, K. S., Clark, T. G., . . . Teo, Y. Y. (2013). Detecting and characterizing genomic signatures of positive selection in global populations. *Am J Hum Genet, 92*(6), 866-881. doi: 10.1016/j.ajhg.2013.04.021

Ma, J. Z., Payne, T. J., Nussbaum, J., & Li, M. D. (2010). Significant association of glutamate receptor, ionotropic N-methyl-D-aspartate 3A (GRIN3A), with nicotine dependence in European- and African-American smokers. *Hum Genet, 127*(5), 503-512. doi: 10.1007/s00439-010-0787-6

Mackay, J., & Eriksen, M. (2002). World Health Organization: The Tobacco Atlas. from http://whqlibdoc.who.int/publications/2002/9241562099.pdf.

Mah, S. P., Nelson, M.R., DeLisi, L.E., Reneland, R.H., Markward, N., James, M.R., . . . Braun, A. (2006). Identification of the Semaphorin Receptor PLXNA2 as a Candidate for Susceptibility to Schizophrenia. *Mol Psychiatry, 11*, 471-478. doi: 10.1038/sj.mp.4001785

Malarcher, A., Dube, S., Shaw, L., Babb., S., & Kaufmann., R. (2011). Quitting Smoking Among Adults, 2001-2010. *Morbidity and Mortality Weekly Report, 60*(44), 1513-1519.

Mathers, C., Stevens, G., Tursan d'Espaignet, E., & Wolfenden L. (2012). World Health Organization global report: mortality attributable to tobacco.

Merikangas, K. R., Stolar, M., Stevens, D. E., Goulet, J., Preisig, M. A., Fenton, B., . . . Rounsaville, B. J. (1998). Familial transmission of substance use disorders. *Arch Gen Psychiatry, 55*(11), 973-979.

Mitchell, H. (2011). Tobacco Consumption in India - An Overview. from http://www.tobaccoasia.com/index.php?option=com_content&view=article&id=389%3Atobacco-consumption-in-india-an-overview&catid=83%3Aarticles-q2-11&Itemid=65

Monakhov, M., Golimbet, V., Abramova, L., Kaleda, V., & Karpov, V. (2008). Association study of three polymorphisms in the dopamine D2 receptor gene and schizophrenia in the Russian population. *Schizophr Res, 100*(1-3), 302-307. doi: 10.1016/j.schres.2008.01.007

Murphy, F. C., & Klein, R. M. (1998). The effects of nicotine on spatial and non-spatial expectancies in a covert orienting task. *Neuropsychologia, 36*(11), 1103-1114.

Mwenifumbo, J. C., & Tyndale, R. F. (2009). Molecular genetics of nicotine metabolism. *Handb Exp Pharmacol*(192), 235-259. doi: 10.1007/978-3-540-69248-5_9

Nagoshi, C. (2011). Behavioral Genetics. *Encyclopedia of Human Behavior, Article No.: 00050*.

NCI. (2010). Smokeless Tobacco or Health: An International Perspective. from http://www.cancer.gov/cancertopics/factsheet/Tobacco/smokeless

Nesse, R. M., & Berridge, K. C. (1997). Psychoactive drug use in evolutionary perspective. *Science, 278*(5335), 63-66.

Newhouse, P. A., Potter, A., Kelton, M., & Corwin, J. (2001). Nicotinic treatment of Alzheimer's disease. *Biol Psychiatry, 49*(3), 268-278.

Noble, E. P., Ozkaragoz, T. Z., Ritchie, T. L., Zhang, X., Belin, T. R., & Sparkes, R. S. (1998). D2 and D4 dopamine receptor polymorphisms and personality. *Am J Med Genet, 81*(3), 257-267.

Osler, M., Holst, C., Prescott, E., & Sorensen, T. I. (2001). Influence of genes and family environment on adult smoking behavior assessed in an adoption study. *Genet Epidemiol, 21*(3), 193-200. doi: 10.1002/gepi.1028

Pan, Y., Wang, K. S., & Aragam, N. (2011). NTM and NR3C2 polymorphisms influencing intelligence: family-based association studies. *Prog Neuropsychopharmacol Biol Psychiatry, 35*(1), 154-160. doi: 10.1016/j.pnpbp.2010.10.016

Perez de los Cobos, J., Baiget, M., Trujols, J., Sinol, N., Volpini, V., Banuls, E., . . . Alvarez, E. (2007). Allelic and genotypic associations of DRD2 TaqI A polymorphism with heroin dependence in Spanish subjects: a case control study. *Behav Brain Funct, 3*, 25. doi: 10.1186/1744-9081-3-25

Pomerleau, C. S. (1997). Co-factors for smoking and evolutionary psychobiology. *Addiction, 92*(4), 397-408.

Post, R. M., & Kalivas, P. (2013). Bipolar disorder and substance misuse: pathological and therapeutic implications of their comorbidity and cross-sensitisation. *Br J Psychiatry, 202*(3), 172-176. doi: 10.1192/bjp.bp.112.116855

Preller, K. H., Herdener, M., Schilbach, L., Stampfli, P., Hulka, L. M., Vonmoos, M., . . . Quednow, B. B. (2014). Functional changes of the reward system underlie blunted response to social gaze in cocaine users. *Proc Natl Acad Sci U S A, 111*(7), 2842-2847. doi: 10.1073/pnas.1317090111

Preuss, U. W., Zill, P., Koller, G., Bondy, B., & Sokya, M. (2007). D2 dopamine receptor gene haplotypes and their influence on alcohol and tobacco consumption magnitude in alcohol-dependent individuals. *Alcohol Alcohol, 42*(3), 258-266. doi: 10.1093/alcalc/agm030

Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. A., & Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet, 38*(8), 904-909. doi: 10.1038/ng1847

Quik, M., Vailati, S., Bordia, T., Kulak, J.M., Fan, H., McIntosh, J.M., . . . Gotti, C. (2005). Subunit Composition of Nicotinic Receptors in Monkey Striatum: Effect

of Treatments with 1-Methyl-4-phenyl-1,2,3,6-tetrahydropyridine or L-DOPA. *Mol Pharmacol, 67*, 32-41.

Radwan, G. N., El-Setouhy, M., Mohamed, M. K., Hamid, M. A., Azem, S. A., Kamel, O., . . . Loffredo, C. A. (2007). DRD2/ANKK1 TaqI polymorphism and smoking behavior of Egyptian male cigarette smokers. *Nicotine Tob Res, 9*(12), 1325-1329. doi: 10.1080/14622200701704889

Regier, D. A., Farmer, M. E., Rae, D. S., Locke, B. Z., Keith, S. J., Judd, L. L., & Goodwin, F. K. (1990). Comorbidity of mental disorders with alcohol and other drug abuse. Results from the Epidemiologic Catchment Area (ECA) Study. *JAMA, 264*(19), 2511-2518.

Rice, J. P., Hartz, S. M., Agrawal, A., Almasy, L., Bennett, S., Breslau, N., . . . Consortium, Geneva. (2012). CHRNB3 is more strongly associated with Fagerstrom test for cigarette dependence-based nicotine dependence than cigarettes per day: phenotype definition changes genome-wide association studies results. *Addiction, 107*(11), 2019-2028. doi: 10.1111/j.1360-0443.2012.03922.x

Rigbi, A., Kanyas, K., Yakir, A., Greenbaum, L., Pollak, Y., Ben-Asher, E., . . . Lerer, B. (2008). Why do young women smoke? V. Role of direct and interactive effects of nicotinic cholinergic receptor gene variation on neurocognitive function. *Genes Brain Behav, 7*(2), 164-172. doi: 10.1111/j.1601-183X.2007.00329.x

Ripke, S., O'Dushlaine, C., Chambert, K., Moran, J. L., Kahler, A. K., Akterin, S., . . . Sullivan, P. F. (2013). Genome-wide association analysis identifies 13 new risk loci for schizophrenia. *Nat Genet, 45*(10), 1150-1159. doi: 10.1038/ng.2742

Sabeti, P. C., Reich, D. E., Higgins, J. M., Levine, H. Z., Richter, D. J., Schaffner, S. F., . . . Lander, E. S. (2002). Detecting recent positive selection in the human genome from haplotype structure. *Nature, 419*(6909), 832-837. doi: 10.1038/nature01140

Saccone, N. L., Culverhouse, R. C., Schwantes-An, T. H., Cannon, D. S., Chen, X., Cichon, S., . . . Bierut, L. J. (2010). Multiple independent loci at chromosome 15q25.1 affect smoking quantity: a meta-analysis and comparison with lung cancer and COPD. *PLoS Genet, 6*(8). doi: 10.1371/journal.pgen.1001053

Saccone, N. L., Neuman, R. J., Saccone, S. F., & Rice, J. P. (2003). Genetic analysis of maximum cigarette-use phenotypes. *BMC Genet, 4 Suppl 1*, S105. doi: 10.1186/1471-2156-4-S1-S105

Saccone, N. L., Schwantes-An, T. H., Wang, J. C., Grucza, R. A., Breslau, N., Hatsukami, D., . . . Bierut, L. J. (2010). Multiple cholinergic nicotinic receptor genes affect nicotine dependence risk in African and European Americans. *Genes Brain Behav, 9*(7), 741-750. doi: 10.1111/j.1601-183X.2010.00608.x

Saccone, N. L., Wang, J. C., Breslau, N., Johnson, E. O., Hatsukami, D., Saccone, S. F., . . . 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 Res, 69*(17), 6848-6856. doi: 10.1158/0008-5472.CAN-09-0786

Saccone, S. F., Hinrichs, A. L., Saccone, N. L., Chase, G. A., Konvicka, K., Madden, P. A., . . . Bierut, L. J. (2007). Cholinergic nicotinic receptor genes implicated in a nicotine dependence association study targeting 348 candidate genes with 3713 SNPs. *Hum Mol Genet, 16*(1), 36-49. doi: 10.1093/hmg/ddl438

Sadler, B., Haller, G., Agrawal, A., Culverhouse, R. C., Bucholz, K., Brooks, A., . . . Goate, A. (2014). Variants near CHRNB3-CHRNA6 are associated with DSM-5 cocaine use disorder: evidence for pleiotropy. *Scientific Reports, 4*, 4497. doi: 10.1038/SREP04497

Salthouse, T. A. (1992). What do adult age differences in the Digit Symbol Substitution Test reflect? *J Gerontol, 47*(3), P121-128.

Santos, G. C., Marin, M. T., Cruz, F. C., Delucia, R., & Planeta, C. S. (2009). Amphetamine- and nicotine-induced cross-sensitization in adolescent rats persists until adulthood. *Addict Biol, 14*(3), 270-275. doi: 10.1111/j.1369-1600.2009.00153.x

Sherva, R., Kranzler, H. R., Yu, Y., Logue, M. W., Poling, J., Arias, A. J., . . . Gelernter, J. (2010). Variation in nicotinic acetylcholine receptor genes is associated with multiple substance dependence phenotypes. *Neuropsychopharmacology, 35*(9), 1921-1931. doi: 10.1038/npp.2010.64

Smith, L., Watson, M., Gates, S., Ball, D., & Foxcroft, D. (2008). Meta-analysis of the association of the Taq1A polymorphism with the risk of alcohol dependency: a HuGE gene-disease association review. *Am J Epidemiol, 167*(2), 125-138. doi: 10.1093/aje/kwm281

Spitz, M. R., Amos, C. I., Dong, Q., Lin, J., & Wu, X. (2008). The CHRNA5-A3 region on chromosome 15q24-25.1 is a risk factor both for nicotine dependence and for lung cancer. *J Natl Cancer Inst, 100*(21), 1552-1556. doi: 10.1093/jnci/djn363

Spitz, M. R., Shi, H., Yang, F., Hudmon, K. S., Jiang, H., Chamberlain, R. M., . . . Wu, X. (1998). Case-control study of the D2 dopamine receptor gene and smoking status in lung cancer patients. *J Natl Cancer Inst, 90*(5), 358-363.

Subhedar, N.K., Nakhate, K.T., Upadhya, M.A., & Kokare, D.M. (2014). CART in the Brain of Vertebrates: Circuits, Functions and Evolution. *Peptides, 54*, 108-130.

Sullivan, P. F., Neale, B. M., van den Oord, E., Miles, M. F., Neale, M. C., Bulik, C. M., . . . Kendler, K. S. (2004). Candidate genes for nicotine dependence via linkage, epistasis, and bioinformatics. *Am J Med Genet B Neuropsychiatr Genet, 126B*(1), 23-36. doi: 10.1002/ajmg.b.20138

Sullivan, R. J., & Hagen, E. H. (2002). Psychotropic substance-seeking: evolutionary pathology or adaptation? *Addiction, 97*(4), 389-400.

Svanes, C., Omenaas, E., Jarvis, D., Chinn, S., Gulsvik, A., & Burney, P. (2004). Parental smoking in childhood and adult obstructive lung disease: results from the European Community Respiratory Health Survey. *Thorax, 59*(4), 295-302.

Swan, G. E., Hops, H., Wilhelmsen, K. C., Lessov-Schlaggar, C. N., Cheng, L. S., Hudmon, K. S., . . . Benowitz, N. (2006). A genome-wide screen for nicotine dependence susceptibility loci. *Am J Med Genet B Neuropsychiatr Genet, 141B*(4), 354-360. doi: 10.1002/ajmg.b.30315

Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics, 123*(3), 585-595.

Tanda, G., Pontieri, F. E., & Di Chiara, G. (1997). Cannabinoid and heroin activation of mesolimbic dopamine transmission by a common mu1 opioid receptor mechanism. *Science, 276*(5321), 2048-2050.

Thiel, C. M., Zilles, K., & Fink, G. R. (2005). Nicotine modulates reorienting of visuospatial attention and neural activity in human parietal cortex. *Neuropsychopharmacology, 30*(4), 810-820. doi: 10.1038/sj.npp.1300633

Thorgeirsson, T. E., Geller, F., Sulem, P., Rafnar, T., Wiste, A., Magnusson, K. P., . . . Stefansson, K. (2008). A variant associated with nicotine dependence, lung cancer and peripheral arterial disease. *Nature, 452*(7187), 638-642. doi: 10.1038/nature06846

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

Tsuang, M. T., Bar, J. L., Harley, R. M., & Lyons, M. J. (2001). The Harvard Twin Study of Substance Abuse: what we have learned. *Harv Rev Psychiatry, 9*(6), 267-279.

Tsuang, M. T., Lyons, M. J., Meyer, J. M., Doyle, T., Eisen, S. A., Goldberg, J., . . . Eaves, L. (1998). Co-occurrence of abuse of different drugs in men: the role of drug-specific and shared vulnerabilities. *Arch Gen Psychiatry, 55*(11), 967-972.

Tyas, S. L. (1996). Are tobacco and alcohol use related to Alzheimer's disease? A critical assessment of the evidence and its implications. *Addict Biol, 1*(3), 237-254. doi: 10.1080/1355621961000124856

Verdejo-Garcia, A. (2014). Social cognition in cocaine addiction. *Proc Natl Acad Sci U S A, 111*(7), 2406-2407. doi: 10.1073/pnas.1324287111

Vink, J. M., Posthuma, D., Neale, M. C., Eline Slagboom, P., & Boomsma, D. I. (2006). Genome-wide linkage scan to identify Loci for age at first cigarette in Dutch sibling pairs. *Behav Genet, 36*(1), 100-111. doi: 10.1007/s10519-005-9012-0

Visscher, P. M., Brown, M. A., McCarthy, M. I., & Yang, J. (2012). Five years of GWAS discovery. *Am J Hum Genet, 90*(1), 7-24. doi: 10.1016/j.ajhg.2011.11.029

Voight, B. F., Kudaravalli, S., Wen, X., & Pritchard, J. K. (2006). A map of recent positive selection in the human genome. *PLoS Biol, 4*(3), e72. doi: 10.1371/journal.pbio.0040072

Wagner, M., Schulze-Rauschenbach, S., Petrovsky, N., Brinkmeyer, J., von der Goltz, C., Grunder, G., . . . Winterer, G. (2013). Neurocognitive impairments in non-deprived smokers--results from a population-based multi-center study on smoking-related behavior. *Addict Biol, 18*(4), 752-761. doi: 10.1111/j.1369-1600.2011.00429.x

Wang, J. C., Cruchaga, C., Saccone, N. L., Bertelsen, S., Liu, P., Budde, J. P., . . . collaborators, Gelcc. (2009). Risk for nicotine dependence and lung cancer is conferred by mRNA expression levels and amino acid change in CHRNA5. *Hum Mol Genet, 18*(16), 3125-3135. doi: 10.1093/hmg/ddp231

Wang, J. C., Kapoor, M., & Goate, A. M. (2012). The genetics of substance dependence. *Annu Rev Genomics Hum Genet, 13*, 241-261. doi: 10.1146/annurev-genom-090711-163844

Wang, J.-C., Spiegel, N., Bertelsen, S., Le, N., McKenna, N., Budde, J.P., . . . Goate, A. (2013). Cis-Regulatory Variants Affect CHRNA5 mRNA Expression in Populations of African and European Ancestry. *PLoS One, 8*, e80204.

Weiss, R. B., Baker, T. B., Cannon, D. S., von Niederhausern, A., Dunn, D. M., Matsunami, N., . . . Leppert, M. F. (2008). A candidate gene approach identifies the CHRNA5-A3-B4 region as a risk factor for age-dependent nicotine addiction. *PLoS Genet, 4*(7), e1000125. doi: 10.1371/journal.pgen.1000125

WHO. (2011). World Health Organization Tobacco Factsheet. from http://www.who.int/mediacentre/factsheets/fs339/en/index.html

Williamson, S. H., Hubisz, M. J., Clark, A. G., Payseur, B. A., Bustamante, C. D., & Nielsen, R. (2007). Localizing recent adaptive evolution in the human genome. *PLoS Genet, 3*(6), e90. doi: 10.1371/journal.pgen.0030090

Winterer, G., Mittelstrass, K., Giegling, I., Lamina, C., Fehr, C., Brenner, H., . . . Rujescu, D. (2010). Risk gene variants for nicotine dependence in the CHRNA5-CHRNA3-CHRNB4 cluster are associated with cognitive performance. *Am J Med Genet B Neuropsychiatr Genet, 153B*(8), 1448-1458. doi: 10.1002/ajmg.b.31126

Wise, R. A. (1996). Addictive drugs and brain stimulation reward. *Annu Rev Neurosci, 19*, 319-340. doi: 10.1146/annurev.ne.19.030196.001535

Wray, N. R., James, M. R., Mah, S. P., Nelson, M., Andrews, G., Sullivan, P. F., . . . Martin, N. G. (2007). Anxiety and comorbid measures associated with PLXNA2. *Arch Gen Psychiatry, 64*(3), 318-326. doi: 10.1001/archpsyc.64.3.318

Xu, K., Lichtermann, D., Lipsky, R. H., Franke, P., Liu, X., Hu, Y., . . . Goldman, D. (2004). Association of specific haplotypes of D2 dopamine receptor gene with vulnerability to heroin dependence in 2 distinct populations. *Arch Gen Psychiatry, 61*(6), 597-606. doi: 10.1001/archpsyc.61.6.597

Yakir, A., Rigbi, A., Kanyas, K., Pollak, Y., Kahana, G., Karni, O., . . . Lerer, B. (2007). Why do young women smoke? III. Attention and impulsivity as neurocognitive predisposing factors. *Eur Neuropsychopharmacol, 17*(5), 339-351. doi: 10.1016/j.euroneuro.2006.09.004

Yang, Z., & Bielawski, J. P. (2000). Statistical methods for detecting molecular adaptation. *Trends Ecol Evol, 15*(12), 496-503.

Young, J. W., Meves, J. M., Tarantino, I. S., Caldwell, S., & Geyer, M. A. (2011). Delayed procedural learning in alpha7-nicotinic acetylcholine receptor knockout mice. *Genes Brain Behav, 10*(7), 720-733. doi: 10.1111/j.1601-183X.2011.00711.x

Yu, Y., Panhuysen, C., Kranzler, H. R., Hesselbrock, V., Rounsaville, B., Weiss, R., . . . Gelernter, J. (2006). Intronic variants in the dopa decarboxylase (DDC) gene are associated with smoking behavior in European-Americans and African-Americans. *Hum Mol Genet, 15*(14), 2192-2199. doi: 10.1093/hmg/ddl144

Zeiger, J. S., Haberstick, B. C., Schlaepfer, I., Collins, A. C., Corley, R. P., Crowley, T. J., . . . Ehringer, M. A. (2008). The neuronal nicotinic receptor subunit genes (CHRNA6 and CHRNB3) are associated with subjective responses to tobacco. *Hum Mol Genet, 17*(5), 724-734. doi: 10.1093/hmg/ddm344

Zeron-Medina, J., Wang, X., Repapi, E., Campbell, M. R., Su, D., Castro-Giner, F., . . . Bond, G. L. (2013). A polymorphic p53 response element in KIT ligand influences cancer risk and has undergone natural selection. *Cell, 155*(2), 410-422. doi: 10.1016/j.cell.2013.09.017

Zhang, Z., Li, J., Zhao, X.-Q., Wang, J., Wong, G. K.-S., & Yu, J. (2006). KaKs Calculator: Calculating Ka and Ks Through Model Selection and Model Averaging. *Geno. Prot. Bioinfo., 4*, 259-263.

Zheng, Y., Kan, M., Yu, L., Niu, X., Zhou, D., He, L., . . . Liu, Y. (2012). GPC5 rs2352028 polymorphism and risk of lung cancer in Han Chinese. *Cancer Invest, 30*(1), 13-19. doi: 10.3109/07357907.2011.630052

APPENDIX A

GENOME-WIDE TOP 50 SNPS FROM FTND NICOTINE DEPENDENCE ASSOCIATION ANALYSIS

CHR	Position	*Gene*	rs number	A1	A2	Freq	INFO	OR	SE	Mega P-value	Meta P-value
15	*78874842*	*CHRNA5-A3-B4*	*chr15:78874842:D*	*AG*	*A*	*0.27*	*1.02*	*1.45*	*0.07*	*2.93E-08*	*3.14E-08*
13	*91955562*	*GPC5*	*rs7995715*	*T*	*G*	*0.32*	*1.07*	*1.41*	*0.06*	*3.27E-08*	*9.20E-08*
15	*78901113*	*CHRNA5-A3-B4*	*rs114205691*	*C*	*T*	*0.31*	*1.03*	*1.41*	*0.06*	*4.35E-08*	*4.14E-08*
15	78867482	*CHRNA5-A3-B4*	rs17486278	A	C	0.34	1.02	1.39	0.06	4.91E-08	5.51E-08
15	78900650	*CHRNA5-A3-B4*	rs147499554	C	T	0.27	0.98	1.45	0.07	5.33E-08	5.05E-08
15	78900647	*CHRNA5-A3-B4*	rs141518190	A	G	0.27	0.98	1.44	0.07	5.33E-08	5.05E-08
15	78851615	*CHRNA5-A3-B4*	rs2036527	G	A	0.32	1.03	1.39	0.06	7.54E-08	5.80E-08
15	78882925	*CHRNA5-A3-B4*	rs16969968	G	A	0.27	1.09	1.43	0.07	7.82E-08	1.02E-07
15	78857986	*CHRNA5-A3-B4*	rs55781567	C	G	0.33	1.02	1.39	0.06	9.70E-08	1.08E-07
15	78898932	*CHRNA5-A3-B4*	rs55676755	C	G	0.30	1.04	1.40	0.06	9.74E-08	9.86E-08
15	78862064	*CHRNA5-A3-B4*	rs11633958	C	T	0.27	1.08	1.43	0.07	9.95E-08	1.31E-07
15	78886198	*CHRNA5-A3-B4*	rs8192482	C	T	0.27	1.09	1.43	0.07	1.03E-07	1.34E-07
15	78886947	*CHRNA5-A3-B4*	rs4887067	G	A	0.27	1.09	1.43	0.07	1.03E-07	1.34E-07
15	78857939	*CHRNA5-A3-B4*	rs55853698	T	G	0.28	1.07	1.42	0.07	1.08E-07	1.06E-07
15	78923987	*CHRNA5-A3-B4*	rs17487223	C	T	0.31	1.03	1.41	0.07	1.21E-07	1.70E-07
15	78849779	*CHRNA5-A3-B4*	rs72740955	C	T	0.29	1.06	1.40	0.06	1.29E-07	1.07E-07
15	78866445	*CHRNA5-A3-B4*	rs140330585	G	A	0.29	1.06	1.40	0.06	1.36E-07	1.41E-07
15	78922638	*CHRNA5-A3-B4*	rs2869548	G	A	0.30	1.07	1.42	0.07	1.48E-07	1.61E-07
15	78865197	*CHRNA5-A3-B4*	rs17486195	A	G	0.29	1.06	1.40	0.06	1.52E-07	1.43E-07
15	78862453	*CHRNA5-A3-B4*	rs7172118	C	A	0.29	1.06	1.40	0.06	1.74E-07	1.41E-07
13	91965134	*GPC5*	rs9515908	C	T	0.29	0.99	1.40	0.07	1.96E-07	3.57E-07
13	91948047	*GPC5*	rs9523295	G	A	0.30	1.05	1.38	0.06	2.15E-07	4.04E-07
13	91931922	*GPC5*	rs9523288	C	T	0.18	1.04	1.49	0.08	2.15E-07	6.79E-07
13	91929736	*GPC5*	rs67147421	A	C	0.18	1.03	1.49	0.08	2.17E-07	6.82E-07
15	78894339	*CHRNA5-A3-B4*	rs1051730	G	A	0.29	1.06	1.40	0.06	2.18E-07	1.63E-07
15	78878541	*CHRNA5-A3-B4*	rs951266	G	A	0.29	1.07	1.40	0.06	2.21E-07	1.75E-07
15	78899003	*CHRNA5-A3-B4*	rs56077333	C	A	0.30	0.99	1.40	0.06	2.24E-07	1.93E-07
15	78873993	*CHRNA5-A3-B4*	rs7180002	A	T	0.29	1.07	1.40	0.06	2.25E-07	1.79E-07
13	91935848	*GPC5*	rs9523289	G	A	0.19	1.02	1.49	0.08	2.26E-07	7.06E-07
13	91952459	*GPC5*	rs9301726	T	C	0.30	1.05	1.38	0.06	2.28E-07	4.44E-07
13	91953042	*GPC5*	rs7994634	C	T	0.30	1.05	1.38	0.06	2.28E-07	4.46E-07
13	91950114	*GPC5*	rs7986895	C	A	0.30	1.05	1.38	0.06	2.31E-07	4.44E-07
13	91957035	*GPC5*	rs7335045	A	G	0.30	1.06	1.38	0.06	2.35E-07	4.84E-07
13	91956188	*GPC5*	rs9589183	C	T	0.30	1.06	1.38	0.06	2.38E-07	4.73E-07

13	91956038	*GPC5*	rs7989842	G	T	0.30	1.06	1.38	0.06	2.38E-07	4.77E-07
13	91955192	*GPC5*	rs9523299	G	A	0.30	1.06	1.38	0.06	2.42E-07	4.57E-07
13	91963080	*GPC5*	rs1475655	A	T	0.29	1.01	1.40	0.06	2.60E-07	5.16E-07
13	91949444	*GPC5*	rs9523296	G	A	0.30	1.05	1.38	0.06	2.77E-07	5.36E-07
15	78900701	*CHRNA5-A3-B4*	rs138544659	T	G	0.27	0.96	1.42	0.07	2.84E-07	2.54E-07
13	91955287	*GPC5*	chr13:91955287:I	A	AT	0.30	1.05	1.38	0.06	3.05E-07	5.81E-07
15	78900908	*CHRNA5-A3-B4*	rs147144681	C	T	0.29	0.99	1.40	0.07	3.18E-07	2.44E-07
13	91952853	*GPC5*	chr13:91952853:D	CA	C	0.30	1.05	1.37	0.06	4.15E-07	7.95E-07
13	91953719	*GPC5*	rs1332216	T	C	0.33	1.07	1.37	0.06	4.57E-07	1.02E-06
13	91925668	*GPC5*	rs73599638	G	A	0.18	1.04	1.47	0.08	5.06E-07	1.63E-06
4	125635331	*ANKRD50*	rs72931230	A	G	0.04	1.04	0.46	0.16	7.27E-07	8.54E-07
11	131852596	*NTM*	rs56361772	G	C	0.53	0.95	1.37	0.06	7.74E-07	8.52E-07
13	91898554	*GPC5*	chr13:91898554:I	C	CT	0.19	1.01	1.47	0.08	8.84E-07	4.91E-06
9	23512657	*ELAV*	rs274918	T	C	0.12	1.07	0.62	0.10	9.00E-07	2.21E-06
12	25471228	*KRAS*	rs11047952	T	A	0.39	1.00	1.36	0.06	9.22E-07	7.95E-07
13	91908725	*GPC5*	rs68126334	C	T	0.18	1.03	1.45	0.08	1.39E-06	4.28E-06

APPENDIX B

ASSOCIATION RESULTS FOR DSM-5 COCAINE USE DISORDER, DSM-5 ALCOHOL USE DISORDER AND FTND NICOTINE DEPENDENCE

SNP	LD Bin	BP	Freq	R-DSM-5 Alc	P-DSM-5 Alc	OR-DSM-5 Coc	P-DSM-5 – Coc	OR-FTND_DX	P-FTND_DX
rs9298626	1	42,647,165	0.04	1.05	0.84	2.62	***2.34E-04***	0.46	***7.52E-04***
rs7844824	1	42,672,170	0.04	1.1	0.67	2.65	***2.43E-04***	0.44	***4.50E-04***
rs4305884	2	42,637,880	0.06	1.11	0.58	2.13	***2.69E-04***	0.53	***7.17E-04***
rs7824160	1	42,705,413	0.04	1.1	0.66	2.5	***4.88E-04***	0.45	***5.11E-04***
rs11986893	4	42,772,016	0.2	1.04	0.7	1.56	***4.92E-04***	0.69	***3.79E-04***
rs7002907	1	42,702,998	0.04	1.1	0.67	2.49	***5.03E-04***	0.45	***5.03E-04***
rs6997994	1	42,702,328	0.04	1.1	0.67	2.49	***5.04E-04***	0.45	***4.96E-04***
rs7815274	1	42,701,740	0.04	1.07	0.77	2.47	***6.41E-04***	0.47	***1.04E-03***
rs4952	1	42,706,222	0.04	1.14	0.57	2.43	***8.12E-04***	0.47	***1.17E-03***
rs10107450	5	42,749,052	0.22	1.11	0.33	1.5	***1.04E-03***	0.64	***1.22E-05***
rs1868859	2	42,634,958	0.07	1.21	0.27	1.85	***1.36E-03***	0.59	***1.95E-03***
rs892413	3	42,733,535	0.2	1.04	0.71	1.48	2.11E-03	0.69	***5.92E-04***
rs4950	6	42,671,790	0.22	1.07	0.54	1.42	5.93E-03	0.66	***7.32E-05***
rs13280604	6	42,678,743	0.22	1.07	0.51	1.41	7.49E-03	0.66	***9.11E-05***
rs1530848	6	42,672,065	0.22	1.06	0.58	1.4	7.53E-03	0.67	***1.08E-04***
rs2196128	3	42,737,443	0.23	1.03	0.79	1.37	8.83E-03	0.75	3.89E-03
rs6474414	6	42,679,493	0.22	1.05	0.64	1.39	9.67E-03	0.67	***1.10E-04***
rs6997909	6	42,679,406	0.22	1.05	0.64	1.39	9.67E-03	0.67	***1.10E-04***
rs4736835	6	42,666,190	0.22	1.06	0.55	1.39	1.03E-02	0.66	***8.22E-05***
rs9298628	3	42,725,148	0.21	1.06	0.62	1.38	1.04E-02	0.72	***1.68E-03***
rs6474415	6	42,682,095	0.22	1.05	0.67	1.38	1.05E-02	0.67	***1.01E-04***
rs1451240	6	42,665,868	0.22	1.06	0.55	1.38	1.14E-02	0.66	***8.03E-05***

rs7004381	6	42,670,318	0.22	1.06	0.58	1.38	1.24E-02	0.66	***1.03E-04***
rs13273442	6	42,663,174	0.22	1.07	0.55	1.38	1.25E-02	0.67	***1.35E-04***
rs1955185	6	42,668,804	0.22	1.06	0.59	1.37	1.30E-02	0.67	***1.02E-04***
rs6474413	6	42,670,221	0.22	1.05	0.62	1.37	1.31E-02	0.66	***8.87E-05***
rs16891620	7	42,744,820	0.13	1.07	0.59	1.42	2.00E-02	0.77	3.54E-02
rs10958726	6	42,655,066	0.21	1.07	0.53	1.34	2.20E-02	0.67	***1.60E-04***
rs6474421	12	42,776,255	0.07	1.31	0.14	1.5	2.62E-02	1.01	9.34E-01
rs10958725	6	42,643,741	0.22	1.07	0.55	1.31	3.37E-02	0.68	***2.22E-04***
rs7012713	13	42,711,460	0.04	1.18	0.49	1.65	4.66E-02	1.26	2.75E-01
rs7825957	10	42,793,695	0.04	0.8	0.35	1.53	1.08E-01	0.75	2.34E-01
rs7013926	9	42,620,874	0.07	1.33	0.1	0.74	1.38E-01	0.78	1.22E-01
rs7017215	15	42,785,365	0.03	0.81	0.38	1.41	2.09E-01	0.71	1.53E-01
rs1376442	8	42,628,754	0.13	1.18	0.21	0.87	3.69E-01	0.78	5.40E-02
rs7844566	21	42,608,999	0.06	1.64	0.01	1.18	4.42E-01	0.63	1.00E-02
rs10105699	14	42,759,051	0.07	0.88	0.46	1.15	4.68E-01	0.94	7.21E-01
rs16891604	11	42,737,870	0.05	1.11	0.61	1.17	4.70E-01	1.25	2.39E-01
rs11785591	16	42,626,897	0.08	1.18	0.3	0.87	4.79E-01	0.82	2.24E-01
rs16891454	17	42,604,697	0.02	2.01	0.04	1.23	5.33E-01	0.71	2.12E-01
rs4737055	19	42,602,483	0.05	1.15	0.46	0.88	6.09E-01	0.83	3.42E-01
rs1530847	21	42,667,396	0.16	1.02	0.85	1.08	6.13E-01	0.77	2.83E-02
rs6989472	22	42,624,623	0.18	1.19	0.12	1.06	6.95E-01	0.68	***5.61E-04***
rs2084706	18	42,619,920	0.07	0.94	0.68	0.95	7.99E-01	0.79	1.59E-01
rs6995005	18	42,617,813	0.07	0.91	0.55	0.97	8.72E-01	0.79	1.71E-01
rs7825128	20	42,778,205	0.08	0.99	0.95	1.03	8.79E-01	0.71	3.50E-02

APPENDIX C

IHS VALUES FOR LD BIN ASSOCIATED WITH COCAINE DEPENDENCE ON CHROMOSOME 8

SNP	dbSNP Functional Annotation	Frequency In 1000 Genomes EUR	iHS Value EUR	Frequency In 1000 Genomes ASN	iHS Value ASN	Frequency In 1000 Genomes AFR	iHS Value AFR
rs62516708	intergenic	0.0396	0.299	0	n/a	0.0955	0.867
rs13439675	intergenic	0.0383	-0.113	0	n/a	0.0163	0.268
rs9298625	intergenic	0.0396	0.019	0	n/a	0.0955	1.217
rs9298626	intergenic	0.9604	0.518	0	n/a	0.9045	-0.447
rs1451241	intergenic	0.0383	0.565	0	n/a	0.0163	0.362
rs1451242	intergenic	0.0396	0.671	0	n/a	0.0955	1.329
rs28616109	intergenic	0.0396	0.729	0	n/a	0.0955	1.317
rs62516714	intergenic	0.9591	-0.233	0	n/a	0.9248	-0.717
rs76304348	intergenic	0.0383	0.803	0	n/a	0.0041	-0.493
rs62516738	intergenic	0.0383	0.976	0	n/a	0.0041	-0.077
rs62516743	intronic	0.0396	1.388	0	n/a	0.0061	0.997
rs62516744	intronic	0.0383	1.292	0	n/a	0.0447	1.339
rs62516747	intronic	0.9617	-0.688	0.9615	0.133	0.8943	**-2.155**
rs62516749	intronic	0.9617	-0.686	0	n/a	0.9024	-1.626
rs80157946	intronic	0.0383	1.362	0	n/a	0.0996	**2.651**
rs74675009	intronic	0.9617	-0.701	0	n/a	0.9004	-1.696
rs62516750	intronic	0.9617	-0.893	0	n/a	0.9004	-1.822
rs62516752	intronic	0.0383	1.573	0	n/a	0.0996	**2.691**
rs62516754	intronic	0.9617	-0.893	0	n/a	0.9004	-1.731
rs62516757	intronic	0.9617	-0.996	0	n/a	0.9004	-1.741
rs62516758	intronic	0.9617	-0.996	0	n/a	0.9004	-1.741
rs62516759	intronic	0.9617	-0.996	0	n/a	0.9004	-1.741
rs62518160	intronic	0.9617	-0.996	0	n/a	0.9004	-1.741
rs62518180	intronic	0.9617	-1.016	0	n/a	0.9004	-1.75
rs62518181	intronic	0.9617	-1.016	0	n/a	0.9004	-1.75
rs62518182	intronic	0.9617	-1.016	0	n/a	0.9004	-1.75
rs62518183	intronic	0.9617	-1.016	0	n/a	0.9146	-1.788
rs62518184	intronic	0.9644	-1.044	0	n/a	0.9085	-1.759
rs76121545	intronic	0.0383	**2.729**	0	n/a	0.0061	**2.138**
rs79103618	intronic	0.0383	**2.729**	0	n/a	0.1016	**2.934**

rs77200520	intronic	0.9617	-1.943	0	n/a	0.9085	-1.719
rs74931311	intronic	0.9617	-1.943	0	n/a	0.9085	-1.719
rs62518192	intronic	0.9617	-1.943	0	n/a	0.9004	-1.774
rs62518193	intronic	0.0383	**2.729**	0	n/a	0.0061	**2.17**
rs62518194	intronic	0.0383	**2.729**	0	n/a	0.0996	**2.698**
rs62518196	intronic	0.0383	2.733	0	n/a	0.0061	2.093
rs62518201	intronic	0.9617	-1.658	0	n/a	0.9045	**-2.623**
rs62518202	intronic	0.9617	-1.657	0	n/a	0.9004	**-2.705**
rs62518203	intronic	0.0383	**2.414**	0	n/a	0.0996	**3.798**
rs62518204	intronic	0.9617	-1.657	0	n/a	0.9004	**-2.705**
rs7008048	intronic	0.0383	**2.414**	0	n/a	0.0996	**3.461**
rs62518206	intronic	0.9617	-1.657	0	n/a	0.9004	**-2.449**
rs62518210	intronic	0.0383	**2.398**	0	n/a	0.0996	**3.529**
rs62518211	intronic	0.0383	**2.398**	0	n/a	0.0996	**3.529**
rs62518212	intronic	0.0383	**2.398**	0	n/a	0.0996	**3.529**
rs62518213	intronic	0.0383	1.77	0.0035	-0.416	0.0061	1.096
rs77930111	intronic	0.9617	-1.058	0	n/a	0.9004	**-2.119**
rs59569717	intronic	0.9617	-1.058	0	n/a	0.8963	**-2.244**
rs58205452	intronic	0.0383	1.755	0	n/a	0.0996	**3.124**
rs7832104	intronic	0.0383	1.721	0	n/a	0.0996	**3.338**
rs118073227	intronic	0.0383	1.722	0	n/a	0.0061	1.093
rs57499403	intronic	0.9617	-1.002	0	n/a	0.9024	-1.979
rs6997994	intronic	0.9617	-1.002	0	n/a	0.9004	-1.911
rs7002613	intronic	0.9617	-1.002	0	n/a	0.9004	**-2.044**
rs7002907	intronic	0.9644	-0.822	0	n/a	0.9045	-1.839
rs77259784	intronic	0.0356	1.495	0	n/a	0.0996	**2.927**
rs55817013	intronic	0.9617	-0.754	0	n/a	0.9004	-1.573
rs4952	intronic	0.0383	1.393	0	n/a	0.0061	0.953

iHS values for all SNPs in the LD bin (r^2=0.9) on chromosome 8 tagged by rs9298626 and associated with increased risk of cocaine dependence. Significant values are bolded and highlighted in yellow.